Tableau Prep 帮助

上次更新 2025/2/10

©2024 Salesforce, Inc.

目录

Tableau Prep 中的新功能 1
相关资源 1
Tableau Prep Builder 入门指南 3
示例文件 3
背景故事... 4
1.连接到数据 4
检查您的工作:观看“连接到数据”的实际运行效果。 8
2.浏览数据 8
3.清理数据 10
清理 Orders_Central 10
查看您的更改 16
检查您的工作:观看“清理 Orders_Central”的实际运行效果。 17
清理 Orders_East 17
清理 Orders_West 19
4.合并数据 23
合并数据 23
检查您的工作:观看“合并数据”的实际运行效果。 28
清理产品退货数据 28
联接数据 33
清理联接结果 37
5.运行流程并生成输出 40
总结和资源 43
关于 Tableau Prep 44
使用 Tableau Prep 44
观看视频:查看 Tableau Prep Builder 的实际运行效果 45

46
Tableau Prep 工作区一览 46
“连接”窗格 (1) 47
“流程”窗格 (2) 48
“配置”窗格 (3) 49
“数据”网格 (4) 51
Tableau Prep 如何存储数据 51
Web 上的 Tableau Prep 52
安装和部署 52
示例数据和处理限制 53
Web 上的可用功能 53
自动保存并使用草稿 55
在 Web 上发布流程 55
嵌入凭据 55
发布流程 57
谁可以执行此操作 57
Tableau Prep 视觉词典 58
启动或打开数据流程 62
启动新流程 62
打开现有流程 66
在 Tableau Prep Builder 中打开流程 66
在 Web 版 Tableau Prep 中打开流程 67
连接到数据 69
通过适用于常见数据类型的内置连接器进行连接 69
使用内置连接器时的注意事项 70
Tableau Prep Builder 70
Web 上的 Tableau Prep 70

配置 SSL 以连接到 Google BigQuery(仅限 MacOS) 73
设置和管理您的 Google BigQuery 凭据 74
使用服务帐户 (JSON) 文件登录 74
使用 OAuth 登录 74
支持的清理操作 77
连接之前 77
连接到空间文件 78
连接之前 79
使用“其他数据库 (ODBC)”进行连接 81
使用自定义连接器进行连接 84
使用合作伙伴构建的连接器 84
连接到发布的数据源 85
关于凭据和权限： 86
在流程中使用已发布数据源 86
连接到虚拟连接 91
连接到虚拟连接时的注意事项： 91
连接到 Tableau 数据提取 93
通过 Tableau Catalog 连接到数据 94
其他连接选项 94
使用自定义 SQL 连接到数据 94
使用初始 SQL 查询连接 96
运行初始 SQL 97
在初始 SQL 语句中包括参数 97
配置数据集 98
包括数据集中的行号 99
将源行号字段添加到您的流程 100
源行号详细信息 100

设置标题和数据起始行 101
配置标题和起始行 102
单个文件中的多个架构 104
合并多个表 105
连接到自定义 SQL 查询 106
在输入步骤中应用清理操作 107
选择要包括在流程中的字段 107
将筛选器应用于输入步骤中的字段 108
应用计算筛选器 108
应用相对日期筛选器 110
更改字段名称 111
更改数据类型 112
配置字段属性 114
配置文本文件中的文本设置 114
选择数据样本大小 114
为采样准备数据 115
更改数据样本设置 115
示例 117
在“输入”步骤中添加更多数据 119
刷新输入步骤数据或更改您的连接 120
刷新数据源 120
替换您的数据源 120
编辑连接 120
替换输入连接 121
在输入步骤中合并文件和数据库表 123
合并文件 123
核心筛选条件 124

额外的筛选器 124
创建输入并集 127
并集数据库表 131
在并集之后合并字段 133
在“输入”步骤中联接数据 133
构建和组织流程 136
添加或插入步骤 136
添加步骤 136
插入步骤 139
对步骤进行分组 142
对步骤进行分组的要求 142
创建组 143
更改流程配色方案 145
从流程中移除步骤 145
为流程步骤和清理动作添加说明 146
为流程步骤添加说明 146
为更改条目添加说明 147
重新组织流程的布局 150
使用流程导航工具 151
检查您的数据 153
查看分配给数据的数据类型 153
查看有关数据的大小详细信息 154
查看值或唯一值的分布 156
搜索字段和值 158
复制数据网格中的字段值 160
对值和字段进行排序 161
对字段进行重新排序 161

突出显示流程中的字段和值 163
在流程中跟踪字段 163
查看相关值 164
突出显示相同的值 164
筛选您的数据 165
保留或移除字段 165
隐藏字段 166
隐藏和取消隐藏字段 167
可用于每种数据类型的筛选器 169
筛选器选项位于何处? 169
计算筛选器 170
“选定值”筛选器 170
“值范围”筛选器 171
“日期范围”筛选器 171
“相对日期”筛选器 172
“通配符匹配”筛选器 172
“Null 值”筛选器 173
移除重复行 173
识别并移除重复项 174
样本数据中的重复项 177
计算字段的排序顺序 177
编写您自己的筛选器计算来查找并移除重复项 177
使用数据角色验证数据 178
为数据分配标准数据角色 179
创建自定义数据角色 181
要求 181
创建自定义数据角色 182

应用自定义数据角色 186
查看和管理自定义数据角色 188
按数据角色对类似值进行分组 189
在流程中创建和使用参数 192
我可以在何处应用参数? 193
创建用户参数 194
更改用户参数默认值 197
编辑用户参数 197
重置用户参数默认值 198
将参数应用于您的流程 200
将参数应用于输入步骤 200
文件名或文件路径 200
数据库表 201
自定义 SQL 202
将用户参数应用于输出步骤 202
文件名或文件路径 203
已发布数据源名称 203
数据库表和自定义 SQL 前后 204
将系统参数应用于输出步骤 205
文件名 205
已发布数据源名称 206
将用户参数应用于筛选器计算 206
将用户参数应用于计算字段 207
删除用户参数 208
运行包含参数的流程 210
手动运行流程 211
按计划运行流程 212

清理和调整数据 213
关于清理操作 213
可用的清理操作 213
操作顺序 215
应用清理操作 217
选择视图 218
暂停数据更新以提高性能 221
应用清理操作 222
批量重命名字段 225
查看您所做的更改 227
合并字段 229
使用建议应用清理操作 231
应用建议 232
编辑字段值 234
编辑单个值 234
编辑多个值 235
使用快速清理操作编辑多个值 235
以内联方式编辑分组或多个值 237
将一个或多个值替换为 Null 237
将多个值手动映射到标准值 238
将多个值映射到单个所选字段 238
通过选择多个值来创建组 239
添加和标识不在数据集中的值 240
使用模糊匹配将值自动映射到标准值 243
使用模糊匹配分组类似值 244
在对字段值进行分组时调整结果 245
复制步骤、操作和字段 247

复制并粘贴步骤 247
复制并粘贴清理操作 249
复制字段 251
创建可重用的流程步骤 253
创建可重用的步骤 253
在流程中插入可重用的步骤 254
填补顺序数据中的空白 256
生成新行 257
创建详细级别、排名和分片计算 259
计算详细级别 259
创建详细级别 (LOD) 计算 260
计算编辑器 260
可视化计算编辑器 262
计算排名或行号 263
支持的分析函数 264
创建排名或行号计算 267
计算编辑器 267
可视化计算编辑器 271
计算分片 274
创建分片计算 276
可视化计算编辑器 276
计算编辑器 279
计算多行的值 280
计算差异 281
可视化计算编辑器 281
计算编辑器 284
计算百分比差异 286

可视化计算编辑器 286
计算编辑器 289
计算移动平均值或总和 291
可视化计算编辑器 291
计算编辑器 294
获取上一个值 296
填充计算 298
Prep 可视化计算编辑器 298
计算编辑器 300
示例 301
运行计算 303
Prep 可视化计算编辑器 304
计算编辑器 306
使用 Tableau Agent 创建计算字段 307
创建计算字段 308
获得最佳结果的技巧 310
帮助 Tableau Agent 改进 310
示例 311
计算示例 1: 311
计算示例 2: 312
转置您的数据 313
将列转置为行 314
观看“对多个字段进行转置”的实际运行效果。 317
使用通配符搜索进行转置 317
将行转置为列 319
在流程中使用 R 和 Python 脚本 322
在您的流程中使用 R (Rserve) 脚本 323

先决条件 323
资源 323
针对 Tableau Server 配置 Rserve Server 323
其他 Rserve 配置(可选) 324
创建 R 脚本 325
连接到您的 Rserve 服务器 326
向流程中添加脚本 327
在您的流程中使用 Python 脚本 330
先决条件 331
为 Tableau Server 配置 Tableau Python (TabPy) 服务器 331
创建 Python 脚本 332
连接到 Tableau Python (TabPy) 服务器 334
向流程中添加脚本 335
聚合、联接或合并数据 339
对值进行聚合和分组 339
联接数据 340
检查联接的结果 343
常见联接问题 344
修复不匹配的字段以及更多操作 345
合并数据 346
检查并集的结果 347
修复不匹配的字段 349
其他合并字段选项 351
将 Einstein Discovery 预测添加到您的流程 352
什么是 Einstein Discovery? 352
先决条件 353
Salesforce 要求 353

Tableau Prep 要求 354
将预测数据添加到流程 354
查看您的结果 359
保存和共享工作 361
保存您的流程 361
在 Web 上自动保存流程 362
自动文件恢复 363
在 Tableau Desktop 中查看流程输出 363
创建数据提取文件和已发布数据源 364
Tableau Prep Builder 364
Tableau Prep Builder 以及在 Web 上 365
在流程输出中包含参数 365
创建数据提取到文件 365
创建数据提取到 Microsoft Excel 工作表 366
创建发布的数据源 368
将流程输出数据保存到外部数据库 370
输出选项 370
其他选项 371
支持的数据库和数据库要求 371
将流程数据保存到数据库 372
将流程输出数据保存到 CRM Analytics 中的数据集 376
先决条件 376
Salesforce 要求 376
Tableau Prep 要求 377
将流程数据保存到 CRM Analytics 377
将流程输出数据保存到 Data Cloud 379
权限先决条件 380

将流程数据保存到 Data Cloud 380
注意事项 383
Data Cloud 设置先决条件 383
设置摄取 API 连接器 384
创建数据流 384
将数据流添加到数据空间 385
将数据湖对象映射到 Salesforce 对象 385
为 Data Cloud 摄取 API 创建已连接应用 386
架构要求 386
YMAL文件示例 387
将流程输出数据保存到 Amazon S3 388
权限 389
将流程数据保存到 Amazon S3 389
使用增量刷新来刷新流程数据 390
流程刷新选项 391
配置增量刷新 391
使用 Append 进行增量刷新 394
配置写入选项 394
运行流程 396
通过命令行刷新流程输出文件 397
运行流程之前 398
凭据 .json 文件要求 398
版本 2020.3.1 及更高版本 399
运行包含参数值的流程 400
示例 401
连接到服务器连接 401
连接到服务器连接和数据库连接输出 402

流程包括 Rserve 和 TabPy 脚本连接并输出到数据库连接 402
连接并输出到不同的数据库连接 403
版本 2020.2.3 及更低版本 404
示例 405
连接到已发布数据源 405
连接到两个数据库 406
流程包括 Rserve 和 TabPy 的脚本步骤，并连接到数据库 406
创建凭据文件的相关提示 407
运行流程 408
在启用增量刷新的情况下运行流程 410
命令选项 410
语法示例 413
流程连接到并发布至本地文件 413
流程连接并发布到本地文件，并为增量刷新使用短形式 413
流程连接到数据库并发布至服务器 414
流程发布至服务器，并且凭据文件存储在网络共享上 414
与 Tableau Prep 的版本兼容性 416
版本号格式 416
查找您的版本 417
Tableau Prep Builder 不同版本之间的兼容性 419
修复 Tableau Prep Builder 的兼容性问题 419
不同版本的 Tableau Prep Builder 和 Tableau Server 之间的兼容性 420
检测不兼容的功能 420
Tableau Prep Builder（版本 2020.1.1 及更高版本） 420
Tableau Prep Builder（版本 2019.3.1 及更高版本） 422
Tableau Prep Builder（所有版本） 422
修复兼容性问题 424

确定不兼容的功能 424
从流程中移除不兼容的功能 426
不兼容的数据源 426
不兼容的功能 426
使流程数据保持最新 428
运行流程 430
流程运行选项 431
手动运行流程 431
将流程发布到 Tableau Server 或 Tableau Cloud 433
发布之前 433
从 Tableau Prep Builder 中发布流程 438
Tableau Server 438
文件 439
数据库 442
Tableau Cloud 446
文件 447
数据库 447
谁可以执行此操作 452
日常生活场景 453
使用 Tableau Prep 分析医院床位使用情况 453
数据 454
初步分析 454
所需的数据结构 455
重构数据 457
床位小时矩阵 457
患者床位使用 459
Tableau Desktop 中的分析 463

总结和资源 466
使用 Tableau Prep 查找第二个日期 467
数据 468
所需的数据结构 468
重构数据 469
第一次违章日期的初始聚合 469
第二次违章日期的第二个聚合 473
为第一次和第二次违章创建完整数据集 475
创建完整数据集 477
总结 478
继续阅读Tableau Desktop 中包含第二个日期的分析 on page 1。 479
Tableau Desktop 中包含第二个日期的分析 479
Tableau Desktop 中的分析 480
更进一步 — 转置的数据 487
转置数据的优点 496
再进一步 — 只使用计算 497
对各种方法的反思 502
Driver Infractions(司机违章) 503
Pivoted Driver Infractions(转置的司机违章) 503
LOD Driver Infractions(LOD 司机违章) 504
Tableau Prep Builder 疑难解答 507
运行 LogShark 507
使用命令行运行流程时的常见错误 507
错误 : “These features were found that prevent this version of the application from using this file”(发现了使此版本的应用程序无法使用此文件的这些功能) 511
使用 Tableau Prep 登录到启用 SSL 的 Tableau Server 时出现“You are using Server version: null...”(您正使用 Server 版本 : null...) 错误 512

维护 Tableau Desktop 和 Tableau Prep 的许可证 512
查看有关许可证的数据 513
使用零停机许可自动刷新产品密钥 514
跟踪 Tableau Desktop 许可证使用情况和过期数据 514
其他资源 515
Tableau Prep 函数参考 517
数字函数 517
字符串函数 521
聚合函数 525
类型转换函数 527
日期函数 528
逻辑函数 531
分析函数 531
关键字 534

Tableau Prep 中的新功能

使用下面的可视化项探索 Tableau Prep 中的新功能。单击某个功能以显示工具提示，其中包含指向该功能详细文档的链接。探索筛选器以优化您的搜索。下载数据以创建自定义列表。

- 使用**“按功能搜索”**仪表板查看产品或版本的新功能列表，或探索功能何时发布。The dashboard currently defaults to Tableau Prep as the product (which includes Prep Builder and Prep Conductor features) for the version Tableau Prep Builder.
- 使用**“升级 Prep”**仪表板查看特定于升级的功能列表。如果您将流程发布到 Tableau Server 以按计划运行，某些新功能需要最低 Tableau Server 版本才能运行。该视图列出了支持计划在特定版本的 Tableau Prep Builder 中创建的流程的最低 Tableau Server 版本，以帮助您快速发现具有兼容性要求的功能。

相关资源

新功能

浏览新功能摘要来了解当前支持的版本。

所有已知问题 | 下载

Tableau Prep Builder 入门指南

注意：从版本 2020.4.1 开始，作为 Creator，您还可以在 Web 上创建和编辑流程。本教程是使用 Tableau Prep Builder 设计的，但也可以在 Web 上完成，不过有一些值得注意的例外。

本教程介绍 Tableau Prep 中的常见操作。您将使用 Tableau Prep 附带的样本数据集来演练为示例 Superstore 创建流程。本教程使用最新版本的 Tableau Prep Builder。如果您使用的是以前的版本，结果可能会不同。

请在进行演练的过程中观看提示，深入了解 Tableau Prep 如何帮助您清理和调整数据以进行分析。

若要在继续本教程之前安装 Tableau Prep Builder，请参见 Tableau Desktop 和 Tableau Prep Builder 部署指南中的通过用户界面安装 Tableau Desktop 或 Tableau Prep Builder。否则您可以下载免费试用版。

示例文件

若要完成本教程中的任务，您需要安装 Tableau Prep Builder，或者如果在服务器版本 2020.4 或更高版本上启用了 Web 制作，还可以尝试 Web 上的步骤。

在计算机上安装 Tableau Prep Builder 后，您还可以在以下位置找到示例文件：

- (Windows) `C:\Program Files\Tableau\Tableau Prep Builder <version>\help\Samples\en_US\Superstore Files`
- (Mac) `/Applications/Tableau Prep Builder <version>.app/Contents/help/Samples/en_US/Superstore Files`

或者，从这些链接下载示例文件，并创建一个 Samples 目录和一个 South 子目录。如果在 Web 上完成本教程，则需要这样做。

下载到 Samples 目录

- Orders_Central
- Orders_East
- Orders_West
- returns_reasons_new

下载到 South 子目录

- Orders_South_2015
- Orders_South_2016
- Orders_South_2017
- Orders_South_2018

背景故事...

您在一家大型零售连锁店的总部工作。您的老板想要分析过去四年公司的产品销售额和利润。您推荐他使用 Tableau Desktop 来完成该操作。您的老板认为这个主意很棒，并希望您立即着手完成该工作。

在开始收集需要的所有数据时，您注意到，有人已经通过其他方式为每个区域收集和跟踪过数据。您还注意到不同文件中有大量创新数据条目，而且一个区域甚至每年都有单独的文件！

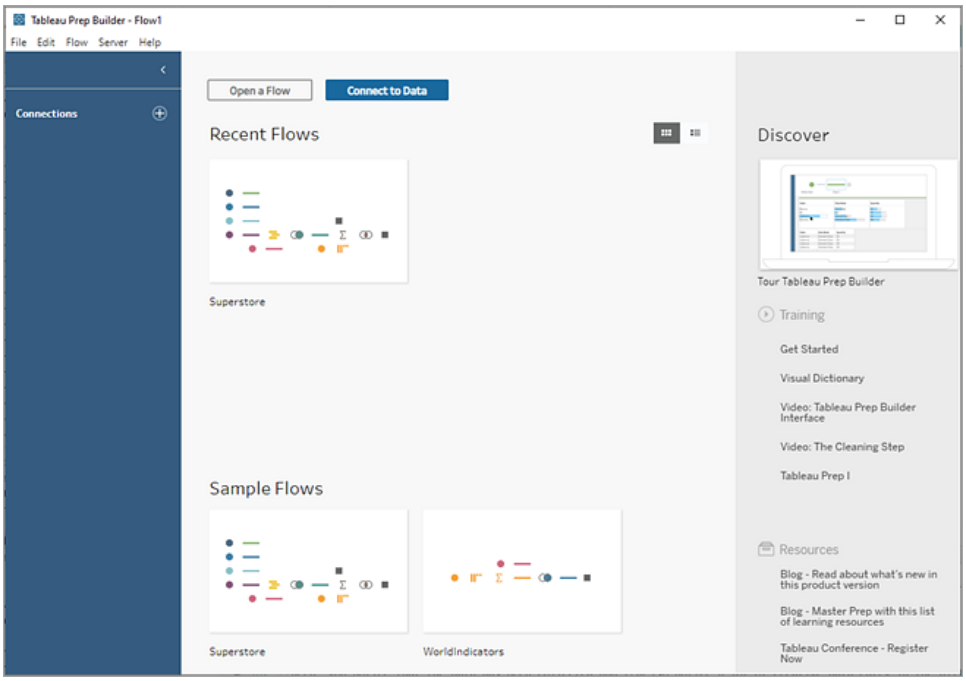
您将不得不先执行大量的数据清理工作，然后才能开始在 Tableau 中分析数据，而这项工作将需要很长时间。

在您翻找餐厅的菜单来订晚餐时，您想起来 Tableau 有一款名为 Tableau Prep 的新产品，该产品可帮助您完成艰巨的数据清理任务。

您下载产品，并注册了免费试用版，决定试用一下。

1.连接到数据

打开 Tableau Prep Builder 时，您首先看到是一个带有**“连接”**窗格的开始页面，就像 Tableau Desktop 一样。



若要开始操作，第一步是连接到数据并创建输入步骤。您将从该处开始创建工作流（在 Tableau Prep 中称之为“流程”），并随着进展添加更多步骤来对数据执行操作。

提示：输入步骤是数据的摄入点和流程的开始点。您可能有多个输入步骤，并且某些步骤可能包括多个数据文件。有关连接到数据的详细信息，请参见**连接到数据** 在本页69。

不同区域的销售额数据文件以不同的格式存储，并且“南部”区域的订单实际上是多个文件。您检查了**“连接”**窗格，并发现您有很多选择来连接到数据。太好了！

由于其他区域全部四年的数据是一个文件，因此您决定先 着手处理“南部”区域的文件。

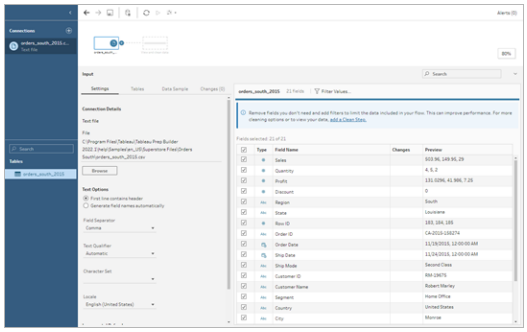
1. 在**“连接”**窗格上，单击**“添加连接”** 按钮。

   在 Web 制作中，从**“主页”**中单击**“创建”**>**“流程”**，或从**“浏览”**页面中单击**“新建”**>**“流程”**。然后单击**“连接到数据”**。

2. 文件是 .csv 文件，因此请在连接列表中选择**“文本文件”**。

3. 导航到您的文件的目录。在 Orders South 子目录中，选择第一个文件 **orders_south_2015.csv**，并单击**“打开”**以将其添加到流程。（有关文件位置，请参见**总结和资源** 在本页43。）

   连接到第一个文件后，Tableau Prep Builder 工作区将打开，您会看到它分为两个主要部分。顶部的“**流程**”窗格和底部的“**输入**”窗格。

   与 Tableau Desktop 很像，“**流程**”窗格是您的工作区，您可以在其中直观地与数据交互并构建流程。“**输入**”窗格包含有关如何摄入数据的配置选项。它还显示数据集中的字段、数据类型以及值的示例。

   我们将在下一部分中探讨您如何才能与此数据交互。

   提示：对于单个表，Tableau Prep 会在您将数据添加到流程时自动为您在**“流程”**窗格中创建输入步骤。否则，您可以使用拖放功能将表添加到**“流程”**窗格。

4. 在南方，您还有另外三个订单文件，如何组合它们取决于您在哪里工作。

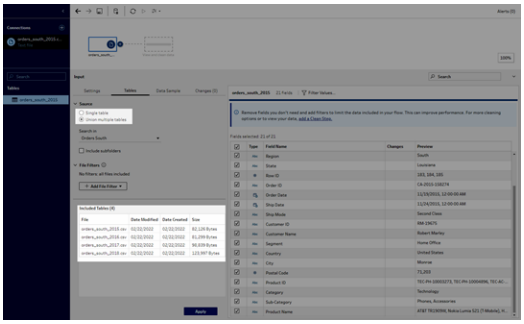
在 Tableau Prep Builder 中：

a. 您可以单独添加每个文件，但您想要将所有文件合并到一个输入步骤中，因此您在“**输入**”窗格中单击**“表”**选项卡。

b. 您会看到一个**“合并多个表”**选项。选择该选项。

您注意到，您在其中选择文件的目录已填充，并且您需要的其他文件列在“输入”窗格的**“已包括文件”**部分中。

提示：使用通配符并集可以非常方便的连接到单一数据源中具有类似名称和结构的多个文件并将其合并。若要使用此选项，文件必须位于同一父目录或子目录中。如果没有立即看到所需的文件，请更改搜索条件。有关详细信息，请参见**在输入步骤中合并文件和数据库表** 在本页 123。



c. 单击**“应用”**以将这些文件中的数据添加到**“orders_south_2015”**输入步骤。

d. 其他区域的文件全部都是单一表文件，因此您可以一次性选择所有文件并将它们添加到流程。

> **注意：**在 Web 上，文件只能单独上载。

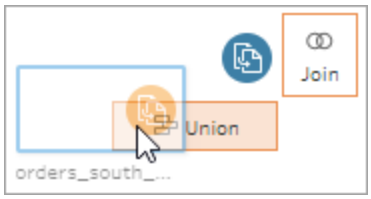
在 Tableau Server 或 Tableau Cloud 中：

通配符选项目前对于 Tableau Server 或 Tableau Cloud 不可用。不过，您还是希望包括来自南方的所有文件并同样处理数据，因此将它们组合在一起是有意义的。

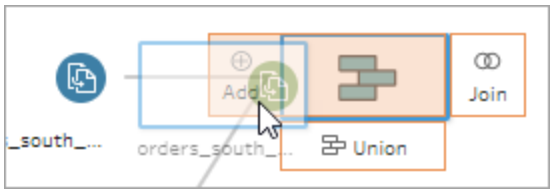
a. 重复步骤 2 和 3 以添加 Orders South 子目录中的其他文件。

b. 将它们与合并步骤结合。(有关更多详细信息，请参见**在输入步骤中合并文件和数据库表** 在本页 123。)

i. 将**“Orders_South_2016”**拖到**“Orders_South_2015”**上，并将其放在**“并集”**选项上。



ii. 将**“Orders_South_2017”**拖到新的合并步骤上，并将其放在**“添加”**上。对最终文件重复此步骤。

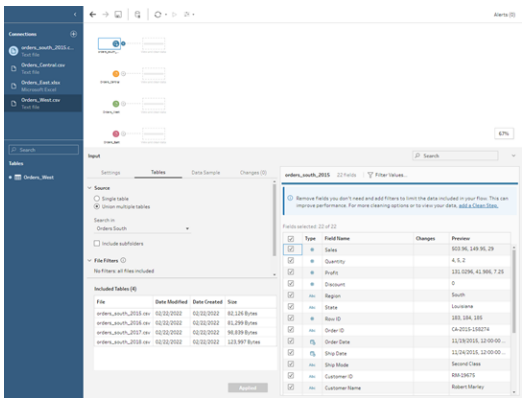


5. 添加其余文件。

在 Tableau Prep Builder 中：

- 打开文件资源管理器或 Finder，并导航到文件的目录。按住 Ctrl 单击或按住 Cmd 单击 (MacOS)，选择以下文件并将它们拖放到“**流程**”窗格上，将这些文件添加到流程。(有关文件位置，请参见**总结和资源** 在本页43。)
  - Orders_Central.csv
  - Orders_East.xlsx
  - Orders_West.csv

> **注意：**这些是不同的文件类型。如果您看不到所有这些文件，请确保您的文件资源管理器或 Finder 设置为查看所有文件类型。

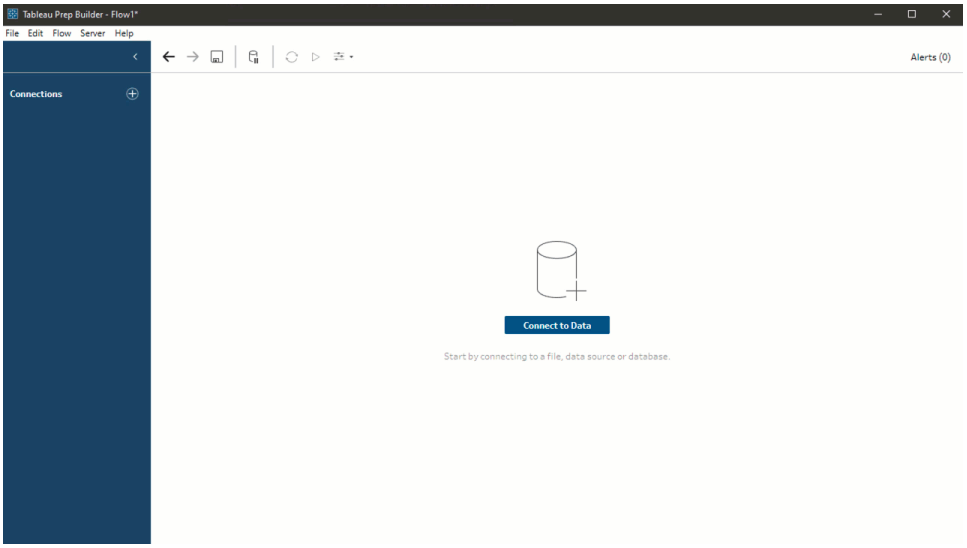


在 Tableau Server 或 Tableau Cloud 中：

- 按照步骤 2 和 3 添加 Orders_Central.csv 和 Orders_West.csv。
- 在**“连接”**窗格上，单击**“添加连接”**按钮。单击 **Microsoft Excel** 并选择“**Orders_East.xlsx**”。

检查您的工作：观看“连接到数据”的实际运行效果。

单击图像可重播



2.浏览数据

既然您已将数据文件加载到 Tableau Prep，您很确定自己想要将所有文件合并在一起。但在执行该操作之前，不妨先检查一下这些文件，看看是否能找出任何问题。

在“**流程**”窗格中选择输入步骤时，您可能会看到用于引入数据的设置、所包括的字段以及值的预览。

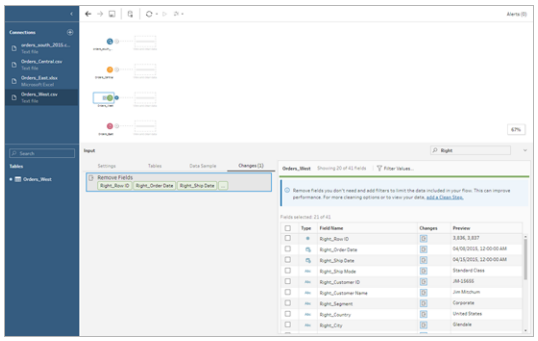
在这里可以很方便地决定要在流程中包括多少数据，以及移除或筛选不需要的字段。您还可以更改未正确分配的任何数据类型。

提示：如果在处理大型数据集，Tableau Prep 会自动引入数据样本，从而最大限度提升性能。如果看不到预期数据，您可能需要调整样本。可在**“数据样本”**选项卡上执行此操作。有关配置数据选项和样本大小的详细信息，请参见选择数据样本大小 在本页 114。

在**“流程”**窗格中，当您选择每个步骤并查看每个数据集时，您注意到需要稍后修复的一些问题，以及一个可立即在**输入**步骤中修复的问题。

- 选择“**Orders_West**”输入步骤。
  - **“State”(州)**字段使用州名称的缩写。其他文件则使用完整拼写，因此您将需要稍后修复该问题。
  - 有许多以“**Right_**”开头的字段。这些字段似乎与其他字段重复。您不想在流程中包括这些重复字段。此问题可在**“输入”**步骤中的此处修复：

    若要立即修复此问题，请清除以“**Right_**”开头的所有字段的复选框。这会指示 Tableau Prep 忽略这些字段，不将它们包括在流程中。

    提示：当您在某个步骤中执行清理操作(比如移除字段)时，Tableau Prep 会在**“更改”**窗格中跟踪您所做的更改，并在“流程”窗格中(以小图标的形式)添加注释来帮助您跟踪对数据所做的操作。对于“输入”步骤，也会向每个字段中添加注释。
- 在**“流程”**窗格中，单击“**Orders_Central**”输入步骤将其选定。在**“输入”**窗格中，您注意到以下问题：
  - 订单日期和发货日期在字段中已针对月、日和年分开。
  - 某些字段的数据类型与其他文件中相同字段的数据类型不同。
  - **“Region”(区域)**没有字段。

  您将需要对这些字段执行一些清理操作，然后才能将此文件与其他文件合并。但您无法在**“输入”**步骤中的此处修复该问题，因此记下来以便稍后进行
- 选择“**Orders_East**”输入步骤。

  此文件中的字段看起来与其他文件完全一致。但**“Sales”(销售额)**值似乎全部都包括了货币代码。您稍后也需要修复该问题。

既然您已经确定了数据集中的一些问题，那么下一步是更仔细地检查您的数据并清理找到的任何问题，以便能够合并和调整数据，并生成可用于分析的输出文件。

3.清理数据

在 Tableau Prep 中，检查和清理数据是一个反复的过程。确定了要处理的数据集之后，下一步是检查该数据，并通过对数据应用各种清理、调整和合并操作来对其进行操作。您通过向流程中添加步骤来应用这些操作。有关清理选项的详细信息，请参见**清理和调整数据** 在本页213。

根据您尝试进行的操作，步骤有多种风格。例如，在想要对字段应用像筛选、合并、拆分、重命名这样的清理操作时随时添加清理步骤。添加聚合步骤来对字段进行分组和聚合，并更改数据的详细级别。有关不同步骤类型及其用法的详细信息，请参见**构建和组织流程** 在本页136。

提示：向流程中添加步骤时，会自动添加一条流程线将步骤相互连接在一起。您可以将这些流程线四处移动，并根据需要移除或添加流程线。

运行流程时，必须要有这些连接点，以便 Tableau Prep 知道哪些步骤已连接以及步骤在流程中的应用顺序。如果缺少流程线，则流程会中断并出现错误。

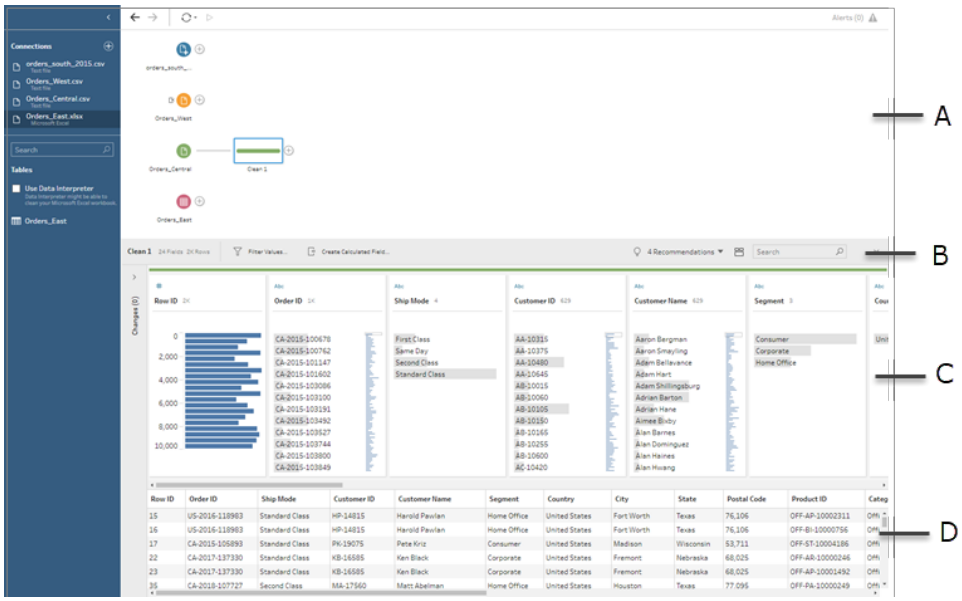
清理 Orders_Central

为了解决之前发现的问题并确定是否还有任何其他问题，您首先向“**Orders_Central**”输入步骤中添加一个清理步骤。

1. 在**“流程”**窗格中，选择**“Orders_Central”**，执行以下操作之一：

   - 单击加号 ⊕ 图标并添加清理步骤。根据您的版本，此菜单选项为**“添加步骤”**、**“添加清理步骤”**或**“清理步骤”**。
   - 单击建议的清理步骤（Tableau Prep Builder 版本 2020.3.3 及更高版本以及在 Web 上）

   向流程中添加清理步骤时，工作区将发生变化，您将看到数据的详细信息。



A. “流程”窗格、B. 工具栏、C. “配置”窗格、D. 数据网格

工作区现在分为三个部分：**“流程”**窗格、带工具栏的**“配置”**窗格和**“数据”**网格。

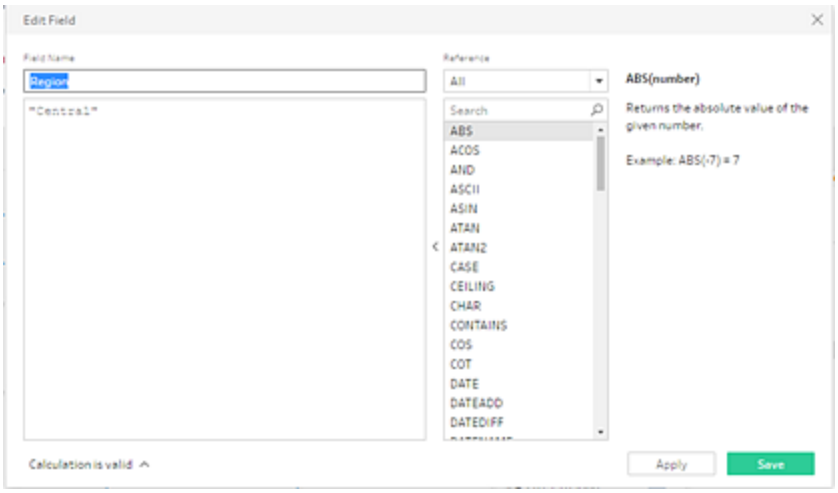
“**配置**”窗格显示数据的结构，同时将字段值汇总到数据桶中，以便您能够快速查看相关值，并找出离群值和 null 值。**“数据”**网格显示字段的行级别详细信息。

提示：“配置”窗格中的每个字段都显示在配置卡上。使用每张卡上的**“更多选项”** ··· 菜单（在以前的版本中为下拉箭头）来查看和选择可用于该字段类型的不同清理选项。您也可以对字段值进行排序、更改数据类型、为字段分配数据角色，或在**“数据”**网格中拖放配置卡和列以重新排列它们。

使用计算字段清理数据

此数据集缺少“**Region**”（区域）字段。由于其他数据集有此字段，因此您将需要添加此字段，以便能够稍后合并数据。您将需要使用计算字段来执行此操作。

2. 在工具栏中，单击**“创建计算字段”**。
3. 在计算编辑器中，将计算字段命名为**“Region”**（区域）。然后输入 **"Central"**（包括引号），并单击“**保存**”。



您能够使用计算字段来调整数据，这种灵活性让您非常喜欢。您非常高兴看到 Tableau Prep 使用与 Tableau Desktop 相同的计算编辑器语言。

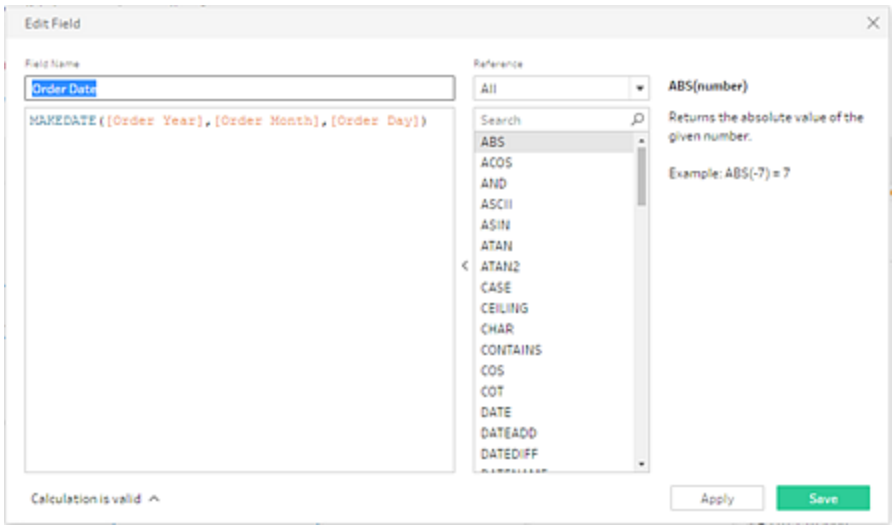
提示：当您对字段和值进行更改时，Tableau Prep 会在左侧的**“更改”**窗格跟踪所做的更改。同时，会向流程中的清理步骤和**“配置”**窗格中的字段添加一个表示更改的图标（注释）。我们将在进行更多更改之后查看**“更改”**窗格。

接下来，您需要解决单独订单日期和发货日期字段的问题。您需要将它们合并为两个单一字段，一个用于“**Order Date**”（订单日期），一个用于“**Ship Date**”（发货日期），使它们与其他数据集中的相同字段保持一致。确保表具有相同的字段将使您能够稍后使用并集来合并表。

您可以再次使用计算字段，通过一个轻松的步骤执行此操作。

4. 在工具栏中，单击**“创建计算字段”**，将**“Order Year”（订单年份）**、**“Order Month”（订单月份）**和**“Order Day”（订单日）**字段合并为一个格式为“MM/DD/YYYY”的字段。
5. 在计算编辑器中，将计算字段命名为“**Order Date**”（订单日期）。然后输入以下计算，并单击**“保存”**：

```
MAKEDATE([Order Year],[Order Month],[Order Day])
```



既然有了新的订单日期字段，那么您需要移除现有字段，因为您不再需要它们。

“配置”窗格中有很多字段。您注意到工具栏右上角的**“搜索”**框。您不知道是否能使用该框来快速查找想要移除的字段。您决定试一下。

6. 在**“配置”**窗格的搜索框中，键入 **Order**。

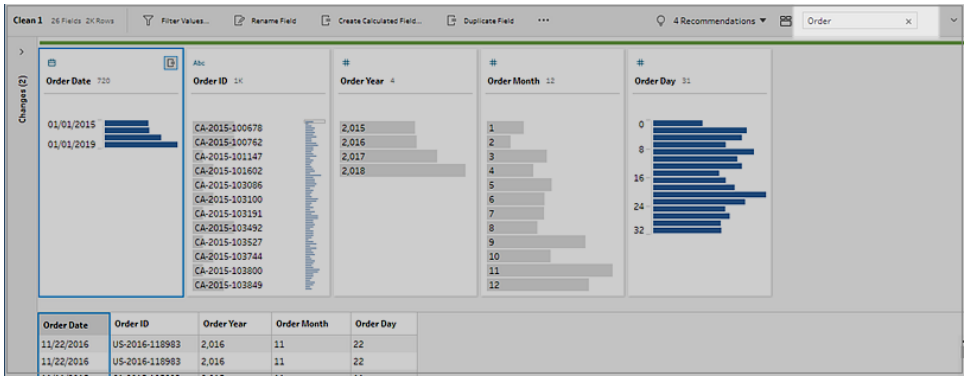
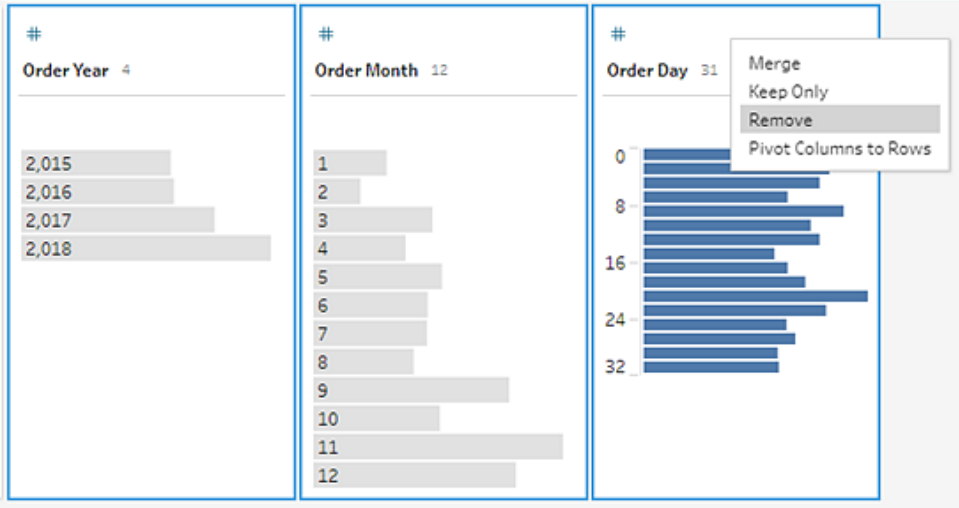


Tableau Prep 将在视图中快速滚动名称中包含**“Order”**的所有字段。太酷了！

7. 按住 Ctrl 单击或按住 Cmd 单击 (MacOS)，选择**“Order Year”(订单年份)**、**“Order Month”(订单月份)**和**“Order Day”(订单日)**字段。然后右键单击所选字段，并从菜单中选择**“移除字段”**(在以前的版本中为**“移除字段”**)来移除它们。



8. 现在重复上面的步骤 4 至 7，为**“Ship Date”(发货日期)**创建单一字段。自己尝试一下，或使用下面的步骤来帮助您。

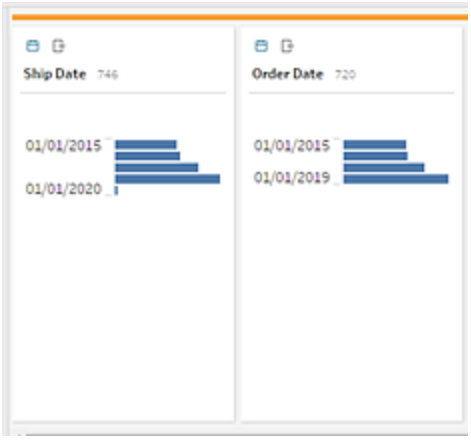
   - 在工具栏中，单击**“创建计算字段”**，将**“Ship Year”(发货年份)**、**“Ship Month”(发货月份)**和**“Ship Day”(发货日)**字段合并为一个格式为“MM/DD/YYYY”的字段。
   - 将该计算字段命名为**“Ship Date”**(发货日期)，并输入计算 `MAKEDATE([Ship Year],[Ship Month],[Ship Day])`。然后单击**“保存”**。
   - 移除**“Ship Year”(发货年份)**、**“Ship Month”(发货月份)**和**“Ship Day”(发货日)**字段。搜索并选择字段，然后从菜单中选择**“移除”**(在以前的版本中为**“移除字段”**)来移除这些字段。

   提示：Tableau Prep 会将“配置”窗格中的数据汇总到数据桶中，帮助您快速查看数据的形状、查找离群值、找出字段之间的关系，等等。

   在此方案中，订单日期和发货日期现在可按年份汇总。每个数据桶都表示从起始年份一月份到下一年一月份的一年，并带有相应标签。由于有处于 2018 年和 2019 年下半年的销售日期及发货日期，因此该数据的数据桶相应标有 2019 年和 2020 年年末。

   若要将此视图更改为实际日期，请单击“配置”卡中的**“更多选项”**··· 菜单(在以前的

版本中为下拉箭头)，并选择**“详细信息”**。



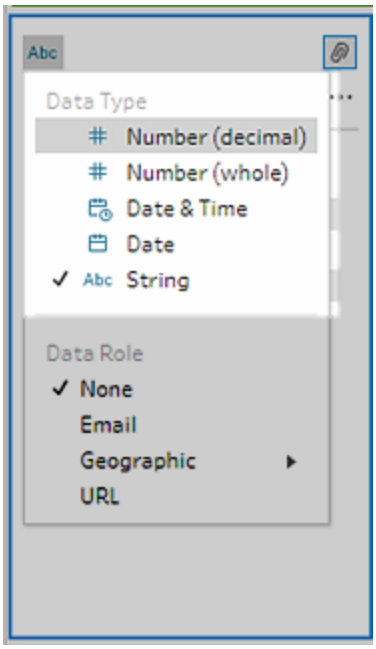
与字段直接交互来清理数据

您的数据一开始看起来很好。但在完成为订单日期和发货日期移除额外字段的操作后，您注意到**“Discounts”(折扣)**字段有一些问题。

- 系统为其分配了**“字符串”**数据类型，而不是**“数字(小数)”**数据类型。
- 对于无折扣的情况，字段值是**“None”(无)**，而不是数字值。

这将在合并文件时导致问题，因此最好也修复该问题。

9. 清除搜索，并在搜索框中输入 **disc** 来查找字段。
10. 选择**“Discounts”(折扣)**字段，双击字段值**“None”(无)**，并将其更改为数字值**“0”**。
11. 若要将**“Discount”(折扣)**字段的数据类型从**“字符串”**更改为**“数字(小数)”**，单击**“Abc”**，并从下拉菜单中选择**“数字(十进制)”**。



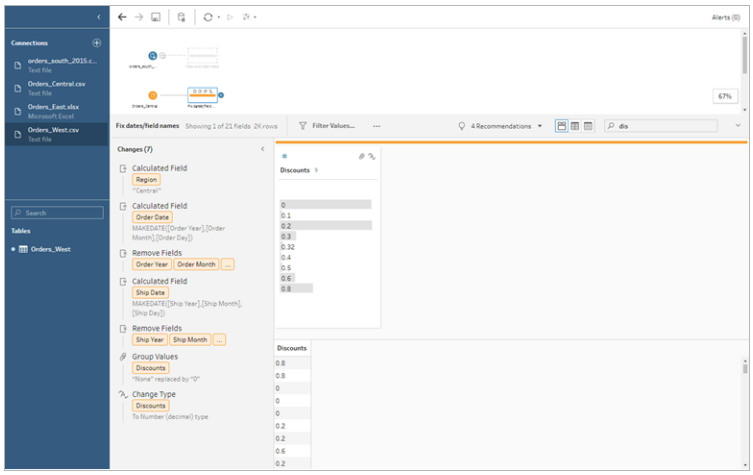
12. 最后为步骤命名，以帮助跟踪您在此步骤中进行的操作。在**“流程”**窗格中，双击步骤名称**“Clean 1”(清理 1)**，并键入**“Fix dates/field names”(修复日期/字段名称)**。

查看您的更改

您对此数据集进行了许多更改，并开始担心自己会不记得所进行的操作。在检查工作时，您看到**“配置”**窗格的左侧有一个名为**“更改”**的列。

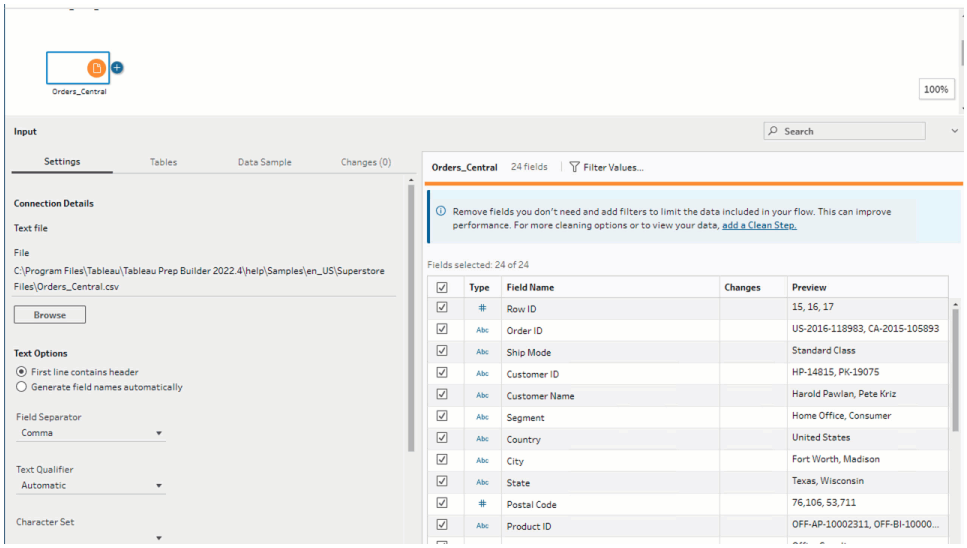
您单击箭头将其打开，看到其中列出了您刚刚进行的每项更改，这让您很高兴。在您滚动浏览列表中的更改时，您注意到，您可以删除或编辑更改，甚至能将更改四处移动来改变所做更改的顺序。

您可以轻松地查找构建流程时在任何步骤中所做的更改，并对这些更改的顺序进行试验来最大程度地利用您的数据，这让您非常喜欢。



检查您的工作：观看“清理 Orders_Central”的实际运行效果。

单击图像可重播



既然已清理了一个文件，那么您将查看其他文件来确定需要修复的其他问题。

接下来您决定查看“**Orders_East**”的 Excel 文件。

清理 Orders_East

在查看“**Orders_East**”文件的字段时，除了“**Sales**”（销售额）之外，大多数字段看起来与其他文件一致。为了进一步查看并确定是否还有任何其他要解决的问题，您向“**Orders_East**”输入步骤中添加一个清理步骤。

1. 在**“流程”**窗格中，选择“**Orders_East**”，并执行以下操作之一：

   - 单击加号 ⊕ 图标并添加清理步骤。根据您的版本，此菜单选项为**“添加步骤”**、**“添加清理步骤”**或**“清理步骤”**。
   - 单击建议的清理步骤（Tableau Prep Builder 版本 2020.3.3 及更高版本以及在 Web 上）。

   通过查看**“Sales”（销售额）**字段，您很快发现随销售额数字一起包括了 **USD** 货币代码，并且 Tableau Prep 已将这些字段值解读为字符串。

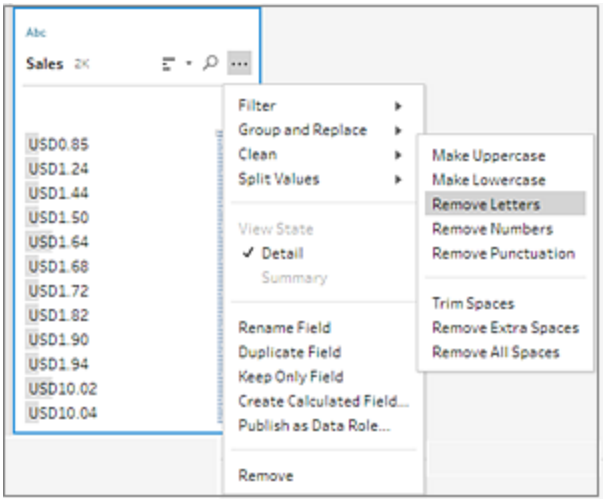
   如果要获得准确的销售额数据，您将需要从此字段中移除货币代码，并更改数据类型。

   修复数据类型很容易，您已经知道如何做了。但唯一的销售额数据行超过 2000 行，而修复每个单独的行来移除货币代码看起来很麻烦。

   但这是 Tableau Prep，因此您决定检查下拉菜单，看看是否有用于修复此问题的选项。

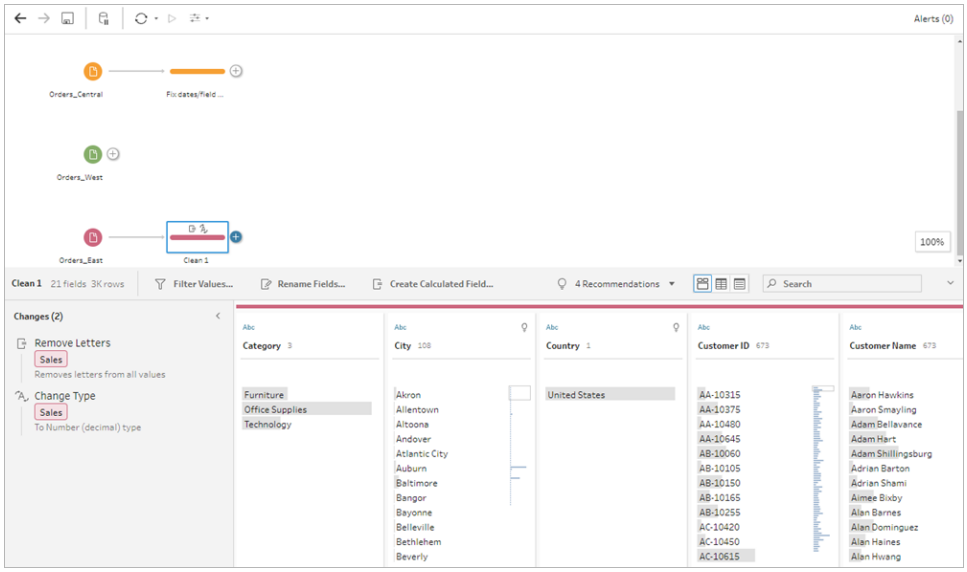
   当您单击“**Sales**”（销售额）字段的**“更多选项”** ⋯（以前版本中的下拉箭头）时，您看到一个名为**“清理”**的菜单选项，以及该选项下一个用于移除字母的选项。您决定尝试一下，看看该选项有什么用。

2. 选择**“Sales”（销售额）**字段。单击**“更多选项”** ⋯ 菜单（在以前的版本中为下拉箭头），并选择**“清理”>“移除字母”**。



哇！该清理选项立即从每个字段中移除了货币代码。现在您只需要将数据类型从**“字符串”**更改为**“数字(小数)”**，此文件看起来就正常了。

3. 单击**“Sales”(销售额)**字段的数据类型，并从下拉列表中选择**“数字(小数)”**来更改数据类型。



4. 文件的其余部分看起来很好。为清理步骤命名以跟踪您的工作。例如，命名为“**Change data type**”(更改数据类型)。

接下来您将查看 **Orders_West** 的最后一个文件，看看是否有任何需要修复的问题。

清理 Orders_West

在您查看“**Orders_West**”文件的字段时，大多数字段看起来都与其他文件一致，但您记得看到过“**States**”(州)字段为值使用了缩写，而不是州名称的完整拼写。为了将此文件与其他文件合并，您将需要修复此问题。因此您向“**Orders_West**”输入步骤中添加一个清理步骤。

1. 在**“流程”**窗格中，选择“**Orders_West**”，并执行以下操作之一：
   - 单击加号 ⊕ 图标并添加清理步骤。
   - 单击建议的清理步骤(Tableau Prep Builder 版本 2020.3.3 及更高版本以及在 Web 上)。
2. 滚动或使用“搜索”来查找“**State**”(州)字段。

您将看到所有州名称值都使用简短缩写。此字段只有 11 个唯一的值。您可以手动更改每个值，但 Tableau Prep 是否可能有另一种方法来完成此操作？

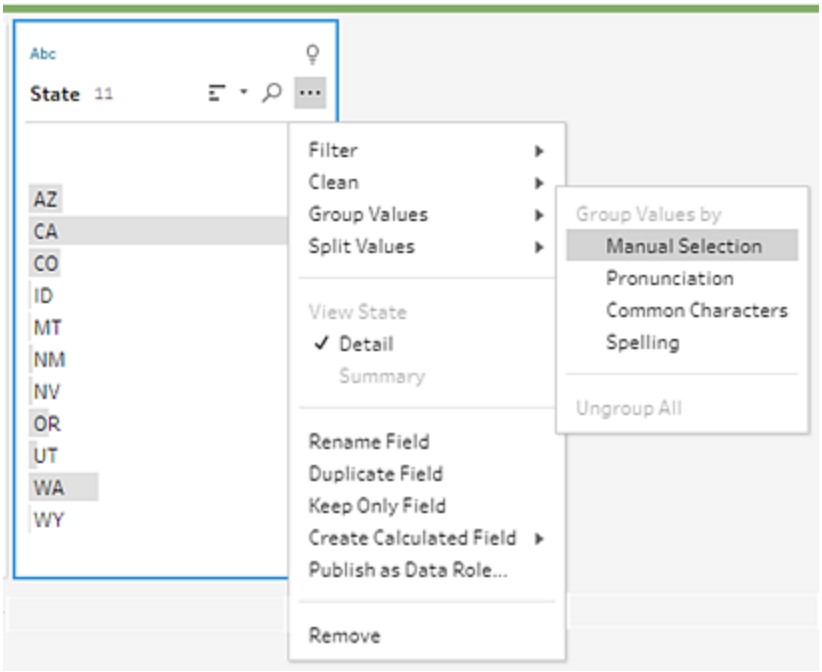
您单击字段的**“更多选项”**⋯ 菜单（以前版本中的下拉箭头），并看到一个名为**“对值进行分组”**（在以前的版本中为“分组和替换”）的选项。选择该选项时，您将看到若干选项：

- 手动选择
- 发音
- 常用字符
- 拼写

州名称的发音不相似，它们的拼写不正确，也没有共用相同的字符，因此您决定尝试**“手动选择”**选项。

提示：您可以双击字段名称或字段值来编辑单个值。若要编辑多个值，您可以选择所有值，并使用右键单击菜单选项**“编辑值”**。但如果要将一个或多个值映射到特定值，请使用下拉菜单中的**“对值进行分组”**选项。
有关对值进行编辑和分组的详细信息，请参见**编辑字段值** 在本页234。

3. 选择**“State”（州）**字段。单击下拉箭头，并选择**“对值进行分组”**（以前版本中的“分组和替换”）**>“手动选择”**。



一个两列的卡将打开。这是**“对值进行分组”编辑器**。左侧的列显示当前字段值，右侧的列显示可映射到左侧字段的字段。

您想要将州缩写映射到州名称的完整拼写版本，但在**“Orders_West”**数据集中没有这些值。您不知道是否能直接编辑名称和在该处添加名称，因此您尝试了一下。

4. 在左侧窗格的**“对值进行分组”编辑器**中，双击**“AZ”**以突出显示值，并键入 **Arizona**。然后按 **Enter** 以添加更改。

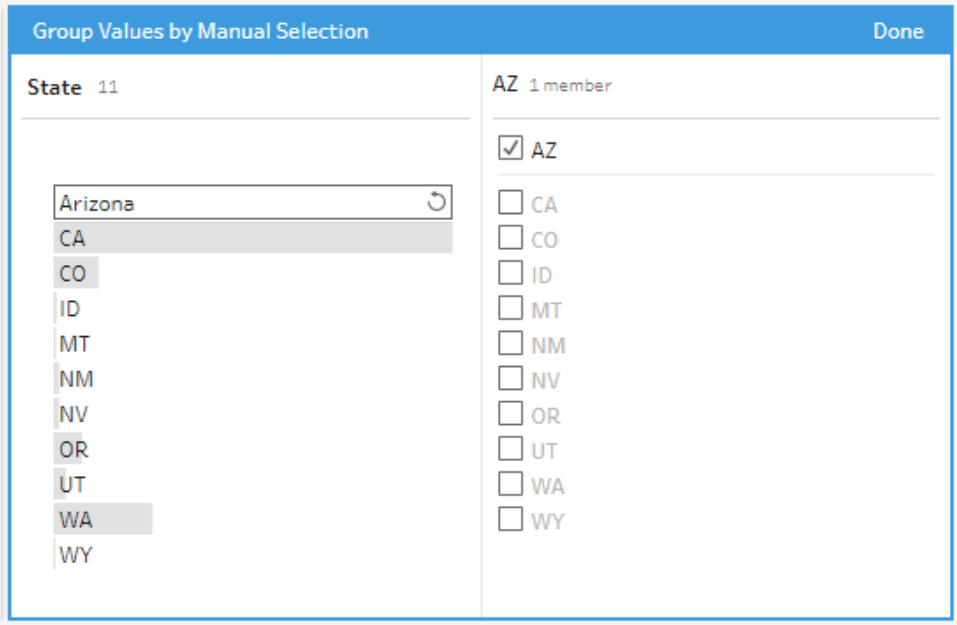


Tableau Prep 为新值**“Arizona”**创建了一个映射值，并自动将旧值**“AZ”**映射到该值。如果您像这样输入了更多此区域的数据，那么为这些值建立映射关系将节省您的时间。

提示：您可以添加数据样本中没有的字段值，建立映射关系来组织数据。如果刷新数据源并添加了新数据，您可以将新数据添加到映射，而不是手动修复每个值。

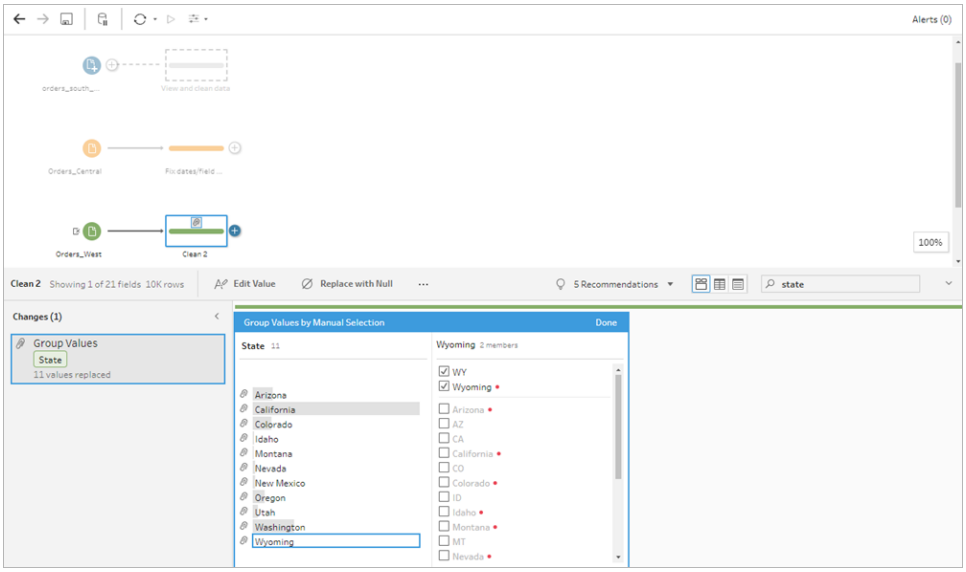
当您手动添加数据样本中没有的值时，值将带有红点标记，可帮助您轻松识别该值。

5. 重复这些步骤，将每个州都映射到其名称的完整拼写版本。

缩写　州名称

AZ	Arizona
CA	California
CO	Colorado
ID	Idaho
MT	Montana
NM	New Mexico
NV	Nevada
OR	Oregon
UT	Utah
WA	Washington
WY	Wyoming

然后，单击**“完成”**关闭**“对值进行分组”**编辑器。



映射了所有州之后，您查看**“更改”**窗格，看到其中只有一个条目，而不是 11 个条目。

Tableau Prep 已将字段的类似操作分组在一起。您喜欢该功能，因为利用它可以更轻松地查找随后对数据集所做的更改。

修复“**State**”(州)字段值是您需要在此处进行的唯一更改。

6. 为清理步骤命名以跟踪您的工作。例如“**Rename states**”(重命名州)。

您在文件中进行了大量的清理，不敢相信操作能如此快捷轻松地完成。您竟然还可以回家吃晚饭！为了确保不会丢失到目前为止所做的工作，请保存您的流程。

> **注意**：如果在 Web 上工作，更改会在您移动时自动保存，从而创建草稿流程。单击草稿标题以命名草稿。有关在 Web 上制作的详细信息，请参见 Tableau Server 或 Tableau Cloud 帮助中的**“Web 上的 Tableau Prep”**。

单击**“文件”>“保存”**或**“文件”>“另存为”**。将文件另存为流程文件 (.tfl) 并为其命名。例如“**My Superstore**”(我的超市)。

提示：在保存流程文件时，您可以将它们另存为流程文件 (.tfl) 或打包文件 (.tflx)，并将本地数据文件与这些文件打包在一起，以便将流程和文件与其他人共享。有关保存和共享流程的详细信息，请参见保存和共享工作 在本页361。

4.合并数据

既然清理了所有文件，那么您就已最终准备好将它们合并在一起。

由于清理后所有文件都有类似的字段，因此若要将所有行合并到单个表中，您需要将这些表合并。

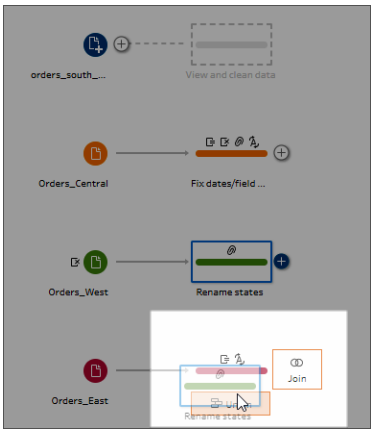
您记得有一个名为**“并集”**的步骤选项，但不知道是否能简单地通过拖放步骤来合并文件。您决定试一下看看。

合并数据

1. 针对您的工作位置执行相关步骤。

   Tableau Prep Builder

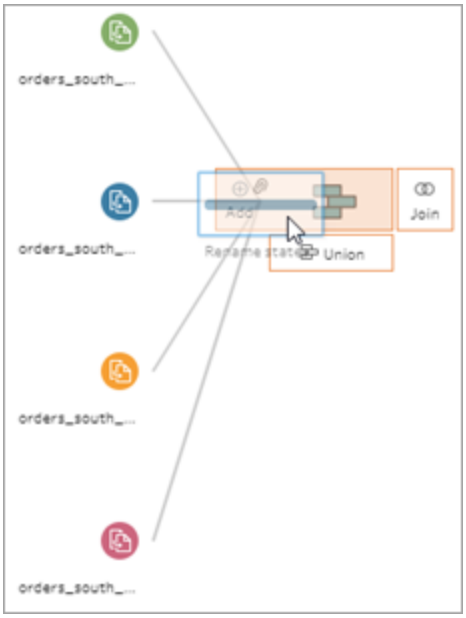
   - 在**“流程”**窗格中，将清理步骤“**Rename states**”(重命名州)拖到清理步骤“**Changed data type**”(更改的数据类型)步骤，并将其放在**“并集”**选项上。



您看到 Tableau Prep Builder 向流程中添加了一个新的**“Union”(合并)**步骤。太好了！现在，您想要向此并集中添加其他文件。

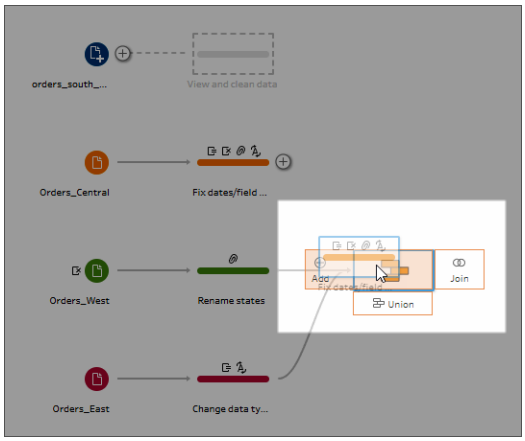
Tableau Server 或 Tableau Cloud

- 在**“流程”**窗格中，将清理步骤**“Rename states”**(重命名州)拖到之前为 South 文件创建的“并集”步骤，并将其放在**“添加”**选项上。



您会看到 Tableau Prep 将新文件添加到以前的并集中。太好了！现在，您想要向此并集中添加其他文件。

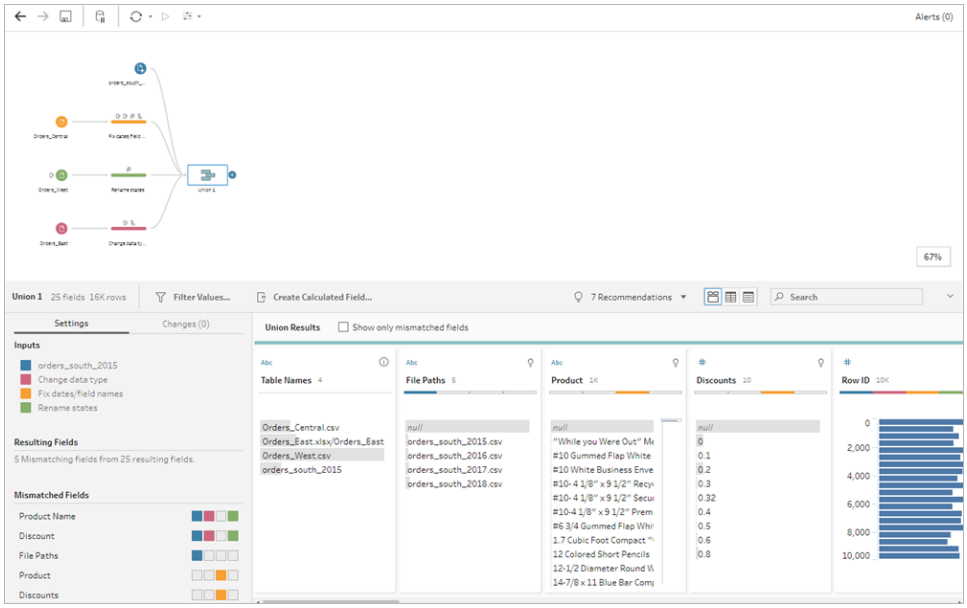
2. 将流程中的下一个清理步骤拖到新的**“并集”**步骤，然后将其放在**“添加”**上以添加到现有并集。



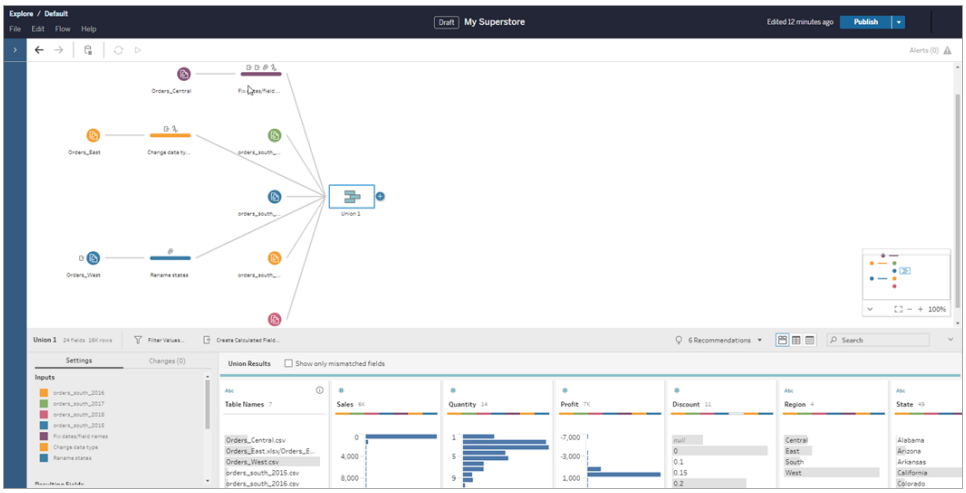
3. 将其余步骤(如果在 Tableau Prep Builder 中工作，为**“orders_south_2015”**输入步骤，如果在 Web 上工作，则为您的清理步骤)拖到新的**“并集”**步骤。将其放在**“添加”**上以添加到现有并集。

   现在，所有文件已合并到一个表中。在**“流程”**窗格中，选择新的**“Union”(合并)**步骤以查看结果。

   在 Tableau Prep Builder 上：



   在 Tableau Server 或 Tableau Cloud 上：



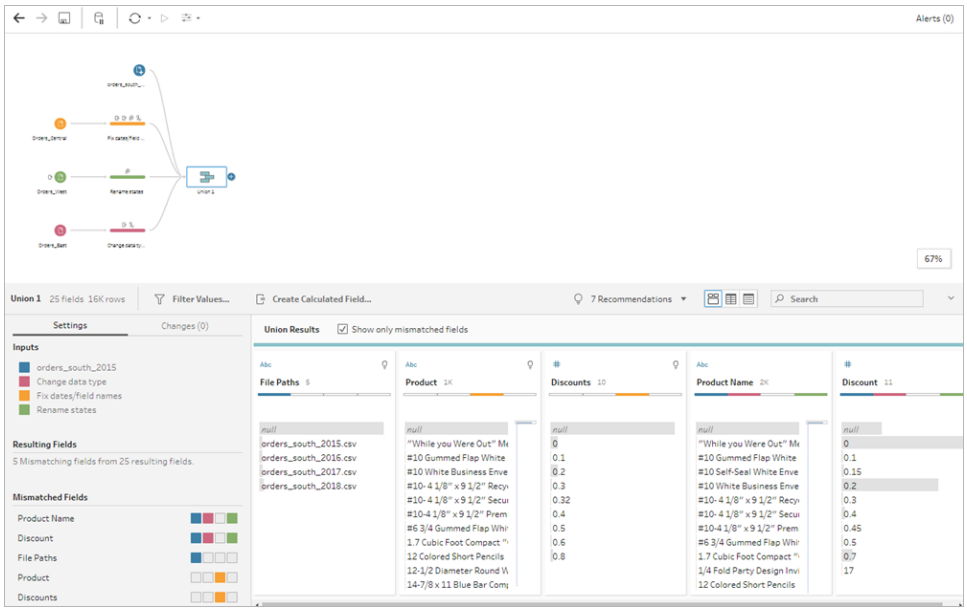
您注意到，Tableau 自动匹配了具有相同名称和类型的字段。

您还看到，分配给流程中步骤的**颜色**用在并集配置中指明字段来自于何处，并同时出现在每个字段顶部的**色带**中，显示该字段是否存在于该表中。

您注意到新增了一个名为**“Table Names”(表名称)**的新字段，该字段列出了并集中所有行的来源表。

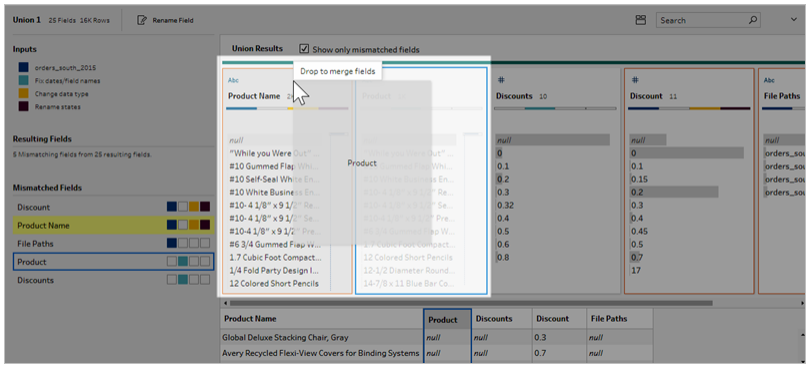
不匹配字段的列表也显示在摘要窗格中，并且您可以立即看到字段**“Product”(产品)**和**“Discounts”(折扣)**只出现在 **Orders_Central** 文件中。

4. 若要更仔细地查看这些字段，请在**“并集结果”**窗格中选中**“仅显示不匹配字段”**复选框。



通过查看字段数据，您很快发现数据相同，但字段名不同。您可以简单地重命名字段，但不知道是否能拖放这些字段来合并它们。您决定试一下看看。

5. 选择**“Product”(产品)**字段，并将其拖放到**“Product Name”(产品名称)**字段上来合并字段。字段合并后将不再出现在窗格中。



6. 重复此步骤，将**“Discounts”(多项折扣)**字段与**“Discount”(折扣)**字段合并。

现在，唯一没有匹配项的字段是**“File Paths”(文件路径)**字段。在 Tableau Prep Builder 中，此字段显示您为“南部”区域销售订单执行的通配符并集的文件路径。您决定将此字段保留在该处，因为它有很好的信息。

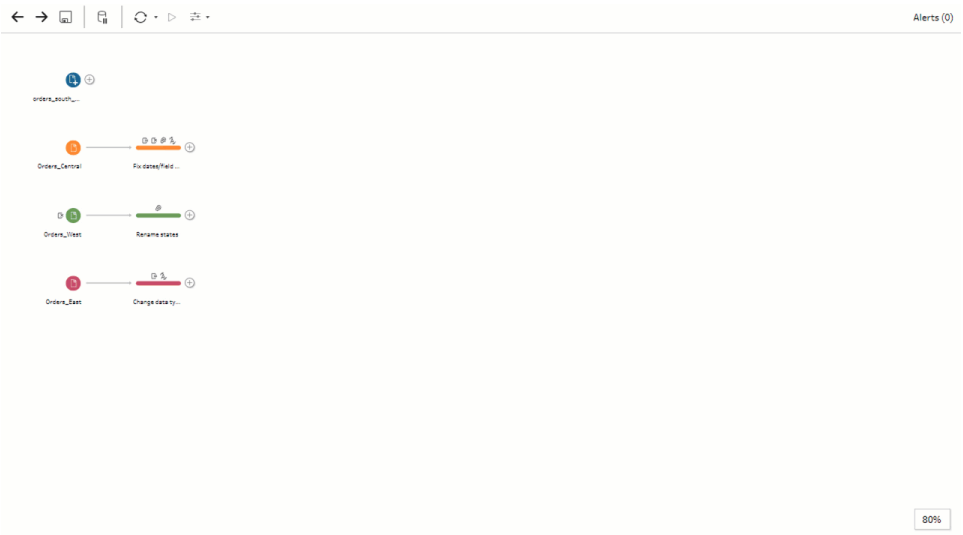
提示：在合并之后修复不匹配的字段时，您有若干选择。如果 Tableau Prep 检测到可能的匹配项，它将以黄色突出显示相应匹配项。若要合并字段，请将光标悬停在突出显示的字段上，并单击出现的加号 ⊕ 按钮。

有关在并集中合并字段的更多方式，请参见修复不匹配的字段 在本页349。

7. 清除**“仅显示不匹配字段”**复选框以显示并集中包括的所有字段。
8. 为合并步骤命名以表示此并集所包括的内容。例如，“**All orders**”(所有订单)。

检查您的工作:观看“合并数据”的实际运行效果。

单击图像可重播



您是清理天才！正当您对结果赞不绝口时，您的老板打来了电话。他忘了说他还想要您在分析中包括任何产品退货。他希望这项工作不会太麻烦。利用工具包中的 Tableau Prep，这完全不成问题！

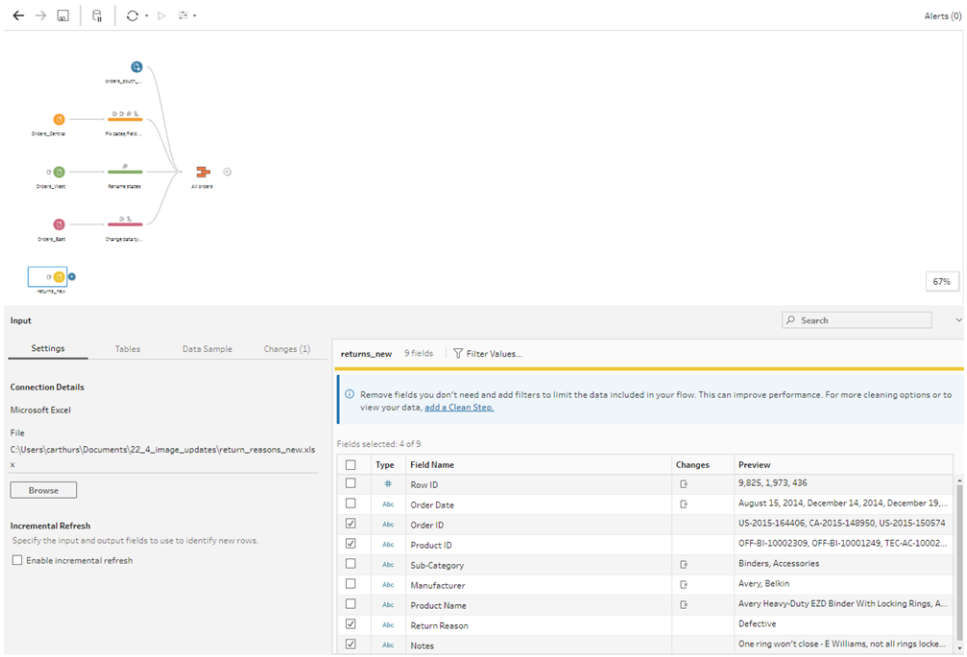
清理产品退货数据

您检查老板发给您的产品退货相关 Excel 文件，发现文件有一点杂乱。您将新文件 **return_reasons new** 添加到您的流程来仔细检查。

1. 在**“连接”**窗格中，单击**“添加连接”**。选择**“Microsoft Excel”**并导航到您用于本练习的示例数据文件。(请参见示例文件 在本页3以下载文件。)
2. 选择 **return reasons_new.xlsx**，然后单击**“打开”**将该文件添加到流程窗格。

   此文件中只有 4 个您想要包括在流程中的字段:**“Order ID”(订单 ID)**、**“Product ID”(产品 ID)**、**“Return Reason”(退货原因)**和**“Notes”(注释)**。

3. 在“**returns_new**”的**“输入”**窗格中，清除最左边的列顶部的复选框以清除所有复选框。然后，选中**“Order ID”(订单 ID)**、**“Product ID”(产品 ID)**、**“Return Reason”(退货原因)**和**“Notes”(注释)**字段的复选框。



4. 重命名输入步骤以更好地反映此输入中包括的数据。在**“流程”**窗格中，双击输入步骤名称**“Returns_new”**，并键入**“Returns (all)”(退货(全部))**。

   通过查看示例字段值，您注意到**“Notes”(注释)**字段似乎有很多不同的数据合并在一起。

   您要在此文件中执行一些清理，然后才能对数据进行进一步处理，因此，您添加了清理步骤来进行检查。

5. 在**“流程”**窗格中，选择输入步骤**“Returns (all)”**(退货(全部))，单击加号 ⊕ 图标或建议的清理步骤，添加一个清理步骤。

   在**“配置”**窗格中，调整**“Notes”(流程)**字段的大小以便更好地能看到各个条目。为此，请单击并将字段的右侧外缘向右拖。

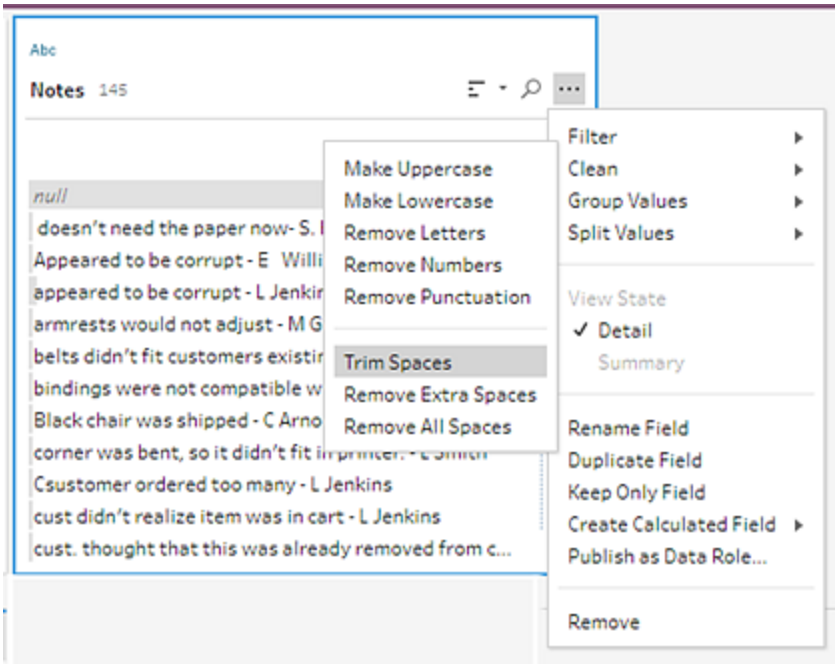
6. 在**“Notes”(注释)**字段中，使用字段值右侧可视的滚动条来浏览各个值。

   您注意到一些内容有问题：

- 某些条目的内容中有额外的空格。这可能会导致字段被读取为 null 值。
- 审批者的名称似乎包括在退货注释条目中。为了更好地使用此数据，您需要将该信息放在一个单独的字段中。

为了处理额外的空格，您记得有一个用于移除尾随空格的清理选项，因此决定尝试一下，看看该选项是否能修复问题。

7. 选择**“Notes”(注释)**字段。单击**“更多选项”**⋯ 菜单(在以前的版本中为下拉箭头)，并选择**“清理”>“剪裁空格”**。

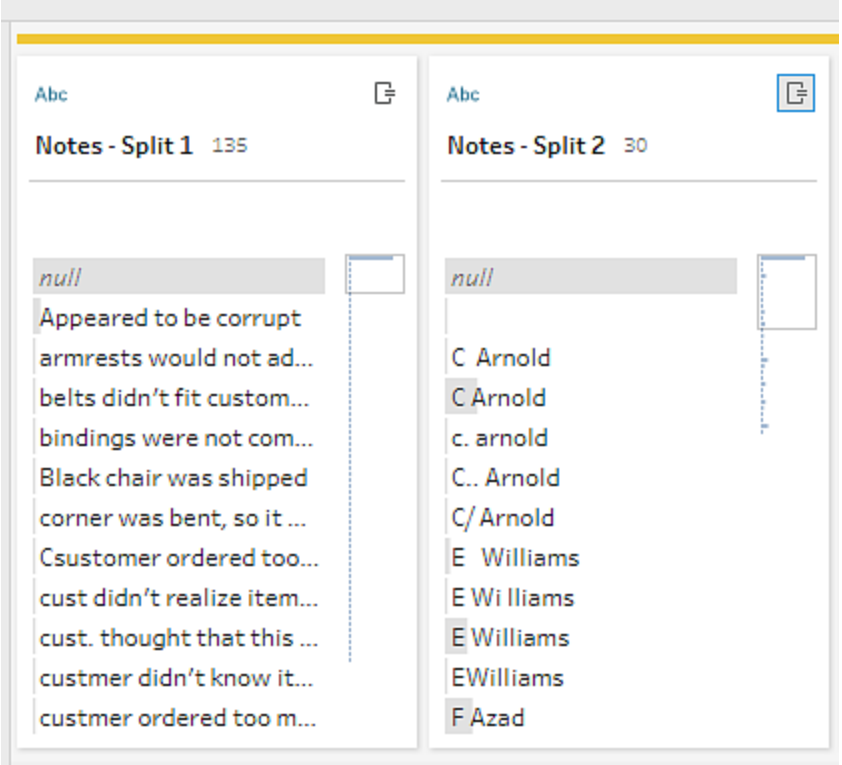


太好了！它准确地完成了所需的操作。额外的空格不见了。

接下来，您想要为审批者名称创建一个单独的字段。您看到菜单中的**“拆分值”**选项，因此决定尝试该选项。

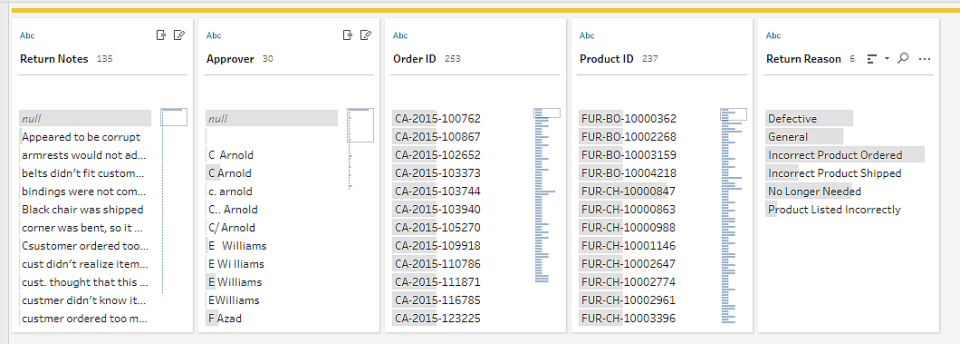
8. 选择**“Notes”(注释)**字段。单击**“更多选项”**⋯ 菜单(在以前的版本中为下拉箭头)，并选择**“拆分值”>“自动拆分”**。

此选项准确地完成了您希望其进行的操作。它自动将退货注释和审批者姓名拆分到单独的字段中。



就像 Tableau Desktop 一样，Tableau Prep 也为这些字段自动分配了名称。因此您将需要将新字段重命名为有意义的名称。

9. 选择字段**“Notes-Split 1”(注释-拆分 1)**。在字段名中双击，并键入**“Return Notes”(退货注释)**。
10. 为第二个字段重复此步骤，并将其重命名为**“Approver”(审批者)**。
11. 最后，在您不再需要原始**“Notes”(注释)**字段时将其移除。选择**“Notes”(注释)**字段，单击**“更多选项”**⋯ 菜单(在以前的版本中为下拉箭头)，并从菜单中选择**“移除”**(在以前的版本中为**“移除字段”**)。

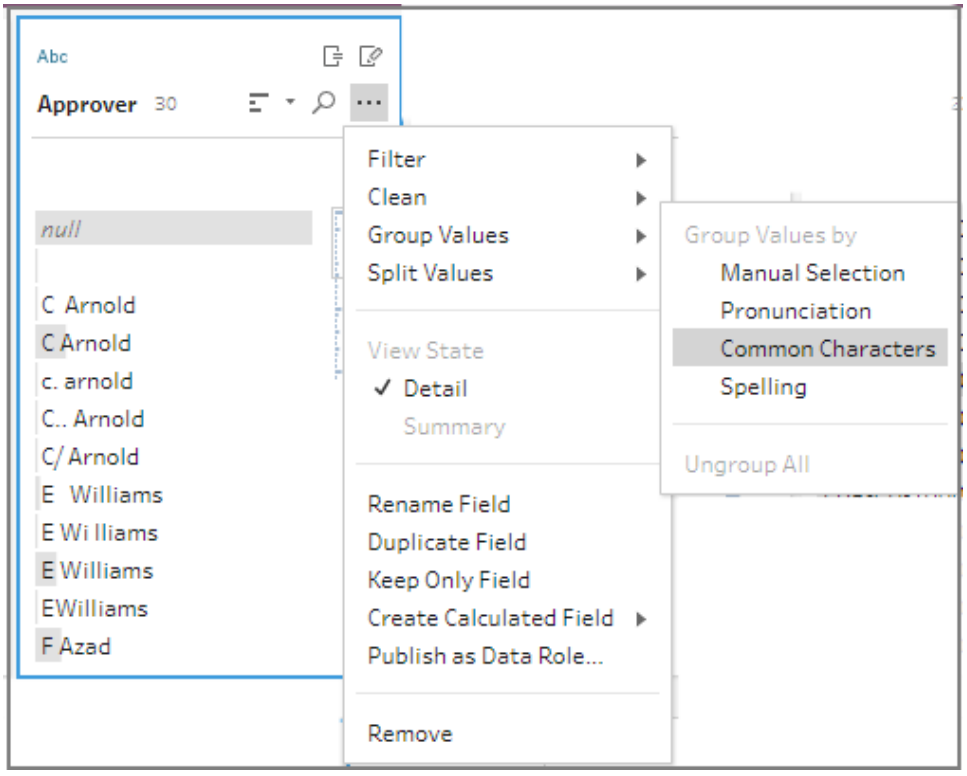


通过查看新的**"Approver"(审批者)**字段，您注意到字段值列出了相同名称，但这些名称的输入方式不同。您想要将这些字段分组，以消除同一值的多个变体。

也许**"对值进行分组"**(以前版本中的"分组和替换")选项可帮助完成该操作？

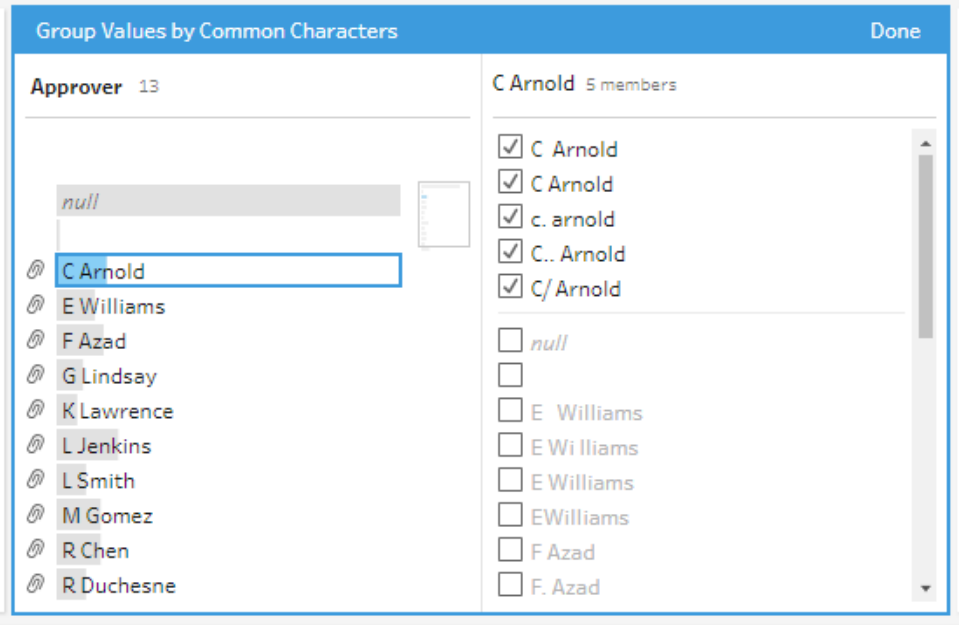
您记得有一个用于**"共同字符"**的选项。由于这些值共用相同字母，因此您决定试一下。

12. 选择**"Approver"(审批者)**字段。单击**"更多选项"**··· 菜单(在以前的版本中为下拉箭头)，并选择**"对值进行分组"**(以前版本中的"分组和替换")**>"共同字符"**。



此选项为您将每个名称的所有变体分组在一起。这完全就是您想要达到的目的。

检查了其他名称来确保其正确分组后，您单击**“完成”**关闭**“对值进行分组”**编辑器。



此文件看起来非常好。

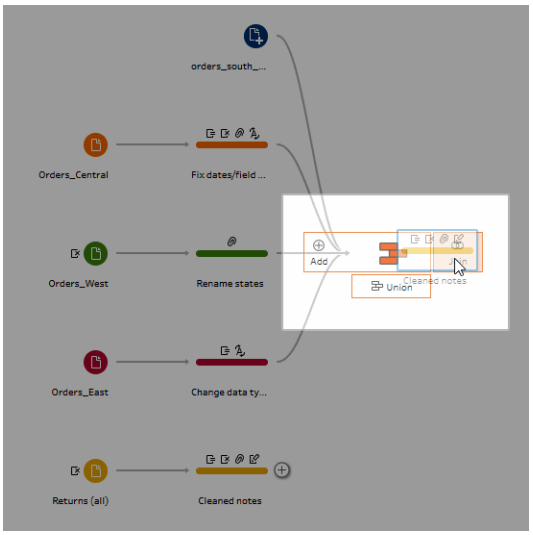
13. 为清理步骤命名以跟踪您的工作。例如，“**Cleaned notes**”(清理的注释)。

既然产品退货数据全部已清理，那么您想要将退货数据添加到合并的文件中的订单数据。但其中许多字段在合并的文件中不存在。若要将这些字段(数据列)添加到合并的数据集，您需要使用联接。

联接数据

在联接数据时，这些文件至少必须有一个共同的字段。您的文件共用**“Order ID”(订单 ID)**和**“Product Id”(产品 Id)**字段，因此您可以在这些字段上联接，以查看共有这些字段的所有行。在您使用拖放创建并集时，您记得有一个用于创建联接的选项，因此您试了一下。

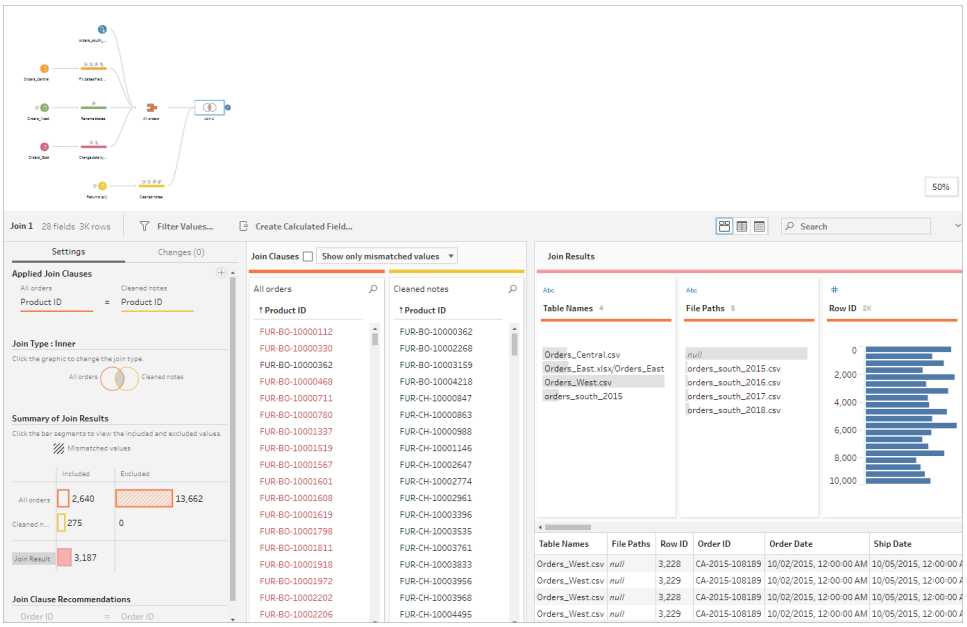
1. 在**“流程”**窗格中，将**“Cleaned notes”(清理的注释)**步骤拖到**“All orders”(所有订单)**合并步骤上，并将其放在**“联接”**上。



在联接文件时，Tableau Prep 会在**“联接配置”**中显示联接的结果。

处理联接可能会很棘手。您通常想要有一个具有联接中所包括因素的清晰视图，例如用于联接文件的字段、结果中的包括的行数，以及未包括或为 null 值的任何字段。

在 Tableau Prep 中查看联接的结果时，您很高兴地看到有如此多的信息，并且可以随时交互。



提示：在联接配置最左侧的窗格中，您可以浏览联接并与之交互。您也可以在**“联接子句”**窗格中直接编辑值，并在**“联接结果”**窗格中执行清理操作。

在**“联接类型”**图表中单击以尝试不同的联接配置，并在**“联接结果汇总”**部分中查看联接中为每个表包括或排除的行数。

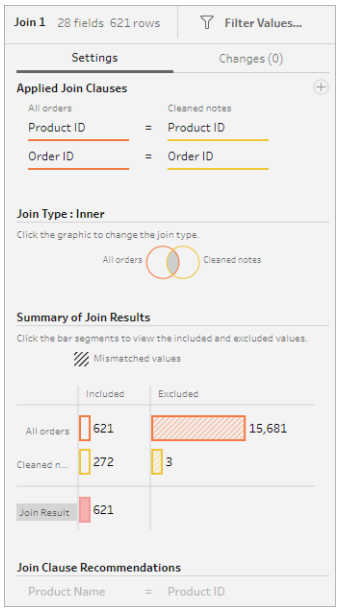
在“已应用联接子句”部分中选择要进行联接的字段，或通过“联接子句建议”部分添加建议的联接子句。

有关使用联接的详细信息，请参见**聚合、联接或合并数据** 在本页339。

您看到有超过 13,000 行被从“**All Orders**”(所有订单)文件中排除。创建联接之后，Tableau Prep 自动对**“Product ID”(产品 ID)**字段进行了联接，但您还想对**“Order ID”(订单 ID)**字段进行联接。

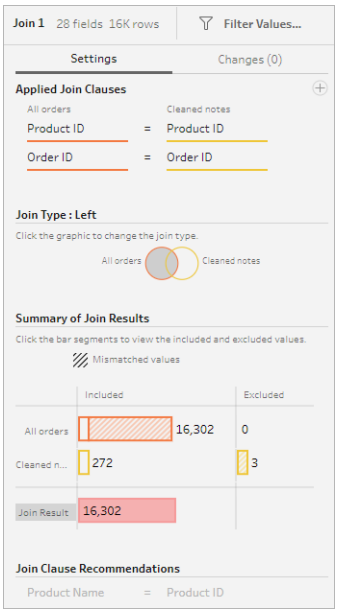
在浏览联接配置的左侧窗格时，您看到建议的联接子句列表中有**“Order ID”(订单 Id)**，因此您从该处快速添加了该字段。

2. 在**联接**配置左侧窗格中的**“联接子句建议”**部分，选择**“Order ID = Order Id”**，并单击加号 按钮添加联接子句。



由于**“联接类型”**设置为内部联接(Tableau Prep 的默认设置),因此联接仅包括两个文件中均存在的值。但您需要**“Orders”(订单)**文件中的所有数据,以及这些文件的退货数据。因此您将需要更改联接类型。

3. 在**“联接类型”**部分,单击图表的一侧以包括所有订单。在下面的示例中,单击图表的左侧以将联接类型更改为**“左”**联接,并包括**“All orders”(所有订单)**合并步骤中的所有数据,以及**“Cleaned notes”(清理的注释)**步骤中的任何匹配数据。



现在,您有了销售订单文件中的所有数据,以及适用于这些订单的任何退货数据。您查看**“联接子句”**窗格,并看到一些其他文件中不存在的不同值。

例如,有很多没有对应退货数据的订单行(显示为红色)。您能够浏览此详细级别的联接相关信息,这让您非常喜欢。

您急于想在 Tableau Desktop 中开始分析此数据,但在开始分析之前,您注意到联接要有一些结果需要清理。幸好您知道怎么做!

提示:不知道数据是否足够干净?从 Tableau Prep Builder 中,您可以从流程的任何步骤中通过 Tableau Desktop 预览数据,对其进行检查。

只需在**“流程”**窗格中右键单击该步骤,并从菜单中选择**“在 Tableau Desktop 中预览”**。

您可能会遇到您的数据和在 Tableau Desktop 中进行的更改未写回 Tableau Prep

Builder 中的数据源的情况。有关详细信息，请参见在 Tableau Desktop 中查看流程输出 在本页363。

4. 在开始清理联接结果之前，将**联接**步骤命名为**“Orders+Returns”(订单+退货)**，并保存流程。

清理联接结果

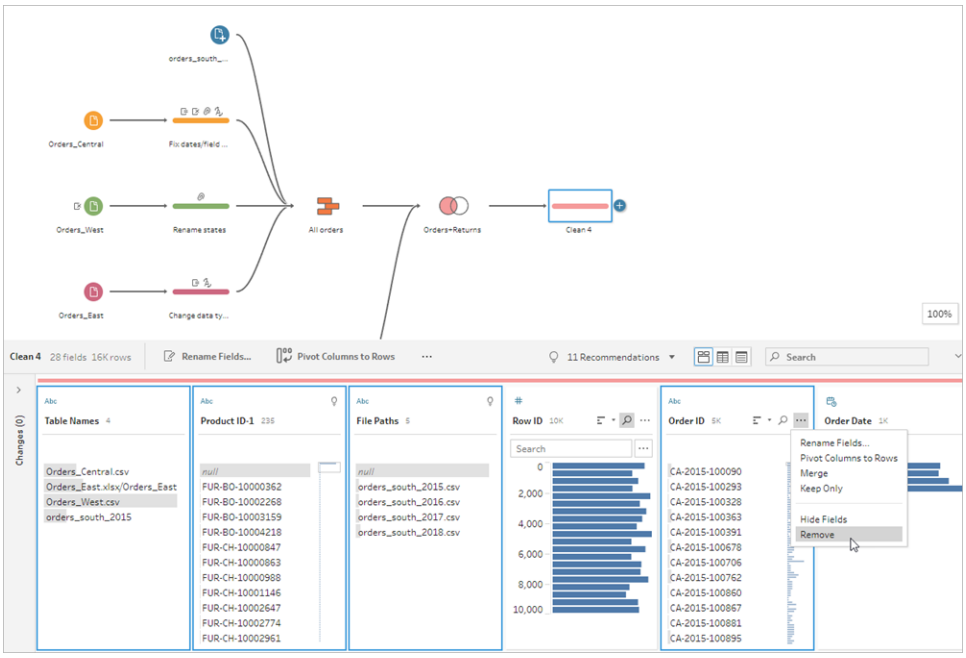
注意：若要清理联接中的字段，您可以直接在“联接”步骤中执行清理操作。对于本教程，我们将添加一个清理步骤，以便您能够清晰地看到您的清理操作。如果要尝试直接在联接步骤中执行这些步骤，请跳过下面的步骤 1 和 3。

联接了两个步骤之后，系统为两个表都添加了“Order ID”(订单 ID) 和“Product ID”(产品 ID) 共同字段。

您想要保留所有订单中的“Product ID”(产品 ID) 字段和退货文件中的“Order ID”(订单 ID) 字段，并移除来自这些字段的重复字段。您也不需要输出文件中的**“File Paths”(文件路径)** 和**“Table Names”(表名称)** 字段，因此也需要移除这些字段。

提示：当您使用两个文件中均存在的字段联接表时，Tableau Prep 将会引入这些字段，并通过向字段名中添加“-1”或“-2”来重命名第二个文件中的重复字段。例如，“Order ID”(订单 ID) 和“Order ID-1”(订单 ID-1)。

1. 在**“流程”**窗格中，选择“**Orders+Returns**”(订单+退货)，单击加号 ⊕ 图标，并添加清理步骤。
2. 在**“配置”**窗格中，选择并移除以下字段：
    - **Table Names(表名称)**
    - **Order ID(订单 ID)**
    - **文件路径**(仅限 Tableau Prep Builder)
    - **Product ID-1(产品 ID-1)**



3. 将字段**“Order ID-1”(订单 Id-1)**重命名为**“Order ID”(订单 ID)**。

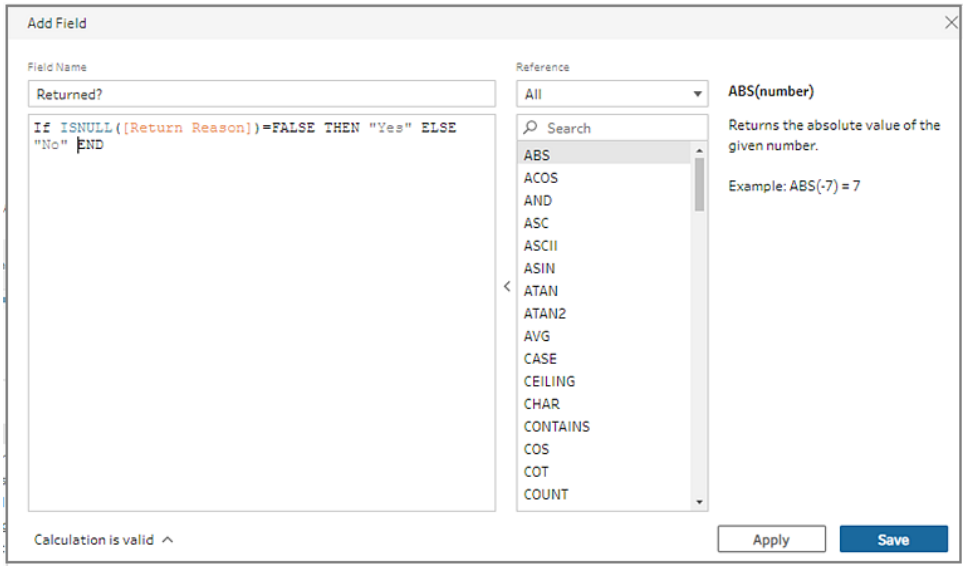
   有很多产品已退货的 null 值，但没有指明退货注释或审批者。为了使此数据更易于分析，您想要添加一个值为**“Yes”**和**“No”**的字段来指明产品是否已退货。

   您没有此字段，但可以通过创建计算字段来添加它。

4. 在工具栏中，单击**“创建计算字段”**。

5. 将字段命名为**“Returned?”(已退货？)**，然后输入以下计算并单击**“保存”**。

   ```
   If ISNULL([Return Reason])=FALSE THEN "Yes" ELSE "No" END
   ```

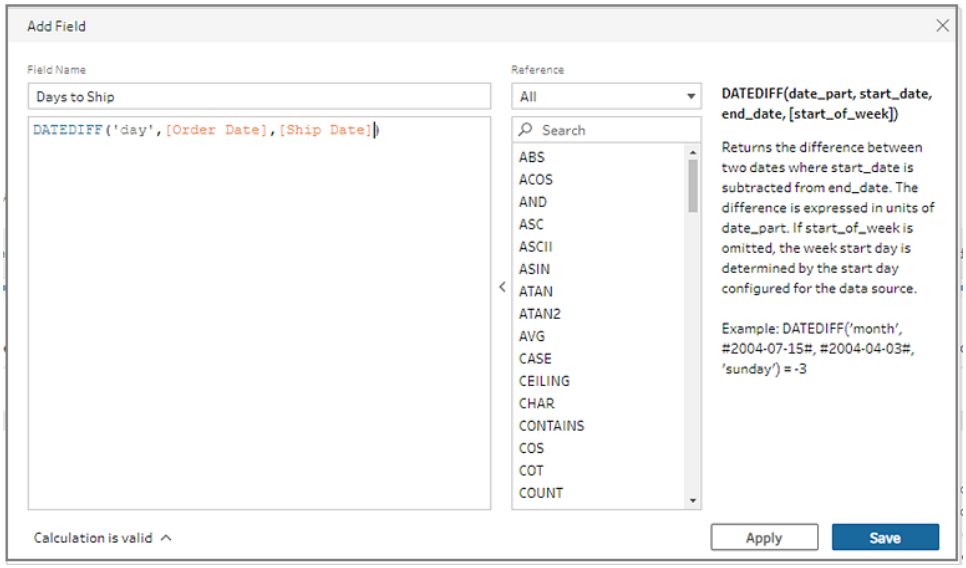


为了进行分析，您还想要知道订单发货天数，但您也没有该字段。

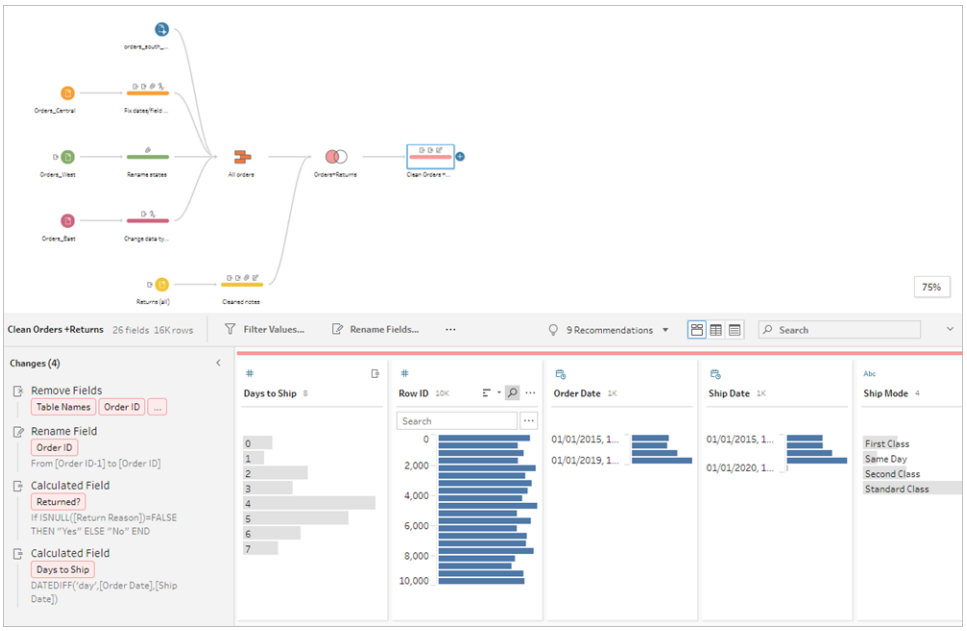
不过您有创建该字段所需的全部信息，因此添加了另一个计算字段来创建它。

6. 在工具栏中，单击**“创建计算字段”**。
7. 将字段命名为**“Days to Ship”(发货天数)**，然后输入以下计算并单击**“保存”**。

```
DATEDIFF('day',[Order Date],[Ship Date])
```



8. 将步骤命名为**“Clean Orders +Returns”(清理订单+退货)**。
9. 保存您的流程。



5.运行流程并生成输出

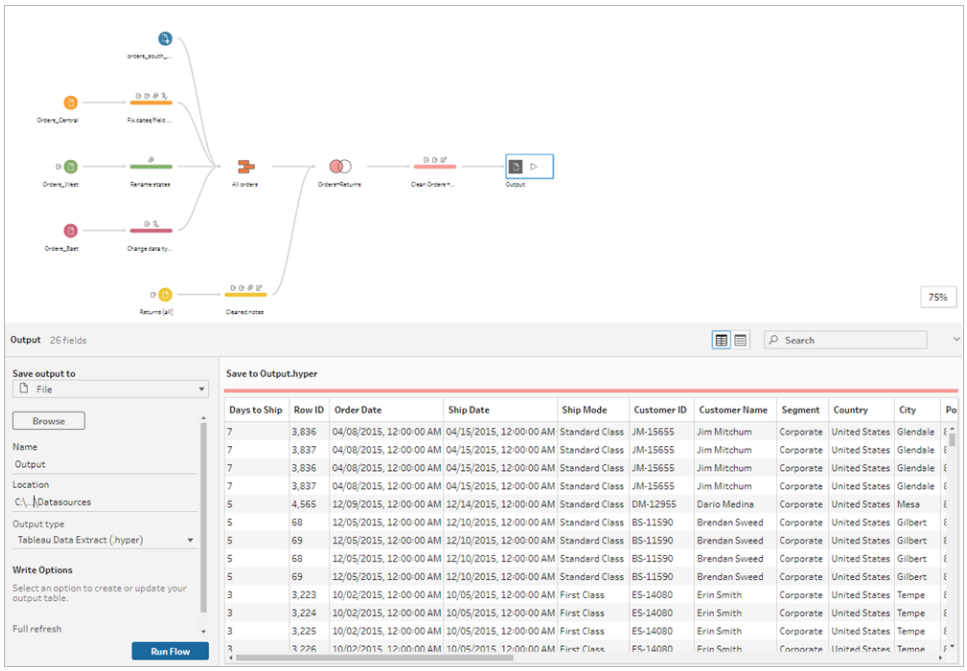
您的数据看起来很好，并且您已准备好生成输出文件以便开始在 Tableau Desktop 中对其进行分析。您只需运行流程并生成数据提取文件即可。为此，您需要添加一个**输出**步骤。

根据工作位置，您可以将流程输出到文件（仅限 Tableau Prep Builder），输出到已发布数据源或数据库。

1. 在**“流程”**窗格中，选择“**Clean Orders+Returns**”（清理订单+退货），单击加号 ⊕ 图标，并选择**“输出”**（以前版本中的**“添加输出”**）。

   添加输出步骤时，**“输出”**窗格将打开，并显示数据的快照。在这里您可以选择要生成的输出的类型，并指定名称和要保存文件的位置。

   默认位置为数据源文件夹中的**“我的 Tableau Prep Builder”**存储库。



2. 在左侧窗格的**“将输出保存到”**下拉列表中，根据工作位置，执行以下操作之一：

   Tableau Prep Builder

   a. 选择**“文件”**（在以前的版本中选择**“保存到文件”**）。
   b. 单击**“浏览”**按钮，然后在**“将数据提取另存为”**对话框中输入文件的名称，例如**“Orders_Returns_Superstore”**，并单击**“接受”**。
   c. 在**“输出类型”**字段中，选择一种输出类型。对于 Tableau Desktop，请选择 Tableau 数据提取 (.hyper)，或者，如果要与第三方共享数据提取，请选择逗号分隔值 (.csv)。

   Tableau Server 或 Tableau Cloud

   a. 选择**“已发布数据源”**。
   b. 选择一个项目。
   c. 输入文件的名称，例如**“Orders_Returns_Superstore”**。

   提示：在通过流程生成输出时，您有一些选择。您可以生成数据提取文件（仅限 Tableau Prep Builder），可以将您的数据以数据源的形式发布到 Tableau Server 或 Tableau Cloud，或者可以将数据写入数据库。有关生成输出文件的详细信息，请参见创建数据提取文件和已发布数据源 在本页364。

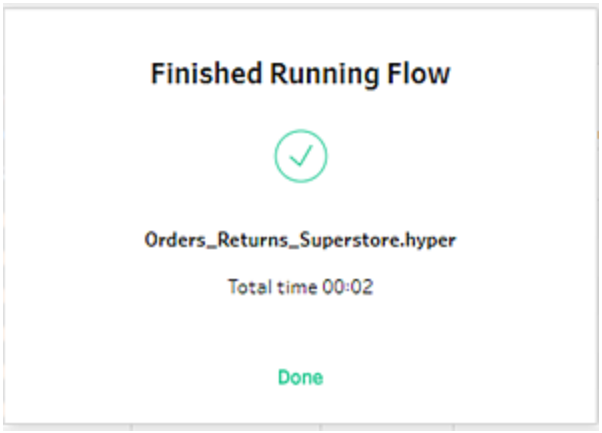
3. 在**“写入选项”**部分，查看将新数据写入文件的选项。您希望使用默认选项（**“创建表”**）并将表替换为流程输出，因而此处没有要更改的内容。

提示：从版本 2020.2.1 开始，您可以选择要如何将流程数据写回表。您可以从两个选项中进行选择；**“创建表”**或**“追加表”**。默认情况下，Tableau Prep 使用**“创建表”**选项，并在运行流程时用新数据覆盖表数据。如果选择**“追加表”**，Tableau Prep 会将流程数据添加到现有表，以便可以在每次流程运行时同时跟踪新数据和历史数据。有关详细信息，请参见**配置写入选项** 在本页394。

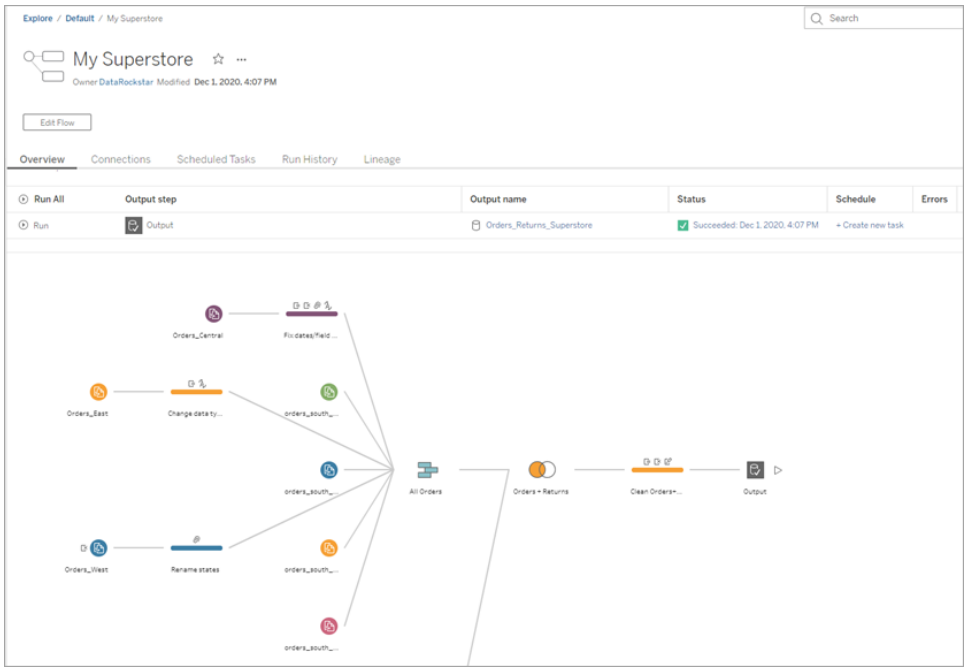
4. 在“输出”窗格中单击**“运行流程”**，或在流程窗格中单击**“运行流程”** ▷ 按钮以生成输出。

   注意：如果在 Web 上工作，请单击“**发布**”以发布草稿流程。只能运行已发布的流程。

5. 流程完成运行时，一个状态对话框将显示流程运行成功，以及运行所花费的时间。单击**“完成”**关闭该对话框。



如果在 Web 上工作，请导航到**“浏览”>“所有流程”**页面，并找到您的流程。您可以在“流程概述”页面上查看流程运行的状态。



若要使数据保持最新，您可以手动或使用命令行来运行流程。如果您有 Data Management 并启用了 Tableau Prep Conductor，还可以在 Tableau Server 或 Tableau Cloud 中按计划运行流程。

从 Tableau Prep Builder 版本 2020.2.1 开始以及在 Web 上，您还可以选择在每次运行流程时刷新所有数据，或使用增量刷新运行流程，并且每次只处理新数据。

有关使数据保持最新的详细信息，请参见以下主题：

- **通过命令行刷新流程输出文件** 在本页 397
- **将流程发布到 Tableau Server 或 Tableau Cloud** 在本页 433
- **使用增量刷新来刷新流程数据** 在本页 390

总结和资源

您是数据准备巨星了！您轻松地找出了不良数据并对其进行了改变！您很快清理并准备好了来自多个数据集的数据，并将其转换为可立即在 Tableau Desktop 中处理来进行分析的井然有序、干净的数据集。

想要进行更多练习？尝试使用下列文件夹中的数据文件重复执行 Superstore 示例流程的其余步骤：

- Orders_South_2015
- Orders_South_2016

- Orders_South_2017
- Orders_South_2018
- Orders_Central
- Orders_East
- Orders_West
- returns_reasons_new
- Quota

在安装 Tableau Prep Builder 后，您还可以在计算机上的以下位置找到这些文件：

- (Windows) `C:\Program Files\Tableau\Tableau Prep Builder <version>\help\Samples\en_US\Superstore Files`
- (Mac) `/Applications/Tableau Prep Builder <version>.app/Contents/help/Samples/en_US/Superstore Files`

想要获得更多培训？查看这些优质资源，或参加现场培训课程。

想要了解我们所讨论主题的详细信息？请查看 Tableau Prep 联机帮助中的其他主题。

关于 Tableau Prep

Tableau Prep Builder 是 Tableau 产品套件中的一个新工具，旨在让您的数据准备工作更加轻松和直观。使用 Tableau Prep Builder 来合并、调整和清理数据，以便在 Tableau 中进行分析。

注意：Tableau Prep 版本 2019.1.2 已更名为 Tableau Prep Builder 并引用 Desktop 应用程序。从版本 2020.4 开始，作为 Creator，您还可以在 Tableau Server 和 Tableau Cloud 上创建和编辑流程。

使用 Tableau Prep

首先，通过各种文件、服务器或 Tableau 数据提取连接到数据。连接到来自多个数据源的数据并合并数据。拖放或双击以将表放入流程窗格，然后添加流程步骤，您可以在其中使用熟悉的操作（例如筛选、拆分、重命名、转置、联接和合并）来清理和调整数据。

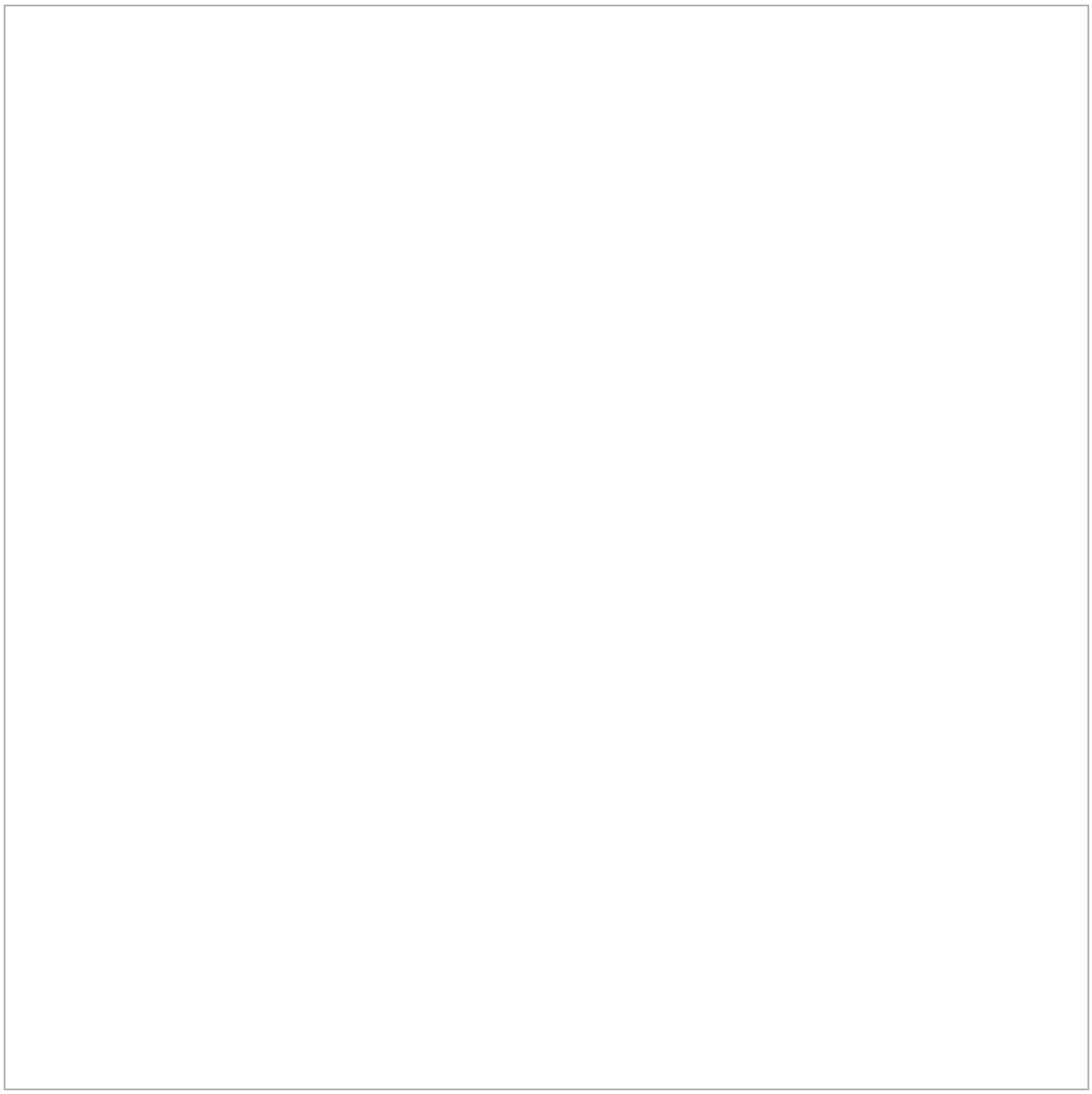
流程中的每个步骤都直观地呈现在您创建和控制的流程图中。Tableau Prep 会跟踪每个操作，以便您能够在流程中的任何位置检查工作和进行更改。

完成流程之后，请运行流程以将操作应用于整个数据集。

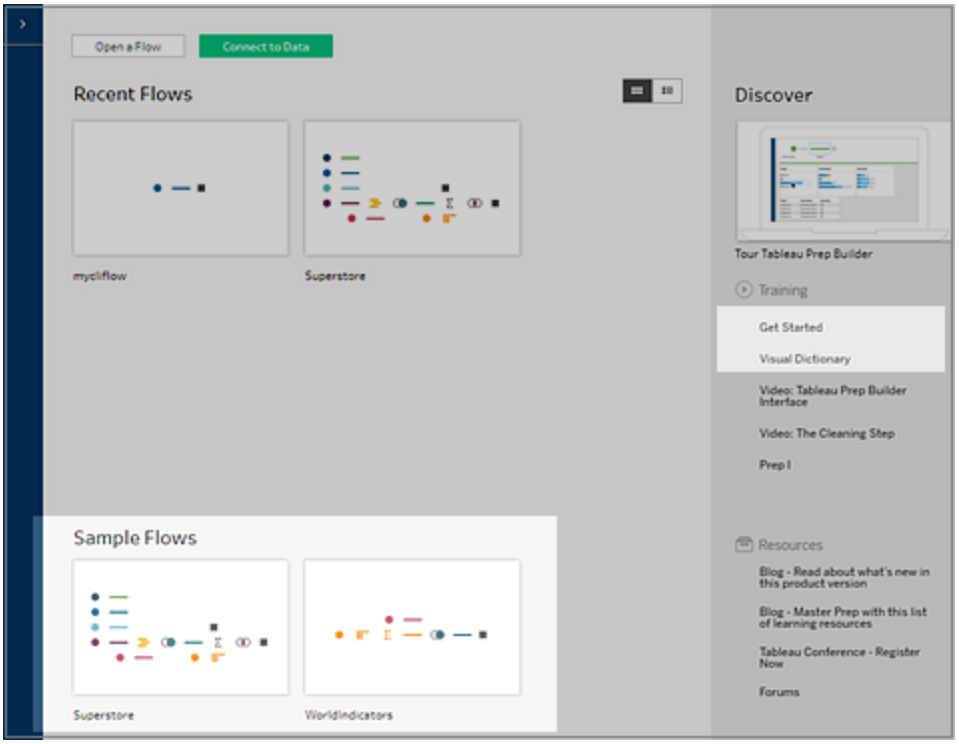
Tableau Prep 可与其他 Tableau 产品无缝配合工作。您可以在流程中的任何位置创建数据提取、将数据源发布到 Tableau Server 或 Tableau Cloud，将流程发布到 Tableau Server 或 Tableau Cloud 以继续在 Web 上编辑或使用计划刷新数据。您还可以直接从 Tableau Prep Builder 中打开 Tableau Desktop 以预览数据。

有关安装 Tableau Prep Builder 的信息，请参见 Tableau Desktop 和 Tableau Prep 部署指南中的安装 Tableau Desktop 或 Tableau Prep Builder。

观看视频：查看 Tableau Prep Builder 的实际运行效果



准备好试用了吗？从“**开始**”页面中，单击示例流程之一以浏览和体验各个步骤，尝试 **Tableau Prep Builder 入门指南** 在本页3实践教程来了解如何创建流程，或者尝试使用 Tableau Prep Builder 逐步执行**日常生活场景** 在本页453。



注意：您可以在以下位置找到流程中使用的示例数据文件：

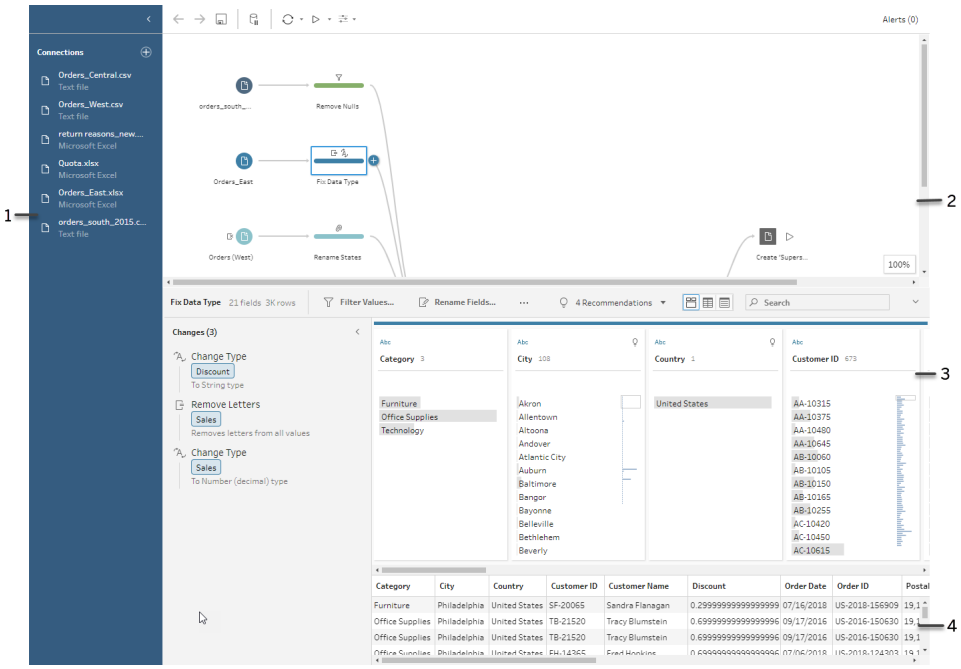
- (Windows) `C:\Program Files\Tableau\Tableau Prep Builder <version>\help\Samples\en_US`
- (Mac) `/Applications/Tableau Prep Builder <version>.app/Contents/help/Samples/en_US`

若要详细了解 Tableau Prep Builder 如何针对性能优化您的数据，请参见 Tableau Prep under the hood(Tableau Prep 后台揭秘）。若要详细了解 Tableau Prep 及其提供的不同特性和功能，请查看本指南中的各个主题。

Tableau Prep 工作区一览

Tableau Prep 工作区由可在其中连接到数据源的**“连接”窗格**(1)和三个协调区域组成，这些区域可帮助您与数据交互以及浏览数据：

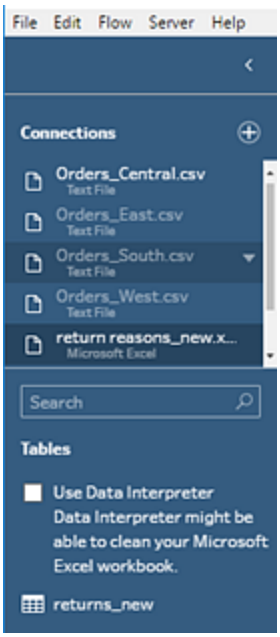
- **“流程”窗格** (2):在您准备数据时直观地呈现操作步骤。这是添加步骤来构建流程的位置。
- **“配置”窗格** (3):汇总了数据样本中的每个字段。查看数据的形状,并快速查找离群值和 null 值。
- **“数据”网格** (4):数据的行级别详细信息。



连接到数据并开始构建流程之后,您将在**“流程”**窗格中添加步骤。这些步骤可深入展示数据的结构,并汇总了应用于数据的操作。每个步骤都表示流程中您定义的一个不同操作类别。

“连接”窗格 (1)

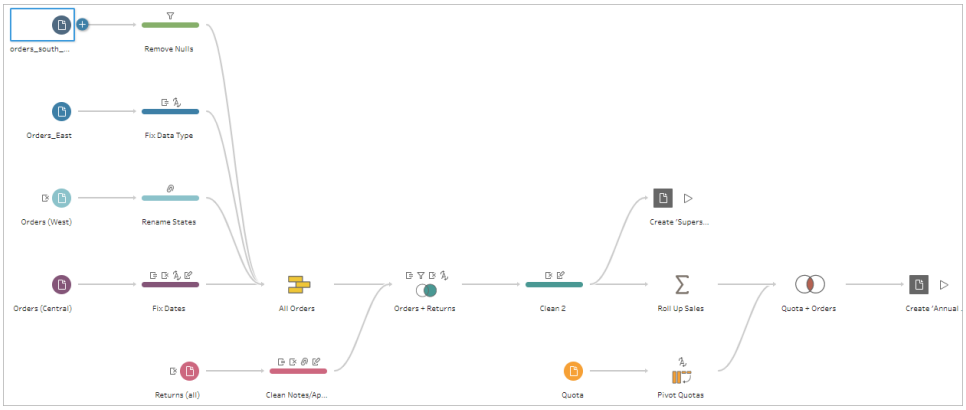
工作区的左侧是**“连接”**窗格,其中显示您连接到的数据库和文件。添加与一个或多个数据源的连接,然后将要使用的表拖到**“流程”**窗格中。有关详细信息,请参见**连接到数据** 在本页69。



如果需要工作区中有更多空间，您可以最小化**“连接”**窗格。

“流程”窗格 (2)

工作区的顶部是**“流程”**窗格。这是构建流程的位置。在您连接、清理、调整和合并数据时，步骤会出现在**“流程”**窗格中，并沿顶部从左到右对齐。这些步骤指出正在应用何种操作、应用操作的顺序以及会对数据产生什么影响。例如，“联接”步骤显示您已应用的联接类型、联接子句、建议使用的联接子句，以及联接中包括的表的字段。

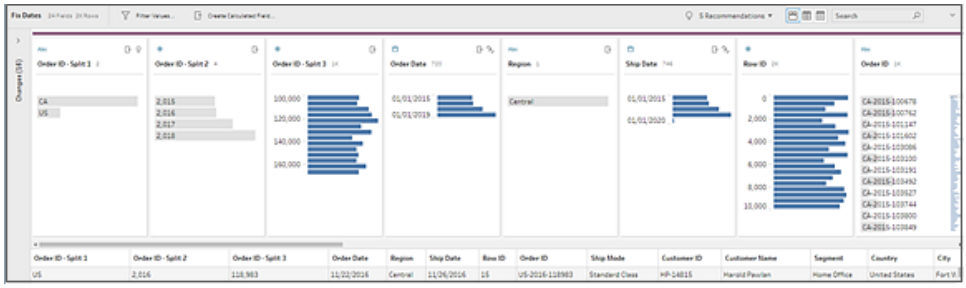


通过将表拖到**“流程”**窗格中来开始流程。在此处，您可以添加其他数据集、对数据进行转置、合并或联接数据、创建聚合，以及用文件(.hyper、.csv、.xlsx)、可在 Tableau 数据库中使用的已发布数据源格式或 CRM Analytics 生成流程输出文件。有关生成输出文件的详细信息，请参见保存和共享工作 在本页361。

注意：如果在 Tableau Desktop 中对数据进行更改，例如重命名字段、更改数据类型等，这些更改将不会写回到 Tableau Prep Builder。

“配置”窗格 (3)

工作区的中心是**“配置”**窗格。**“配置”**窗格显示流程中任何位置的数据的结构。可以用多种方式表示数据的结构，具体情况取决于要对数据执行的操作，或您在**“流程”**窗格中选择的步骤。

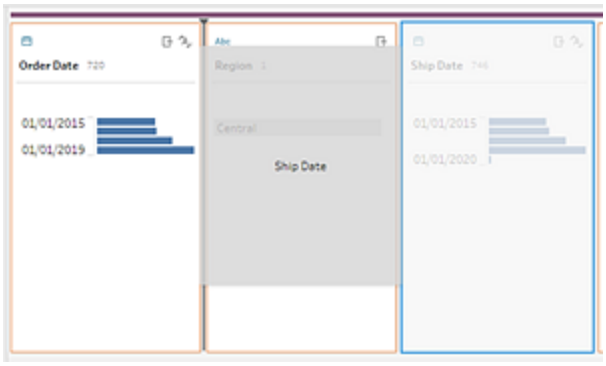


“配置”窗格顶部是一个工具栏，其中显示您可为流程中的每个步骤执行的清理操作。**“配置”**窗格中的每个卡上还将显示一个选项菜单，您可在其中选择可对数据执行的不同操作。

例如：

- 搜索字段、对字段排序以及拆分字段
- 筛选值、包括值或排除值
- 查看并修复 null 值
- 重命名字段
- 使用对值进行分组或快速清理操作清理数据条目错误
- 使用自动数据解析来更改数据类型

- 通过将字段列拖放到所需的位置，重新排列字段列的顺序



在“配置”卡上选择一个或多个字段值，并右键单击或按住 Ctrl 单击 (MacOS) 以查看用于保留或排除值、将选定值分组或将值替换为 Null 的其他选项。

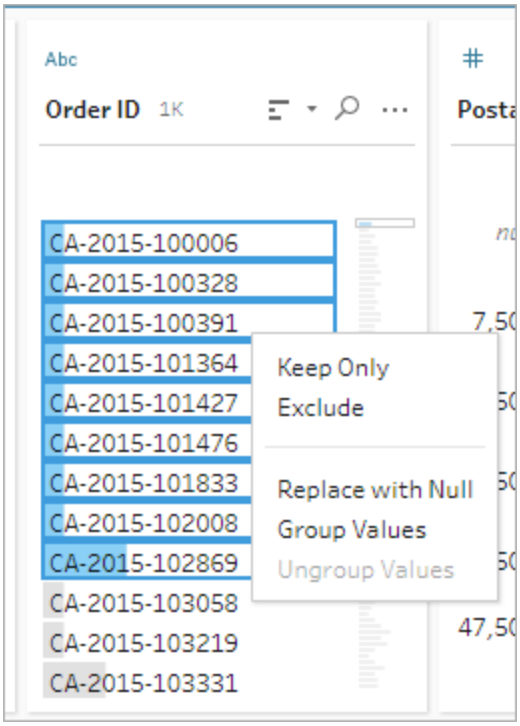
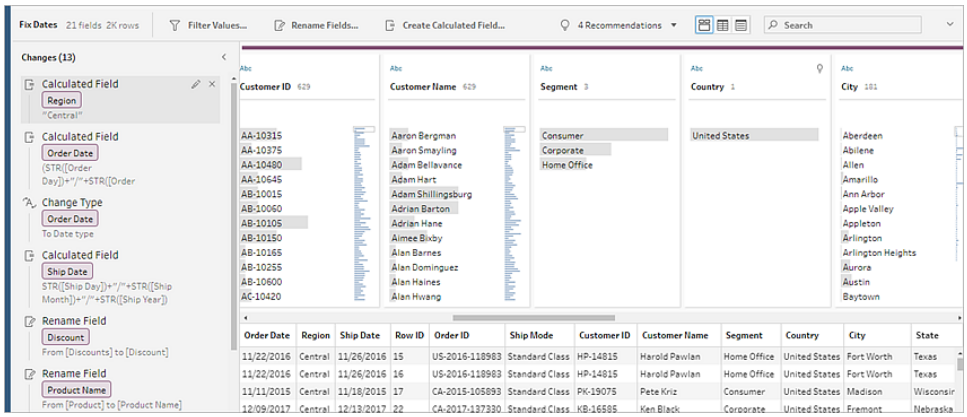


Tableau Prep 会按您进行更改的顺序来跟踪您所做的更改，以便您始终能够在需要时返回并查看或编辑这些更改。使用拖放对列表中的操作进行重新排序，按不同的顺序试验和应用更改。



单击窗格右上方的箭头展开和折叠“**更改**”窗格，腾出更多空间来在“**配置**”窗格中处理数据。

有关对数据应用清理操作的详细信息，请参见**清理和调整数据** 在本页213。

“数据”网格 (4)

工作区的底部是“**数据**”网格，其中显示数据中的行级别详细信息。“**数据**”网格中显示的值反映“**配置**”窗格中定义的操作。如果更愿意在更详细的级别工作，您可以像在“配置”窗格中一样在此处执行相同的清理操作。

单击工具栏上的“**折叠配置**” 图标折叠(和展开)“**配置**”窗格，以查看您的选项。

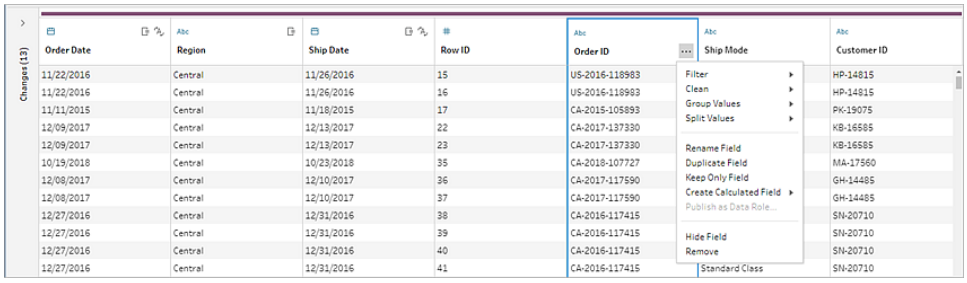


Tableau Prep 如何存储数据

当您将 Tableau Prep 连接到数据并创建流程时，它会将经常使用的数据存储在一个 .hyper 文件中。对于大型数据集，这可能是一个数据样本。任何存储的数据均保存在一个安全的临时文件目录下名为 Prep BuilderXXXXX 的文件中，其中 XXXXX 表示一个全局唯一标识符 (UUID)。保存流程之后，将会删除该文件。有关 Tableau Prep 如何进行数据抽样的详细信息，请参见**选择数据样本大小** 在本页114。

Tableau Prep Builder 还会将数据保存在 Tableau 流程 (.tfl) 文件中，以便支持以下操作（可捕获输入的数据值）：

- 输入步骤中使用的自定义 SQL
- 筛选（针对数据条目）
- 对值进行分组（针对数据条目）
- 计算

Web 上的 Tableau Prep

不支持 Windows 上的 Internet Explorer 11 和 Internet Explorer 的兼容模式。

从版本 2020.4 开始，Tableau Prep 支持流程的 Web 制作。现在，您可以创建流程以使用 Tableau Prep Builder、Tableau Server 或 Tableau Cloud 清理和准备数据。您也可以手动在 Web 上运行流程，而不需要 Data Management。

虽然 Web 上也支持大多数相同的 Tableau Prep Builder 功能，但创建和处理流程存在一些差异。

重要信息：若要在 Web 上创建和编辑流程，您必须具有 Creator 许可证。只有当您希望使用 Tableau Prep Conductor 按计划运行流程时，才需要 Data Management。有关配置和使用 Tableau Prep Conductor 的详细信息，请参见 Tableau Server 或 Tableau Cloud 帮助中的“Tableau Prep Conductor”。

安装和部署

若要使用户能够在 Web 上创建和编辑流程，您将需要在服务器上配置若干设置。有关其中每个设置的详细信息，请参见在 Web 上创建流程并与其进行交互。

- **Web 制作**：此选项默认情况下已启用，控制用户是否可以在 Tableau Server 或 Tableau Cloud 上创建和编辑流程。
- **立即运行**：控制用户或只有管理员可以使用**“立即运行”**选项手动运行流程。不需要 Data Management 即可在 Web 上手动运行流程。
- **Tableau Prep Conductor**：如果 Data Management 已获得许可，则启用此选项，以便允许用户计划和跟踪流程。
- **Tableau Prep 扩展程序**（版本 2021.2.0 及更高本）：控制用户是否可以连接到 Einstein Discovery 以针对其流程中的数据应用和运行预测模型。
- **自动保存**：（默认情况下启用）此功能每隔几秒钟自动保存用户的流程工作。

示例数据和处理限制

若要在 Web 上处理流程时保持性能，请对可以在流程中包含的数据量应用限制。

以下限制适用：

- 连接到文件时，最大文件大小为 1GB。
- 包含所有数据的数据采样选项不可用。默认示例数据限制为 100 万行。
- 用户在使用大型数据集时可以选择的最大行数由管理员配置。作为用户，您可以选择最多达到该限制的行数。

有关设置数据样本的详细信息，请参见 Tableau Prep 帮助中的设置数据样本大小。

Web 上的可用功能

当您在 Web 上创建和编辑流程时，您可能会注意到导航和某些功能可用性的一些差异。虽然大多数功能在所有平台上都可用，但 Tableau Server 或 Tableau Cloud 中某些功能受限或尚未受支持。下表标识可能存在差异的功能。

功能区域	例外	Tableau Prep Builder	Tableau Server	Tableau Cloud
连接到数据	某些连接器在 Web 上可能不受支持。打开服务器上的**"连接"**窗格以查看支持的连接器。	✓	✓	✓
构建和组织流程		✓	✓	✓
选择数据样本大小	在 Tableau Server 和 Tableau Cloud 中，数据样本大小受管理员设置的限制所限	✓	✓	✓
在输入步骤中合并文件和数据库表	无法在 Tableau Server 或 Tableau Cloud 中编辑或创建输入并集。仅在 Tableau Prep Builder 中提供。	✓	✓	✓
清理和调整数据		✓	✓	✓

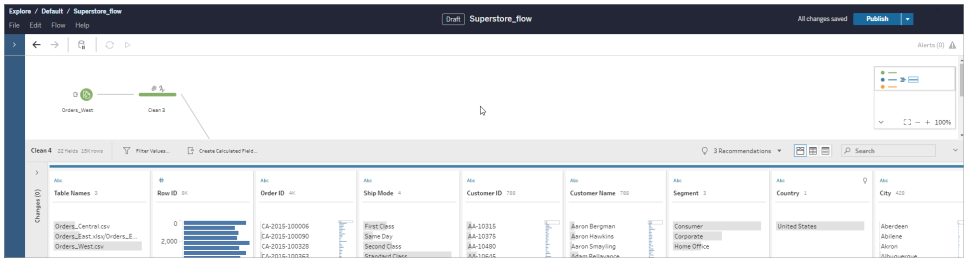
复制数据网格值	在 Tableau Prep Builder、Tableau Server(从版本 2022.3 开始) 和 Tableau Cloud(从 2022.2(八月) 开始) 中提供。	✓	✓	✓
聚合、联接或合并数据		✓	✓	✓
在流程中使用 R 和 Python 脚本	在 Tableau Cloud 中创建或编辑流程时，无法添加脚本步骤。这当前仅在 Tableau Prep builder 和 Tableau Server 中受支持。	✓	✓	⊘
创建可重用的流程步骤		✓	⊘	⊘
在 Web 上自动保存流程		不可用	✓	✓
自动文件恢复		✓	不可用	不可用
在 Tableau Desktop 中查看流程输出		✓	⊘	⊘
创建数据提取到文件		✓	⊘	⊘
创建数据提取到 Microsoft Excel 工作表		✓	⊘	⊘
连接到自定义 SQL 查询		✓	✓	✓
创建发布的数据源		✓	✓	✓

将流程输出保存到外部数据库		✓	✓	✓
将 Einstein Discovery 预测添加到您的流程		✓	✓	✓

自动保存并使用草稿

当您在服务器上创建或编辑流程时，您的工作每隔几秒钟会自动保存为草稿，这样在发生崩溃时，或在意外关闭选项卡时，您不会丢失工作。

草稿将保存到您登录的服务器和项目。您无法保存草稿或将草稿发布到其他服务器，但您可以使用**“文件”>“发布为”**菜单选项将流程保存到该服务器上的另一个项目。



草稿内容只能由您查看，直到您发布它。如果发布更改并需要还原更改，可以使用**“修订历史记录”**对话框查看并还原到以前发布的版本。有关在 Web 上保存流程的详细信息，请参见自动在 Web 上保存流程。

在 Web 上发布流程

无论您是在 Web 上从头开始创建流程还是编辑现有流程，在运行流程之前都需要发布流程。

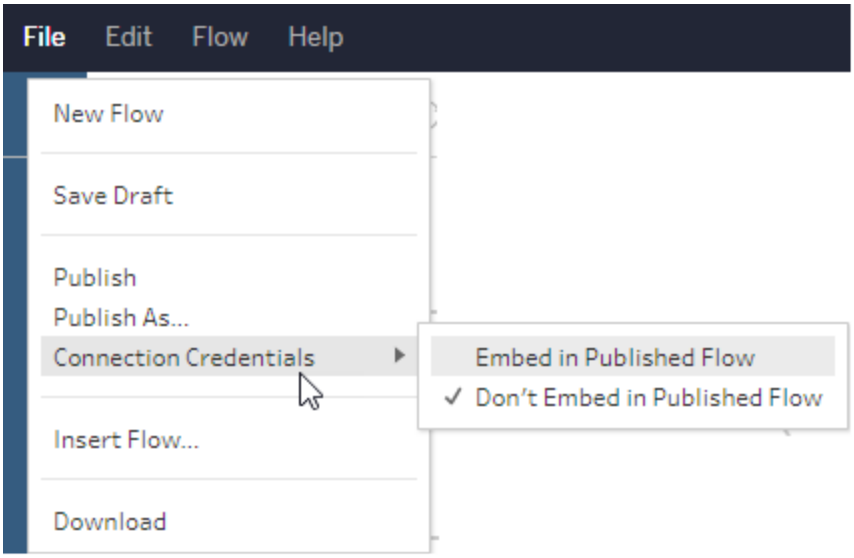
- 您只能将草稿流程发布到已登录的同一服务器。
- 您可以使用**“文件”**菜单并选择**“发布为”**将草稿发布到其他项目。
- 您可以为流程的数据库连接嵌入凭据，以使流程能够运行，而无需在流程运行时手动输入凭据。如果您打开流程进行编辑，您将需要重新输入您的凭据。

嵌入凭据

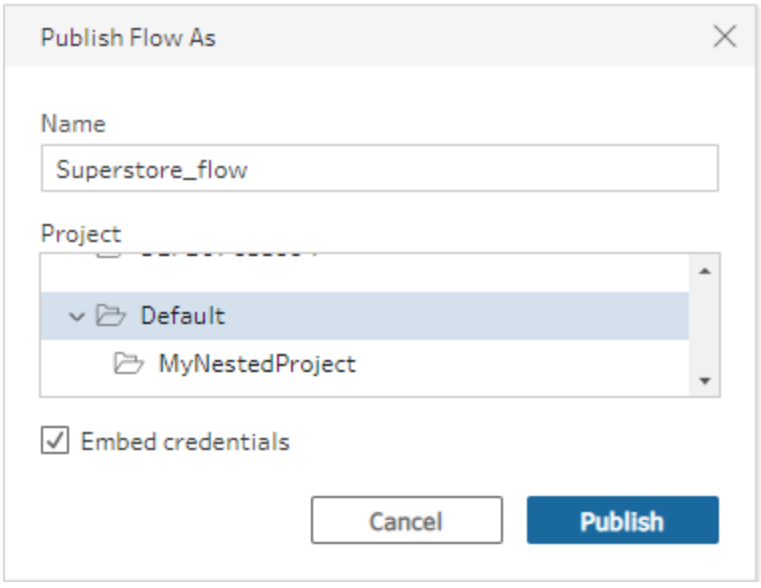
嵌入凭据仅适用于在服务器上运行流程。目前，在编辑连接到数据库的流程时，需要手动输入凭据。嵌入凭据只能在流程级别设置，而不能在服务器或站点级别设置。

执行以下操作之一：

- 从顶部菜单中，选择**“文件”>“连接凭据”>“嵌入在已发布流程中”**。



- 发布流程时，选择**“嵌入凭据”**复选框。此选项显示何时选择**“发布为”**以首次将流程发布到新项目，或者何时编辑上次由其他人发布的流程。

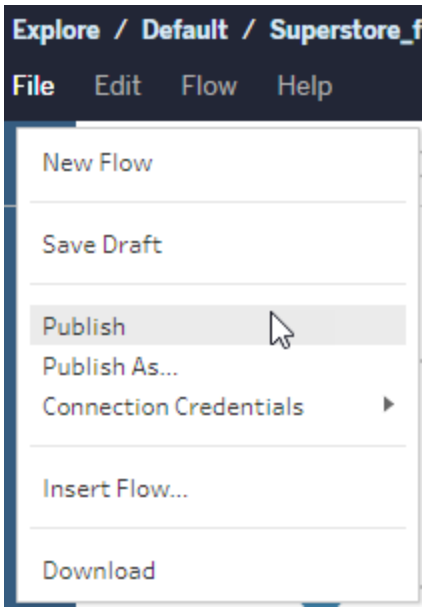


发布流程

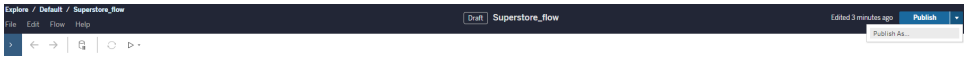
发布流程时，它将成为流程的当前版本，并可以由有权访问您的项目的其他人运行和看到。在发布流程之前，您只能看到从未发布过的流程或对草稿所做的流程更改。有关流程状态的详细信息，请参见自动在 Web 上保存流程。

若要发布您的流程，请执行以下操作之一：

- 从顶部菜单中，选择**“文件”>“发布”**或**“文件”>“发布为”**



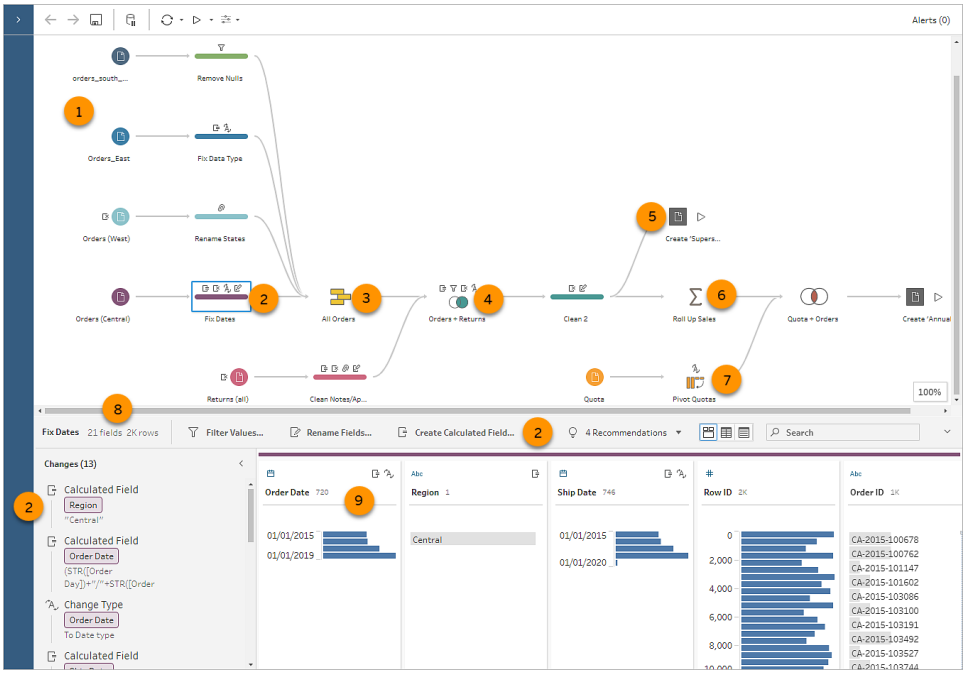
- 在顶部栏中，单击**“发布”**按钮或单击放置箭头以选择**“发布为”**。



谁可以执行此操作

- 服务器管理员、站点管理员 Creator 和 Creator 允许完全连接和发布访问权限。
- Creator 可以执行 Web 制作任务。

Tableau Prep 视觉词典



1	**Input Step** Start your flow by dragging tables or files to the Flow pane to create an Input step. Icons show you the type of data source. Data Source Data Source with input union File (CSV, Excel, Extract, PDF, Spatial, etc) File with input union

2

Clean Step, Changes Pane, Toolbar

Annotations help you keep track of changes you make to your data. While most changes can be made in a Clean step, you can also find these icons on other step types, in the Changes pane, and on the toolbar.

- Calculated Field
- Change Data Type
- Rename Field
- Group Values
- Filter, Exclude, or Keep Only Values
- Remove Fields
- Hidden Field (only visible in List view)
- Recommendations (suggested cleaning options)
- Merge Fields
- Show Profile Pane
- Show Data Grid
- Show List View
- Search

3

Union Step

Combine up to ten sources of data with similar fields in a single Union step.

Union Data

4

Join Step

Connect two steps to join your data on one or more common fields. Select one of these options to choose the join type.

- Full Anti Join
- Inner Join
- Left Inner Join
- Left Outer Join
- Full Outer Join
- Right Inner Join
- Right Outer Join

5	**Output Step** Add an output step to run your flow, apply changes to your complete data set, and generate output data to use as a data source or share with others. File Published Data Source Database or CRM Analytics Run Flow
6	**Aggregate Step** Add an Aggregate step to group and aggregate your data, which changes the level of detail of your data. Aggregate Data
7	**Pivot Step** Add a Pivot step to change columns to rows. Pivot Data

8

Profile Pane

See the exact row count of your data and know when your data is sampled.

Sampled — Only shows when data is sampled

Hover over row summary to see exact row count

9

Profile Card

Change data types and apply cleaning operations to a selected field in the Profile pane.

Data Types

- Number (Decimal and Whole)
- Date
- Date and Time
- Text

Cleaning Operations

- Calculated Field, Split, Duplicate, Clean
- Change Data Type
- Rename Field
- Group Values
- Filter, Exclude, or Keep Only Values
- Remove Fields
- Keep Only Field
- Hidden Field (only visible in List view)
- Merge Fields
- Search

启动或打开数据流程

若要使用 Tableau Prep Builder 开始准备您的数据，您可以：

- 启动新流程
- 打开现有流程

注意：从版本 2020.4.1 开始，您还可以在 Tableau Server 和 Tableau Cloud 中创建和编辑流程。除非特别指出，本主题中的信息适用于所有平台。有关在 Web 上制作流程的详细信息，请参见 Tableau Server 和 Tableau Cloud 帮助中的 **Web 上的 Tableau Prep**。

您可以打开多个 Tableau Prep Builder 工作区以同时处理多个流程。在 Tableau Prep Builder 版本 2019.3.1 及更低版本中，如果选择**“文件”>“打开”**，Tableau Prep Builder 会将当前打开的流程替换为您选择的新流程。

启动新流程

通过连接到数据启动新流程，就像在 Tableau Desktop 中一样。

注意：如果在连接器不受支持的版本中打开流程，流程可能会打开，但是，除非移除了数据连接，否则流程可能会出现错误或无法运行。
某些连接器可能会要求您下载并安装驱动程序，然后才能连接到数据。请参见 Tableau 网站上的驱动程序下载页面获取驱动程序下载链接以及安装说明。

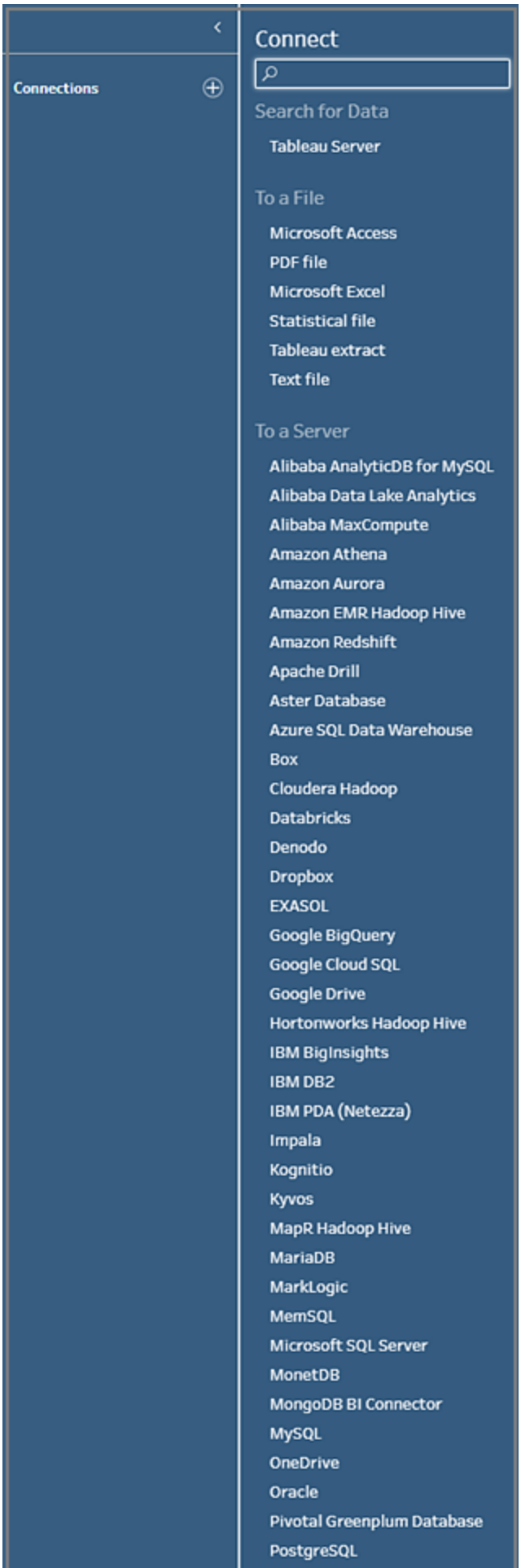
1. 打开 Tableau Prep Builder 并单击**“添加连接”**按钮。

   在 Web 制作中，从**“主页”**中单击**“创建”>“流程”**，或从**“浏览”**页面中单击**“新建”>“流程”**。然后单击**“连接到数据”**。

   从版本 2021.4 开始，如果您有启用了 Catalog 的 Data Management，则还可以从 Web 上的**“外部资产”**页面单击**“新建”>“流程”**，以创建具有支持 Catalog 的连接的流程。有关详细信息，请参见 Tableau Server 或 Tableau Cloud 帮助中的**“Tableau Catalog”**。

2. 从连接器列表中，选择文件类型或托管您的数据的服务器。如果出现提示，请输入所需的信息以登录并访问数据。

注意：在 Web 制作中，文件连接器列表可能不同。

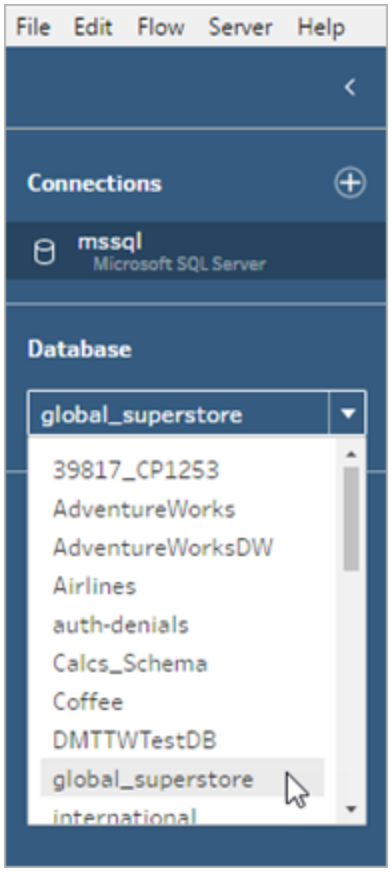

Connections
Connect
Search for Data
Tableau Server
To a File
Microsoft Access
PDF file
Microsoft Excel
Statistical file
Tableau extract
Text file
To a Server
Alibaba AnalyticDB for MySQL
Alibaba Data Lake Analytics
Alibaba MaxCompute
Amazon Athena
Amazon Aurora
Amazon EMR Hadoop Hive
Amazon Redshift
Apache Drill
Aster Database
Azure SQL Data Warehouse
Box
Cloudera Hadoop
Databricks
Denodo
Dropbox
EXASOL
Google BigQuery
Google Cloud SQL
Google Drive
Hortonworks Hadoop Hive
IBM BigInsights
IBM DB2
IBM PDA (Netezza)
Impala
Kognitio
Kyvos
MapR Hadoop Hive
MariaDB
MarkLogic
MemSQL
Microsoft SQL Server
MonetDB
MongoDB BI Connector
MySQL
OneDrive
Oracle
Pivotal Greenplum Database
PostgreSQL

3. 从**“连接”**窗格中，执行以下操作之一：

   - 如果连接到文件，请双击表或将表拖到**“流程”**窗格以开始流程。对于单个表，Tableau Prep 会在您将数据添加到流程时自动为您在**“流程”**窗格中创建输入步骤。

     注意：在 Web 制作中，对于文件连接，您一次只能下载一个文件。当前不支持直接到文件网络共享。

   - 如果连接到数据库，请选择数据库或架构，然后双击表或将表拖到**“流程”**窗格以开始流程。



注意：在 Tableau Prep Builder 中，可以使用通配符搜索将输入步骤中的单个数据源中的多个文件或数据库表合并。在 Web 制作中，您无法创建或编辑输入并集，但在从 Tableau Prep Builder 发布的流程中则支持它

们。有关详细信息，请参见**在输入步骤中合并文件和数据库表** 在本页 123。

打开现有流程

在 Tableau Prep Builder 中，您可以直接在开始页面上查看和访问最近的流程，因此可以轻松地找到正在进行的工作。在 Web 上处理流程时，所有流程都方便地列在**“浏览”**页面的**“所有流程”**菜单下。打开一个流程

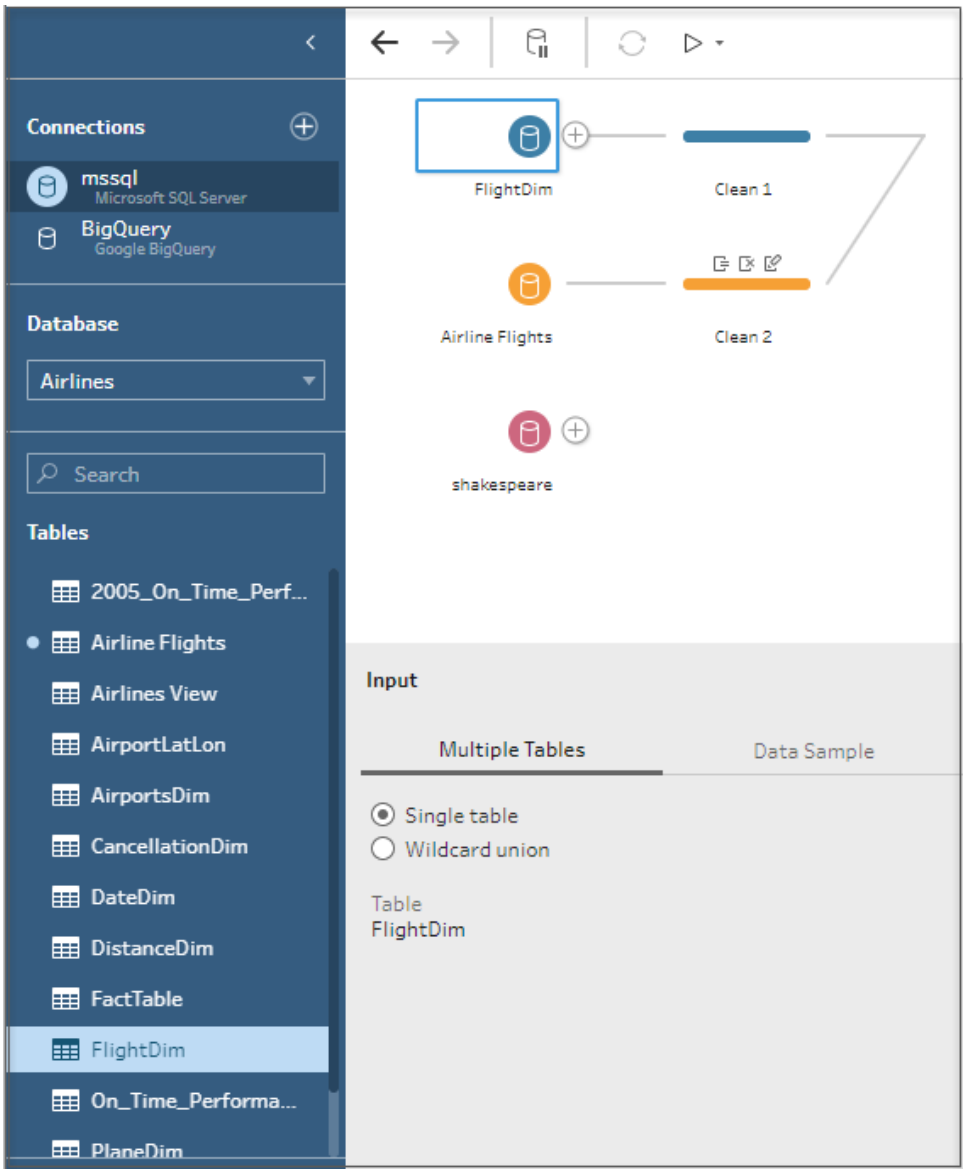
在 Tableau Prep Builder 中打开流程

在**“开始”**页面中执行以下操作之一：

- 在**“最近的流程”**下，选择一个流程。
- 单击**“打开流程”**，导航到流程文件并将其打开。

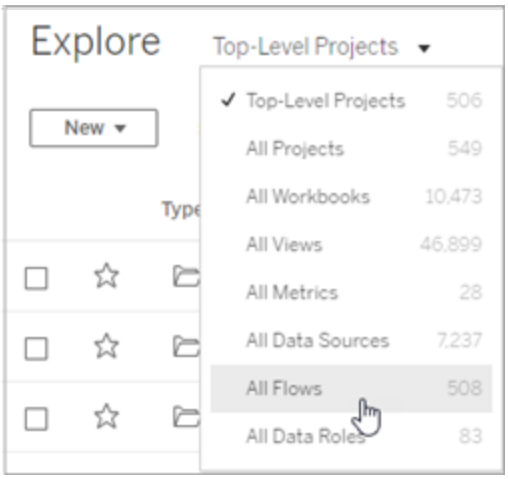
连接到数据之后，在输入步骤中使用不同的选项来确定要在流程中处理的数据。然后您可以添加清理步骤或其他步骤类型来检查、清理和调整数据。

当您的流程包含多个数据源连接器时，Tableau Prep 可帮助您轻松查看哪些连接器和表与输入步骤关联。单击“输入”步骤时，关联的连接器和数据表会在“连接”窗格中突出显示。此选项是 Tableau Prep Builder 版本 2020.1.1 中新增的，在 Web 上编辑流程时也受支持。

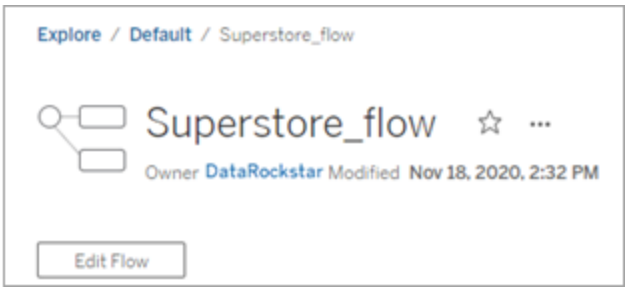


在 Web 版 Tableau Prep 中打开流程

1. 若要打开和编辑现有流程，请在**“浏览”**页面上从顶部下拉菜单中选择**“所有流程”**，然后从列表中选择流程。



2. 在**“流程概述”**页面上，单击**“编辑”**以编辑流程。



您的流程将在新选项卡中打开。一旦您开始进行更改，Tableau 就会每隔几秒钟自动保存一次更改，并将修改后的流程保存为草稿。草稿仅对您和管理员可见。

完成后，可以关闭流程并稍后继续进行更改，或者发布流程以应用更改，从而创建流程的新版本。

与其他工具一样，流程发布使用先入方法。如果其他用户在您之前修改并重新发布流程，则首先提交其更改。但是，您可以使用**“修订历史记录”**页面跟踪并还原到以前的版本。有关详细信息，请参见 Tableau Desktop 帮助中的处理内容修订。

连接到数据

Tableau Prep 可帮助您清理和调整数据以进行分析。此过程的第一步是确定您将使用的数据。

注意：从版本 2020.4.1 开始，您还可以在 Tableau Server 和 Tableau Cloud 中创建和编辑流程。除非特别指出，本主题中的信息适用于所有平台。有关在 Web 上制作流程的详细信息，请参见 Tableau Server 帮助中的 Web 上的 Tableau Prep。

您可以使用以下任一方式连接到您的数据：

- 适用于常见数据类型的内置连接器
- 适用于其他数据类型的自定义连接器
- 已发布数据源
- Tableau 数据提取
- Tableau Catalog

通过适用于常见数据类型的内置连接器进行连接

连接到数据的最常见方法是使用 Tableau Prep Builder 中的内置连接器。这些连接器适用于大多数常见数据类型，并且随着新版本的 Tableau Prep Builder 会经常添加新的连接器。有关可用连接器的列表，请打开 Tableau Prep Builder 或在 Web 上启动流程，并单击**“添加连接”**按钮，以在左侧窗格中的**“连接”**下查看列出的可用连接器。

大多数内置连接器在我们所有平台上的工作方式都相同，并且在 Tableau Desktop 帮助中的支持的连接器主题中进行了描述。

注意：连接到数据源时，不要为字段名称或计算字段使用保留名称“Number of Records”(记录数)。使用此保留名称将导致与权限相关的错误。

使用内置连接器时的注意事项

- 如果在连接器不受支持的版本中打开流程，流程可能会打开，但是，除非移除了数据连接，否则流程可能会出现错误或无法运行。
- 使用基于 MySQL 的连接器时，默认行为是启用 SSL 时连接是安全的。但是，Tableau Prep Builder 不支持为基于 MySQL 的连接器使用基于自定义证书的 SSL 连接。
- 以下部分中详细介绍的一些连接器在与 Tableau Prep Builder 配合使用时有不同的要求。

通过 Tableau Server 或 Tableau Cloud 连接到云数据源

您可以在 Tableau Prep 中像 Tableau Desktop 那样连接到云数据源，但如果计划发布连接到云数据源的流程并在服务器中按计划运行这些流程，则需要在 Tableau Server 或 Tableau Cloud 中配置凭据。

您可以在**“我的帐户设置”**页面的**“设置”**选项卡中设置凭据，然后使用这些相同的凭据连接到云连接器输入。

Tableau Prep Builder

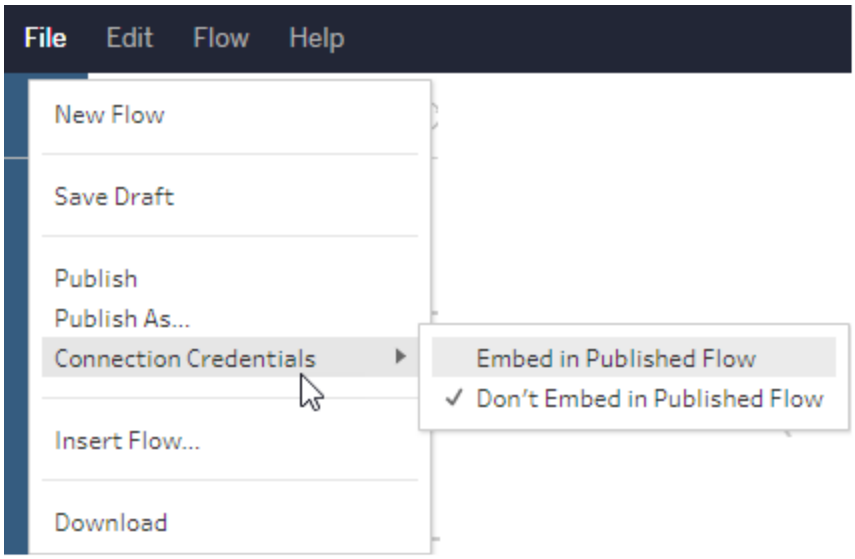
发布流程时，在**“发布”**对话框上单击**“编辑”**以编辑连接，然后在**“身份验证”**下拉列表中，选择**“嵌入”**<您的凭据>。

发布流程时，您还可以通过发布对话框(Tableau Prep Builder 版本 2020.1.1 及更高版本)直接添加凭据，然后在发布时自动将它们嵌入到流程中。有关详细信息，请参见**从 Tableau Prep Builder 中发布流程** 在本页 438。

如果尚未设置已保存的凭据并在**“身份验证”**下拉列表中选择 **“提示用户”**，则在发布流程后，您必须编辑连接并在Tableau Server 或 Tableau Cloud 中的**“连接”**选项卡中输入凭据，否则流程将在运行时失败。

Web 上的 Tableau Prep

在 Web 制作中，可以从顶部菜单中的**“文件”>“连接凭据”**下嵌入凭据。有关详细信息，请参见 Tableau Server 帮助中的发布流程。



在 Tableau Prep Builder 版本 2019.4.1 中，添加了以下云连接器，在 Web 上创建或编辑流程时也可用：

- Box
- DropBox
- Google 云端硬盘
- OneDrive

有关如何使用这些连接器连接到数据的详细信息，请参见 Tableau Desktop 帮助中特定于连接器的帮助主题。

连接到 Salesforce 数据

在 Tableau Prep Builder 版本 2020.2.1 及更高版本中受支持，以及从 Tableau Server 和 Tableau Cloud 版本 2020.4 开始，在 Web 上制作流程时受支持。

Tableau Prep Builder 就像 Tableau Desktop 一样，支持使用 Salesforce 连接器来连接到数据，但有几点不同。

- Tableau Prep Builder 支持您要执行的任何联接类型。
- 可以在 Tableau Prep Builder 2022.1.1 或更高版本中创建自定义 SQL。可以运行使用自定义 SQL 的流程，并且可以在 2020.2.1 或更高版本中编辑现有步骤。
- 当前不支持使用标准连接创建自己的自定义连接。
- 不能将默认数据源名称更改为唯一名称或自定义名称。

- 如果计划将在 Tableau Server 上发布流程并希望使用已保存的凭据，则服务器管理员必须使用 OAuth 客户端 ID 和连接器上的密文配置 Tableau Server。有关详细信息，请参见 Tableau Server 帮助中的将 Salesforce.com OAuth 更改为使用已保存凭据。
- 若要对使用 Salesforce 连接器的流程输入运行增量刷新，您必须使用 Tableau Prep Builder 版本 2021.1.2 或更高版本。有关使用增量刷新的详细信息，请参见**使用增量刷新来刷新流程数据** 在本页390。

Tableau Prep 通过创建数据提取来导入数据。Salesforce 当前仅支持数据提取。初始数据提取可能需要一些时间加载，具体取决于包含的数据量。在加载数据时，您将在“输入”步骤中看到计时器。

有关使用 Salesforce 连接器的详细信息，请参见 Tableau Desktop 和 Web 制作帮助的 Salesforce。

连接到 Salesforce Data Cloud

您可以使用 Salesforce Data Cloud 连接器(于 2023 年 10 月在 Tableau Cloud 中发布)连接到使用 Salesforce Data Cloud 的数据。有关详细信息，请参见将 Tableau Cloud Web 制作连接到 Salesforce Data Cloud。

若要连接到 Salesforce Data Cloud 数据，请使用 Salesforce Data Cloud 连接器。Salesforce Data Cloud 连接器可识别数据空间、呈现更清晰的对象标签并由加速查询提供支持。

1. 在“连接”窗格中，从“服务器连接器”列表中选择**“Salesforce Data Cloud”**。
2. 在**“Salesforce Data Cloud”**对话框中，单击**“登录”**。
3. 使用您的用户名和密码登录到 Salesforce。
4. 选择**“Allow”(允许)**。
5. 关闭浏览器窗格。
6. 在 Tableau Prep 中，选择“数据空间”以查看表。
7. 选择一个表。

连接到 Google BigQuery 数据

就像 Tableau Desktop 一样，Tableau Prep Builder 支持使用 Google BigQuery 连接到数据。

您必须配置凭据，使 Tableau Prep 能够与 Google BigQuery 通信。如果打算将流程发布到 Tableau Server 或 Tableau Cloud，还必须为这些应用程序配置 OAuth 连接。

> **注意：**Tableau Prep 当前不支持使用 Google BigQuery 自定义属性。

- 为 Google 设置 OAuth - 为 Tableau Server 配置 OAuth 连接。
- OAuth 连接 - 为 Tableau Cloud 配置 OAuth 连接。

配置 SSL 以连接到 Google BigQuery(仅限 MacOS)

如果您在 Mac 上使用 Tableau Prep Builder，并且正在使用代理连接到 Big Query，则可能需要修改 SSL 配置以连接到 Google BigQuery

> **注意：**Windows 用户不需要执行额外的步骤。

若要为 Google BigQuery OAuth 连接配置 SSL，请完成以下步骤：

1. 将代理的 SSL 证书导出到文件，例如 proxy.cer。您可以在`Applications > Utilities > Keychain Access >System > Certificates (under Category)`中找到您的证书。
2. 找到用于运行 Tableau Prep Builder 的 java 版本。例如：`/Applications/Tableau Prep Builder 2020.4.app/Plugins/jre/lib/security/cacerts`
3. 打开“终端”命令提示符，并针对您的 Tableau Prep Builder 版本运行以下命令：

   > **注意：**keytool 命令必须从包含用于运行 Tableau Prep Builder 的 java 版本的目录中运行。在运行此命令之前，您可能需要更改目录。例如，`cd /Users/tableau_user/Desktop/SSL.cer -keystore Tableau Prep Builder 2020.1.1/Plugins/jre/bin`。然后运行 keytool 命令。

   ```
   keytool -import -trustcacerts -file /Users/tableau_user/Desktop/SSL.cer -keystore Tableau Prep Builder <version>/Plugins/jre/lib/security/cacerts -storepass changeit
   ```

   示例： `keytool -import -trustcacerts -file /Users/tableau_user/Desktop/SSL.cer -keystore Tableau Prep Builder 2020.4.1/Plugins/jre/lib/security/cacerts -storepass changeit`

如果在运行 keytool 命令时收到 **FileNotFoundexception**(拒绝访问)，请尝试使用提升的权限运行该命令。例如：`sudo keytool -import -trustcacerts -file /Users/tableau_user/Desktop/SSL.cer -keystore Tableau Prep Builder 2020.4.1/Plugins/jre/lib/security/cacerts -storepass changeit`。

设置和管理您的 Google BigQuery 凭据

在“输入”步骤中用于连接到 Google BigQuery 的凭据必须与您在 Tableau Server 或 Tableau Cloud 中适用于 Google BigQuery 的**“我的帐户设置”**页面的**“设置”**选项卡中设置的凭据匹配。

如果发布流程时在身份验证设置中选择不同的凭据或不选择凭据，则在 Tableau Server 或 Tableau Cloud 中编辑流程连接之前，该流程将失败。

若要编辑凭据，请执行以下操作：

1. 在 Tableau Server 或 Tableau Cloud 中的**“连接”**选项卡上，针对 Google BigQuery 连接单击**“更多操作”** ··· 。
2. 选择**“编辑连接”**。
3. 选择在**“我的帐户设置”**页面的**“设置”**选项卡中设置的已保存凭据。

使用服务帐户 (JSON) 文件登录

在 Tableau Prep Builder 版本 2021.3.1 及更高版本中受支持。在 Web 上制作流程时，服务帐户访问权限不可用。

1. 添加服务帐户作为保存的凭据。有关详细信息，请参见将 Google OAuth 更改为已保存凭据。
2. 使用您的电子邮件或电话号码登录 Google BigQuery，然后选择**“下一步”**。
3. 在“身份验证”中，选择**“使用服务帐户 (JSON) 文件登录”**。
4. 输入文件路径或使用**“浏览”**按钮进行搜索。
5. 单击**“登录”**。
6. 输入您的密码以继续。
7. 选择**“接受”**，以允许 Tableau 访问您的 Google BigQuery 数据。系统将提示您关闭浏览器。

使用 OAuth 登录

在 Tableau Prep Builder 版本 2020.2.1 及更高版本中受支持，以及从 Tableau Server 和 Tableau Cloud 版本 2020.4 开始，在 Web 上制作流程时受支持。

1. 使用您的电子邮件或电话号码登录 Google BigQuery，然后选择**“下一步”**。
2. 在“身份验证”中，选择**“使用 OAuth 登录”**。
3. 单击**“登录”**。
4. 输入您的密码以继续。

5. 选择**“接受”**，以允许 Tableau 访问您的 Google BigQuery 数据。系统将提示您关闭浏览器。

有关设置和管理凭据的详细信息，请参见以下主题：

Tableau Desktop 和 Web 制作帮助中的管理您的帐户设置。

从 Tableau Prep Builder 中发布流程 在本页438，了解有关发布流程时设置身份验证选项的信息。

查看和解决错误，了解有关在 Tableau Server 或 Tableau Cloud 中解决连接器错误的信息。

连接到 SAP HANA 数据

在 Tableau Prep Builder 版本 2019.2.1 及更高版本中受支持，以及从 Tableau Server 和 Tableau Cloud 版本 2020.4 开始，在 Web 上制作流程时受支持。

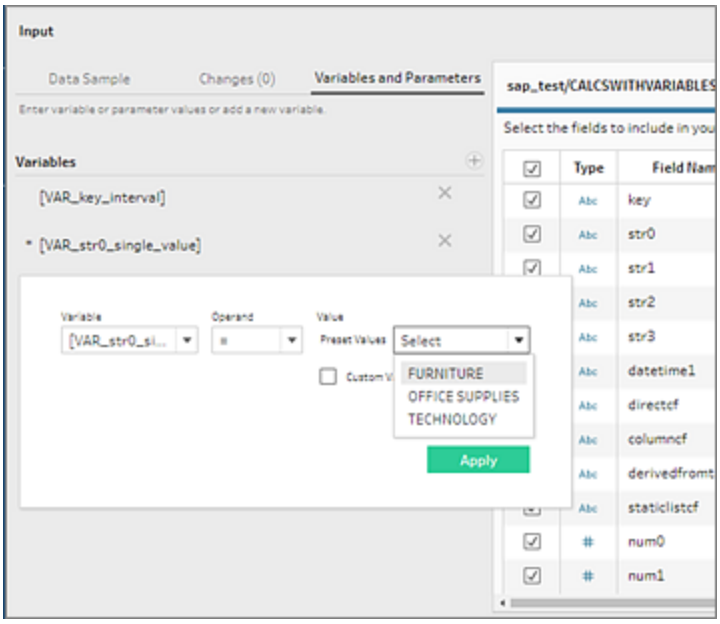
Tableau Prep Builder 就像 Tableau Desktop 一样，支持使用 SAP HANA 连接到数据，但有几点不同。

使用在 Tableau Desktop 中使用的相同过程连接到数据库。有关详细信息，请参见SAP HANA。连接并搜索表之后，将表拖到画布上以开始构建流程。

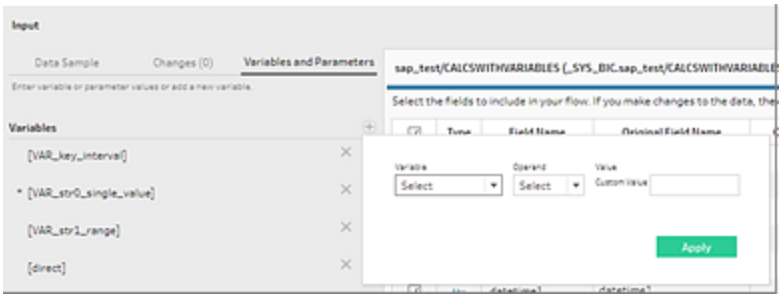
不支持在 Tableau Prep 中打开流程时的变量和参数提示。相反，在“输入”窗格中，单击**“变量和参数”**选项卡并选择要使用的变量和操作数，然后从预设值列表中选择，或者输入自定义值来查询数据库并返回所需的值。

连接到 SAP HANA 已发布数据源时，Tableau Prep Builder 和 Tableau Prep Conductor 不支持变量和参数。

> **注意**：从 Tableau Prep Builder 版本 2019.2.2 以及在 Web 上从 2020.4.1 开始，您可以使用初始 SQL 来查询您的连接。如果有多个变量值，您可以从下拉列表中选择所需的值。



也可以添加其他变量。单击**"变量"**部分中的加号按钮 ⊕ 并选择一个变量或操作数，然后输入自定义值。



注意：此连接器需要 Tableau Server 版本 2019.2 及更高版本才能按计划运行流程。如果您使用的是较早的服务器版本，则可以使用命令行界面刷新流程数据。有关如何通过命令行运行流程的详细信息，请参见**通过命令行刷新流程输出文件** 在本页 397。有关版本兼容性的详细信息，请参见**与 Tableau Prep 的版本兼容性** 在本页 416。

连接到空间文件和数据库

在 Tableau Prep Builder 版本 2020.4.1 及更高版本中受支持，以及从 Tableau Server 和 Tableau Cloud 版本 2020.4 开始，在 Web 上制作流程时受支持。

您可以在 Tableau Prep Builder 中或在 Web 上创建或编辑渡海中时连接到空间文件和空间数据源。

Tableau Prep 支持以下连接类型：

- 空间文件格式
  - Tableau Prep Builder：Esri Shapefile、Esri 文件地理数据库、KML、TopoJSON、GeoJSON、数据提取、MapInfo MID/MIF、TAB 文件和压缩的 shapefile。
  - Tableau Server 和 Tableau Cloud：压缩的 shapefile、KML、TopoJSON、GeoJSON、Esri 文件地理数据库和数据提取。
- 空间数据库（Amazon Redshift、Microsoft SQL Server、Oracle 和 PostgreSQL）。

还可以使用标准联接将空间表与非空间表合并，并将空间数据输出到数据提取 (.hyper) 文件。当前不支持 Tableau Prep 中地图视图早的空间函数、通过相交进行空间联接和可视化空间数据。

支持的清理操作

使用形状文件数据时，不支持某些清理操作。在使用形状文件数据时，Tableau Prep 中只有以下清理操作可用。

- 筛选器：仅移除 Null 或未知值
- 重命名字段
- 复制字段
- 仅保留字段
- 移除字段
- 创建计算字段

连接之前

在连接到空间文件之前，请确保以下文件在同一目录中：

- **Esri Shapefile：**文件夹必须包含 .shp、.shx、.dbf 和 .prj 文件，以及 Esri shapefile 的 .zip 文件。
- **Esri 文件地理数据库：**文件夹必须包含文件地理数据库的 .gdb 或者文件地理数据库的 .gdb 的 .zip。
- **KML 文件**：文件夹必须包含 .kml 文件。（不需要其他文件。）
- **GeoJSON 文件**：文件夹必须包含 .geojson 文件（不需要其他文件）。
- **TopoJSON 文件：**文件夹必须包含 .json 或 .topojson 文件。（不需要其他文件。）

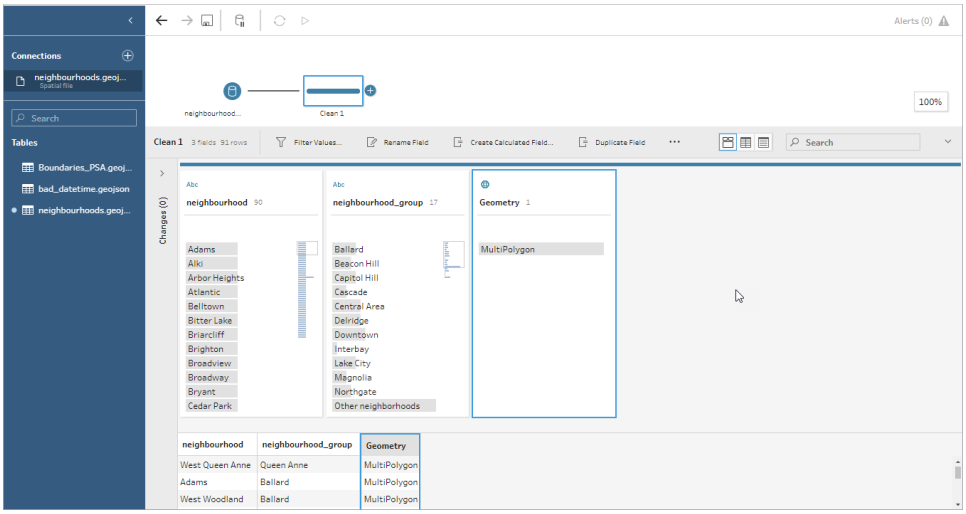
连接到空间文件

1. 执行以下操作之一：

   - 打开 Tableau Prep Builder 并单击**“添加连接”**按钮。
   - 打开 Tableau Server 或 Tableau Cloud。从**“浏览”**菜单中，单击**“”>“流程”**。

2. 从连接器列表中，选择**“空间文件”**。

   将为空间字段分配空间数据类型，并且无法更改。如果字段来自空间文件，则会为该字段分配默认字段名称“Geometry”。如果字段来自空间数据库，将显示数据库字段名称。如果 Tableau 无法确定数据类型，则该字段将显示为“Null”。



使用 ODBC 进行连接

在 Tableau Prep Builder 版本 2019.2.2 及更高版本中受支持。在 Web 上制作流程时，此连接器类型尚不受支持。

如果需要连接到**“连接”**窗格中未列出的数据源，您可以使用支持 SQL 标准并实现 ODBC API 的**“其他数据库 (ODBC)”**连接器连接到任何数据源。使用**“其他数据库 (ODBC)”**连接器连接到数据的工作方式与您在 Tableau Desktop 中使用该连接器的方式类似，但有一些不同之处：

- 您只能使用 DSN(数据源名称)选项进行连接。
- 若要在 Tableau Server 中发布并运行流程，必须使用匹配的 DSN 配置服务器。

> **注意：**目前不支持从包含“其他数据库 (ODBC)”连接器的命令行中运行流程。

- Windows 和 MacOS 上的连接体验是一致的。不支持 ODBC 驱动程序 (Windows) 的提示连接属性。
- Tableau Prep Builder 仅支持 64 位驱动程序。

连接之前

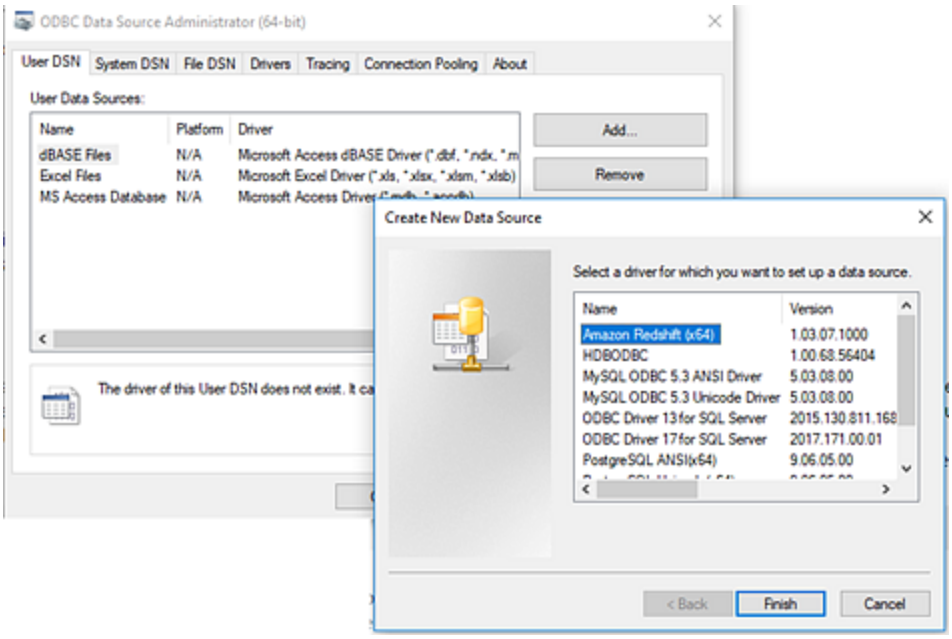
若要使用**“其他数据库 (ODBC)”**连接器连接到数据，您必须安装数据库驱动程序，并设置和配置 DSN（数据源名称）。若要将流程发布到 Tableau Server 并运行流程，也必须使用匹配的 DSN 配置服务器。

重要信息：Tableau Prep Builder 仅支持 64 位驱动程序。如果已设置并配置了 32 位驱动程序，则您可能需要将其卸载，然后安装 64 位版本（如果驱动程序不允许同时安装两个版本）。

1. 使用 **ODBC 数据源管理器（64 位）** (Windows) 或 ODBC 管理器实用工具 (MacOS) 创建 DSN。

   如果 Mac 上未安装该实用工具，您可以下载一个（例如从 www.odbcmanager.net 下载），或者可以手动编辑 odbc.ini 文件。

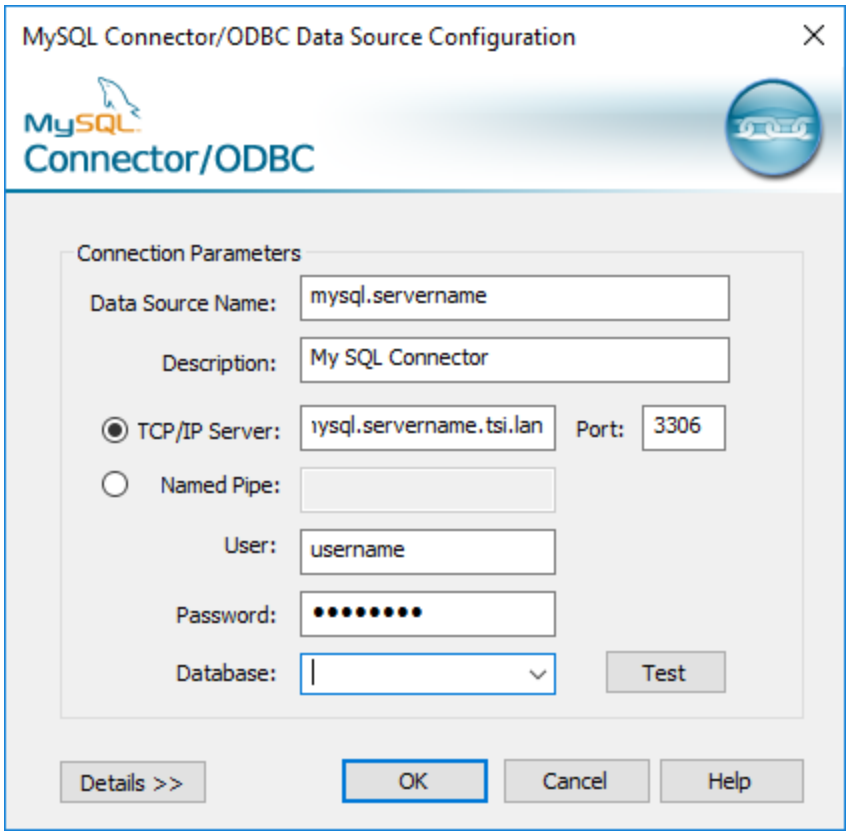
2. 在 **ODBC 数据源管理器（64 位）** (Windows) 或 ODBC 管理器实用工具 (MacOS) 中，添加一个新数据源，然后选择适用于数据源的驱动程序，并单击**“完成”**。



3. 在**“ODBC 驱动程序设置”**对话框中，输入诸如服务器名称、端口、用户名和密码之类的配置信息。单击**“测试”**（如果对话框有该选项）验证您的连接是否设置正确，然后保存您的配置。

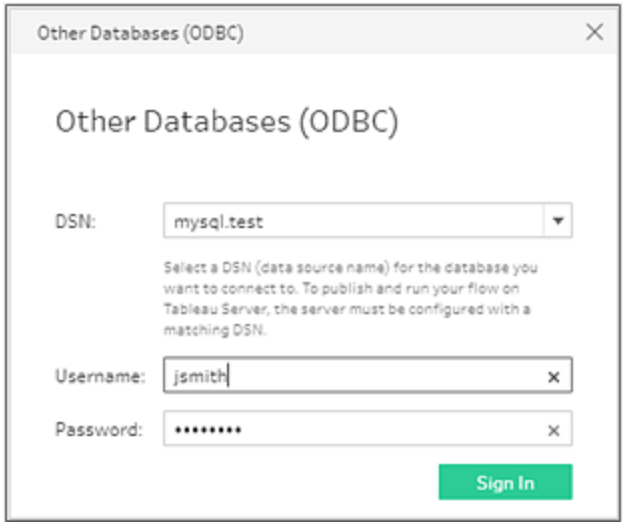
   注意：Tableau Prep Builder 不支持提示连接属性，因此您必须在配置 DNS 时设置此信息。

   此示例显示 MySQL 连接器的配置对话框。

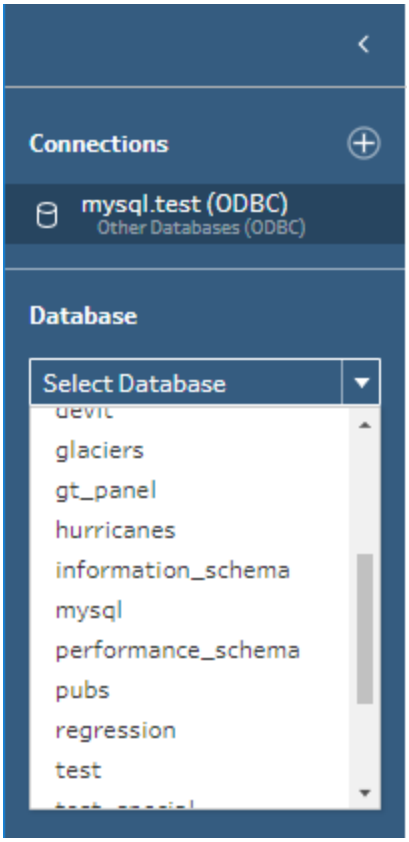


使用“其他数据库 (ODBC)”进行连接

1. 打开 Tableau Prep Builder 并单击**“添加连接”**按钮。
2. 从连接器列表中，选择**“其他数据库 (ODBC)”**。
3. 在**“其他数据库 (ODBC)”**对话框中，从下拉列表中选择一个 DSN，并输入用户名和密码。然后单击**“登录”**。



4. 从**"连接"**窗格中，从下拉列表中选择您的数据库。



连接到 Microsoft Excel 数据并使用 Data Interpreter 进行清理

仅支持直接 Microsoft Excel 连接。Data Interprete目前不适用于存储在云驱动器中的 Excel 文件。

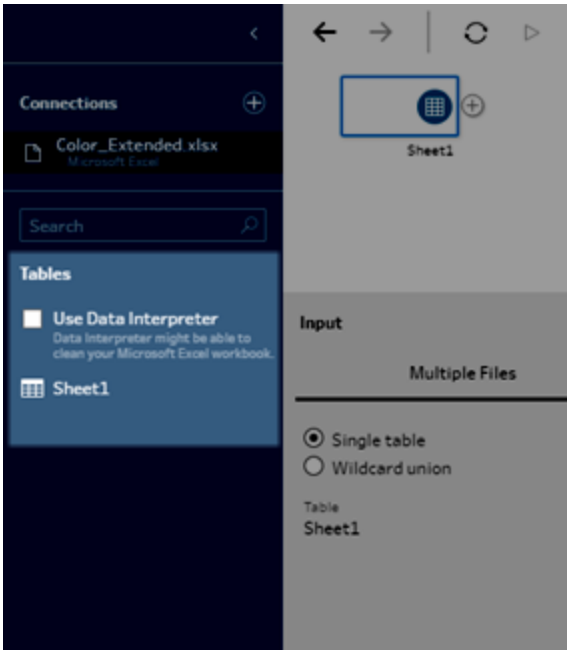
处理 Microsoft Excel 文件时，您可以使用 Data Interpreter 来检测数据中的子表，以及移除无关信息来帮助准备数据进行分析。打开 Data Interpreter 时，它将检测这些子表，并将其列为**"连接"**窗格的**"表"**部分中的新表。然后，您可以将它们拖到**"流程"**窗格中。

如果关闭 Data Interpreter，则会从**"连接"**窗格中移除这些表。如果已在流程中使用了这些表，则会由于缺少数据而导致流程错误。

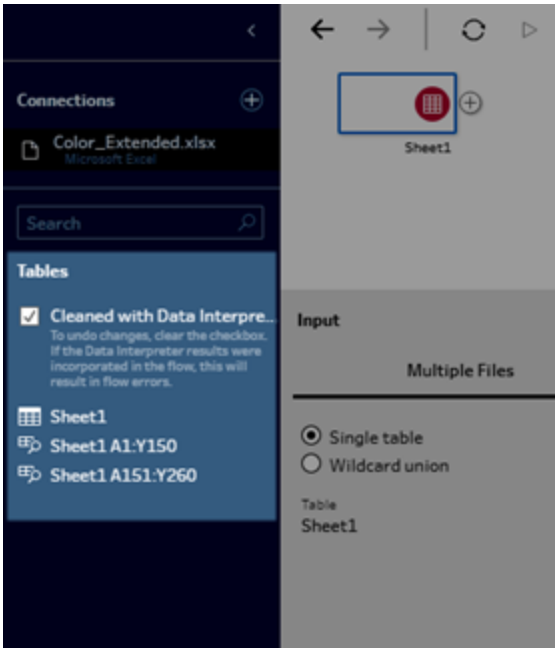
注意：当前 Data Interpreter 只会检测 Excel 电子表格中的子表，不支持为文本文件和电子表格指定起始行。此外，Data Interpreter 检测到的表不包括在通配符并集搜索结果中。

下面的示例显示在 **"连接"**窗格中对 Excel 电子表格使用 Data Interpreter 的结果。Data Interpreter 检测到两个额外的子表。

使用 Data Interpreter 之前



使用 Data Interpreter 之后



若要使用 Data Interpreter，请完成以下步骤：

1. 选择**"连接到数据"**，然后选择**"Microsoft Excel"**。
2. 选择您的文件，并单击**"打开"**。

3. 选中**“使用 Data Interpreter”**复选框。
4. 将新表拖到**“流程”**窗格以将其包括在流程中。若要移除旧表，请右键单击旧表的输入步骤，并选择**“移除”**。

使用自定义连接器进行连接

当 Tableau Prep 没有为基于 ODBC 和 JDBC 的数据提供内置连接器时，您可以使用自定义连接器。您可以：

- 使用合作伙伴构建的连接器。有关 Exchange 中的连接器的详细信息，请参见**使用合作伙伴构建的连接器** 向下。
- 使用通过 Tableau 连接器 SDK 构建的自定义连接器连接器 SDK 提供了为基于 ODBC 或 JDBC 的数据构建自定义连接器的工具。有关详细信息，请参见 Tableau Desktop 帮助中的使用 Tableau 连接器 SDK 构建的连接器。

Tableau Prep Builder 版本 2020.4.1 及更高版本支持用于基于 ODBC 和 JDBC 的数据的自定义连接器。

对于连接器开发人员，请参见运行您的连接器以了解更多信息。

某些自定义连接器需要安装额外的驱动程序。如果在连接过程中出现提示，请按照提示下载并安装所需的驱动程序。自定义连接器当前不能与 Tableau Cloud 一起使用。

使用合作伙伴构建的连接器

“连接”窗格中提供了合作伙伴构建的连接器或其他自定义连接器。这些连接器列在“其他连接”下，也可从 Tableau Exchange 连接器页面获得。

1. 单击左侧窗格中的**“连接”**。
2. 从“连接”窗格中的**“其他连接器”**部分，单击要使用的连接器。
3. 单击**“安装并重新启动 Tableau”**。

   安装连接器后，它将出现在“连接”窗格的**“到服务器”**部分。

注意：如果您收到连接器无法加载的警告，请从 Tableau Exchange 连接器页面安装您需要的 .taco 文件。如果系统提示您安装驱动程序，请转到 Tableau Exchange 了解驱动程序下载说明和位置。

连接到发布的数据源

已发布数据源是您可以与他人共享的数据源。当您希望将数据源提供给其他用户时，您可以将其从 Tableau Prep Builder（版本 2019.3.1 及更高版本）发布到 Tableau Server 或 Tableau Cloud，或作为流程的输出发布。

无论您是在 Tableau Prep Builder 中还是在 Web 上工作，您都可以使用已发布数据源作为流程的输入数据源。

注意：在发布包括发布的数据源作为输入的流程时，系统会将发布者指派为默认流程所有者。当流程运行时，它将为**“运行身份”**帐户使用流程所有者。有关“运行身份”帐户的详细信息，请参见运行身份服务帐户。只有站点或服务器管理员才能在 Tableau Server 或 Tableau Cloud 中更改流程所有者，并且只能更改为自身。

Tableau Prep Builder 支持：

- 包含用户筛选器或函数的已发布数据源（从 Tableau Prep Builder 版本 2021.1.3 开始）。
- 与单一服务器和站点的连接。不支持登录到不同的服务器或同一服务器和不同的站点。必须使用相同的服务器或站点连接来执行以下操作：
  - 连接到发布的数据源。
  - 将流程输出发布到 Tableau Server 或 Tableau Cloud。
  - 计划流程以在 Tableau Server 或 Tableau Cloud上运行。

  如果您的流程使用已发布数据源并且您注销服务器，这将中断流程连接。流程将处于错误状态，您将无法在配置窗格或数据网格中看到来自已发布数据源的数据。

注意：Tableau Prep Builder 不支持包含多维（多维数据集）数据、多服务器连接或具有相关表的已发布数据源的已发布数据源。

Tableau Server 和 Tableau Cloud 支持：

- 包含用户筛选器或函数的已发布数据源（从 Tableau Server 和 Tableau Cloud 版本 2021.2 开始）。
- 使用已发布数据源（Tableau Server 或 Tableau Cloud 版本 2020.4 及更高版本）在 Web 上创建或编辑流程
- 连接到已发布数据源（Tableau Server 和 Tableau Cloud 版本 2019.3 及更高版本）

  注意：早期版本的 Tableau Server 可能不支持已发布数据源的所有功能。

关于凭据和权限：

- 您必须在您登录到的服务器站点中分配有 Explorer 或更高角色，才能连接到已发布数据源。只有 Creator 才能在 Web 上创建或编辑流程。有关站点角色的详细信息，请参见 Tableau Server 帮助中的设置用户的站点角色。
- 在 Tableau Prep Builder 中，数据源访问权限是基于登录到服务器的用户的身份授予的。您将只看到您有权访问的数据。
- 在 Prep Web 制作（Tableau Server 和 Tableau Cloud）中，数据源访问权限也是基于登录到服务器的用户的身份授予的。您将只看到您有权访问的数据。

  但是，当您手动或使用计划运行流程时，将根据流程所有者的身份授予数据源访问权限。最后一个发布流程的用户成为新的流程所有者。
- 站点和服务器管理员可以更改流程所有者，但只能更改为他们自己。
- 必须嵌入凭据才能连接到已发布的数据源。

  提示：如果没有为数据源嵌入凭据，请更新数据源以包括嵌入的凭据。

在流程中使用已发布数据源

若要连接到已发布数据源并在流程中使用它，请按照 Tableau Prep 版本的说明进行操作：

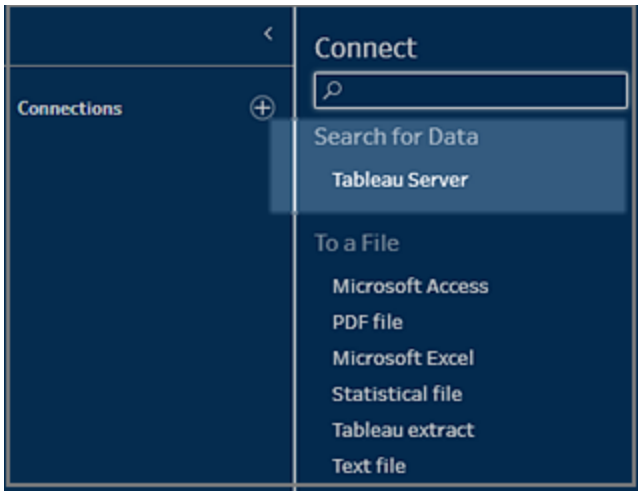
Tableau Prep Builder 版本 2020.2.2 及更高版本以及在 Web 上

您可以直接从“连接”窗格连接到存储在 Tableau Server 或 Tableau Cloud 上的已发布数据源及更多数据源。如果有启用了 Tableau Catalog 的 Data Management，则还可以搜索和连接到数据库和表，以及查看或按有关数据源的元数据（例如说明、数据质量警告和认证）进行筛选。

有关 Tableau Catalog 的详细信息，请参见 Tableau Server 或 Tableau Cloud 帮助中的“关于 Tableau Catalog”。

1. 打开 Tableau Prep Builder 并单击**“添加连接”**按钮。

   在 Web 制作中，从**“主页”**中单击**“创建”>“流程”**，或从**“浏览”**页面中单击**“新建”>“流程”**。然后单击**“连接到数据”**。
2. 在**“连接”**窗格上的**“搜索数据”**下，选择**“Tableau Server”**。



3. 登录以连接到服务器或站点。

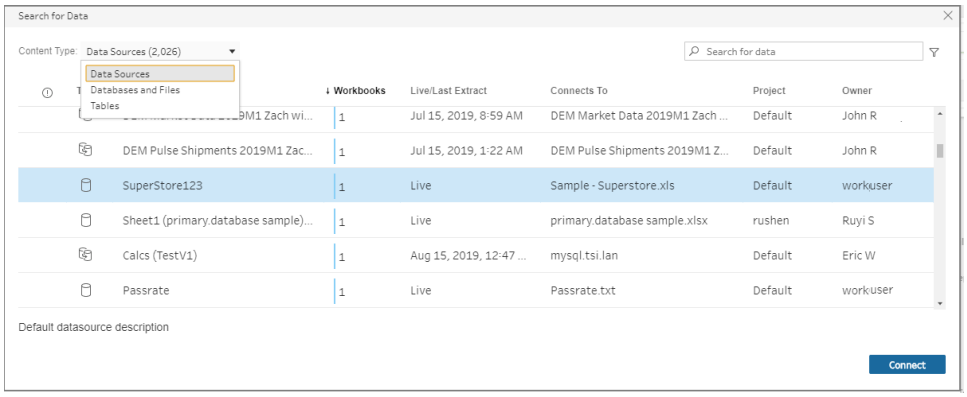
   在 Web 制作中，**“搜索数据”**对话框将针对您登录的服务器打开。

4. 在**“搜索数据”**对话框中，从可用已发布数据源的列表中进行搜索。使用筛选器选项按连接类型和认证数据源进行筛选。

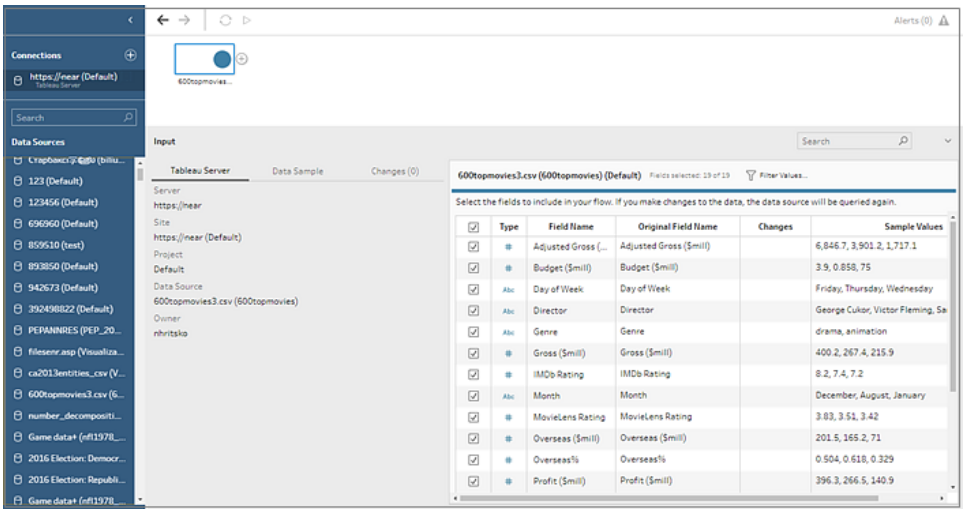
5. 选择要使用的数据源，然后单击**“连接”**。

   如果您没有连接到数据源的权限，则行和**“连接”**按钮显示为灰色。

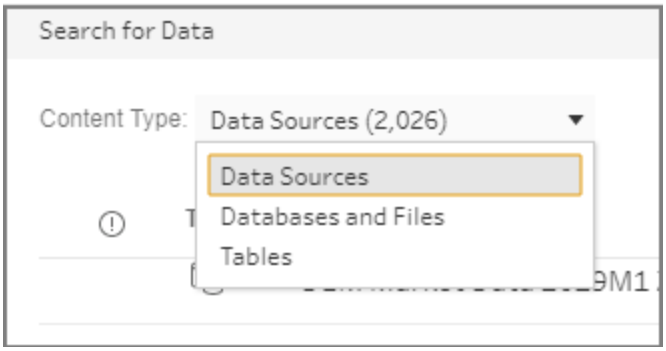
   > **注意**：如果没有启用了 Tableau Catalog 的 Data Management，则不会显示**“内容类型”**下拉列表。列表中只会显示已发布数据源。



6. 数据源已添加到“流程”窗格中。在**“连接”**窗格中，您可以选择其他数据源或使用搜索选项查找数据源，并将其拖到流程窗格以构建流程。“输入”窗格中的**“Tableau Server”**选项卡显示有关已发布数据源的详细信息。



7. (可选)如果有启用了 Tableau Catalog 的 Data Management，请使用**"内容类型"**下拉列表来搜索数据库和表。



您可以使用右上角的筛选器选项按连接类型、数据质量警告和认证来筛选结果。

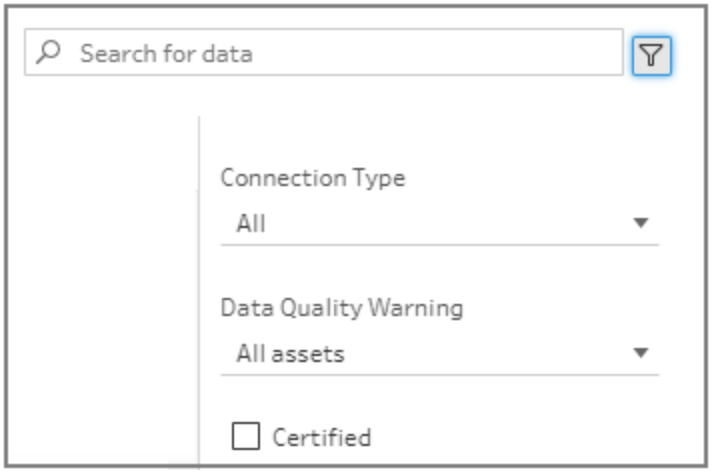
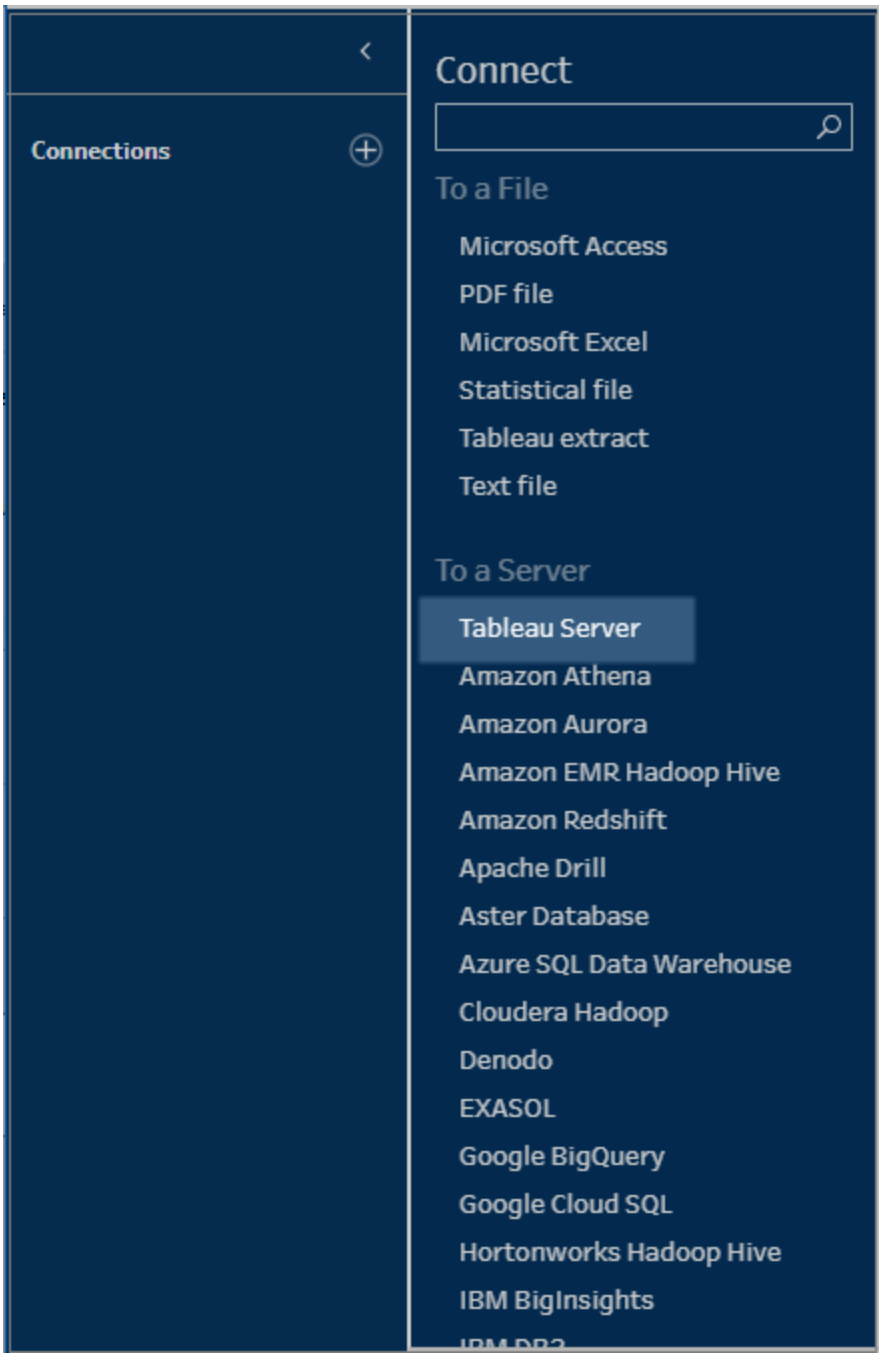
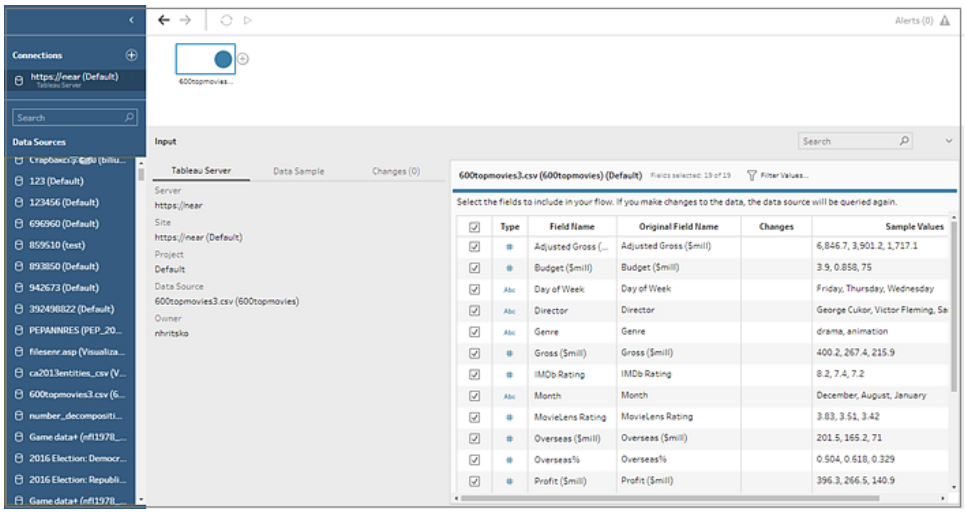


Tableau Prep Builder 版本 2020.2.1 及更高版本

1. 打开 Tableau Prep Builder 并单击**“添加连接”**按钮。
2. 从连接器列表中，选择**“Tableau Server”**。



3. 登录以连接到服务器或站点。
4. 选择数据源或使用搜索选项查找数据源，并将其拖到流程窗格以启动流程。“输入”窗格中的**“Tableau Server”**选项卡显示有关已发布数据源的详细信息。



连接到虚拟连接

在 Tableau Prep Builder 版本 2021.4.1 及更高版本以及 Tableau Server 和 Tableau Cloud 版本 2021.4 及更高版本中受支持。使用此功能需要 Data Management。

您可以为流程使用虚拟连接来连接到数据。虚拟连接是一种可共享的资源，可提供对数据的中央访问点。

连接到虚拟连接时的注意事项：

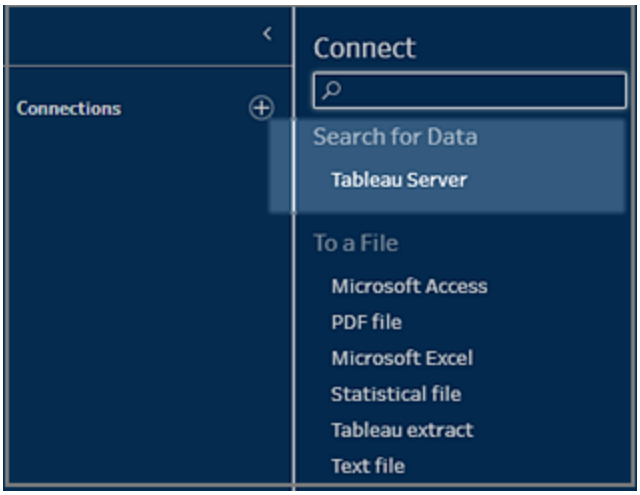
- 数据库凭据嵌入在虚拟连接中。您只需登录您的服务器即可访问虚拟连接中的表。
- 应用行级安全性的数据策略可以包含在虚拟连接中。使用和运行流程时，只会显示您有权访问的表、字段和值。
- 虚拟连接中的行级安全性不适用于流程输出。有权访问流程输出的所有用户都会看到相同的数据。
- 不支持自定义 SQL 和初始 SQL。
- 不支持参数。有关在流程中使用参数的详细信息，请参见在流程中创建和使用参数 在本页 192。

有关虚拟连接和数据策略的详细信息，请参见 Tableau Server 或 Tableau Cloud 帮助。

1. 打开 Tableau Prep Builder 并单击**“添加连接”**按钮。

   在 Web 制作中，从**“主页”**中单击**“创建”>“流程”**，或从**“浏览”**页面中单击**“新建”>“流程”**。然后单击**“连接到数据”**。

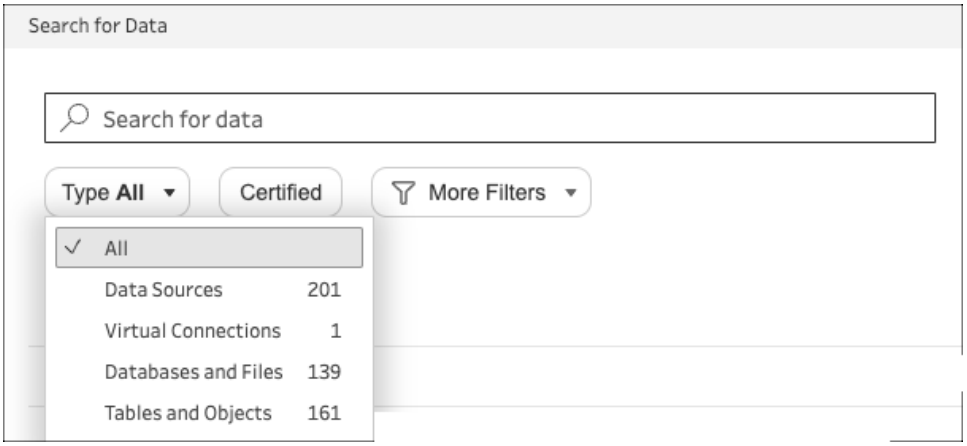
2. 在**“连接”**窗格上的**“搜索数据”**下，选择**“Tableau Server”**。



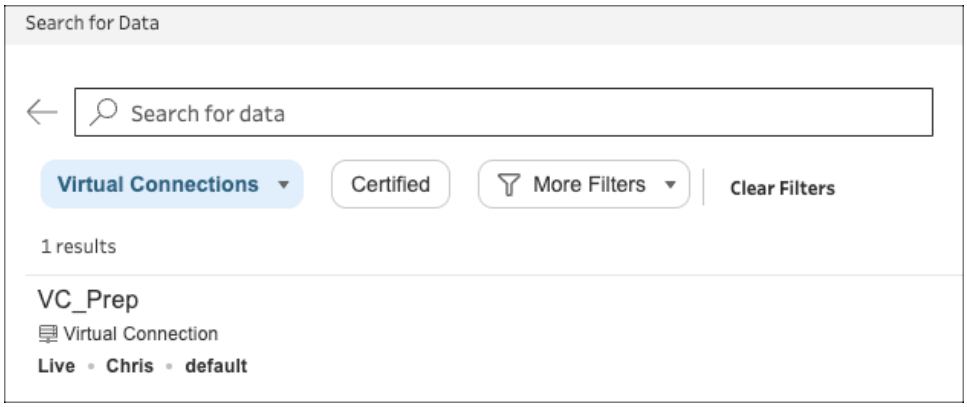
3. 登录以连接到服务器或站点。

   在 Web 制作中，**“搜索数据”**对话框将针对您登录的服务器打开。

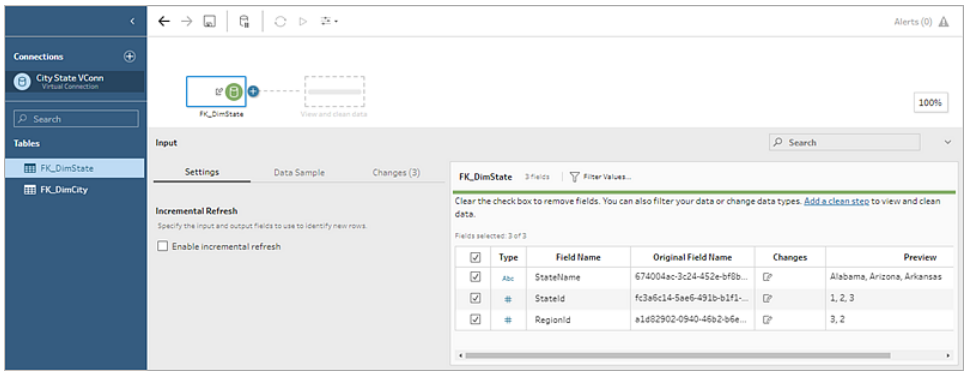
4. 单击**“输入全部”**。



5. 选择**“虚拟连接”**。



6. 选择要使用的数据源，然后单击**“连接”**。
7. 数据源已添加到“流程”窗格中。在**“连接”**窗格中，您可以从虚拟连接中包含的表列表中进行选择，然后将它们拖到流程窗格中以开始您的流程。



注意：如果您在连接到虚拟连接时在**“更改”**窗格中看到**“重命名”**操作，请不要移除这些操作。Tableau Prep 会自动生成这些操作以映射并显示字段的用户友好名称。

连接到 Tableau 数据提取

您可以连接到数据提取作为数据流程的输入。数据提取是您可通过使用筛选器和配置其他限制来创建的已保存数据子集。数据提取保存为 .hyper 文件。

有关将数据提取与 Tableau Prep Builder 配合使用的详细信息，请参见保存和共享工作 在本页361。

通过 Tableau Catalog 连接到数据

如果有启用了 Tableau Catalog 的 Data Management，您还可以搜索并连接到存储在 Tableau Server 或 Tableau Cloud 上的数据库和表。

有关 Tableau Catalog 的详细信息，请参见 Tableau Server 或 Tableau Cloud 帮助中的“关于 Tableau Catalog”。

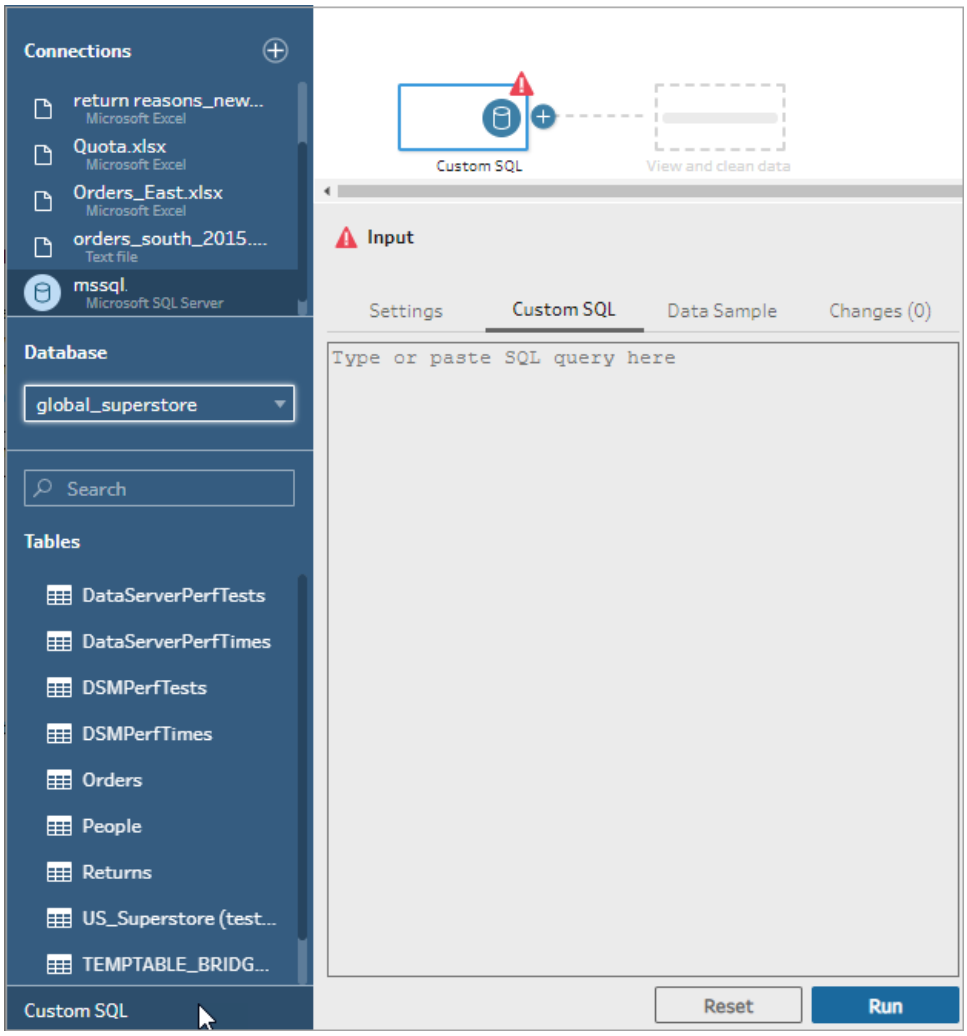
其他连接选项

进行连接时，您可能还会看到以下选项，具体取决于您选择的连接。

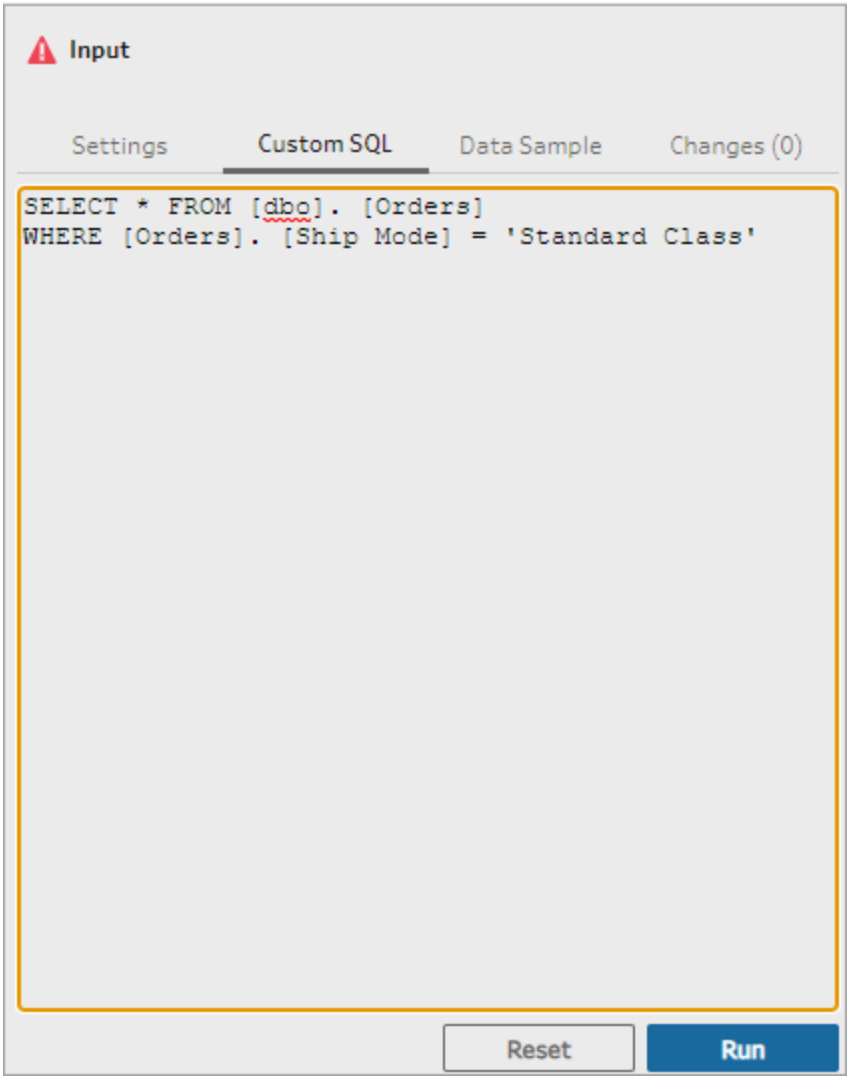
使用自定义 SQL 连接到数据

如果您确切地知道数据库中需要的信息并了解如何编写 SQL 查询，可以使用自定义 SQL 查询连接到数据，就像在 Tableau Desktop 中一样。您可以使用自定义 SQL 跨表合并数据、重新转换字段以执行跨数据库联接、重构或减小数据大小以进行分析等。

1. 连接到您的数据源，并在“连接”窗格的“数据库”字段中，选择一个数据库。
2. 单击**“自定义 SQL”**链接以打开“自定义 SQL”选项卡。



3. 键入查询或将查询粘贴到文本框中，然后单击“运行”以运行查询。



4. 在流程窗格中添加一个清理步骤，以便仅将自定义 SQL 查询中的相关字段添加到流程中。

使用初始 SQL 查询连接

在 Tableau Prep Builder 版本 2019.2.2 及更高版本中受支持，以及从版本 2020.4.1 开始，在 Web 上制作流程时受支持

您可以指定初始 SQL 命令，该命令将在连接到支持该命令的数据库时运行。例如，在连接到 Amazon Redshift 时，您可以输入 SQL 语句，以便在连接到数据库时应用筛选器，就像在“输入”步骤中添加筛选器一样。SQL 命令将在数据抽样之前应用，并加载到 Tableau Prep 中。

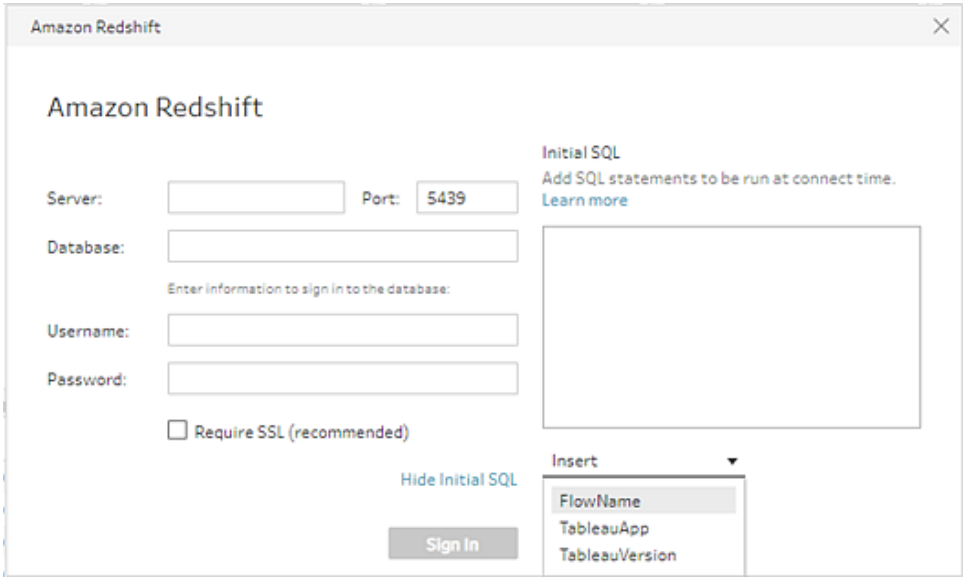
从 Tableau Prep Builder(版本 2020.1.3) 开始以及在 Web 上，还可以包括参数来传递应用程序名称、版本和流程名称数据，以在查询数据源时包括跟踪数据。

运行初始 SQL

若要刷新数据并运行初始 SQL 命令，请执行以下操作之一：

- 更改初始 SQL 命令并通过重新建立连接刷新“输入”步骤。
- 运行流程。初始 SQL 命令在处理所有数据之前运行。
- 在 Tableau Server 或 Tableau Cloud 上运行流程。初始 SQL 在每次流程运行时加载数据的过程中

注意：需要 Data Management 以在 Tableau Server 或 Tableau Cloud 上按计划运行流程。有关 Data Management 的详细信息请参见关于 Data Management。



1. 在“连接”窗格的列表中选择一个支持初始 SQL 的连接器。
2. 单击**“显示初始 SQL”**链接展开对话框，并输入您的 SQL 语句。

在初始 SQL 语句中包括参数

在 Tableau Prep Builder 版本 2020.1.3 及更高版本中受支持，以及从版本 2020.4.1 开始，在 Web 上制作流程时受支持。

您可以将以下参数传递给数据源，以添加有关 Tableau Prep 应用程序、版本和流程名称的其他详细信息。当前不支持 **TableauServerUser** 和 **TableauServerUserFull** 参数。

参数	描述	返回的值
TableauApp	用于访问数据源的应用程序。	Prep Builder Prep Conductor
TableauVersion	应用程序版本号。	Tableau Prep Builder：返回确切的版本。例如 2020.4.1 Tableau Prep Conductor：返回已启用 Tableau Prep Conductor 的主要服务器版本。例如 2020.4
FlowName	Tableau Prep Builder 中 .tfl 文件的名称	示例：Entertainment Data_Cleaned

配置数据集

注意：从版本 2020.4.1 开始，您现在可以在 Tableau Server 和 Tableau Cloud 中创建和编辑流程。除非另有说明，否则本主题中的内容适用于所有平台。有关在 Web 上制作流程的详细信息，请参见 Tableau Server 和 Tableau Cloud 帮助中的 **Web 上的 Tableau Prep**。

若要确定在流程中使用多少数据集，您可以配置数据集。当您连接到数据或将表拖到**“流程”**窗格中时，会自动向流程中添加一个输入步骤。

在“输入”步骤中，您可以决定要在流程中包含哪些数据以及包含多少数据。这始终是流程中的第一步。

如果连接到 Excel 或文本文件，您还可以从输入步骤中刷新数据。有关详细信息，请参见 **在“输入”步骤中添加更多数据** 在本页 119。

在“输入”步骤中，您可以：

- 在流程窗格中右键单击或按住 Cmd 并单击 (MacOS) 输入步骤，以重命名或移除该步骤。

- 合并位于同一父目录或子目录中的多个文件。有关详细信息，请参见**在输入步骤中合并文件和数据库表** 在本页123。
- (版本 2023.1 及更高版本)包括根据数据集的原始排序顺序自动生成的行号。请参见**包括数据集中的行号** 向下。
- 搜索字段。
- 查看字段值的预览。
- 通过更改字段名称来配置字段属性，或为文本文件配置文本设置。

  > **注意：**包括方括号的字段值会自动转换为圆括号。

- 配置接收到您的流程中的数据样本。请参见**选择数据样本大小** 在本页114。
- 移除不需要的字段。您始终可以返回到输入步骤并在以后包含它们。
- 隐藏不需要清理，但仍希望将其包括在流程输出中的字段。如果需要，您可以随时取消隐藏它们。
- 将筛选器应用于选定的字段。
- 更改支持它的数据连接的字段数据类型。
- (版本 2023.3 及更高版本)您可以设置 CSV 文件的标题和起始行。
- (版本 2024.1 及更高版本)您可以设置 Excel 文件的标题和起始行。

包括数据集中的行号

在 Tableau Prep Builder 版本 2023.1 及更高版本中以及在 Web 上支持 Microsoft Excel 和文本 (.csv) 文件。

> **注意：**当前不支持为输入并集中包含的文件使用此选项。

从版本 2023.1 开始，Tableau Prep 会根据数据的原始排序顺序自动生成行号，您可以将其作为新字段包含在流程中。这仅适用于 Microsoft Excel 或文本 (.csv) 文件类型。

在以前的版本中，如果您想包含这些行号，则必须在将数据集添加到流程之前手动将它们添加到源中。

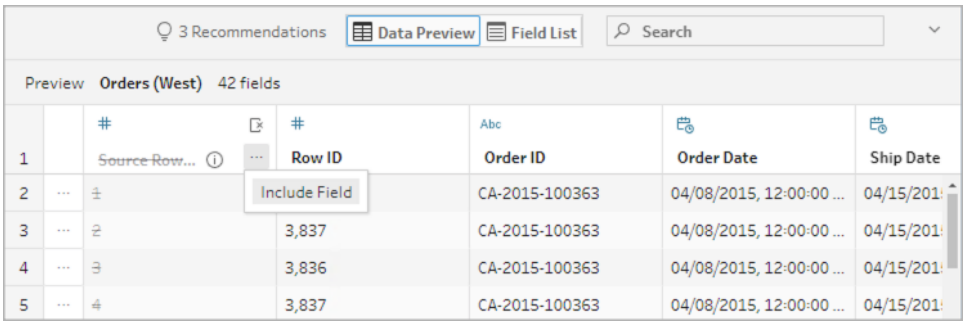
当连接到您的数据时，该字段在输入步骤中生成。默认情况下，它被排除在流程之外，但您可以单击一下将其包括在内。如果选择将其包括在内，它的行为与任何其他字段一样，可用于您的流程操作和计算字段。

Tableau Prep 还支持计算字段的 ROW_NUMBER 函数。当您的数据集中有可以定义排序的字段(例如 Row ID 或 Timestamp)时，此函数很有用。有关使用此函数的详细信息，请参见**创建详细级别、排名和分片计算** 在本页259。

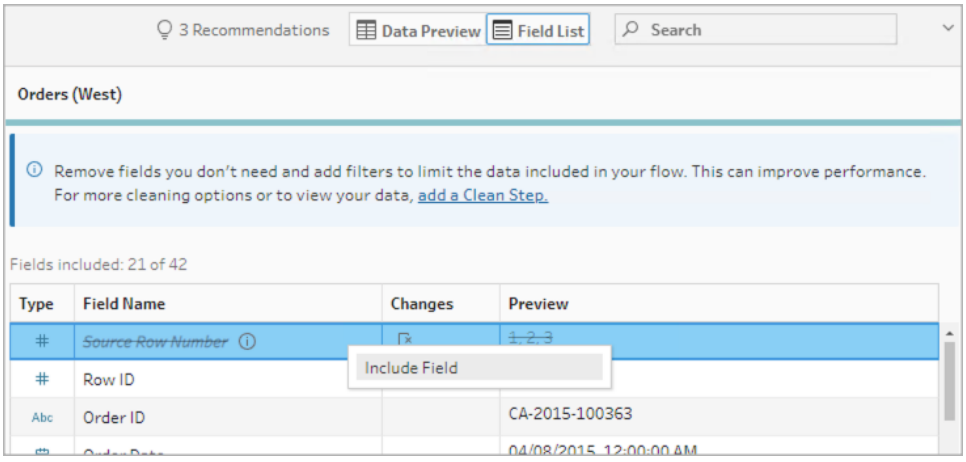
将源行号字段添加到您的流程

1. 右键单击或按住 Cmd 单击 (MacOS) 字段，或单击**“更多选项”**··· 菜单并选择 **“包括字段”**。

   数据预览：



   字段列表：



   更改列表被清除，该字段现在是流程数据的一部分，您可以在后续流程步骤中看到生成的行号。

源行号详细信息

当您在数据集中包括**“Source Row Number”**时，以下选项和注意事项适用。

- 数据源行号在任何数据采样或筛选器之前应用。
- 这将创建一个名为**“Source Row Number”**的新字段，该字段在整个流程中持续存在。此字段名未本地化，但可以随时重命名。

- 如果已存在具有此名称的字段，则新字段名称递增 1。例如**“Source Row Number-1”**、**“Source Row Number-2”**，等等。
- 您可以在后续步骤中更改字段的数据类型。
- 您可以在流程操作和计算中使用此字段。
- 每次刷新输入数据或运行流程时，都会为整个数据集重新生成此值。
- 此字段不可用于输入并集。

设置标题和数据起始行

Tableau Prep Builder 版本 2023.3 及更高版本和 Web 上支持文本 (.csv) 文件，版本 2024.1 及更高版本支持 Excel (.xls) 文件。

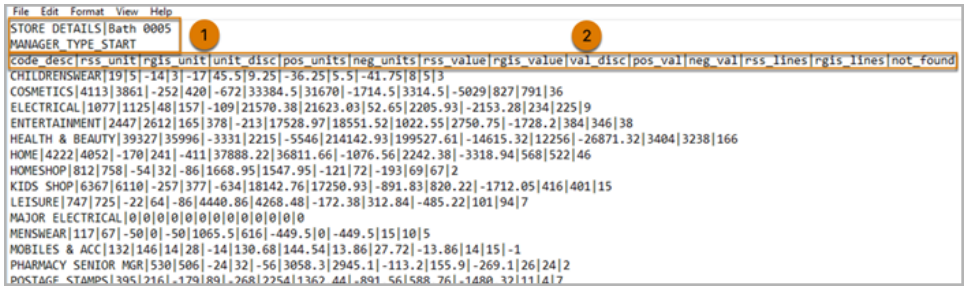
对于 Excel 和文本 (.csv) 文件，您可以将特定行设置为字段标题行以及数据起始行。

连接到 Excel 或文本文件时，常见的情况是这些文件在前几行中使用元信息进行格式化，以使其易于阅读。默认情况下，Tableau Prep 将 CSV 文件的第一行解释为字段标题行。Excel 文件根据字段类型和空行进行解释。Tableau Prep 可以选择一行作为标题，也可以不包含标题行。

例如，在以下文件中，`STORE DETAILS` 被解释为标题行。

您可以排除元数据信息 (1)，并通过将第 3 行设置为标题 (2)，将第 4 行设置为数据起始行，以此来提供数据的正确架构结构。

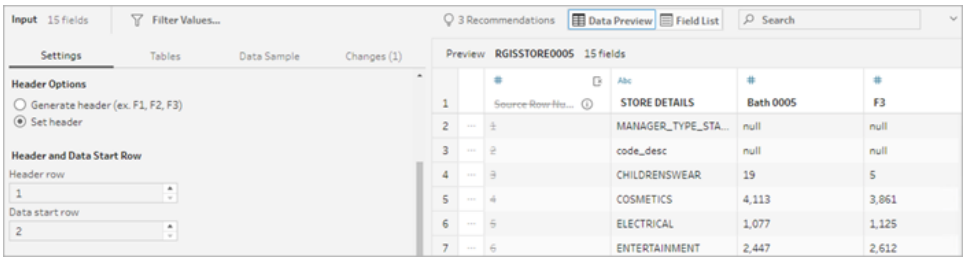
CSV 文件：



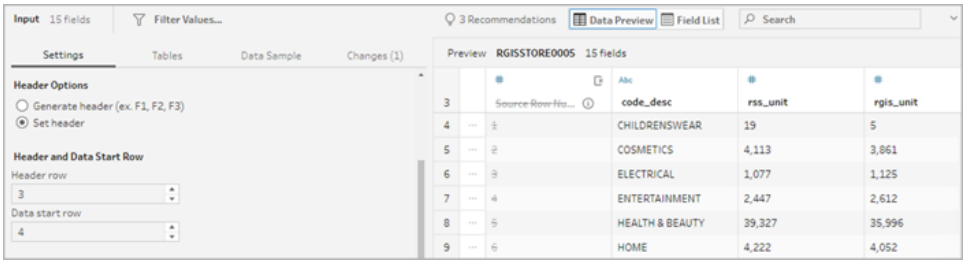
Excel 文件：

1	STORE DETAILS|Bath 0005
2	MANAGER_TYPE_START
3	code_desc|rss_unit|rgis_unit|unit_disc|pos_units|neg_units|rss_value|rgis_value|val_disc|pos_val|neg_val|rss_lines|rgis_lines|not_found
4	CHILDRENSWEAR|19|5|-14|3|-17|45.5|9.25|-36.25|5.5|-41.75|8|5|3
5	COSMETICS|4113|3861|-252|420|-672|33384.5|31670|-1714.5|3314.5|-5029|827|791|36
6	ELECTRICAL|1077|1125|48|157|-109|21570.38|21623.03|52.65|2205.93|-2153.28|234|225|9
7	ENTERTAINMENT|2447|2612|165|378|-213|17528.97|18551.52|1022.55|2750.75|-1728.2|384|346|38
8	HEALTH & BEAUTY|39327|35996|-3331|2215|-5546|214142.93|199527.61|-14615.32|12256|-26871.32|3404|3238|166
9	HOME|4222|4052|-170|241|-411|37888.22|36811.66|-1076.56|2242.38|-3318.94|568|522|46
10	HOMESHOP|812|758|-54|32|-86|1668.95|1547.95|-121|72|-193|69|67|2
11	KIDS SHOP|6367|6110|-257|377|-634|18142.76|17250.93|-891.83|820.22|-1712.05|416|401|15
12	LEISURE|747|725|-22|64|-86|4440.86|4268.48|-172.38|312.84|-485.22|101|94|7
13	MAJOR ELECTRICAL|0|0|0|0|0|0|0|0|0|0|0|0|0
14	MENSWEAR|117|67|-50|0|-50|1065.5|616|-449.5|0|-449.5|15|10|5
15	MOBILES & ACC|132|146|14|28|-14|130.68|144.54|13.86|27.72|-13.86|14|15|-1
16	PHARMACY SENIOR MGR|530|506|-24|32|-56|3058.3|2945.1|-113.2|155.9|-269.1|26|24|2
17	POSTAGE STAMPS|395|216|-179|89|-268|2254|1362.44|-891.56|588.76|-1480.32|11|4|7

例如，以下显示行标题和起始行的默认设置：



排除元数据后的数据如下：



注意：数据预览不反映数据样本设置的更改。

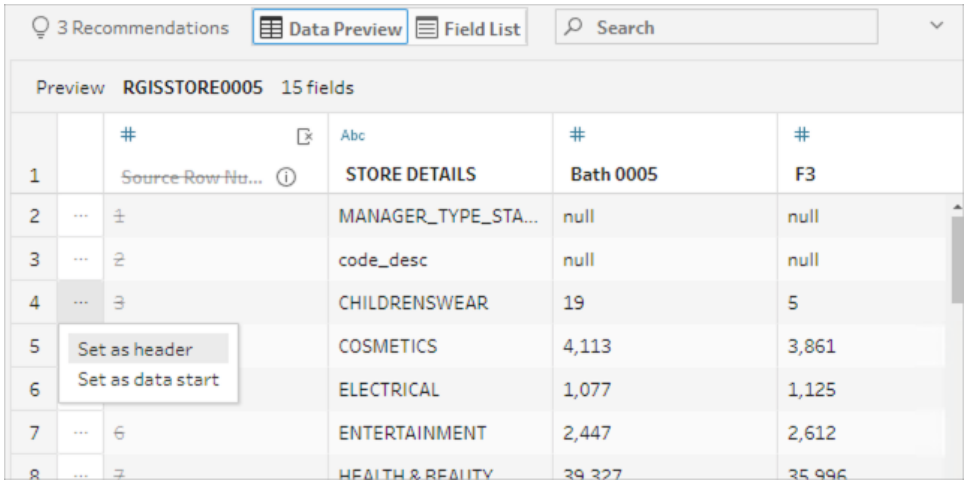
配置标题和起始行

使用数据预览输入视图，您可以直观地检查数据的架构结构，并设置标题和起始行以从输入源数据中排除元数据。

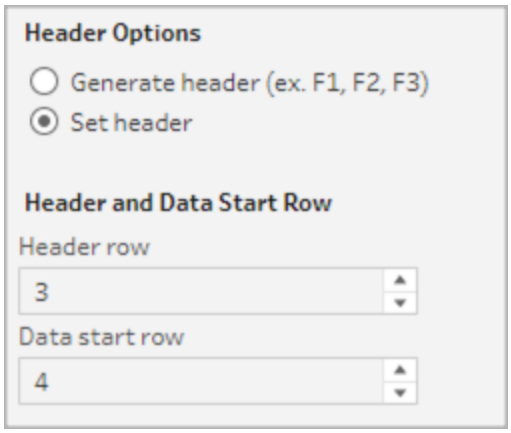
您可以将数据起始行设置为大于标题行值的任何值。默认情况下，Tableau Prep 将数据起始行设置为标题行后的下一个连续数字。标题行和数据起始行之间的任何行都将被忽略。

注意：数据预览和 Data interpreter 是互斥的。Data Interpreter 只会检测 Excel 电子表格中的子表，不支持为文本文件和电子表格指定起始行。

1. 选择输入步骤。
2. 从工具栏中，单击**“数据预览”**输入视图。
3. 在要设置为标题的行上，单击**“更多选项”**⋯ 菜单并选择**“设置为标题”**。
4. 在要设置为数据起始行的行上，单击**“更多选项”**⋯ 菜单并选择**“设置为数据开始”**。默认情况下，数据起始行设置为下一个连续行号。

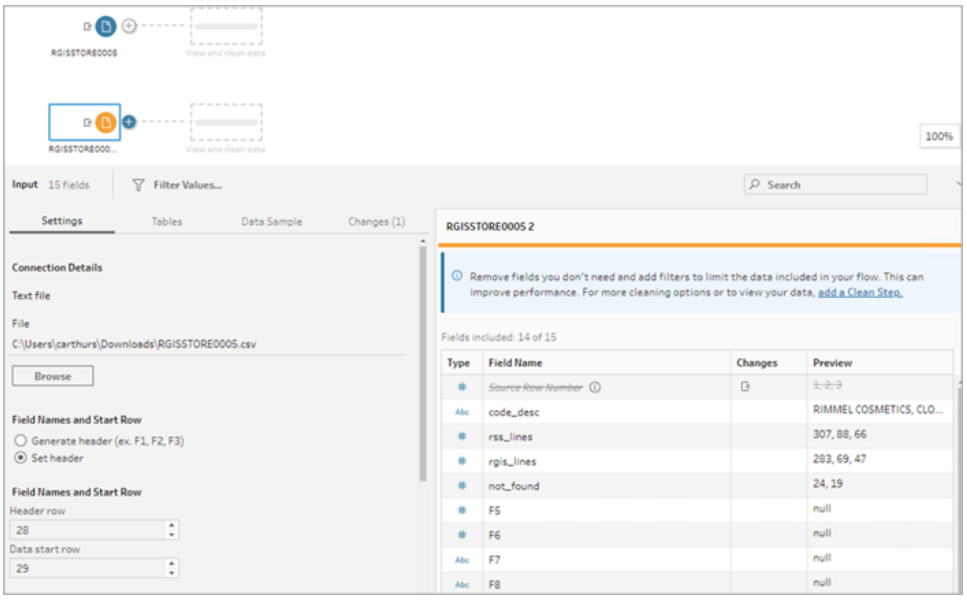


“标题选项”菜单显示标题行和数据起始行的行号。或者，您可以直接在“标题选项”对话框中设置标题和起始行。



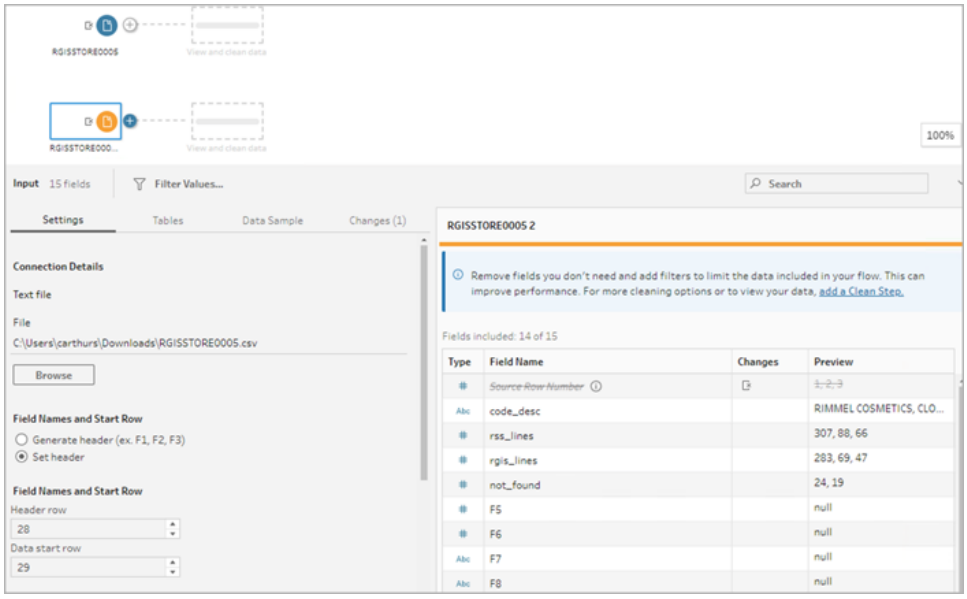
单个文件中的多个架构

如果单个文件包含多个数据源，您可以通过连接到同一数据源来创建附加输入步骤，然后为第二个数据源设置标题和数据起始行。例如，以下文件包含一个从第 3 行 (1) 开始的数据源，另一个单独的架构从第 28 行 (2) 开始。



对于此类数据源，请按照以下步骤操作。

1. 选择第一个输入步骤。
2. 从工具栏中，单击**“数据预览”**输入视图。
3. 在要设置为标题的行上，单击**“更多选项”**⋯ 菜单并选择**“设置为标题”**。
4. 在要设置为数据起始行的行上，单击**“更多选项”**⋯ 并选择**“设置为数据开始”**。默认情况下，数据起始行设置为下一个连续行号。
5. 选择下一个输入步骤。
6. 重复上述步骤，为其他数据源设置标题和起始行。



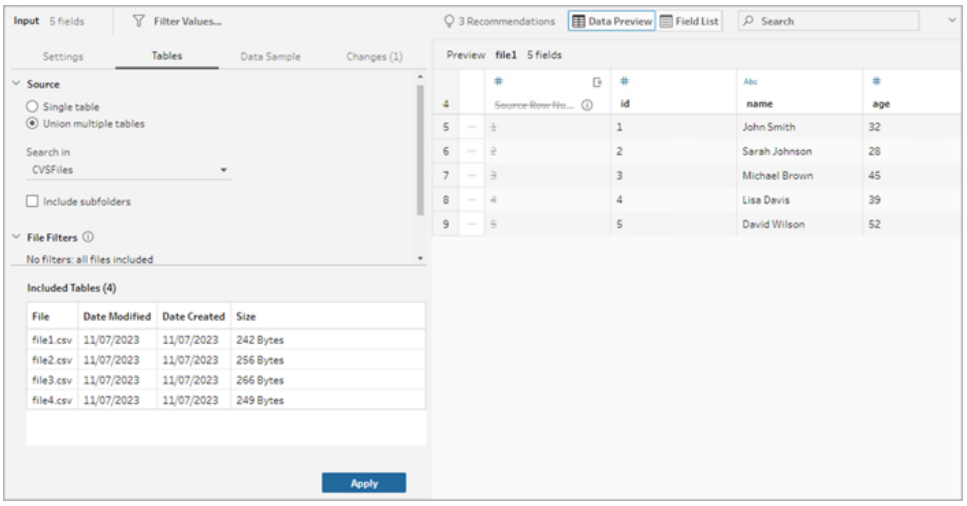
标题行和数据起始行之间的任何行都将被忽略。

合并多个表

在 Tableau Prep Builder 版本 2024.1 及更高版本中以及在 Web 上支持文本 (.csv) 文件。

您可以合并具有相同架构结构和元数据行的数据源中的多个表。

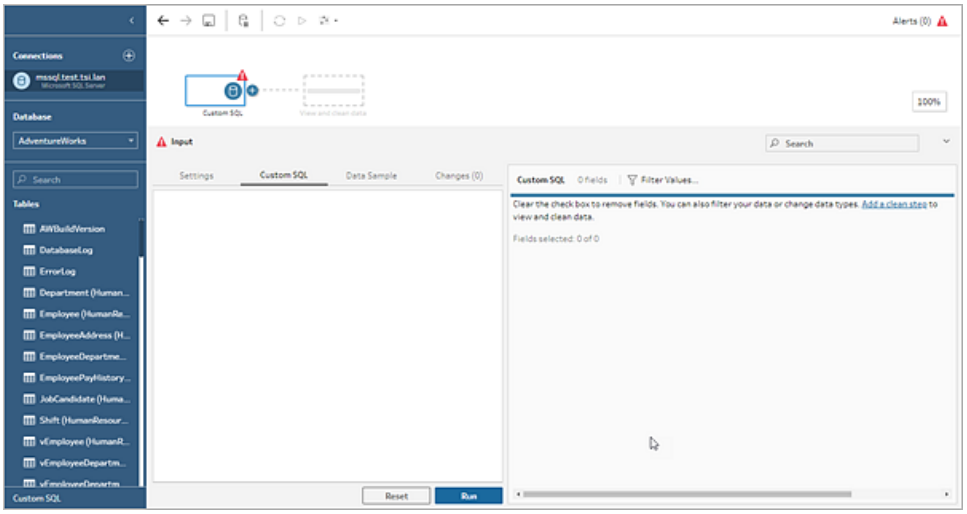
1. 连接到文件并选择第一个输入步骤。
2. 从工具栏中，单击**“数据预览”**输入视图。
3. 在要设置为标题的行上，单击**“更多选项”**⋯ 并选择**“设置为标题”**。
4. 在要设置为数据起始行的行上，单击**“更多选项”**⋯ 并选择**“设置为数据开始”**。
5. 单击**“表”**选项卡并选择**“合并多个表”**。



6. 单击**“应用”**合并文件，并保留输入并集中所有文件的标题和行选择。这一操作的假设是输入合并文件之间的文件结构和架构是相同的。

连接到自定义 SQL 查询

如果数据库支持使用自定义 SQL，您会看到**“自定义 SQL”**显示在**“连接”**窗格靠近底部的位置。双击**“自定义 SQL”**打开**“自定义 SQL”**选项卡，您可以在其中输入查询来预选择数据和使用特定于来源的操作。查询检索数据集之后，您可以选择要包括的字段、应用筛选器，或在将数据添加到流程之前更改数据类型。



有关使用自定义 SQL 的详细信息，请参见**使用自定义 SQL 连接到数据** 在本页94。

在输入步骤中应用清理操作

只能在“输入”步骤中执行部分清理操作。您可以在“输入”字段列表中进行以下任何更改。系统会在**“更改”**窗格中跟踪您所做的更改，并会在**“流程”**窗格中“输入”步骤的左侧以及“输入”字段列表中添加注释。

- **隐藏字段**：隐藏字段而不是移除它们以减少流程中的混乱。如果需要，您可以随时取消隐藏这些字段。当您运行流程时，隐藏字段仍将包含在输出中。
- **筛选**：使用计算编辑器筛选值，或者从版本 2023.1 开始，您也可以使用**“相对日期筛选器”**对话框以快速指定任何日期或日期和时间字段的日期范围。
- **重命名字段：**在**“字段名”**字段中，双击或按住 Ctrl 并单击 (MacOS)字段名，并输入一个新字段名。
- **更改数据类型：**单击字段的数据类型，并从菜单中选择一个新数据类型。Microsoft Excel、文本和 PDF 文件、Box、Dropbox、Google Drive 和 OneDrive 数据源目前支持此选项。所有其他数据源都可以在清理步骤中更改。

选择要包括在流程中的字段

注意：从版本 2023.1 开始，您可以选择多个字段来隐藏、取消隐藏、移除或包含它们。在以前的版本中，您一次可以处理一个字段并选中或清除复选框以包含或移除字段。

“输入”窗格显示数据集中的字段列表。默认情况下，除自动生成的字段**“Source Row Number”**外，所有字段都包括在内。使用“数据预览”或“列表”视图来管理您的字段。

- **搜索**：查找字段。
- **隐藏字段**：隐藏要包含在流程输出中但不需要清理的字段。
    - 在字段列表中，单击眼睛 图标或从**“更多选项”** ··· 菜单中选择**“隐藏字段”**。
    - 在数据预览中，从**“更多选项”** ··· 菜单中选择**“隐藏字段”**。

  字段在运行时由流程处理。如果需要，您也可以随时**“取消隐藏”**字段。有关详细信息，请参见隐藏字段。
- **包含字段**：将已标记为已移除的字段添加到流程中。
    - 在字段列表中，选择一行或多行并单击鼠标右键，按住 Cmd 单击 (MacOS)，或单击**“更多选项”** ··· 菜单并选择**“包含字段”**，添加回标记为已移除的字段。

- 在数据预览中，在要包含在流程中的字段上单击**“更多选项”**··· 菜单，然后选择**“包含字段”**。

- **移除字段**：
  - 在字段列表中，选择一行或多行并单击鼠标右键，按住 Cmd 单击 (MacOS)，单击“X”，或单击**“更多选项”**··· 菜单并选择 **“移除字段”**，移除您不想包含在流程中的字段。
  - 在数据预览中，在要移除的字段上单击**“更多选项”**··· 菜单，然后选择**“移除字段”**。

将筛选器应用于输入步骤中的字段

在输入步骤中应用筛选器以减少您从数据源摄取的数据量。通过消除运行流程时不想处理的数据，您可以获得交互式性能效率和更有用的数据样本。

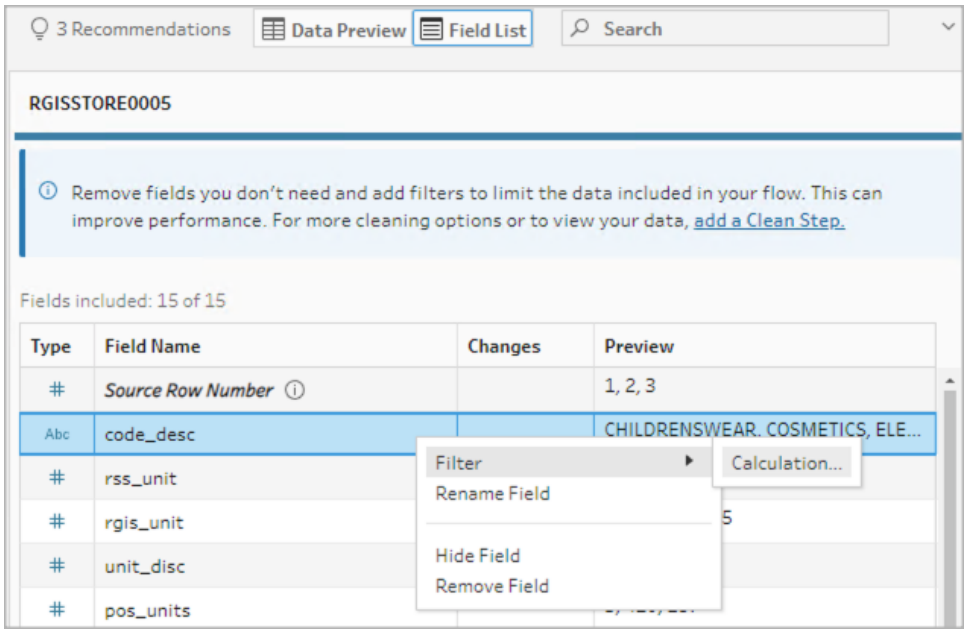
在输入步骤中，您可以使用计算编辑器应用筛选器。从版本 2023.1 开始，您还可以使用**“相对日期筛选器”**对话框以指定要包含在日期和日期和时间字段类型中的确切日期值范围。有关详细信息，请参见**筛选您的数据** 在本页165中的“相对日期筛选器”。

您可以在“清理”步骤或其他步骤类型中使用其他筛选器选项。有关详细信息，请参见**筛选您的数据** 在本页165。

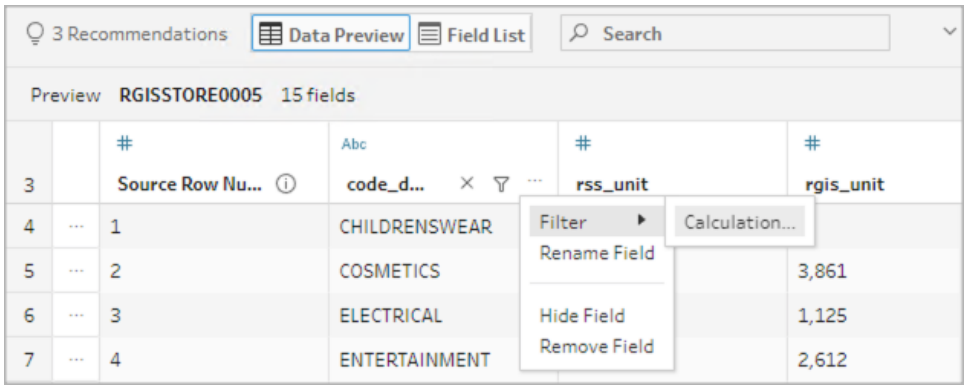
应用计算筛选器

1. 在工具栏中，单击**“筛选值”**。使用以下方法之一筛选数据：

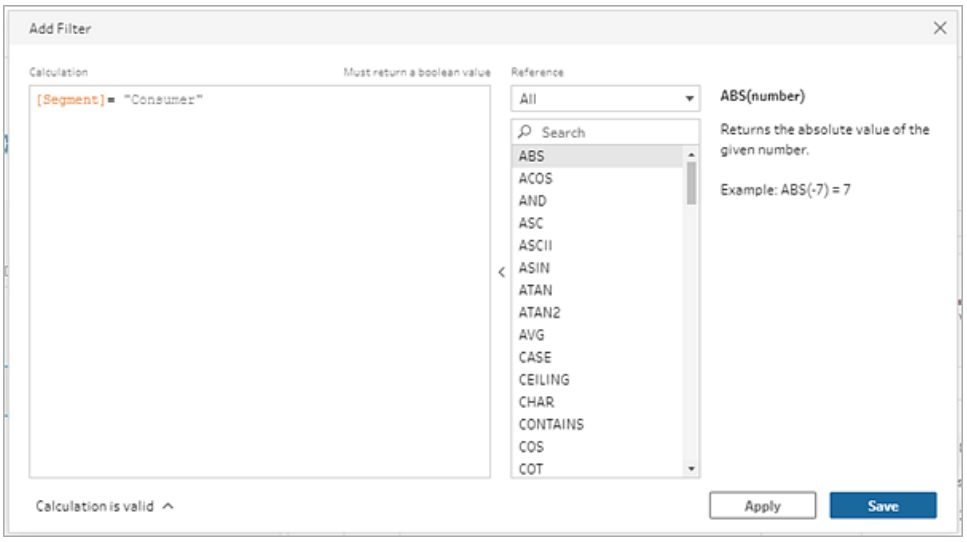
   - 在字段列表中，从字段名称中单击**“更多选项”**··· 菜单，然后选择**“筛选器”>“计算...”**。



- 在数据预览中，从字段名称中单击**“更多选项”**⋯ 菜单，然后选择**“筛选器”>“计算...”**。

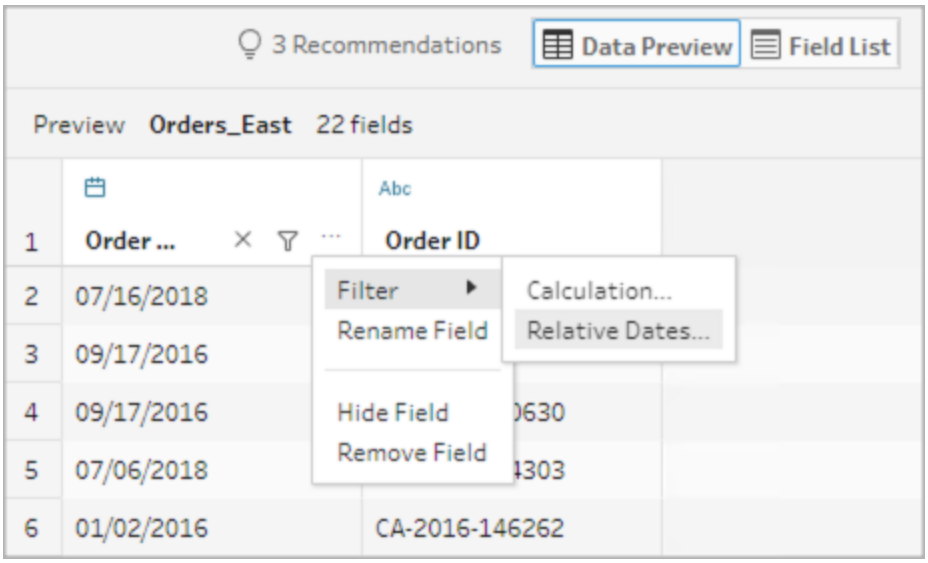


2. 在计算编辑器中输入筛选条件。



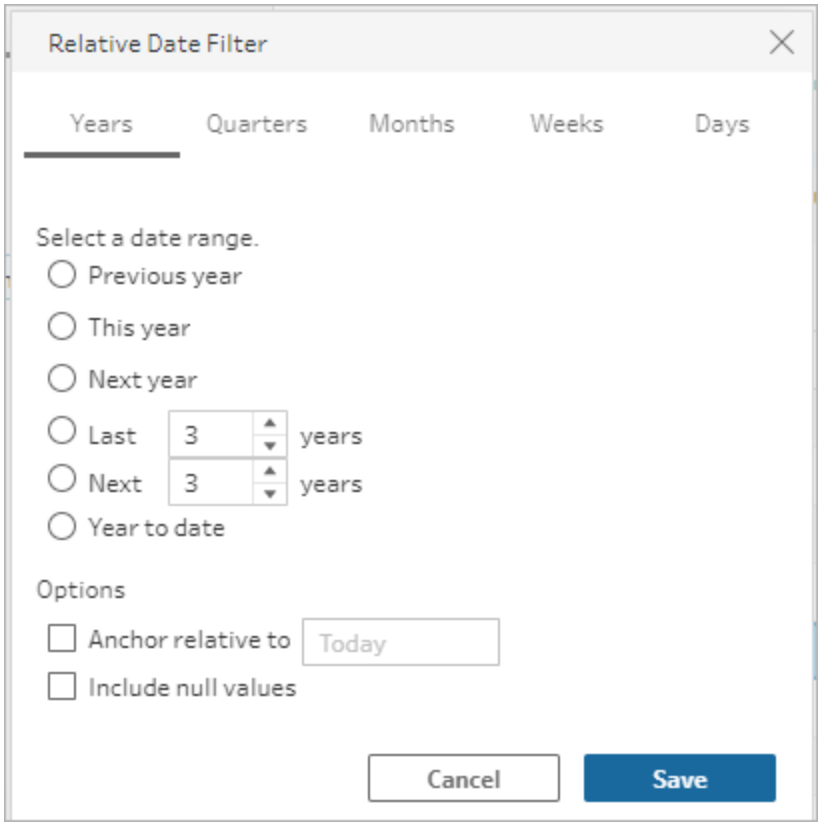
应用相对日期筛选器

1. 选择数据类型为“日期”或“日期和时间”的字段，并使用以下方法之一应用相对日期筛选器。
   - 在字段列表中，右键单击、按住 Cmd 单击 (MacOS)，或从“字段名称”列中单击“**更多选项**”··· 菜单，然后选择“**筛选器**”>“**相对日期**”。
   - 在数据预览中，从字段中单击“**更多选项**”··· 菜单，然后选择“**筛选器**”>“**相对日期**”。



2. 使用“相对日期筛选器”对话框，指定要包含在流程中的年、季度、月、周或天的确切范围。您也可以配置相对于特定日期锚点，并包括 null 值。

   注意：默认情况下，筛选器相对于流在创作体验中运行或预览的日期运行。

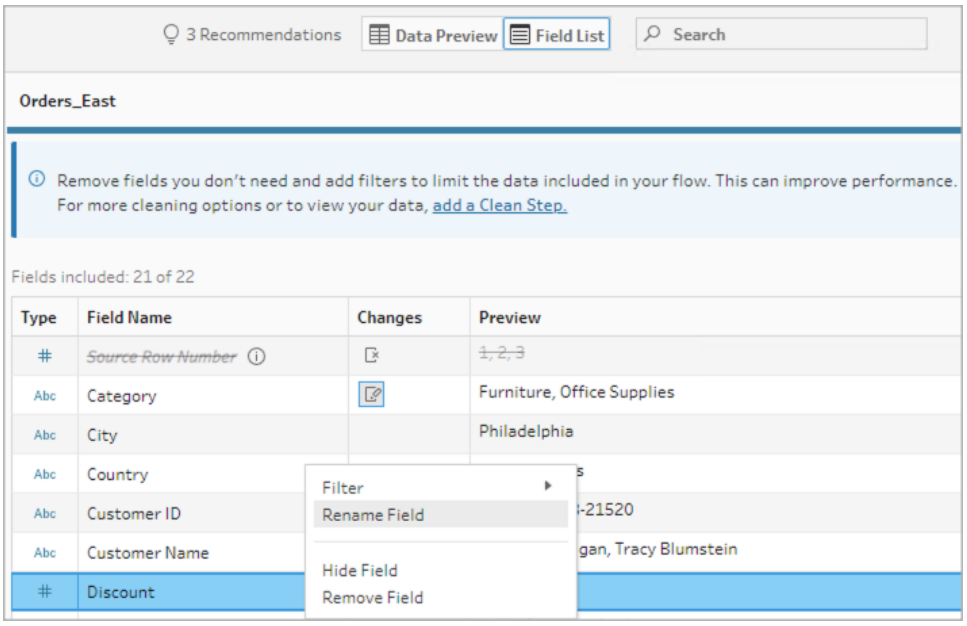


更改字段名称

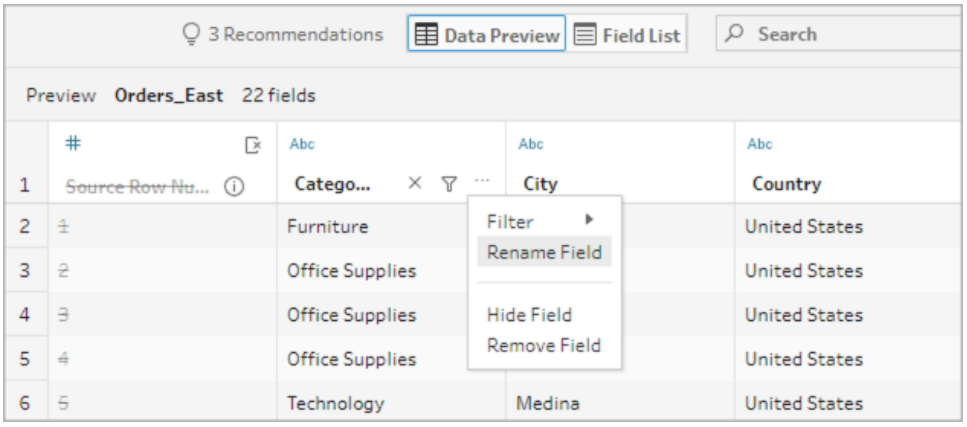
使用以下方法之一更改字段的名称。

将会在字段网格以及“输入”步骤左侧的流程窗格中添加注释。系统也会在“更改”窗格中跟踪您所做的更改。

- 在字段列表中，从“字段名称”列中选择一个字段，然后单击**“重命名字段”**。在字段中键入新名称。



- 在数据预览中，选择一个字段并单击**"重命名字段"**。在字段中键入新名称。



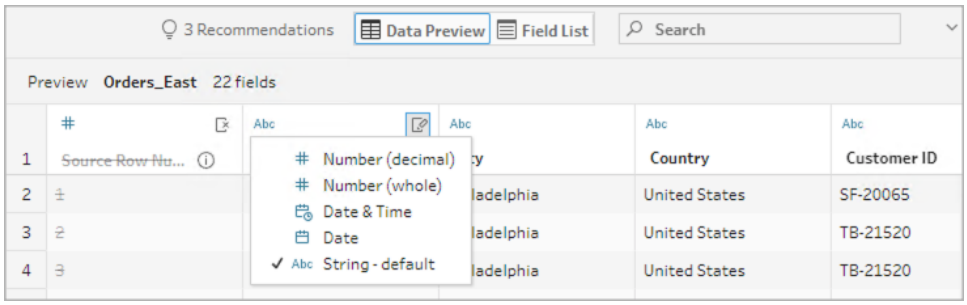
更改数据类型

Microsoft Excel、文本和 PDF 文件、Box、Dropbox、Google Drive 和 OneDrive 数据源目前支持。所有其他数据源都可以在清理步骤中更改。

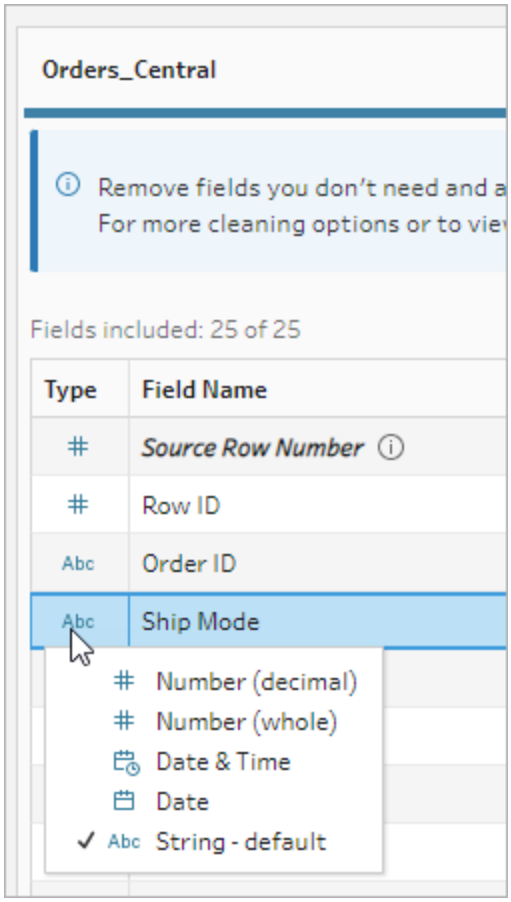
注意：源行号（版本 2023.1 及更高版本）的数据类型只能在清理步骤或其他步骤类型中更改。

若要更改字段的数据类型，请执行以下操作：

1. 单击字段的数据类型。
2. 从菜单中选择新数据类型。
    - 数据预览：



    - 字段视图：



您还可以更改流程中其他步骤类型中的字段的数据类型，或分配数据角色来帮助验证字段值。有关更改数据类型或使用数据角色的详细信息，请参见**查看分配给数据的数据类型** 在本页153和**使用数据角色验证数据** 在本页178。

配置字段属性

处理文本文件时，您会看到“**设置**”选项卡，在该选项卡中，您可以编辑连接和配置文本属性，例如文本文件的字段分隔符。您也可以在“连接”窗格中编辑文件连接或配置增量刷新设置。有关为流程设置增量刷新的详细信息，请参见**使用增量刷新来刷新流程数据** 在本页390。

处理文本或 Excel 文件时，您可以纠正在开始流程之前错误推断的数据类型。在开始流程之后，始终可以在后续步骤中通过**“配置”**窗格更改数据类型。

配置文本文件中的文本设置

若要更改用于解析文本文件的设置，请从以下选项中选择：

- **第一行包含标题**(默认值)：选择此选项以使用第一行作为字段标签。
- **自动生成字段名称**：如果希望 Tableau Prep Builder 自动生成字段标题，请选择此选项。字段命名约定采用与 Tableau Desktop 相同的模式。例如，**F1**、**F2** 等。
- **字段分隔符**：从列表中选择一个字符用于分隔各个列。选择**“其他”**以输入自定义字符。
- **文本限定符**：选择用于在文件中将值引起来的字符。
- **字符集**：选择用于描述文本文件编码的字符集。
- **区域设置**：选择要用于解析文件的区域设置。此设置指明要使用哪个小数分隔符和千位分隔符。

选择数据样本大小

Tableau Prep Builder 版本 2023.3 及更高版本支持分层行选择。

默认情况下，Tableau Prep 确定为数据集的代表性样本有效浏览和准备数据所需的最大行数。根据 Tableau Prep 示例算法，输入数据中的字段越多，允许的行数就越少。对数据进行采样时，得到的样本可能包含您需要的所有行，也可能不包含，这取决于样本是如何计算和返回的。例如，默认情况下，Tableau Prep 使用快速选择方法对数据进行采样。使用此方法，会加载最上面的行，如果您的数据集很大并且数据按时间顺序排列，您可能会

看到最早的采样数据，但不是所有数据的完整表示。如果看不到预期数据，您可以更改数据样本设置以再次运行查询。

使用 Web 制作创建或编辑流程时，用户在使用大型数据集时可以选择的最大行数由管理员配置。有关详细信息，请参见 Tableau Server 或 Tableau Cloud 帮助中的*"示例数据和处理限制"*。

为采样准备数据

如果您知道分析不需要某些值，请移除输入步骤中的字段，以便在制作或运行流程时不包含数据。

如果您有触发采样的大型数据集，则在输入步骤中移除字段会增加 Tableau Prep 加载的行数。未应用采样时，移除输入步骤中的字段会减少 Tableau Prep 加载的数据量。

从数据集中移除不必要的字段和值后，您可以更改为采样加载的数据量或采样方法。

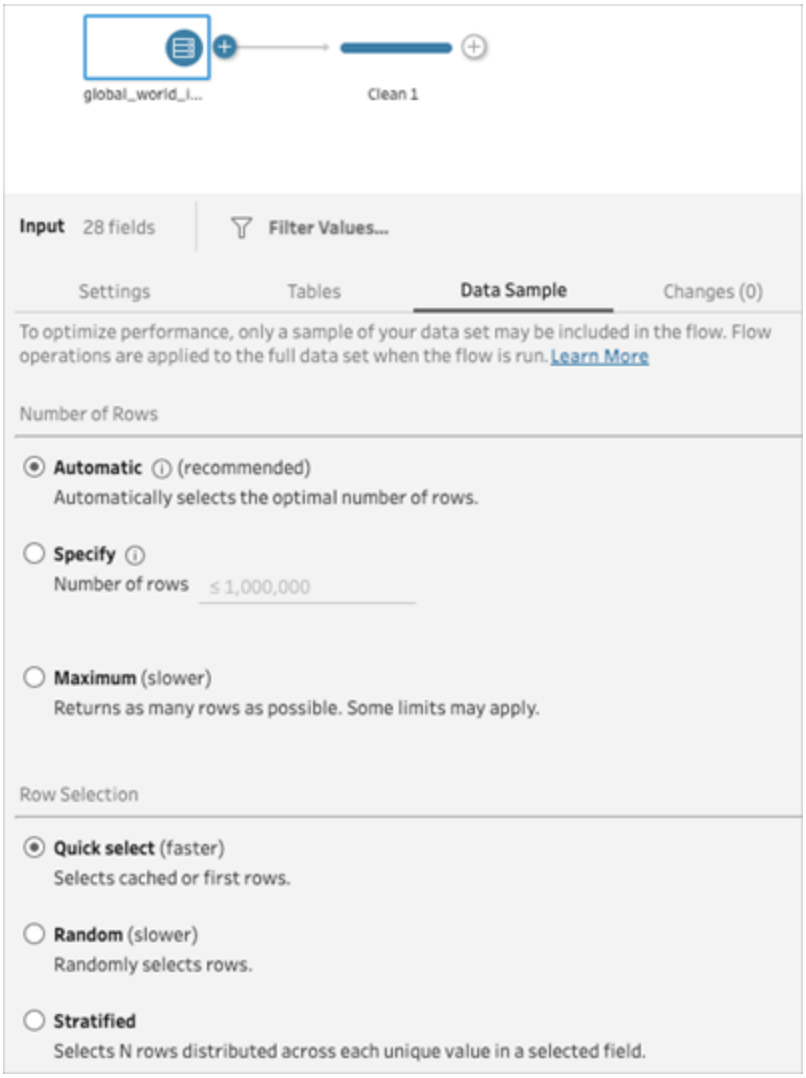
更改数据样本设置

与在工作时分析所有数据并将更改应用到更大的数据集相比，样本数据有助于提供交互式体验，并使编辑流程更加高效。运行流程时将使用所有数据。您在样本部分中所做的任何更改都会应用于当前流程。

若要在清理和调整后验证数据，请运行流程并在 Tableau Desktop 中查看输出。

注意：运行完整流程而不是"Tableau Desktop 中的视图样本"，以便能够查看完整的数据。如果您看到样本中没有的意外或不正确的值，您可以返回 Tableau Prep 来解决此问题。

1. 从数据集中删除不必要的字段和值。
2. 选择一个输入步骤，然后单击**“数据样本”**选项卡。



3. 选择要为数据采样加载的行数。您选择的行数会影响性能。
   - **自动**：(默认)快速加载数据并自动计算行数，以便为样本提供足够的数据。加载的行数等于或小于 393,216。
   - **指定**：通常用于加载少量行，以便您可以理解数据的结构并快速加载。指定小于 100 万行的行数。

> 注意：在 Web 制作中：用户在使用大型数据集时可以选择的最大行数由管理员配置。作为用户，您可以选择最多达到该限制的行数。

- **最大值**：为行选择加载尽可能多的数据，等于或小于 1,048,576。确保您满足大型数据集的高性能要求。

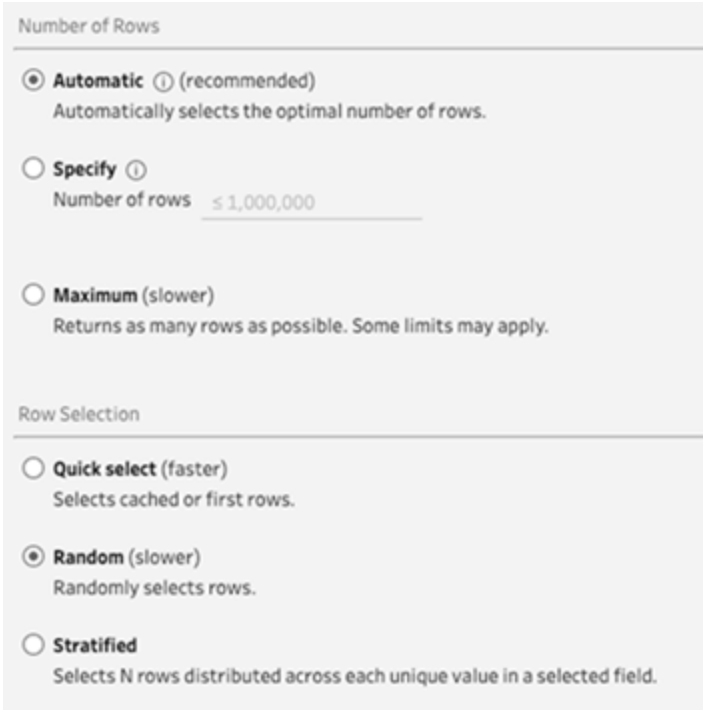
4. 选择用于采样返回的行数的方法。选择“随机”或“分层”时，性能可能会受到影响。

> **注意**：仅当您的输入数据源支持随机采样时才支持行选择。如果您的数据源不支持随机采样，则使用默认的“快速选择”方法。

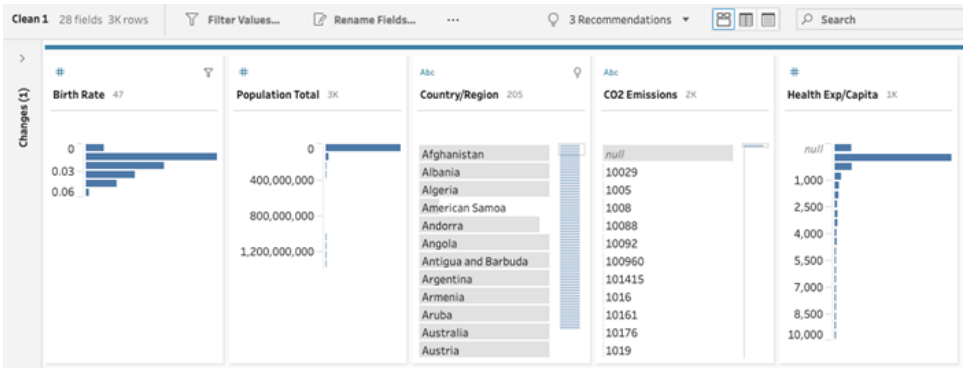
- **快速选择**：（默认）根据性能对数据进行采样，因为行会尽快返回。样本中可能不包括某些行。用于采样的行可能是前 N 行，或是上一次查询中数据库缓存在内存中的行数。虽然这几乎总是比随机抽样更快的结果，但它可能会返回有偏差的样本（例如，如果记录按时间顺序排序，则只返回一年的数据，而不是数据中存在的所有年份的数据）。
- **随机**：允许您对大型数据集进行采样，并返回整个选定行的一般表示形式。Tableau Prep 根据加载的所有选定行返回随机行。在首次检索数据时，此选项可能会影响性能。
- **分层**：允许您按指定字段进行分组，然后对每个子组内的数据进行采样。Prep 返回分布在所选字段中的请求行数，以便尽可能均匀地进行分组。在某些情况下，根据数据源，这可能会导致字段的某些值比其他值具有更多行。

示例

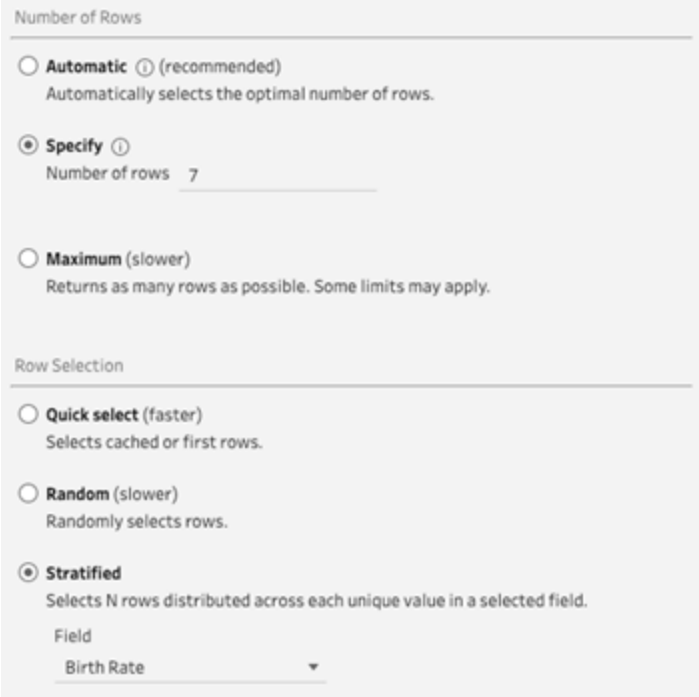
这些示例基于 Tableau Prep 附带的全球世界指标数据集。第一个样本使用**“自动”**作为行数，使用**“随机”**作为采样方法或要返回的采样行数。



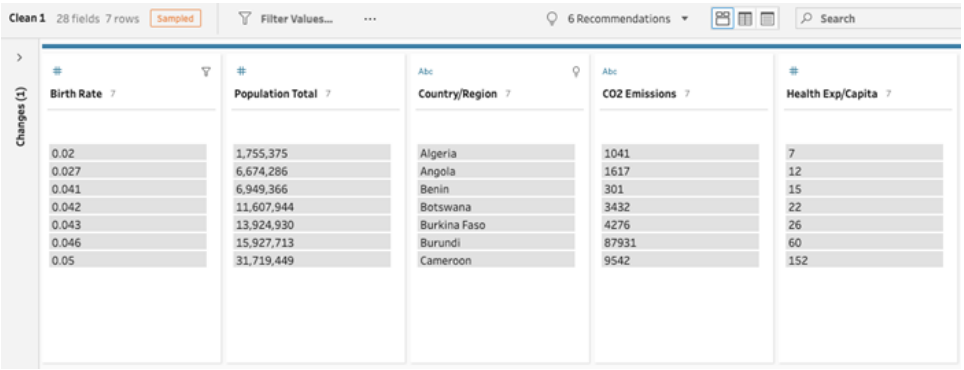
选择这些值后，将随机选择 3000 行并用于表示整个数据集。



第二个样本使用**“指定”**作为行数，并使用**“分层”**作为样本方法。指定的行数设置为值 7，并且*“Birth Rate”(出生率)*字段用于分组。



新的样本值显示所有字段中 7 行唯一值的统一分布。



在“输入”步骤中添加更多数据

注意：从版本 2020.4.1 开始，您现在可以在 Tableau Server 和 Tableau Cloud 中创建和编辑流程。除非特别指出，本主题中的内容适用于所有平台。有关在 Web 上制作流程的详细信息，请参见 Tableau Server 和 Tableau Cloud 帮助中的 **Web 上的 Tableau Prep**。

连接到数据源并开始构建流程后，您可能需要在新数据进入时刷新数据连接。您还可以在输入步骤中联接或合并数据集，以便更有效地处理更大的数据源。

刷新输入步骤数据或更改您的连接

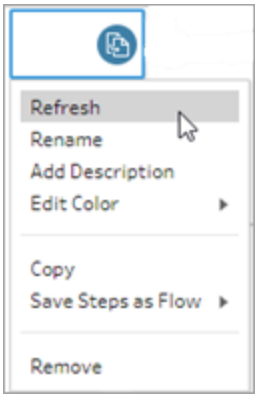
如果开始处理流程后输入文件或表中的数据发生变化，您可以刷新“输入”步骤为引入新数据，或者，您可以轻松地更改和更新单个输入步骤连接，而不会中断您的流程。

刷新数据源

适用于文件类型。Web 上尚不支持。

执行以下操作之一：

- 在流程窗格中，右键单击要刷新的“输入”步骤，并从菜单中选择**“刷新”**。



- 在流程窗格中的顶部菜单上，单击**“刷新”**按钮以刷新所有“输入”步骤。若要刷新单个“输入”步骤，请单击刷新按钮旁边的下拉箭头，并从列表中选择该“输入”步骤。

替换您的数据源

适用于 Tableau Prep Builder 和 Web 上的文件类型、数据源和数据提取。

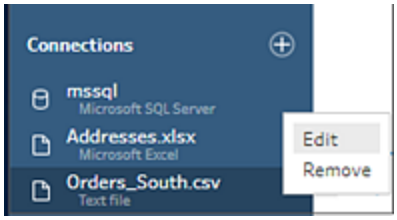
通过编辑单个输入连接或用不同的数据源替换单个流程数据源来刷新您的数据源。

编辑连接

使用此选项可以轻松刷新您的凭据或将数据源替换为相同的数据源类型。

注意：为了保持性能，Tableau Prep 会对大型数据集取样。如果数据已取样，则在配置窗格中，您可能会看到也可能不会看到新数据。您可以在“输入”步骤中的**“数据样本”**选项卡中更改数据的取样方式，但可能会影响性能。有关设置数据样本大小的详细信息，请参见选择数据样本大小 在本页114。

1. 在**“连接”**窗格中，右键单击或按住 Ctrl 单击 (MacOS) 数据源，并选择**“编辑”**。



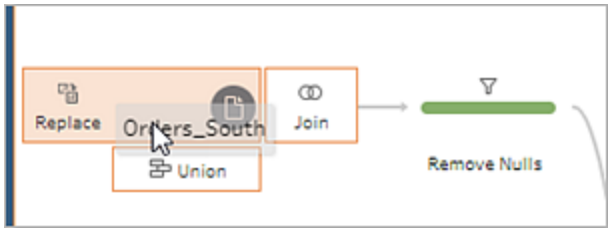
2. 通过登录数据库或者重新选择文件或 Tableau 数据提取来重建连接。

替换输入连接

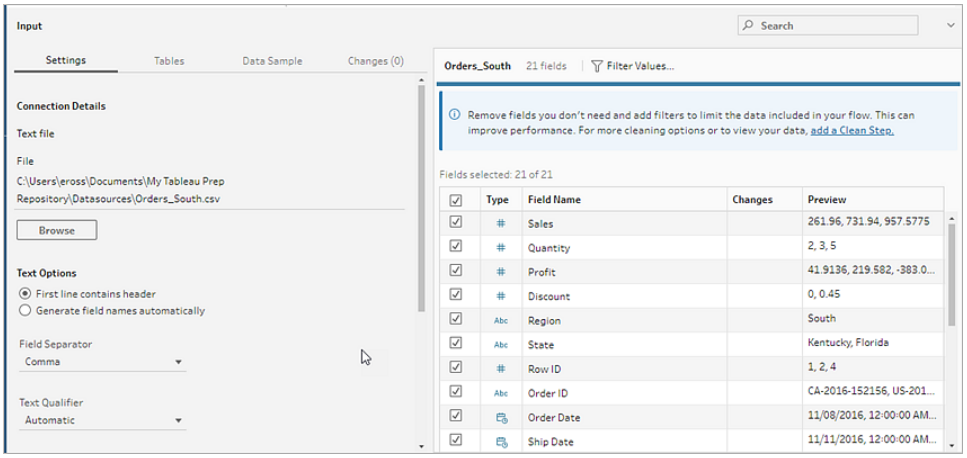
在不中断流程连接的情况下，使用任何新数据源轻松替换流程中的现有数据源。根据您的 Tableau Prep 版本，您可以将新数据源拖放到旧数据源上，或者手动断开并重新连接数据源。

拖放以替换输入连接(版本 20224. 及更高版本)

1. 从**“连接”**窗格中，将新表拖到要替换的输入步骤顶部的流程窗格中，然后将其放在**“替换”**选项上。

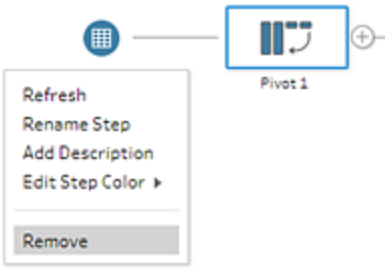


2. 根据需要重新配置任何设置并修复任何错误。



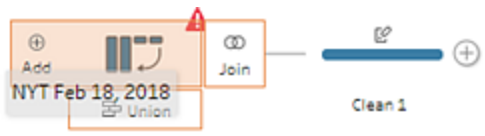
手动断开并重新连接输入数据源(版本 2022.3 及更早版本)

1. 在流程窗格中，右键单击要刷新的“输入”步骤，并从菜单中选择**“移除”**。



这会使您的流程暂时处于错误状态。

2. 连接到新数据源或更新的数据源。
3. 将表拖到流程窗格中要向其中添加“输入”步骤的流程中的第二个步骤上。将其放在**“添加”**选项上，以将其重新连接到流程。



在输入步骤中合并文件和数据库表

输入并集只能在 Tableau Prep Builder 中编辑和创建，但可以计划为在 Web 上运行。

使用来自单一数据源的多个文件或数据库表时，您可以应用筛选器以搜索文件或使用通配符搜索来查找表，然后合并数据以将所有文件或表数据包括在“输入”步骤中。若要合并文件，文件必须位于同一目录或子目录中。

注意：此选项不可用于 Tableau 数据提取。

在下次您打开流程或通过命令行运行流程时，添加到同一文件夹中、与筛选条件匹配的新文件会自动包括在并集中。

打包流程文件 (.tflx) 不会自动获取新文件，因为文件已与流程打包在一起。若要为打包流程包括新文件，请在 Tableau Prep Builder 中打开流程文件 (.tfl) 以获取新文件，然后重新打包流程以包括新文件数据。

若要合并数据库表，表必须位于同一数据库中，并且数据库连接必须支持通配符搜索。以下数据库支持这种类型的并集：

- Amazon Redshift
- Microsoft SQL Server
- MySQL
- Oracle
- PostgreSQL

如果在创建并集之后添加或移除文件或表，您可以刷新输入步骤，用新数据或更改的数据来更新流程。

如果需要合并来自不同数据源的数据，您可以使用合并步骤执行该操作。有关创建合并步骤的详细信息，请参见**合并数据** 在本页 346。

合并文件

默认情况下，Tableau Prep Builder 会合并您连接到的 .csv 文件所在同一目录中的所有 .csv 文件，或者合并您连接到的 Excel 文件中的所有工作表。

如果要更改默认并集，您可以指定额外的筛选条件以查找要包括在并集中的文件或工作表。

核心筛选条件

在 Tableau Prep Builder 版本 2022.1.1 及更低版本中，您可以从以下条件中进行选择：

- **搜索范围**：选择要用于搜索文件的目录。选中**“包括子文件夹”**复选框以包括父文件夹的子目录中的文件。
- **文件**：选择是否包括或排除与通配符搜索条件匹配的文件。
- **匹配模式(xxx*)**：输入通配符搜索模式以查找文件名中包含这些字符的文件。举例来说，如果输入 order*，则会返回文件名中包含“order”的所有文件。将此字段留空以包括指定目录中的所有文件。

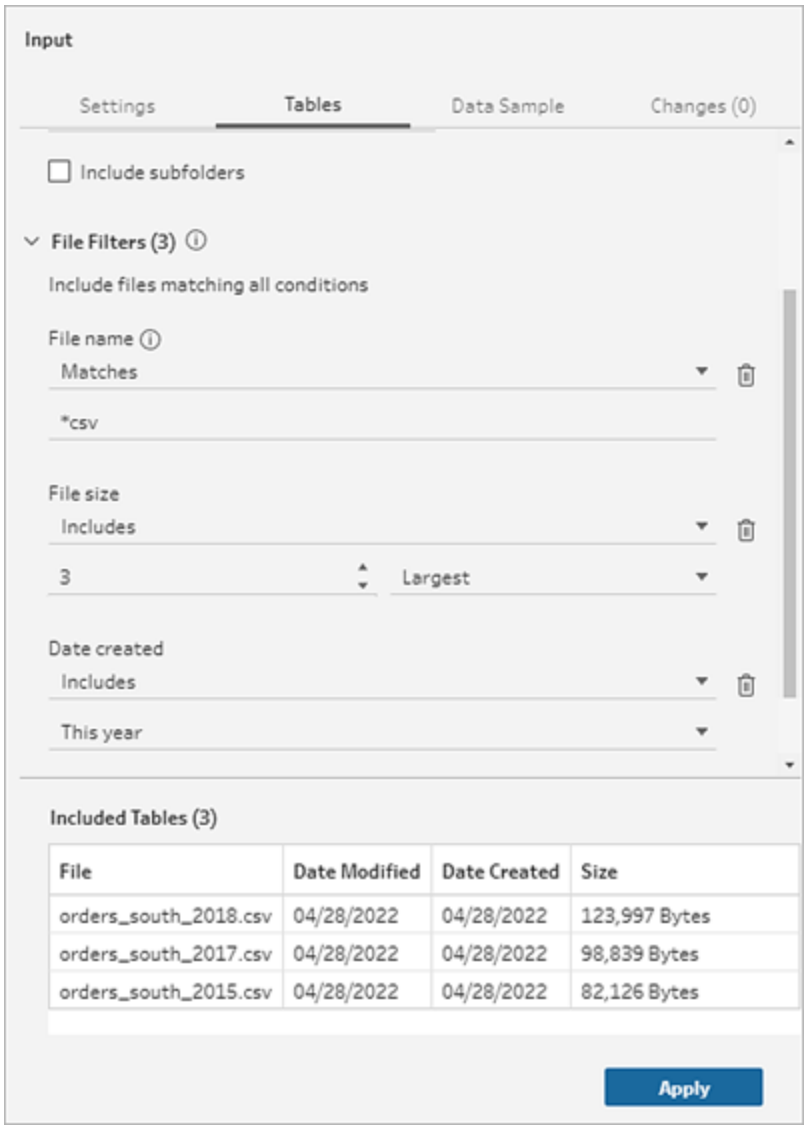
额外的筛选器

在 Tableau Prep Builder 版本 2022.2.1 及更高版本中受支持，并且适用于发布到 Tableau Cloud 的流程。

> **注意**：如果您在流程中使用额外的筛选器，流程计划当前只能使用 Tableau Cloud 进行。您可以在 Tableau Prep Builder 中或通过命令行界面手动运行流程。此功能与 Tableau Server 版本 2022.1 及更低版本不兼容。

从 Tableau Prep Builder 版本 2022.2.1 及更高版本开始，搜索要合并的文件时的筛选选项已更改。虽然您仍然指定要搜索的目录和子目录，但您现在可以设置多个筛选器来执行更精细的搜索。

这些筛选选项适用于文本、Microsoft Excel 和统计文件类型。您可以选择多个筛选器。每个筛选器都是按照您选择它们的顺序从上到下单独应用的。筛选器在添加后目前无法移动，但您可以根据需要删除和添加筛选器。



从以下筛选器中进行选择：

筛选器	描述
文件名	为文件名模式选择**“匹配”**或**“不匹配”**。例如“orders*”。
文件大	通过选择**“大小范围”**或**“按大小排名”**来筛选文件。 **大小范围**：可从以下选项中选择：

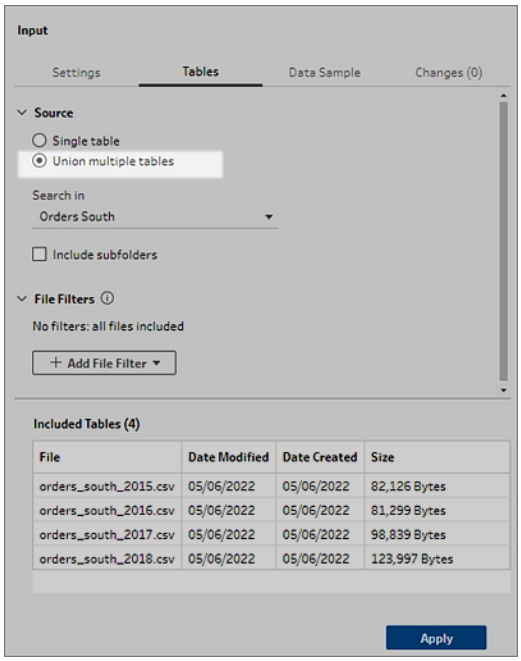
小	• 指定值范围。 • 选择**“小于”**、**“小于或等于”**、**“大于或等于”**或**“大于”**运算符并将其应用于单个值。 **按大小排名**：包括或排除 N 个最大或最小的文件。
创建日期	通过选择**“日期范围”**、**“相对日期”**或**“按日期排名”**来筛选文件。 **日期范围**：可从以下选项中选择： • 指定日期和时间范围。 • 选择**“早于”**、**“早于或等于”**、**“迟于或等于”**或**“迟于”**运算符并将其应用于单个值。 **相对日期**：包括或排除精确的年、季度、月、周或日范围。您也可以配置相对于特定日期的锚点。 **注意**：“前”日期时间段包括完整的当前时间单位，即使某些日期尚未发生。举例来说，如果选择前一个月，且当前日期为 1 月 7 日，则 Tableau 将显示 1 月 1 日至 1 月 31 日的日期。 **按日期排名**：包括或排除 N 个最新或最旧的文件。
修改日期	通过选择**“日期范围”**、**“相对日期”**或**“按日期排名”**来筛选文件。 **日期范围**：可从以下选项中选择： • 指定日期和时间范围。 • 选择**“早于”**、**“早于或等于”**、**“迟于或等于”**或**“迟于”**运算符并将其应用于单个值。 **相对日期**：包括或排除精确的年、季度、月、周或日范围。您也可以配置相对于特定日期的锚点。 **注意**：“前”日期时间段包括完整的当前时间单位，即使某些日期尚未发生。举例来说，如果选择前一个月，且当前日期为 1 月 7 日，则 Tableau 将显示 1 月 1 日至 1 月 31 日的日期。 **按日期排名**：包括或排除 N 个最新或最旧的文件。

注意：以下说明因您的 Tableau Prep Builder 版本而异。

创建输入并集

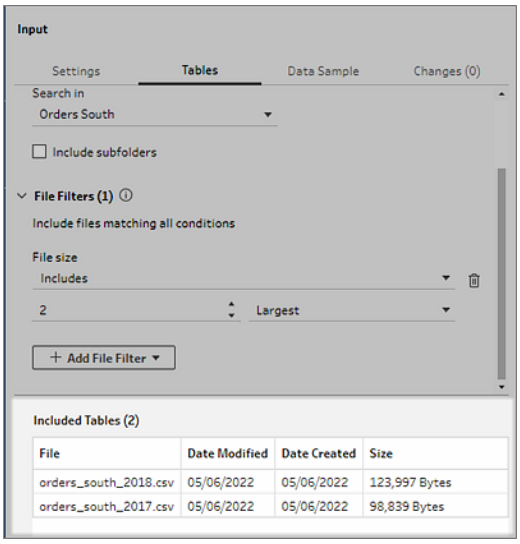
适用于 Tableau Prep Builder 版本 2022.2.1 及更高版本

1. 单击**“添加连接”**按钮，并在**“连接”**下单击**“文本文件”**（适用于 .csv 文件）、**“Microsoft Excel”**（适用于 Excel 文件）或**“统计文件”**（适用于统计文件，然后选择要打开的文件。
2. 在**“输入”**窗格中，选择**“表”**选项卡，然后选择**“合并多个表”**。



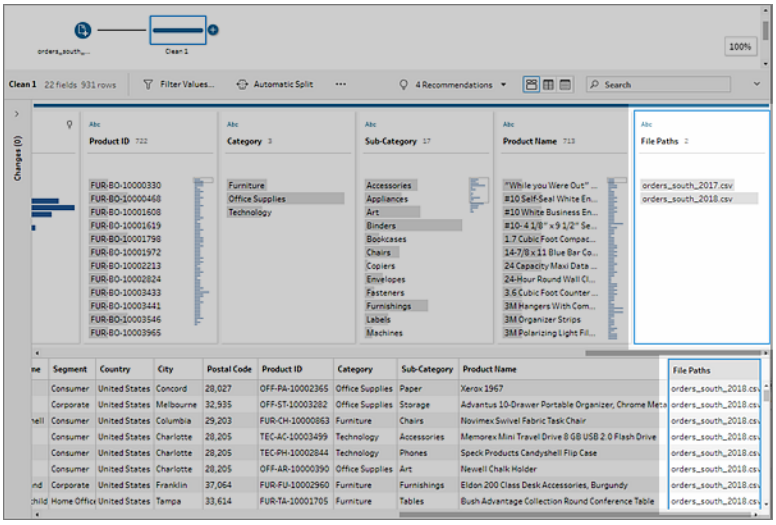
3. 选择要搜索的文件夹。您还可以包括给定目录下列出的所有子文件夹以扩展搜索。
4. 单击**“添加文件筛选器”**并从以下选项中进行选择：
   - **文件名**：输入要搜索的名称模式。
   - **文件大小**：按大小范围或按大小排名搜索。
   - **创建日期**：按日期范围、相对日期或按日期排名搜索。
   - **修改日期**：按日期范围、相对日期或按日期排名搜索。
5. 再次单击**“添加文件筛选器”**以添加更多筛选器。

   筛选结果显示在**“包含的表”**部分。



6. 单击**“应用”**以合并文件。

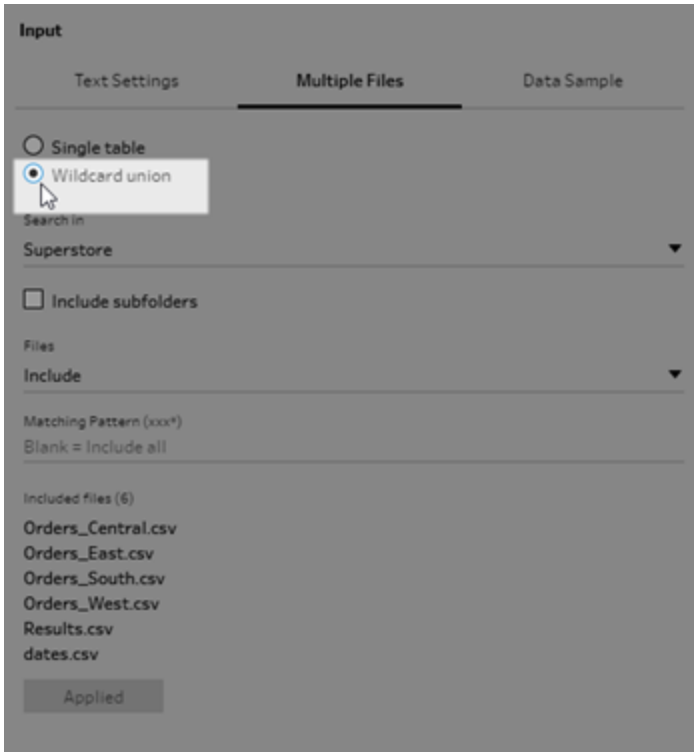
向流程中添加新步骤时，您可以在“配置”窗格的**“文件路径”**字段中看到添加到数据集的所有文件。此字段是自动添加的。



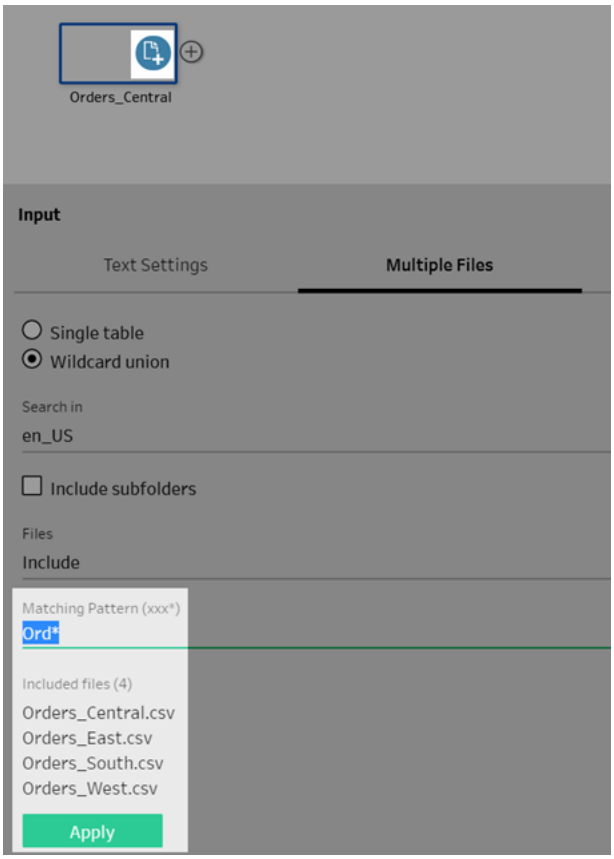
创建输入并集(版本 2022.1.1 及更低版本)

1. 单击**“添加连接”**按钮，并在**“连接”**下单击**“文本文件”**(适用于 .csv 文件)或**“Microsoft Excel”**(适用于 Excel 文件)，然后选择要打开的文件。

2. 在**“输入”**窗格中，选择**“多个文件”**选项卡，然后选择**“通配符并集”**。

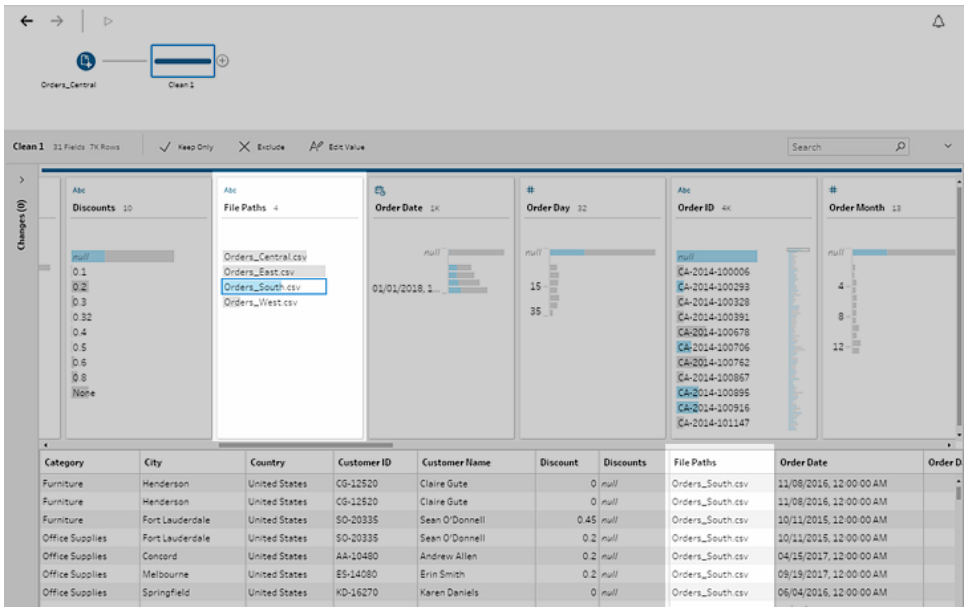


下面的示例显示一个使用匹配模式的输入并集。**“流程”**窗格中“Orders_Central”输入步骤上文件图标上的加号表明该步骤包括一个输入并集。并集中的文件列在**“已包括文件”**下。



3. 使用搜索、文件和匹配模式选项来查找要合并的文件。
4. 单击**“应用”**以合并文件。

向流程中添加新步骤时，您可以在“配置”窗格的**“文件路径”**字段中看到添加到数据集的所有文件。此字段是自动添加的。



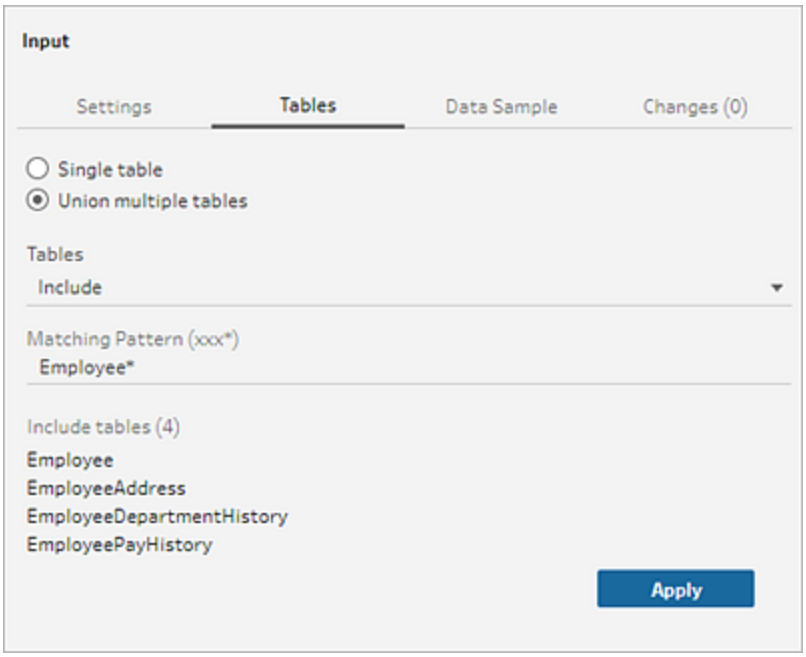
并集数据库表

在 Tableau Prep Builder 版本 2018.3.1 及更高版本中受支持

注意：数据库表的输入并集接口已在 Tableau Prep Builder 版本 2022.2.1 中更新。根据您的版本，您的选项可能看起来不同。

1. 单击**“添加连接”**按钮，并在**“连接”**下连接到支持输入并集的数据库。
2. 将表拖到流程窗格。
3. 在**“输入”**窗格中，选择**“表”**选项卡，然后选择**“合并多个表”**。

   *在以前的版本中，选择**“多个表”**选项卡，然后选择**“通配符并集”**。*



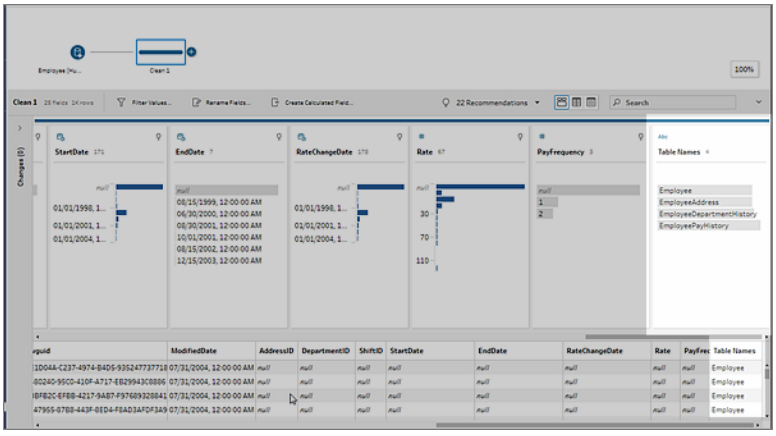
4. 在**“表”**字段中，从下拉选项中选择**“包括”**或**“排除”**，然后输入匹配模式以查找要合并的表。

   *在以前的版本中，使用搜索、**“表”**和**“匹配模式”**选项。*

   只有显示在**“连接”**窗格的**“表”**部分中的表才能包括在并集中。输入并集搜索不会跨架构或跨数据库连接进行搜索来查找表。

5. 单击**“应用”**以合并表数据。

   向流程中添加新步骤时，您可以在“配置”窗格的**“表名称”**字段中看到添加到数据集的所有表。此字段是自动添加的。



在并集之后合并字段

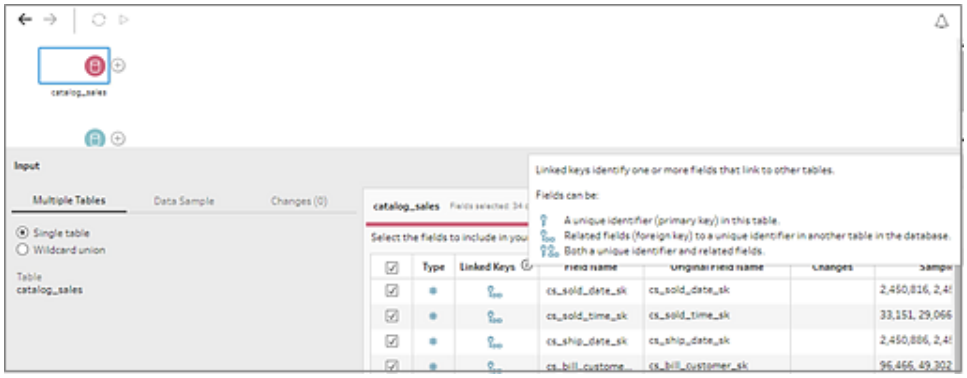
在输入步骤中创建并集后，您可能需要合并字段。您可以在除“输入”或“输出”步骤之外的任何后续步骤中执行此操作。有关详细信息，请参见其他合并字段选项 在本页351。

在“输入”步骤中联接数据

在 Tableau Prep Builder（版本 2019.3.1 及更高版本）中和 Web 上，当连接到的数据库包含具有关系数据的表时，Tableau Prep 可以检测和显示表中的哪些字段被标识为唯一标识符，那些字段被标识为相关字段，并为这些字段显示相关的表名称。

一个名为**“链接键”**的新列将显示在“输入”窗格中，并显示以下关系（如果存在）：

- 唯一标识符。此字段唯一标识表中的每一行。表中可以有多个唯一标识符。字段中的值必须唯一，不能为空白或 null。
- 相关字段。此字段将表与数据库中的另一个表关联。表中可以有多个相关字段。
- 唯一标识符和相关字段。字段是此表中的唯一标识符，并将表与数据库中的另一个表关联。



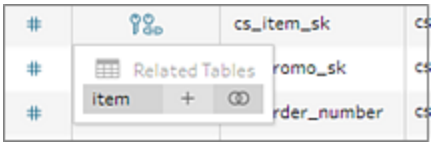
您可以利用这些关系快速查找并向流程中添加相关表，或者通过“输入”步骤创建联接。此功能可用于定义了表关系的任何受支持的数据库连接器。

1. 连接到包含字段关系数据（例如唯一标识符或相关字段）的数据库（例如 Microsoft SQL Server）。

2. 在“输入”窗格中，单击标记为 相关字段或 同时标记为唯一标识符和相关字段的字段。

   此时将打开一个对话框，其中显示相关表的列表。

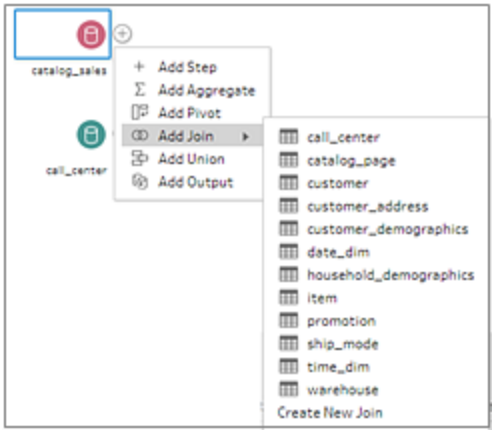
3. 将鼠标指针悬停在要添加或联接的表上，然后单击加号按钮将该表添加到流程中，或者单击联接按钮创建与所选表的联接。



   如果创建联接，Tableau Prep 使用定义的字段关系来联接表，并显示将用于创建联接的联接子句的预览

4. 或者，您可以通过“流程”窗格的菜单来联接相关表。单击加号 图标，然后选择 **“添加联接”**查看相关表的列表。Tableau Prep 基于组成两个表之间的关系的字段创建联接。

   注意：如果表没有定义表关系，则此选项不可用。



有关使用联接的详细信息，请参见**联接数据** 在本页340。

构建和组织流程

注意：从版本 2020.4.1 开始，您现在可以在 Tableau Server 和 Tableau Cloud 中创建和编辑流程。除非特别指出，本主题中的内容适用于所有平台。有关在 Web 上制作流程的详细信息，请参见 Tableau Server 和 Tableau Cloud 帮助中的 **Web 上的 Tableau Prep**。

连接到要包括在流程中的数据后，您可以首先通过向流程中添加新步骤或在现有步骤之间插入步骤来清理和调整数据。

若要组织流程，您可以更改默认步骤颜色、添加说明以提供步骤或清理操作的上下文，或者重新组织流程布局以便能更轻松地跟进复杂流程。

添加或插入步骤

构建流程时，您可以添加不同的步骤类型来执行所需的操作。例如，添加**“清理步骤”**来执行像拆分字段、应用筛选器这样的操作，或执行各种其他操作来清理不良数据。使用**“联接”**或**“合并”**步骤类型来合并数据表，或者添加**“脚本”**步骤类型，以将 R 或 Python 脚本合并到流程中。

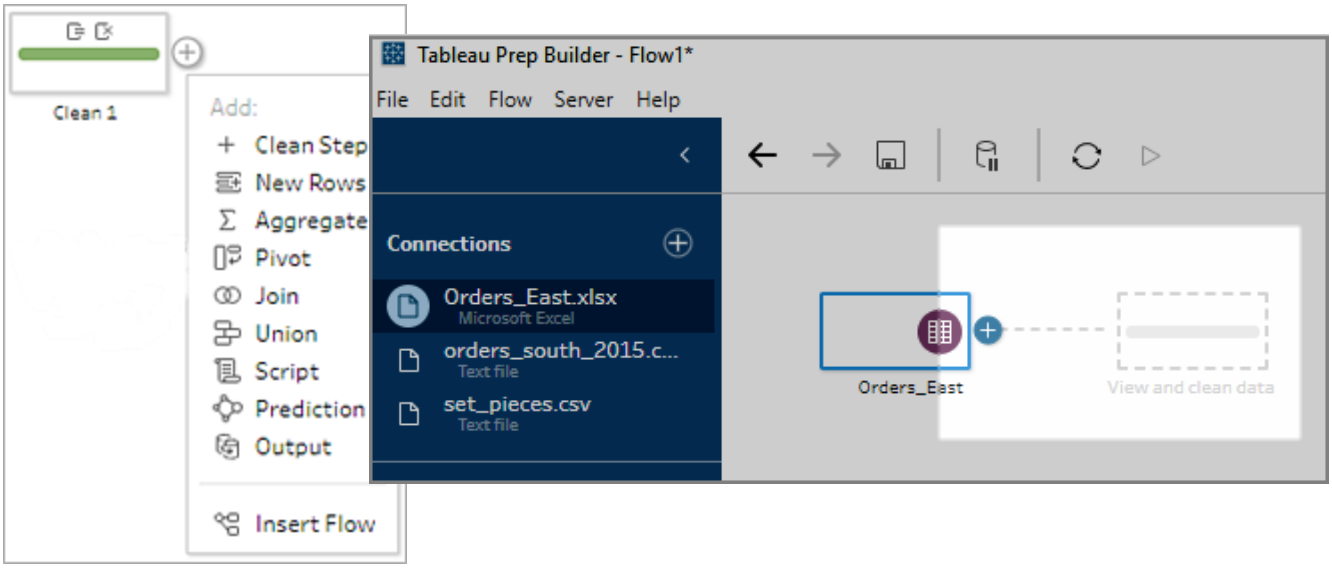
当流程开始成形时，您可能需要回到流程中较早的步骤，并插入不同的步骤类型来执行各种操作，比如添加其他清理步骤，或者聚合数据以使用与后续步骤相同的详细级别。

注意：您看到的菜单选项将因 Tableau Prep Builder 版本以及是添加步骤在流程中构建新步骤还是在现有步骤中插入步骤而异。如果使用的是 Tableau Prep Builder 版本 2019.3.1 或更低版本，请参考该部分来查看菜单选项。

您无法使用这些菜单添加输入步骤。作为替代，您需要将表从**“连接”**窗格拖到“流程”窗格。有关详细信息，请参见**连接到数据** 在本页 69。

添加步骤

连接到数据并将表格拖到画布上后，单击加号按钮 从菜单中选择步骤类型，或单击建议的清理步骤（Tableau Prep Builder 版本 2020.3.3 及更高版本以及在 Web 上）以自动向流程中添加清理步骤。



选择一种步骤类型:

- **清理步骤**:添加清理步骤以执行各种清理操作。有关可使用的不同清理操作的详细信息,请参见**清理和调整数据** 在本页213。

  注意:在 Tableau Prep Builder 版本 2019.4.2 中,**"添加分支"**选项已替换为**"清理步骤"**选项。若要将流程拆分为不同的分支,请单击两个现有步骤之间的加号 按钮,然后从**"添加"**菜单中选择步骤类型。

- **新建行**:生成新行以填补顺序数据集中的空白。有关详细信息,请参见**填补顺序数据中的空白** 在本页256。

  聚合:创建**"聚合"**步骤以选择字段并更改其详细级别。有关详细信息,请参见**对值进行聚合和分组** 在本页339。

- **转置**:创建一个**"转置"**步骤以执行各种转置选项,例如将列数据转换为行,或将行数据转换为列。您还可以设置通配符转置以自动向转置添加新数据。有关详细信息,请参见**转置您的数据** 在本页313。

- **联接**:创建一个**"联接"**步骤以合并数据表。通过菜单选项创建联接时,您必须手动将其他输入添加到联接,并添加联接子句。作为替代方案,您可以通过拖放步骤(如下所示)来自动联接文件。有关创建联接的详细信息,请参见**联接数据** 在本页340。

如果连接到的数据库包含具有关系数据的表，您也可以通过“流程”窗格中的菜单创建联接。有关使用此方法联接表的详细信息，请参见**在“输入”步骤中联接数据** 在本页133。

- **并集**：创建一个“并集”步骤。通过将表拖动到步骤并将其放在显示的**“添加”**选项上，将表添加到并集中。作为替代方案，您可以将步骤拖放到另一个步骤上来合并文件。有关创建并集的详细信息，请参见**合并数据** 在本页346。
- **脚本**（Tableau Prep Builder 版本 2019.3.1 及更高版本以及在 Web 上）：创建一个“脚本”步骤，以便在流程中包含 R 和 Python 脚本。Tableau Cloud 当前不支持脚本步骤。有关详细信息，请参见**在流程中使用 R 和 Python 脚本** 在本页322。
- **预测**：使用 Einstein Discovery 驱动的模型对流程中的数据预测进行批量评分。有关详细信息，请参见**将 Einstein Discovery 预测添加到您的流程** 在本页352。
- **输出**：创建一个“输出”步骤，以将输出保存到数据提取文件 (.hyper)，一个 .csv 文件，将输出作为数据源发布到服务器，或将流程输出写入数据库。

  在 Web 上，当前不支持将输出步骤保存到文件。有关输出类型的详细信息，请参见**保存和共享工作** 在本页361。
- **粘贴**：添加从同一流程中复制的步骤。有关在同一流程中复制和粘贴步骤的详细信息，请参见**清理和调整数据** 在本页213。
- **插入流程**（Tableau Prep Builder 版本 2019.3.2 及更高版本以及在 Web 上）：将从其他流程保存的流程步骤添加到当前流程中。您可以将它们添加到步骤的末尾，或将其插入到现有步骤之间。有关在流程步骤中使用已保存流程步骤的详细信息，请参见**创建可重用的流程步骤** 在本页253

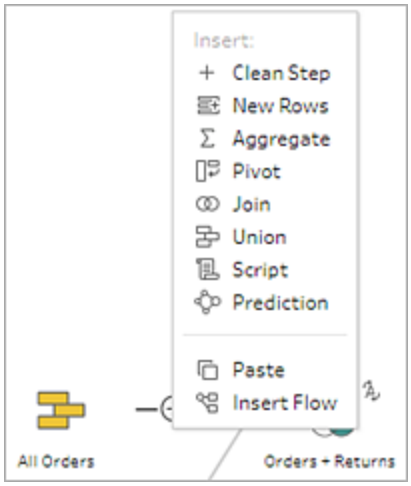
> **注意：**在 Tableau Prep Builder 版本 2019.4.2 中，此选项已添加到此菜单中。在以前的版本中，您可以通过流程窗格的空白区域中右键单击或按住 Ctrl (MacOS) 单击来添加流程步骤。

插入步骤

在现有步骤之间插入一个步骤。“输入”和“输出”步骤类型在此菜单中不可用。选项因产品版本而异。如果使用 Tableau Prep Builder 的早期版本，请参考下面的“**版本 2019.3.1 及更低版本**”部分。

1. 将光标悬停在要在其中插入流程的流程线的中间，直至出现加号 ⊕ 图标。然后单击该图标并选择一种步骤类型。

   注意：您的选项可能看起来不同，具体取决于您的产品版本。例如，Tableau Prep Builder 版本 2019.4.2 的此菜单中增加了**“插入流程”**。



2. 选择一种步骤类型：

   - **清理步骤**：在现有步骤之间插入清理步骤以执行各种清理操作。有关可使用的各种清理操作的详细信息，请参见**清理和调整数据** 在本页213。
   - **新建行**：生成新行以填补顺序数据集中的空白。有关详细信息，请参见**填补顺序数据中的空白** 在本页256。
   - **聚合**：在现有步骤之间插入一个**“聚合”**步骤，以选择字段并更改其详细级别。有关详细信息，请参见**对值进行聚合和分组** 在本页339。
   - **转置**：在步骤之间创建一个**“转置”**步骤以执行各种转置选项，例如将列数据转换为行，或将行数据转换为列。您还可以设置通配符转置以自动向转置添加新数据。有关详细信息，请参见**转置您的数据** 在本页313。

- **联接**：在现有步骤之间插入一个**“联接”**步骤。通过菜单选项创建联接时，您必须手动将其他输入添加到联接，并添加联接子句。作为替代方案，您可以通过拖放步骤（如下所示）来自动联接文件。

  有关创建联接的详细信息，请参见**联接数据** 在本页340。

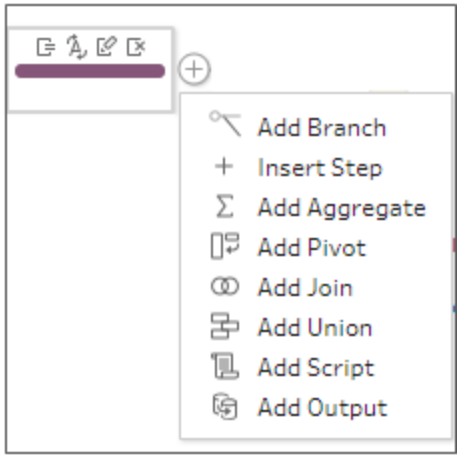
  如果连接到的数据库包含具有关系数据的表您也可以通过“流程”窗格中的菜单创建联接。有关使用此方法联接表的详细信息，请参见**在“输入”步骤中联接数据** 在本页133。

- **并集**：插入一个“并集”步骤。通过将表拖动到步骤并将其放在显示的**“添加”**选项上，将表添加到并集中。作为替代方案，您可以将步骤拖放到另一个步骤上来合并文件。有关创建并集的详细信息，请参见**合并数据** 在本页346。
- **脚本**（Tableau Prep Builder 版本 2019.3.1 及更高版本以及在 Web 上）：插入一个“脚本”步骤，以便在流程中包含 R 和 Python 脚本。Tableau Cloud 当前不支持脚本步骤。有关详细信息，请参见**在流程中使用 R 和 Python 脚本** 在本页322。
- **预测**：使用 Einstein Discovery 驱动的模型对流程中的数据预测进行批量评分。有关详细信息，请参见**将 Einstein Discovery 预测添加到您的流程** 在本页352。
- **粘贴**：在现有步骤之间插入从同一流程中复制的步骤。有关在同一流程中复制和粘贴步骤的详细信息，请参见**清理和调整数据** 在本页213。
- **插入流程**（Tableau Prep Builder 版本 2019.3.2 及更高版本以及在 Web 上）：将从其他流程保存的流程步骤插入到当前流程中。您可以将它们添加到步骤的末尾，或将其插入到现有步骤之间。有关在流程步骤中使用已保存流程步骤的详细信息，请参见**创建可重用的流程步骤** 在本页253。

  注意：在 Tableau Prep Builder 版本 2019.4.2 中，此选项已添加到此菜单中。在以前的版本中，您可以通过流程窗格的空白区域中右键单击或按住 Ctrl (MacOS) 单击来插入流程步骤。

版本 2019.3.1 及更低版本

1. 将光标悬停在步骤上，直至出现加号 ⊕ 图标，然后单击该图标并选择一种步骤类型。**插入步骤**：在步骤之间插入一个清理步骤。所有其他选项将从流程中创建一个分支。



2. 从以下选项中选择：

   - **添加分支**：将流程拆分为不同的分支。
   - **插入步骤**：在现有步骤之间插入清理步骤以执行各种清理操作。有关可使用的各种清理操作的详细信息，请参见**清理和调整数据** 在本页213。
   - **添加聚合**：创建一个**"聚合"**步骤，您可以在其中选择要聚合或分组的字段。有关详细信息，请参见**对值进行聚合和分组** 在本页339。
   - **添加转置**：创建一个**"转置"**步骤，您可以在其中执行各种转置选项，将列数据转换为行，或将数据行转换为列。有关详细信息，请参见**转置您的数据** 在本页313。
   - **添加联接**：创建一个**"联接"**步骤，您可以在其中手动将其他输入添加到联接，并添加联接子句。作为替代方案，您可以通过拖放步骤来联接文件。以下示例中显示了将"Orders_Central"输入步骤拖放到**"联接"**上的操作：

有关创建联接的详细信息，请参见**联接数据** 在本页340。

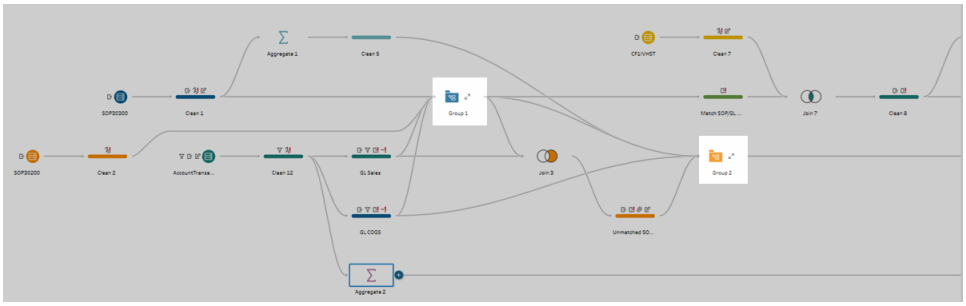
在 Tableau Prep Builder 版本 2019.1.3 及更高版本中，如果连接到的数据库包含具有关系数据的表，您也可以通过“流程”窗格中的菜单创建联接。有关使用此方法联接表的详细信息，请参见**在“输入”步骤中联接数据** 在本页133。

- **添加并集**：创建一个“并集”步骤。通过将表拖动到步骤并将其放在显示的**“添加”**选项上，将表添加到并集中。作为替代方案，您可以将步骤拖放到另一个步骤上来合并文件。有关创建并集的详细信息，请参见**合并数据** 在本页346。
- **添加脚本**（版本 2019.3.1 及更高版本）：创建一个“脚本”步骤以在流程中包含 R 和 Python 脚本。有关详细信息，请参见**在流程中使用 R 和 Python 脚本** 在本页322。
- **添加输出**：选择此选项以将输出保存到数据提取文件 (.hyper)、.csv 文件，或者将输出作为数据源发布到服务器。

对步骤进行分组

在 Tableau Prep Builder 版本 2020.3.3 及更高版本中受支持，以及从版本 2020.4 开始，在 Tableau Server 或 Tableau Cloud 上受支持。

使用“分组”选项将大型复杂流程的各个部分划分到文件夹中，以便更轻松地跟踪流程并对其进行故障排查，或者与他人共享流程。您可以更改组的颜色、添加说明、将分组步骤复制并粘贴到流程的其他区域，（或者在 Tableau Prep Builder 中）甚至将分组的步骤保存到服务器上的文件以在其他流程中重用这些步骤。



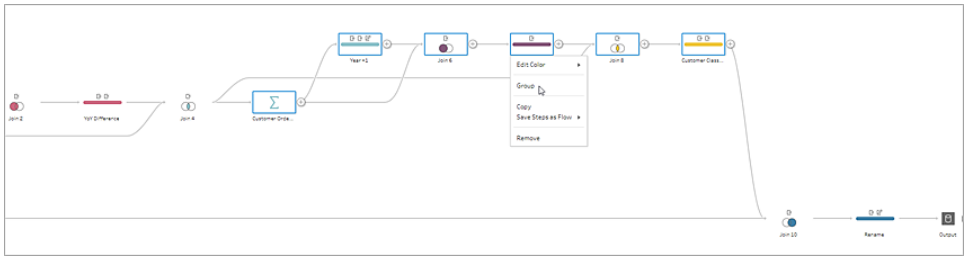
对步骤进行分组的要求

- 步骤必须与流程线直接连接。
- 组中一次只能包括一个步骤。
- 无法嵌套组。

- 只要维护组中各步骤之间的流程线连接，即可随时在组中添加或移除步骤。这也适用于从已包含在组中的流程中移除步骤。在这种情况下，系统会自动将组取消分组。

创建组

在流程中选择一组连接的步骤(您也可以拖动以一键选择多个步骤)，然后在所选步骤上右键单击或按住 Cntrl 单击 (MacOS)，然后从菜单中选择**“分组”**。

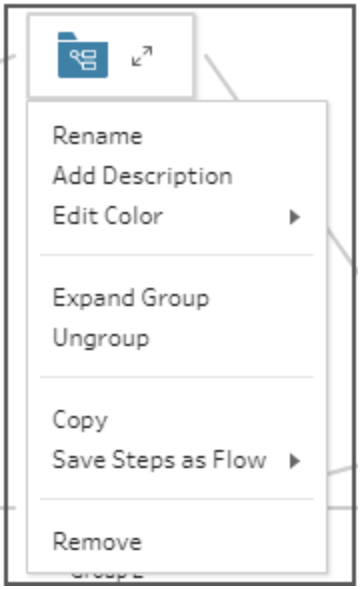


创建组后，可以执行以下任意操作：

- 单击双箭头可随时展开 或折叠 组。
- 通过拖动连接的步骤并将其拖放到折叠的文件夹中，向组添加更多步骤。
- 从组中移除步骤。在展开状态中，右键单击或按住 Cntrl 单击 (Mac OS) 一个步骤，然后选择**“从组中移除”**。

  > **注意：**如果您尝试移除中断组连续性的步骤，则此选项不可用。

- 在折叠状态下，右键单击或按住 Cntrl 单击 (MacOS) 可打开菜单，然后从以下选项中进行选择：



- **重命名**:更改组名称。
- **添加说明**:输入组的说明
- **编辑颜色**:更改组文件夹的颜色。这不会更改组中各个步骤的颜色。
- **展开组**:显示组中的所有步骤。您还可以单击双箭头 展开组。
- **取消分组**:从组中移除所有步骤并删除组。
- **复制**:将组和组中的所有步骤复制到剪贴板,以粘贴到流程中的其他地方。有关使用复制和粘贴的信息,请参见**复制并粘贴步骤** 在本页247。
- **将步骤另存为流程**(仅限 Tableau Prep Builder):将分组的步骤本地保存到计算机上的文件,或将其发布到 Tableau Server 或 Tableau Cloud 以与他人共享,或在其他流中使用它。有关保存步骤以供重用的详细信息,请参见**创建可重用的流程步骤** 在本页253。
- **移除**:从流程中移除组和组中的所有步骤。

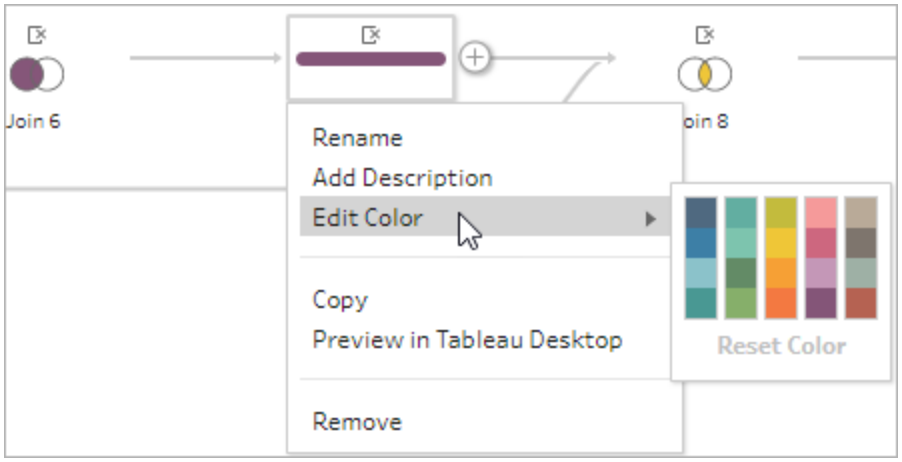
- (版本 2021.1.2 及更高版本)在展开状态中,在展开的组区域中右键单击或按住 Cntrl 单击 (MacOS) 打开菜单以折叠组或取消步骤分组。

更改流程配色方案

默认情况下，Tableau Prep 会为流程中的每个步骤分配一种颜色。此配色方案将在整个流程中应用，帮助您在应用清理步骤、联接、合并或聚合数据时在整个流程中跟踪数据，让您知道哪些文件受到操作的影响。

若要为步骤选择其他配色方案，请执行以下操作：

1. 选择一个或多个步骤。
2. 右键单击或按住 Ctrl 单击 (MacOS) 选定步骤，并选择**“编辑颜色”**。



3. 单击调色板中的一种颜色来应用该颜色。

若要将步骤颜色重置回默认颜色，请执行以下操作之一：

- 从顶部菜单中单击**“撤消”**。
- 按 Cntrl+Z 或 Command-Shift-Z (MacOS)。
- 选择更改的步骤，右键单击所选的步骤，并选择**“编辑颜色”**，然后从调色板的底部选择**“重置步骤颜色”**。

从流程中移除步骤

在流程中的任何位置，您可以移除步骤或步骤之间的流程线。

注意：您无法移除进入或退出折叠步骤组的流程线。您必须先展开组或取消对步骤进行分组。

- 若要移除步骤或流程线，请选择要移除的步骤或流程线，右键单击该元素，然后选择**“移除”**。
- 若要移除多个步骤或流程线，请执行以下操作之一：
  - 使用鼠标拖动并选择一整段流程。然后右键单击或按住 Ctrl 单击 (Mac OS) 某个选定的步骤，并选择**“移除”**。
  - 按 Ctrl+A 或 Cmd+A (MacOS) 选择流程中的所有元素，或按住 Ctrl 并单击或者按住 Cmd 并单击 (MacOS) 以选择特定元素，然后按 **Delete** 键。

为流程步骤和清理动作添加说明

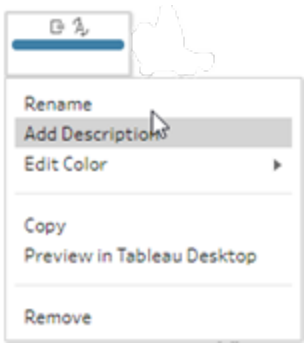
在构建流程并执行各种清理操作时，您可能需要添加说明，帮助稍后可能会查看或使用您的流程的其他人更好地理解您的步骤。您可以直接在“流程”窗格中为流程中的任何单独步骤、任何步骤组添加说明，或为**“更改”**窗格中的任何清理动作添加说明来为更改提供其他上下文。说明最多可为 200 个字符长。

有关在“更改”窗格中查看更改的详细信息，请参见**查看您所做的更改** 在本页 227。

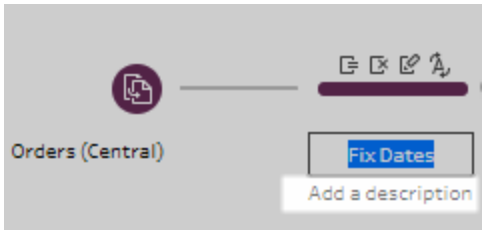
为流程步骤添加说明

当您添加说明时，会在步骤下方添加一个消息 图标。单击图标以在“流程”窗格中显示或隐藏说明文本。

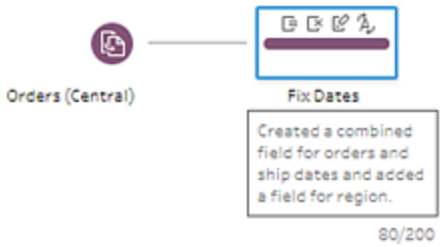
1. 在“流程”窗格中，选择一个步骤。
2. 执行以下操作之一：
   - 在步骤上右键单击或按住 Ctrl 并单击 (MacOS)，然后从菜单中选择**“添加描述”**。



- 在步骤的名称字段中双击，然后单击**“添加说明”**。

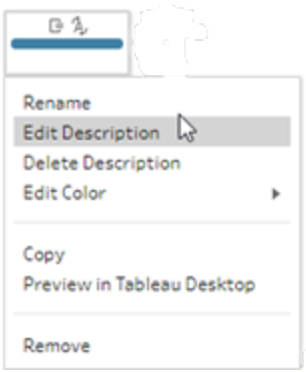


3. 在文本框中键入说明。



4. 在文本框外部单击，或按 Enter 应用您所做的更改。默认情况下，说明显示在步骤的下方。若要隐藏说明，请单击消息 图标。

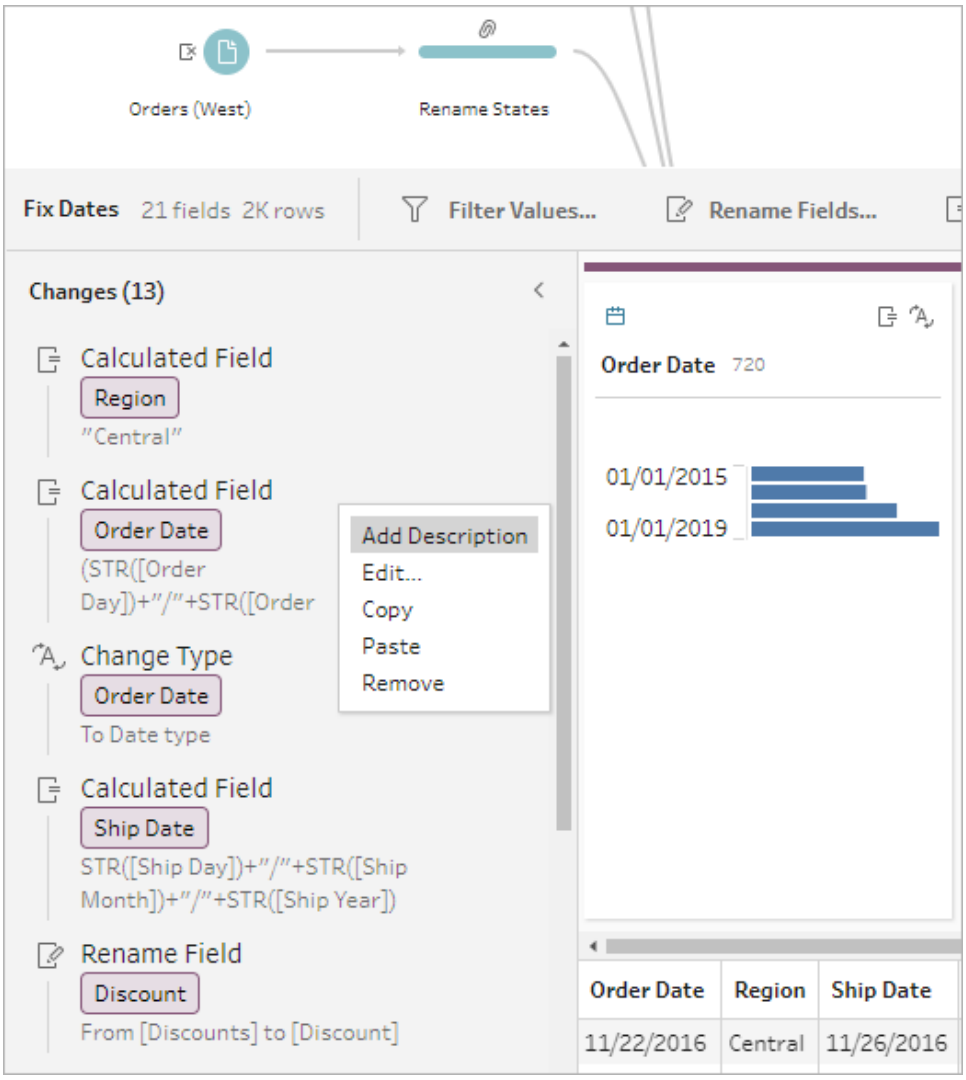
5. 若要编辑或删除描述，请右键单击或按住 Ctrl 并单击 (MacOS) 步骤或描述。然后从菜单中选择**“编辑描述”**或**“删除描述”**。



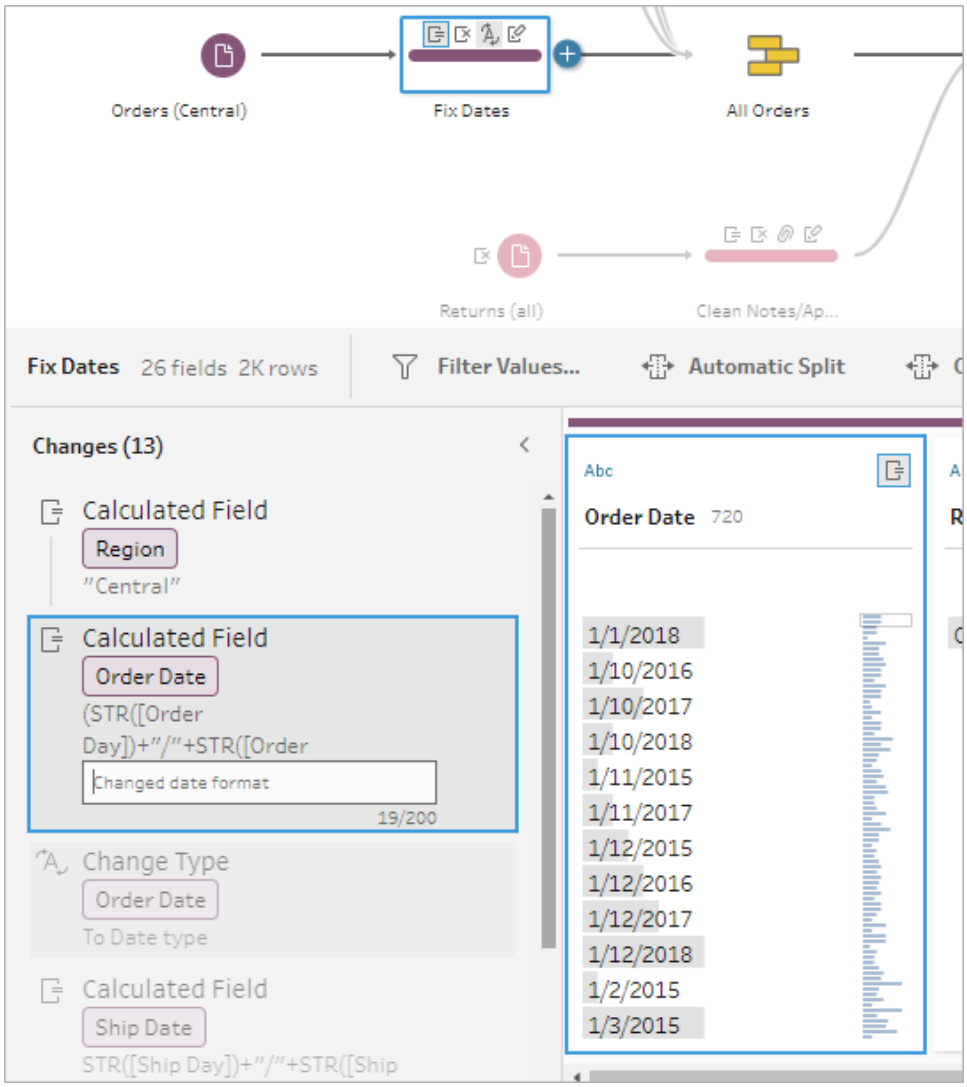
为更改条目添加说明

从 Tableau Prep Builder 版本 2019.1.1 开始以及在 Web 上，您可以在“更改”窗格中向条目添加描述。

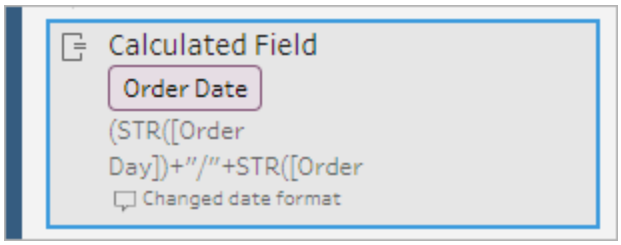
1. 在流程窗格中选择一个步骤。
2. 打开**“更改”**窗格或**“更改”**选项卡。
3. 在**“更改”**窗格中右键单击或按住 Ctrl 并单击 (MacOS) 一个条目，并选择**“添加说明”**。



4. 为更改动作输入说明。



说明将出现在为带有注释 图标的更改生成的文本的下方。



5. 若要编辑或删除说明，请右键单击或按住 Ctrl 并单击 (MacOS) 更改项，然后选择**“编辑说明”**或**“删除说明”**。

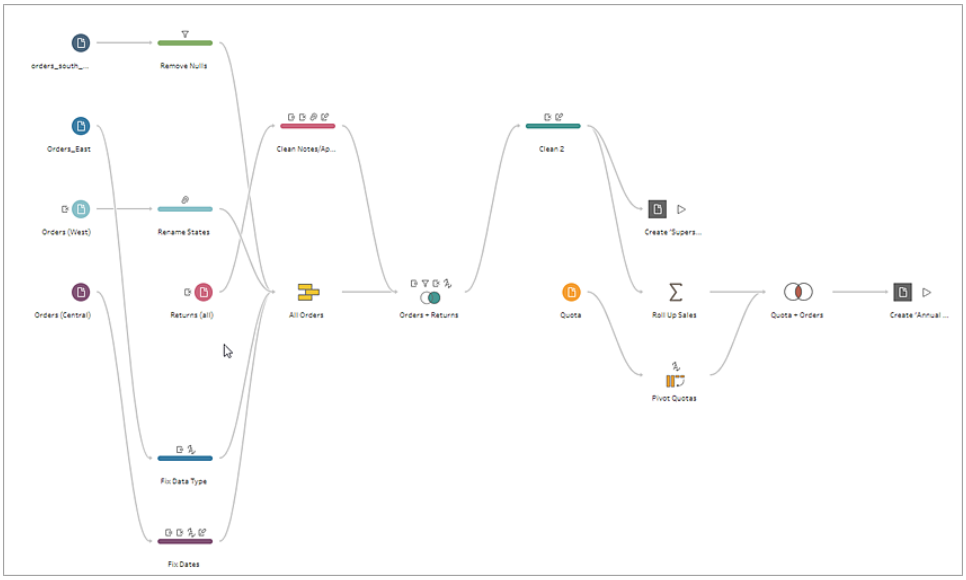
重新组织流程的布局

在 Tableau Prep Builder 版本 2019.2.2 及更高版本中受支持，以及从版本 2020.4 开始，在 Tableau Server 或 Tableau Cloud 上受支持。

当您构建流程时，Tableau Prep Builder 使用默认布局。每个流程按从左到右的顺序布局和处理，“输入”步骤位于画布最左侧的开始位置，“输出”步骤位于画布右侧的结束位置。但是，如果构建复杂的大型流程，则流程会很快变得难以执行。

您可以选择并移动步骤，使流程布局按对您有意义的方式进行组织，从而清理流程的布局。例如，您可以修复交叉的流程线条，移动流程步骤以清理额外的空白，或者重新排列流程步骤以显示清晰的事件序列。

例如，下面的流程混乱不清，难以执行：



若要清理此流程，请选择并将步骤向上、向下、向左或向右拖放到画布中的新位置。无法将流程步骤移到会中断从左到右处理流程的位置。例如，您不能将位于联接步骤之前的合并步骤拖到流程中联接步骤之后的位置。

将流程步骤拖到允许的位置时，将显示一个橙色的框。如果不允许该位置，则不会显示橙色框，当您尝试放置步骤时，它们将返回到原始位置。

若要在流程中移除步骤，请执行以下操作：

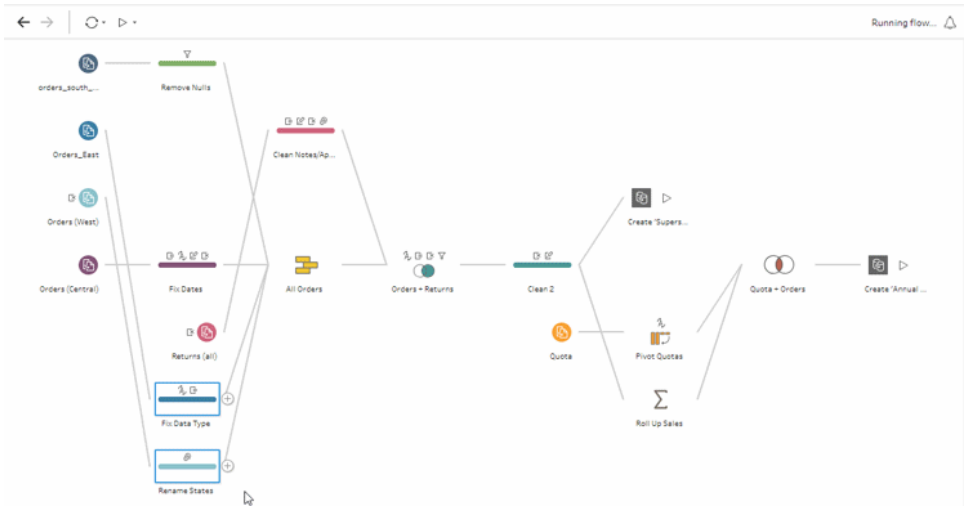
1. 在“流程”窗格中，选择要移除的步骤。您可以单击特定步骤，拖动以选择多个步骤，或按住 Ctrl 单击或按住 Cmd 单击 (MacOS) 以选择不相邻的步骤。

2. 将步骤拖放到新位置。

> **注意：**如果您不喜欢所进行的重新组织移动操作，您可以在顶部菜单中单击**“撤消”**来撤消这些操作。但是，如果在移动步骤之间执行清理操作，您也可以撤消这些操作。**“撤消”**选项会按您执行它们时的步骤撤消您的操作。

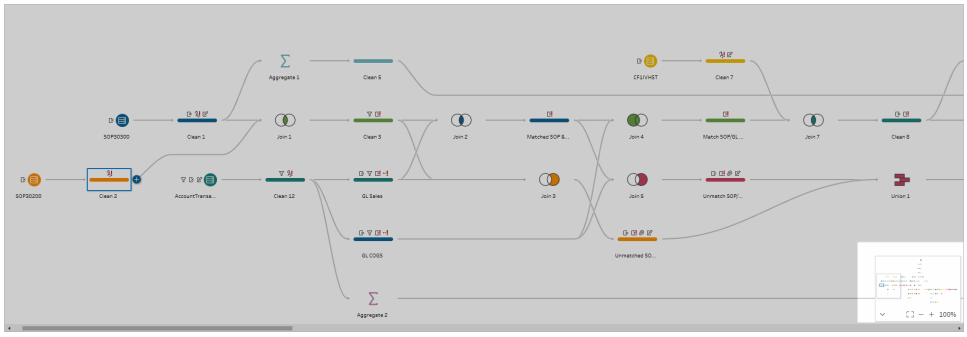
观看“重新组织流程步骤”的实际运行效果

下面的示例演示使用拖放来重新排列流程的操作。

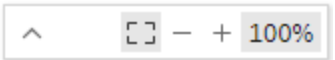


使用流程导航工具

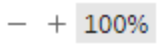
在处理大型流程时，来回滚动以搜索要返回到的特定流程区域可能很困难。利用流程导航工具可以更轻松地完成此操作。流程导航器是流程的微缩版本，显示在画布的右下角。



单击图形的任何区域可跳转到流程的该区域，或使用以下工具栏



选项进行导航：

工具栏选项	描述
	折叠流程导航器图形。在折叠状态下，您可能只看到百分比指示器。只需将光标悬停在此项上来展开工具栏，然后单击向上箭头 即可再次展开图形。
	展开流程导航器图形。
	更改流程的大小以适合屏幕。
	放大和缩小您的流程。您可以单击百分比指示器将视图还原到 100% 大小。

检查您的数据

注意：从版本 2020.4.1 开始，您现在可以在 Tableau Server 和 Tableau Cloud 中创建和编辑流程。除非特别指出，本主题中的内容适用于所有平台。有关在 Web 上制作流程的详细信息，请参见 Tableau Server 和 Tableau Cloud 帮助中的 **Web 上的 Tableau Prep**。

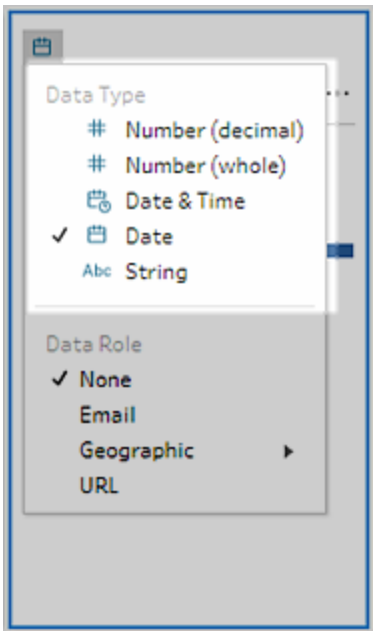
使用本主题中的选项来透彻理解数据的组成，从而更好地了解所需进行的更改，以及流程中包括的操作的效果。

查看分配给数据的数据类型

像 Tableau Desktop 一样，Tableau Prep 会在您将连接拖到**"流程"**窗格中时对数据进行解读，并自动为其分配一个数据类型。由于不同的数据库可能会以不同的方式处理数据，因此 Tableau Prep 的解读可能不会始终正确。

若要更改数据类型，请单击数据类型图标，并从上下文菜单中选择正确的数据类型。您可以将"字符串"或"整数"数据类型更改为**"日期"**或者**"日期和时间"**，Tableau Prep 将触发"Auto DateParse"来更改这些数据类型。像 Tableau Desktop 一样，如果更改未成功，您将会在字段中看到 Null 值，并且您可以创建计算来进行更改。

有关使用 DateParse 的详细信息，请参见 Tableau Desktop 和 Web 制作帮助中的将字段转换为日期字段。



在连接到以下数据源中的数据后，可以在输入步骤中更改数据类型：

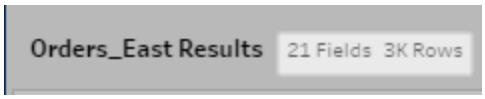
- Microsoft Excel
- 文本文件
- PDF 文件
- Box
- Dropbox
- Google 云端硬盘
- OneDrive

对于所有其他数据源，请添加清理步骤或其他步骤类型以进行此更改。若要查看不同步骤类型中可用的清理选项列表，请参见关于清理操作 在本页213。

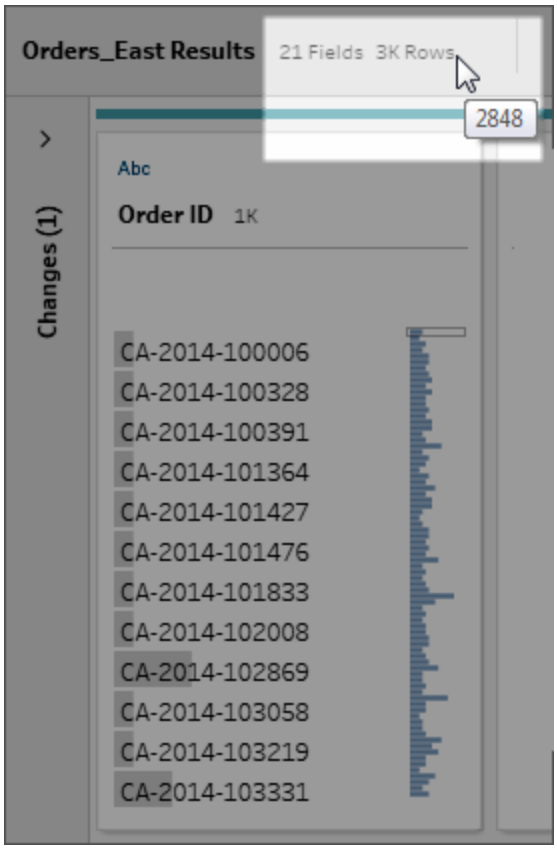
查看有关数据的大小详细信息

连接到数据后，向流程中添加一个表，然后添加一个步骤。您可以使用**“配置”**窗格来查看数据的当前状态和结构，并找出 null 值和离群值。

- **字段数和行数**：在**“配置”**窗格的左上角，您可以找到在流程的某个特定点数据中的字段数和行数的汇总。Tableau Prep 会舍入到最接近的千位。在下面的示例中，数据集中有 21 个字段和 3000 行。



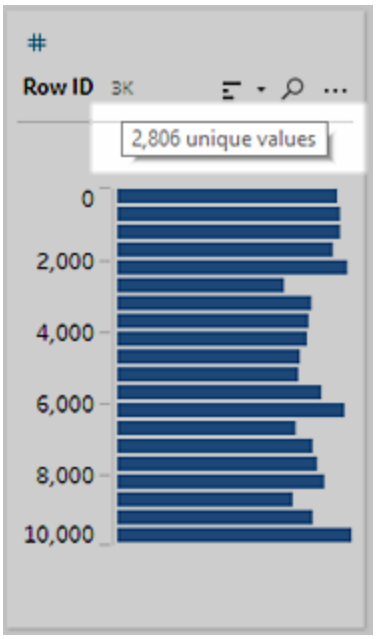
将光标悬停在字段数和行数上时，您可以看到确切的行数(本例中为 2848)。



- **数据集大小**：通过在**“输入”**窗格的**“数据样本”**选项卡中指定要包括的行数来使用数据子集。
- **已抽样**：为了让您能够与数据直接交互，Tableau Prep 会使用原始数据的子集。行数由数据类型和所呈现的字段数确定。与整数相比，字符串字段会占用更多的存储空间，因此，如果数据集中有 10 个字符串字段，与有 10 个整数字段的情况相比，您得到的行数可能会更少。

“已抽样” Sampled 标记显示在**“配置”**窗格中大小详细信息的旁边，用于指明这是数据集的子集。您可以修改流程中包含的数据量。在 Web 上创建或编辑流程时，将应用其他数据限制。有关详细信息，请参见**选择数据样本大小** 在本页 114。

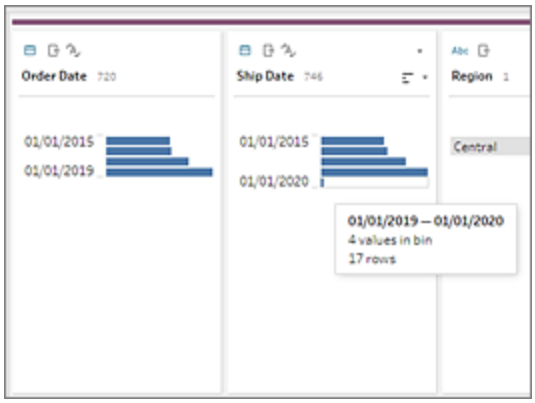
- **唯一值的数量**：每个字段标题旁边的数字表示该字段内包含的不同值。Tableau Prep 会舍入到最接近的千位。在下面的示例中，“说明”字段中呈现的不同值为 3,000 个，但如果将光标悬停在数字上，您可以看到唯一值的确切数量。



查看值或唯一值的分布

默认情况下，Tableau Prep 会将字段中的数字、日期和日期时间值分组为存储段。这些存储段也称为数据桶。数据桶确保您能查看值的整体分布状况，并快速确定离群值和 null 值。数据桶大小是基于字段中的最小值和最大值计算得出的，并且 null 值始终显示在分布的顶部。

例如，将按年份对订单和发货日期进行汇总或“分桶”。每个数据桶都表示从起始年份一月份到下一年一月份的一年，并带有相应标签。由于有处于 2018 年和 2019 年下半年的销售日期及发货日期，因此将针对这些值为以下年度创建数据桶。



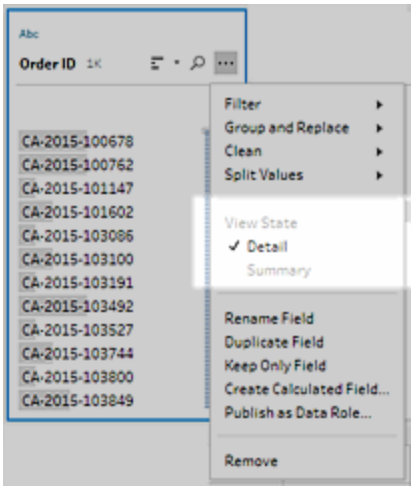
如果离散(或分类)数据字段包含许多行,或者其具有的分布太大,如果不滚动就无法显示在字段中,您可能会在字段的右侧看到汇总分布。您可以在分布中单击和滚动来定位特定值。

如果数据包含数字或日期字段,您可以切换显示值的详细(离散)版本或值的汇总(连续)版本。汇总视图显示字段中的值范围,以及某些值的出现频率。

这种切换可帮助您隔离唯一值(比如字段中的“3”记录数)或值的分布(比如字段中所有“3”记录的总和)

切换视图:

1. 在“配置”窗格、“结果”窗格或数据网格中，单击数值或日期字段的**“更多选项”** ··· 菜单。



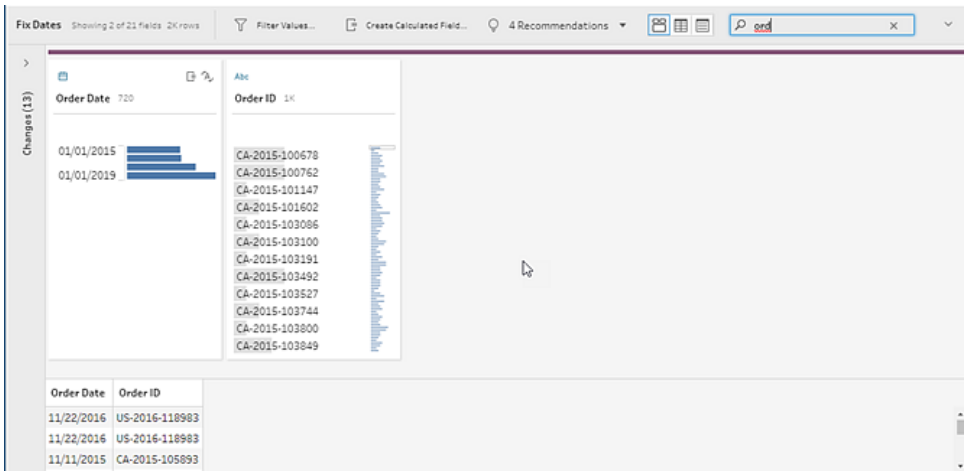
2. 在上下文菜单中，选择**“详细信息”**以查看值的详细版本，或选择**“摘要”**以查看值的分布式版本。

搜索字段和值

在“配置”窗格或“结果”窗格中，您可以搜索您特别感兴趣的字段或值，并使用搜索结果来筛选数据。

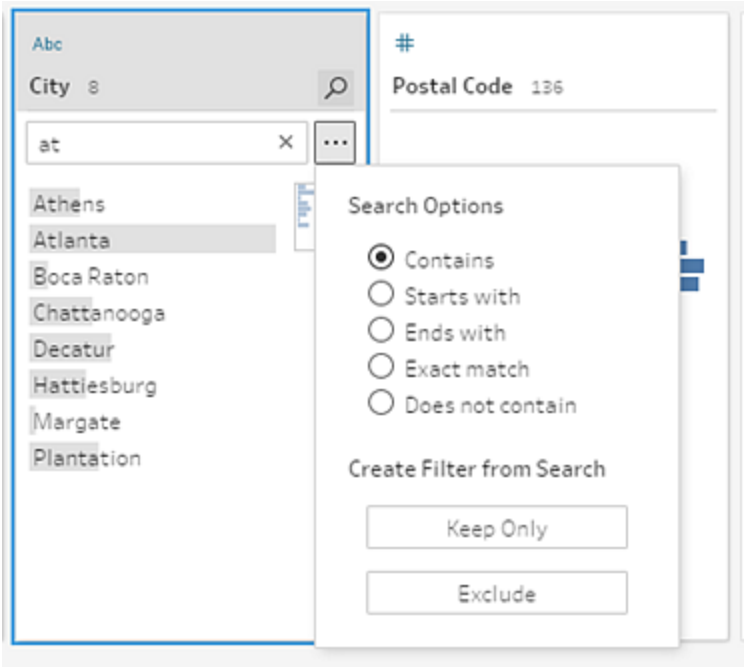
从版本 2021.1.1 开始，当您搜索字段时，将显示一个新指示符，告诉您找到的字段数，以便您更好地了解搜索结果。如果未找到字段，将显示其他消息。

若要搜索字段，请在工具栏上的搜索框中输入全部或部分搜索词。

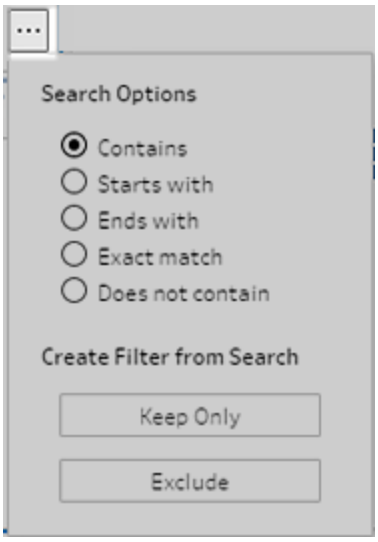


在字段中搜索值：

1. 单击字段的搜索图标 ，并输入一个值。



2. 若要使用高级搜索选项，请单击**“搜索选项...”**按钮。



3. 若要使用搜索结果来筛选数据，请选择**“只保留”**或**“排除”**。

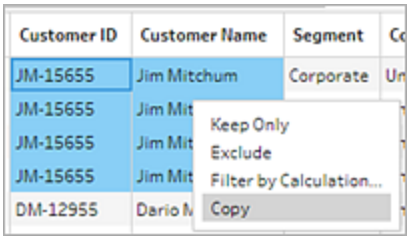
   在**“流程”**窗格中，受影响的步骤上方将出现一个筛选器图标。

复制数据网格中的字段值

在 Tableau Prep Builder、Tableau Server 版本 2022.3 及更高版本以及 Tableau Cloud 版本 2022.2(八月) 及更高版本中受支持。

轻松地从数据网格复制一组选定的值，并将它们粘贴到任何文档中，如 Microsoft Excel、文本 (.csv) 文件、电子邮件等。您甚至可以将它们复制并粘贴到 SQL 编辑器中，以快速运行 SQL 查询。

1. 在数据网格中，选择一个或多个要复制的字段值。
2. 在所选字段值上右键单击或按住 cmd 并单击 (MacOS)，然后从菜单中选择**“复制”**。您也可以使用键盘快捷键 Ctrl + C 或 cmd+C (MacOS)，或者从**“...”**工具栏菜单中选择**“复制”**。



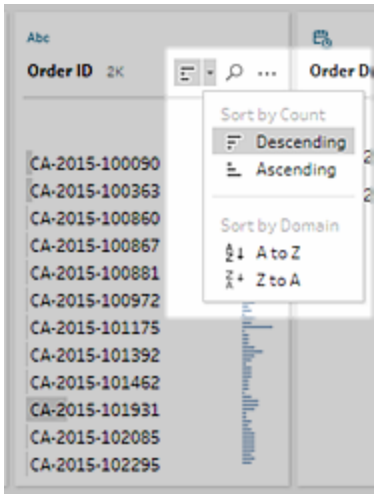
3. 将复制的字段粘贴到您的文档或位置。

> **注意：“编辑”>“复制”**当前不从数据网格复制字段值。

对值和字段进行排序

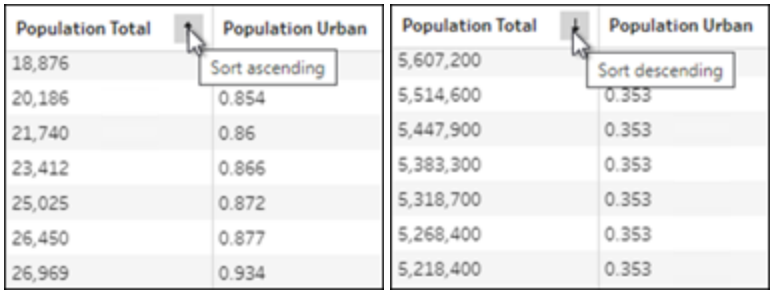
“配置”窗格

“配置”卡上的排序选项使您能按升序或降序对数据桶(由分布条形表示的值计数)进行排序或按字母顺序对单独的字段值进行排序。



“数据”网格

按升序或降序对字段的行级详细信息进行排序。



对字段进行重新排序

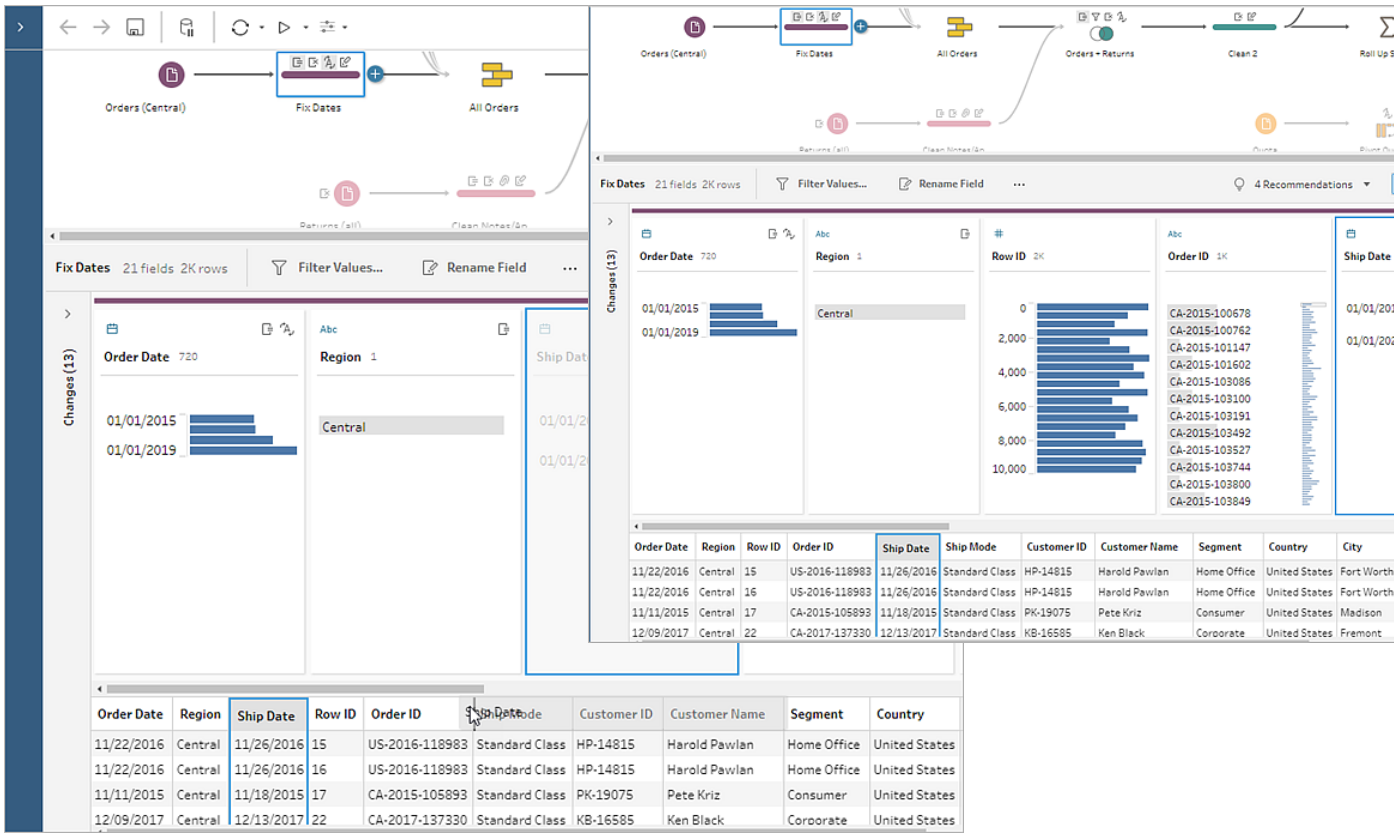
2022.2.1 及更高版本支持使用列表视图更改字段顺序。

您可以通过将配置窗格、数据网格或列表视图中的字段拖放到新位置来更改它们的顺序。

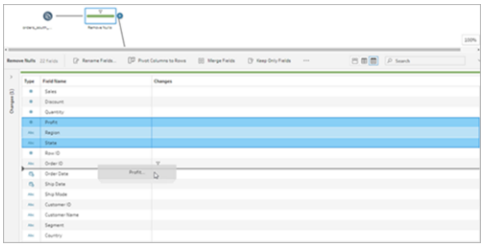
若要重新排列字段的顺序，请执行以下操作：

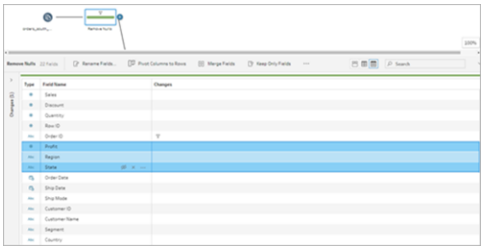
1. 从**“配置”**窗格、**“结果”**窗格、**“数据”**网格或**“列表”**视图中，选择一个或多个配置卡或字段。
2. 拖动配置卡或字段，直到看到黑色目标线出现。
3. 将配置卡或字段放置到位。
   “配置”窗格、数据网格和列表视图是同步的，因此字段将以相同的顺序出现在两个位置中。在运行和计划流程时，字段的新顺序在 Tableau 产品中保持不变。

数据网格重新排序



列表视图重新排序





突出显示流程中的字段和值

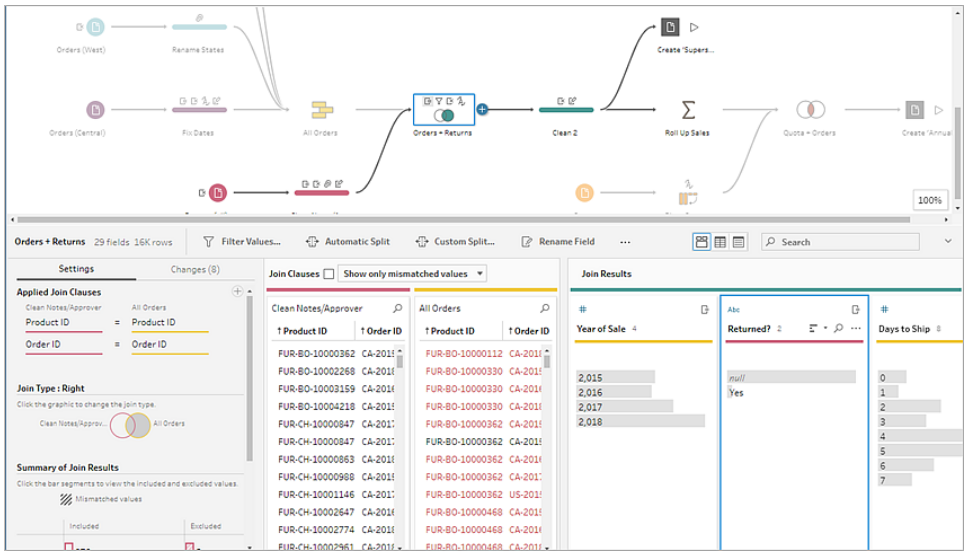
利用 Tableau Prep 可轻松地查找流程数据中的字段和值。在流程窗格中跟踪字段的来源以及在流程中的何处使用了字段，或者在配置卡或数据网格中单击单独的值以突出显示相关值或相同的值。

在流程中跟踪字段

在 Tableau Prep 中，您可以突出显示字段的每处使用位置，甚至可以跟踪字段的来源，从而可帮助您在未看到预期结果时找出缺少的值或排查流程故障。

在清理步骤的**“配置”**窗格中或任何其他步骤类型的**“结果”**窗格中单击某个字段，流程窗格将突出显示其中使用了该字段的路径。

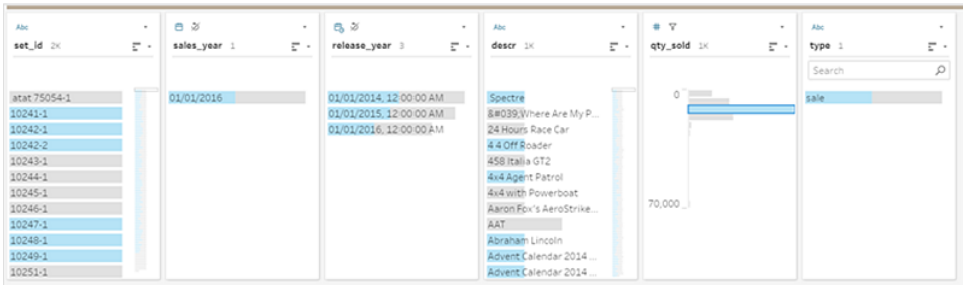
注意：此选项不可用于“输入”或“输出”步骤类型。



查看相关值

您可以使用突出显示来查找字段之间的相关值。当您在“配置”窗格或“结果”窗格的**“配置”**卡中单击某个值时，其他字段中的所有相关值会以蓝色突出显示。蓝色显示您选择的值与其他字段中的值之间的关系分布。

例如，若要突出显示相关值，请在**“配置”**窗格中单击字段中的一个值。其他字段中的相关值将变为蓝色，以蓝色突出显示的条形比例表示关联程度。



突出显示相同的值

在数据网格中选择值时，所有相同的值也会突出显示。这些突出显示可帮助您确定数据中的模式或不规则情况。

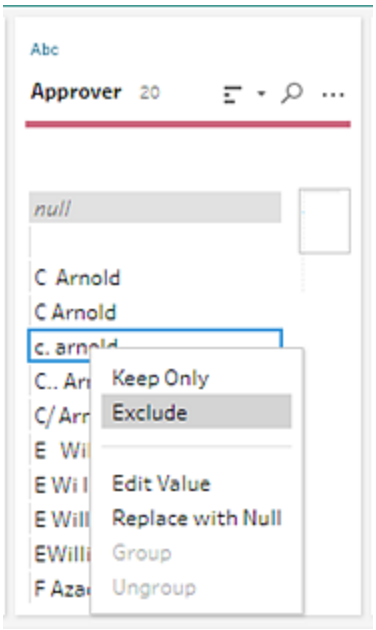
Type	Customer	Purchases	Date
Cash	Wei	5	08/18/2016
Cash	Jim	7	07/15/2016
Credit	Arnold	5	06/29/2016
Credit	Lee	1	08/07/2016
Cash	Maria	2	08/30/2016
Cash	Wendy	1	07/21/2016
Credit	Max	2	07/02/2016
Credit	Juan	1	05/10/2016
Cash	Isaac	4	06/28/2016
Credit	Philip	1	08/09/2016
Credit	Lane	5	05/04/2016

筛选您的数据

注意：从版本 2020.4.1 开始，您现在可以在 Tableau Server 和 Tableau Cloud 中创建和编辑流程。除非特别指出，本主题中的内容适用于所有平台。有关在 Web 上制作流程的详细信息，请参见 Tableau Server 和 Tableau Cloud 帮助中的 **Web 上的 Tableau Prep**。

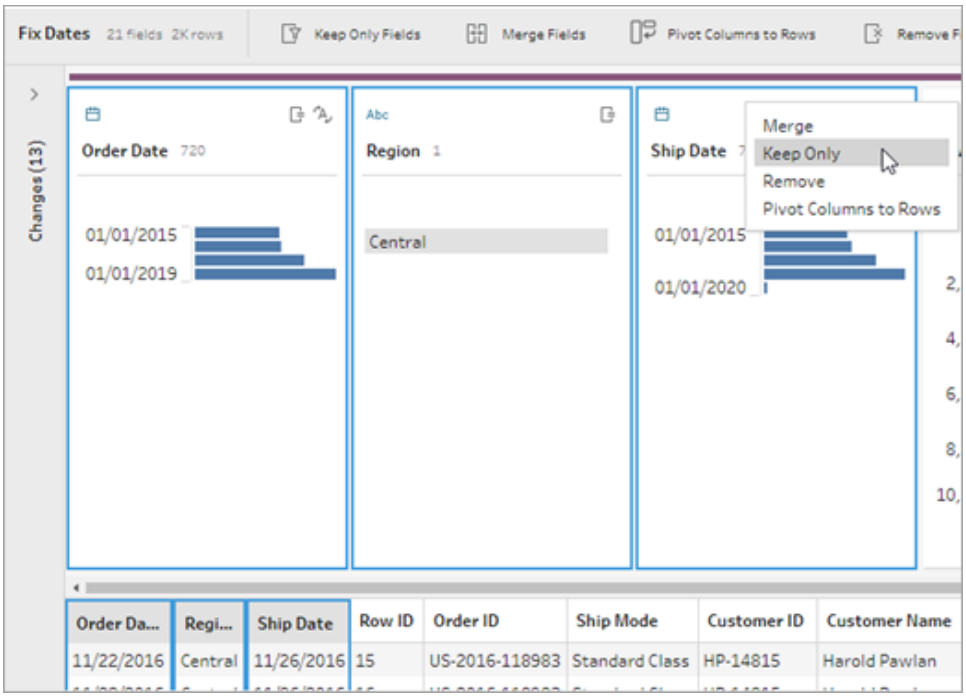
Tableau Prep 提供了可用于筛选数据的各种选项。例如，使用**“只保留”**或**“排除”**以对配置卡、数据网格或结果卡中的字段的特定值进行单击筛选，或者针对更复杂的筛选需求从各种筛选器选项中进行选择。也可以保留或移除整个字段。

在流程的任何步骤中对数据进行筛选。如果要只更改特定值，您可以选择**“编辑值”**以内联方式编辑值，或将值替换为 Null。有关编辑字段值的详细信息，请参见**编辑字段值** 在本页 234。



保留或移除字段

在处理流程中的数据时，您可能需要移除不需要的字段。在任何清理或操作步骤的“配置”窗格或数据网格中，选择一个或多个字段，并右键单击或按住 Ctrl 单击 (MacOS)，然后选择**“移除”**以移除所选字段，或选择**“只保留”**(Tableau Prep Builder 版本 2019.2.2 及更高版本以及 Web 上) 以仅保留所选字段，并移除所有未选择的字段。



隐藏字段

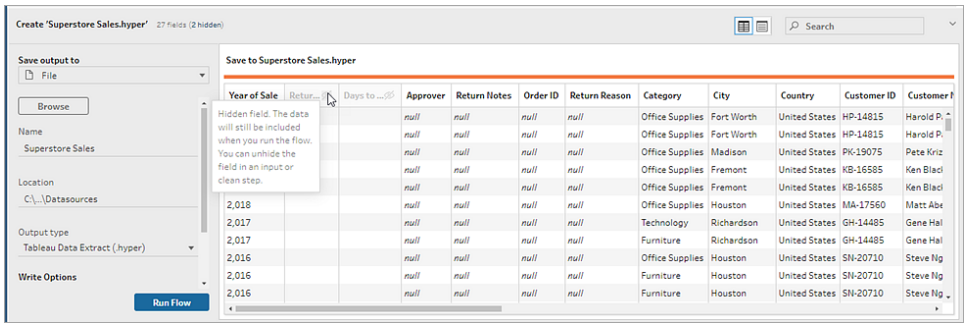
在 Tableau Prep Builder 版本 2021.1.4 及更高版本中受支持，以及从版本 2020.1 开始，在 Tableau Server 或 Tableau Cloud 上受支持。

如果您的流程中有不需要清理的字段，但是您仍想将它们包括在流程中，则可以隐藏字段而不是移除它们。在取消隐藏字段或运行流程以生成输出之前，不会加载这些字段的数据。

隐藏字段时，会自动将一个名为**“隐藏字段”**的新配置卡添加到“配置”窗格中，使您可以根据需要轻松地从列表中取消隐藏字段。

您可以在大多数操作中包括隐藏的字段，但是联接、聚合和数据转置需要取消隐藏该字段才能在这些步骤类型之一中使用它。如果您在其中一项操作中使用了该字段后将其隐藏，则该字段将显示为隐藏状态，并且该操作不会受到影响。

用眼睛 图标标记所有隐藏字段。



隐藏和取消隐藏字段

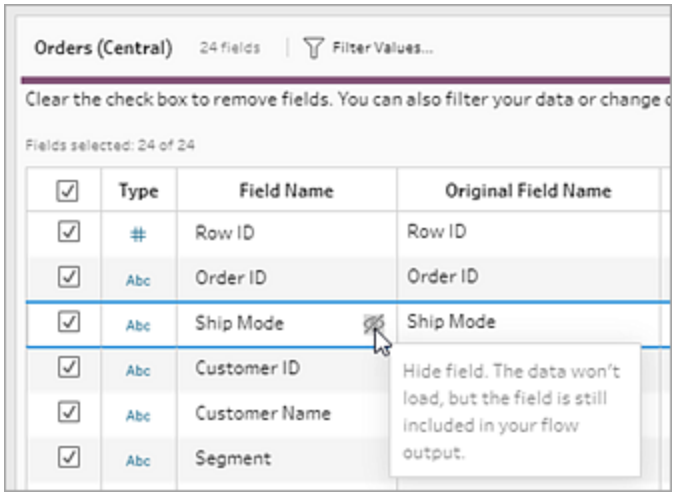
若要隐藏或取消隐藏字段，您必须处于“输入”步骤或“清理”步骤中。在“清理”步骤中，您可以从“配置”窗格、日期网格和“列表”视图中隐藏或取消隐藏字段。

从“输入”步骤中

1. 连接到数据。
2. 在“输入”步骤中选择要隐藏或取消隐藏的字段。

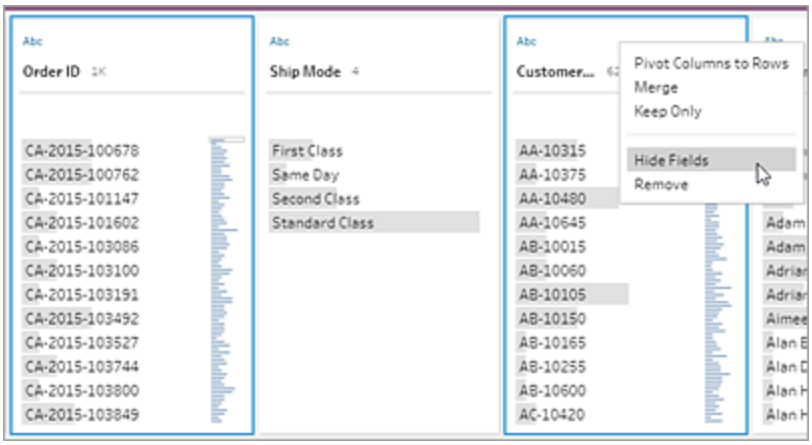
3. 点击眼睛 图标以隐藏或取消隐藏该字段。

   从版本 2023.1 开始，不支持在“输入”步骤中多选字段。



从“配置”窗格中

1. 选择要隐藏的字段。
2. 右键单击或按住 Ctrl 单击 (MacOS)，从**“更多选项”**⋯ 菜单或从工具栏菜单中选择**“隐藏字段”**或**“隐藏字段”**。



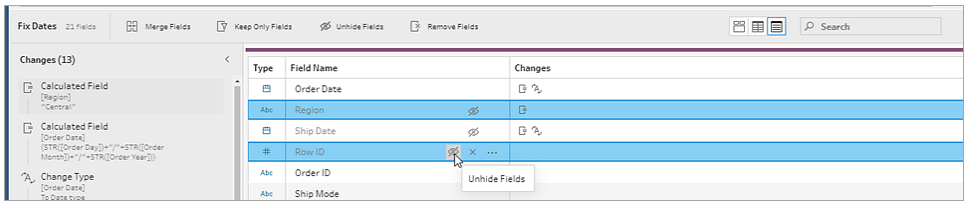
3. 将生成一个新的配置卡，其中显示了您的隐藏字段。

4. 若要取消隐藏字段，请在**“隐藏字段”**配置卡中选择一个或多个字段，然后单击眼睛 图标，右键单击或按住 Ctrl 单击 (MacOS)，然后从菜单中选择**“取消隐藏字段”**。

从“列表”视图中

1. 在“清理”步骤中的工具栏上，单击“列表”视图 图标以更改为列表视图。
2. 选择一个或多个要隐藏或取消隐藏的字段。

3. 单击眼睛 图标以隐藏或取消隐藏字段。

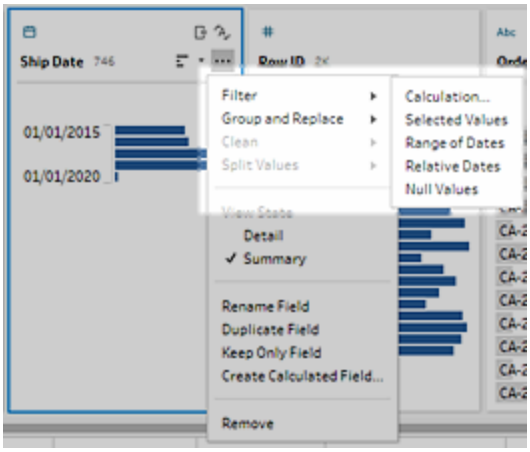


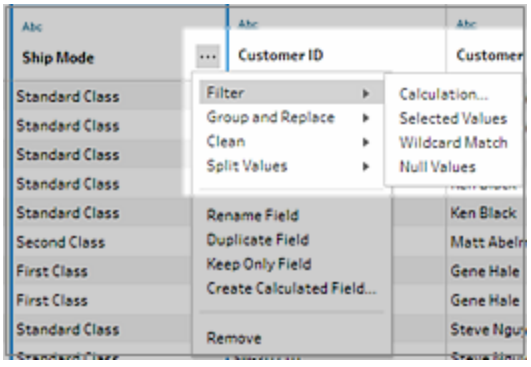
可用于每种数据类型的筛选器

数据类型	可用筛选器
字符串	计算、通配符匹配、Null 值、选定值
数字	计算、值范围、Null 值、选定值
日期及日期和时间	计算、值范围、相对日期、Null 值、选定值

筛选器选项位于何处？

若要查看可用于字段的不同筛选器选项，请在配置卡、数据网格或结果窗格中单击**“更多选项”**⋯菜单。若要查看数据网格上的菜单，您必须先单击**“隐藏配置窗格”**按钮，然后单击**“更多选项”**⋯。

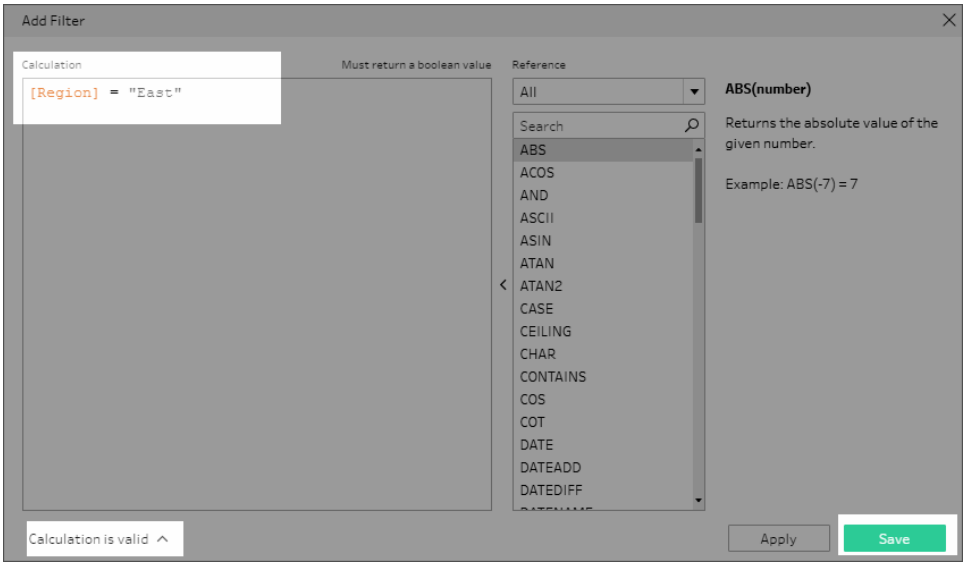




计算筛选器

当您选择**"计算"**时，**"添加筛选器"**对话框将打开。输入计算，验证其是否有效，并单击**"保存"**。从版本 2021.4.1 开始，您还可以在计算筛选器中包括参数。有关详细信息，请参见将用户参数应用于筛选器计算 在本页206。

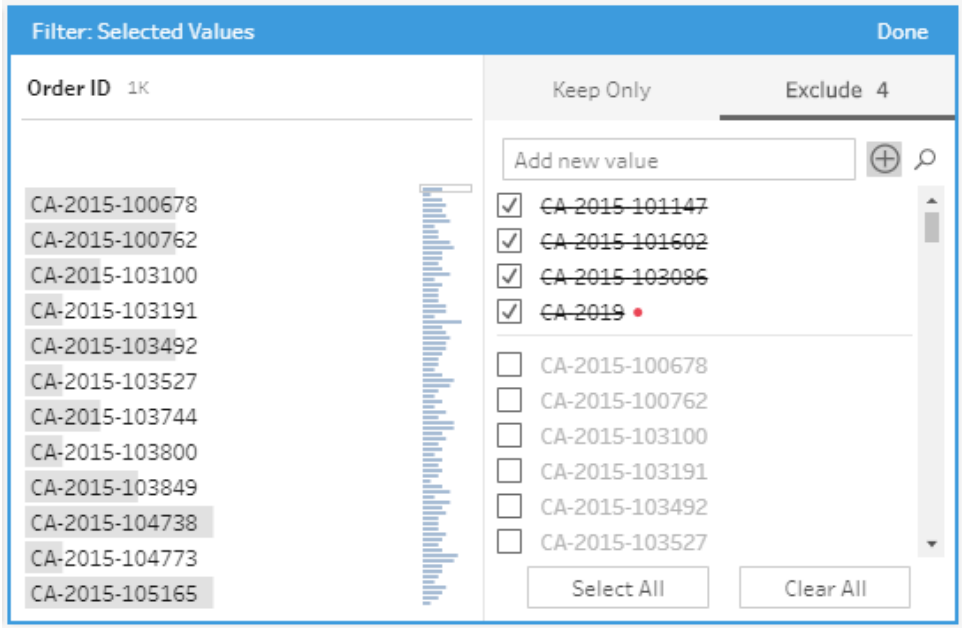
注意：在"输入"步骤中，这是唯一可用的筛选器类型。所有其他筛选器类型都可以在配置卡、数据网格或结果窗格中使用。



"选定值"筛选器

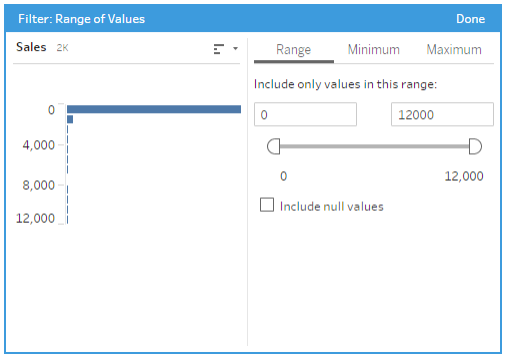
在 Tableau Prep Builder 版本 2019.2.3 及更高版本和 Web 上，您可以使用**"选定值"**筛选器选取和选择要为字段保留或排除的值，即使这些值不在样本中。在右侧窗格中，单击**"只保留"**或**"排除"**选项卡选择操作，然后输入搜索词以搜索值，或单击 **"添加值"** 以添加位于数据集中但未包括在样本中的值。单击**"完成"**应用筛选器。

注意：此筛选器选项对于**"聚合"**或**"转置"**步骤类型不可用。



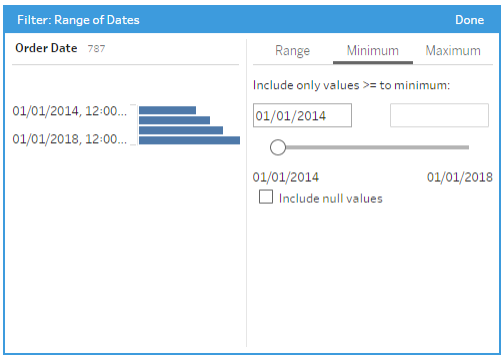
“值范围”筛选器

筛选出位于特定范围内的值。当您选择**“值范围”**时，您可以指定范围，或者设置最小值或最大值。



“日期范围”筛选器

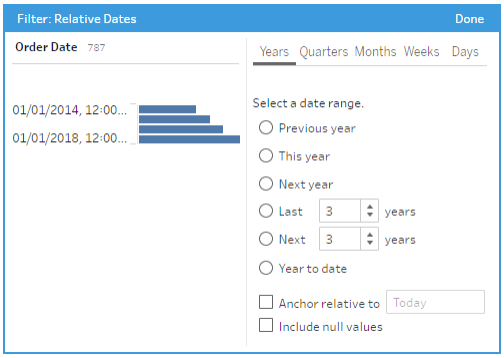
筛选出位于特定日期范围内的值。当您选择**“日期范围”**时，您可以指定日期范围，或者设置最早日期或最晚日期。



“相对日期”筛选器

使用**“相对日期”**筛选器来指定要在数据中查看的年、季度、月、周或天的确切范围。您也可以配置相对于特定日期锚点，并包括 null 值。

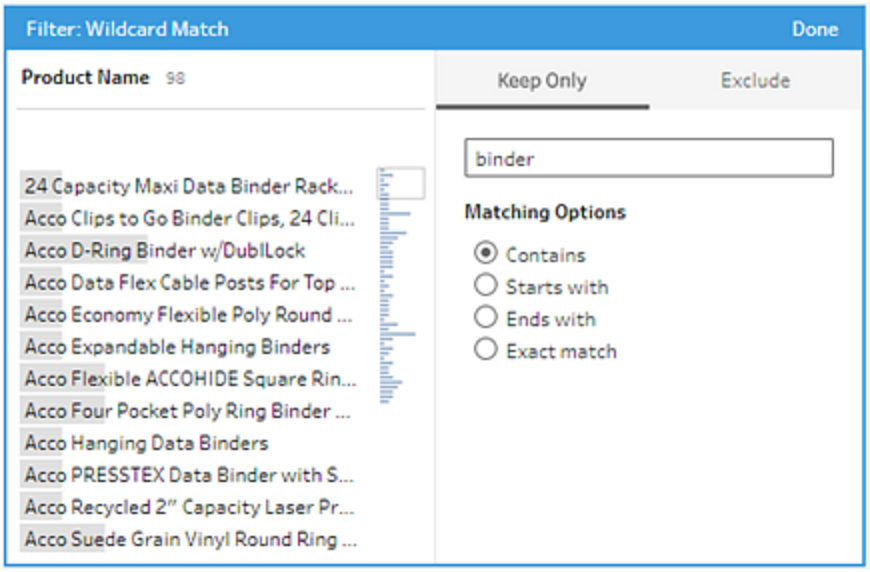
注意：“前”日期时间段包括完整的当前时间单位，即使某些日期尚未发生。举例来说，如果选择前一个月，且当前日期为 1 月 7 日，则 Tableau 将显示 1 月 1 日至 1 月 31 日的日期。



“通配符匹配”筛选器

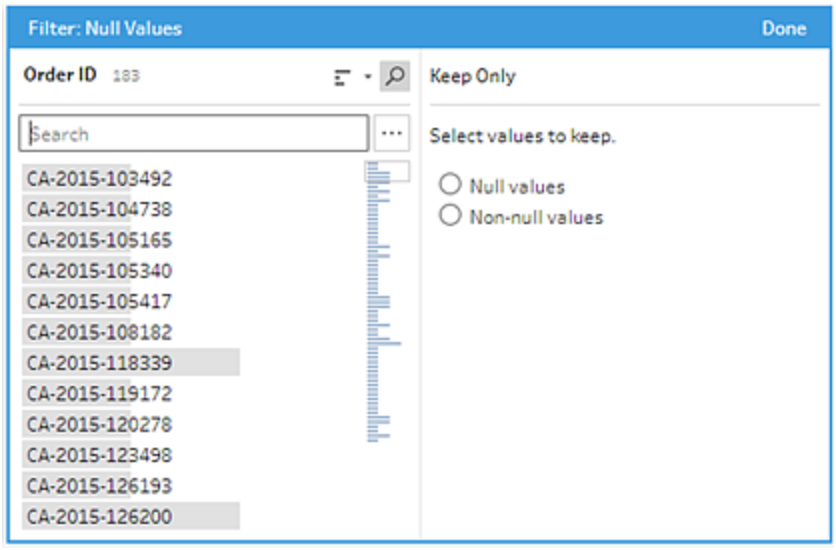
如果选择**“通配符匹配”**，您可以筛选字段值以保留或仅排除符合某个模式的值。在筛选编辑器中，选择**“只保留”**或**“排除”**选项卡，输入要匹配的值，然后设置**“匹配选项”**条件以返回所寻找的值。

筛选的结果显示在筛选编辑器的左侧窗格中，以便您能够查看和试验结果。有了所需的结果之后，单击**“完成”**应用更改。



“Null 值”筛选器

如果选择**“Null 值”**，您可以筛选所选字段中的值，以仅显示 null 值或排除所有 null 值。



移除重复行

在 Tableau Prep Builder 版本 2024.1 及更高版本以及 Tableau Cloud 中的 Web 上受支持。

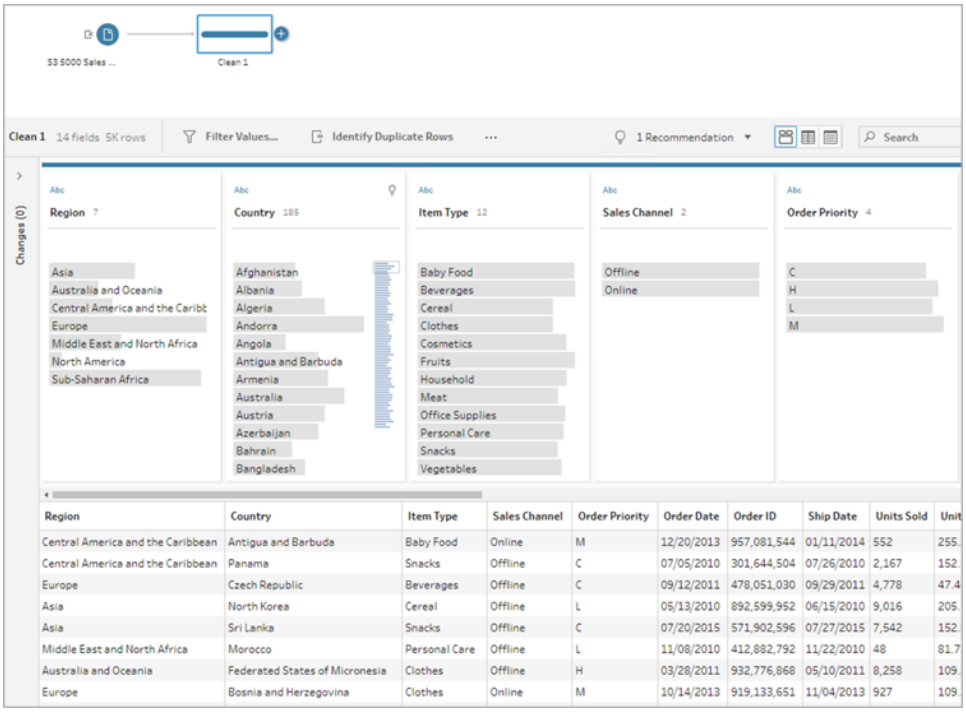
重复数据可能会影响数据质量、预测结果存在偏差并浪费存储空间。移除重复项可确保每个值都是唯一的，从而提供数据的准确表示形式。移除重复项的第一步是识别数据中的重复行。识别出重复项后，您随后可以将它们从工作流程中移除。

识别并移除重复项

Tableau Prep 将重复项标识为至少有一个其他行具有相同值的任何行。

识别并移除重复的行。

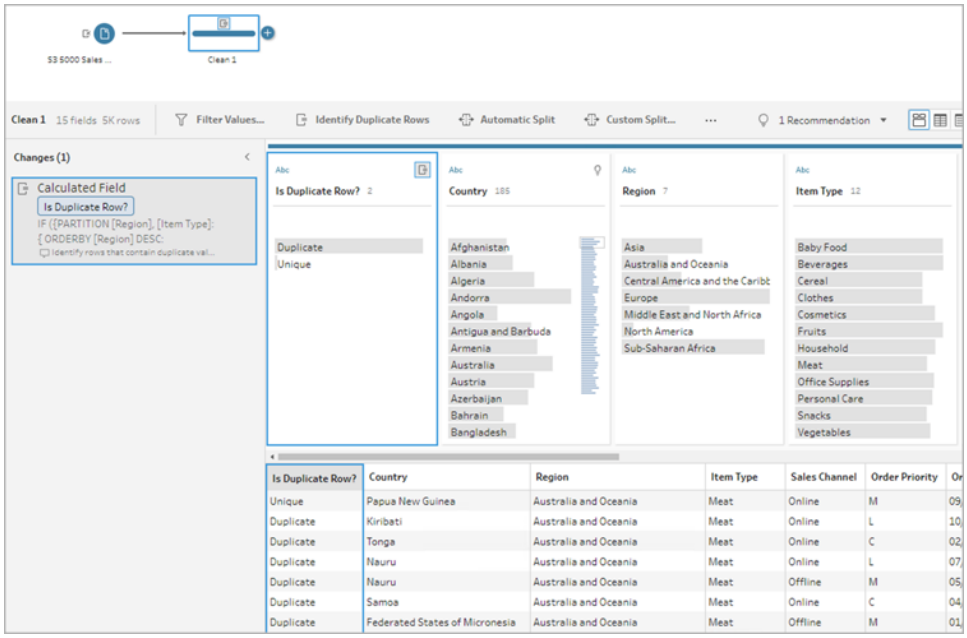
1. (可选)对于 Excel 和文本文件，您可以单击“输入”步骤并添加“源行号”字段以按原始数据源对行进行排序。有关详细信息，请参见将源行号字段添加到您的流程和**计算字段的排序顺序** 在本页 177。
2. 单击“输入”或“输出”步骤以外的任何步骤。



3. 识别重复的行。
   - 若要识别所有字段中的重复行，请从工具栏中单击**“识别重复行”**。
   - 若要识别特定字段中的重复行，请选择一个或多个字段，然后单击**“识别重复行”**。(可选)在配置窗格中，您可以单击从所选字段中选择**“更多选项”** ··· 菜单，并选择**“识别重复行”**。

将创建一个计算字段并在“更改”窗格中列出。计算字段提供重复行跨越的字段的描述和名称。数据网格显示哪些行是唯一的，哪些行是重复的。

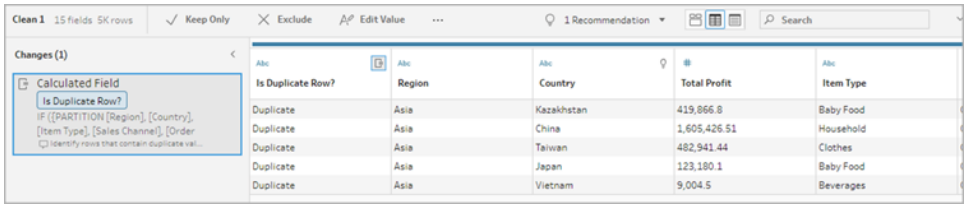
下图显示了选择“*区域*”和“*项目类型*”两个字段以仅识别那些选定字段中的重复行的结果。



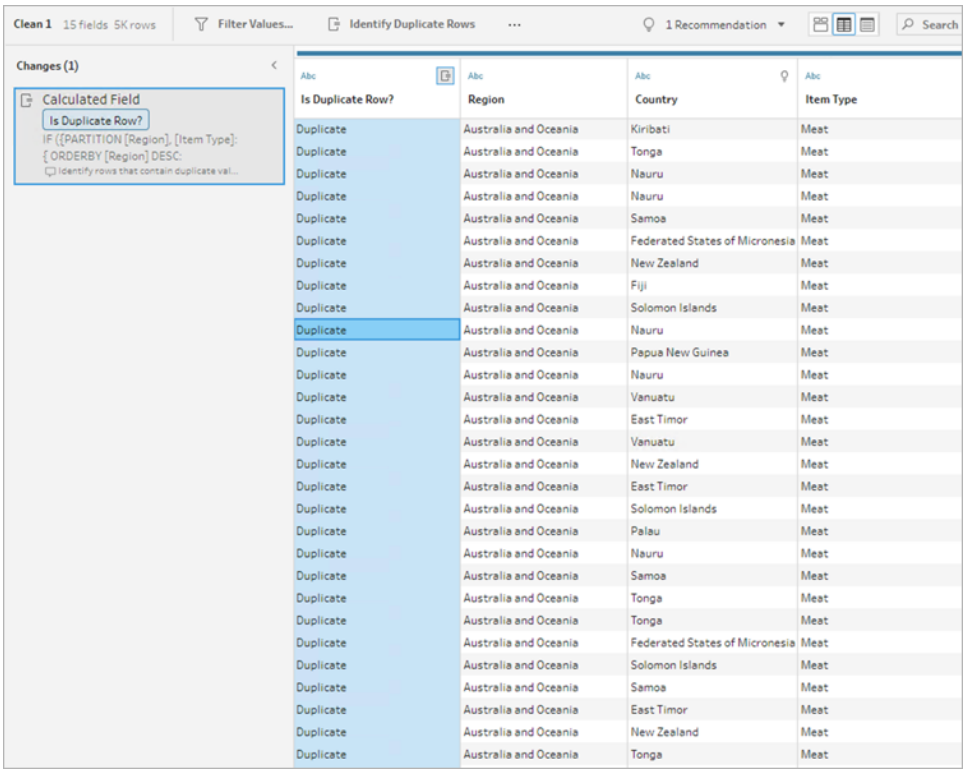
如果源行号可用，则它用于对行的顺序进行排序。否则，行将根据表架构中的第一个字段进行排序。若要更改计算字段中的值，您可以直接编辑计算或编写自己的计算。请参见**计算字段的排序顺序** 在本页 177。

4. 若要评估重复行，请单击“是否为重复行？”字段或数据网格中的**“重复”**或**“唯一”**。

下图显示了所有字段中的重复行。



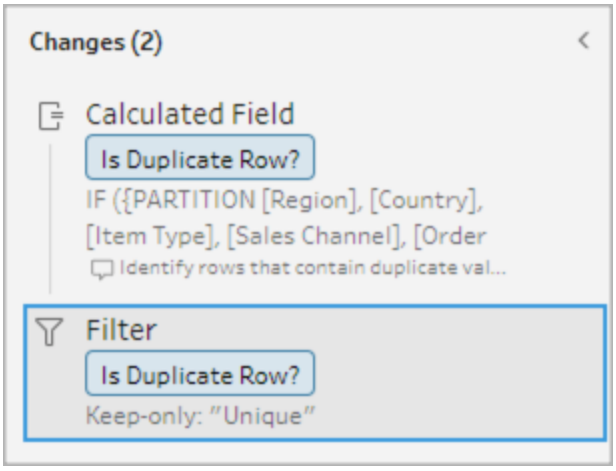
下图仅显示了“*区域*”和“*项目类型*”字段中的重复行。



5. 若要筛选并仅保留唯一行，请单击“数据”网格或“是否为重复行？”字段中的**“唯一”**，然后从工具栏中选择**“仅保留”**。或者，您可以单击“数据”网格或“是否为重复行？”字段中的**“唯一”**，然后右键单击以选择**“仅保留”**。

   您还可以选择**“重复”**，然后选择**“排除”**以筛选重复项数据。这会产生与选择“唯一”和“仅保留”相同的结果，并且不会影响排除或保留哪些行。

   将会在更改窗格中创建一个筛选器，显示仅保留唯一行并排除重复行。



6. 从“是否为重复行？”计算字段中，单击**“更多选项”**⋯ 菜单，然后选择**“移除”**。

   当您运行流程并创建输出时，重复的行将被移除。

   > **注意**：不要移除“是否为重复行？”计算字段（如果您要调整采样数据）。

样本数据中的重复项

重复的行可能不会出现在采样数据中，并且可能会在模型中引入偏差。

若要解决重复行采样数据，请执行以下操作：

1. 不要移除**“是否为重复行？”**计算字段（如果您要调整采样数据）。这使您可以选择更改样本大小或调整输入步骤中的样本数据。
2. 调整样本大小以加载尽可能多的数据用于行选择。请参见设置数据样本大小。
3. 根据样本的大小，重复项可能不会出现在采样数据中。对唯一行使用**“仅保留”**选项允许您移除重复行，即使它们在采样数据中不可见。

计算字段的排序顺序

默认情况下，如果多行包含相同的值，它们将按第一个字段或源行号（如果可用）排序。这些字段用于将第一行标识为*“唯一”*，将其余行标识为*“重复”*。更改字段的排序顺序允许您指定要将哪个重复行标识为“唯一”。您可以通过编辑计算并更改“ORDERBY ”字段来更改顺序。

例如：

```
IF ({PARTITION [Field1], [Field2], [Field3]: { ORDERBY [FieldName]
ASC: ROW_NUMBER() } } = 1) THEN 'Unique' ELSE 'Duplicate' END
```

对于 Excel 和文本文件，您可以通过添加源行号字段按原始数据源排序。

例如：

```
IF ({PARTITION [Source Row Number], [Field1], [Field2], [Field3]: {
ORDERBY [Source Row Number] DESC: ROW_NUMBER() } } = 1) THEN 'Unique'
ELSE 'Duplicate' END
```

有关详细信息，请参见将源行号字段添加到您的流程。

编写您自己的筛选器计算来查找并移除重复项

您可以跨多个字段编写自己的筛选器计算，以查找并移除重复项。

以下筛选器计算将仅返回返回 True 的结果。根据用于 PARTITION 的字段，计算返回 True 或 False

```
{PARTITION [Field1], [Field2], [Field3]: { ORDERBY [Field1] DESC:
ROW_NUMBER() } } = 1
```

使用 CASE 或者 IF 用于识别重复计算

您可以在计算编辑器中使用 IF 或 CASE 函数。例如：

```
CASE {PARTITION [Field1], [Field2], [Field3] : { ORDERBY [[Field3]]:
ROW_NUMBER() } } = 1 WHEN TRUE THEN 'UNIQUE' ELSE 'DUPLICATE' END
```

```
IF ({PARTITION [[Field1]], [[Field2]], [[Field3]]: { ORDERBY
[[Field3]] DESC: ROW_NUMBER() } } = 1) THEN 'Unique' ELSE 'Duplicate'
END
```

使用数据角色验证数据

注意：数据源所有者和 Tableau 管理员可以为“数据问答”(Ask Data) 功能的特定数据字段名称和值添加同义词。有关为“数据问答”(Ask Data) 功能使用数据角色的信息，请参见 Tableau Desktop 帮助中的为“数据问答”(Ask Data) 功能添加同义词。

注意：Tableau 的‘数据问答’(Ask Data) 功能于 2024 年 2 月在 Tableau Cloud 和 Tableau Server 版本 2024.2 中已停用。

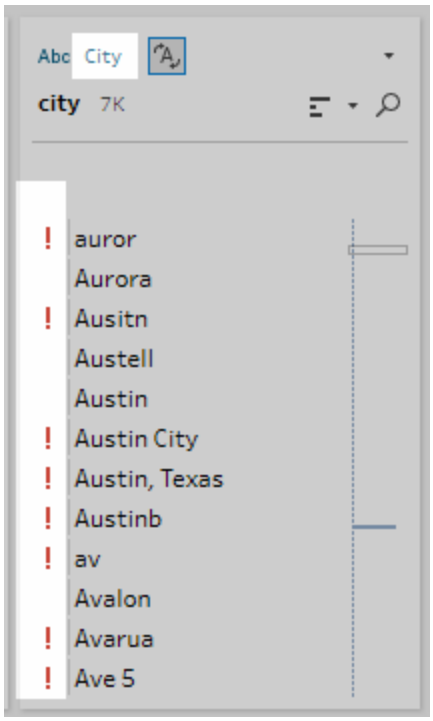
使用数据角色来快速确定字段中的值是否有效。Tableau Prep 提供了一组您可从中选择的标准数据角色，或者，您可以使用数据集中的唯一字段值来创建自己的数据角色。

分配数据角色时，Tableau Prep 会将为数据角色定义的标准值与您的字段中的值进行比较。任何不匹配的值都用红色感叹号标记。您可以筛选字段以仅查看有效或无效的值，并采取相应的措施来修复这些值。将数据角色分配给字段后，您可以使用**“对值进行分组”**选项来进行分组，并基于拼写和发音将无效值与有效值匹配。

注意：从版本 2020.4.1 开始，您现在可以在 Tableau Server 和 Tableau Cloud 中创建和编辑流程。除非特别指出，本主题中的内容适用于所有平台。有关在 Web 上制作流程的详细信息，请参见 Tableau Server 帮助中的 Web 上的 Tableau Prep。

为数据分配标准数据角色

采用与分配数据类型相同的方式将 Tableau Prep 提供的数据角色分配给字段。数据角色确定数据值所代表的意义，使 Tableau Prep 能够自动验证值并突出显示对于该角色无效的值。



举例来说，如果您有地理数据字段值，您可以分配数据角色**“城市”**，Tableau Prep 即会将这些字段中的值与一组已知的域值或模式进行比较来确定不匹配的值。

注意：系统会对每个字段进行单独分析，因此国家/地区“美国”中的州“华盛顿”中的城市值“波特兰”可能不是有效的城市和州组合，但并不会将其标识为无效，因为它是有效的城市名称。

Tableau Prep Builder 提供以下数据角色：

- 电子邮件
- URL
- 地理角色（基于当前地理数据，并且与 Tableau Desktop 使用的数据相同）
  - 机场
  - 地区代码 (美国)

- CBSA/MSA
- 城市
- 国会选区 (美国)
- 国家/地区
- 郡县
- NUTS 欧洲
- 州/省/市/自治区
- 邮政编码

提示：在 Tableau Prep Builder 版本 2019.1.4 及更高版本中以及在 Web 上，如果为字段分配地理角色，则还可以使用该数据角色，通过数据角色定义的标准值来匹配值和对值进行分组。有关使用数据角色对值进行分组的详细信息，请参见**清理和调整数据** 在本页 213。

若要为字段分配数据角色，请执行以下操作：

1. 在“配置”窗格、“结果”窗格或数据网格中，单击字段的数据类型。
2. 为字段选择数据角色。

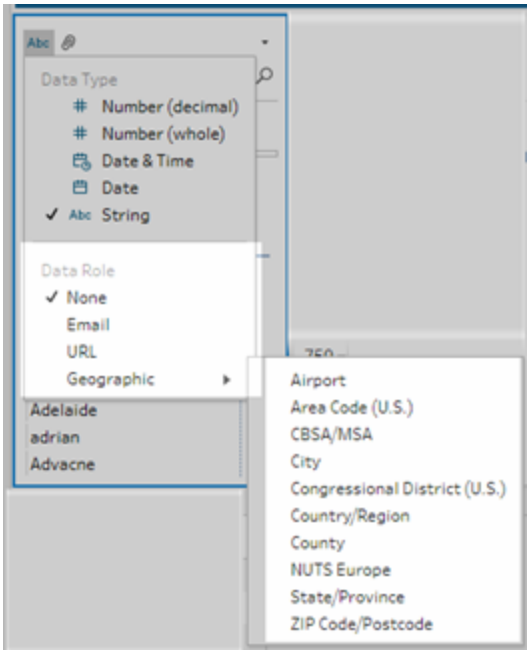
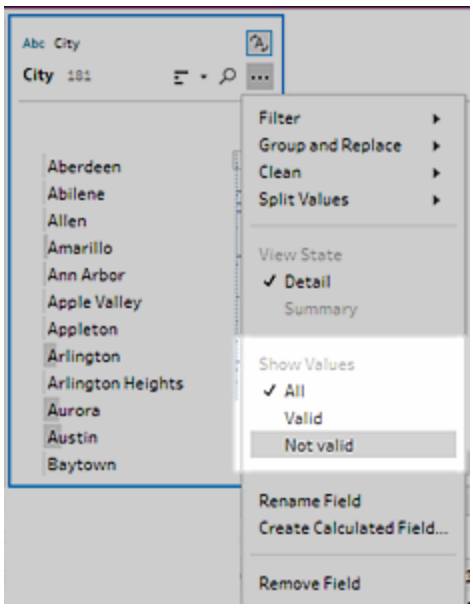
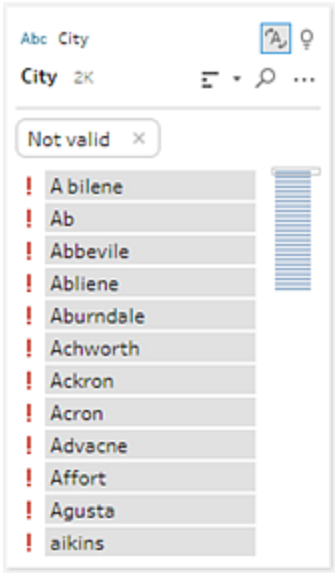


Tableau Prep 会将字段的数据值与所选数据角色的已知域值或模式(对于电子邮件或 URL)进行比较，并用红色感叹号标记任何不匹配的值。

3. 单击字段的下拉箭头，并从**“显示值”**部分选择一个选项，以显示所有值或仅显示对于数据角色有效或无效的值。





4. 使用字段的**"更多选项"**··· 菜单中的清理选项来更正无效的任何值。有关如何清理字段值的详细信息，请参见**关于清理操作** 在本页213。

创建自定义数据角色

重要信息：自定义数据角色将于 2024 年 2 月在 Tableau Cloud 和 Tableau Server 版本 2024.2 中停用。

从 Tableau Prep Builder 版本 2019.3.1 开始以及在 Web 上，您可以使用数据集中的字段值创建自己的自定义数据角色，来为您或他人可用于在清理数据时验证字段值的一组标准值。选择要使用的字段，将任何清理操作应用于该字段（如果需要），然后将该字段发布到 Tableau Server 或 Tableau Cloud 以在流程中使用它或与他人共享您的数据角色。

如果在 Web 上编辑流程时创建自定义数据角色，可以将自定义数据角色直接发布到已登录的服务器。

要求

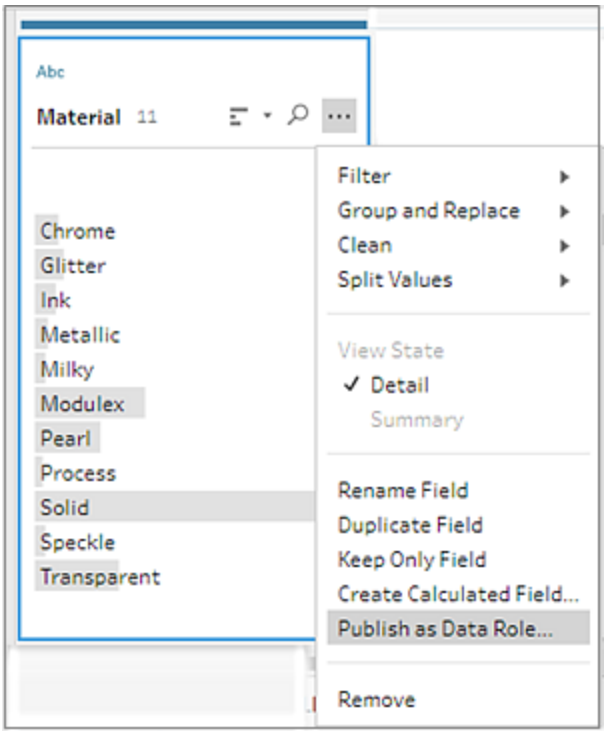
- 您可以依据数据集中的单个字段创建自定义数据角色。不支持依据字段组合创建自定义数据角色。
- 不支持将数据角色发布到具有锁定权限的项目。
- 只能为分配有**"字符串"**和**"数字(整数)"**数据类型的字段创建自定义数据角色。

- 当您创建自定义数据角色时，Tableau Prep 会在您的流程中创建一个专用于发布数据角色的输出步骤。
- 不支持将自定义数据角色发布到同一流程中的多个站点。如果发布流程，您必须将自定义数据角色发布到在其中发布流程的同一站点或服务器。
- 自定义数据角色特定于您在其中发布这些角色的站点、服务器和项目。具有该位置权限的所有用户都可以使用自定义数据角色，但必须登录到站点或服务器才能选择或应用该角色。将针对新项目为自定义数据角色分配**“所有用户”**组的默认权限，而不是**“无”**。
- 自定义数据角色不是特定于版本的。应用自定义数据角色时，将应用最新版本。
- 发布到 Tableau Server 或 Tableau Cloud 后，具有站点、服务器和项目访问权限的用户可以查看该位置中的所有数据角色。
    - 具有适当权限的用户可以移动、删除或编辑数据角色的权限。
    - 您可对自定义数据角色设置的权限以及执行的操作与可对流程设置的权限和执行的操作类似。有关详细信息，请参见 Tableau Server 帮助中的管理流程和权限能力。
- 若要编辑数据角色，必须在 Tableau Prep Builder 中或 Web 上的流程中进行更改，然后使用相同的名称重新发布数据角色以覆盖它。此过程类似于编辑已发布的数据源。

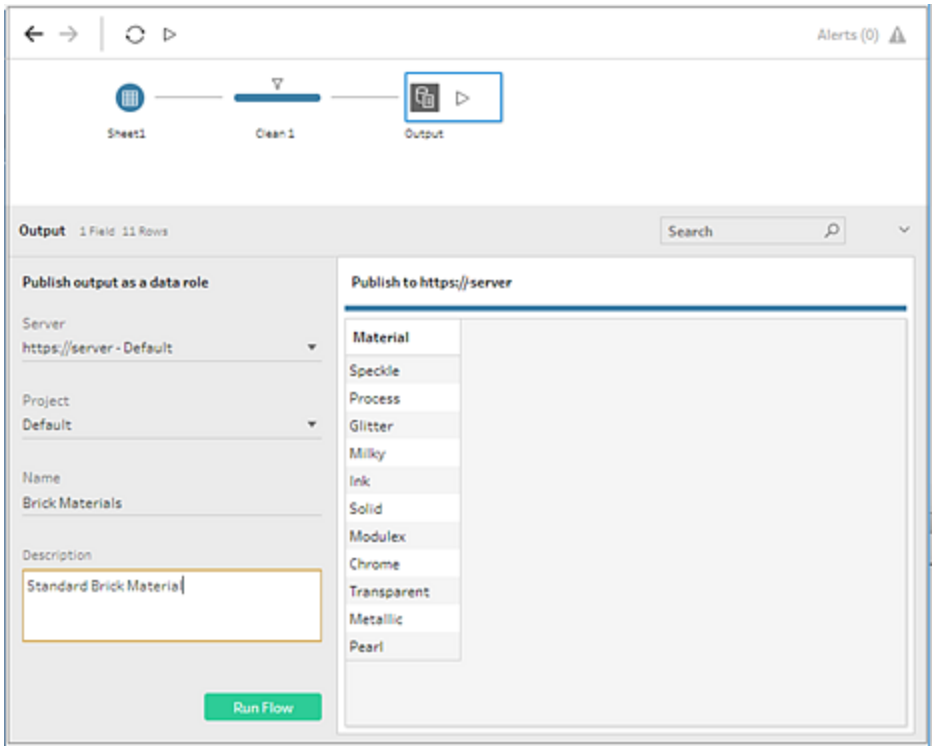
创建自定义数据角色

1. 在**“配置”**窗格、数据网格或“结果”窗格中，选择要用于创建自定义数据角色的字段。

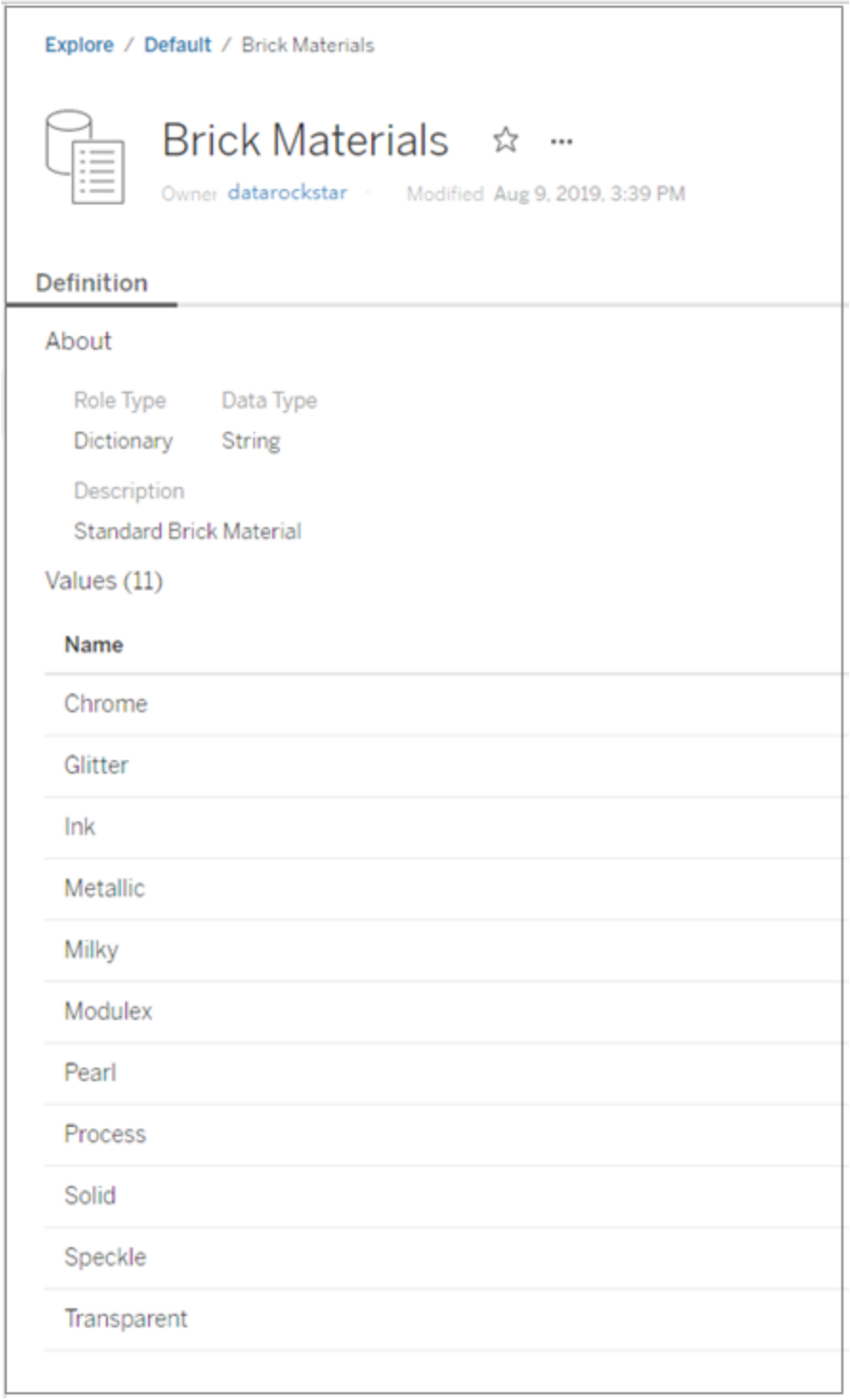
2. 为该字段单击**“更多选项”**…，并选择**“以数据角色形式发布”**。



3. 选择要在其中发布数据角色的服务器和项目。



4. 单击**“运行流程”**以创建数据角色。发布过程成功完成后，您可以在 Tableau Server 或 Tableau Cloud 中查看数据角色。根据 Tableau Server 或 Tableau Cloud 站点上的负载，处理数据角色可能需要一段时间。如果您的数据角色无法立即使用，请等待几分钟，然后再次尝试选择。


Explore / Default / Brick Materials
Brick Materials
Owner datarockstar
Modified Aug 9, 2019, 3:39 PM
Definition
About
Role Type
Dictionary
Data Type
String
Description
Standard Brick Material
Values (11)
Name
Chrome
Glitter
Ink
Metallic
Milky
Modulex
Pearl
Process
Solid
Speckle
Transparent

应用自定义数据角色

1. 在“配置”窗格、“结果”窗格或数据网格中，单击要在其中应用自定义数据角色的字段的数据类型。
2. 选择**“自定义”**，然后选择要应用于字段的数据角色。

   重要信息：在 Tableau Prep Builder 中，请确保登录到在其中发布了数据角色的站点或服务器，否则您将看不到此选项。

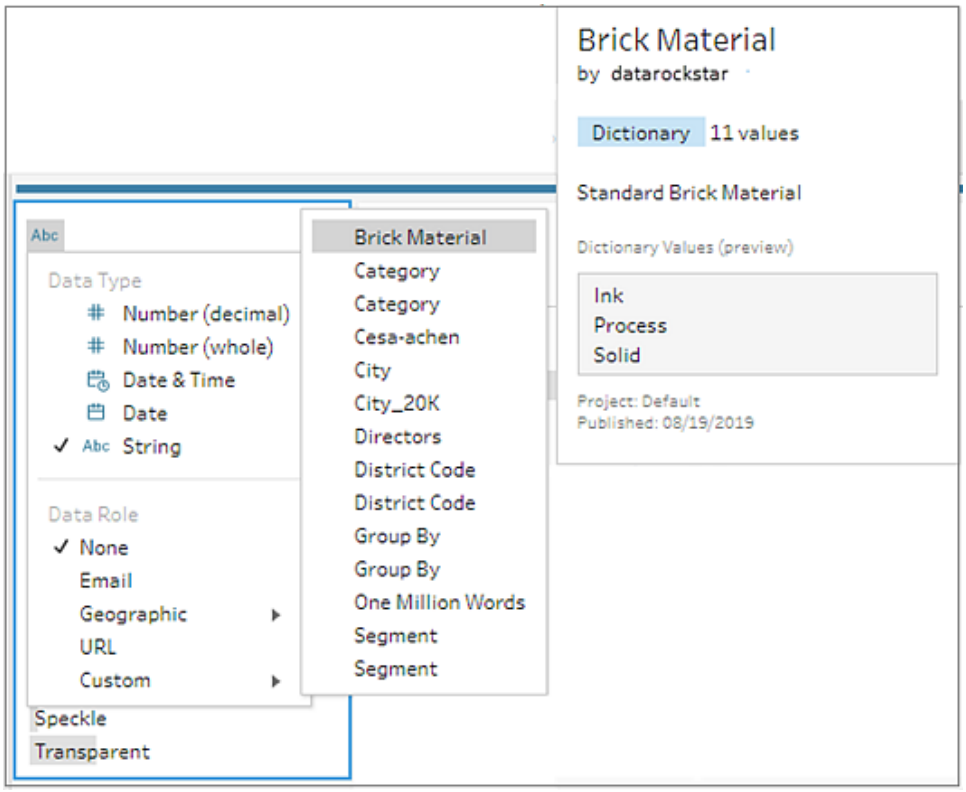
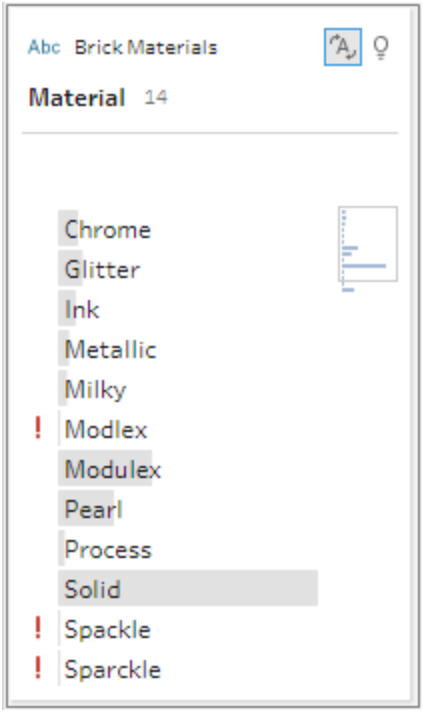
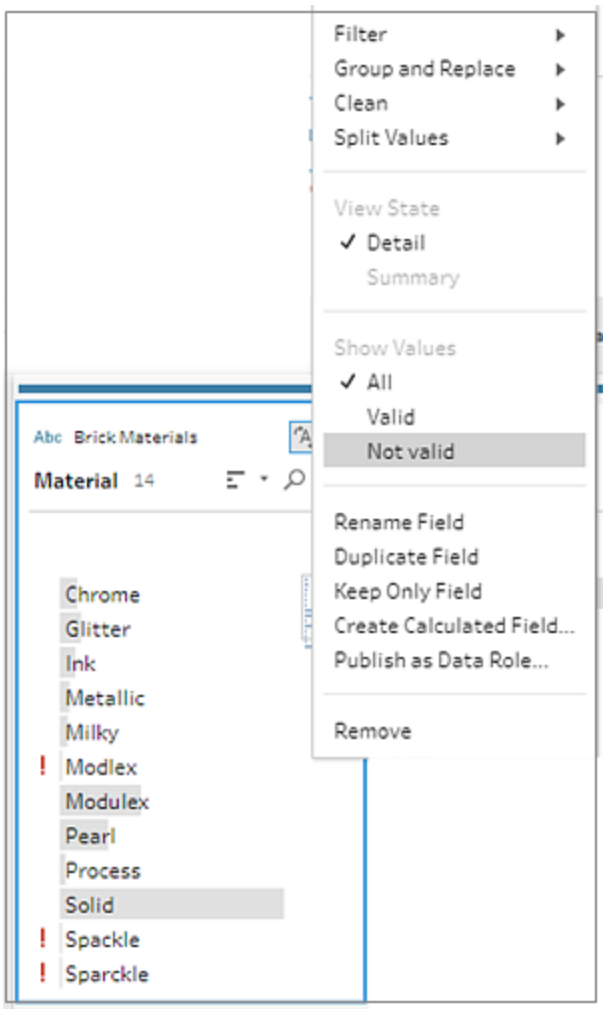


Tableau Prep 会将字段的数据值与所选数据角色的已知域值进行比较，并用红色感叹号标记任何不匹配的值。



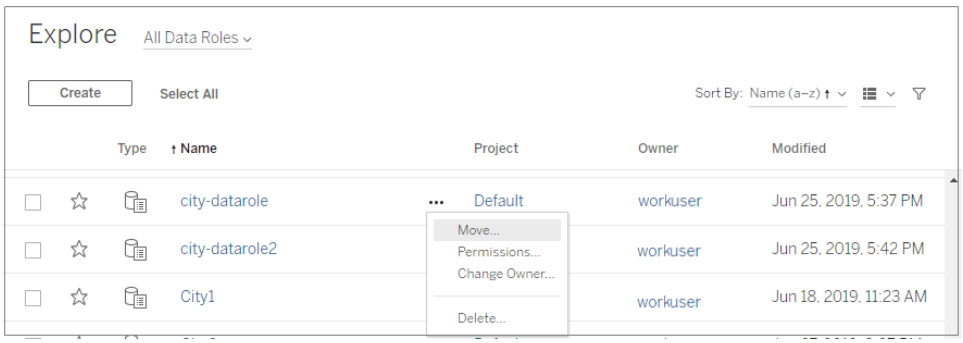
3. 单击字段的下拉箭头，并从**“显示值”**部分选择一个选项，以显示所有值或仅显示对于数据角色有效或无效的值。



4. 使用字段的**"更多选项"**··· 菜单中的清理选项来更正无效的任何值。有关如何清理字段值的详细信息，请参见关于清理操作 在本页213。

查看和管理自定义数据角色

您可以在 Tableau Server 和 Tableau Cloud 上查看及管理发布的自定义数据角色。您可以查看发布到您的站点或服务器的所有自定义数据角色。针对所选数据角色单击**"更多操作"**···，以将其移到其他项目、更改权限或将其删除。



按数据角色对类似值进行分组

注意：在 Tableau Prep Builder 版本 2019.1.4 和 2019.2.1 中，此选项标为**“数据角色匹配”**。

如果为字段分配地理数据角色，您可以使用数据角色中的值，基于拼写和发音对数据字段中的值进行分组和匹配，从而使值标准化。您可以使用**“拼写”**或**“拼写 + 发音”**对无效值进行分组，并将其与有效值匹配。

这些选项使用数据角色定义的标准值。如果数据集示例中没有标准值，Tableau Prep 会自动添加该值，并将值标记为不在原始数据集中。有关为字段分配数据角色的详细信息，请参见**为数据分配标准数据角色** 在本页 179。

若要使用数据角色对值进行分组，请完成以下步骤。

1. 在“配置”窗格、“结果”窗格或数据网格中，单击字段的数据类型。
2. 为字段选择以下数据角色之一：
   - 机场
   - 城市
   - 国家/地区
   - 郡县
   - 州/省/市/自治区

   从 Tableau Prep Builder 版本 2019.3.2 开始以及在 Web 上，您还可以从自定义数据角色中进行选择

标准数据角色(版本 2019.1.4 及更高版本)	自定义数据角色(版本 2019.3.2 及更高版本)

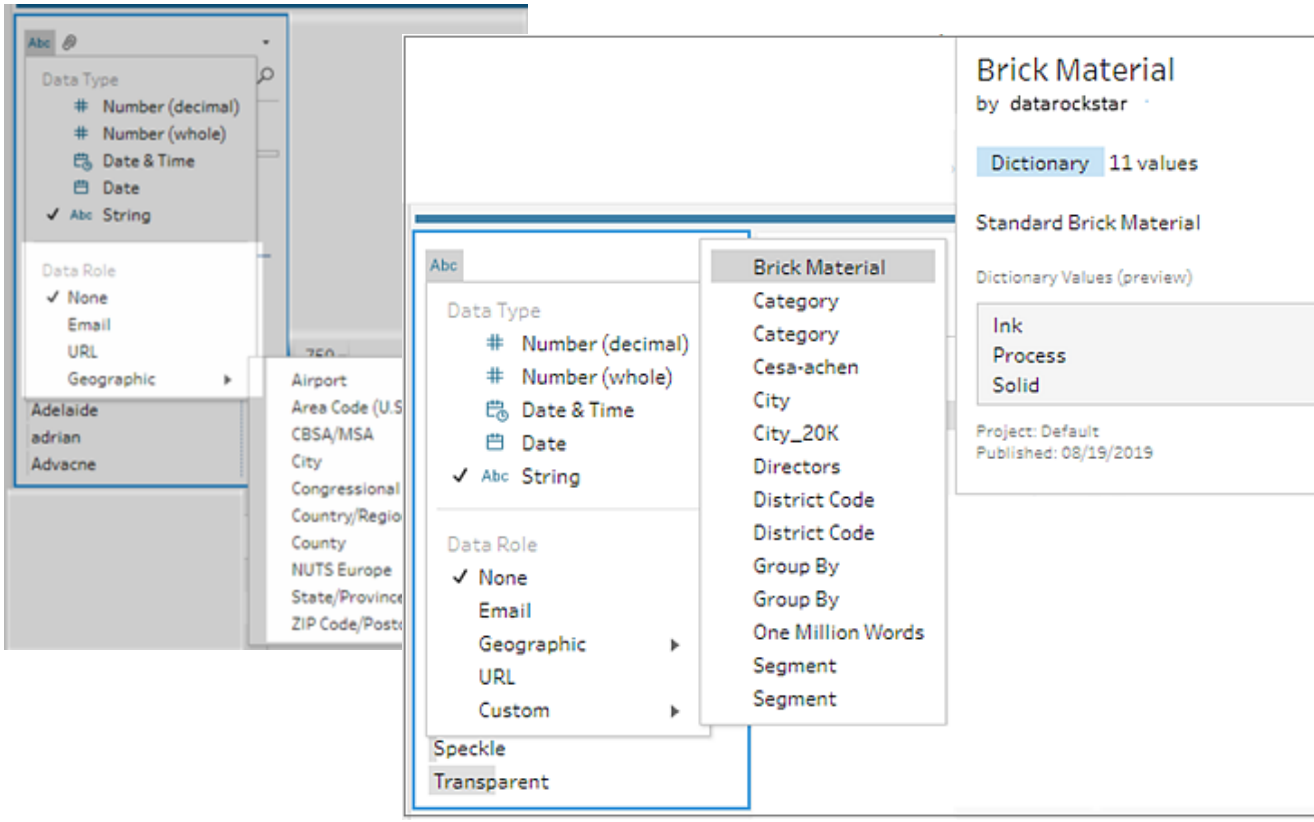
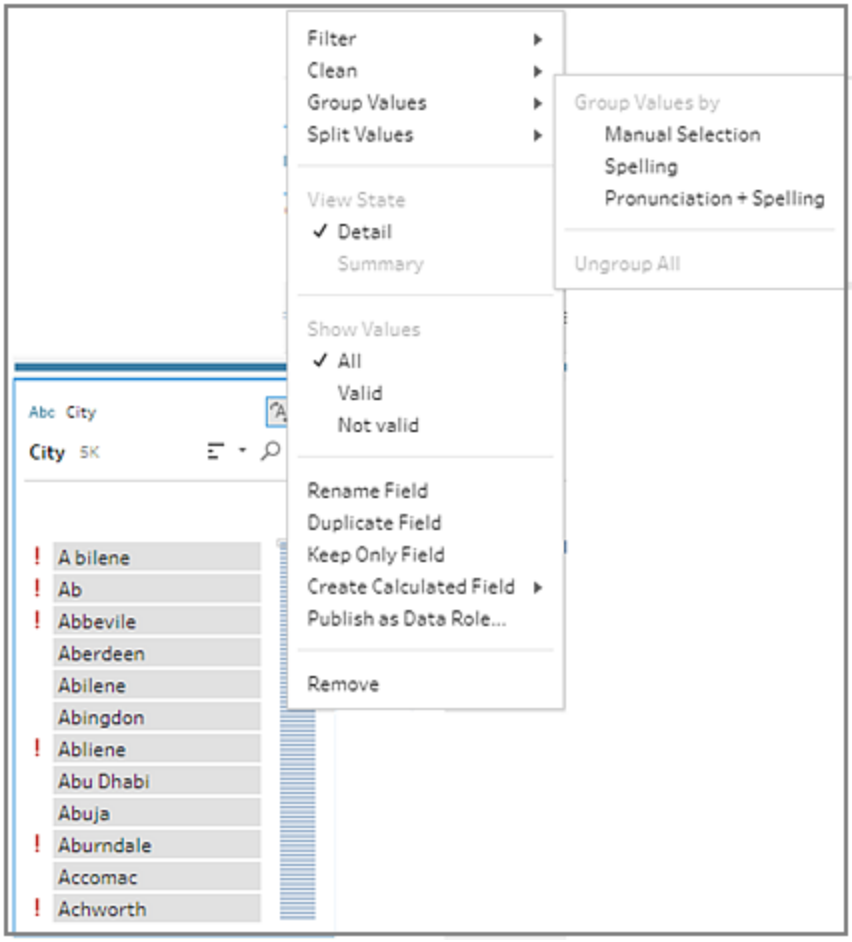


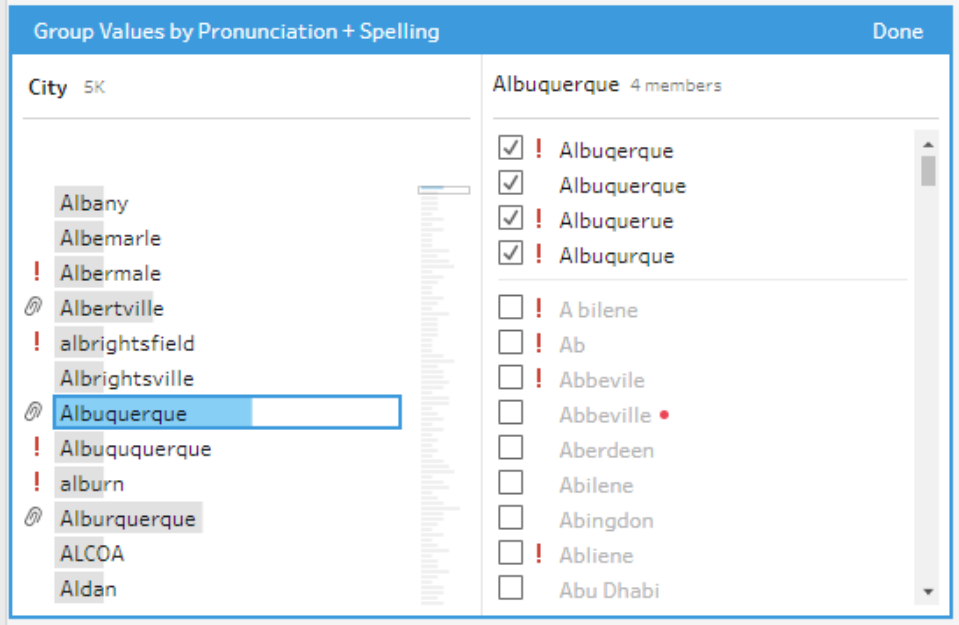
Tableau Prep 会将字段的数据值与所选数据角色的已知域值进行比较，并用红色感叹号标记任何不匹配的值。

3. 单击**“更多选项”**…，选择**“对值进行分组”**(在以前的版本中为**“分组和替换”**)，然后选择以下选项之一：
   - **拼写**：将无效值与因为添加、移除或替换了字符而不同的最接近有效值匹配。
   - **发音 + 拼写**：根据拼写和发音将无效值与最相似的有效值匹配。



您还可以单击字段上的**“建议”** 图标将建议应用于组，并将无效值替换为有效的值。此选项使用**“发音 + 拼写”**值分组选项。

Tableau Prep 按拼写或者拼写和发音比较各个值，然后依据数据角色的标准化值对类似的值进行分组。如果数据集中没有标准化值，则系统会添加该值，并使用红点加以标记。



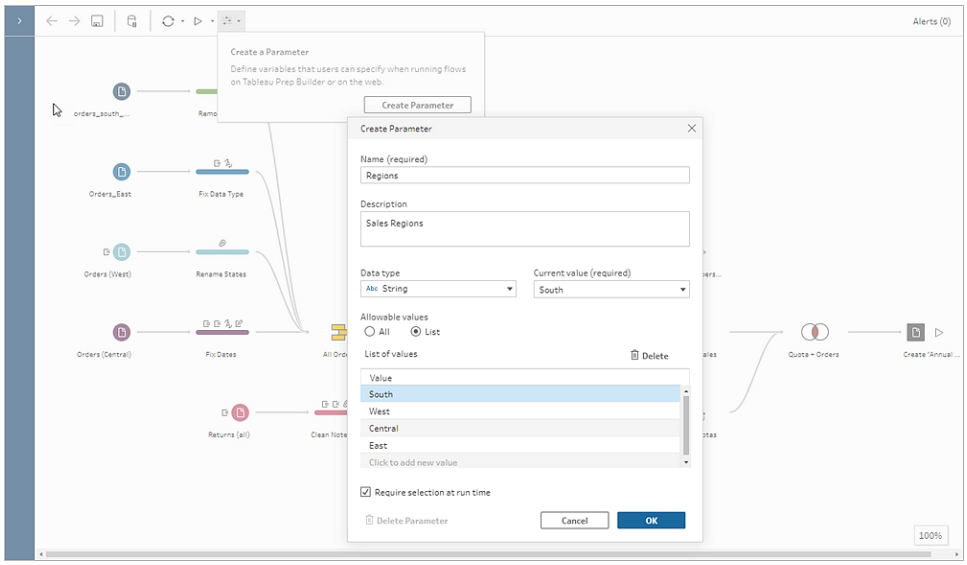
在流程中创建和使用参数

在 Tableau Prep Builder 版本 2021.4.1 及更高版本以及 Web 版 Tableau Cloud 和 Tableau Server 版本 2021.4.0 及更高版本中受支持

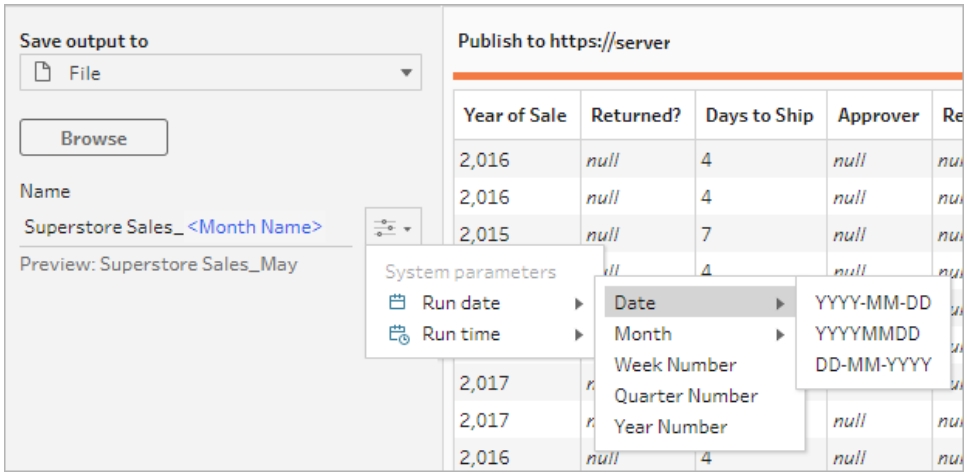
注意：除非特别说明，否则本主题中的内容适用于 Tableau Prep Builder 中和 Web 上的制作流程。有关在 Web 上制作流程的详细信息，请参见 Tableau Server 和 Tableau Cloud 帮助中的“Web 上的 Tableau Prep”.

如果您经常使用具有相同架构的不同数据重复使用流程，则可以创建用户参数并将其应用于流程，以轻松地在场景之间转换。参数是全局占位符值，例如可替换流程中的常量值的数字、文本值或布尔值。

您现在可以构建一个流程并使用参数来运行具有不同数据集的流程，而不是构建和维护多个流程。例如，您可以为不同的销售区域创建一个参数，然后将参数值应用于输入文件路径以仅使用该区域的数据运行流程。



从 Tableau Prep Builder 和 Tableau Cloud 版本 2023.2 开始，您还可以将系统参数添加到文件或已发布的数据源输出名称，以便在每次运行流程时自动添加时间戳。



我可以在何处应用参数？

您可以将用户参数应用于文件名、路径、表名、筛选器表达式和计算字段，具体取决于步骤类型。从版本 2022.1.1 开始，您甚至可以在使用 REST API 运行流程时包含参数覆盖值。有关详细信息，请参见 Tableau REST API 帮助中的流程方法。

您可以将系统参数(版本 2023.2 及更高版本)应用于文件和已发布数据源输出类型的输出名称。

下表列出了您可以为每个步骤类型应用参数的位置。

步骤类型	参数位置
输入	用户参数： • 连接到文件：在文件名或文件路径中使用参数 • 连接到数据库：在表名和自定义 SQL 中使用参数 • 表达式编辑器：筛选器
输出	用户或系统参数： • 输出到文件：将用户参数应用于文件名或文件路径，并从版本 2022.1.1 开始，应用于在 Microsoft Excel 工作表名称。将系统参数应用于文件名。 • 输出到服务器：将用户或系统参数应用于已发布数据源名称 • 输出到数据库：将用户参数应用于表名称，并从版本 2022.1.1 开始，应用于将流程输出写入数据库之前或之后运行的 SQL 脚本。
清理、新建行、转置、联接、合并	用户参数： • 表达式编辑器：筛选器和计算字段值
聚合	用户参数： • 表达式编辑器：筛选器
脚本	用户参数： • 表达式编辑器：筛选器和计算字段值
预测	用户参数： • 表达式编辑器：筛选器和计算字段值

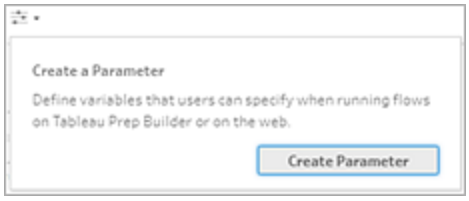
创建用户参数

用户参数特定于使用参数的流程。通过顶部菜单创建参数，然后定义适用于它们的值。您还可以定义接受所有值的参数，这意味着任何流程用户在运行流程时可以输入任何值。

您可以将流程参数值设为必需或可选。运行流程时，系统会提示用户输入参数值。必须先输入所需的参数值，然后用户才能运行流程。可以输入可选参数值，也可以接受当前(默认)值。然后将参数值应用于使用该参数的任何地方的流程运行。

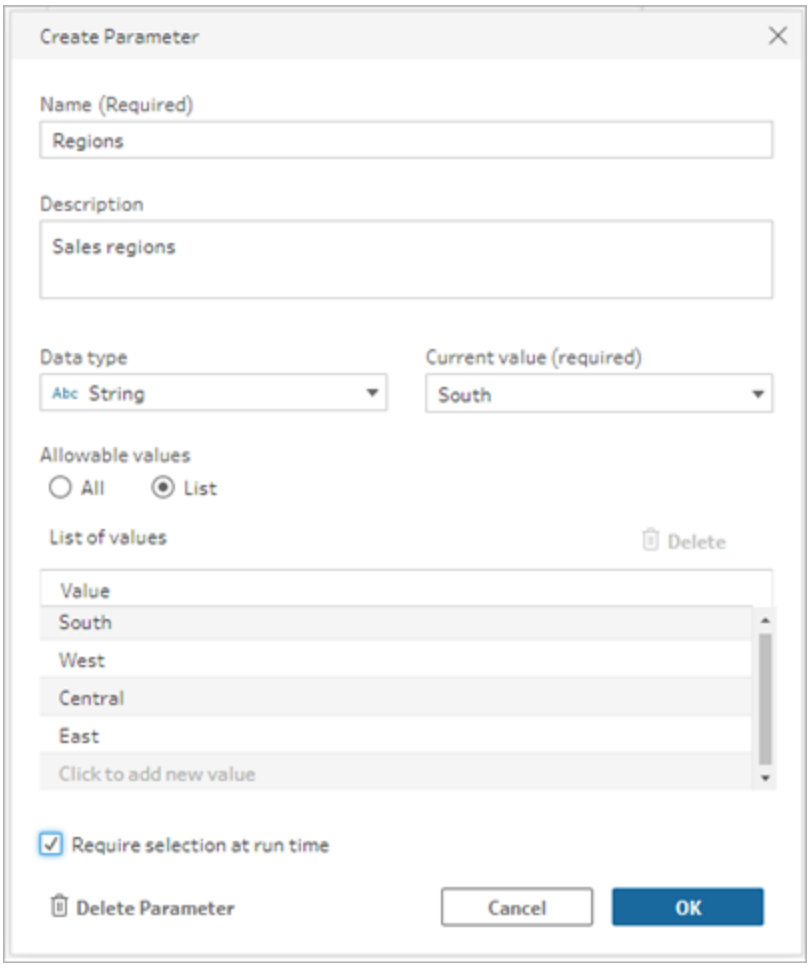
注意：若要在 Tableau Server 或 Tableau Cloud 上运行或计划包含参数的流程，您的管理员必须在服务器上启用**"流程参数"**设置。有关详细信息，请参见 Tableau Server 或 Tableau Cloud 帮助中的**"在 Web 上创建流程并与其进行交互"**。

1. 从顶部菜单中，单击**"参数"** 图标，然后单击**"创建参数"**。



2. 在**"创建参数"**对话框中，输入名称和描述(可选)。参数名称必须唯一。这是添加参数时显示在用户界面中的值。

   如果包括描述，用户可以在将鼠标悬停在参数列表中或参数使用位置中时看到此信息(从版本 2022.1.1 开始)。



3. 选择以下数据类型之一。参数值必须与您选择的数据类型相匹配。
   - 数字(整数或小数)
   - 字符串
   - 布尔值
4. 指定**"允许的值"**。这些是用户可以在参数中输入的值。
   - **全部**:此选项允许用户为参数键入任何值,即使在运行流程时也是如此。

     注意:将此选项用于可在输入和输出步骤中使用的参数可能存在安全风险。例如,允许输入任何值的自定义 SQL 查询可能会使您的数据资产遭到 SQL 注入攻击。
   - **列表**:输入用户可在应用参数时从中进行选择的值列表。若要输入多个值,请在每次输入后**按 Enter**。

5. (可选)选择**“运行时需要选择”**(在以前的版本中为**“运行时提示输入值”**)。这会将参数设为必需。用户在运行或计划流程时需要输入一个值。
6. 输入**“当前值”**。这是一个必需的值,用作参数的默认值。
   - **全部**:输入一个值。
   - **列表**:Tableau 使用列表中的第一个值。使用下拉选项进行更改。
   - **布尔值**:选择**“True”**或**“False”**。
7. 单击**“确定”**保存参数。

更改用户参数默认值

创建用户参数时,必须指定当前(默认)值。如果流程中包含参数,则此值用于:

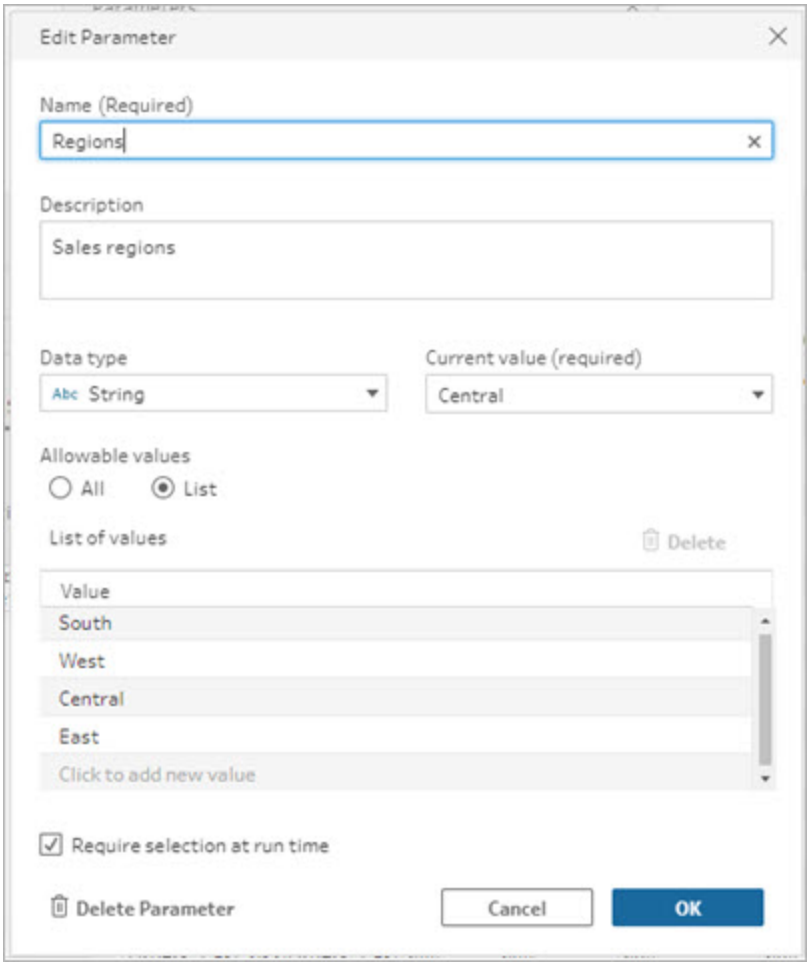
- 运行在输入步骤中定义的自定义 SQL 查询。
- 填写运行时未指定的可选参数。
- 将参数替换为已保存步骤中的静态值(版本 2022.1.1 及更高版本)。
- 使用打包数据集发布流程时,将参数替换为文件路径中的静态值。

可以随时更改值。您可以从顶部菜单编辑参数,或使用参数列表上的**“设置”**按钮。在流程中,您可以在应用参数的任何位置使用**“设置”**按钮。执行此操作时,它会在使用该参数的任何地方重置参数的当前(默认)值,即使在自定义 SQL 查询中也是如此。

编辑用户参数

1. 从顶部菜单中,单击**“参数”** 图标。
2. 单击**“编辑参数”**。

3. 在**“编辑参数”**对话框中，进行任何更改，然后单击**“确定”**。



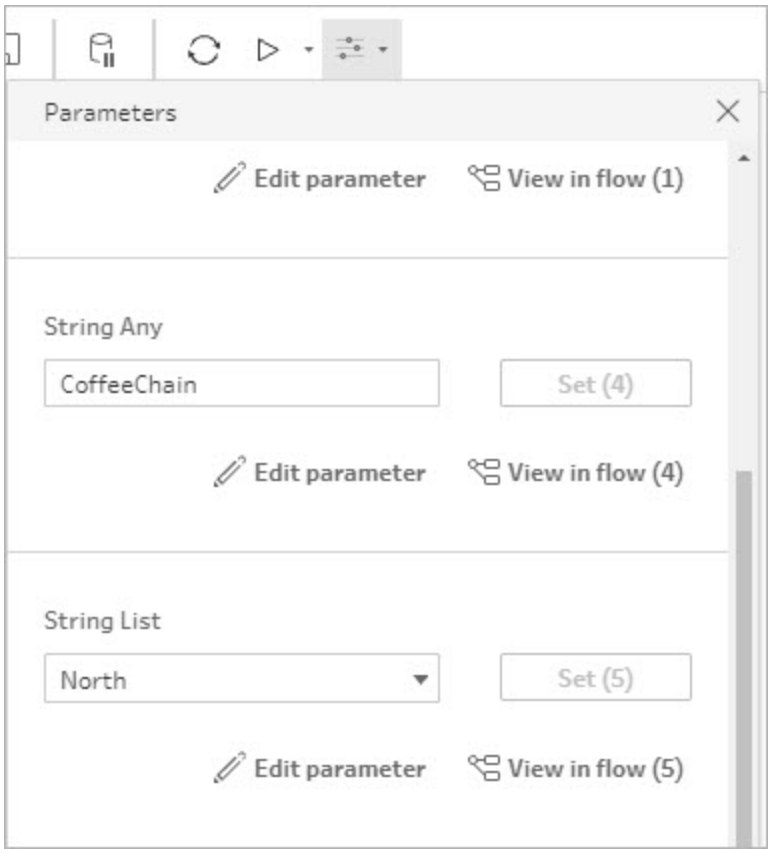
重置用户参数默认值

若要快速重置参数默认值，请使用**“设置”**按钮。该按钮向您显示一个计数，指示流程中使用该参数的位置数。

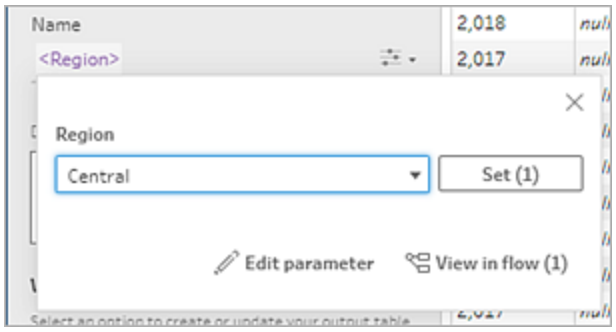
若要突出显示流程中使用参数的步骤，请单击参数对话框上的**“在流程中查看”**。如果只有一个地方使用了该参数，您将直接进入该步骤并打开配置窗格。

1. 执行以下操作之一：

   - 从顶部菜单中，单击**“参数”** 图标。使用此选项可重置流程中任何位置使用的参数值，或用于筛选器和计算字段时的参数值。



- 单击在流程中应用的参数。您可以将此选项用于文件名、文件路径、表名称、自定义 SQL 以及前后 SQL 脚本中使用的参数。



2. 选择或输入参数值。
3. 单击**“设置”**以应用更改。

将参数应用于您的流程

创建用户参数后，您可以将它们应用到整个流程中的不同位置，具体取决于步骤类型。当流程运行时，参数值将应用于该流程运行以生成适用于特定数据场景的输出。

运行流程时会自动生成系统参数(版本 2023.2 及更高版本)。只需将它们应用到您的输出步骤名称，每次运行流程时，参数都会随流程运行开始日期或时间动态更新。

将参数应用于输入步骤

在输入步骤中，您可以使用用户参数来替换文件名、文件路径部分、数据库表名，或者在使用自定义 SQL 时使用参数。

文件名或文件路径

在 Web 上编辑或制作流程时，此选项不可用。

您可以在文件路径中包含用户参数，但有一些例外。从版本 2022.1.1 开始，您还可以查看参数值的预览。

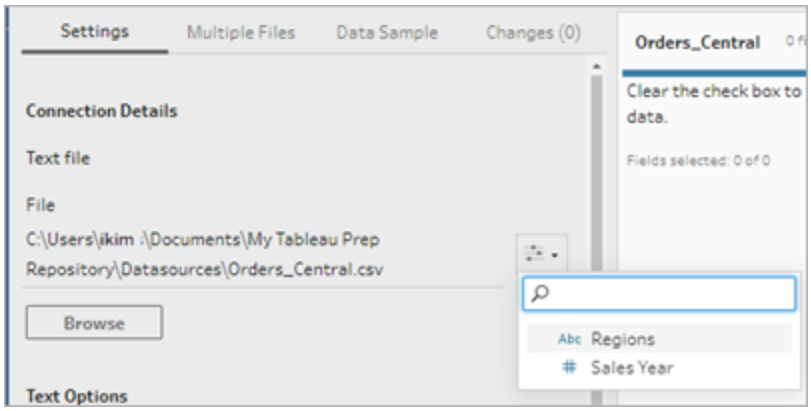
例外

- 从版本 2022.1.1 开始，您可以在 Web 上计划和运行在输入文件路径中包含参数的流程。如果使用更低版本，请在 Tableau Prep Builder 中或从命令行运行流程。
- 若要在将流程发布到 Web 时在文件路径中包含参数，则需要直接文件连接。否则，使用**“当前值”**将参数转换为静态值。

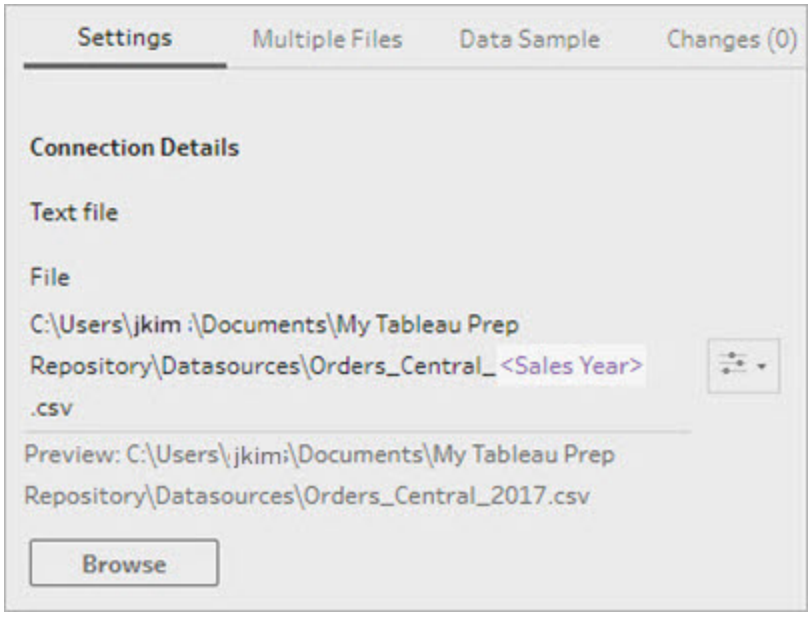
> **注意**：直接文件连接要求文件位置包含在您组织的允许列表中。有关详细信息，请参见 Tableau Server 帮助中的允许列表输入和输出位置。

将用户参数应用于文件名或路径

1. 在**“设置”**选项卡的文件路径中，将光标置于要添加参数的位置。
2. 单击参数 图标，并选择您的参数。



3. 查看参数值的预览。当前(默认)值显示在预览中。运行流程时，系统会提示您选择或输入参数值。

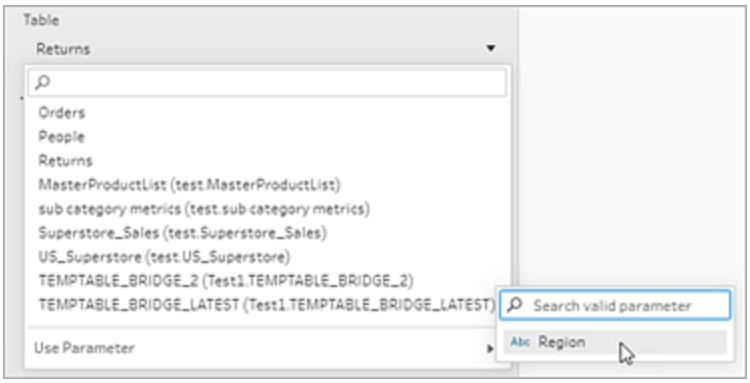


数据库表

在表名中使用用户参数时，整个表名必须是参数。当前不支持为部分表名使用参数。

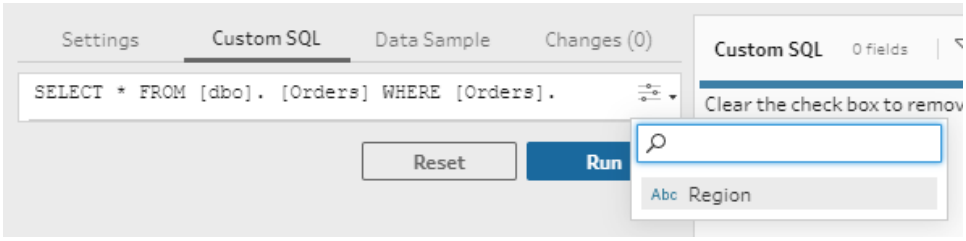
> **注意**：尚不支持为 Google BigQuery 输入连接中的表名使用参数。

1. 在**“设置”**选项卡的**“表”**字段中，单击下拉菜单。
2. 选择**“使用参数”**，然后从列表中选择参数。



自定义 SQL

1. 在**“连接”**窗格中，单击**“自定义 SQL”**。
2. 在**“自定义 SQL”**选项卡的文本框中键入或粘贴查询。
3. 单击参数 图标，并选择您的参数。



4. 单击**“运行”**以运行查询。在您运行流程之前，系统不会提示您输入参数值。相反，查询最初将使用参数的**“当前值”**运行。

> **注意**：如果该参数在流程中的其他位置使用并且**“当前值”**已重置，则该更改可能会影响您的查询。

将用户参数应用于输出步骤

在输出步骤中，您可以将用户参数应用于以下位置：

- 文件名
- 文件路径的部分
- 已发布数据源名称
- 数据库表名称
- Microsoft Excel 工作表名称(版本 2022.1.1 及更高版本)
- 在将流程输出数据写入数据库之前或之后运行的自定义 SQL 脚本(版本 2022.1.1 及更高版本)

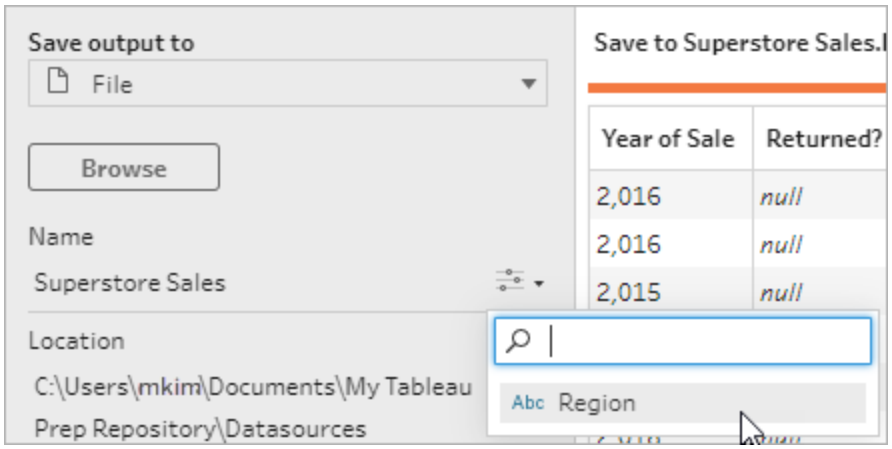
文件名或文件路径

在 Web 上创建或编辑流程时，此输出选项不可用

1. 在**“输出”**窗格中，从**“将输出保存到”**下拉列表中选择**“文件”**。

2. 在**“名称”**或**“位置”**字段中，单击参数 图标并选择您的参数。

   对于文件路径，将光标置于要添加参数的位置。

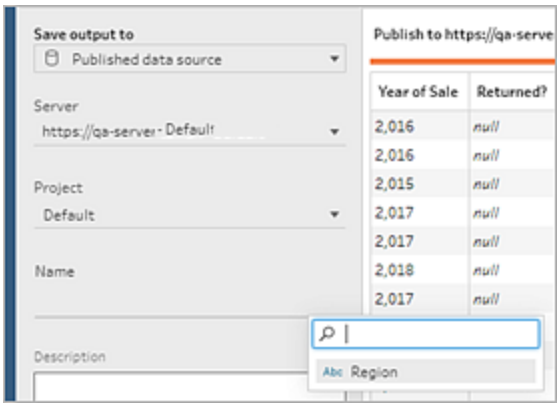


   运行流程时，系统会提示您输入参数值。

已发布数据源名称

1. 在**“输出”**窗格的**“将输出保存到”**下拉列表中，选择**“已发布数据源”**。

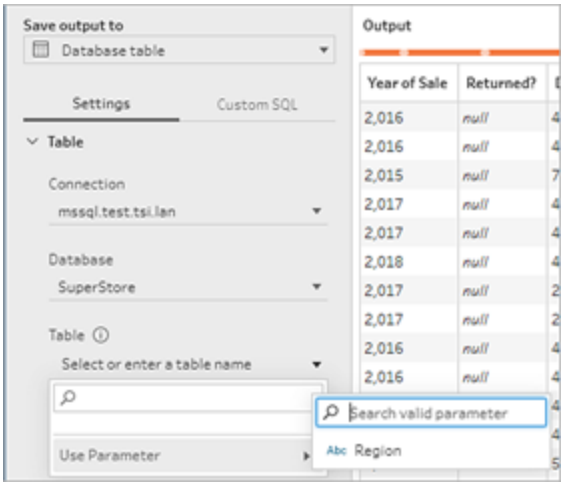
2. 在**“名称”**字段中，单击参数 图标并选择您的参数。



运行流程时，系统会提示您输入参数值。

数据库表和自定义 SQL 前后

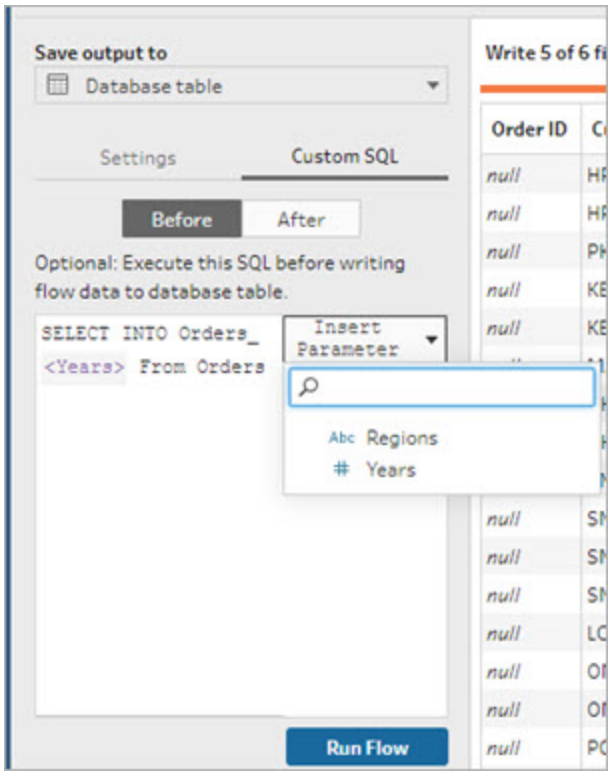
1. 在**“输出”**选项卡的**“将输出保存到”**下拉列表中，选择**“数据库表”**。
2. 在**“表”**字段中，选择**“使用参数”**，然后从列表中选择参数。



3. （可选）单击**“自定义 SQL”**选项卡。从版本 2022.1.1 开始，您可以输入带参数的 SQL 脚本，以在数据写入表**之前**和**之后**运行。若要包含参数，请单击**“插入参数”**，然后选择您的参数。

   有关在将输出写入数据库时使用 SQL 脚本的详细信息，请参见**将流程输出数据保存到外部数据库** 在本页370。

   注意：SQL 脚本中使用的参数必须手动删除。有关详细信息，请参见**手动删除用户参数** 在本页209。



运行流程时，系统会提示您输入参数值。

将系统参数应用于输出步骤

在输出步骤中，您可以将日期和时间系统参数应用于以下位置：

- 文件名
- 已发布数据源名称

文件名

在 Web 上创建或编辑流程时，此输出选项不可用

1. 在**“输出”**窗格中，从**“将输出保存到”**下拉列表中选择**“文件”**。

2. 在**“名称”**字段中，单击参数 图标并从以下运行日期或运行时间参数中进行选择。您可以组合多个系统参数来创建您需要的任何时间戳。

   运行日期
   - **日期**：YYYY-MM-DD、YYYMMDD、DD-MM-YYYY
   - **月份**：月份名称、月份编号
   - **周数**

- **季度编号**
- **年数**

运行时间

- YYYY-MM-DD_HH-MM-SS(24 小时)
- YYYYMMDD_HHMMSS(24 小时)

当您运行流程时 , Tableau Prep 使用您的本地时区或服务器时区应用流程开始运行时间。

已发布数据源名称

1. 在**“输出”**窗格的**“将输出保存到”**下拉列表中 , 选择**“已发布数据源”**。

2. 在**“名称”**字段中 , 单击参数 图标并从以下运行日期或运行时间参数中进行选择。您可以组合多个系统参数来创建您需要的任何时间戳。

   运行日期

   - **日期**:YYYY-MM-DD、YYYMMDD、DD-MM-YYYY
   - **月份**:月份名称、月份编号
   - **周数**
   - **季度编号**
   - **年数**

   运行时间

   - YYYY-MM-DD_HH-MM-SS(24 小时)
   - YYYYMMDD_HHMMSS(24 小时)

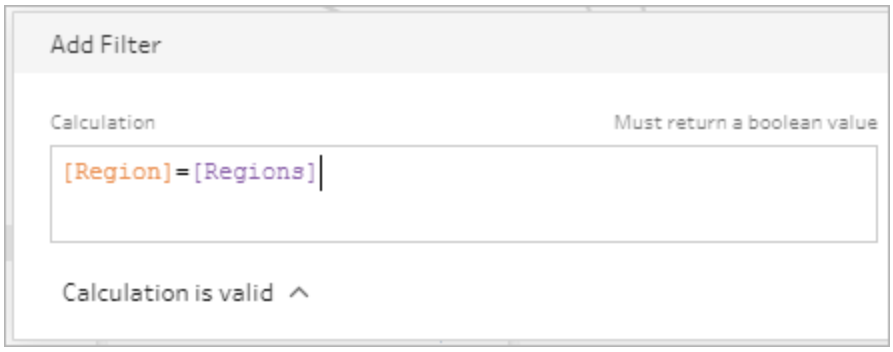
   当您运行流程时 , Tableau Prep 使用您的本地时区或服务器时区应用流程开始运行时间。

将用户参数应用于筛选器计算

使用用户参数来筛选整个流程中的数据。在输入步骤中筛选数据集或在步骤或字段值级别应用筛选器参数。例如 , 使用筛选器参数以便只为特定区域输入数据 , 或将步骤中的数据筛选到特定部门。

注意:从版本 2022.1 开始 , 当存在具有相同名称和数据类型的相同参数时 , 您可以使用复制和粘贴在其他流程中重用带有参数的筛选器计算。

1. 从配置窗格上的输入步骤或工具栏中，选择**“筛选值”**。若要将参数筛选器添加到字段，请从**“更多选项”**··· 菜单中选择**“筛选器”>“计算”**。
2. 在**“添加筛选器”**计算编辑器中，输入参数名称以从列表中选择它（参数显示为紫色），然后单击**“保存”**以保存您的筛选器。



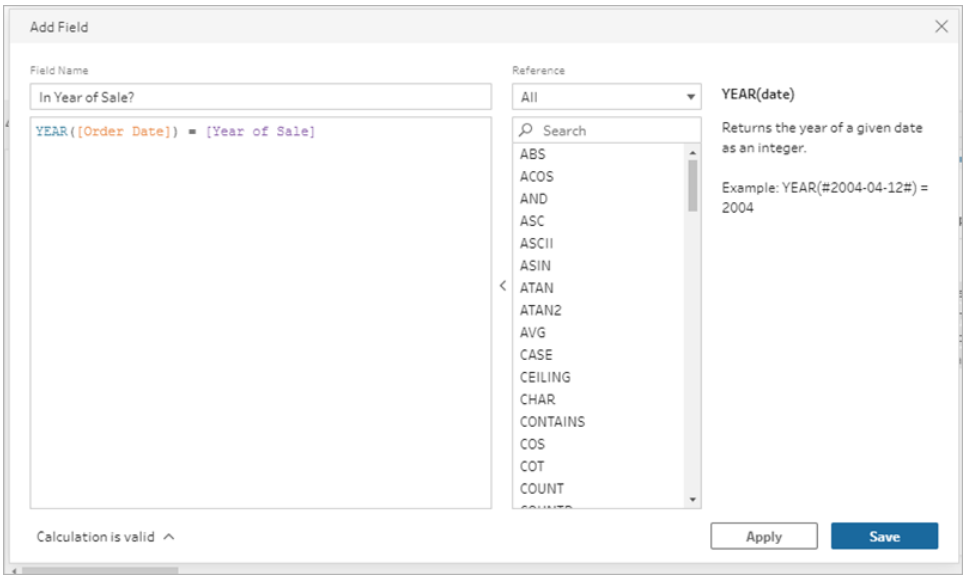
运行流程时，系统会提示您输入参数值。

将用户参数应用于计算字段

使用用户参数替换您在整个流程中使用的计算中的常量值。您可以在步骤或字段值级别应用计算参数。

注意：从版本 2022.1 开始，当存在具有相同名称和数据类型的相同参数时，您可以使用复制和粘贴在其他流程中重用带有参数的计算。

1. 从配置窗格的工具栏中，单击**“创建计算字段”**。若要将参数添加到字段的计算中，请从**“更多选项”**··· 菜单中选择**“创建计算字段”>“自定义计算”**。
2. 在**“添加字段”**计算编辑器中，输入您的计算，键入参数的名称以从列表中选择它，然后单击**“保存”**以保存您的计算。



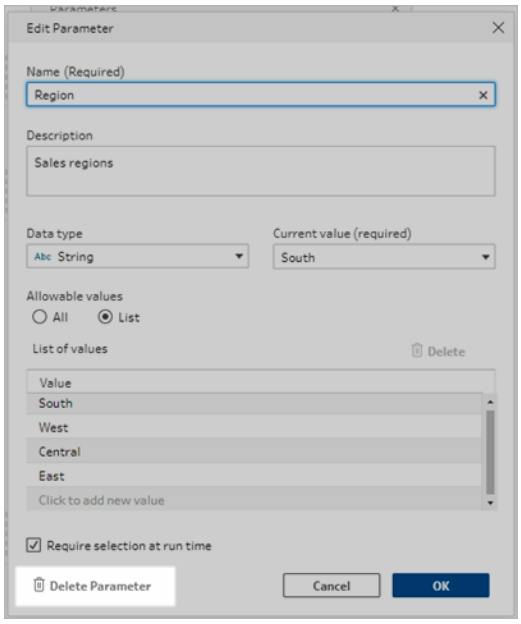
运行流程时，系统会提示您输入参数值。

删除用户参数

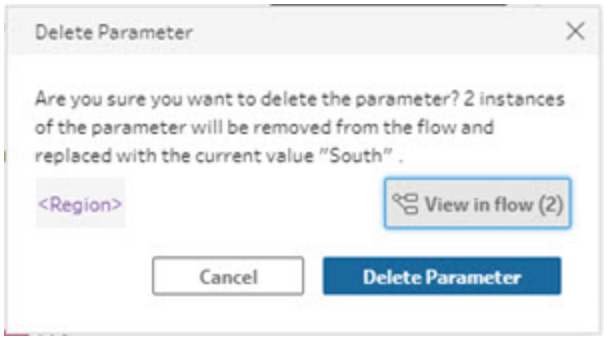
若要删除不再需要的用户参数，请单击**“编辑参数”**对话框中的**“删除参数”**。这将删除整个流程中使用的参数的任何实例，并将其替换为参数的**“当前值”**。此操作无法撤消。

注意：用于删除渡海中中参数的选项因版本而异。对于 2022.1 及更高版本，请使用以下说明。对于以前的版本，或者若要删除将输出写入数据库之前或之后运行的自定义 SQL 脚本中使用的参数，请使用手动删除用户参数 在对页。

1. 从顶部菜单中，单击参数 图标下拉菜单，然后为要删除的参数单击**“编辑参数”**。
2. 在**“编辑参数”**对话框中，单击**“删除参数”**。



3. 在确认对话框中，再次单击**“删除参数”**。您可以单击**“在流程中查看”**以突出显示步骤并调查删除参数之前参数的使用位置。



手动删除用户参数

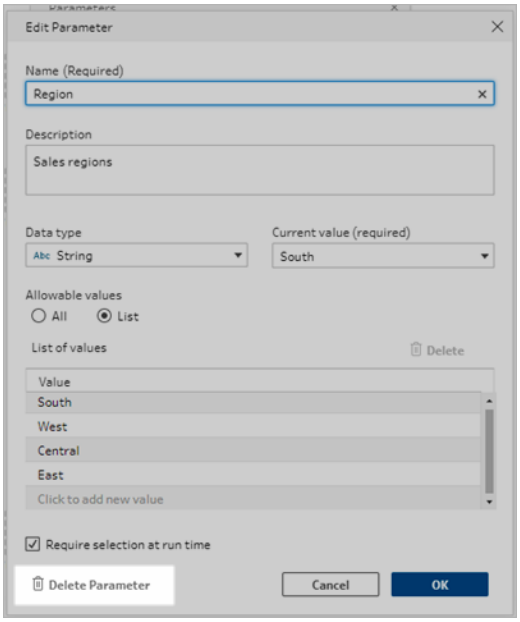
适用于版本 2021.4.4 及更早版本以及自定义 SQl 脚本前后使用的参数

在您可以从参数列表中删除用户参数之前，您必须首先从流程中找到并删除所有参数实例，甚至必须从**“更改”**窗格中执行此操作。

1. 从顶部菜单中，单击参数 图标下拉菜单。
2. 对于要删除的参数，单击**“在流程中查看”**以查找流程中使用该参数的所有实例。

如果该参数未在流程中的任何地方使用，请跳至步骤 4。

3. 对于使用参数的每个步骤，移除参数，包括删除**“更改”**窗格中列出的任何更改。
4. 从顶部菜单中，单击参数 图标下拉菜单，对于要删除的参数，单击**“编辑参数”**。
5. 在**“编辑参数”**对话框中，单击**“删除参数”**。



该参数将替换为参数的**当前值**。

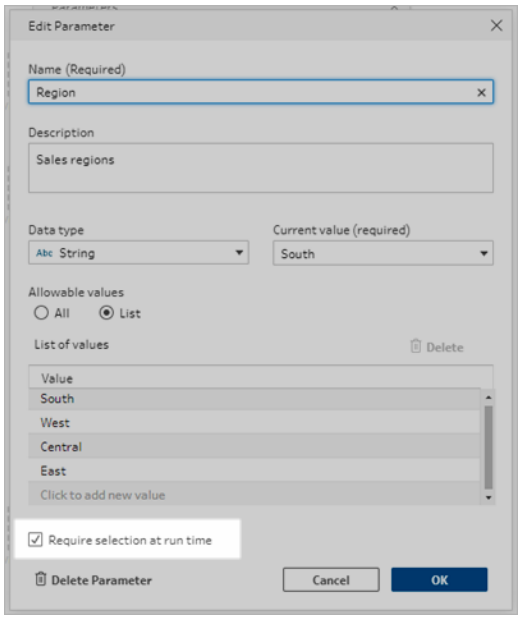
运行包含参数的流程

运行包含参数的流程与运行不包含参数的流程相同，只是在运行时或在 Tableau Server 或 Tableau Cloud.中将流程添加到计划时，会提示用户输入用户参数值。

流程运行时会自动应用系统参数。

如果用户参数被标记为必需，用户必须输入一个值才能运行流程。如果参数是可选的，用户可以输入一个值或默认接受参数的**当前值**。

必需参数是指选中了**“运行时需要选择”**(在以前的版本中为**“运行时提示输入值”**)复选框的参数。

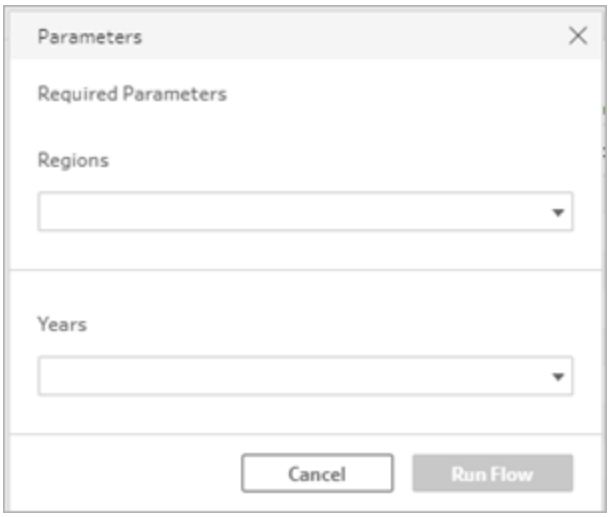


如果您使用命令行界面运行流程并希望覆盖当前(默认)参数值，请创建参数覆盖 .json 文件，并在命令行中包含 -p --parameters 语法。有关详细信息，请参见**通过命令行刷新流程输出文件** 在本页397。

手动运行流程

当您从 Tableau Prep Builder 中运行流程，或者在 Tableau Server 或 Tableau Cloud 中手动运行流程时，**“参数”**对话框会在您单击**“运行”**时打开。

1. 输入或选择用户参数值。如果流程中有可选参数，您可以在此时输入值或接受当前(默认)参数值。
2. 单击**“运行流程”**以运行流程。



有关运行流程的详细信息，请参见**将流程发布到 Tableau Server 或 Tableau Cloud** 在本页433。

按计划运行流程

当您计划流程以在 Tableau Server 或 Tableau Cloud 上运行时，您需要在计划流程时输入任何必需的用户参数值。

1. 在**“新建任务”**或**“链接任务”**选项卡上的**“设置参数”**部分，输入或选择参数值。如果流程中有可选参数，您可以在此时输入值，或将该字段留空以使用当前(默认)参数值。
2. 单击**“创建任务”**以计划您的流程。

清理和调整数据

注意：从版本 2020.4.1 开始，您现在可以在 Tableau Server 和 Tableau Cloud 中创建和编辑流程。除非特别指出，本主题中的内容适用于所有平台。有关在 Web 上制作流程的详细信息，请参见 Tableau Server 和 Tableau Cloud 帮助中的 **Web 上的 Tableau Prep**。

Tableau Prep 提供了您可用来清理和调整数据的各种清理操作。通过清理不良数据，可以更轻松地合并和分析数据，或可以让其他人在共享您的数据集时更轻松地理解您的数据。

您还可以使用转置步骤或脚本步骤来清理数据，以将 R 或 Python 脚本应用于流程。Tableau Cloud 不支持脚本步骤。有关详细信息，请参见**转置您的数据** 在本页 313或**在流程中使用 R 和 Python 脚本** 在本页 322。

关于清理操作

您通过应用诸如筛选、添加、重命名、拆分、分组或移除字段等清理操作来清理数据。您可以在流程中的大多数步骤类型中执行清理操作。您还可以在数据网格的清理步骤中执行清理操作。

您可以在输入步骤中应用有限的清理操作，并且不能在输出步骤中应用清理操作。有关在输入步骤中应用清理操作的详细信息，请参见**在输入步骤中应用清理操作** 在本页 107。

可用的清理操作

下表显示了在每个步骤类型中可以执行哪些清理操作：

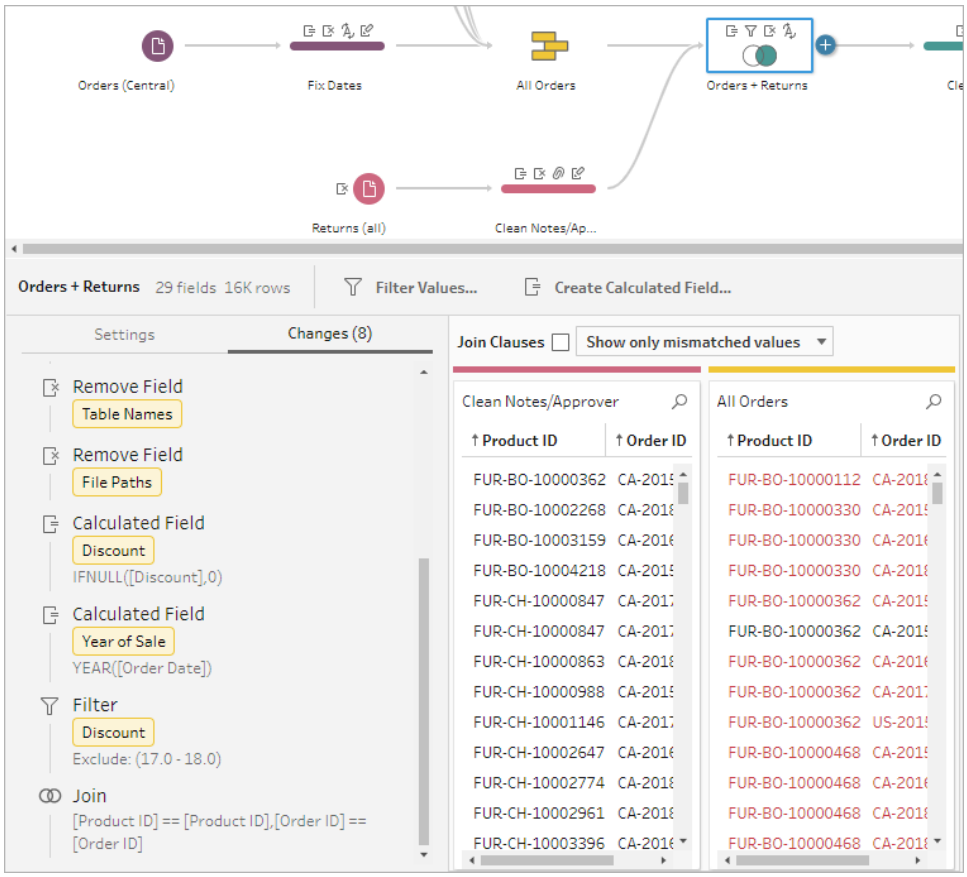
	输入	清理	聚合	转置	联接	并集	新建行	输出
筛选	X	X	X	X	X	X	X	
对值进行分组		X		X		X	X	
清理		X		X	X	X	X	
转换日期		X	X	X	X	X	X	
拆分值		X		X	X	X	X	

重命名字段	X	X		X	X	X	X	
重命名字段(批量)		X						
复制字段		X		X	X	X	X	
仅保留字段	X	X	X	X	X	X	X	
移除字段	X	X	X	X	X	X	X	
创建计算字段		X		X	X	X	X	
编辑值		X		X	X	X	X	
更改数据类型	X	X	X	X	X	X	X	

对数据进行更改时，将会向**“流程”**窗格中的对应步骤中添加注释，并会在**“更改”**窗格中添加一个条目来跟踪您的操作。如果在“输入”步骤中进行更改，则注释会显示在**“流程”**窗格中步骤的左侧，并会显示在字段列表的**“输入配置”**中。

您应用更改的顺序很重要。在“聚合”、“转置”、“联接”和“合并”步骤类型中进行的更改会在这些清理操作之前或之后执行，具体情况取决于您进行更改时字段的位置。进行更改的位置显示在步骤的**“更改”**窗格中。

以下示例显示了联接步骤中对若干字段所做的更改。将在联接操作之前执行更改以生成更正的结果。



操作顺序

下表显示了“聚合”、“转置”、“联接”和“合并”步骤类型中清理操作的执行位置，具体情况取决于字段在步骤中的何处。

操作	步骤类型：	聚合	聚合	转置	转置	联接	联接	并集	并集	新建行
	字段位置：	*分组字段*	*聚合字段*	*不在转置中*	*通过转置创建*	*包含在一个表中**	*同时包含在两个表中**	*不匹配的字段*	*合并字段*	*用于生成行的字段*
筛选		聚合之前	聚合之后	转置之	转置之后	联接之前	联接之后	合并之前	合并之	新建行之后

				前					后	
对值进行分组		NA	NA	转置之前	转置之后	联接之前	联接之后	合并之前	合并之后	新建行之后
清理		NA	NA	转置之前	转置之后	联接之前	联接之后	合并之前	合并之后	新建行之后
转换日期		聚合之前	聚合之后	转置之前	转置之后	联接之前	联接之后	合并之前	合并之后	新建行之后
拆分值		NA	NA	转置之前	转置之后	联接之前	联接之后	合并之前	合并之后	新建行之后
重命名字段		NA	NA	转置之前	转置之后	联接之前	联接之后	合并之前	合并之后	新建行之前
复制字段		NA	NA	转置之前	转置之后	联接之前	联接之后	合并之前	合并之后	新建行之后
仅保留字段		聚合之后	聚合之后	转置之前	转置之后	联接之前	联接之后	合并之前	合并之后	新建行之后
移除字段		从聚合中	从聚合中	转置	转置之后	联接之前	联接之后	合并之前	合并	新建行之后

		移除	移除	之前					之后	
创建计算字段		NA	NA	转置之前	转置之后	联接之后	联接之后	合并之前	合并之后	新建行之后
编辑值		NA	NA	转置之前	转置之后	联接之前	联接之后	合并之前	合并之后	新建行之后
更改数据类型		聚合之前	聚合之后	转置之前	转置之后	联接之前	联接之前	合并之前	合并之后	新建行之前

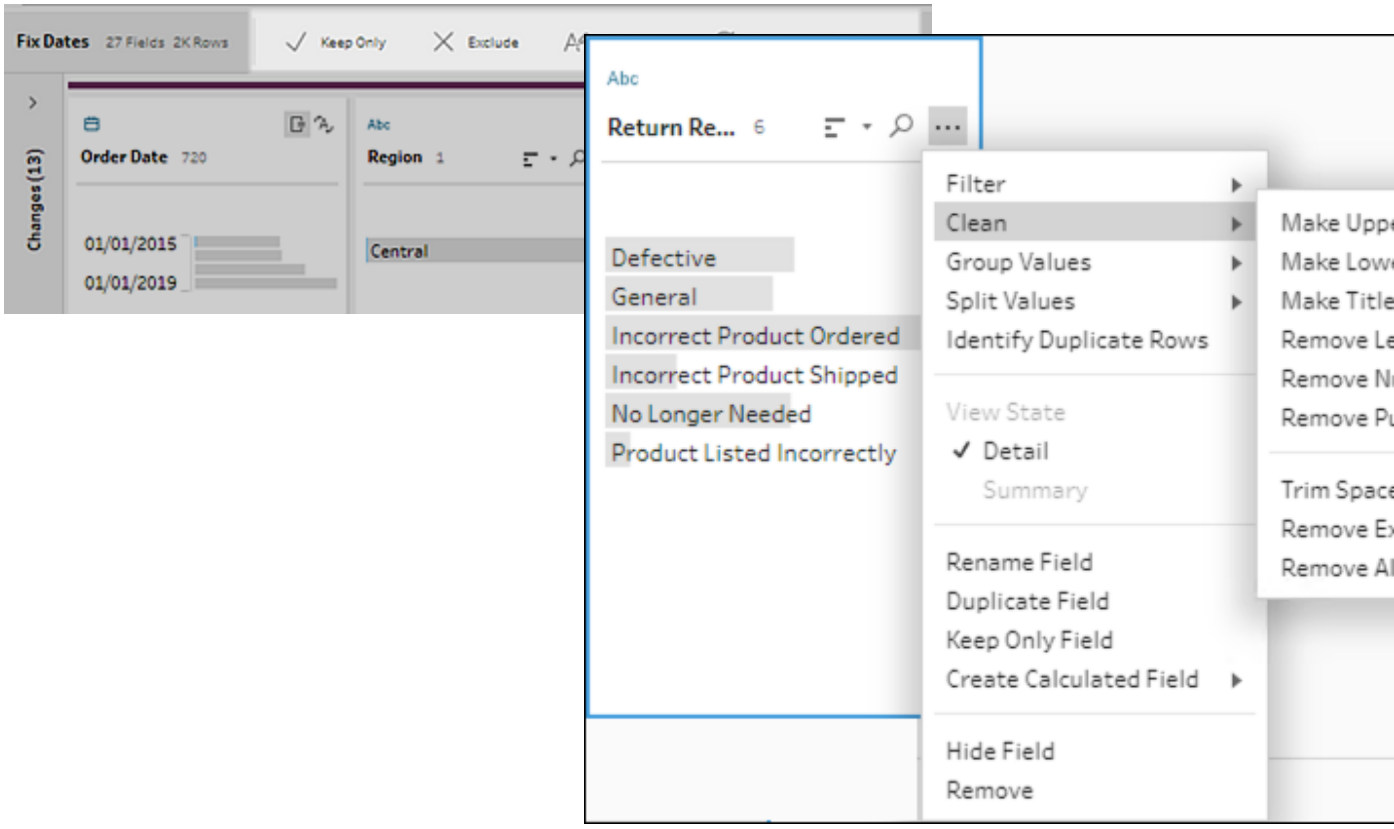
注意：对于联接，如果字段是使用一个表中的字段创建的计算字段，则在联接之前应用更改。如果字段是同时使用两个表中的字段创建的，则在联接之后应用更改。

应用清理操作

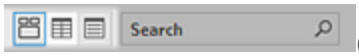
若要向字段应用清理操作，请使用工具栏选项，或在字段配置卡、数据网格或“结果”窗格上单击**“更多选项”** ··· 打开菜单。

在“聚合”、“转置”、“联接”和“合并”步骤类型中，可以在“结果”窗格和对应数据网格中的配置卡上找到**“更多选项”** ··· 菜单。如果您在整个流程中反复执行相同的清理操作或操作，则可以复制和粘贴步骤、操作甚至字段。有关详细信息，请参见**复制步骤、操作和字段** 在本页247。

“配置”窗格工具栏 **下拉菜单**

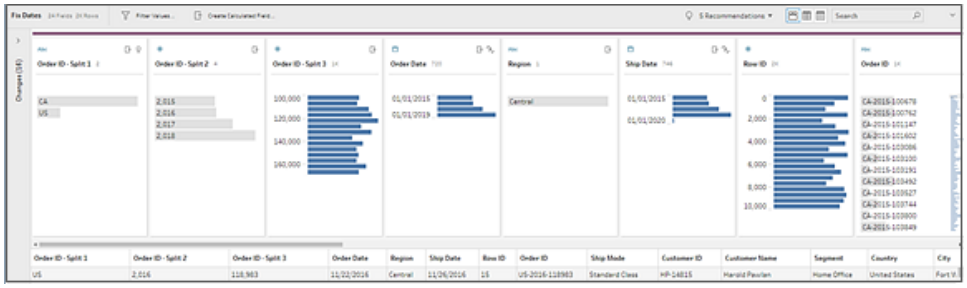


选择视图

您可以在数据网格或列表视图中的配置窗格或结果窗格外部中执行清理操作。使用视图工具栏 

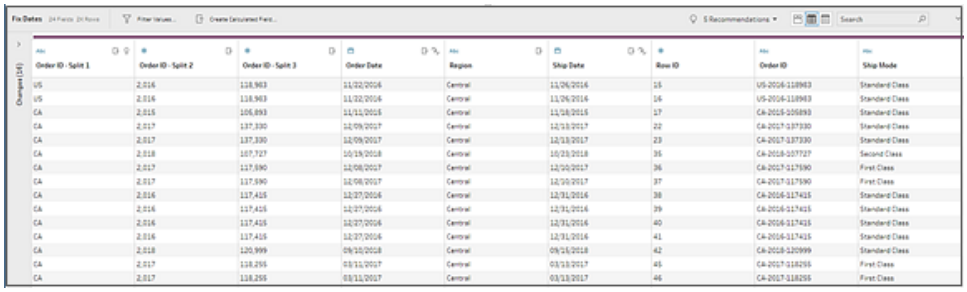
(Tableau Prep Builder 版本 2019.3.2 及更高版本以及在 Web 上)更改视图，然后在字段上单击**“更多选项”**⋯ 打开清理菜单。

- **显示配置窗格** :这是默认视图。选择此按钮可返回配置窗格或结果窗格视图。



- **显示数据网格** :折叠配置窗格或结果窗格以展开并仅显示数据网格。此视图提供了更详细的数据视图,在您需要处理特定字段值时非常有用。选择此选项后,此视图状态将在流程的所有步骤中保持不变,但您可以随时对其进行更改。

  注意:并非所有清理操作都可在数据网格中使用。举例来说,如果对值进行内联编辑,则必须使用"配置"窗格。



- **显示列表视图** (Tableau Prep Builder 版本 2019.3.2 及更高版本以及在 Web 上):将配置窗格或结果窗格转换为列表。选择此选项后,此视图状态将在流程的所有步骤中保持不变,但您可以随时对其进行更改。

  在此视图中,您可以:

  - 使用**"X"**选项选择和移除多行。
  - (版本 2021.1.4 及更高版本)使用 选项选择并隐藏或取消隐藏多行。
  - (版本 2021.2.1 及更高版本)批量重命名字段。

- 使用**“更多选项”**··· 菜单将操作应用于所选字段。

  如果为字段分配数据角色，或选择**“筛选”**、**“对值进行分组”**、**“清理”**或**“拆分值”**，则将返回到“配置”或“结果”视图来完成这些操作。可以在列表视图中执行所有其他选项。

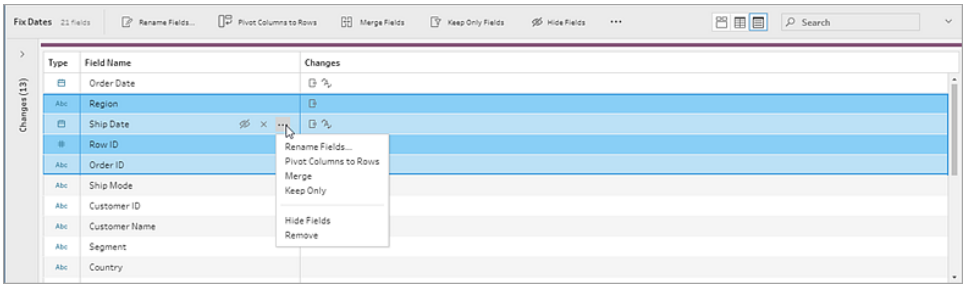
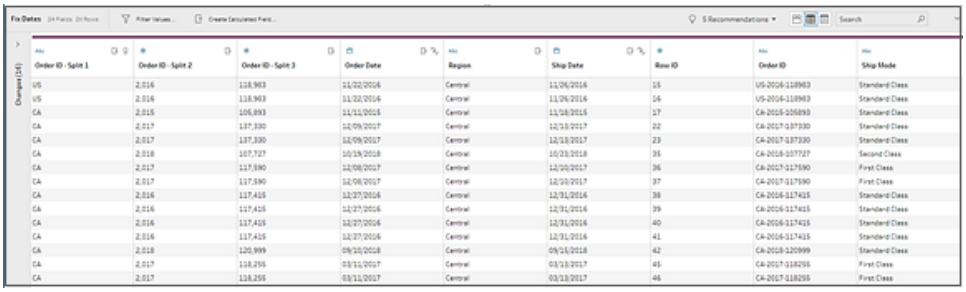


Tableau Prep Builder 版本 2019.3.1 及更低版本

使用视图工具栏

隐藏“配置”窗格，并只显示数据网格。然后在数据网络中的某个字段上单击**“更多选项”**··· 以打开清理菜单。此视图显示了更详细的数据视图，在您需要处理特定字段值时非常有用。选择此选项后，此视图状态将在流程的所有步骤中保持不变，但您可以随时对其进行更改。

注意：并非所有清理操作都可在数据网格中使用。举例来说，如果对值进行内联编辑，则必须使用“配置”窗格。



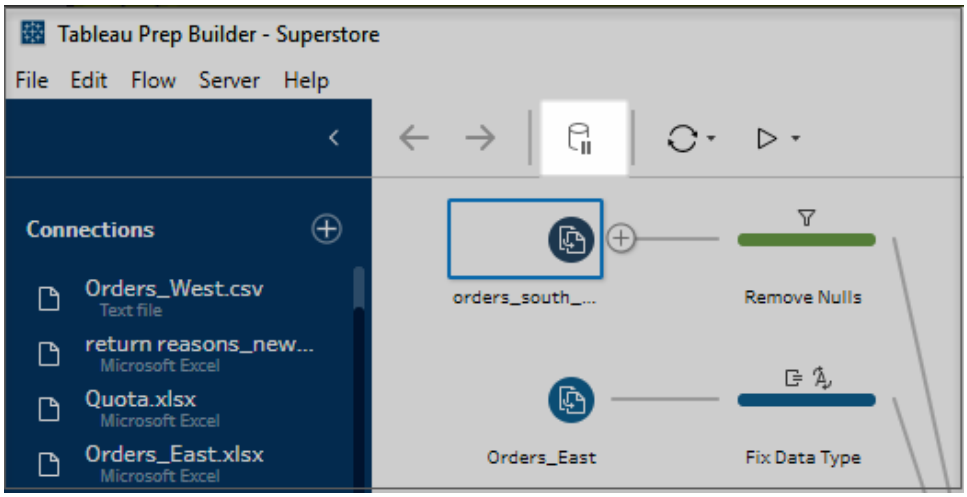
暂停数据更新以提高性能

对数据执行清理操作时，Tableau Prep 会随着进展应用更改，以便立即向您显示结果。当您知道需要进行的更改并且在进行每次更改时不需要立即反馈时，为了节省宝贵的处理时间，您可以通过暂停数据更新来提高性能。

暂停数据更新时，可以一次进行所有更改，然后恢复更新以查看结果。您可以随时恢复数据更新并启用所有可用操作。

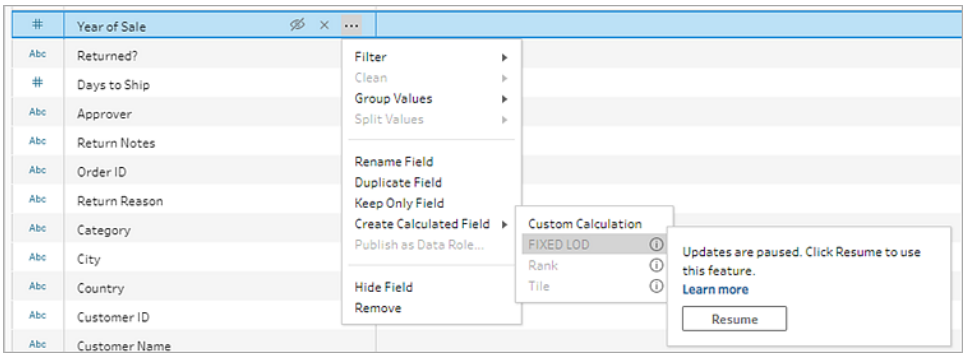
注意：暂停数据更新时，将禁用任何要求您查看值的操作。举例来说，如果要将筛选器应用于所选值，则需要查看要排除的值。

1. 在顶部菜单中，单击**“暂停数据更新”**以暂停更新。



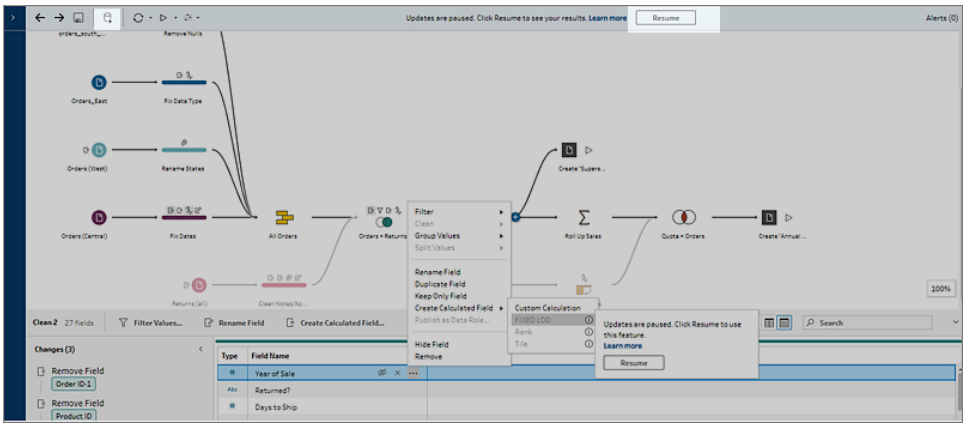
2. Tableau Prep 会将“配置”窗格转换为“列表”视图。在“列表”视图中，使用**“更多选项”** ··· 菜单将操作应用于所选字段。如果操作要求您查看值，它将被禁用。若要启用该操作，您必须恢复数据更新。

   有关使用“列表”视图模式的详细信息，请参见选择视图 在本页218。



3. 若要查看更改的结果或启用禁用的功能，请恢复数据更新。单击**“恢复数据更新”**按钮，单击菜单对话框中或“流程”窗格顶部消息横幅中的**“恢复”**按钮。

> **注意**：Tableau Prep Builder 为您提供直接从菜单中恢复更新的选项。如果在 Web 上编辑流程，您需要从顶部菜单恢复更新。



应用清理操作

若要将清理操作应用于字段，请执行以下操作：

> **注意：**从 Tableau Prep Builder 版本 2019.3.2 开始以及在 Tableau Server 和 Tableau Cloud 上从版本 2020.4 开始，您可以在列表视图中执行清理操作。

1. 在**“配置”**窗格、数据网格、“结果”窗格或列表视图中，选择要进行更改的字段。

2. 从工具栏或字段的**“更多选项”** ··· 菜单中，从以下选项中进行选择：

- **筛选**或**筛选值**：选择筛选器选项之一，右键单击或按住 Ctrl 并单击 (MacOS) 字段值以保留或排除值。您也可以使用**“选定值”**筛选器来选取和选择要筛选的值，包括流程样本中没有的值。有关筛选器选项的详细信息，请参见**筛选您的数据** 在本页 165。
- **对值进行分组**（以以前的版本中为**“分组和替换”**）：手动选择值或使用自动分组。您也可以在“配置”卡中选择多个值，并右键单击或按住 Ctrl 并单击 (MacOS) 以对值进行分组或取消分组，或者编辑组值。有关使用**“对值进行分组”**的详细信息，请参见**使用模糊匹配将值自动映射到标准值** 在本页 243。
- **清理**：从快速清理操作列表中进行选择，以便应用于字段中的所有值。
- **转换日期**（Tableau Prep Builder 版本 2020.1.4 及更高版本以及在 Web 上）：对于分配给“日期”或“日期和时间”数据类型的字段，从 DATEPART 快速清理操作列表中选择，将日期字段值转换为表示年份、季度、月份、周、日或日期和时间值的整数值。

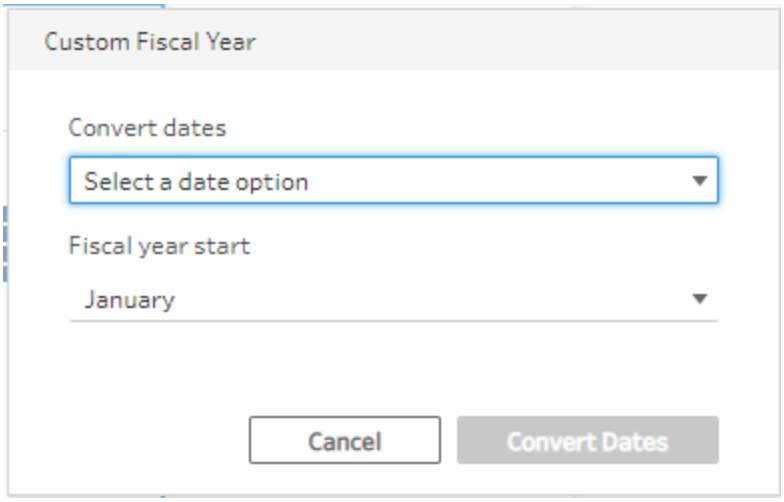
  从版本 2021.1.4 开始，您还可以从两个 DATENAME 快速清理操作中选择星期几或月份名称，以转换日期字段值。

  - **自定义会计年度**（Tableau Prep Builder 版本 2020.3.3 及更高版本以及在 Web 上）：如果您的会计年度未在 1 月开始，则可以设置自定义会计月份以使用该月份而不是默认的 1 月转换日期。

    此设置基于每个字段，因此如果要将自定义会计年度应用于其他字段，请重复此相同的步骤。

    若要打开对话框，请从**“更多选项”** ··· 菜单中选择**“转换日期”>“自定义**

会计年度”。



- **拆分值**：基于常见分隔符自动拆分值，或使用自定义拆分来指定要如何拆分字段值。

  自动拆分和自定义拆分的工作方式与它们在 Tableau Desktop 中的工作方式相同。有关详细信息，请参见 Tableau Desktop 和 Web 制作帮助中的将字段拆分为多个字段。

- **重命名字段**：编辑字段名称。
- **复制字段**（Tableau Prep Builder 版本 2019.2.3 及更高版本以及在 Web 上）：创建字段和值的副本。
- **仅保留字段**（Tableau Prep Builder 版本 2019.2.2 及更高版本以及在 Web 上）：仅保留所选字段并排除步骤中的所有其他字段。
- **创建计算字段**：在计算编辑器中编写自定义计算，或使用可视化计算编辑器（Tableau Prep Builder 版本 2020.1.1 及更高版本以及在 Web 上）创建详细级别、排名或行号计算。有关详细信息，请参见**创建详细级别、排名和分片计算** 在本页 259。
- **以数据角色形式发布**：创建自定义数据角色，然后可将其应用于您的字段，以在清理数据时验证字段值。有关此选项的详细信息，请参见**创建自定义数据角色** 在本页 181。
- **隐藏字段**：如果您有字段想要保留在流程中但不需要清理，您可以将这些字段隐藏起来，而不是删除它们。有关详细信息，请参见**隐藏字段** 在本页 166。
- **移除**（在以前的版本中为**“移除字段”**）：从流程中移除字段。

3. 若要编辑值，请右键单击或按住 Ctrl 并单击 (MacOS) 一个或多个值，选择**“编辑值”**，然后输入新值。您也可以选择**“替换为 Null”**将值替换为 Null 值，或在单个字段中双击以直接对其进行编辑。有关编辑字段值的详细信息，请参见**编辑字段值** 在本页 234。
4. 在**“配置”**窗格、“摘要”窗格或数据网络中查看这些操作的结果。

批量重命名字段

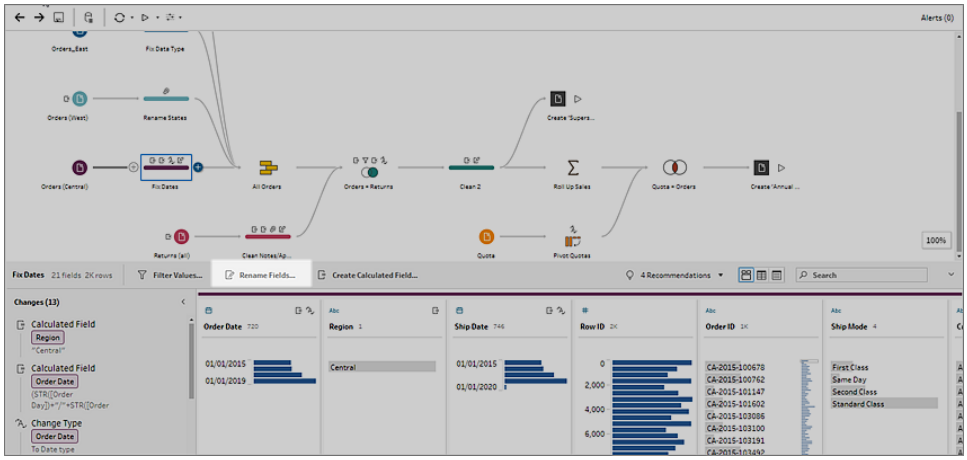
在 Tableau Prep Builder 版本 2021.2.1 及更高版本中受支持。在 Tableau Server 和 Tableau Cloud 版本 2021.2 及更高版本的 Web 版 Tableau Prep 中受支持。

使用**“重命名字段”**选项批量重命名多个字段。搜索字段名称的一部分以替换或移除它，或者为数据集中的所有或选定字段添加前缀或后缀。

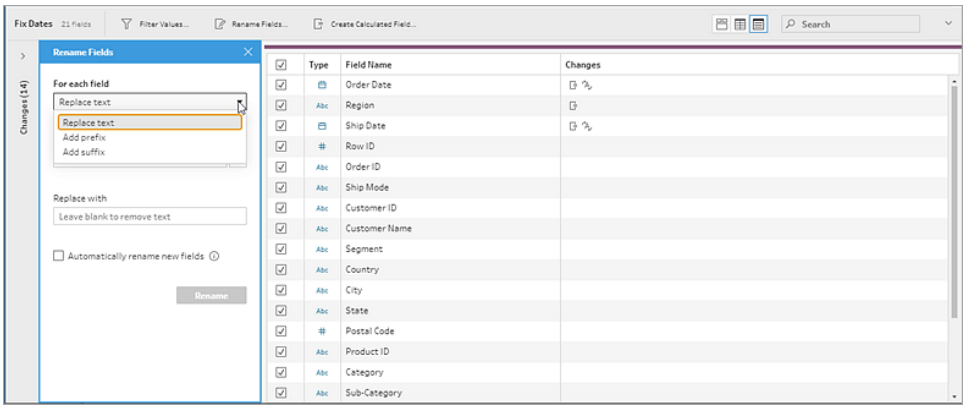
通过在进行更改时选中**“自动重命名新字段”**复选框，您还可以自动将相同的更改应用于将来添加的任何符合您条件的字段。

注意：此选项仅在**“清理”**步骤类型中可用。

1. 在**“清理”**步骤中，从工具栏中选择**“重命名字段”**。



您的视图会自动转换为显示流程中所有字段的**“列表视图”**。您可以使用工具栏中的**“搜索”**选项来缩小结果范围。

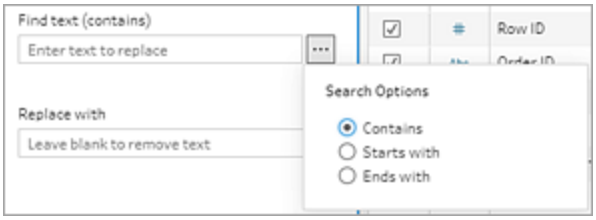


所有字段默认情况下处于选定状态。清除顶部复选框可清除所有字段的选择，从而仅手动选择要更改的字段。

2. 在**“重命名字段”**窗格中，从以下选项中进行选择：

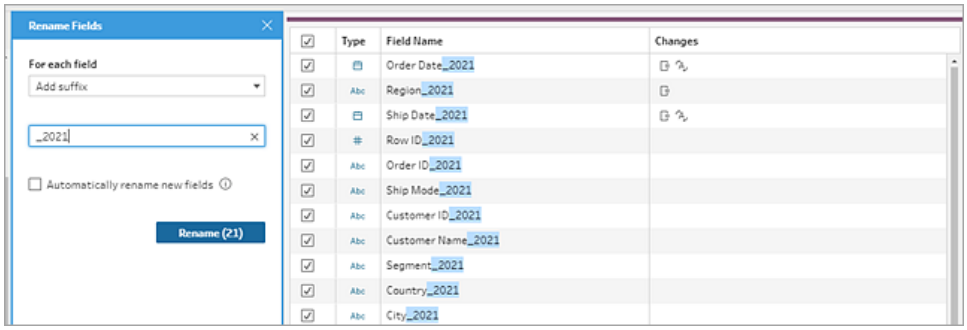
   - **替换文本**：在**“查找文本”**字段中，使用**“搜索选项”**查找匹配文本，然后在**“替换为”**字段中输入替换文本。若要查找空格，请在**“查找文本”**字段中按空格键。

     > **注意**：重命名字段不会导致空白或重复的字段名称。

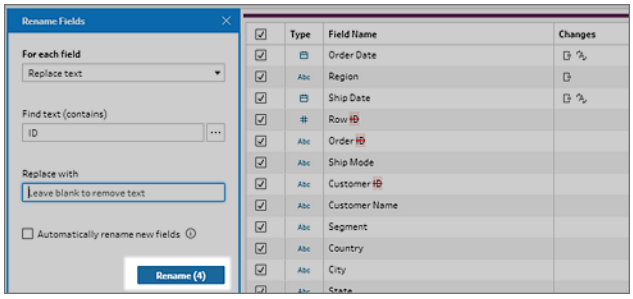


   - **添加前缀**：将文本添加到所有选定字段名称的开头。
   - **添加后缀**：在所有选定字段名称的末尾添加文本。

当您进行输入时，您的结果将显示在**“列表视图”**窗格中。



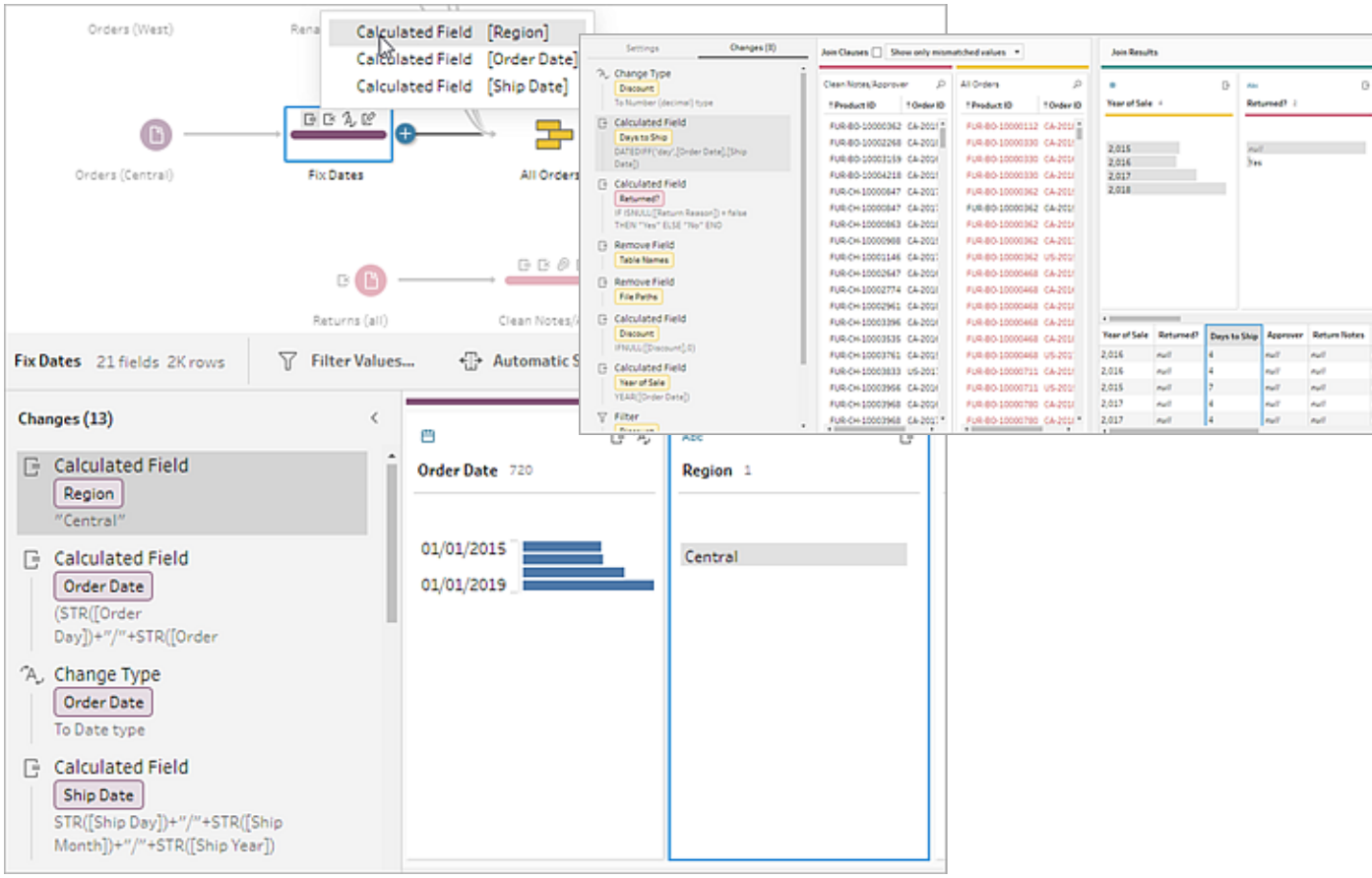
3. (可选)选择**“自动重命名新字段”**以在刷新数据时自动将这些相同的更改应用到符合替换条件的新字段。
4. 单击**“重命名”**应用您的更改并关闭窗格。**“重命名”**按钮显示受更改影响的字段数。



查看您所做的更改

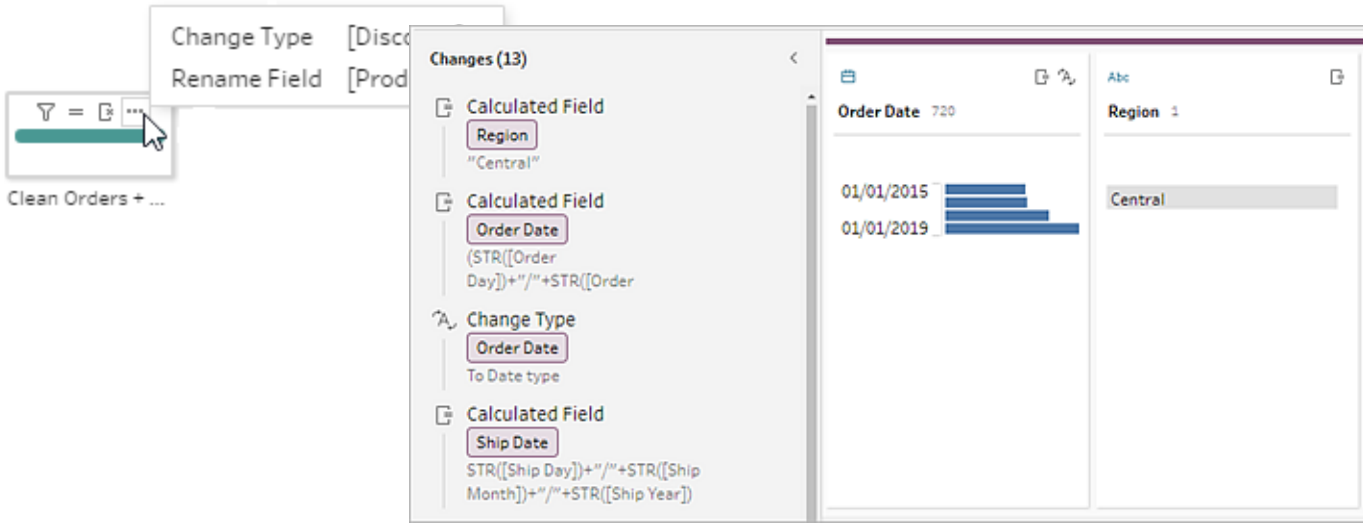
不同类型的清理操作由流程中步骤上的图标表示。如果向步骤应用了超过四种类型的操作，则步骤上会显示省略号。将光标悬停在这些图标上即可查看显示所应用的操作及其执行顺序的注释。

从 Tableau Prep Builder 版本 2019.1.3 及更高版本开始以及在 Web 上，您可以在“流程”窗格中某个步骤上单击更改图标上的注释，或者单击“配置”或“结果”窗格中的某个配置卡，更改及其影响的字段将会在**“更改”**窗格以及**“配置”**或**“结果”**窗格中突出显示。

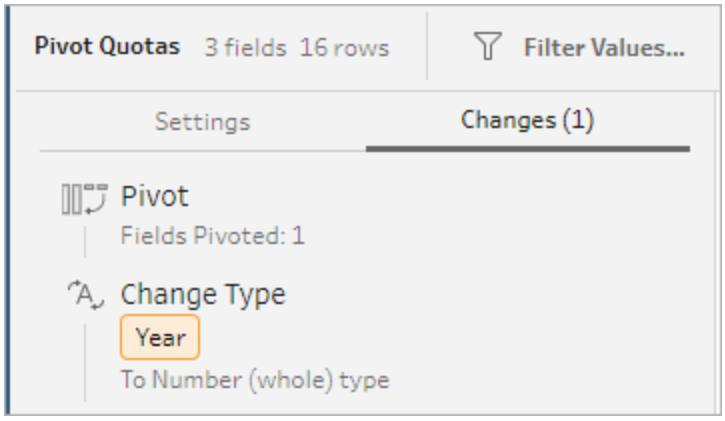


您也可以选择步骤，然后展开**“更改”**窗格以查看每个更改的详细信息、编辑或移除更改，或者将更改向上或向下拖动以更改其应用顺序，并且您可以添加描述，以便为其他用户提供上下文。有关为更改添加描述的详细信息，请参见为流程步骤和清理动作添加说明 在本页 146。

清理注释 **“更改”窗格**



在“聚合”、“转置”、“联接”或“合并”步骤中查看更改时，更改的应用顺序会在调整操作之前或之后显示。这些更改的顺序由系统应用，无法更改。您可以编辑和移除更改。



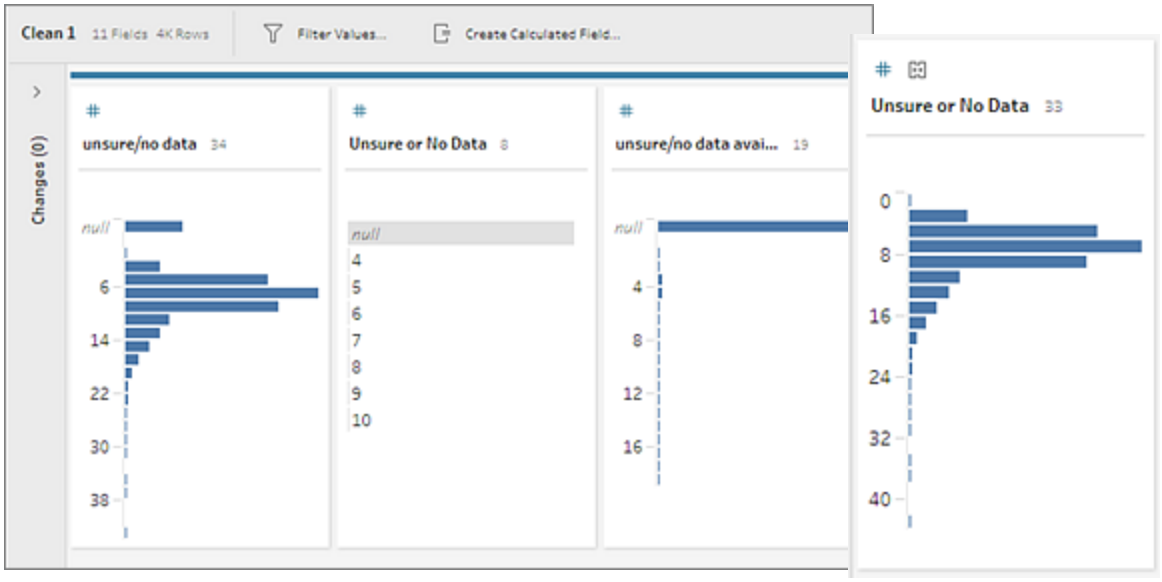
合并字段

如果字段包含的值相同，但名称不同，您可以通过将一个字段拖到另一个字段上，轻松地将它们合并为单个字段。在合并字段时，目标字段将成为主字段，并且目标字段的字段名称保持不变。合并到目标字段的字段会被移除。

示例：

输入并集生成 3 个具有相同值的字段

将 3 个字段合并为 1 个字段



合并字段时，Tableau Prep 会保留目标字段中的所有字段，并将该字段中的任何 null 值替换为您与目标字段合并的源字段中的值。源字段会被移除。

示例

Name	Contact_Phone	Business_Phone	Cell_Phone	Home_Phone
Bob	123-4567	123-4567	null	null
Sally	null	null	456-7890	789-0123
Fred	null	null	null	567-8901
Emma	null	234-5678	345-6789	null

如果将 **Business _Phone**、**Cell_Phone** 和 **Home_Phone** 字段与 **Contact_phone** 字段合并，则会移除其他字段并生成以下内容：

Name	Contact_Phone
Bob	123-4567
Sally	456-7890
Fred	567-8901

Emma	234-5678

若要合并字段，请执行以下操作之一：

- 将一个字段拖放到另一个字段上。将显示**“放置以合并字段”**指示符。
- 选择多个字段，并在所选范围内右键单击以打开上下文菜单，然后单击**“合并字段”**。
- 选择多个字段，然后在工具栏上单击**“合并字段”**。

有关如何修复合并所产生的不匹配字段的信息，请参见修复不匹配的字段 在本页349。

使用建议应用清理操作

有时可能很难确定您需要使用哪些清理操作来修复数据中的问题。Tableau Prep 可对数据进行分析并提出清理操作建议，您可以自动应用这些操作来快速修复数据字段中的问题，或者帮助确定问题以便能对其进行修复。除“输入”、“输出”和“联接”步骤类型外的所有步骤类型中提供了此功能。

注意：在 Tableau Prep Builder 中，如果不想使用此功能，您可以将其关闭。从顶部菜单中，转到**“帮助”>“设置和性能”**。然后单击**“启用建议”**以清除设置旁边的复选标记。

建议类型包括：

- 数据角色
- 筛选
- 对值进行分组（从 Tableau Prep Builder 版本 2019.2.3 开始以及在 Web 上，也适用具有数据角色的字段）
- 将列转置为行（Tableau Prep Builder 版本 2019.4.2 及更高版本以及在 Web 上）
- 将值替换为 Null 值
- 移除字段
- 拆分（Tableau Prep Builder 版本 2019.1.1 及更高版本以及在 Web 上）

  注意：此选项专用于处理固定宽度类型文本文件中的数据。若要将拆分建议与此文件类型结合使用，请在连接到数据源后，在“输入”步骤的**“文本设置”**选项

卡中选择一个数据中未使用的**“字段分隔符”**字符，使数据以单一字段形式加载。

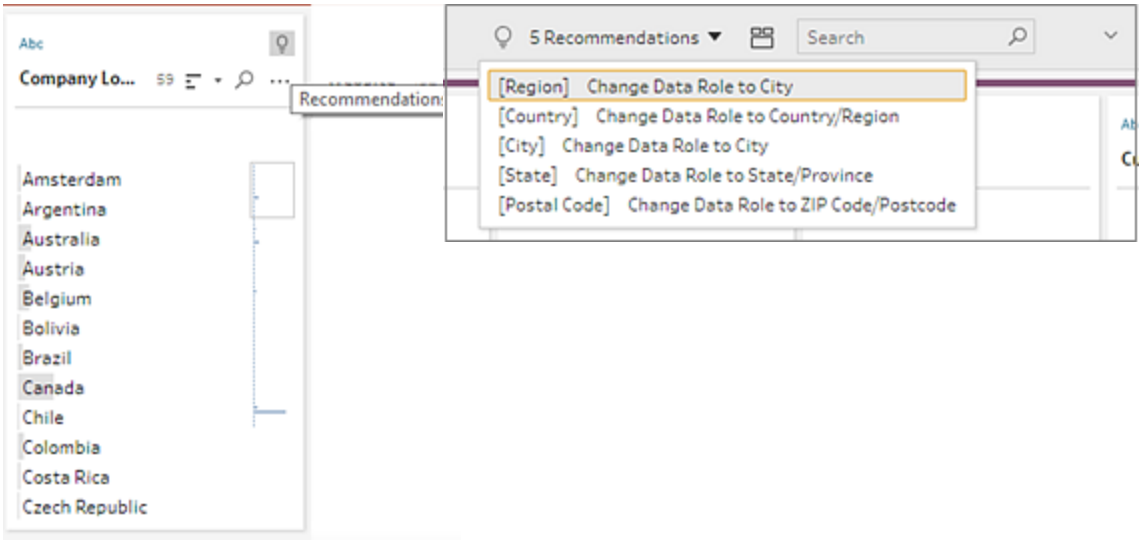
- 剪裁空格

应用建议

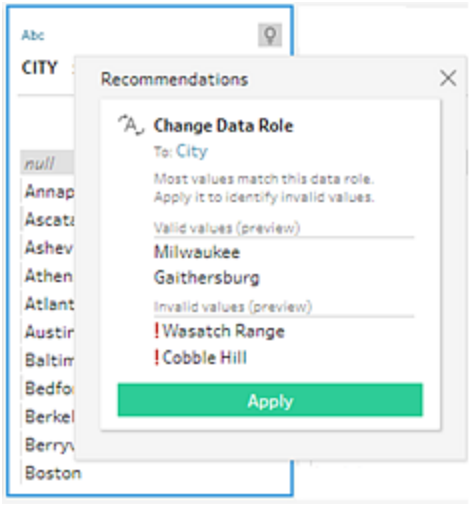
1. 执行以下操作之一：

   - 单击配置卡右上角中的灯泡 图标。
   - 从工具栏中单击**“建议”**下拉箭头以查看适用于您的数据集的所有建议，并从列表中选择一个建议。

   只有在 Tableau Prep 确定了建议的更改时，此选项才会出现。



2. 若要应用建议，请将鼠标指针悬停在“建议”卡上，然后单击**“应用”**。

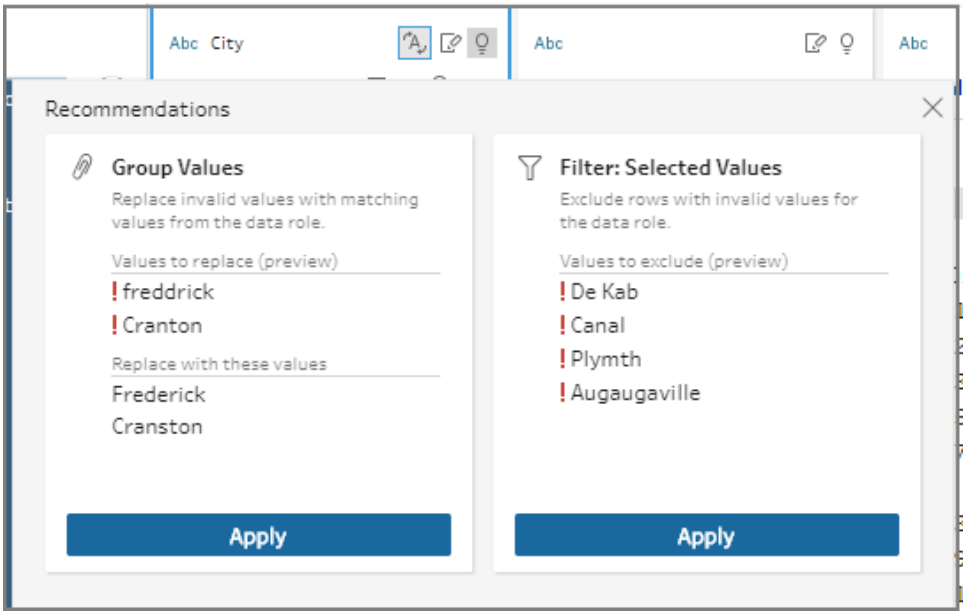


系统会自动应用更改，并向**“更改”**窗格中添加一个条目。若要移除更改，请在顶部菜单中单击**“撤消”**，或将鼠标指针悬停在**“更改”**窗格中的更改上，并单击**“X”**将其移除。

如果将建议应用于转置字段，将自动创建一个“转置”步骤，您可以在其中执行任何其他转置操作，如重命名转置字段或对其他字段进行转置。

3. 如果 Tableau Prep 因为所做更改而确定了进一步的建议，则灯泡图标会一直显示在“配置”卡上，直至找不到进一步的建议为止。

   重复上面的步骤以应用任何其他更改或忽略建议的更改，并使用其他清理工具来解决数据问题。



编辑字段值

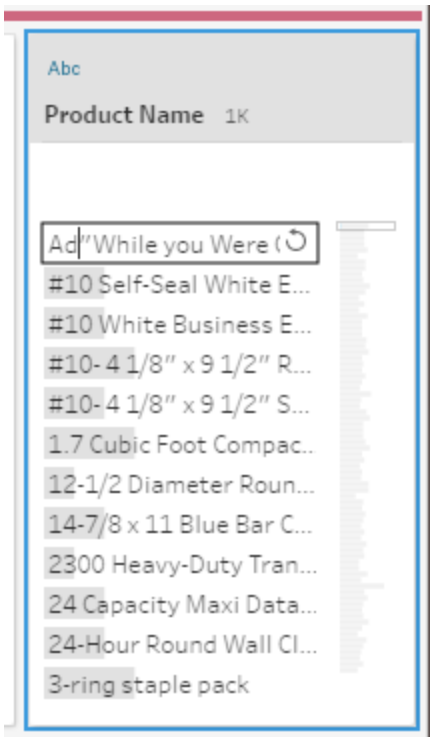
相同值的多个变体可能会让您无法准确地汇总数据。可以使用以下选项快捷轻松地纠正这些变化。

注意:对值进行的任何编辑都必须与字段数据类型兼容。

编辑单个值

1. 在**"配置"**卡中,单击要编辑的值,并输入新值。值旁边将显示一个组图标 。

   或者,右键单击值并单击**"编辑值"**。更改记录在屏幕左侧的**"更改"**窗格中。



2. 在**"配置"**窗格和数据网络中查看结果。

编辑多个值

可通过多个选项来一次性编辑多个值。例如，使用快速清理操作来移除字段中所有值的标点符号、使用多选手动将值分组、使用可找到类似值的模糊匹配算法自动将值分组在一起，或者选择多个值并将它们替换为 Null。

注意：将多个值映射到单个值时，原始字段会在值旁边显示一个分组图标，显示哪些值分组在一起。

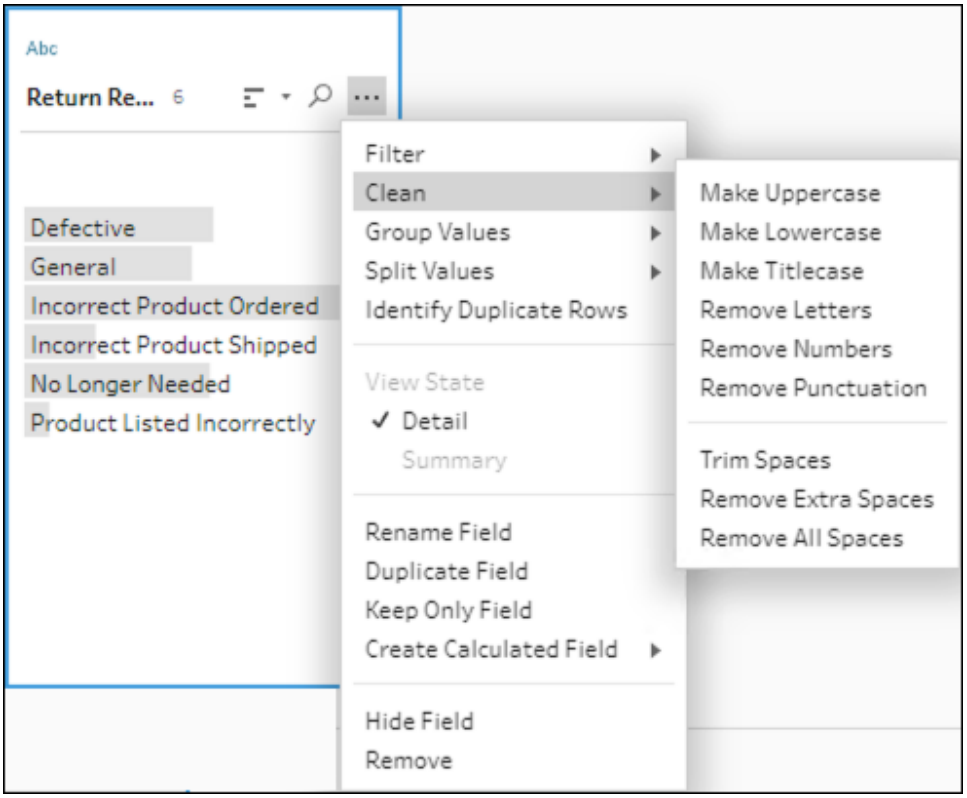
使用快速清理操作编辑多个值

此选项仅适用于文本字段。

1. 在**"配置"**窗格、"结果"窗格或数据网格中，选择要编辑的字段。

2. 单击**"更多选项"** ···，选择**"清理"**，然后选择以下选项之一：

- **设为大写**:将所有值更改为大写文本。
- **设为小写**:将所有值更改为小写文本。
- **设为首字母大写**:将所有值更改为首字母大写文本。
- **移除字母**:移除所有字母并仅保留其他字符。
- **移除数字**:移除所有数字并保留字母和其他字符。
- **移除标点符号**:移除所有标点符号。
- **剪裁空格**:移除前导和尾随空格。
- **移除额外的空格**:移除前导和尾随空格,以及将字符之间的额外空格替换为单一空格。
- **移除所有空格**:移除所有空格,包括前导和尾随空格以及字符之间的任何空格。



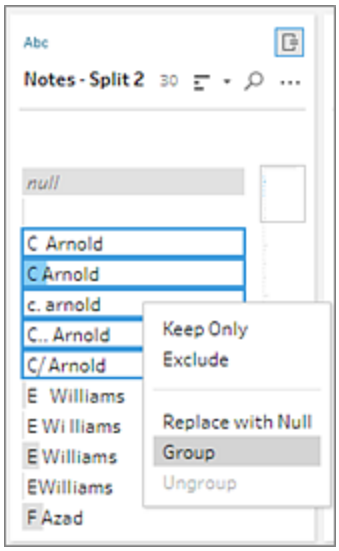
您可以累加操作,以将多个清理操作应用于字段。例如,首先选择**"清理">"移除数字"**,然后选择**"清理">"移除标点符号"**以从字段值中移除所有数字和标点符号。

3. 若要撤消更改,请单击**"流程"**窗格顶部的**"撤消"**箭头,或从更改列表中移除更改。

以内联方式编辑分组或多个值

使用此选项可手动选择多个值，并将它们分组在配置卡中的某个标准值下。若要使用其他方法对值进行分组，请参见**将多个值手动映射到标准值** 下一页和**使用模糊匹配将值自动映射到标准值** 在本页243。

1. 在**"配置"**卡中，选择要编辑的字段。
2. 按 Ctrl 或按住 Shift 并单击，或者按 Command 或按住 Shift 并单击 (MacOS)，并选择要分组的值。
3. 右键单击，并从上下文菜单中选择**"分组"**。您右键单击的所选内容中的值将成为新组的默认名称，但您可以对此名称进行内联编辑。



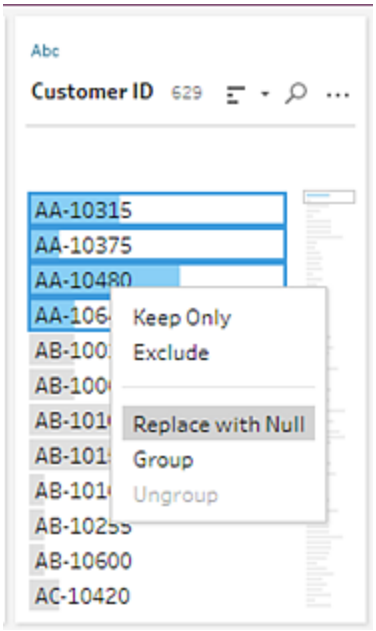
4. 若要编辑组名称，请选择分组的字段并编辑值，或者右键单击或按住 Ctrl 单击 (Mac) 分组的字段并从上下文菜单中选择**"编辑值"**。
5. 若要取消分组已分组的字段值，请右键单击分组的字段，并从上下文菜单中选择**"取消分组"**。

将一个或多个值替换为 Null

如果有要包括在分析中的数据行，但想要排除特定字段值，您可以将这些值更改为 Null 值。

1. 在**"配置"**卡中，按 Ctrl 或按住 Shift 并单击，或者按 Command 或按住 Shift 并单击 (Mac 上)，并选择要更改的值

2. 右键单击或按住 Ctrl 单击 (Mac)，并从菜单中选择**“替换为 Null”**。值将更改为 Null，并且值旁边会显示分组图标 。



将多个值手动映射到标准值

使用**“对值进行分组”**(以前的版本中为**“分组和替换”**)将字段的值从一个值映射到另一个值，或者手动选择多个值以对它们进行分组。您甚至可以添加新值，设置映射关系来组织数据。

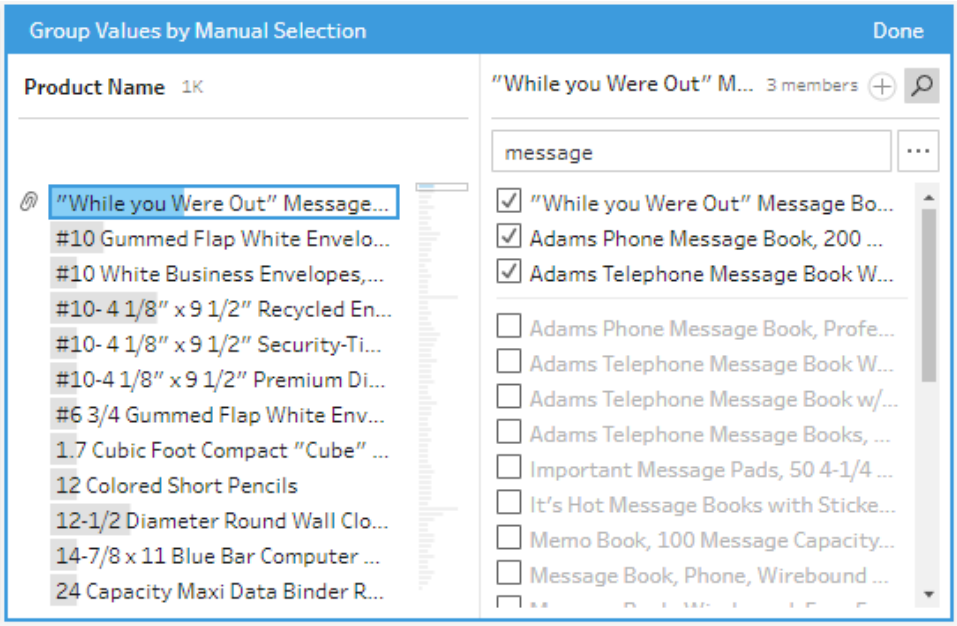
例如，假设字段中有三个值：“My Company”、“My Company Incorporated”和“My Company Inc”。所有这些值都表示同一公司，即“My Company”。您可以使用**“对值进行分组”**将值“My Company Incorporated”和“My Company Inc”映射到“My Company”，以便所有三个值在字段中都显示为“My Company”。

将多个值映射到单个所选字段

1. 在**“配置”**窗格或“结果”窗格中，选择要编辑的字段。

2. 单击**“更多选项”**…，并从菜单中选择**“对值进行分组”**(以前的版本中为**“分组和替换”**)>**“手动选择”**。

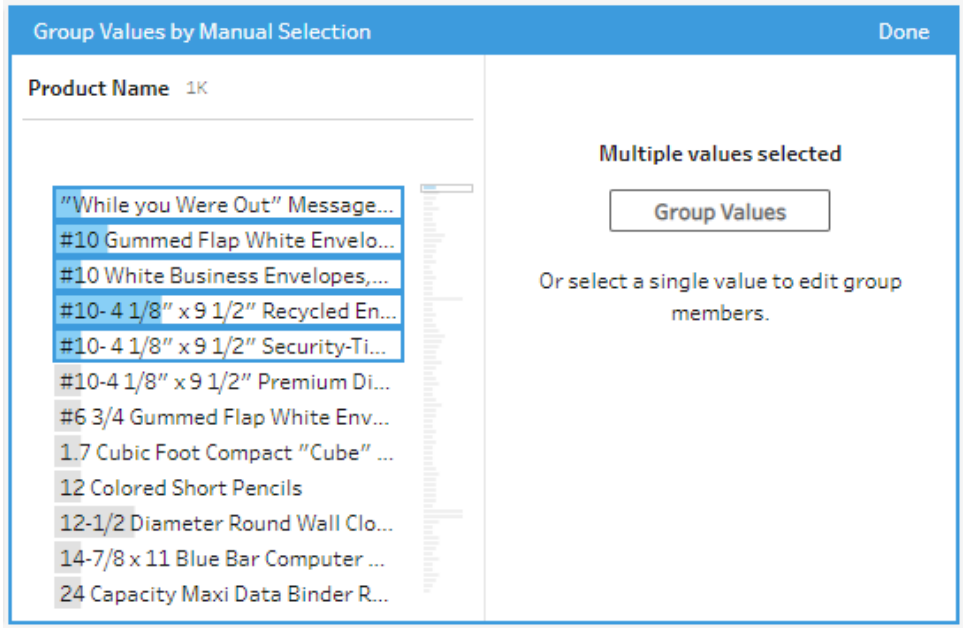
3. 在**“对值进行分组”**编辑器的左侧窗格中，选择要用作分组值的字段值。此值现在显示在右侧窗格的顶部。
4. 在**“对值进行分组”**编辑器右侧窗格的下半部分中，选择要添加到组的值。



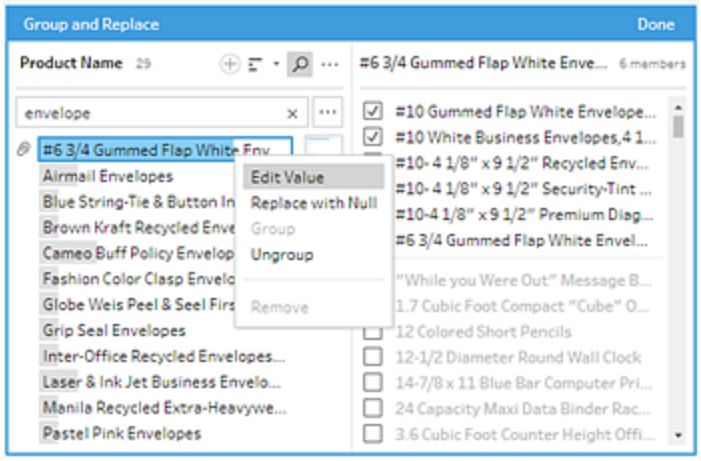
若要从组中移除值，请在**“对值进行分组”**编辑器右侧窗格的上半部分中，清除值旁边的复选框。

通过选择多个值来创建组

1. 在**“配置”**窗格或“结果”窗格中，选择要编辑的字段。
2. 单击**“更多选项”**…，并从菜单中选择**“对值进行分组”**（以前的版本中为**“分组和替换”**）>**“手动选择”**。
3. 在**“对值进行分组”**编辑器的左侧窗格中，选择要分组的多个值。
4. 在**“对值进行分组”**编辑器的右侧窗格中，单击**“对值进行分组”**。



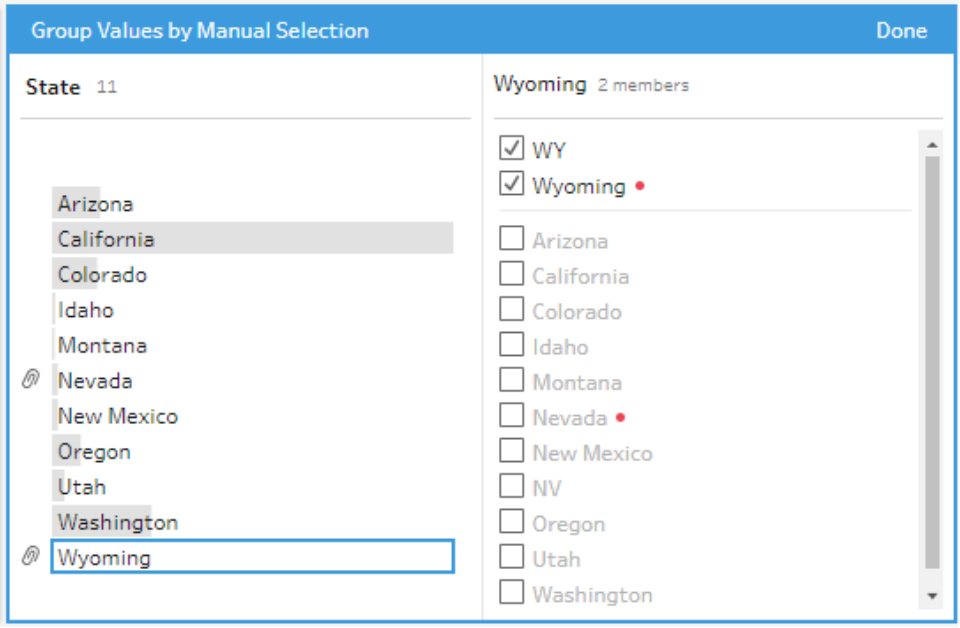
将会使用最后一个所选的值作为组名称创建一个新组。若要编辑组名称，请选择分组的字段并编辑值，或者右键单击或按住 Ctrl 单击 (MacOS) 分组的字段并从菜单中选择**"编辑值"**。



添加和标识不在数据集中的值

如果要将数据集中的值映射到不存在的新值，您可以使用**"对值进行分组"**(以前的版本中为**"分组和替换"**)来添加该值。为了轻松识别数据集中没有的任何值，这些值在**"对值进行分组"**编辑器中的值名称旁边带有红点标记。

例如在下图中，Wyoming 和 Nevada 不在数据集中。

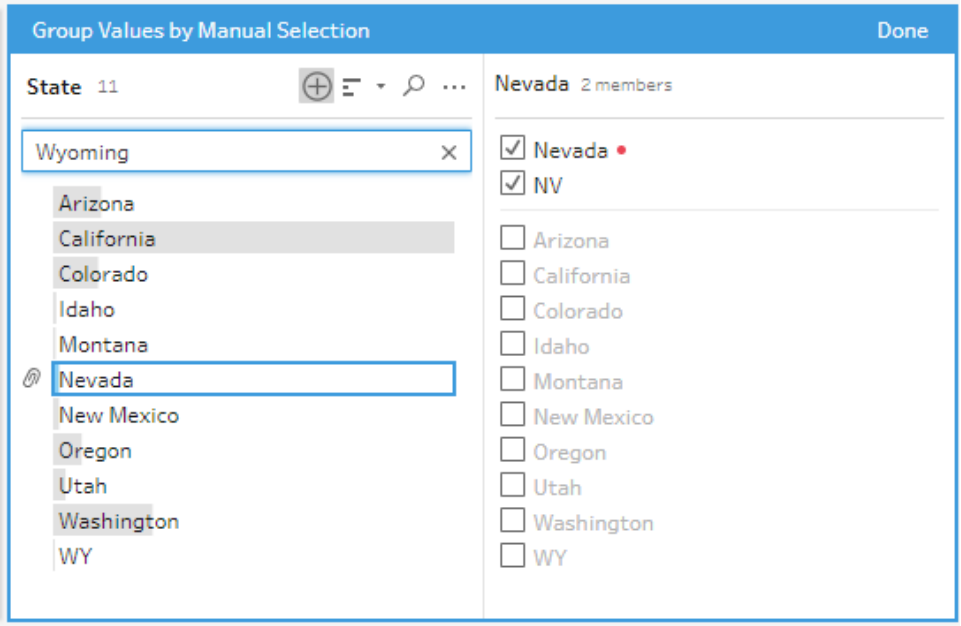


数据集中之所以可能没有某个值，可能有以下一些原因：

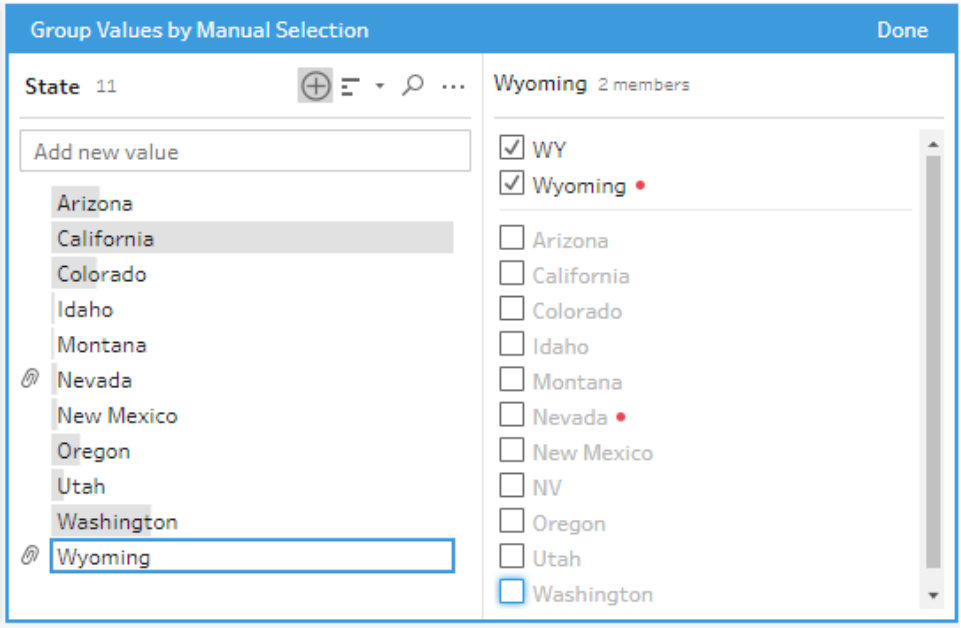
- 您刚刚手动添加了新值。
- 值不再在数据中。
- 值在数据中，但不在抽样数据集中。

添加新值：

1. 在**“配置”**窗格或“结果”窗格中，选择要编辑的字段。
2. 单击**“更多选项”**…，并从上下文菜单中选择**“对值进行分组”**（以前的版本中为**“分组和替换”**）>**“手动选择”**。
3. 在**“对值进行分组”**编辑器的左侧窗格中，单击加号 ⊕ 以添加新值。
4. 在字段中键入新值，并按 Enter 添加该值。



5. 在右侧窗格中，选择要映射到新值的值。



6. （可选）若要向映射的值中添加其他新值，请单击**“对值进行分组”**编辑器右侧窗格中的加号 ⊕ 按钮。

使用模糊匹配将值自动映射到标准值

若要搜索类似的值并对其进行自动分组，请使用模糊匹配算法之一。字段值依据最频繁出现的值进行分组。查看分组的值，并根据需要在组中添加或移除值。

如果使用数据角色来验证字段值，则可以使用**“对值进行分组”**（以前的版本中为**“分组和替换”**）选项将无效值与有效值匹配。有关详细信息，请参见**按数据角色对类似值进行分组** 在本页189

选择以下选项之一对值进行分组：

- **发音**：查找发音类似的值并对其进行分组。此选项使用 Metaphone 3 算法，该算法按单词的发音对单词编制索引，最适合于英语单词。许多常用的拼写检查器都使用这种类型的算法。此选项对于数据角色不可用。
- **常用字符**：查找具有共同字母或数字的值并对其进行分组。此选项使用 ngram 指纹算法，该算法按单词移除标点符号、重复字符和空格之后的唯一字符对单词编制索引。此算法适合于任何支持的语言。此选项对于数据角色不可用。

  例如，此算法将匹配表示为“John Smith”和“Smith, John”的名称，因为它们均生成键“hijmnost”。由于此算法不考虑发音，因此值“Tom Jhinois”将具有相同的键“hijmnost”，并也会包括在组中。
- **拼写**：查找拼写类似的值并对其进行分组。此选项使用 Levenshtein 距离算法，使用固定默认阈值来计算两个文本值之间的编辑距离。当编辑距离小于阈值时，则将这两个值分组在一起。此算法适合于任何支持的语言。

  从 Tableau Prep Builder 版本 2019.2.3 开始以及在 Web 上，此选项可在应用了数据角色之后使用。在这种情况下，它会使用编辑距离将无效值与最近的有效值匹配。如果数据集示例中没有标准值，Tableau Prep 会自动添加该值，并将值标记为不在原始数据集中。
- **发音 + 拼写**：（Tableau Prep Builder 版本 2019.1.4 及更高版本以及在 Web 上）如果为字段分配数据角色，您可以使用该数据角色，通过数据角色定义的标准值来匹配值和对值进行分组。此选项随后将根据拼写和发音将无效值与最相似的有效值匹配。如果数据集示例中没有标准值，Tableau Prep 会自动添加该值，并将值标记为不在原始数据集中。此选项最适合于英文单词。

  有关详细信息，请参见**清理和调整数据** 在本页213想了解更多关于这些模糊匹配算法的信息吗？ 在 Tableau.com 上查看 Tableau Prep Builder 中的自动分组

注意：在 Tableau Prep Builder 版本 2019.1.4 和 2019.2.1 中，此选项标为**“数据角色匹配”**。

使用模糊匹配分组类似值

1. 在**“配置”**窗格或“结果”窗格中，选择要编辑的字段。
2. 单击**“更多选项”**⋯ 并选择**“对值进行分组”**，然后选择以下选项之一：
   - **发音**
   - **常用字符**
   - **拼写**

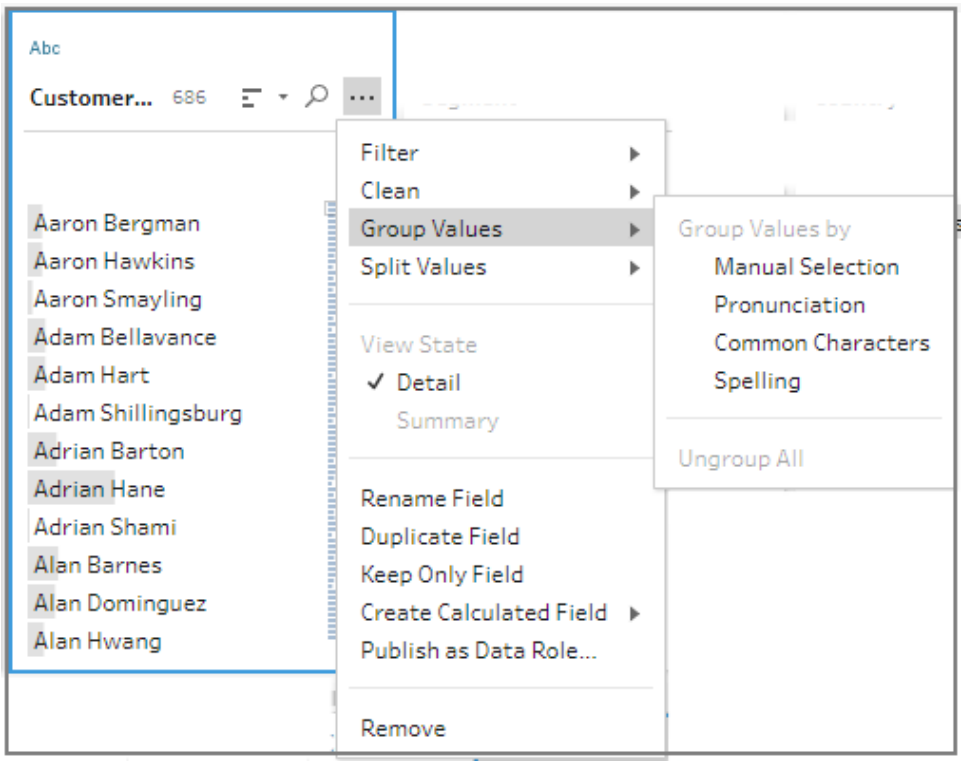
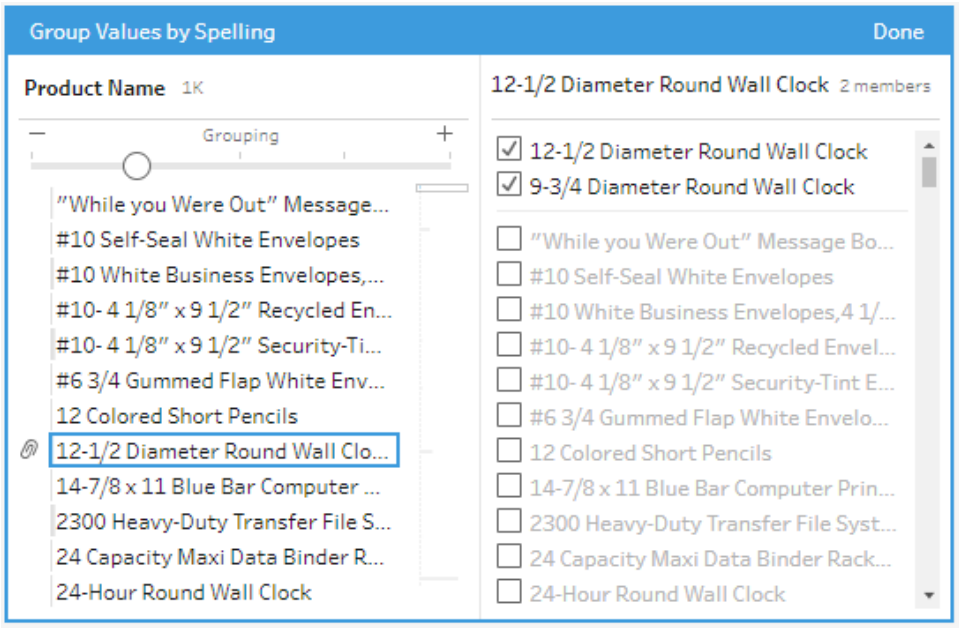


Tableau Prep Builder 将查找匹配的值并对其进行分组，并将它们替换为组中出现最频繁的值。

3. 查看分组，并根据需要手动添加或移除值或对其进行编辑。然后单击**“完成”**。



在对字段值进行分组时调整结果

如果按**“拼写”**或**“发音”**对类似值进行分组，您可以使用字段上的滑块来调整分组参数的严格程度，从而更改结果。

根据滑块的设置方式，您可以更大程度地控制组中包括的值数量以及创建的组数量。默认情况下，Tableau Prep 会检测最佳分组设置，并将滑块显示在该位置。

更改阈值时，Tableau Prep 会对值的样本进行分析来确定新的分组。依据设置生成的组保存并记录在**“更改”**窗格中，但不会保存阈值设置。下次通过编辑现有更改或进行新更改打开**“对值进行分组”**编辑器时，阈值滑块会显示在默认位置，使您能够根据当前数据集进行任何调整。

1. 在**“配置”**窗格或**“结果”**窗格中，选择要编辑的字段。

2. 单击**“更多选项”**…，并选择**“对值进行分组”**(在前的版本中为**“分组和替换”**)，然后选择以下选项之一：

   - **发音**
   - **拼写**

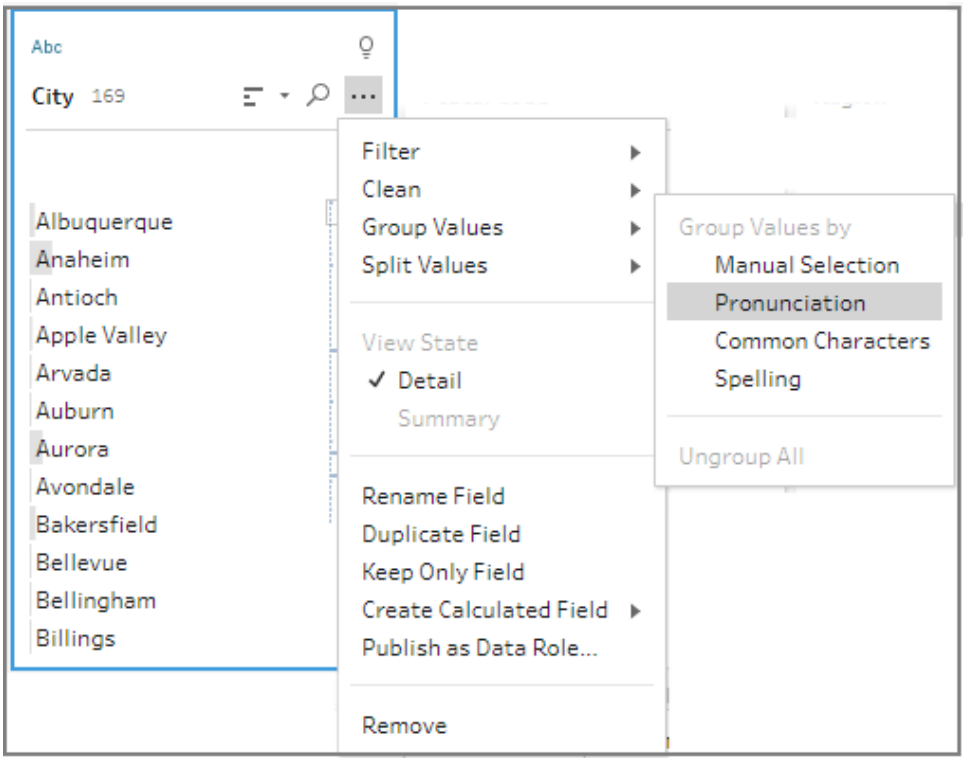
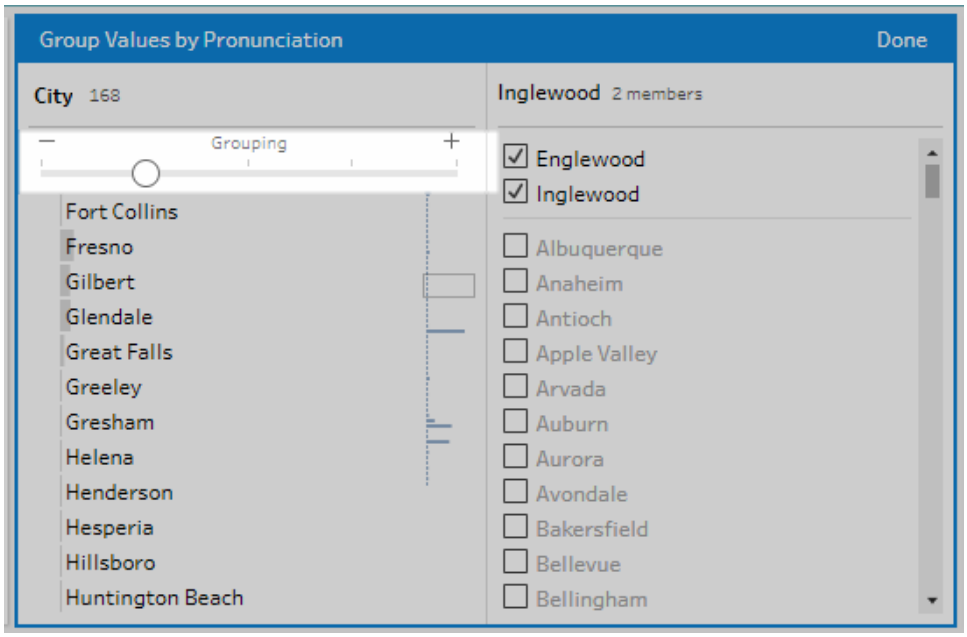


Tableau Prep 将查找匹配的值并对其进行分组，并将它们替换为组中出现最频繁的值。

3. 在**“对值进行分组”**编辑器的左侧窗格中，将滑块拖到 5 个阈值级别之一以更改结果。



若要设置更严格的阈值，请将滑块向左移动。这会使匹配数更少，并创建更少的组。
若要设置更宽松的阈值，请将滑块向右移动。这会使匹配数更多，并创建更多的组。

4. 单击**"完成"**保存您所做的更改。

复制步骤、操作和字段

注意：从版本 2020.4.1 开始，您现在可以在 Tableau Server 和 Tableau Cloud 中创建和编辑流程。除非特别指出，本主题中的内容适用于所有平台。有关在 Web 上制作流程的详细信息，请参见 Tableau Server 和 Tableau Cloud 帮助中的 **Web 上的 Tableau Prep**。

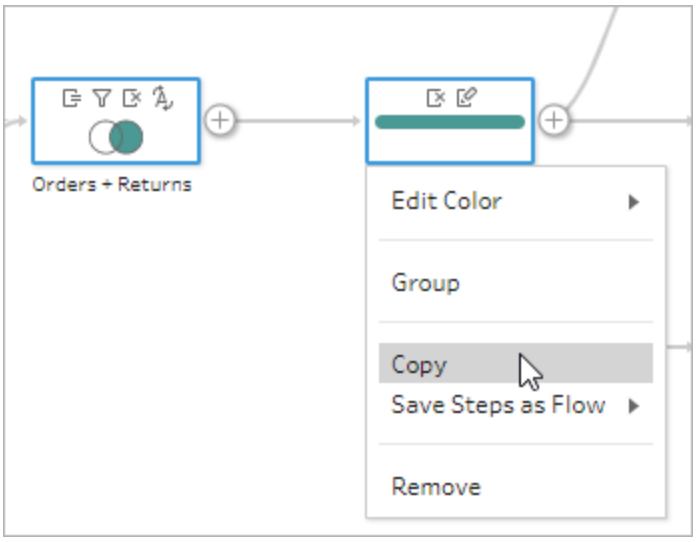
在清理数据时，您通常会在整个流程中反复执行相同的清理操作或动作。为了使数据的清理和调整更加高效，您可以在整个流程中复制和粘贴这些操作或动作，或者甚至复制选定的步骤或组并保存它们，这样您就能执行清理操作或动作一次，然后在需要的地方重用该操作或动作。您甚至可以复制字段来体验不同的清理操作。

有关在流程中创建组的详细信息，请参见**对步骤进行分组** 在本页 142。

复制并粘贴步骤

复制一个或多个步骤以在同一流程的另一个区域中使用它们。此选项不可用于在输入步骤中包括并集的输入步骤。

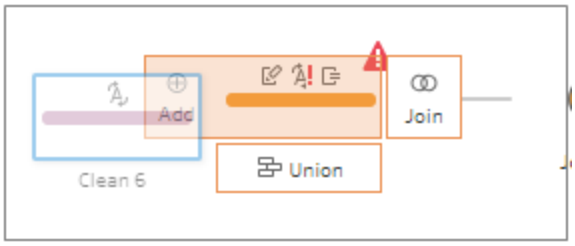
1. 在“流程”窗格中，选择流程中的一个或多个步骤或组。
2. 右键单击或按住 Ctrl 单击 (MacOS) 选定的步骤，然后选择**“复制”**。



3. 若要粘贴复制的步骤，请执行下列操作之一：
   - 将光标悬停在步骤或流程线上，直至出现加号 ⊕ 图标，然后从菜单中单击该图标并选择**“粘贴”**。
   - 在画布的任何空白处右键单击或按住 Ctrl 并单击 (MacOS)，然后单击**“粘贴”**。

4. 如果在流程空白处粘贴步骤，请将步骤拖放到流程中要放置步骤的位置。如果将步骤添加到流程步骤的末尾，则这些步骤将自动添加到步骤的末尾。如果在现有流程步骤之间插入步骤。将步骤移到流程中所需的位置并修复任何错误。

如果需要，您可以移除流程行或四处移动步骤。例如，若要将某个步骤连接到复制的步骤，请移除现有流程行(如果有)，然后将现有步骤拖到新步骤，并放在**“添加”**上。



有关组织流程的详细信息，请参见**重新组织流程的布局** 在本页150。

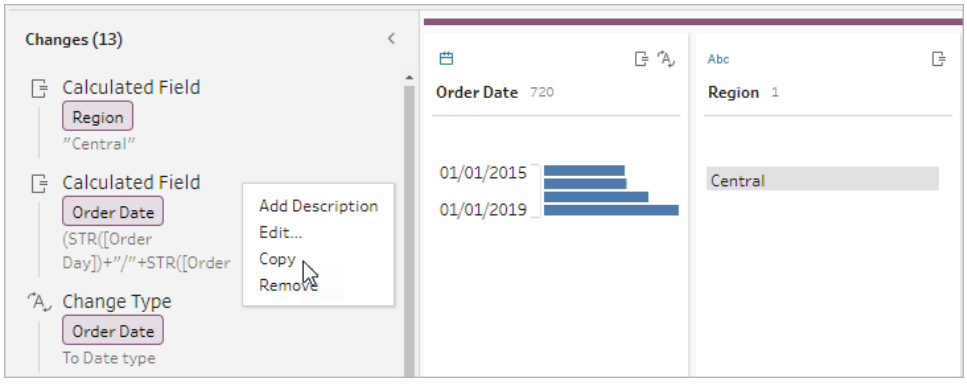
复制并粘贴清理操作

您可以在同一流程中复制和粘贴清理操作，以便通过以下选项之一来重用操作：

- 从一个步骤的**“更改”**窗格中复制操作，然后将其粘贴在同一步骤或另一个步骤的**“更改”**窗格中，从而在该步骤中应用该相同操作。
- 将一个操作从**“更改”**窗格拖放到“配置”窗格中的其他字段，以将该步骤应用于多个字段。此选项不可用于影响多个字段(例如计算字段)的操作。

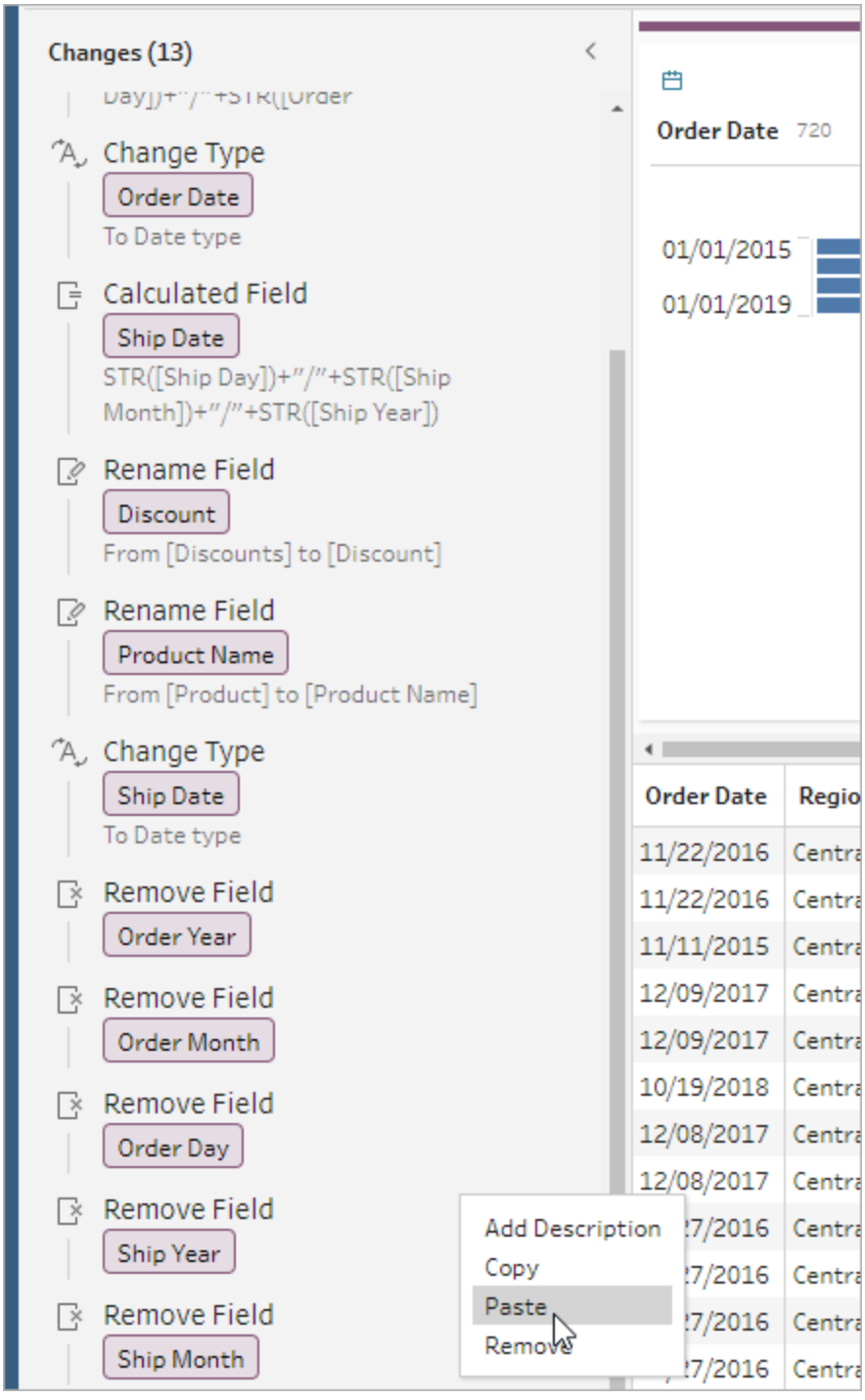
若要将步骤中的更改复制并粘贴到同一步骤或其他步骤，请执行以下操作：

1. 在**“更改”**窗格中，选择要复制的更改。
2. 在更改项上右键单击或按住 Ctrl 并单击 (MacOS)，然后从菜单中选择**“复制”**。



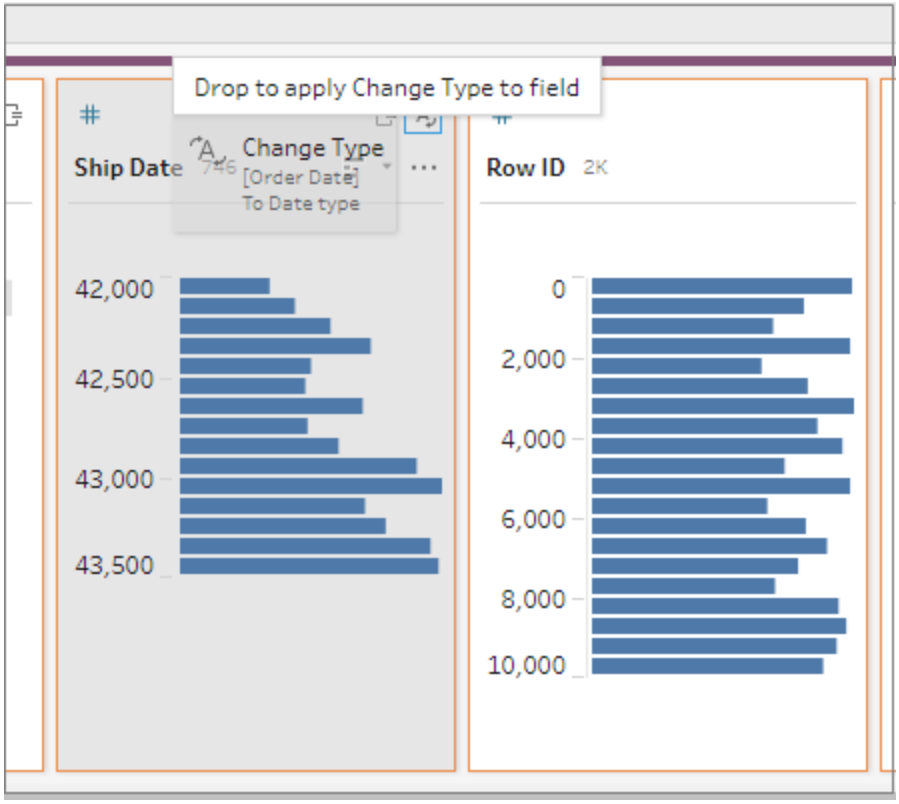
3. 在要在其中粘贴更改的**“更改”**窗格中，右键单击或按住 Ctrl 并单击 (MacOS)，并选

择**“粘贴”**。选择更改，并根据需要单击**“编辑”**进行任何调整。



若要将更改拖放到步骤中的其他字段，请执行以下操作：

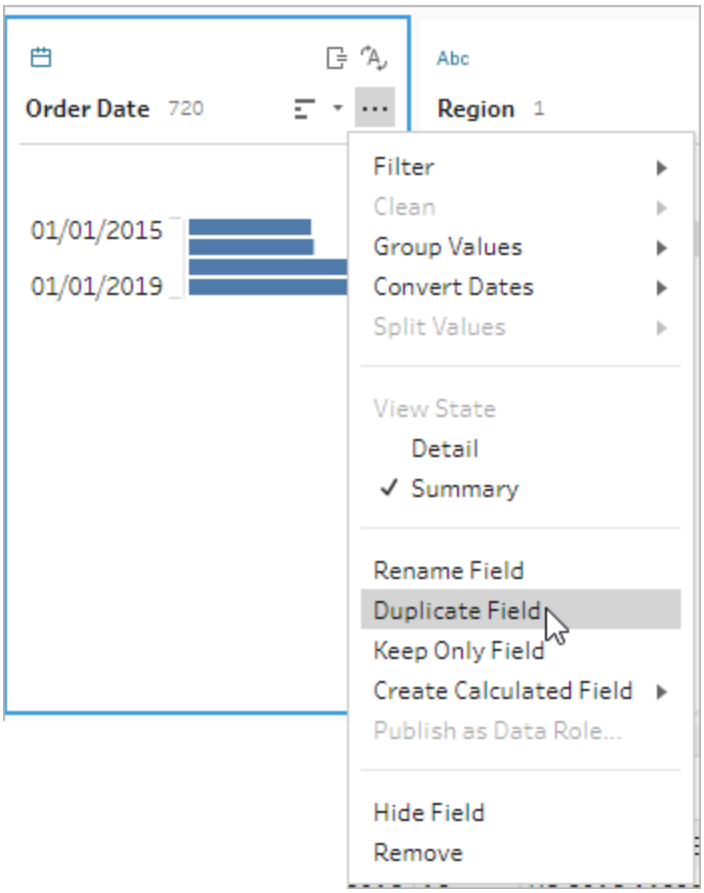
1. 在**“更改”**窗格中，选择要复制的更改。
2. 将更改拖放到要应用更改的字段上。根据需要重复此操作。



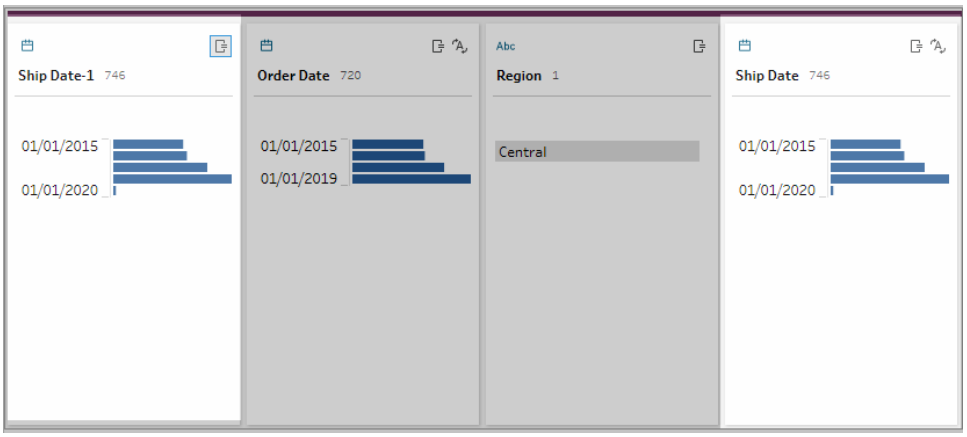
复制字段

从 Tableau Prep Builder 版本 2019.2.3 及更高版本以及在 Web 上，如果想要在一个字段上试验清理操作但不想更改原始数据，则可以复制字段。

1. 在“配置”窗格、数据网格、“结果”窗格或列表视图中，选择要复制的字段。
2. 从**“更多选项”**菜单中，选择**“复制字段”**。



系统将使用相同的名称和修饰符创建一个新字段。例如，“Ship Date -1”(发货日期 -1) 。



创建可重用的流程步骤

在 Tableau Prep Builder 版本 2019.3.2 及更高版本中受支持。

注意：无法在 Web 上创建可重用的流程步骤，但您可以在 Web 流程中使用它们。Web 上尚不支持包含基于文件的输入步骤的可重用步骤。

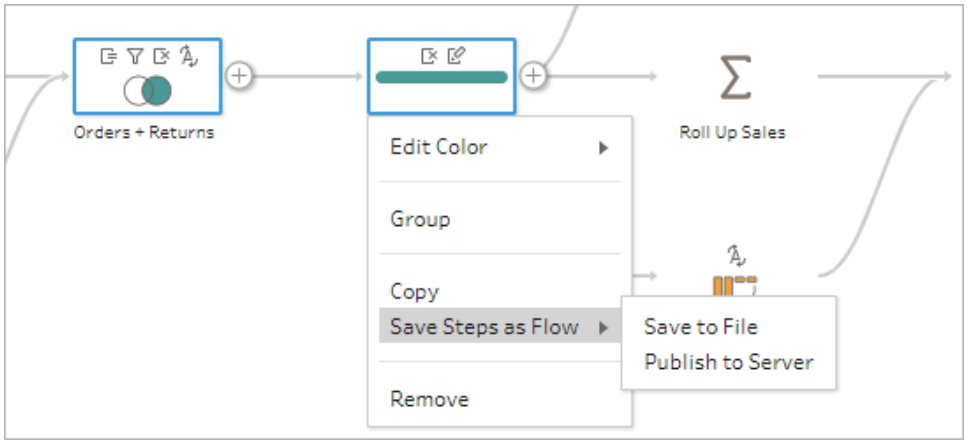
如果您通常对数据一遍又一遍地执行相同的操作，并且想要在其他流程中应用这些相同的步骤，在 Tableau Prep Builder 版本 2019.3.2 及更高版本中，您可以选择一个或多个流程步骤或组及其关联的操作或整个流程，并将其以本地方式保存到计算机上的文件。您也可以将其发布到 Tableau Server 或 Tableau Cloud 以与其他人共享。

将流程步骤发布到服务器时，将自动添加**"已保存的步骤"**标记，以便在将这些步骤添加到流程中时可以轻松地搜索和找到它们。

从版本 2022.1.1 开始，您可以创建包含参数的可重用步骤。保存步骤后，使用参数的**当前值**将参数转换为静态值。有关在流程中使用参数的详细信息，请参见**在流程中创建和使用参数** 在本页 192。

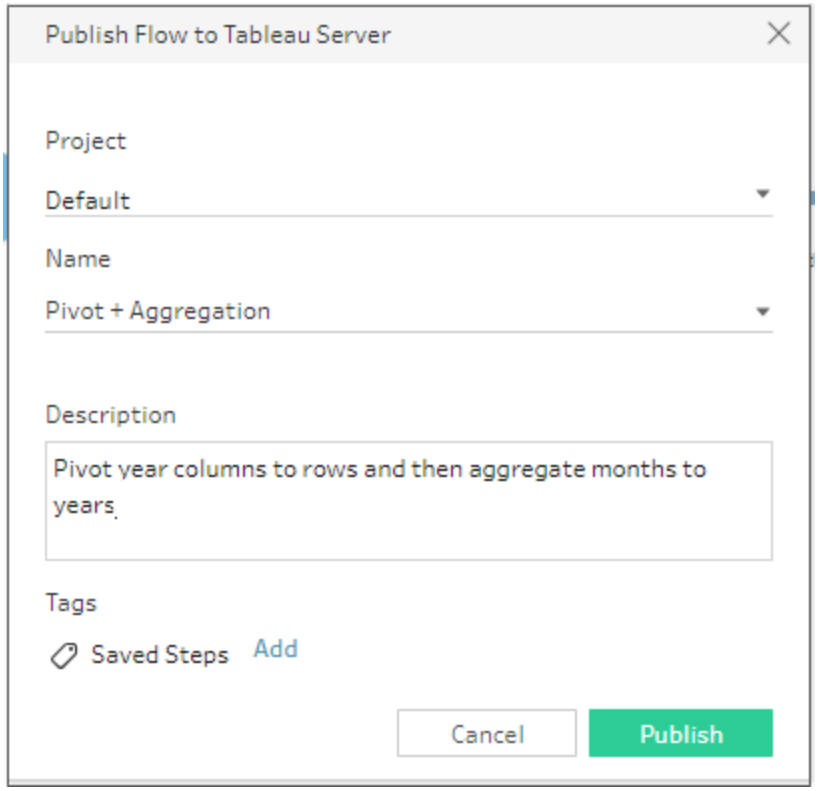
创建可重用的步骤

1. 选择一个或多个步骤。
2. 右键单击或按住 Ctrl 单击 (MacOS) 选定步骤，并选择**"将步骤另存为流程"**。



3. 选择**"保存到文件"**以本地方式保存流程，或选择**"发布到服务器"**以将流程发布到 Tableau Server 或 Tableau Cloud。

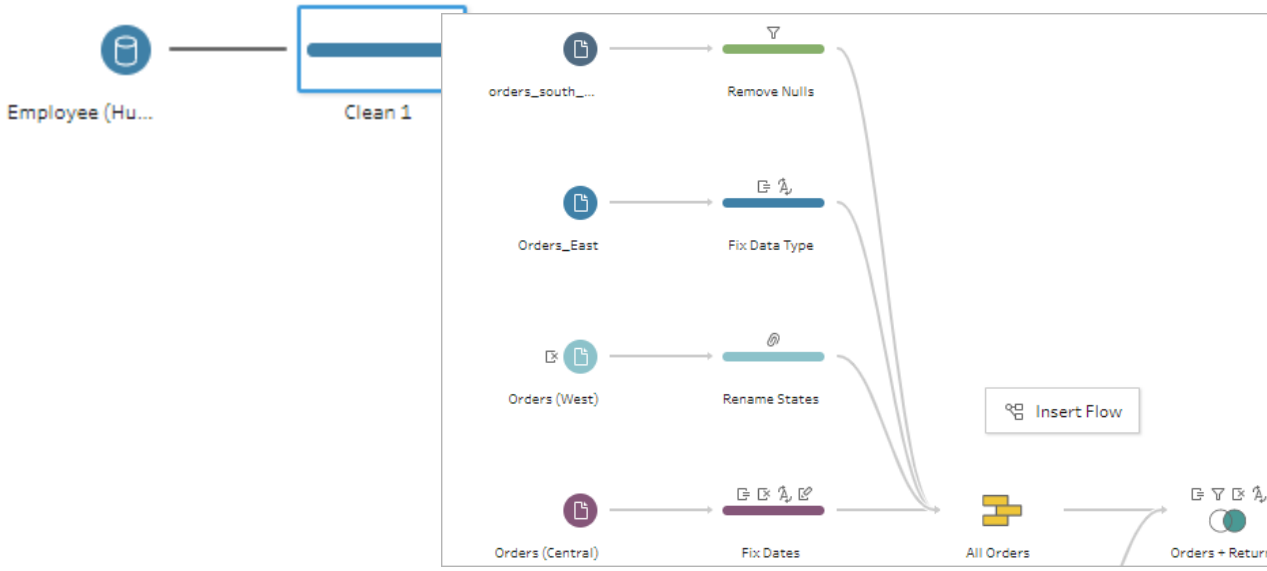
4. 如果将流程发布到 Tableau Server 或 Tableau Cloud，请登录到服务器（如果需要），然后填写**“发布流程”**对话框中的字段，再单击**“发布”**。



在流程中插入可重用的步骤

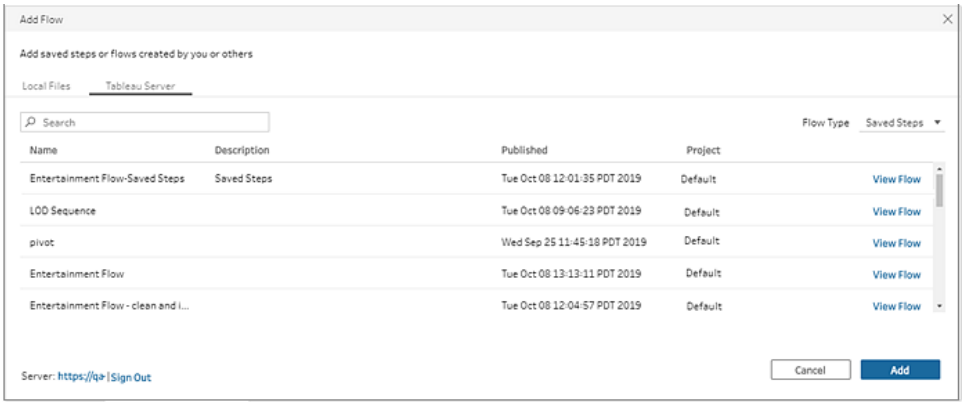
1. 打开一个流程。
2. 在流程窗格中，执行以下操作之一：
   - 将光标悬停在步骤或流程线上，直至出现加号 ⊕ 图标，然后选择**“插入流程”**。
   - 在画布的白色区域中，右键单击或按住 Ctrl 并单击 (MacOS)，然后单击 **“插入流程”**，或者从顶部菜单中单击**“编辑”>“插入流程”**。

“流程步骤”菜单 **“画布”菜单**

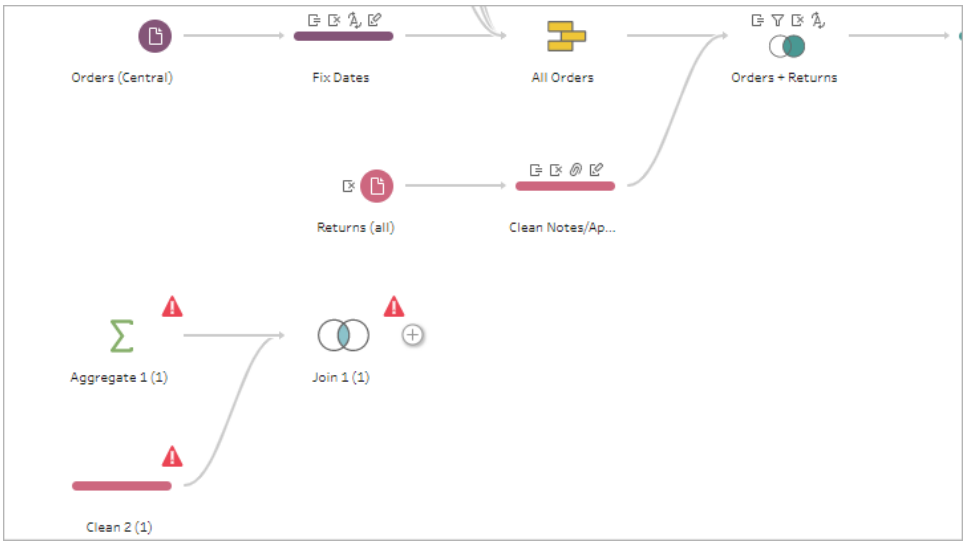


3. 在**“添加流程”**对话框中，从保存到本地文件或服务器的流程中进行选择，然后单击**“添加”**。系统将自动筛选流程列表，以显示带有**“已保存的步骤”**标记的流程。若要插入其他流程，请将**“流程类型”**更改为**“所有流程”**。

   在 Tableau Prep Builder 版本 2019.4.2 及更高版本中以及在 Web 上，可以单击**“查看流程”**以打开并查看已登录服务器中的已发布流程。



4. 流程已添加到流程窗格中。如果将流程添加到流程步骤的末尾，则流程步骤将自动添加到步骤的末尾。如果在现有流程步骤之间插入流程步骤。将步骤移到流程中所需的位置并修复任何错误。



填补顺序数据中的空白

在 Tableau Prep Builder 版本 2021.3.1 及更高版本以及 Web 版 Tableau Server 和 Tableau Cloud 版本 2021.3.0 及更高版本中受支持。

当您的顺序数据集存在空白时，您可能需要用新行填充这些空白，以有效地分析您的数据或执行趋势分析。您可以使用**“新建行”**步骤类型来生成缺失的行，并设置配置选项以获得您需要的结果。

可以为具有数字(整数)或日期值的字段生成新行。配置选项包括：

- 使用来自单个字段或两个字段的值生成行
- 使用字段中的所有数据或选择一个值范围
- 使用结果创建新字段或将新行添加到现有字段
- 设置生成新行时要使用的增量(最多为 10,000)
- 将新行的值设置为零、Null 或复制前一行的值。

示例

- **示例 1**：您有一个销售数据表，但有些日子没有记录销售。您每天都需要一行，而不仅仅是有销售额的那几天。使用**“新建行”**，您可以为缺失的几天生成行，并将它们添加到现有字段“星期几”中。由于这几天没有记录销售，您希望销售数量值为零。

Original Table

Day of the week	Quantity Sold
1	100
3	23
5	75
7	14

Table with missing rows added

Day of the week	Quantity Sold
1	100
2	0
3	23
4	0
5	75
6	0
7	14

- **示例 2**:您有一个销售数据表,其中使用日期范围记录已执行的订单。您每天都需要一行。由于您不知道每天执行了多少订单,因此您希望新行的值为 Null。使用**"新建行"**,您可以生成两个日期之间的缺失行,并创建一个名为"所有日期"的新字段以保留您的原始数据。

Original Table

Start Date	End Date	Orders Filled
1/1/2020	1/3/2020	100
1/4/2020	1/7/2020	75

Table with missing rows added

Start Date	End Date	All Dates	Orders Filled
1/1/2020	1/3/2020	1/1/2020	100
1/1/2020	1/3/2020	1/2/2020	Null
1/1/2020	1/3/2020	1/3/2020	Null
1/4/2020	1/7/2020	1/4/2020	75
1/4/2020	1/7/2020	1/5/2020	Null
1/4/2020	1/7/2020	1/6/2020	Null
1/4/2020	1/7/2020	1/7/2020	Null

生成新行

1. 在**"流程"**窗格中,单击加号 ⊕ 图标,并选择**"新建行"**。**"新建行"**步骤显示在**"流程"**窗格中。

   完成以下步骤来配置您的选项以生成新行。
2. **您想要如何添加新行?** 使用以下选项之一来选择缺少行的一个或多个字段。
   a. **来自一个字段的值**:依据单个字段中的值生成缺失的行。将此选项用于**"数字(整数)"**或**"日期"**数据类型。

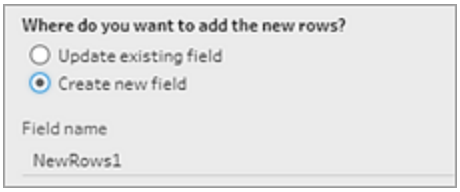
      默认情况下,使用最小值和最大值生成缺失行。此选项使用字段中的所有值。如果您只想使用值范围来生成缺失的行,请设置**"开始值"**和**"结束值"**。

      注意:**"起始值"**和**"结束值"**字段不能用于生成当前数据集之外的行。

b. **两个字段的值范围**：使用两个日期字段之间的值范围生成新行。此选项仅适用于**“日期”**数据类型，使用字段中的所有值，并要求两个字段具有相同的数据类型。

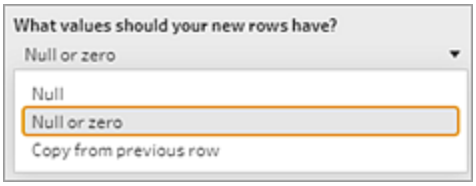
3. **您想在哪里添加新行？** 使用单个字段时，您可以将新行添加到现有字段，或创建新字段以保留原始数据。使用来自两个字段的值范围时，您必须创建一个新字段。

   - **字段名称**：输入新字段的名称。

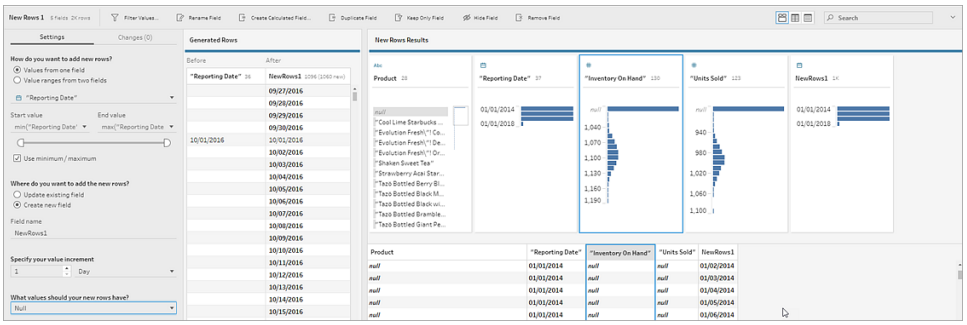


4. **指定您的增量值**：输入 1-10,000 之间的值。每个新行都会按您选择的值递增。如果您选择的值大于值之间的差距，则不会生成新行。
   - **数字字段**：选择一个数字值。
   - **日期字段**：选择一个数字值，并选择**“日”**、**“周”**或**“月”**。

5. **您的新行应该有什么值？** ：选择一个选项以填写新行的其他字段值。
   - **Null**：用 Null 填充所有字段值。
   - **Null 或零**：用 Null 填充所有文本值，用零填充所有数字值。
   - **从前一行复制**：使用前一行的值填充所有字段值。



当您输入配置设置时，新行以**粗体**显示在**“生成的行”**窗格中。行详细信息显示在**“新行结果”**窗格中。



创建详细级别、排名和分片计算

注意：从版本 2020.4.1 开始，您现在可以在 Tableau Server 和 Tableau Cloud 中创建和编辑流程。除非特别指出，本主题中的内容适用于所有平台。有关在 Web 上制作流程的详细信息，请参见 Tableau Server 和 Tableau Cloud 帮助中的 **Web 上的 Tableau Prep**。

可以使用计算字段，通过数据源中已存在的数据来创建新数据。Tableau Prep 支持的许多计算类型与 Tableau Desktop 相同。有关创建计算的一般信息，请参见 Tableau 中的计算入门指南。

注意：创建计算字段时，不要使用保留名称“Number of Records”(记录数)。使用此保留名称将导致与权限相关的错误。

从 Tableau Prep Builder 版本 2020.1.3 开始以及在 Web 上，您可以使用 FIXED 详细级别 (LOD) 以及 RANK and ROW_NUMBER 分析函数来执行更复杂的计算。

例如，添加 FIXED LOD 计算来更改表中字段的粒度、使用 ROW_NUMBER() 分析函数快速查找重复行，或使用 RANK() 函数之一来查找具有类似数据的所选行的前 N 个或后 N 个值。如果希望在构建这些类型的表达式时获得更多指引，您可以使用可视化计算编辑器。

从版本 2021.4.1 Tableau Prep Builder 开始以及在 Web 上，您可以使用分片功能将行分布到指定数量的存储桶中。

注意：Tableau Desktop 中支持的某些函数在 Tableau Prep 中可能不受支持。若要查看可用于 Tableau Prep 的函数，请查看计算编辑器中的函数列表。

计算详细级别

当您需要在同一表中以多个粒度级别计算数据时，可以编写详细级别 (LOD) 表达式来执行此操作。举例来说，如果要查找每个区域的总销售额，可以编写类似于 `{FIXED [Region] : SUM([Sales])}` 的计算。

Tableau Prep 支持 FIXED 详细级别表达式，并使用语法 `{FIXED [Field1],[Field2] : Aggregation([Field)}`。

LOD 表达式有两个用冒号分隔的等式部分。

- **FIXED [字段]**(必需):这是您要为其计算值的一个或多个字段。举例来说,如果要查找客户和区域的总销售额,请输入 `FIXED [Customer ID], [Region]:`。如果未选择任何字段,这相当于执行在冒号右侧定义的聚合,并为每一行重复该值。
- **聚合([字段])**(必需):选择要计算的内容以及所需的聚合级别。举例来说,如果要查找总销售额,请输入 `SUM([Sales]`。

在 Tableau Prep 中使用此功能时,以下要求适用:

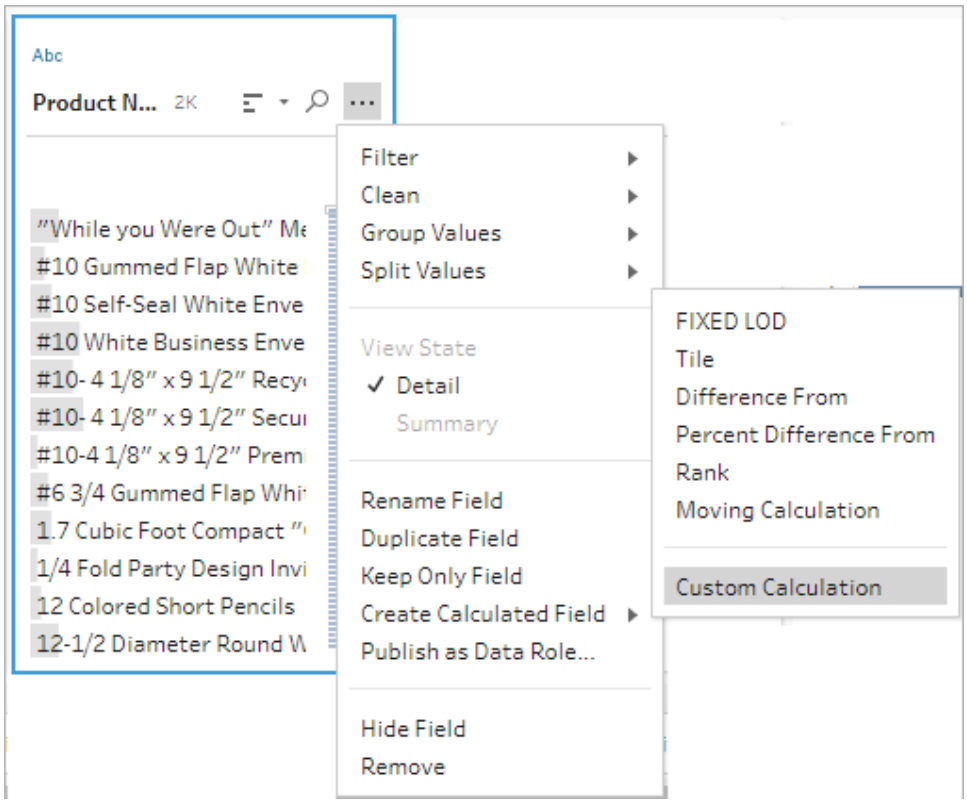
- 不支持 `INCLUDE` 和 `EXCLUDE` LOD 表达式。
- 仅支持在 LOD 表达式内部使用聚合计算。例如,`SUM([Sales])` 将无效,但 `{FIXED [Region] : SUM([Sales])}` 有效。
- 不支持在 LOD 表达式内嵌套表达式。例如,不支持 `{ FIXED [Region] : AVG( [Sales] ) / SUM( [Profit] )}`。
- 不支持将 LOD 表达式与其他表达式结合使用。例如,`[Sales]/{ FIXED [Country / Region]:SUM([Sales])}` 不受支持。

创建详细级别 (LOD) 计算

若要创建详细级别计算,可以使用计算编辑器自行编写计算,或者,如果您希望获得更多指引,则可以使用可视化计算编辑器,您可以在其中选择字段,Tableau Prep 将为您编写计算。

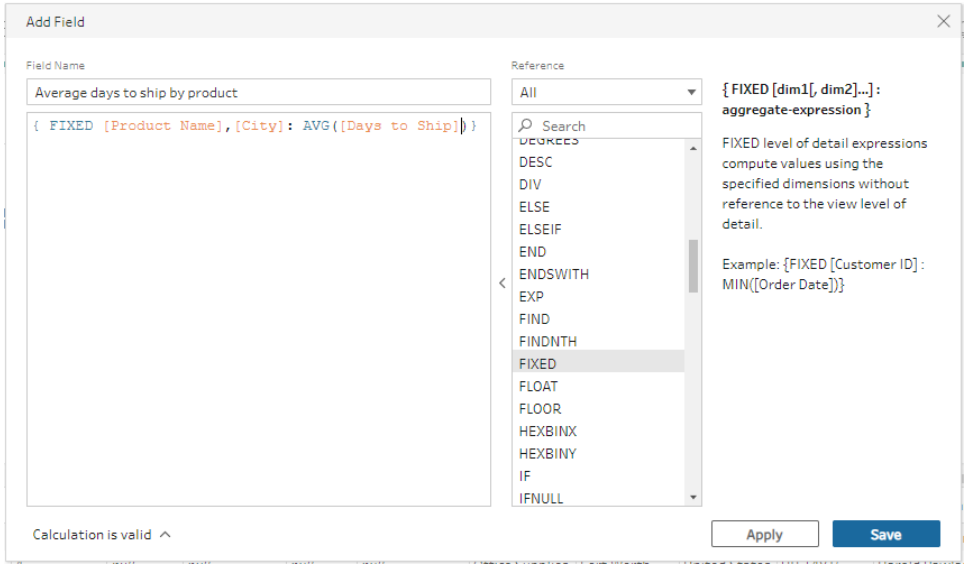
计算编辑器

1. 在"配置"窗格工具栏中单击**"创建计算字段"**,或在配置卡或数据网格中单击**"更多选项"** ··· 菜单,并选择**"创建计算字段">"自定义计算"**。



2. 在计算编辑器中，输入计算的名称，并输入表达式。

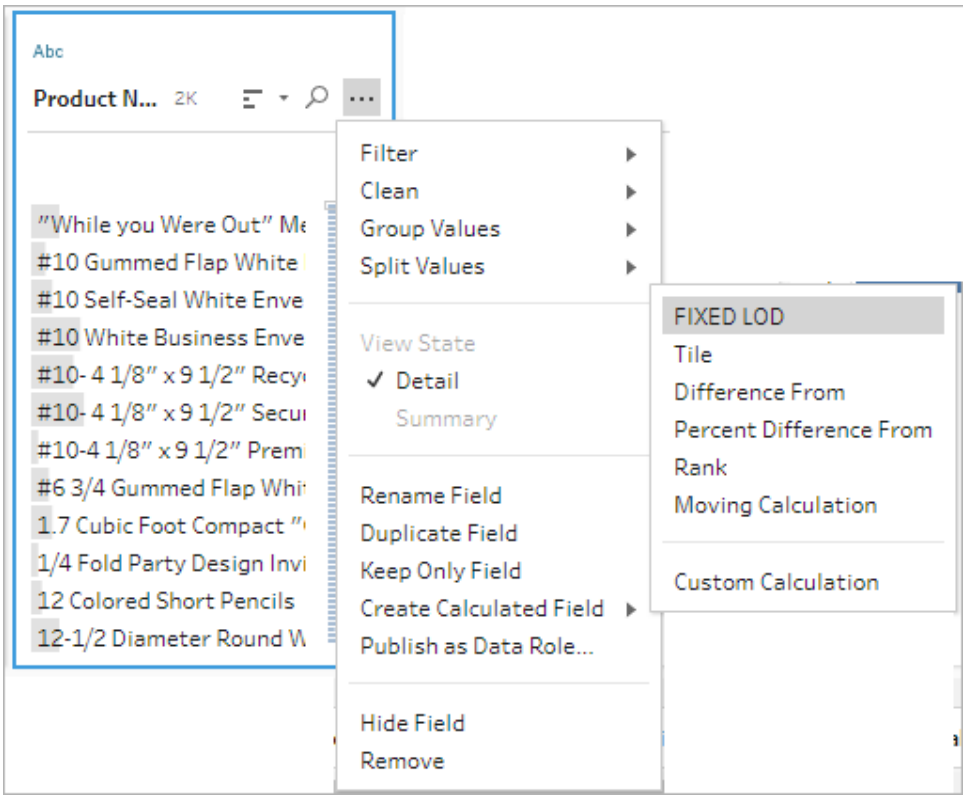
   例如，若要按城市查找产品平均发货天数，请创建如下所示的计算。



可视化计算编辑器

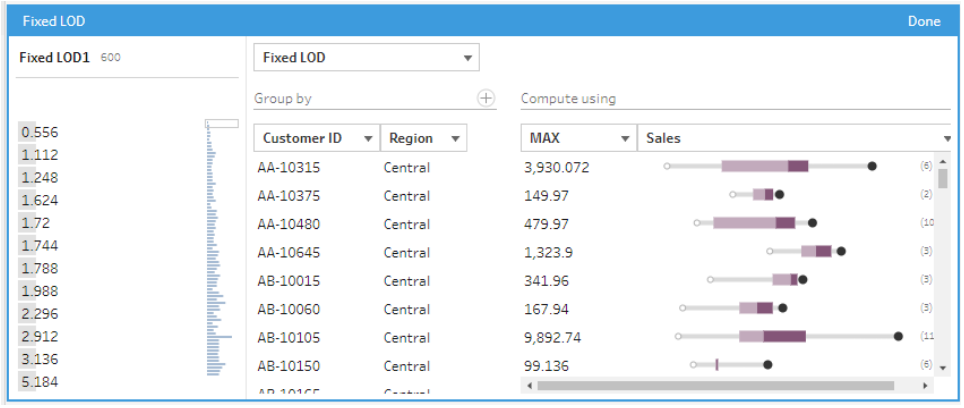
从列表中选择字段，Tableau Prep 将在您进行选择时为您构建计算。结果的预览显示在左侧窗格中，这样您就能够随时查看所选内容的结果。

1. 在配置卡或结果窗格中，单击**“更多选项”**… 菜单并选择**“创建计算字段”>“FIXED LOD”**。



2. 在可视化计算编辑器中，执行以下操作：
   - 在**“分组依据”**部分，选择要为其计算值的字段。默认情况下，将添加您在其中选择了**“创建计算字段”>“FIXED LOD”**菜单选项的字段。单击加号 ⊕ 图标以向计算中添加任何其他字段。这将填充等式的左侧，即 `{FIXED [Field1],[Field2] :`。
   - 在**“计算依据”**部分，选择要用于计算新值的字段。然后选择您的聚合。这将填充等式的右侧，即 `Aggregation([Field])}`。

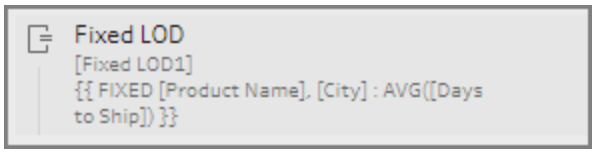
字段下方的图形显示值的分布和每个值组合的总计数。根据数据类型，这可以是盒形图、值范围或实际值。



注意：可用聚合值因分配给字段的数据类型而异。

- 若要移除字段，请在**“分组依据”**部分的字段下拉框中右键单击或按住 Ctrl 单击 (MacOS)，然后选择**“移除字段”**。
- 在左侧窗格中，双击字段标题并输入用于计算的名称。

3. 单击**“完成”**以添加新的计算字段。在**“更改”**窗格中，您可以看到 Tableau Prep 生成的计算。单击**“编辑”**打开可视化计算编辑器以进行任何更改。



计算排名或行号

分析函数（有时称为窗口计算）使您能够对整个表或数据集中的所选行（分区）执行计算。例如，在将排名应用于所选行时，可以使用以下计算语法：

```
{PARTITION [field]: {ORDERBY [field]: RANK() }}
```

- **PARTITION**（可选）：指定要对其执行计算的行。可以指定多个字段，但如果要使用整个表，请省略函数的此部分，Tableau Prep 会将所有行视为分区。例如，`{ORDERBY [Sales] : RANK() }`。
- **ORDERBY**（必需）：指定要用于生成排名序列的一个或多个字段。

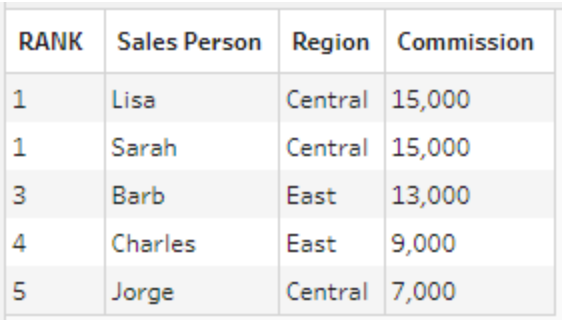
- **Rank ()**(必需) : 指定要计算的排名类型或 ROW_NUMBER ()。Tableau Prep 支持 rank ()、RANK_DENSE()、RANK_MODIFIED()、RANK_PERCENTILE() 和 ROW_NUMBER() 函数。
- **DESC** 或 **ASC**(可选) : 表示降序 (DESC) 或升序 (ASC) 顺序。默认情况下，排名按降序排序，因此无需在表达式中指定此项。如果要更改排序顺序，请向表达式中添加 **ASC**。

  也可以在函数中同时包括这两个选项。举例来说，如果要对所选行进行排名，但希望按升序对行进行排序，然后按降序应用排名，则应在表达式中包括这两个选项。例如：`{PARTITION [Country], [State]: {ORDERBY [Sales] ASC, [Customer Name] DESC: RANK() }}`

使用此功能时，以下要求适用：

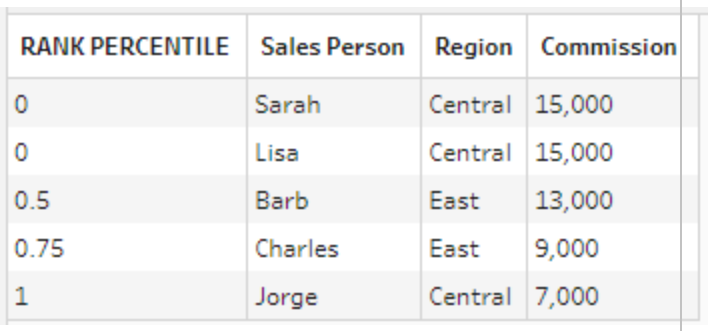
- 不支持在 RANK () 函数中嵌套表达式。例如，不支持 `[Sales]/{PARTITION [Country]: {ORDERBY [Sales]: RANK() }} / SUM( [Profit] )}`。
- 不支持将 RANK () 函数与其他表达式结合使用。例如，`[Sales]/{PARTITION [Country]: {ORDERBY [Sales]: RANK() }}` 不受支持。

支持的分析函数

函数	描述	结果
RANK()	从 1 开始按升序或降序为每一行分配整数排名。如果行具有相同的值，则它们共享分配给该值的第一个实例的排名。在计算下一行的排名时，将添加具有相同排名的行数，因此您可能无法获得连续排名值。 **示例计算**： `{ORDERBY`	

RANK	Sales Person	Region	Commission
1	Lisa	Central	15,000
1	Sarah	Central	15,000
3	Barb	East	13,000
4	Charles	East	9,000
5	Jorge	Central	7,000

<table>
<tr><td></td><td>[Commission] DESC: RANK()}</td><td></td></tr>
<tr><td>RANK_ DENSE()</td><td>从 1 开始按升序或降序为每一行分配整数排名。如果行具有相同的值，则它们共享分配给该值的第一个实例的排名，但不会跳过任何排名值，因此您会看到连续排名值。
示例计算：
{ORDERBY [Commission] DESC: RANK_ DENSE()}</td><td><table><tr><th>DENSE RANK</th><th>Sales Person</th><th>Region</th><th>Commission</th></tr><tr><td>1</td><td>Lisa</td><td>Central</td><td>15,000</td></tr><tr><td>1</td><td>Sarah</td><td>Central</td><td>15,000</td></tr><tr><td>2</td><td>Barb</td><td>East</td><td>13,000</td></tr><tr><td>3</td><td>Charles</td><td>East</td><td>9,000</td></tr><tr><td>4</td><td>Jorge</td><td>Central</td><td>7,000</td></tr></table></td></tr>
<tr><td>RANK_ MODIFIED()</td><td>从 1 开始按升序或降序为每一行分配整数排名。如果行具有相同的值，则它们共享分配给该值的最后一个实例的排名。Rank_ Modified 的计算方式为 Rank + (Rank + Number of duplicate rows - 1)。
示例计算：
{ORDERBY</td><td><table><tr><th>RANK MODIFIED</th><th>Sales Person</th><th>Region</th><th>Commission</th></tr><tr><td>1</td><td>Sarah</td><td>Central</td><td>15,000</td></tr><tr><td>1</td><td>Lisa</td><td>Central</td><td>15,000</td></tr><tr><td>3</td><td>Barb</td><td>East</td><td>13,000</td></tr><tr><td>4</td><td>Charles</td><td>East</td><td>9,000</td></tr><tr><td>5</td><td>Jorge</td><td>Central</td><td>7,000</td></tr></table></td></tr>
</table>

	`[Commission] DESC: RANK_ MODIFIED()}`	
RANK_ PERCENTILE ()	从 0 到 1 按升序或降序为每一行分配百分比排名。RANK_ PERCENTILE 的计算方式为 `(Rank-1)/ (Total rows-1)`。 **示例计算**： `{ORDERBY [Commission] DESC: RANK_ PERCENTILE()}` **注意：**如果出现平局，Tableau Prep 会向下舍队，类似于 SQL 中的 PERCENT_ RANK()。	

RANK PERCENTILE	Sales Person	Region	Commission
0	Sarah	Central	15,000
0	Lisa	Central	15,000
0.5	Barb	East	13,000
0.75	Charles	East	9,000
1	Jorge	Central	7,000

ROW_NUMBER()	为每个唯一行分配连续的行 ID。不会跳过任何行号值。如果具有重复行并使用此计算，则每次运行流程时，如果行的顺序发生更改，结果可能会发生变化。 **示例计算**： `{ORDERBY [Commission] DESC: ROW_NUMBER()}`	

ROW NUMBER	Sales Person	Region	Commission
1	Sarah	Central	15,000
2	Lisa	Central	15,000
3	Barb	East	13,000
4	Charles	East	9,000
5	Jorge	Central	7,000

以下示例显示了应用于同一数据集的上述每个函数的比较。

RANK	DENSE RANK	RANK MODIFIED	RANK PERCENTILE	ROW NUMBER	Sales Person	Region	Commission
1	1	1	0	1	Lisa	Central	15,000
1	1	1	0	2	Sarah	Central	15,000
3	2	3	0.5	3	Barb	East	13,000
4	3	4	0.75	4	Charles	East	9,000
5	4	5	1	5	Jorge	Central	7,000

创建排名或行号计算

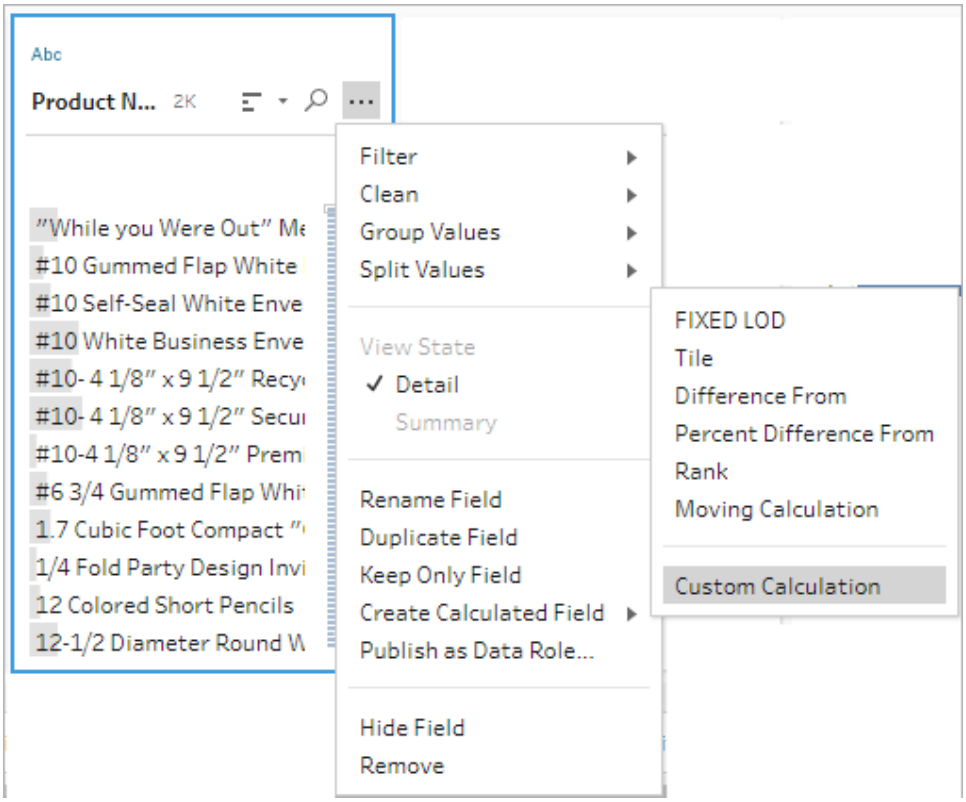
若要创建排名或行号计算，可以使用计算编辑器自行编写计算，或者，如果您希望获得更多指引，则可以使用可视化计算编辑器，您可以在其中选择字段，Tableau Prep 将为您编写计算。

注意：RROW_NUMBER() 计算在可视化计算编辑器中不可用。

计算编辑器

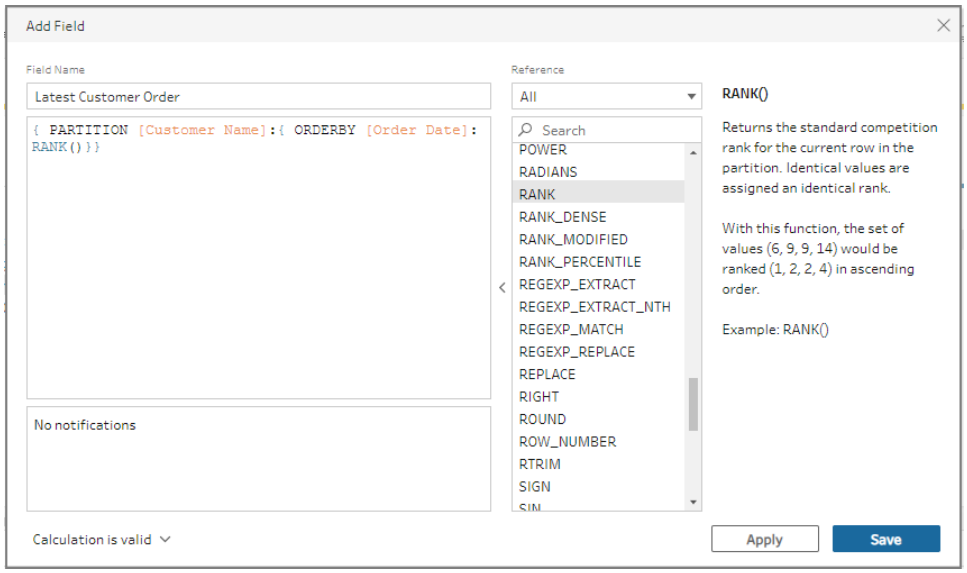
使用计算编辑器创建任何支持的 RANK() 或 ROW_NUMBER() 计算。支持的分析计算列表显示在计算编辑器中**“分析”**下的**“参考”**下拉列表中。

1. 在“配置”窗格工具栏中单击**“创建计算字段”**，或在配置卡或数据网格中单击**“更多选项”** ··· 菜单，并选择**“创建计算字段”>“自定义计算”**。



2. 在计算编辑器中，输入计算的名称，并输入表达式。

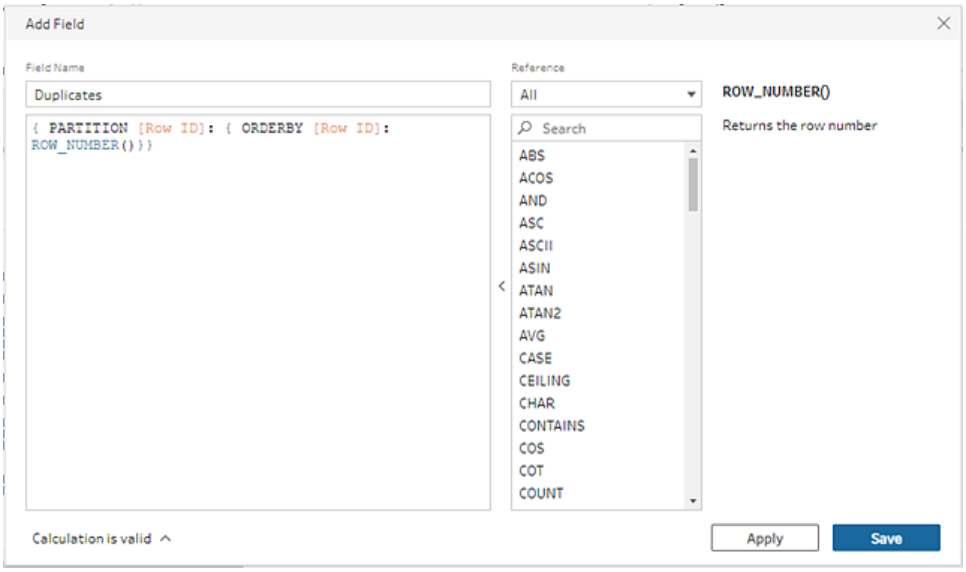
   例如，要查找最新的客户订单，请创建如下所示的计算，然后仅保留使用数字 **1** 排名的客户订单行。



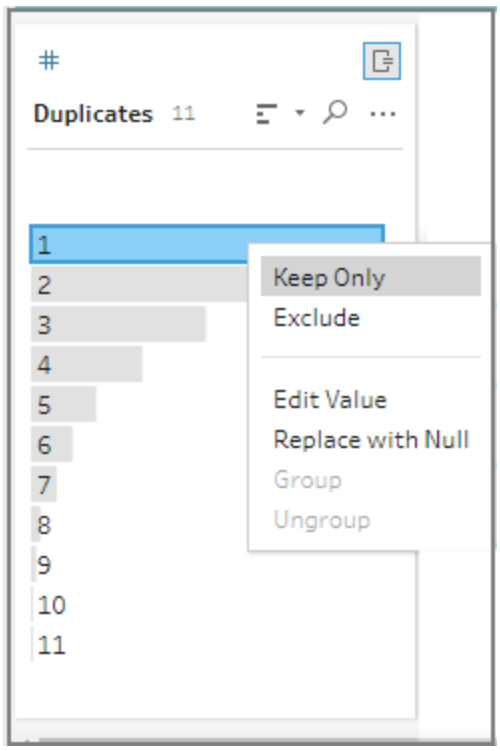
示例：使用 ROW_NUMBER 查找和移除重复值。

本示例使用 Tableau Prep Builder 中的 Superstore 示例数据集，使用 ROW_NUMBER 函数查找和移除字段**“Row ID”(行 ID)** 的确切重复值。

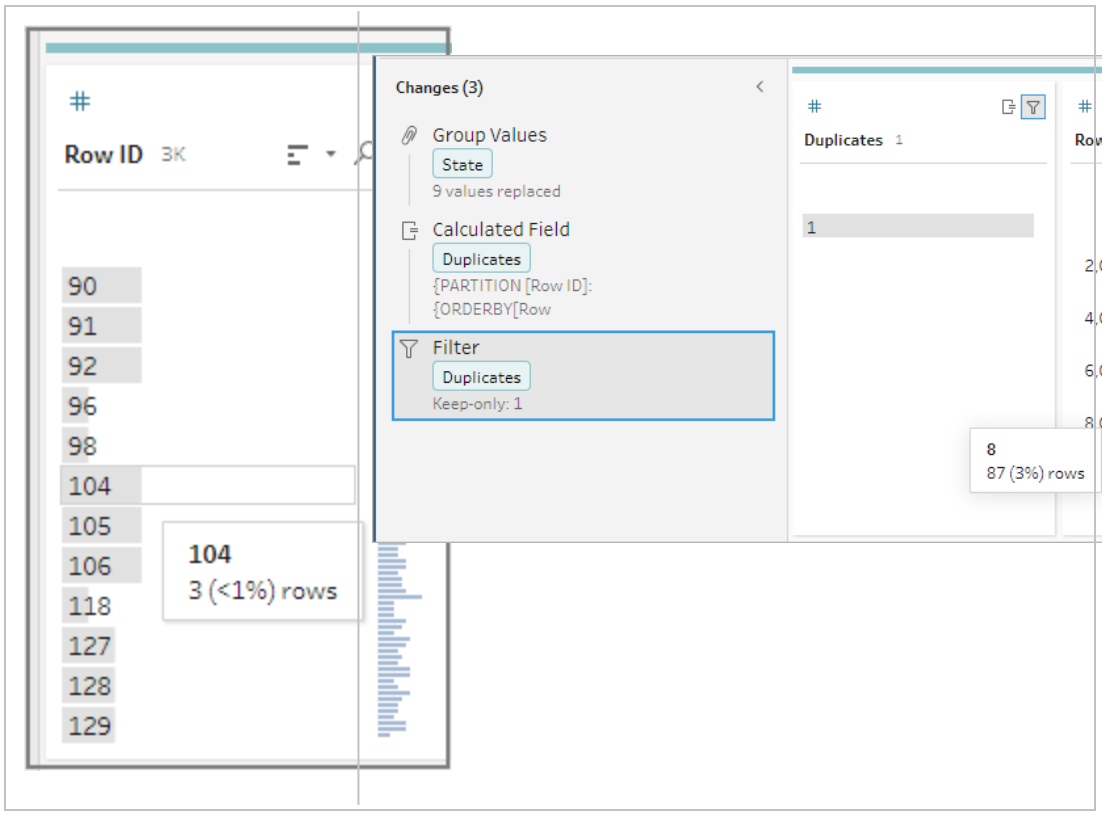
1. 打开**“Sample Superstore”(示例超市)**流程。
2. 在“流程”窗格中，对于“输入”步骤**“Orders West”(对西部排序)**，单击“清理”步骤**“Rename States”(重命名状态)**。
3. 在工具栏中，单击**“创建计算字段”**。
4. 在计算编辑器中，将新字段命名为“Duplicates”(重复项)，并使用 `ROW_NUMBER` 函数，通过 `{PARTITION [Row ID]: {ORDERBY[Row ID]:ROW_NUMBER()}}` 表达式向字段**“Row Id”(行 ID)** 中添加行号，然后单击**“保存”**。



5. 在新的计算字段中，在字段值**“1”**上右键单击或按住 Cmd 单击 (MacOS)，然后从菜单中选择**“只保留”**。



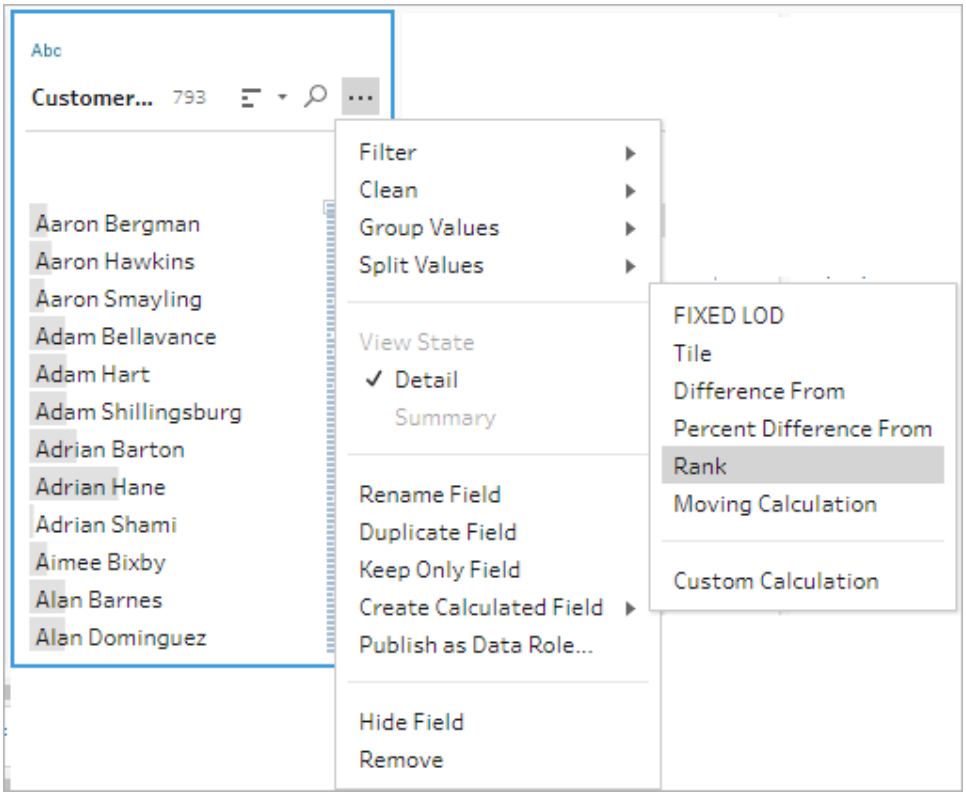
之前	之后



可视化计算编辑器

就像创建详细级别计算时一样，您可以使用可视化计算编辑器来生成排名计算。选择要在计算中包括的字段，然后选择要用于对行进行排名的字段，以及要计算的排名类型。结果的预览显示在左侧窗格中，这样您就能够随时查看所选内容的结果。

1. 在配置卡或结果窗格中，单击**“更多选项”**… 菜单并选择**“创建计算字段”>“排名”**。



2. 在可视化计算编辑器中，执行以下操作：

   - 在**“分组依据”**部分，选择包含要为其计算值的行的字段。这将创建计算的 Partition 部分。

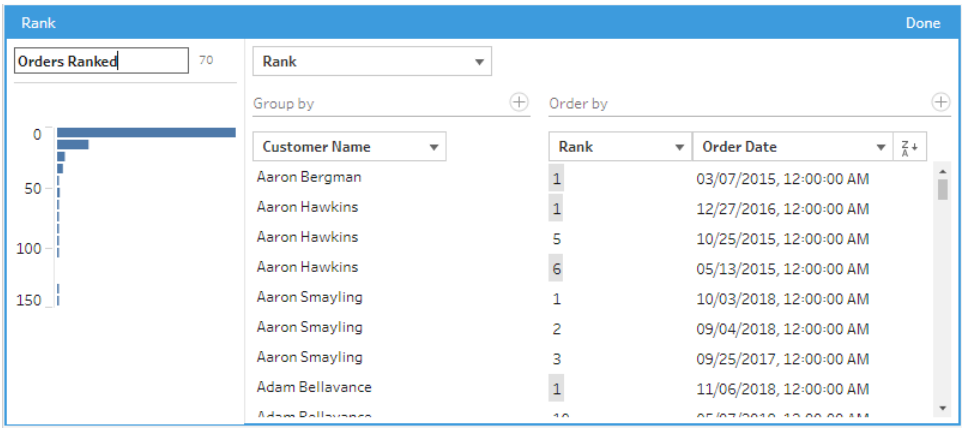
     选择第一个字段后，单击加号 ⊕ 图标以将任何其他字段添加到您的计算中。如果要包括所有行或移除所选行，请在**“分组依据”**部分的字段下拉框中右键单击或按住 Cmd 单击 (MacOS)，然后选择**“移除字段”**。

   - 在**“排序依据”**部分，选择要用于对新值进行排名的字段。默认情况下，将添加您在其中选择了**“创建计算字段”>“排名”**菜单选项的字段。

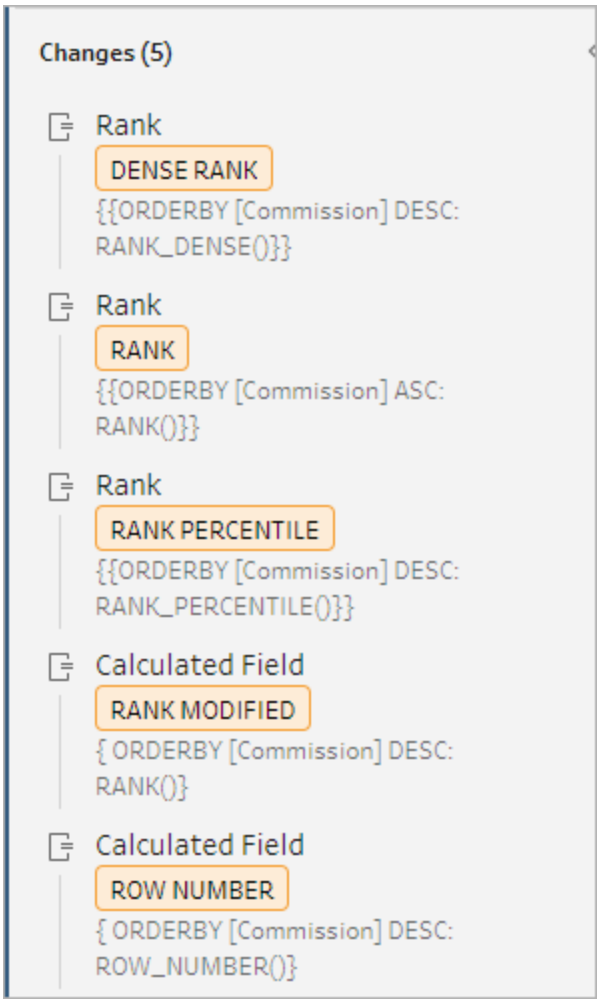
     单击加号 ⊕ 图标以向计算中添加任何其他字段，然后选择**“排名”**类型。单击排序 图标将排名顺序从降序 (DESC) 更改为升序 (ASC)。

> **注意**：排名值因分配给字段的数据类型而异。

- 在左侧窗格中，双击字段标题并输入用于计算的名称。



3. 单击**“完成”**以添加新的计算字段。在**“更改”**窗格中，您可以看到 Tableau Prep Builder 生成的计算。单击**“编辑”**打开可视化计算编辑器以进行任何更改。



计算分片

通过创建计算字段，使用“分片”功能将行分布到指定数量的存储桶中。您选择要作为分布依据的字段，以及要使用的组(分片)的数量。您还可以选择其他字段来创建分区，其中分片的行分布在组中。使用计算编辑器手动输入语法，或使用可视化计算编辑器选择字段，Tableau Prep 会为您编写计算。

举例来说，如果您有几行学生数据并想查看哪些学生在前 50% 和后 50% 中，您可以将数据分组为两个分片。

Student	Subject	Grade
George	English	65
Christina	Science	80
Brian	English	89
Jennifer	Math	55
Daniel	Math	70
Ian	Science	60
Amanda	English	90
Heather	Math	65
Frank	Science	70

以下示例显示了学生成绩上半部分和下半部分的两个组。此方法的语法是:

```
{ORDERBY [Grade] DESC:NTILE(2)}
```

Calculation1	Student	Subject	Grade
1	Amanda	English	90
1	Brian	English	89
1	Christina	Science	80
1	Daniel	Math	70
1	Frank	Science	70
2	George	English	65
2	Heather	Math	65
2	Ian	Science	60
2	Jennifer	Math	55

您还可以创建一个分区,其中字段的每个值都是一个单独的分区,并将数据划分为每个分区的组。

以下示例显示为“Subject”(科目)字段创建分区。为每个科目创建一个分区,并为“Grade”(成绩)字段创建两个组(分片)。然后将行均匀地分配到三个分区的两个组中。此方法的语法是:

```
{PARTITION [Subject]:{ORDERBY [Grade] DESC:NTILE(2)}}
```

Calculation1	Student	Subject	Grade
1	Daniel	Math	70
1	Heather	Math	65
2	Jennifer	Math	55
1	Christina	Science	80
1	Frank	Science	70
2	Ian	Science	60
1	Amanda	English	90
1	Brian	English	89
2	George	English	65

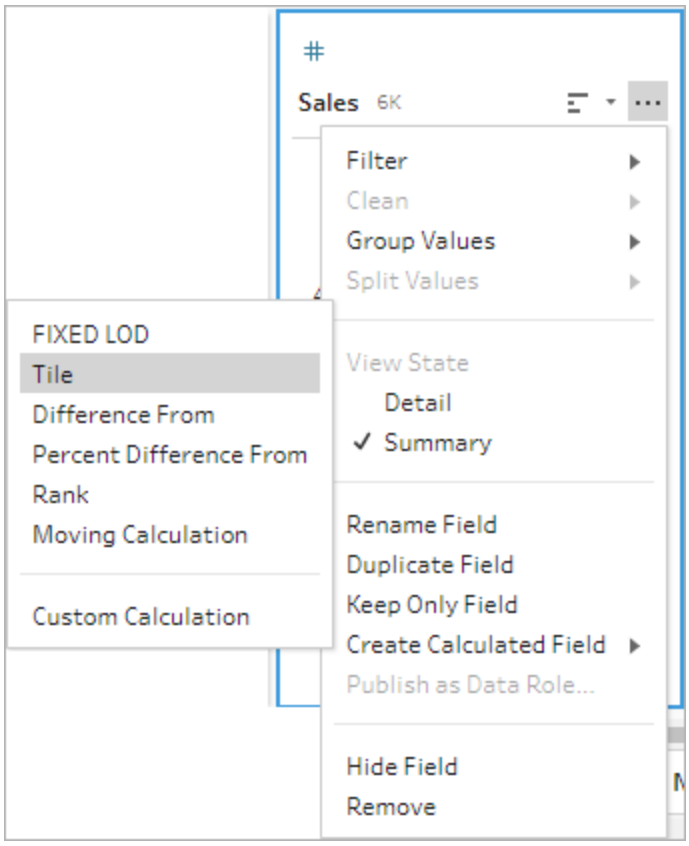
创建分片计算

若要创建分片计算，可以使用计算编辑器自行编写计算，或者，如果您希望获得更多指引，则可以使用可视化计算编辑器，您可以在其中选择字段，Tableau Prep 将为您编写计算。

可视化计算编辑器

当您使用可视化计算编辑器创建分片计算时，结果的预览显示在左侧窗格中。

1. 选择配置卡以创建分片计算。
2. 单击**“更多”**选项菜单，并选择**“创建计算字段”>“分片”**。



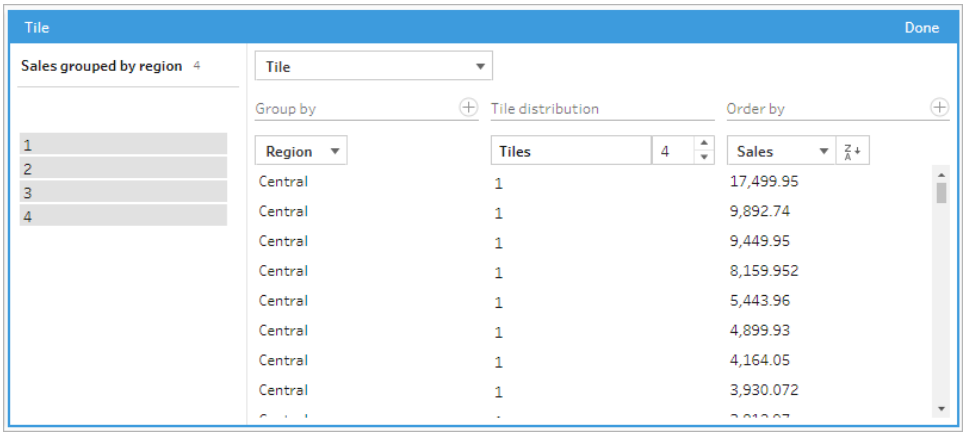
所选的配置卡以 ORDERBY 字段形式添加。

3. 在可视化计算编辑器中，执行以下操作：

   - 选择所需的分片分组数。“分片”的默认值为 1。
   - 在**“分组依据”**部分，选择要针对其各个行计算值的字段。这将创建计算的 PARTITION 部分。对于单个计算，您可以有多个“分组依据”字段。

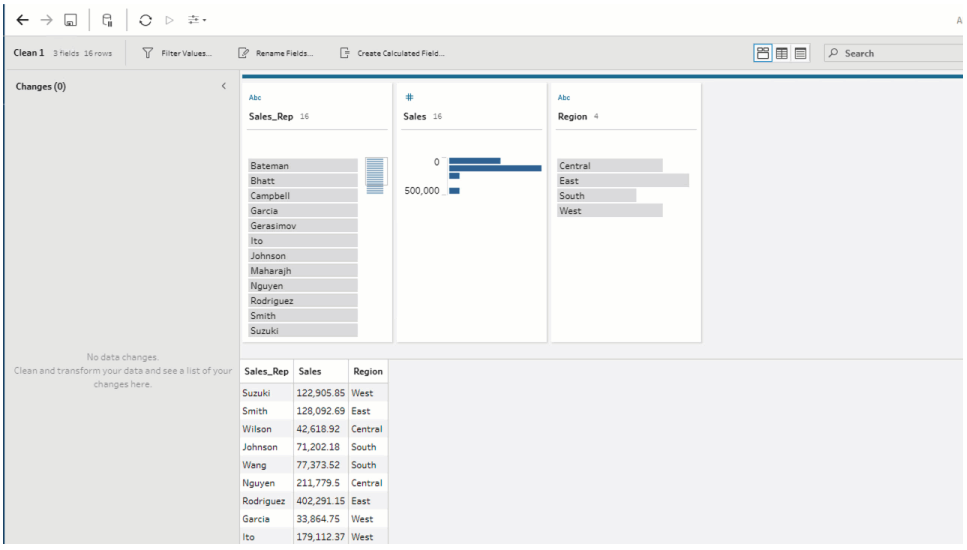
     单击加号 ⊕ 图标以向计算中添加任何其他字段。如果要包括所有行或移除所选行，请在**“分组依据”**部分的字段下拉框中右键单击或按住 Cmd 单击 (MacOS)，然后选择**“移除字段”**。
   - 在左侧窗格中，双击字段标题并输入用于计算的名称。
   - 在**“排序依据”**部分，选择要用于分组和分布新值的字段。您必须至少有一个“排序依据”字段。默认情况下，将添加您在其中选择了**“创建计算字段”>“分**

片”菜单选项的字段。



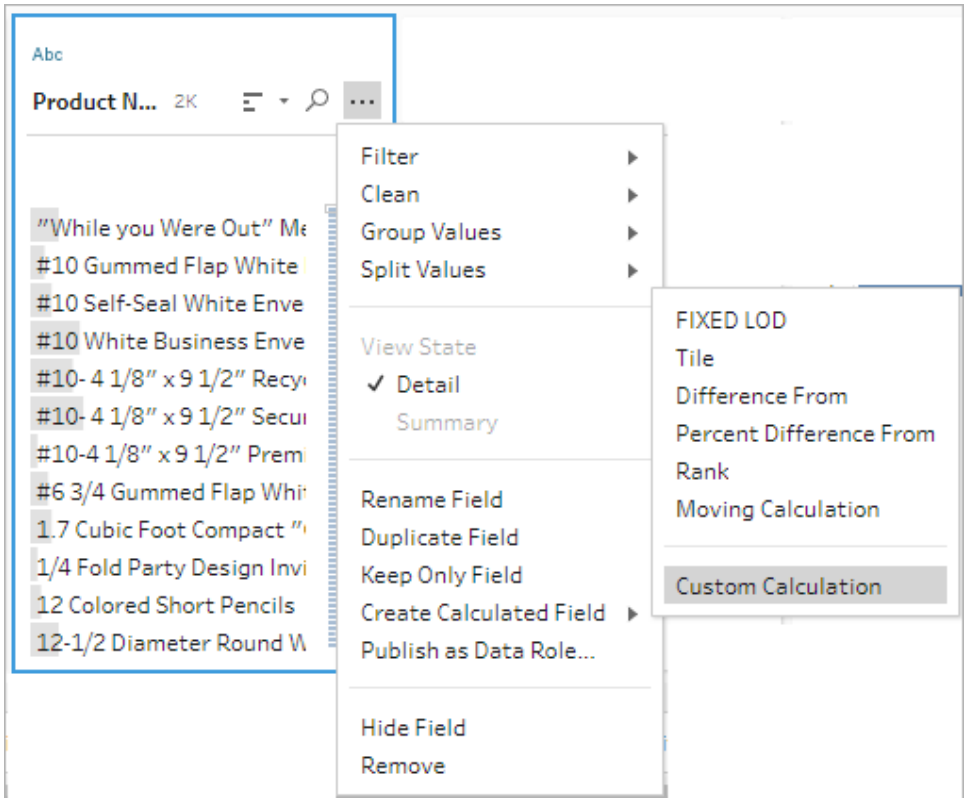
4. 若要对结果进行排序，请执行以下操作：
   - 单击任何计算行以筛选所选分组的结果
   - 更改排序依据字段的升序或降序顺序。
5. 单击**“完成”**以添加新的计算字段。
6. 在**“更改”**窗格中，您可以看到 Tableau Prep Builder 生成的计算。单击**“编辑”**打开可视化计算编辑器以进行任何更改。

   以下示例显示了行的四分位数划分。基于四个美国地区创建一个分区，然后将“Sales”(销售额)字段数据均匀地分组到这些分区中。

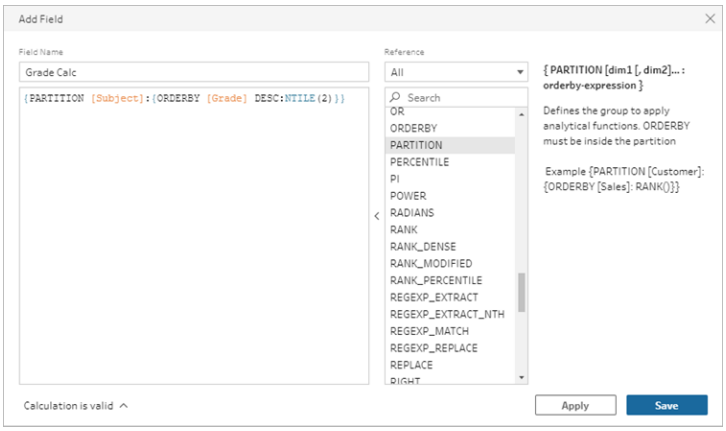


计算编辑器

1. 在“配置”窗格工具栏中单击“创建计算字段”，或在配置卡或数据网格中单击**“更多选项”**… 菜单，并选择**“创建计算字段”>“自定义计算”**。



2. 在计算编辑器中，输入计算的名称，并输入表达式。例如，若要按成绩将学生行分为两组，然后按科目将它们分组，请使用：`{PARTITION [Subject]:{ORDERBY [Grade] DESC:NTILE(2)}}`。



分片计算包括以下元素：

- **PARTITION**（可选）：partition 子句将结果集的行划分为使用了 NTILE() 函数的分区。
- **ORDERBY**（必需）：ORDER BY 子句定义使用 NTILE() 的每个分区中的行分布。
- **NTILE**（必需）：NTILE 是将行划分为的整数。

  > **注意**：当所有行都可以被 NTILE 子句划分时，该功能会在一定数量的分区中均匀划分各行。当行数不能被 NTILE 子句划分时，生成的组将划分成不同大小的数据桶。

- **DESC** 或 **ASC**（可选）：表示降序 (DESC) 或升序 (ASC) 顺序。默认情况下，分片按降序排序，因此无需在表达式中指定此项。如果要更改排序顺序，请向表达式中添加 **ASC**。

3. 单击**“保存”**。

   生成的字段显示与表中每一行相关联的分片分组（数据桶）分配。

计算多行的值

> **注意**：从版本 2020.4.1 开始，您可以在 Tableau Server 和 Tableau Cloud 中创建和编辑流程。除非另有说明，否则本主题中的内容适用于所有平台。有关在 Web 上制作流程的详细信息，请参见 Tableau Server 和 Tableau Cloud 帮助中的 **Web 上的 Tableau Prep**。

多行计算可让您计算流程中多行数据之间的值。虽然与 Tableau 中的表计算类似，但在运行流程时，多行计算适用于整个数据集。您还可以使用其他类型的计算来构建结果。

注意：创建计算字段时，不要使用保留名称“Number of Records”(记录数)。使用此保留名称将导致与权限相关的错误。

在 Tableau 中，表计算仅适用于可视化项中的值。虽然您可以基于结果进行构建，但必须使用另一个表计算来执行此操作。有关在 Tableau 中使用表计算的详细信息，请参见 Tableau 帮助中的使用表计算转换值。

在数据准备期间执行表计算可以在 Tableau 中分析数据时提供更大的灵活性。在构建视图时，您可以轻松地重复使用计算，并且基础计算不会受到筛选的影响。由于在查询运行后不会重新计算表计算，因此大型数据集的工作簿加载时间可能会更快。

Tableau Prep 目前支持以下多行计算：

- **差异**：计算当前行值与另一个值之间的差异。
- **百分比差异**：以百分比形式计算当前行值与另一个值之间的差异。
- **移动计算**：返回一组灵活的行中数字字段的总和或平均值。
- **填充计算**：计算并填充上面行中缺失的数据。
- **运行计算**：计算列中值的总和或平均值。

使用可视化计算编辑器快速生成计算，或在计算编辑器中编写您自己的自定义计算。

计算差异

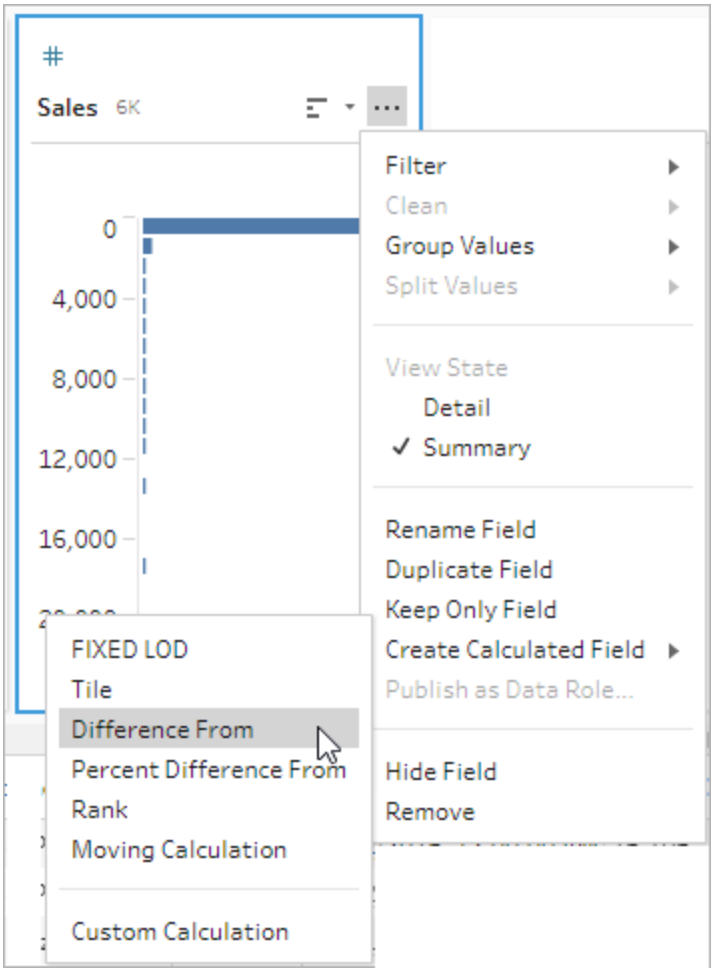
在 Tableau Prep Builder 2023.2 及更高版本以及 Tableau Cloud 中的 Web 上受支持。在 Tableau Server 2023.3 及更高版本中也受支持。

“差异”计算对当前值与当前行之前或之后 N 行的值之间的差异进行运算。

可视化计算编辑器

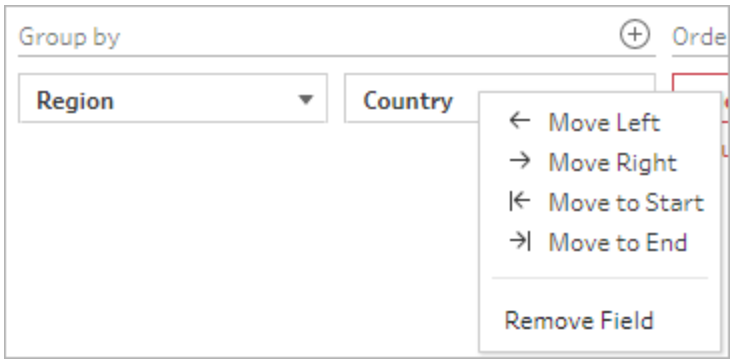
从列表中选择字段，Tableau Prep 将在您进行选择时为您构建计算。新字段结果的预览显示在左侧窗格中，您可以在窗格最右侧查看计算结果。

1. 在配置卡或结果窗格中，单击**“更多选项”**··· 菜单并选择**“创建计算字段”>“差异”**。



2. 在**“分组依据”**部分，选择具有要包括在计算中的行的字段。这会在执行计算时对表进行分区。若要将计算应用于表中的所有行，请接受默认值**“完整表”**。

   选择第一个字段后，单击加号 ⊕ 图标以将任何其他**“分组依据”**字段添加到分区。若要重新排序或移除字段，请右键单击或按住 Ctrl 键单击 (MacOS)，然后从菜单中选择一个操作。

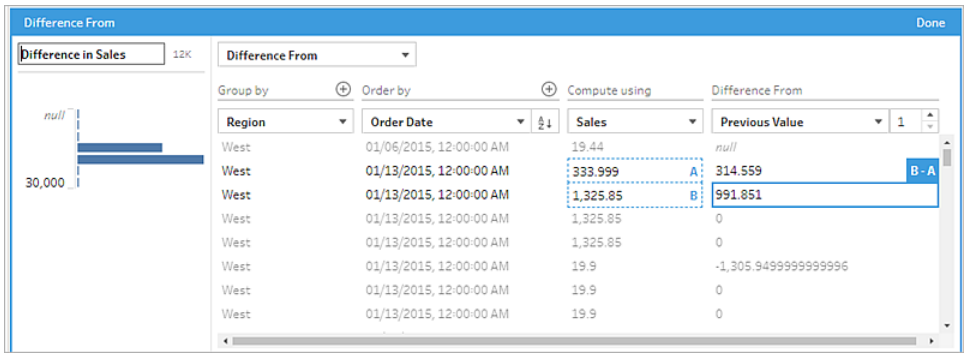


3. 在**“排序依据”**部分，选择要用作排序依据的字段。此字段用于指定 LOOKUP 函数如何对表中的行进行排序。

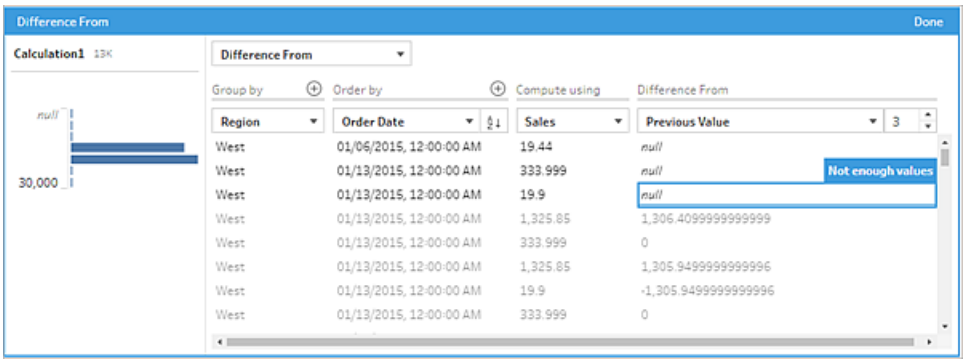
   如果您在其中选择**“创建计算字段”>“差异”**菜单选项的字段是日期或时间字段，则默认情况下会添加此字段，但您可以对其进行更改。

   单击加号 ⊕ 图标以向计算中添加任何其他**“排序依据”**字段。单击排序图标 将顺序从升序 (ASC) 更改为降序 (DESC)。您还可以右键单击或按住 Ctrl 键单击 (MacOS)，然后从菜单中选择一个操作来重新排序或移除字段。
4. 在**“计算依据”**部分，选择包含要用于计算结果的值的字段。
5. 在**“差异”**部分，选择用于计算差异的行。例如，选择**前一个值** 2 可计算当前值与该值之前 2 行的值之间的差异。注释突出显示用于执行计算的行。

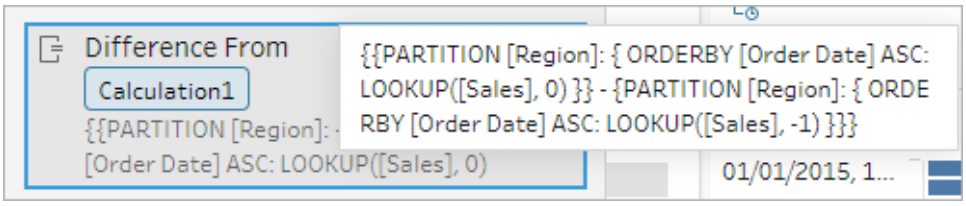
   默认情况下，计算预览将显示第一个非 null 行。但是，您可以单击结果表中的任意行并查看所选值的更新预览。



如果无法使用当前设置执行计算，则会显示注释**“值不足”**。为了解决此问题，请选择不同的当前值或更改**“差异”**部分中的配置。



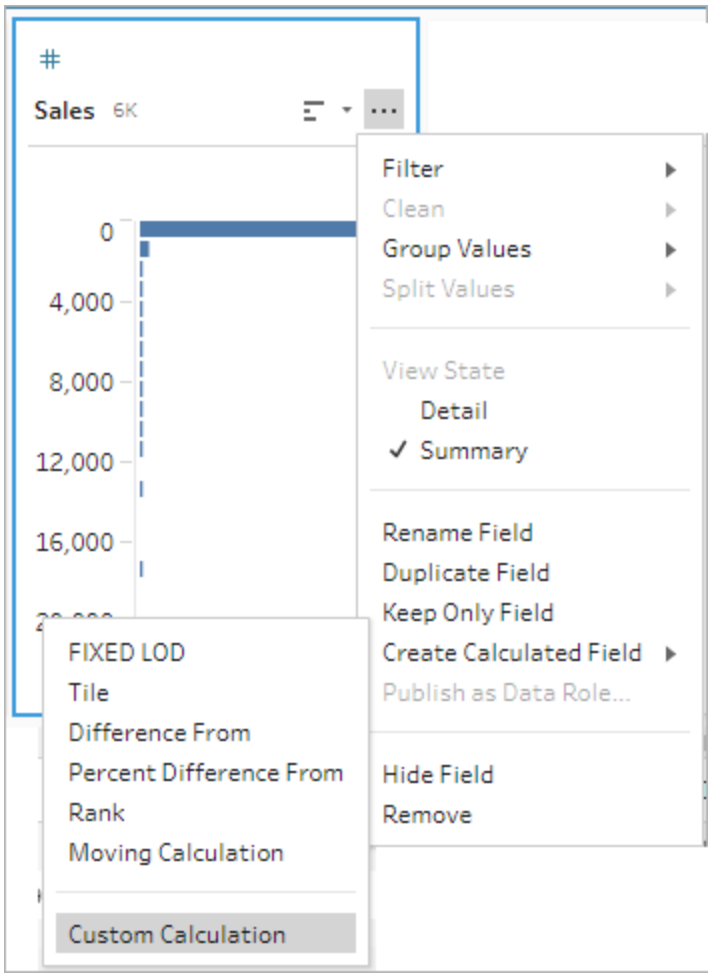
6. 在左侧窗格中，双击字段标题并输入用于计算的名称。
7. 单击**“完成”**以添加新的计算字段。在**“更改”**窗格中，您可以看到 Tableau Prep 生成的计算。单击**“编辑”**打开可视化计算编辑器以进行任何更改。



计算编辑器

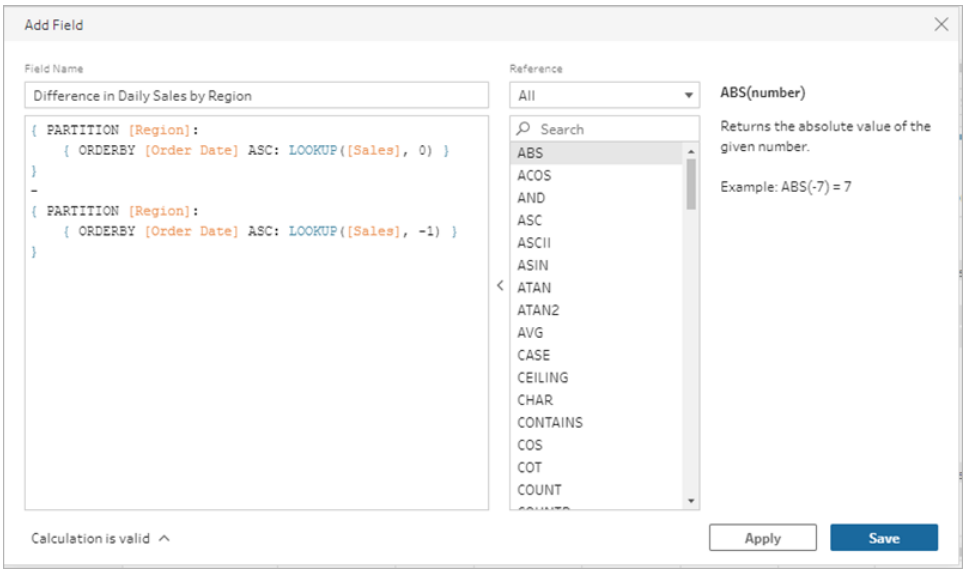
如果您想编写自己的计算来计算两个值之间的差异，请在计算编辑器中使用 LOOKUP 函数。

1. 在“配置”窗格工具栏中单击**“创建计算字段”**，或在配置卡或数据网格中单击**“更多选项”** ··· 菜单，并选择**“创建计算字段”>“自定义计算”**。



2. 在“计算编辑器”中输入表达式。例如，若要按区域查找当前销售额与前一天销售额之间的差异，请创建如下所示的计算。

```
{PARTITION [Region]:{ ORDERBY [Order Date]ASC:LOOKUP([Sales],0)}}
-
{ PARTITION [Region]:{ ORDERBY [Order Date]ASC:LOOKUP
([Sales],1)}}
```



3. 为计算输入一个名称，并单击**“保存”**：

计算百分比差异

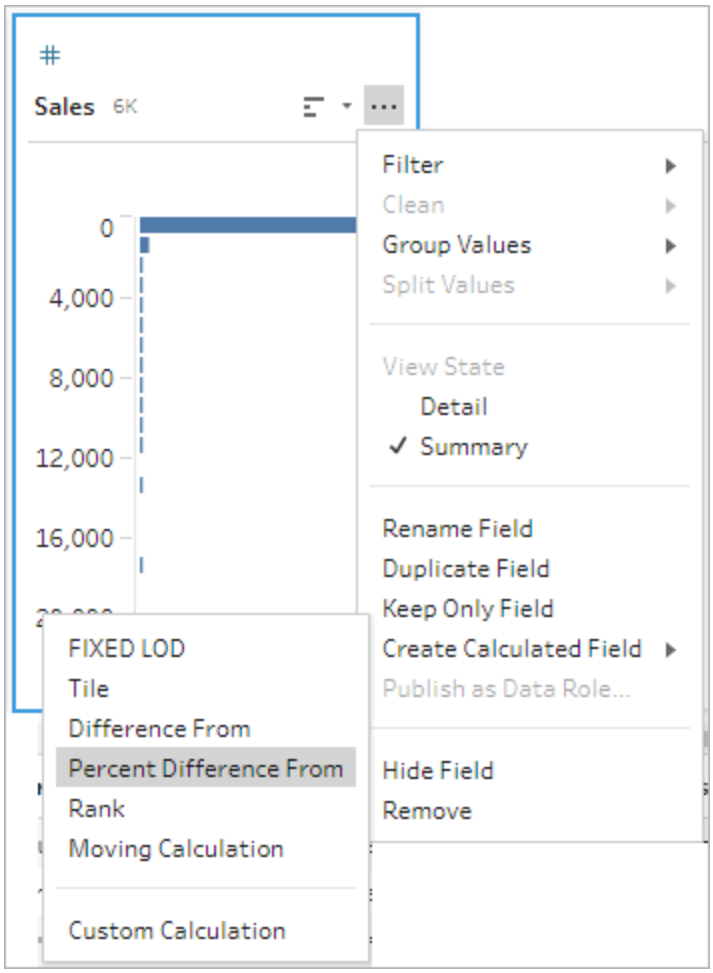
在 Tableau Prep Builder 2023.2 及更高版本以及 Tableau Cloud 中的 Web 上受支持。在 Tableau Server 2023.3 及更高版本中也受支持。

百分比差异计算以百分比形式计算当前值与当前行之前或之后 N 行的值之间的差异。例如，`Value1-Value2/Value2`。

可视化计算编辑器

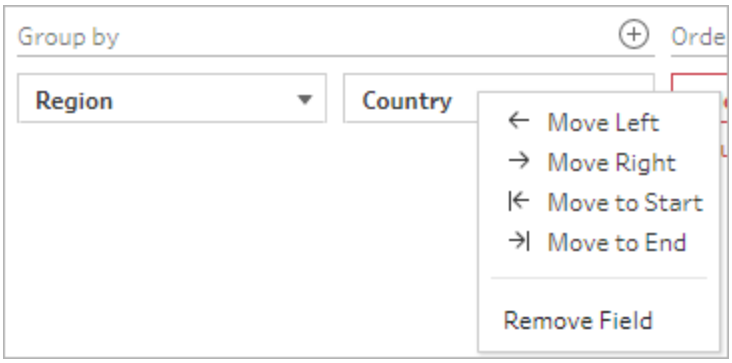
从列表中选择字段，Tableau Prep 将在您进行选择时为您构建计算。新字段结果的预览显示在左侧窗格中，您可以在窗格最右侧查看计算结果。

1. 在配置卡或结果窗格中，单击**“更多选项”**⋯ 菜单并选择**“创建计算字段”>“百分比差异”**。



2. 在**“分组依据”**部分，选择具有要包括在计算中的行的字段。这会在执行计算时对表进行分区。若要将计算应用于表中的所有行，请接受默认值**“完整表”**。

   选择第一个字段后，单击加号 ⊕ 图标以将任何其他**“分组依据”**字段添加到分区。若要重新排序或移除字段，请右键单击或按住 Ctrl 键单击 (MacOS)，然后从菜单中选择一个操作。

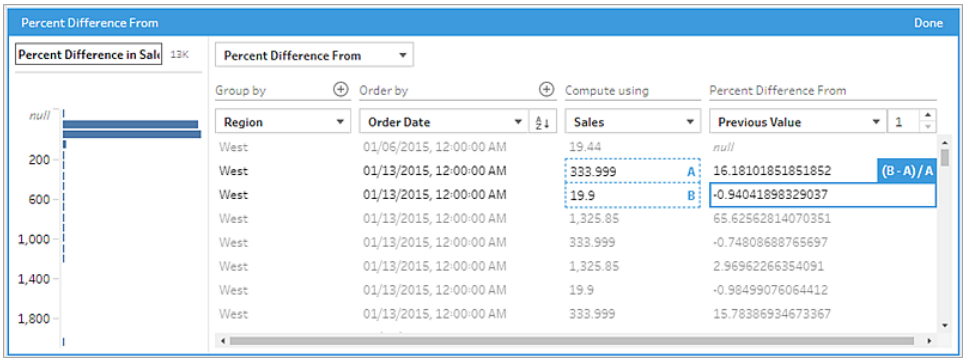


3. 在**“排序依据”**部分，选择要用作排序依据的字段。此字段用于指定 LOOKUP 函数如何对表中的行进行排序。

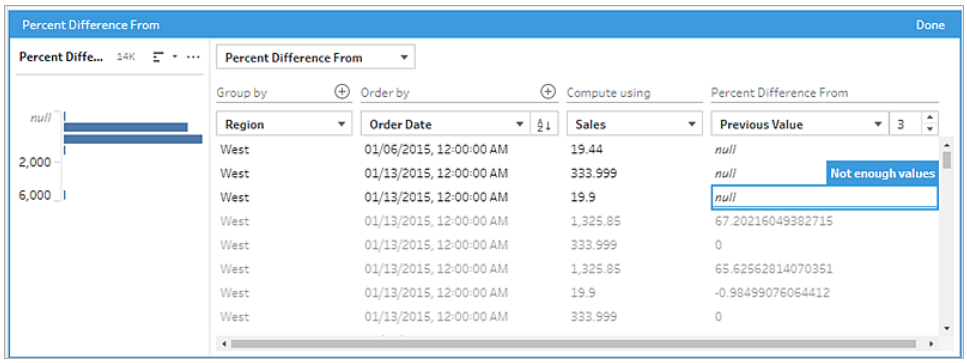
   如果您在其中选择**“创建计算字段”>“百分比差异”**菜单选项的字段是日期或时间字段，则默认情况下会添加此字段，但您可以对其进行更改。

   单击加号 ⊕ 图标以向计算中添加任何其他**“排序依据”**字段。单击排序图标 将顺序从升序 (ASC) 更改为降序 (DESC)。您还可以右键单击或按住 Ctrl 键单击 (MacOS)，然后从菜单中选择一个操作来重新排序或移除字段。
4. 在**“计算依据”**部分，选择包含要用于计算结果的值的字段。
5. 在**“百分比差异”**部分，选择用于计算结果的行。例如，选择**前一个值** 2 可计算当前值与该值之前 2 行的值之间的百分比差异。注释突出显示用于执行计算的行。

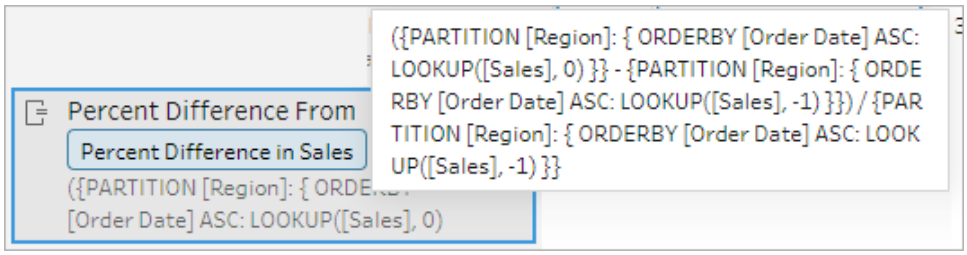
   默认情况下，计算预览将显示第一个非 null 行。但是，您可以单击结果表中的任意行并查看所选值的更新预览。



如果无法使用当前设置执行计算，您将看到注释**“值不足”**。为了解决此问题，请选择不同的当前值或更改**“百分比差异”**部分中的配置。



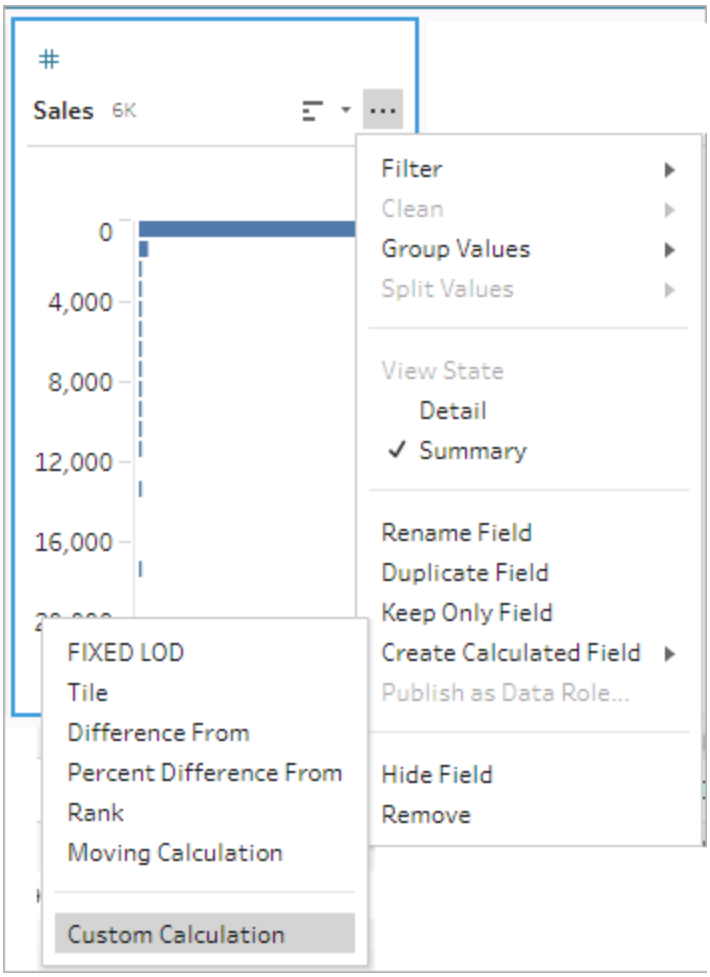
6. 在左侧窗格中，双击字段标题并输入用于计算的名称。
7. 单击**“完成”**以添加新的计算字段。在**“更改”**窗格中，您可以看到 Tableau Prep 生成的计算。单击**“编辑”**打开可视化计算编辑器以进行任何更改。



计算编辑器

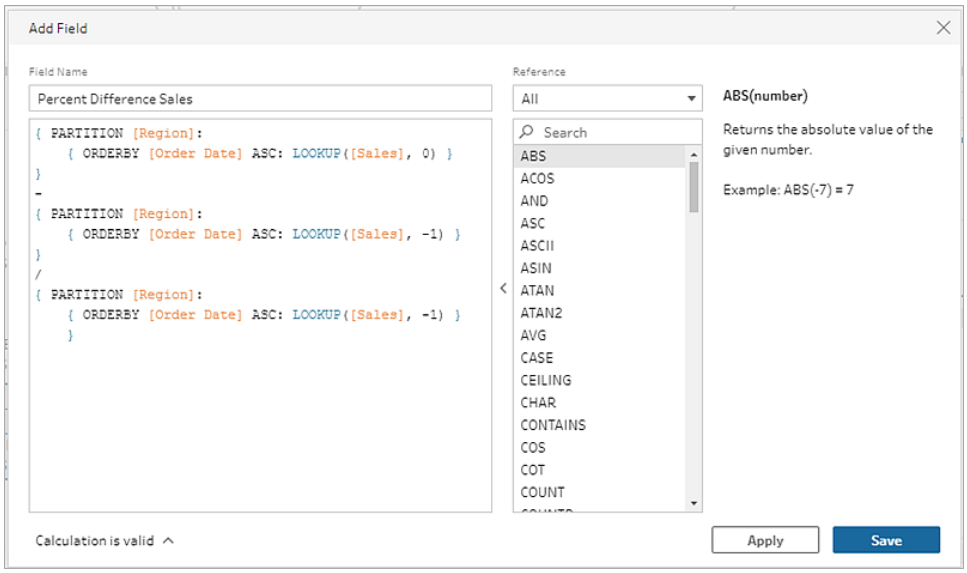
如果您想编写自己的计算来计算两个值之间的百分比差异，请在计算编辑器中使用 LOOKUP 函数。

1. 在“配置”窗格工具栏中单击**“创建计算字段”**，或在配置卡或数据网格中单击**“更多选项”**⋯ 菜单，并选择**“创建计算字段”>“自定义计算”**。



2. 在“计算编辑器”中输入表达式。例如，若要按区域查找当前销售额与前几天销售额之间的百分比差异，请创建如下所示的计算。

```
{ PARTITION [Region]:{ ORDERBY [Order Date]ASC:LOOKUP
([Sales],0)}}
-
{ PARTITION [Region]:{ ORDERBY [Order Date]ASC:LOOKUP([Sales],-
1)}}
/
{ PARTITION [Region]:{ ORDERBY [Order Date]ASC:LOOKUP([Sales],-
1)}}
```



3. 为计算输入一个名称，并单击**“保存”**：

计算移动平均值或总和

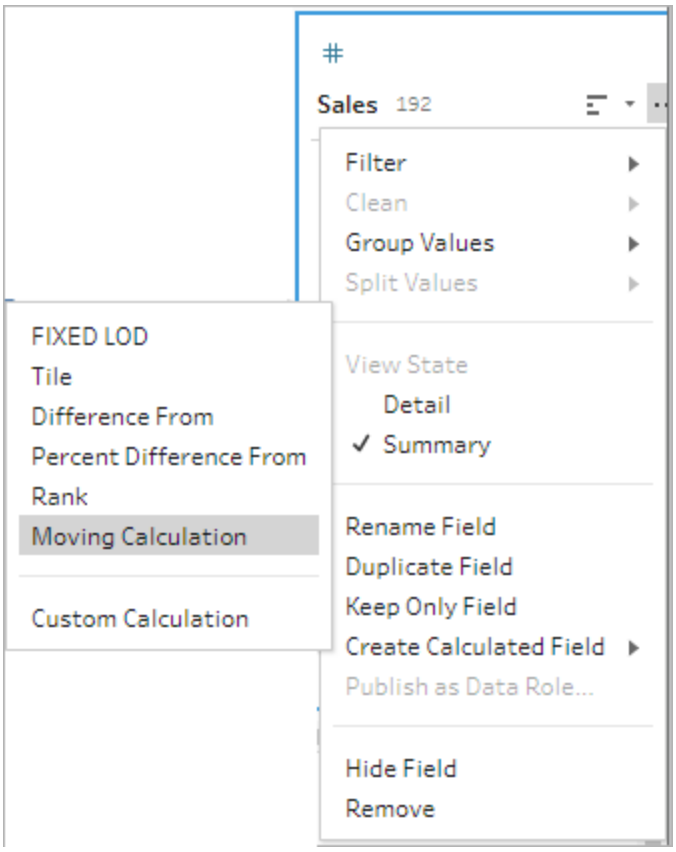
在 Tableau Prep Builder 2023.2 及更高版本以及 Tableau Cloud 中的 Web 上受支持。在 Tableau Server 2023.3 及更高版本中也受支持。

创建移动计算以更好地了解数据趋势并减少总体波动。在 Tableau Prep 中，您可以计算当前值之前或之后指定数量的值的移动平均值或总和。例如，跟踪每个地区销售额的三个月移动平均值。

可视化计算编辑器

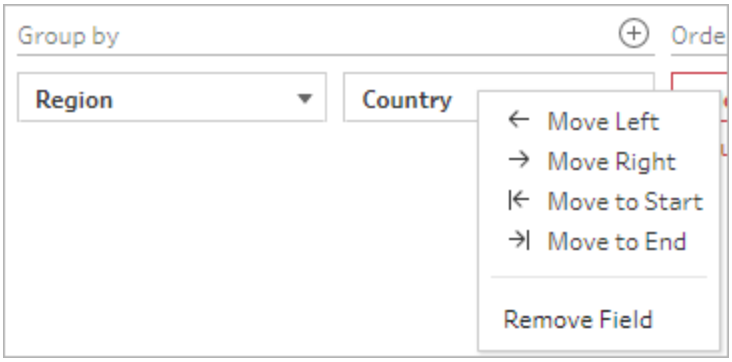
从列表中选择字段，Tableau Prep 将在您进行选择时为您构建计算。新字段结果的预览显示在左侧窗格中，您可以在窗格最右侧查看计算结果。

1. 在配置卡或结果窗格中，单击**“更多选项”**⋯ 菜单并选择**“创建计算字段”>“移动计算”**。



2. 在**"分组依据"**部分，选择具有要包括在计算中的行的字段。这会在执行计算时对表进行分区。若要将计算应用于表中的所有行，请接受默认值**"完整表"**

   选择第一个字段后，单击加号 ⊕ 图标以将任何其他**"分组依据"**字段添加到计算。若要重新排序或移除字段，请右键单击或按住 Ctrl 键单击 (MacOS)，然后从菜单中选择一个操作。

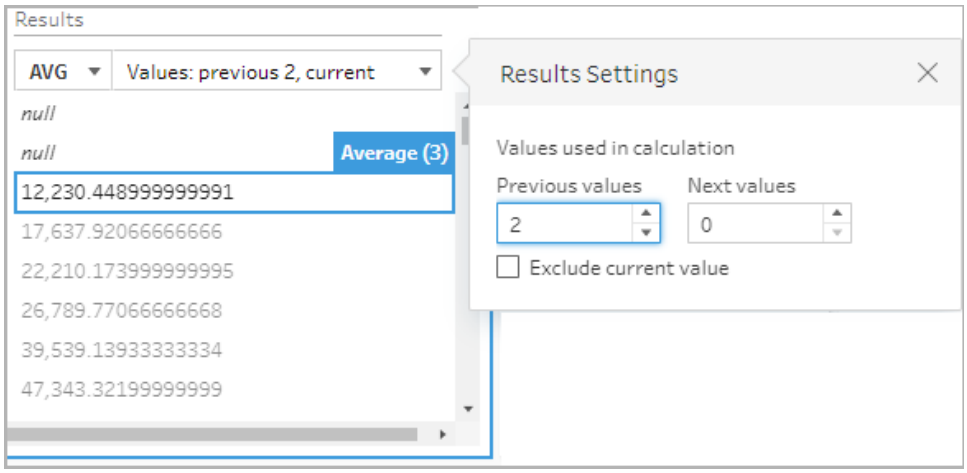


3. 在**“排序依据”**部分，选择要用作排序依据的字段。此字段用于指定 LOOKUP 函数如何对表中的行进行排序。

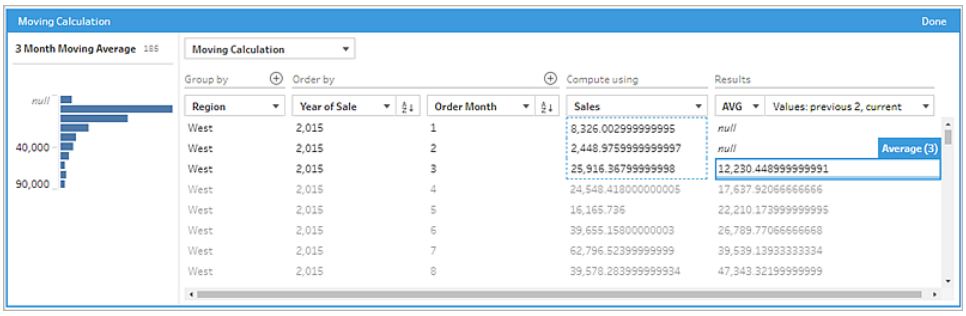
   如果您在其中选择**“创建计算字段”>“移动计算”**菜单选项的字段是日期或时间字段，则默认情况下会添加此字段，但您可以对其进行更改。

   单击加号 ⊕ 图标以向计算中添加任何其他**“排序依据”**字段。单击排序图标 将顺序从升序 (ASC) 更改为降序 (DESC)。您还可以右键单击或按住 Ctrl 键单击 (MacOS)，然后从菜单中选择一个操作来重新排序或移除字段。
4. 在**“计算依据”**部分，选择包含要用于计算结果的值的字段。
5. 在**“结果”**部分，选择要执行的聚合（总和或平均值）、要包含在计算中的行数以及是包含还是排除当前行。

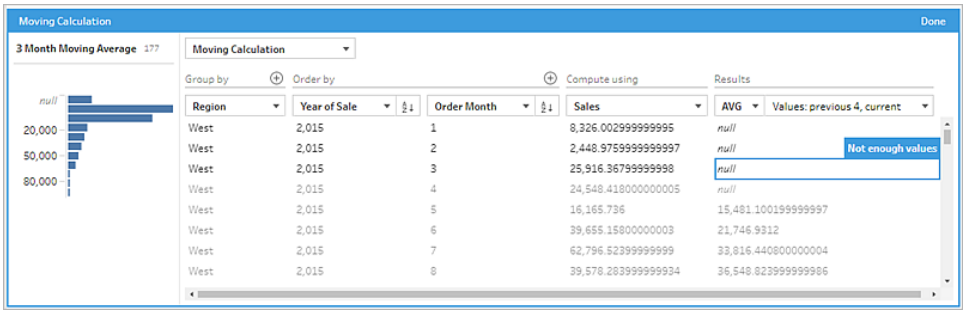
   若要更改结果设置，请单击**“值”**字段中的下拉菜单。例如，若要计算当月和前 2 个月的销售额移动平均值，请将**“上一个值”**设置为 2 并关闭对话框。



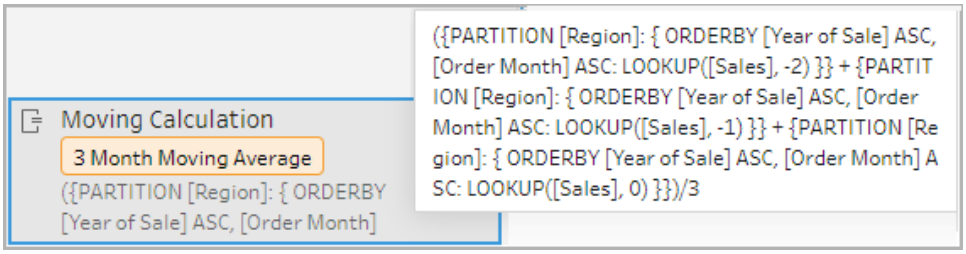
6. 默认情况下，计算预览将显示第一个非 null 行。但是，您可以单击结果表中的任意行并查看所选值的更新预览。注释突出显示用于执行计算的行。



如果无法使用当前设置执行计算，您将看到注释**“值不足”**。为了解决此问题，请单击**“值”**字段中的下拉菜单以更改**“结果设置”**中的配置。



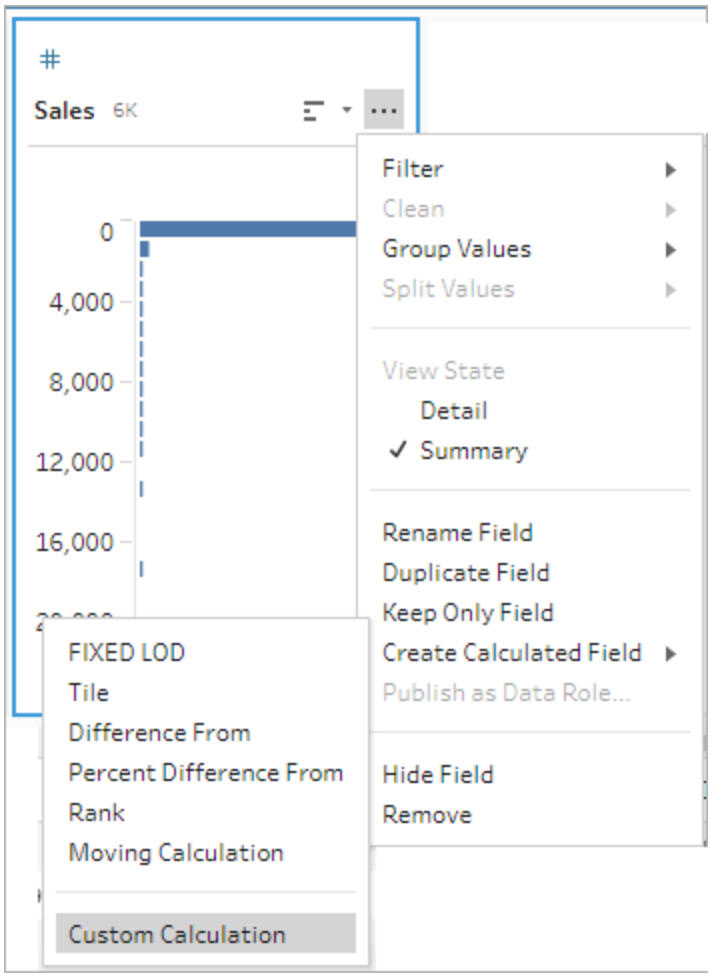
7. 在左侧窗格中，双击字段标题并输入用于计算的名称。
8. 单击**“完成”**以添加新的计算字段。在**“更改”**窗格中，您可以看到 Tableau Prep 生成的计算。单击**“编辑”**打开可视化计算编辑器以进行任何更改。



计算编辑器

如果您想编写自己的计算来计算移动平均值或总和，请在计算编辑器中使用 LOOKUP 函数。

1. 在“配置”窗格工具栏中单击**“创建计算字段”**，或在配置卡或数据网格中单击**“更多选项”** ··· 菜单，并选择**“创建计算字段”>“自定义计算”**。

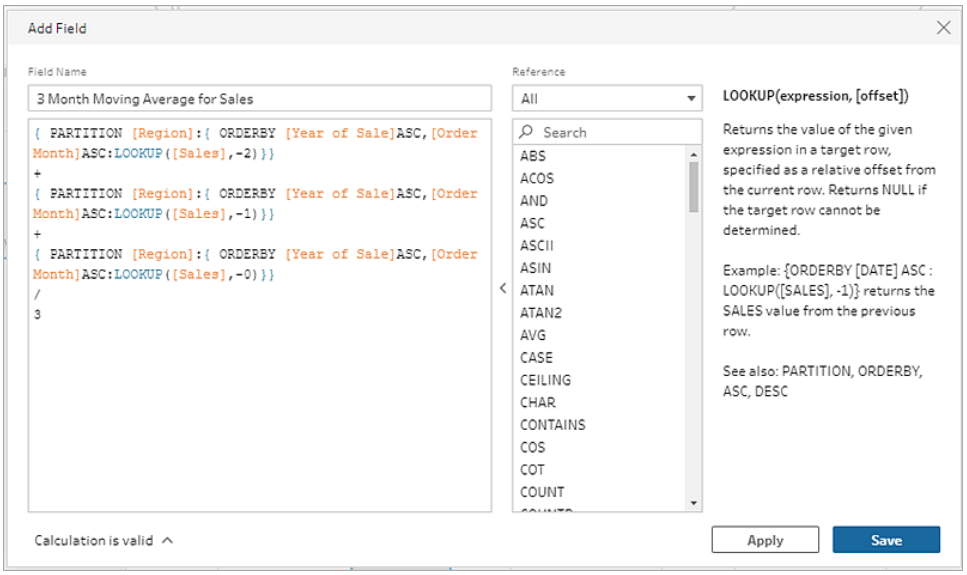


2. 在“计算编辑器”中输入表达式。例如，若要按区域查找销售额三个月的移动平均值，请创建如下所示的计算。

> **注意**：此示例假设数据集的详细级别正确，每月一行。如果您的数据集的详细级别不正确，请考虑在应用计算之前使用聚合步骤来更改此设置。

```
{ PARTITION [Region]:{ ORDERBY [Year of Sale]ASC,[Order
Month]ASC:LOOKUP([Sales],-2)}}
+
{ PARTITION [Region]:{ ORDERBY [Year of Sale]ASC,[Order
Month]ASC:LOOKUP([Sales],-1)}}
+
{ PARTITION [Region]:{ ORDERBY [Year of Sale]ASC,[Order
```

```
Month]ASC:LOOKUP([Sales],-0)}}
/
3
```



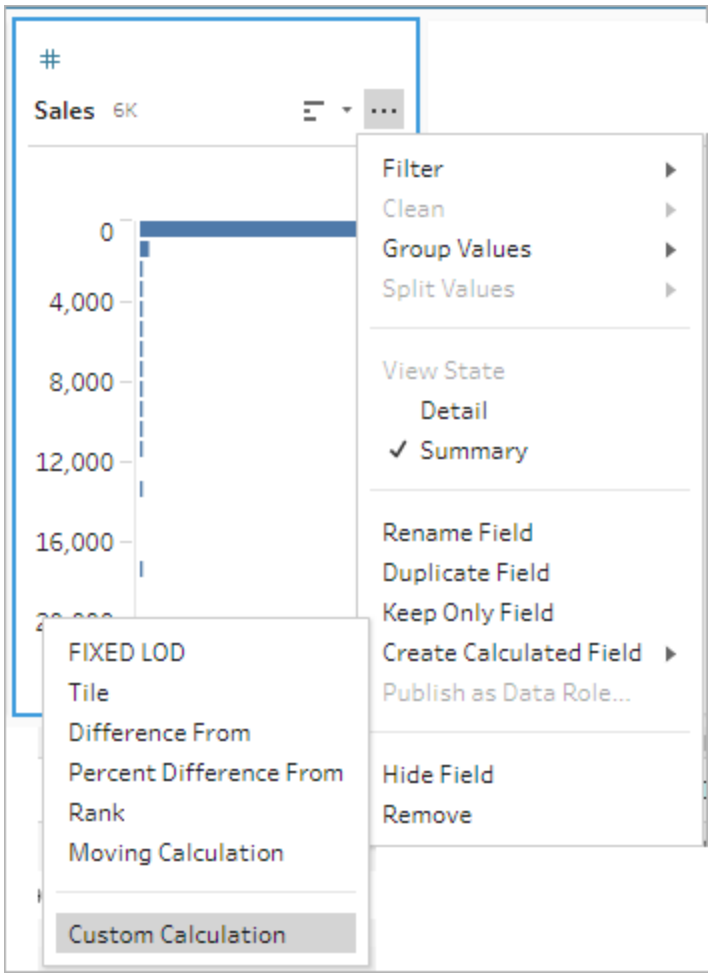
3. 为计算输入一个名称，并单击**“保存”**：

获取上一个值

在 Tableau Prep Builder 2023.2 及更高版本以及 Tableau Cloud 中的 Web 上受支持。在 Tableau Server 2023.3 及更高版本中也受支持。

如果您需要使用前一行的值创建一个字段，您可以使用 LOOKUP 函数创建自定义计算。

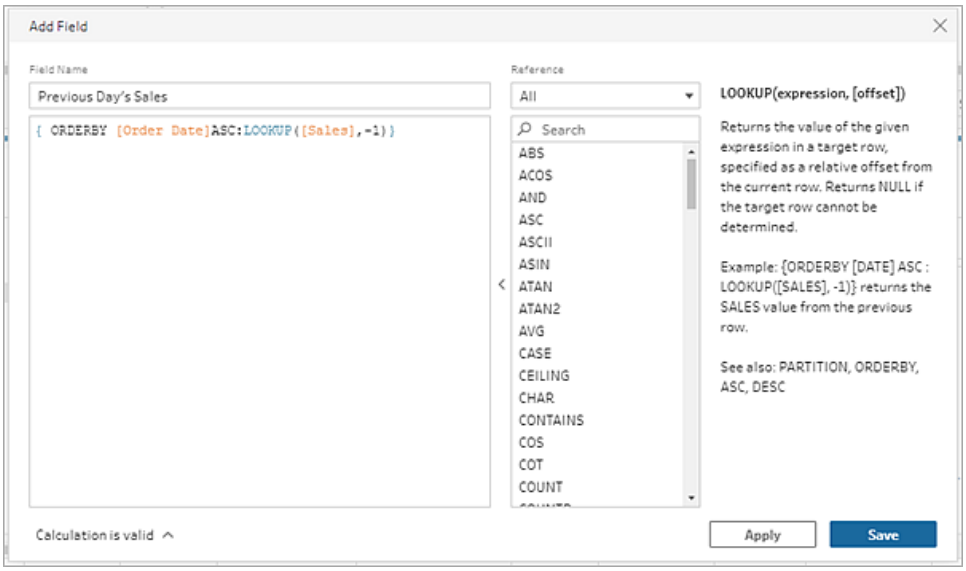
1. 在“配置”窗格工具栏中单击**“创建计算字段”**，或在配置卡或数据网格中单击**“更多选项”** ··· 菜单，并选择**“创建计算字段”>“自定义计算”**。



2. 在“计算编辑器”中输入表达式。例如，若要按订单日期查找以前的销售额值，请创建如下所示的计算。

> **注意**：此示例假设数据集的详细级别正确，每天一行。如果您的数据集的详细级别不正确，请考虑在应用计算之前使用聚合步骤来更改此设置。

```
{ ORDERBY [Order Date]ASC:LOOKUP([Sales],-1)}
```



3. 为计算输入一个名称，并单击**“保存”**：

填充计算

在 Tableau Prep Builder 2023.2 及更高版本以及 Tableau Cloud 中的 Web 上受支持。在 Tableau Server 2023.3 及更高版本中也受支持。

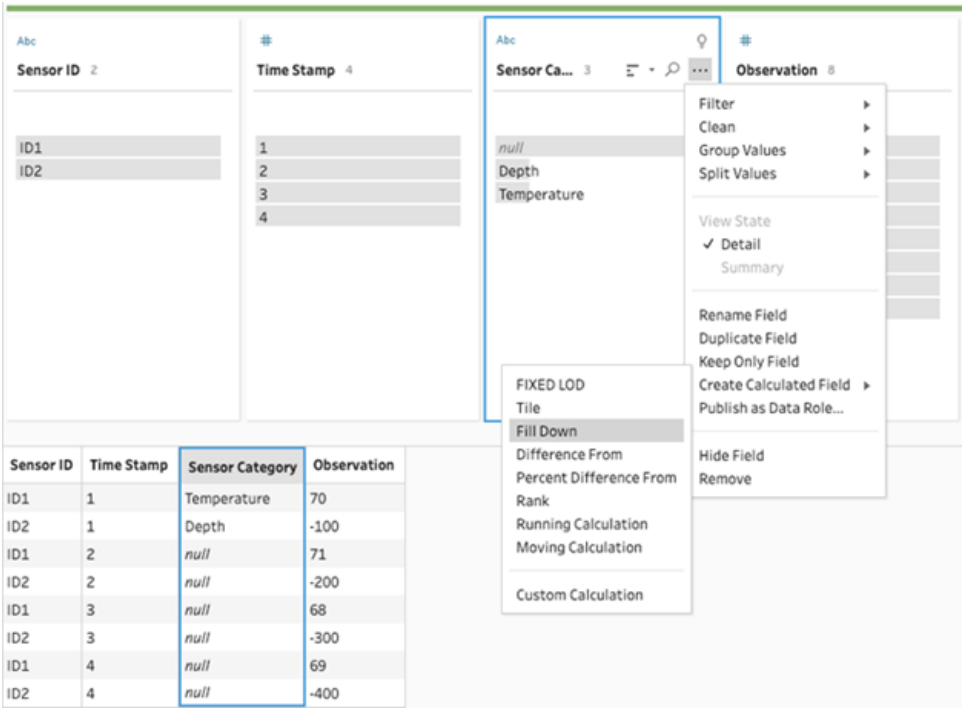
填充计算会计算并填充缺失的数据，以便将 null 值替换为上面行中的值，从而使数据易于可视化。下面的示例显示了填充计算的结果，方法是将“Student”(学生)字段中的 null 值替换为基于上面行的值。

Student	Subject	Grade
Amanda	English	90
	Math	89
	Science	80
Daniel	English	70
	Math	70
	Science	65
Ian	English	65
	Math	60
	Science	55

Student	Subject	Grade
Amanda	English	90
Amanda	Math	89
Amanda	Science	80
Daniel	English	70
Daniel	Math	70
Daniel	Science	65
Ian	English	65
Ian	Math	60
Ian	Science	55

Prep 可视化计算编辑器

1. 在配置卡或结果窗格中，单击**“更多选项”**⋯ 菜单，并选择**“创建计算字段”>“填充”**。



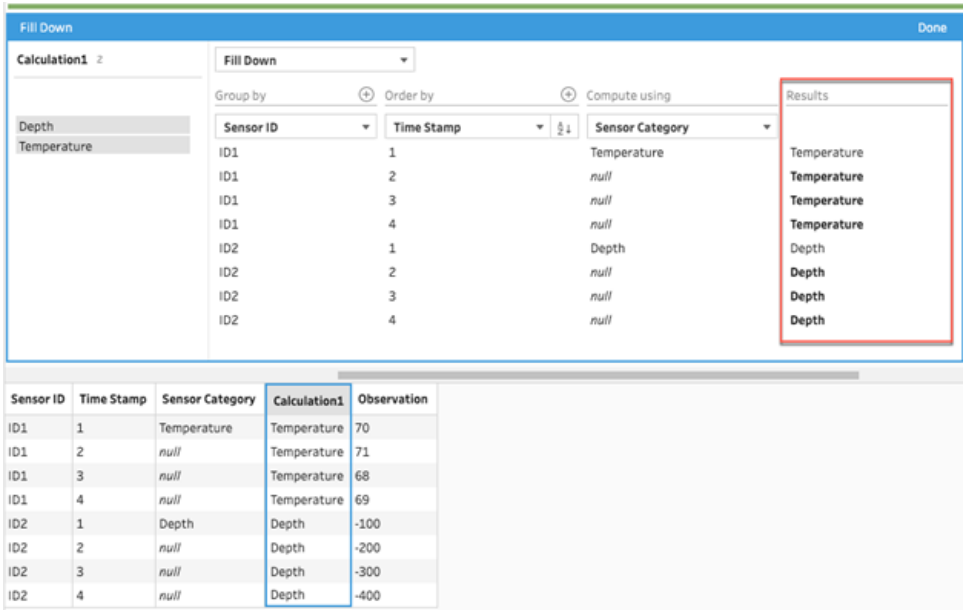
2. 在**“分组依据”**部分，您可以在执行计算时选择字段来对数据进行分区。如果您不希望对数据进行分区，请接受默认值**“完整表”**。

   选择第一个字段后，您可以通过单击加号 ⊕ 图标将其他“分组依据”字段添加到分区中。若要重新排序或移除字段，请右键单击或按住 Ctrl 键单击 (MacOS)，然后从菜单中选择一个操作。

3. 在**“排序依据”**部分，选择要用于排序顺序的一个或多个字段。此字段用于指定表中行的顺序。

   您还可以在文本、CSV 或 Excel 文件的输入步骤中添加源行号，以按数据源顺序对数据进行排序。

4. 单击加号 ⊕ 图标以向计算中添加任何其他“排序依据”字段。您至少必须选择一个字段。单击排序图标 将顺序从升序 (ASC) 更改为降序 (DESC)。您还可以右键单击或按住 Ctrl 键单击 (MacOS)，然后从菜单中选择一个操作来重新排序或移除字段。
5. 在**“计算依据”**部分，选择包含要填充的值的字段。**“结果”**字段以粗体显示填充的值。



6. 在左侧窗格中，双击字段标题并输入用于计算的名称。
7. 单击**“完成”**以添加新的计算字段。在“更改”窗格中，您可以看到 Tableau Prep 生成的计算。单击**“编辑”**打开可视化计算编辑器以进行任何更改。

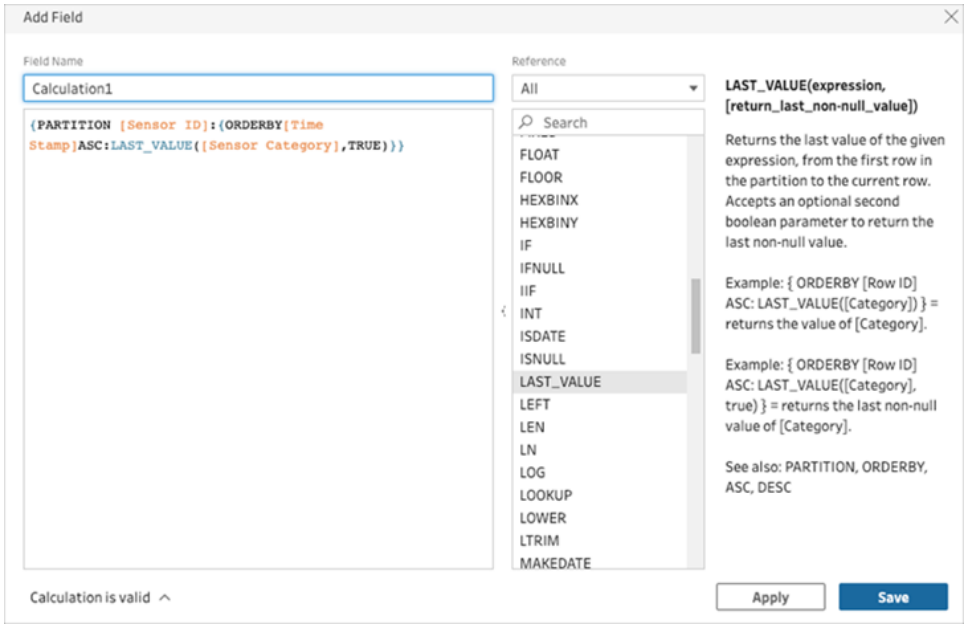
计算编辑器

1. 在“配置”窗格工具栏中单击**“创建计算字段”**，或在配置卡或数据网格中单击**“更多选项”** ··· 菜单，并选择**“创建计算字段”>“自定义计算”**。
2. 在“计算编辑器”中输入表达式。
3. 为计算输入一个名称，并单击**“保存”**：

示例

带分区的填充示例：根据以下数据，您可以使用 LAST_VALUE 函数填充“Sensor Category”（传感器类别）字段中的缺失值。在此示例中，*“Sensor ID”（传感器 ID）*用于对数据进行分区，*“Time Stamp”（时间戳）*用于对数据进行排序，*“Sensor Category”（传感器类别）*用于对数据进行运算。

Sensor ID	Time Stamp	Sensor Category	Observation
ID1	1	Temperature	70
ID2	1	Depth	-100
ID1	2		71
ID2	2		-200
ID1	3		68
ID2	3		-300
ID1	4		69
ID2	4		-400

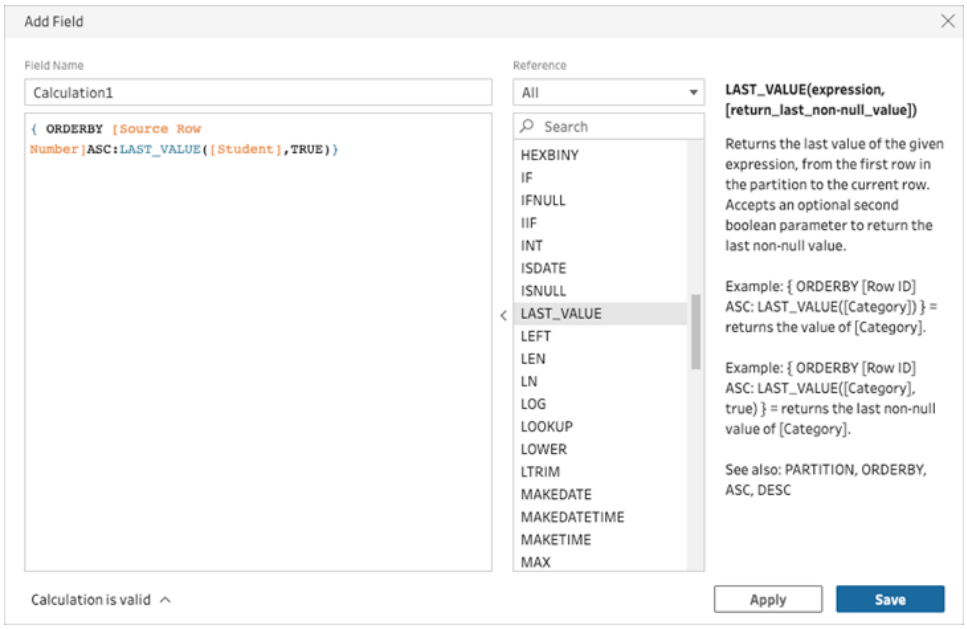


填充结果在计算窗格中突出显示。

Calculation1	Sensor ID	Time Stamp	Sensor Category	Observation
Temperature	ID1	1	Temperature	70
Temperature	ID1	2	*null*	71
Temperature	ID1	3	*null*	68
Temperature	ID1	4	*null*	69
Depth	ID2	1	Depth	-100
Depth	ID2	2	*null*	-200
Depth	ID2	3	*null*	-300
Depth	ID2	4	*null*	-400

不带分区的填充示例：根据以下数据，您可以填充*“Student”(学生)*字段的缺失值。*“Source Row Number”(源行号)*字段已添加到输入步骤中，用于按源顺序对数据进行排序。*“Student”(学生)*用于对数据进行运算。

Student	Subject	Grade
Amanda	English	90
	Math	89
	Science	80
Daniel	English	70
	Math	70
	Science	65
Ian	English	65
	Math	60
	Science	55



填充结果在计算窗格中突出显示。

Calculation1	Source Row Number	Student	Subject	Grade
Amanda	1	Amanda	English	90
Amanda	2	*null*	Math	89
Amanda	3	*null*	Science	80
Daniel	4	Daniel	English	70
Daniel	5	*null*	Math	70
Daniel	6	*null*	Science	65
Ian	7	Ian	English	65
Ian	8	*null*	Math	60
Ian	9	*null*	Science	55

运行计算

在 Tableau Prep Builder 2023.2 及更高版本以及 Tableau Cloud 中的 Web 上受支持。在 Tableau Server 2023.3 及更高版本中也受支持。

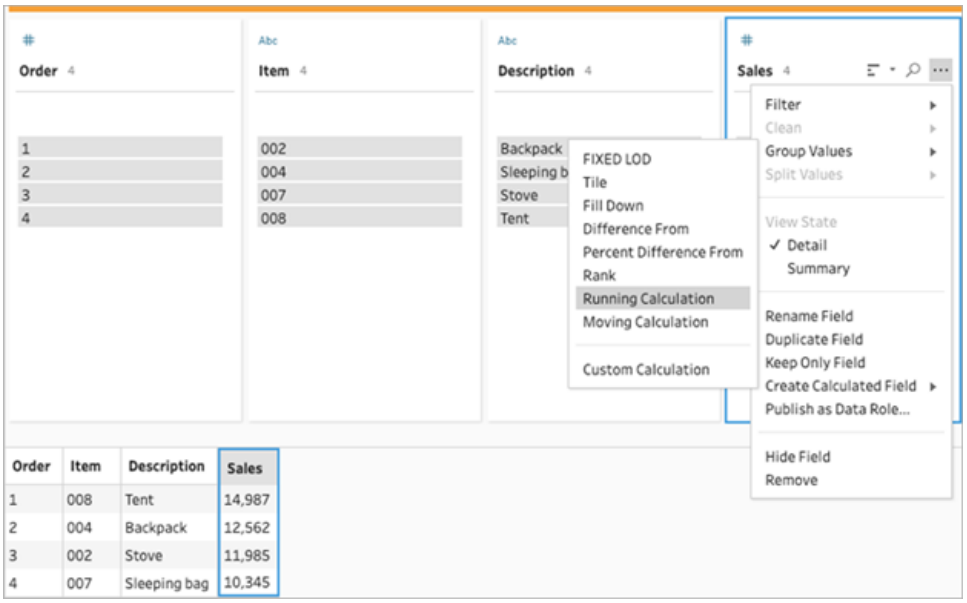
运行计算运算某个值与列中所有先前值的累积总和或平均值。下面的示例显示了随时间变化的销售额总和的运行计算结果。

Order	Item	Description	Sales
1	008	Tent	$14,987.00
2	004	Backpack	$12,562.00
3	002	Stove	$11,985.00
4	007	Sleeping bag	$10,345.00

Order	Item	Description	Sales	Running Sum of Sales
1	008	Tent	$14,987.00	$14,987.00
2	004	Backpack	$12,562.00	$27,549.00
3	002	Stove	$11,985.00	$39,534.00
4	007	Sleeping bag	$10,345.00	$49,879.00

Prep 可视化计算编辑器

1. 在配置卡或结果窗格中，单击**“更多选项”**⋯ 菜单，并选择**“创建计算字段”>“填充”**。



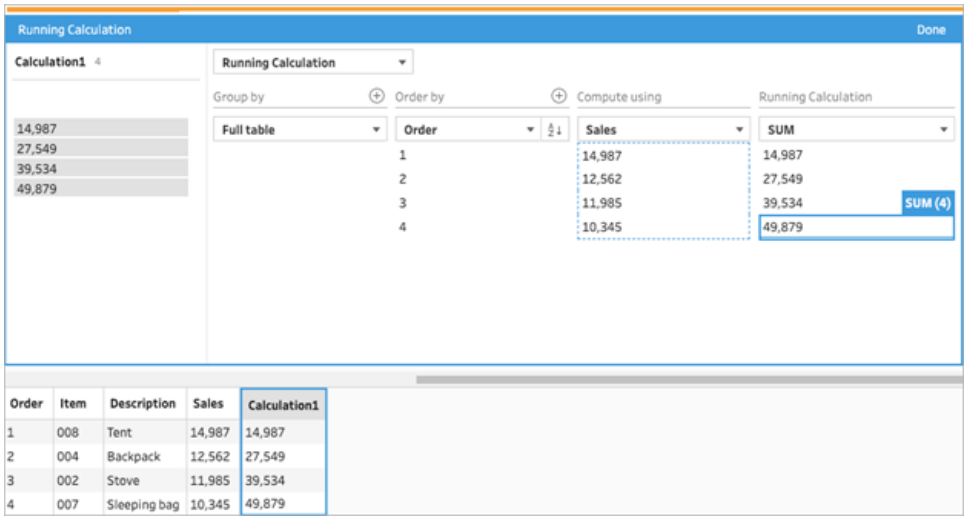
2. 在“分组依据”部分，您可以在执行计算时选择字段来对数据进行分区。如果您不希望对数据进行分区，请接受默认值**“完整表”**。

选择第一个字段后，您可以通过单击加号 ⊕ 图标将其他“分组依据”字段添加到分区中。若要重新排序或移除字段，请右键单击或按住 Ctrl 键单击 (MacOS)，然后从菜单中选择一个操作。

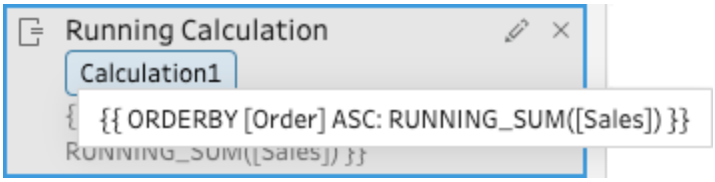
3. 在**“排序依据”**部分，选择要用于排序顺序的一个或多个字段。此字段用于指定表中行的顺序。

   您还可以在文本、CSV 或 Excel 文件的输入步骤中添加源行号，以按数据源顺序对数据进行排序。

4. 单击加号 ⊕ 图标以向计算中添加任何其他“排序依据”字段。您至少必须选择一个字段。单击排序图标 将顺序从升序 (ASC) 更改为降序 (DESC)。您还可以右键单击或按住 Ctrl 键单击 (MacOS)，然后从菜单中选择一个操作来重新排序或移除字段。
5. 在**“结果”**部分，选择要执行的聚合（SUM 或 AVG），以获得运行总和或运行平均值
6. 单击结果表中的各行并查看所选值的更新预览。注释突出显示用于执行计算的行。



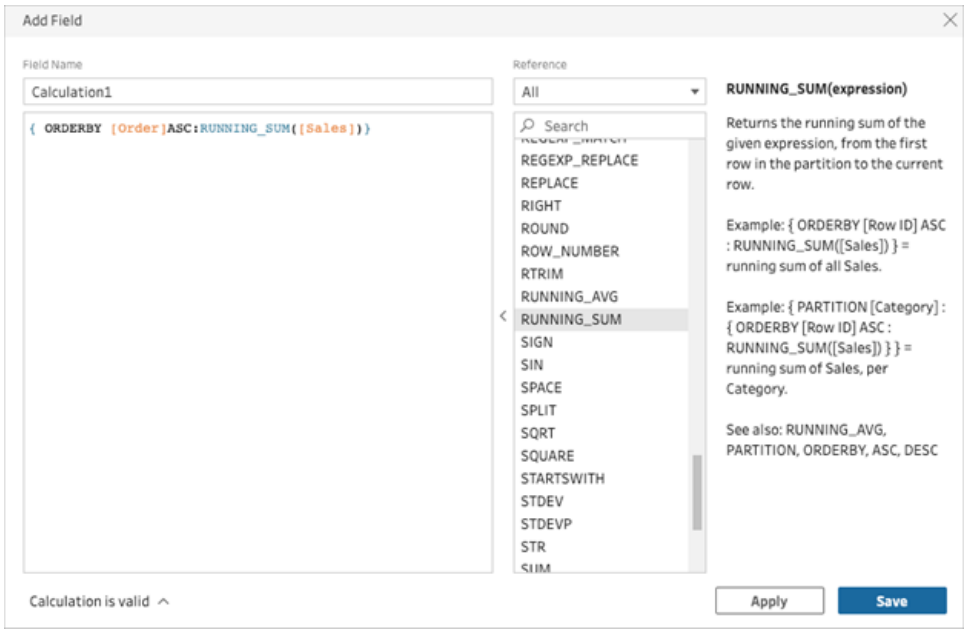
7. 在左侧窗格中，双击字段标题并输入用于计算的名称。
8. 单击**“完成”**以添加新的计算字段。在“更改”窗格中，您可以看到 Tableau Prep 生成的计算。单击**“编辑”**打开可视化计算编辑器以进行任何更改。



计算编辑器

1. 在“配置”窗格工具栏中单击**“创建计算字段”**，或在配置卡或数据网格中单击**“更多选项”** ··· 菜单，并选择**“创建计算字段”>“自定义计算”**。
2. 在“计算编辑器”中输入表达式。例如，根据以下数据，您可以使用 RUNNING_SUM 函数计算一段时间内的*“Sales”(销售额)* 总计。数据未分区，而是使用**“完整表”**进行分组。*“Order”(顺序)* 字段用于按时间顺序排序，*“Sales”(销售额)* 用于对数据进行运算。

Order	Item	Description	Sales
1	008	Tent	$14,987.00
2	004	Backpack	$12,562.00
3	002	Stove	$11,985.00
4	007	Sleeping bag	$10,345.00



3. 为计算输入一个名称，并单击**“保存”**：

   运行总和结果在计算窗格中突出显示。

Calculation1	Order	Item	Description	Sales
14,987	1	008	Tent	14,987
27,549	2	004	Backpack	12,562
39,534	3	002	Stove	11,985
49,879	4	007	Sleeping bag	10,345

使用 Tableau Agent 创建计算字段

在 Tableau Prep Web 制作版本 2024.2 中受支持。

注意：随着平台扩展以容纳更多 AI 代理功能，Einstein Copilot 已更名为 Tableau Agent。从 2024 年 10 月开始，您将看到 Tableau Prep、Tableau Catalog 和 Tableau Cloud Web 制作中的页面部分、字段名称和其他 UI 文本的更新。帮助内容和 Trailhead 模块也正在更新以反映这些变化。

现在，使用 Tableau Agent 创建计算字段比以往更加简单。只需描述一个计算，Tableau Agent 就会创建公式，让您只需单击一下即可向 Prep 流程中添加强大的计算字段。以前，

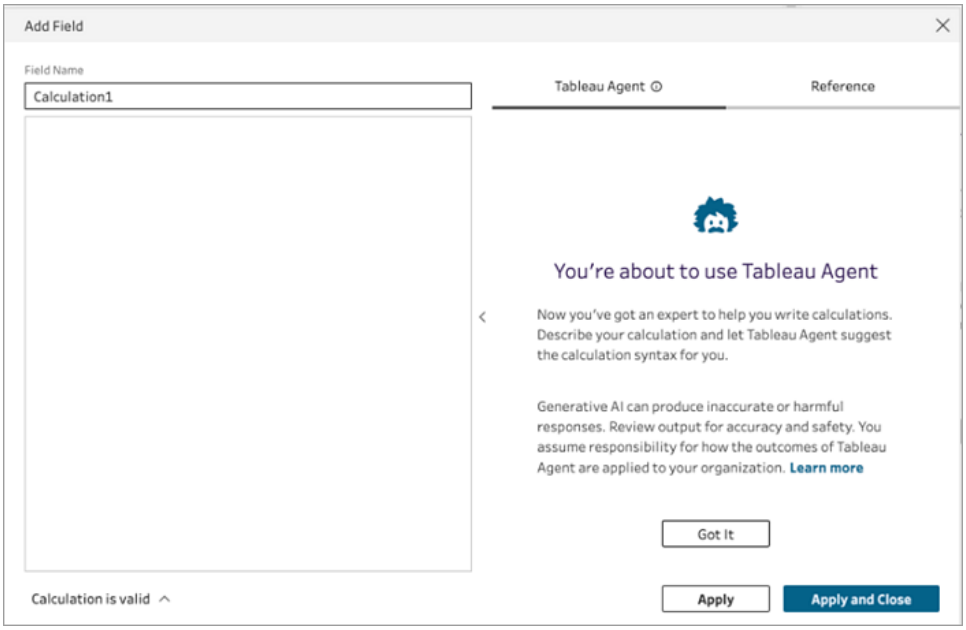
创建计算字段的公式需要了解对象和字段、可用函数以及基于数据类型、公式类型等的各种限制。

Tableau Agent 大大简化了创建自定义函数的任务。现在，技术和非技术用户都可以使用简单的语言轻松地创建计算。通过从自然语言描述中快速高效地创建行级和摘要公式字段的能力，您可以执行数据准备，从而更轻松、更快地组合、塑造和清理数据以进行分析。

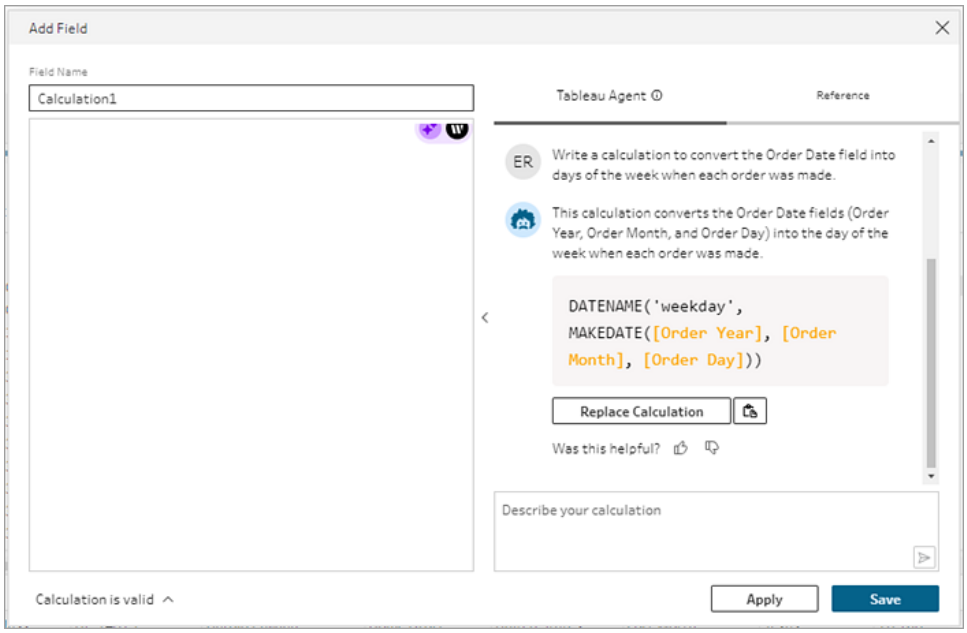
注意：您必须拥有 Tableau+ 并连接到设置了 Einstein 生成式 AI 的 Salesforce 组织才能在 Tableau Prep 中使用 Tableau Agent。有关详细信息，请参见为您的站点启用 Tableau 中的 AI。

创建计算字段

1. 在“流程”窗格中单击一个清理步骤。
2. 在“配置”窗格工具栏中单击**“创建计算字段”**，或在配置卡或数据网格中单击**“更多选项”** ··· 菜单，并选择**“创建计算字段”>“自定义计算”**。

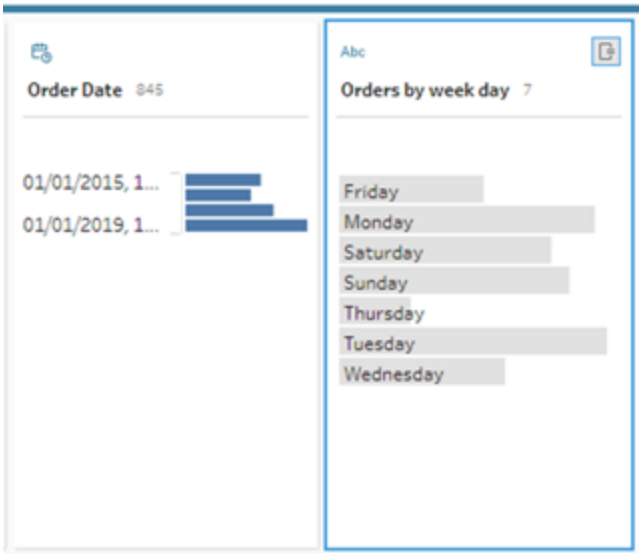


3. 为字段提供一个名称，并在“Tableau Agent”文本框框中描述您想要的计算。提供尽可能多的细节以尝试获得更好的结果。
4. 当您的请求准备就绪后，单击文本框中的箭头将其传递给 Tableau Agent。



注意：如果 Tableau Agent 不能提供有效的计算，您可以提供反馈。

5. 评估输出以确保它是您想要的。
6. 单击**“替换计算”**将计算添加到编辑器中。然后，您可以编辑计算并根据需要进行自定义。您还可以添加额外的提示来优化计算，然后替换原始计算。
7. 单击**“应用”**创建一个基于计算的字段。
8. 单击**“保存”**保存字段并关闭计算编辑器。

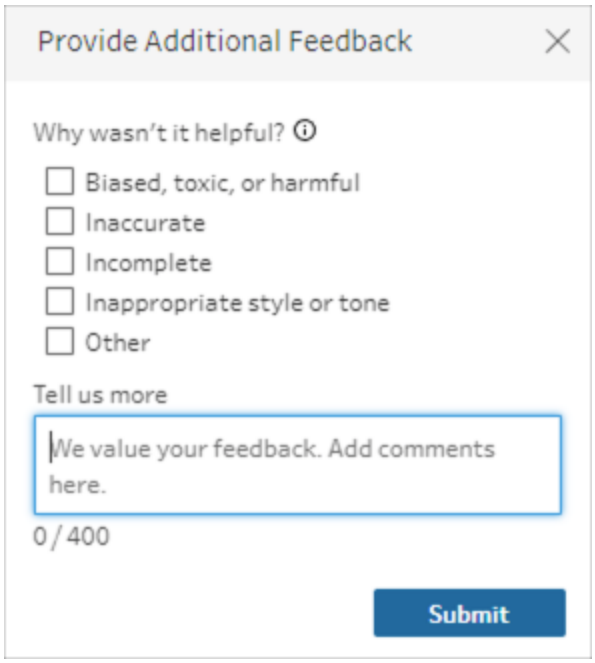


获得最佳结果的技巧

- 遵循生成计算语法的基本指令。
- 使用动词-名词结构。
  - 首选:编写一个计算来从字段中提取字符...
  - 非首选:对于提取字符,我该如何改变方式...
- 使用正确的语法和拼写。

帮助 Tableau Agent 改进

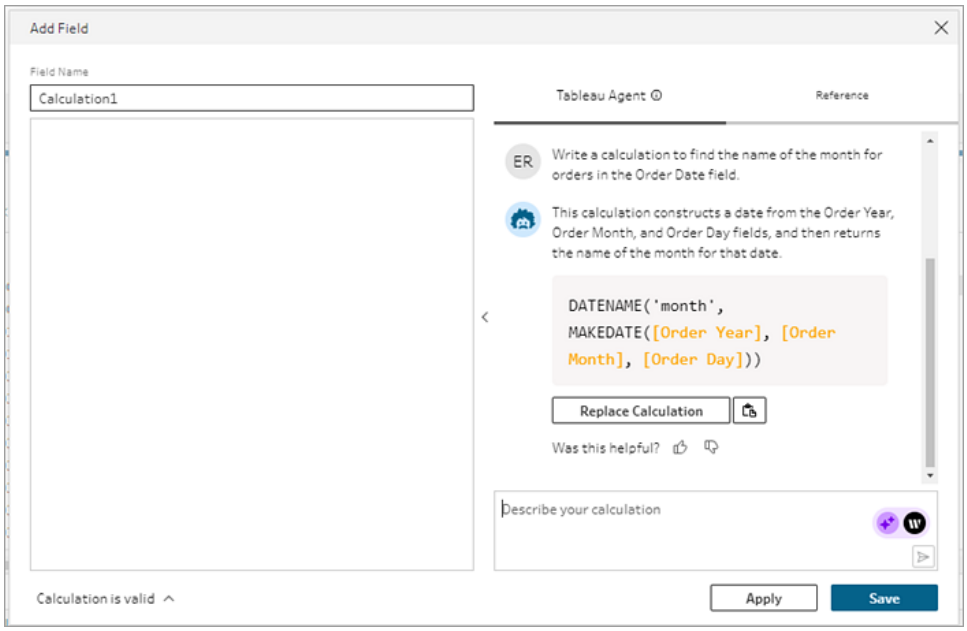
Tableau Agent 仍在学习如何构建准确的 Tableau Prep 计算。Tableau Agent 创建的一些计算不会按照您预期的方式编写。当这种情况发生时,您对计算错误的反馈有助于 Tableau Agent 继续改进并建立越来越准确的计算。

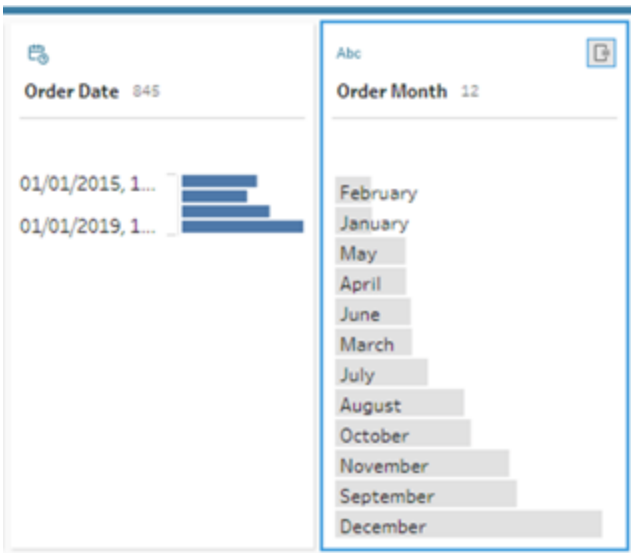


示例

计算示例 1:

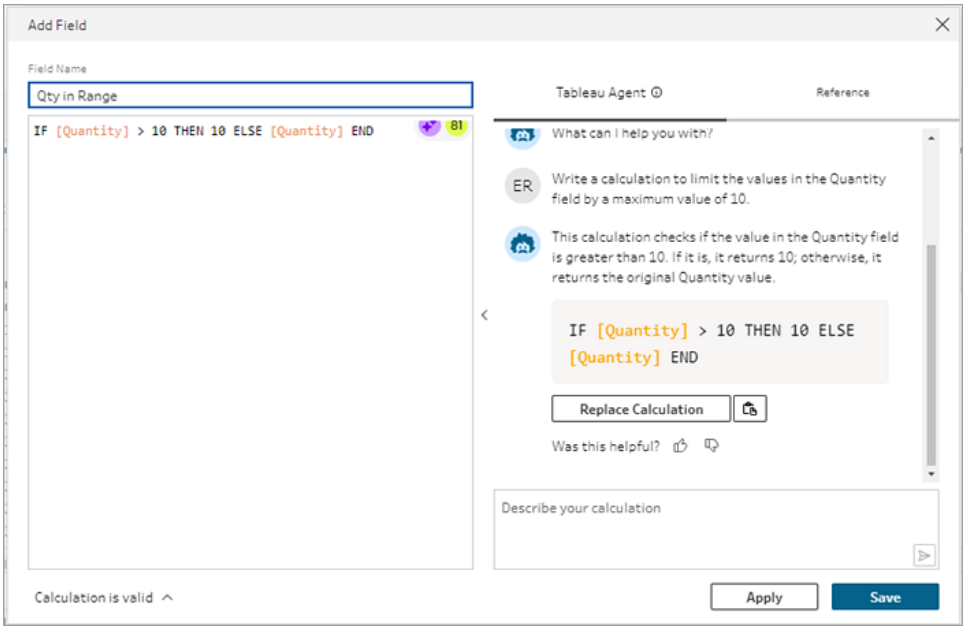
提示:*编写一个计算来查找订单日期字段中订单的月份名称。*





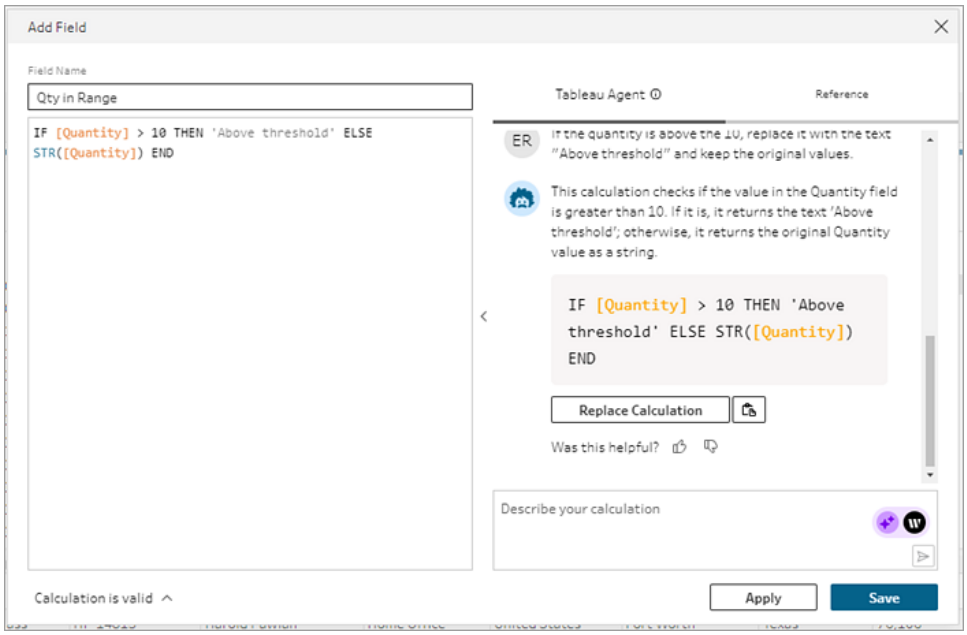
计算示例 2:

提示:*编写一个计算，将限制数量字段中的值的最大值限制为 10。*



在此示例中，添加了额外的提示来优化计算以保留高于阈值的值。

提示:*如果数量超过 10，则将其替换为文本“超过阈值”并保留原始值。*



转置您的数据

注意：从版本 2020.4.1 开始，您现在可以在 Tableau Server 和 Tableau Cloud 中创建和编辑流程。除非特别指出，本主题中的内容适用于所有平台。有关在 Web 上制作流程的详细信息，请参见 Tableau Server 和 Tableau Cloud 帮助中的 **Web 上的 Tableau Prep**。

有时候，使用Tableau分析电子表格或者横向表格（交叉数据）的数据会遇到一些困难。Tableau更倾向于数据是一行一行的“高高的数据”（行数据），而不是一列一列的“很宽很长的数据”（列数据），这就需要我们把列数据转置为行数据。

但是，可能还会有一些数据表既长又窄，并且太过标准化而无法正常分析的情况。例如，销售部门在两个列中跟踪广告支出，一个列名为“**Advertising**”（广告），其中包含表示“**radio**”（广播广告）、“**television**”（电视广告）和“**print**”（印刷广告）的行，而另一个列则表示总支出。在这种情况下，若要将此数据作为单独的度量进行分析，您将需要将该行数据转置为列。

但如果转置较大的数据集或在一段时间内频繁更改的数据，会怎么样？ 您可以使用通配符模式匹配来搜索与模式匹配的字段，并自动转置数据。

在转置数据时，使用以下选项之一：

- 将列转置为行
- 使用通配符搜索，根据模式匹配即时转置字段（Tableau Prep Builder 版本 2019.1.1 及更高版本和在 Web 上）。
- 将行转置为列（Tableau Prep Builder 版本 2019.1.1 及更高版本和在 Web 上）。

不管如何转置字段，您都可以直接与结果交互，并执行任何额外的清理操作，让数据看起来恰如您所需。您还可以使用 Tableau Prep 的智能默认命名功能来自动重命名已转置的字段和值。

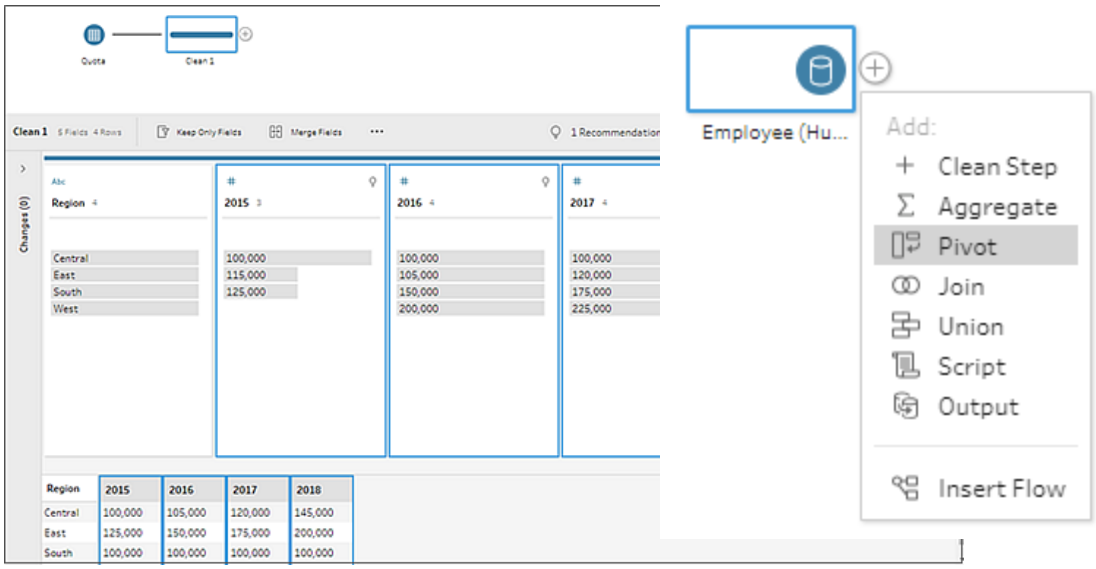
将列转置为行

使用此转置选项将宽数据转换为高数据。在一个或多个字段组上将列转置为行。选择要处理的字段，并将数据从列转置为行。

1. 连接到您的数据源。
2. 将要进行转置的表拖到**“流程”**窗格。
3. 执行以下操作之一：
   - **Tableau Prep Builder 版本 2019.4.2 及更高**版本和在 Web 上：在“配置”窗格中，选择要转置的字段，然后右键单击或按住 Ctrl 单击（MacOS），并从菜单中选择**“将列转置为行”**。如果使用此选项，请跳到步骤 7。
   - **所有版本：**单击加号 ⊕ 图标，并从上下文菜单中选择**“添加转置”**。

选择字段（Tableau Prep Builder 版本 2019.4.2 及 流程步骤菜单（所有版本）

更高版本和在 Web 上)



4. (可选)在**“字段”**窗格的**“搜索”**字段中输入一个值,以在字段列表中搜索要转置的字段。
5. (可选)选中**“自动重命名转置的字段和值”**复选框,使 Tableau Prep 能够使用数据中的常用值重命名新的转置字段。如果找不到常用值,则使用默认名称。
6. 从左侧窗格中选择一个或多个字段,并将这些字段拖到**“转置的字段”**窗格中的**“转置 1 值”**列。
7. (可选)在**“转置的字段”**窗格中,单击加号 ⊕ 图标添加要进行转置的更多列,然后重复上面的步骤以选择要进行转置的更多字段。结果将立即出现在**“转置结果”**窗格和数据窗格中。

   注意:您选择的字段数必须与在步骤 5 中选择的字段数相同。举例来说,如果选择了 3 个字段对其进行最初转置,则您进行转置的每个后续列也必须包含 3 个字段。
8. 如果未启用默认命名选项或者 Tableau Prep 无法自动检测名称,请编辑字段的名称。您也可以在此窗格中编辑原始字段的名称,以对数据进行最恰当地描述。
9. (可选)重命名新的“转置”步骤以跟踪更改。例如“转置月份”。

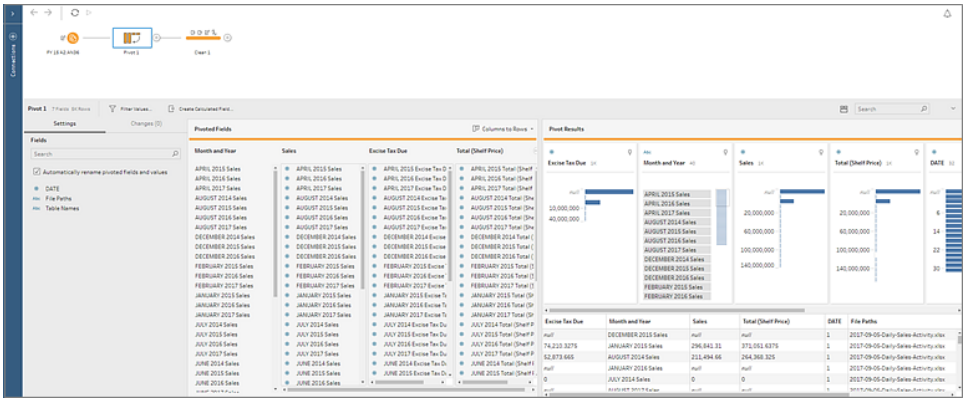
10. 若要在数据更改时刷新转置数据，请运行流程。如果将需要添加到转置的新字段添加到数据源，请将它们手动添加到转置。

示例：对多个字段进行转置

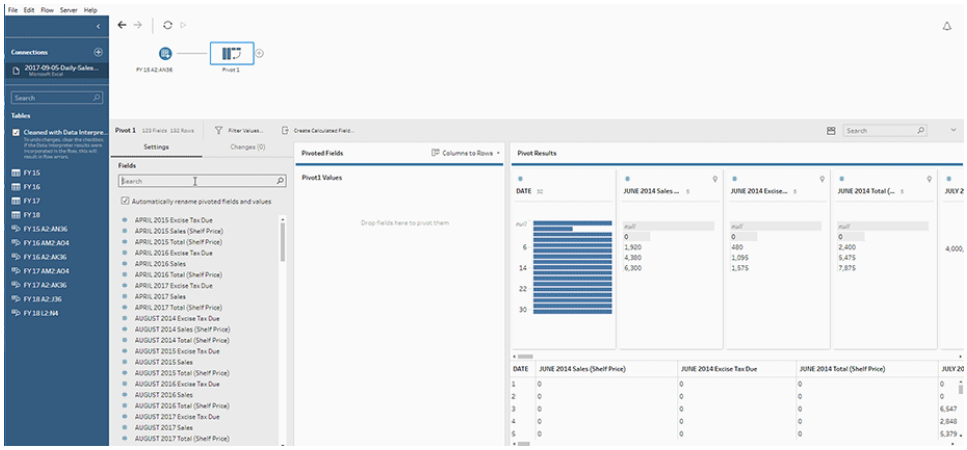
此示例按月和年显示药品销售额、税款和总计的电子表格。

	A	T	U	V	W	X	Y	Z	AA	AB
2	DATE	DECEMBER 2014			JANUARY 2015			FEBRUARY 2015		
3		Sales (Shelf Price)	Excise Tax Due	Total (Shelf Price	Sales (Shelf Price)	Excise Tax Due	Total (Shelf Price	Sales (Shelf Price)	Excise Tax Due	Total (Shelf Price
4	1	$ 448,111	$ 112,028	$ 560,139	$ 296,841	$ 74,210	$ 371,052	$ 212,919	$ 53,230	$ 266,148
5	2	$ 425,472	$ 106,368	$ 531,840	$ 754,061	$ 188,515	$ 942,577	$ 449,897	$ 112,474	$ 562,371
6	3	$ 435,525	$ 108,881	$ 544,406	$ 482,497	$ 120,624	$ 603,121	$ 627,711	$ 156,928	$ 784,639
7	4	$ 634,765	$ 158,691	$ 793,456	$ 332,228	$ 83,057	$ 415,284	$ 688,263	$ 172,066	$ 860,329
8	5	$ 695,425	$ 173,856	$ 869,282	$ 601,529	$ 150,382	$ 751,912	$ 789,233	$ 197,308	$ 986,541
9	6	$ 436,720	$ 109,180	$ 545,899	$ 527,374	$ 131,843	$ 659,217	$ 867,501	$ 216,875	$ 1,084,377
10	7	$ 238,481	$ 59,620	$ 298,101	$ 560,102	$ 140,026	$ 700,128	$ 554,459	$ 138,615	$ 693,074
11	8	$ 421,422	$ 105,356	$ 526,778	$ 539,974	$ 134,993	$ 674,967	$ 448,846	$ 112,211	$ 561,057
12	9	$ 543,816	$ 135,954	$ 679,770	$ 683,408	$ 170,852	$ 854,260	$ 768,266	$ 192,067	$ 960,333
13	10	$ 616,271	$ 154,068	$ 770,339	$ 442,352	$ 110,588	$ 552,940	$ 719,637	$ 179,909	$ 899,546
14	11	$ 756,542	$ 189,135	$ 945,677	$ 288,605	$ 72,151	$ 360,756	$ 1,154,919	$ 288,730	$ 1,443,649
15	12	$ 726,270	$ 181,567	$ 907,837	$ 674,121	$ 168,530	$ 842,651	$ 1,019,936	$ 254,984	$ 1,274,921
16	13	$ 477,208	$ 119,302	$ 596,510	$ 526,451	$ 131,613	$ 658,064	$ 951,242	$ 237,811	$ 1,189,053
17	14	$ 245,898	$ 61,475	$ 307,373	$ 573,842	$ 143,461	$ 717,303	$ 798,392	$ 199,598	$ 997,991
18	15	$ 456,254	$ 114,064	$ 570,318	$ 658,952	$ 164,738	$ 823,690	$ 453,091	$ 113,273	$ 566,364

通过对数据进行转置，您可以为每个月份和年度创建行，并为销售额、税款和总计创建单独的列，以便 Tableau 能够更轻松地解读此数据来进行分析。



观看“对多个字段进行转置”的实际运行效果。

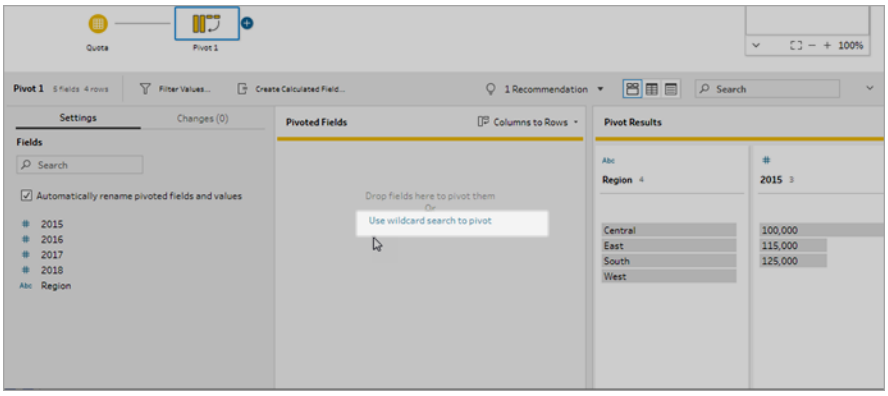


使用通配符搜索进行转置

如果使用大数据集或者数据在一段时间内频繁更改，从 Tableau Prep Builder 版本 2019.1.1 开始和在 Web 上，您可以在将列转置为行时使用通配符搜索，以便根据通配符模式匹配即时转置数据。

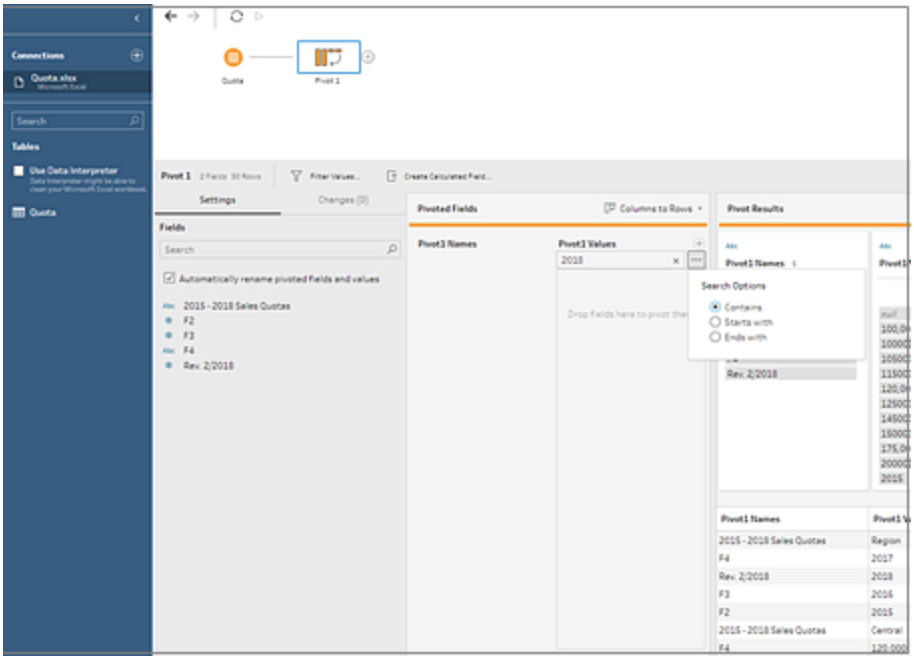
如果添加或移除了与模式匹配的新字段，Tableau Prep 会在运行流程时检测到架构更改，并且转置结果会自动更新。

1. 连接到您的数据源。
2. 将要进行转置的表拖到**“流程”**窗格。
3. 单击加号 ⊕ 图标，并从上下文菜单中选择**“添加转置”**。
4. 在**“转置的字段”**窗格中，单击**“使用通配符搜索进行转置”**链接。



5. 输入要搜索的值或部分值。例如，输入“**Sales_**”以匹配标为“**sales_2017**”(2017 年销售额)、“**sales_2018**”(2018 年销售额)和“**sales_2019**”(2019 年销售额)的字段。

   不要使用星号来匹配模式，除非星号是所搜索的字段值的一部分。相反，请单击**“搜索选项”** 按钮来选择所需的值匹配方式。然后按 **Enter** 应用搜索并转置匹配的值。



6. (可选)在**“转置的字段”**窗格中，单击加号 图标添加要进行转置的更多列，然后重复上面的步骤以选择要进行转置的更多字段。

7. 如果未启用默认命名选项或者 Tableau Prep 无法自动检测名称，请编辑字段的名称。
8. 若要在数据更改时刷新转置数据，请运行流程。系统会自动检测到添加到数据源并与通配符模式匹配的任何新字段，并将其添加到转置。
9. 如果结果不是预期结果，请尝试以下选项之一：
    - 在“搜索”字段中输入其他值模式，并按 Enter。转置将自动刷新并显示新结果。
    - 将其他字段手动拖到**“转置的字段”**窗格中的**“转置 1 值”**列。您也可以移除手动添加的字段，方法是将它们拖离**“转置 1 值”**列并放到**“字段”**窗格中。

      注意：对于依据通配符搜索结果添加的字段，无法通过将它们拖离**“转置 1 值”**列来移除。作为替代，请尝试使用更具体的模式来匹配所查找的搜索结果。

将行转置为列

在 Tableau Prep Builder 版本 2019.1.1 及更高版本中和 Web 上，如果数据过于规范化，并且需要创建新列，请将行转置为列 - 从高数据转置为更宽的数据。

举例来说，假设您有每个月的广告成本数据，其中每个月都在一列中包括了所有广告类型。如果您将数据从行转置为列，则每种广告类型会变为拥有单独的一列，从而使数据更易于分析。

Month	Cost	Advertising Type
January	12187	Radio
January	45698	Television
January	9762	Print
February	4598	Print
February	23987	Television
February	8907	Radio
March	10987	Print

Month	Radio	Television	Print
January	12187	45698	9762
February	8907	23987	4598
March	Null	Null	10987

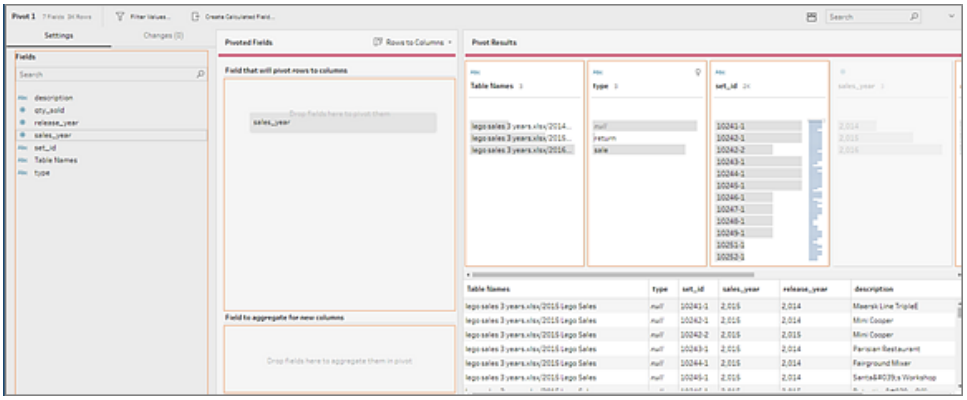
您可以选择一个要转置的字段。随后将使用该字段的字段值来创建新列。然后，选择一个要用于填充新列的字段。这些字段值是聚合的，您可要选择要应用的聚合类型。

由于应用了聚合，因此将列转置回行将不会反向进行此转置操作。若要逆转行到列转置类型，您将需要撤消操作。在顶部菜单中单击**“撤消”**按钮、从**“转置的字段”**窗格中移除字段，或删除转置步骤。

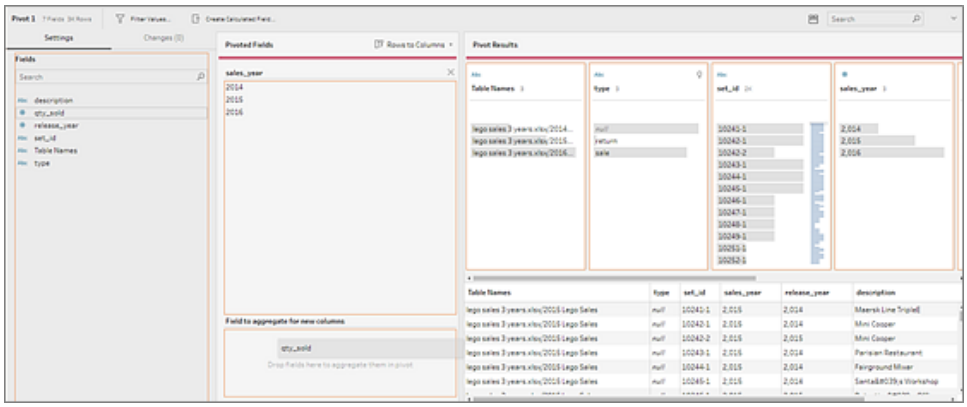
1. 连接到您的数据源。
2. 将要进行转置的表拖到**“流程”**窗格。
3. 单击加号 ⊕ 图标，并从上下文菜单中选择**“添加转置”**。
4. 在**“转置的字段”**窗格中，从下拉列表中选择**“行转置为列”**。
5. (可选)在**“字段”**窗格的**“搜索”**字段中输入一个值，以在字段列表中搜索要转置的字段
6. 从左侧窗格中选择一个字段，并将其拖到**“转置的字段”**窗格中的**“将从行转置为列的字段”**部分。

   注意：如果要转置的字段具有日期或日期时间数据类型，则需要将数据类型更改为字符串才能对其进行转置。

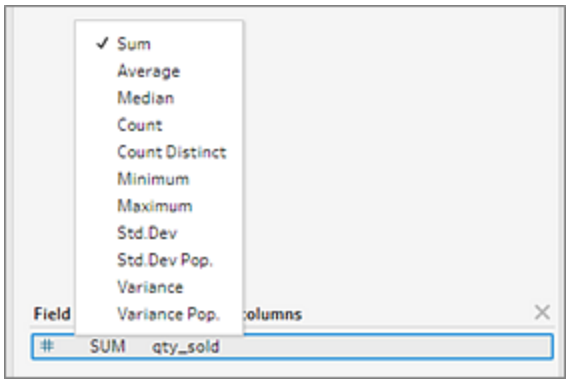
   此字段中的值将用于创建和命名新列。您可以稍后在**“转置结果”**窗格中更改列名称。



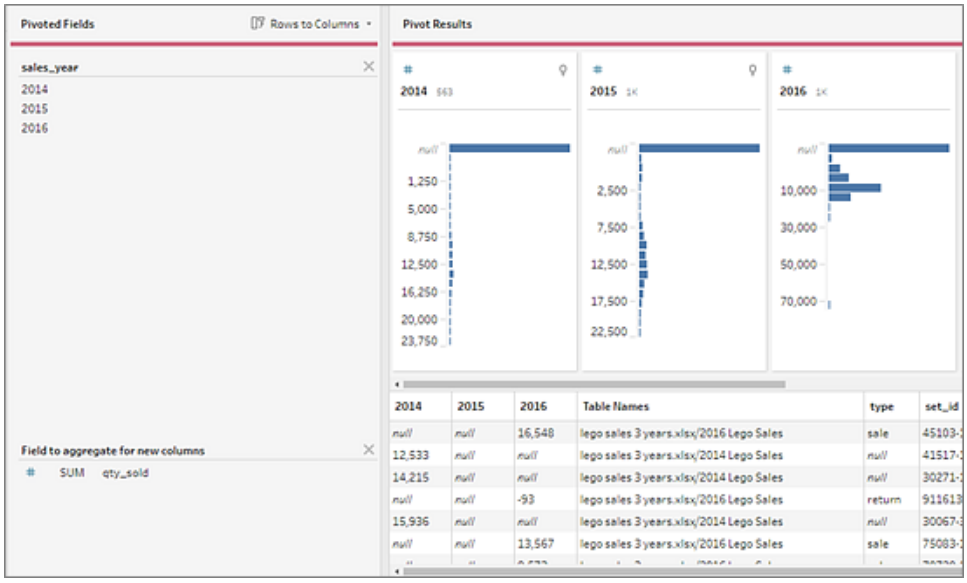
7. 从左侧窗格中选择一个字段，并将其拖到**“转置的字段”**窗格中的**“要为新列聚合的字段”**部分。此字段中的值用于填充通过上一步创建的新列。



系统会为字段分配一个默认聚合类型。单击聚合类型可对其进行更改。



8. 在**"转置结果"**窗格中，查看结果并向您创建的新列应用任何清理操作。

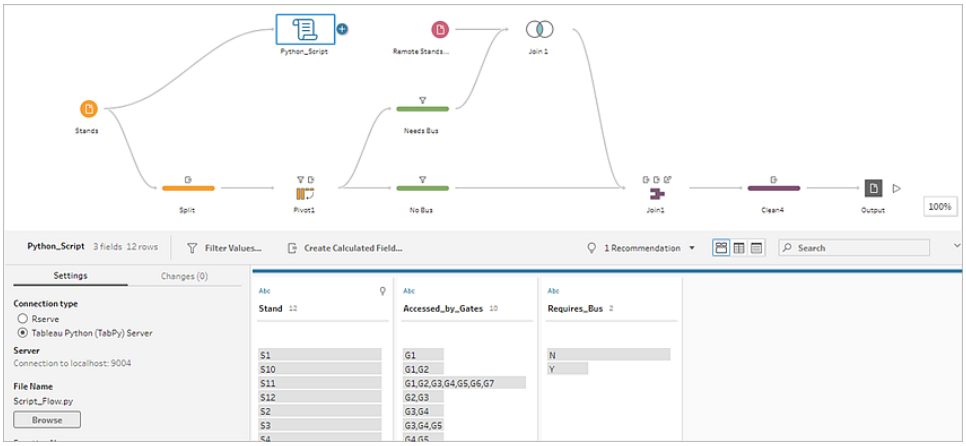


9. 如果正在转置的字段的行数据发生变化，请在流程窗格中右键单击或按住 Ctrl 并单击 (MacOS)“转置”步骤，并选择**“刷新”**。

在流程中使用 R 和 Python 脚本

从版本 2019.3.1 开始，您可以使用 R 和 Python 脚本来执行更复杂的清理操作，或者在流程中并入预测建模数据。数据通过 R 或 Python 脚本步骤以输入形式从流程中传递，然后以您可以使用 Tableau Prep Builder 的功能和函数继续清理的输出数据的形式返回。

注意：尚不支持以流程输入步骤的形式连接到脚本。此外，对于制作或发布到 Tableau Cloud 的流程，尚不支持脚本步骤。



配置 Rserve 服务器或 Tableau Python (TabPy) 服务器，并将脚本步骤添加到流程中。Tableau Prep 将数据传递到 Rserve for R 或用于 Python 的 Tableau Python 服务器 (TabPy)，然后以表的形式将生成的数据返回流程。您可以继续将清理操作应用到结果并生成分析输出。

创建脚本时，需要包含一个函数，该函数将数据框架指定为函数的参数。如果要返回不同于输入的字段，则需要在脚本中包含定义输出和数据类型的 getOutputSchema 函数。否则，输出将使用输入数据中的字段。

如果您在 Tableau Server(版本 2020.4.1 及更高版本) 中制作或编辑包含脚本步骤的流程，Tableau Server 还必须连接到 Rserve 或 TabPy 服务器才能运行脚本步骤。有关如何配置要在流程中使用的 R 或 Python 以及如何创建脚本的信息，请参见**在您的流程中使用 R (Rserve) 脚本** 在对页或**在您的流程中使用 Python 脚本** 在本页 330。

在您的流程中使用 R (Rserve) 脚本

免责声明：本主题包括有关第三方产品的信息。请注意，尽管我们努力保持第三方内容的准确性，但我们在此处提供的信息可能会在 R 和 Rserve 发生变化时更改，恕不另行通知。有关最新信息，请查阅 R 和 Rserve 文档及支持。

R 是一种用于统计计算和图形的开源软件编程语言和软件环境。为了扩展 Tableau Prep Builder 的功能，您可以用 R 创建脚本以在通过 Rserve 服务器运行的流程中使用，生成可在流程中进一步处理的输出。

例如，您可能想要使用 R 语言脚本向流程中已有的数据中添加统计建模数据或预测数据，然后使用 Tableau Prep Builder 的功能清理生成的数据集以进行分析。

若要在流程中包括 R 脚本，您需要在 Tableau Prep Builder 和 Rserve 服务器之间配置连接。然后，您可以使用 R 脚本，通过 R 表达式将支持的函数应用于流程中的数据。输入配置详细信息并将 Tableau Prep Builder 指向要使用的文件和函数后，系统会将数据安全地传递到 Rserve 服务器，应用表达式，并以可根据需要清理或输出的表 (R data.frame) 的形式返回结果。

只要配置了与 Rserve 服务器的连接，就可以在 Tableau Server 中运行包含脚本步骤的流程。当前不支持在 Tableau Cloud 中运行包含脚本步骤的流程。若要配置 Tableau Server，请参见**针对 Tableau Server 配置 Rserve Server** 向下。

先决条件

若要在流程中包括 R 脚本步骤，请安装 R 并配置与 Rserve 服务器的连接。

资源

- 下载并安装 R：下载并安装适用于 Linux、Mac 或 Windows 的最新版本的 R。
- R 实施说明（社区帖子）。针对 Windows 安装并配置与 R 和 Rserve 的连接。
- 安装并配置 Rserve：适用于所有平台的常规安装和配置说明。
    - Rserve for Windows（版本说明）：本主题讨论在 Windows 上以本地方式安装 Rserve 时的限制。

针对 Tableau Server 配置 Rserve Server

使用以下说明配置 Rserve 服务器和 Tableau Server 之间的连接。

- **版本 2019.3 及更高版本**：您可以在 Tableau Server 中运行包含脚本步骤的已发布流程。

- **版本 2020.4.1 及更高版本**：您可以在 Tableau Server 中创建、编辑和运行包含脚本步骤的流程。
- **Tableau Cloud**：当前不支持创建或运行包含脚本步骤的流程。

1. 打开 TSM 命令行。
2. 输入以下命令以设置主机地址、端口值和连接超时：

   ```
   tsm security maestro-rserve-ssl enable --connection-type
   {maestro-rserve-secure/maestro-rserve} --rserve-host <Rserve IP
   address or host name> --rserve-port <Rserve port> --rserve-
   username <Rserve username> --rserve-password <Rserve password> --
   rserve-connect-timeout-ms <RServe connect timeout>
   ```

   - 选择 {maestro-rserve-secure} 以启用安全连接，或选择 {maestro-rserve} 以启用不安全的连接。
   - 如果选择 {maestro-rserve-secure}，请在命令行中指定证书文件 -cf<certificate file path>。
   - 以毫秒为单位指定 --rserve-connect-timeout-ms <RServe connect timeout>。例如，--rserve-connect-timeout-ms 900000。
3. 若要禁用 Rserve 连接，请输入以下命令

   ```
   tsm security maestro-rserve-ssl disable
   ```

其他 Rserve 配置（可选）

您可以创建一个名为 Rserv.cfg 的文件以设置默认配置值来自定义 Rserve，并将其放在 /etc/Rserve.conf 安装位置中。为了提高 Rserve 服务器和 Tableau Prep Builder 的稳定性，您可以向 Rserve 配置中添加其他值。启动 Rserve 时，您可以引用此文件来应用配置选项。例如：

- Windows：`Rserve(args="--RS-conf C:\\folder\\Rserv.cfg")`
- MacOS 和 Linux：`Rserve(args=" --no-save --RS-conf ~/Documents/Rserv.cfg")`

下面的示例演示了一些可包括在 Rserve.conf 配置文件中的其他选项：

```
# If your data includes characters other than ASCII, make it explicit
that data should be UTF8 encoded.
encoding utf8
# Disable interactive behavior for Rserve or Tableau Prep Builder
will stall when trying to run the script as it waits for an input
```

```
response.
interactive no
```

有关设置 Rserve.conf 文件的信息，请参见 R 实施说明（社区帖子）中的“高级 Rserve 配置”部分。

创建 R 脚本

创建脚本时，请包括一个指定数据框架作为函数参数的函数。这将从 Tableau Prep Builder 中调用您的数据。您还需要使用支持的数据类型在数据框架中返回结果。

例如：

```
postal_cluster <- function(df) {
  out <- kmeans(cbind(df$Latitude, df$Longitude), 3, iter.max=10)
  return(data.frame(Latitude=df$Latitude, Longitude=df$Longitude,
Cluster=out$cluster))
 }
```

支持以下数据类型：

Tableau Prep Builder 中的数据类型	**R 中的数据类型**
字符串	标准 UTF-8 字符串
十进制	双精度
整数	整数
布尔值	逻辑
日期	ISO_DATE 格式“YYYY-MM-DD”的字符串，带有可选区域偏移。例如，“2011-12-03+01:00”是有效日期。
日期时间	ISO_DATE_TIME 格式“YYYY-MM-DDT:HH:mm:ss”的字符串，带有可选区域偏移。例如，“2011-12-03T10:15:30+01:00”是有效日期。

注意：日期和日期时间必须始终以有效字符串形式返回。不支持将 R 中的本机日期（日期时间）类型作为返回值，但可在脚本中使用。

如果要返回不同于输入的字段，则需要在脚本中包含定义输出和数据类型的 getOutputSchema 函数。否则，输出将使用输入数据中的字段，这些字段是从流程中脚本步骤之前的步骤中获取的。

在 getOutputSchema 中指定字段的数据类型时，请使用以下语法：

R 中的函数	生成的数据类型
prep_string ()	字符串
prep_decimal ()	十进制
prep_int ()	整数
prep_bool ()	布尔值
prep_date ()	日期
prep_datetime ()	日期时间

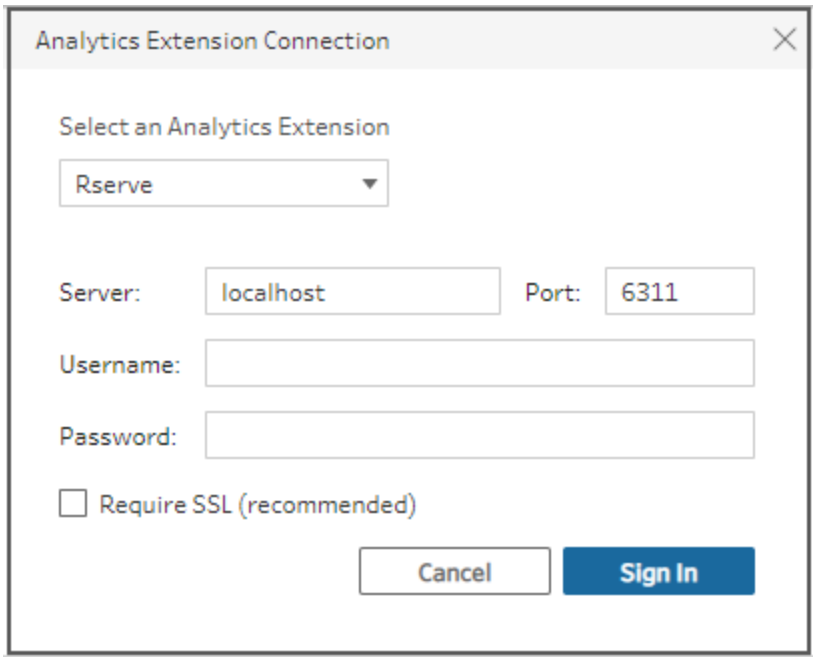
以下示例显示 postal_cluster 脚本的 getOutputSchema 函数：

```
getOutputSchema <- function() {
  return (data.frame (
    Latitude = prep_decimal (),
    Longitude = prep_decimal (),
    Cluster = prep_int ()));
}
```

连接到您的 Rserve 服务器

重要信息：从 Tableau Prep Builder 版本 2020.3.3 开始，通过顶部**“帮助”**菜单配置一次服务器连接，而不是通过在“脚本”步骤中单击**“连接到 Rserve 服务器”**并输入连接详细信息，设置每个流程的连接。对于在版本 2020.3.3 中打开的 Tableau Prep Builder 旧版本中创建的任何流程，您需要使用此新菜单重新配置连接。

1. 选择**“帮助”>“设置和性能”>“管理分析扩展程序连接”**。
2. 在**“选择分析扩展程序”**下拉列表中，选择**“Rserve”**。



3. 输入您的凭据：
   - 端口 6311 是纯文本 Rserve 服务器的默认端口。
   - 端口 4912 是 SSL 加密 Rserve 服务器的默认端口。
   - 如果服务器需要凭据，请输入用户名和密码。
   - 如果服务器使用 SSL 加密，请选中**“需要 SSL”**复选框，然后单击**“自定义配置文件”**链接，为连接指定证书。

     注意：Tableau Prep Builder 未提供测试连接的方式。如果连接存在问题，则会在您尝试并运行流程时显示一条错误消息。

向流程中添加脚本

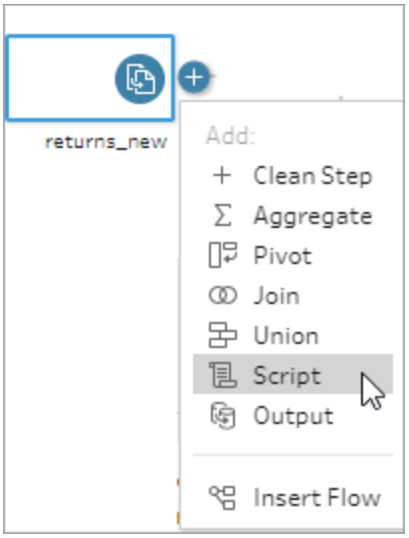
启动 Rserve 服务器，然后完成以下步骤：

1. 打开 Tableau Prep Builder 并单击**“添加连接”**按钮。

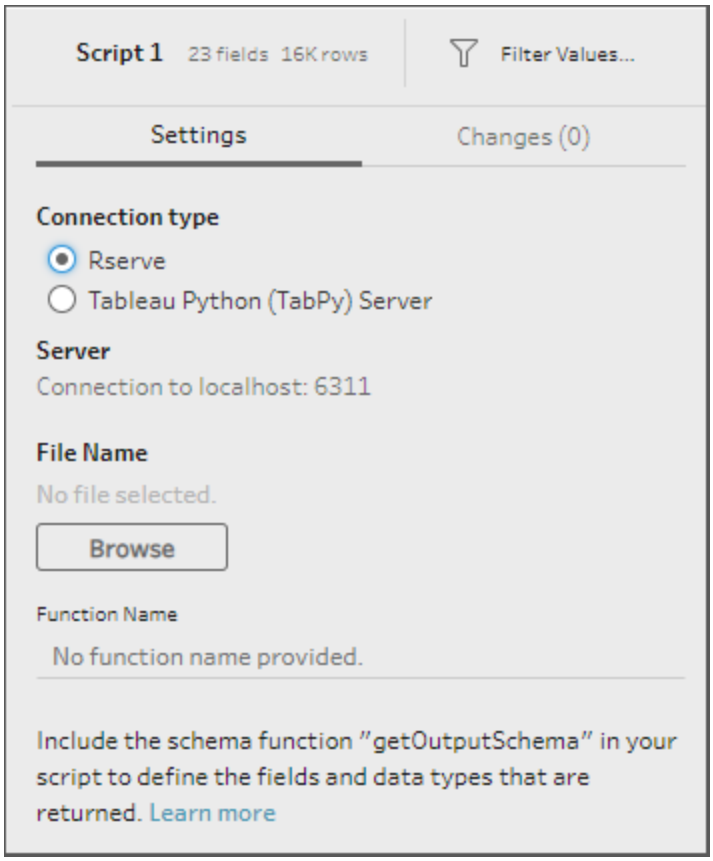
   在 Web 制作中，从**“主页”**中单击**“创建”>“流程”**，或从**“浏览”**页面中单击**“新建”>“流程”**。然后单击**“连接到数据”**。

2. 从连接器列表中，选择文件类型或托管您的数据的服务器。如果出现提示，请输入所需的信息以登录并访问数据。

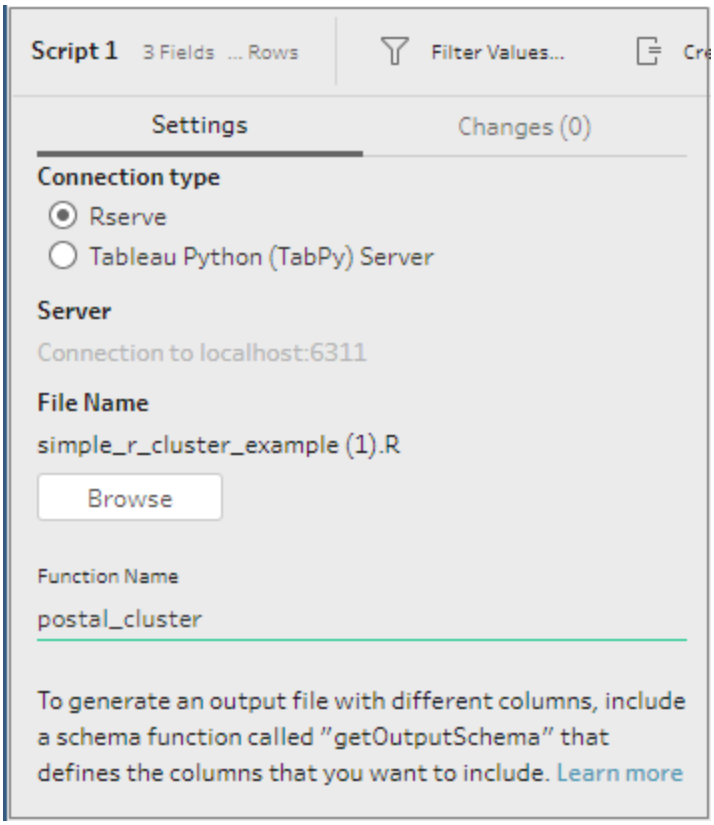
3. 单击加号 ⊕ 图标，并从上下文菜单中选择**“添加脚本”**。



4. 在“脚本”窗格中的**“连接类型”**下，选择**“Rserve”**。



5. 在**“文件名”**部分，单击**“浏览”**选择您的脚本文件。
6. 输入**“函数名称”**，然后按 **Enter** 运行脚本。



在您的流程中使用 Python 脚本

免责声明：本主题包括有关第三方产品的信息。请注意，尽管我们努力保持第三方内容的准确性，但我们在此处提供的信息可能会在 Python 发生变化时更改，恕不另行通知。有关最新信息，请查阅 Python 文档和支持。

Python 是广泛使用的高级编程语言，用于一般用途编程。通过借助 Tableau Prep Builder 将 Python 命令发送到外部服务，您可以执行诸如添加行号、进行字段排名、填写字段，以及执行可使用计算字段以其他方式执行的清理操作，从而轻松地扩展数据准备选项。

若要在流程中包括 Python 脚本，您需要在 Tableau 和 TabPy 服务器之间配置连接。然后，您可以使用 Python 脚本，通过 pandas 数据框架将支持的函数应用于流程中的数据。向流程中添加脚本步骤并指定要使用的配置详细信息、文件和函数时，系统会将数据安全传递给 TabPy 服务器，应用脚本中的表达式，并以可根据需要清理或输出的表的形式返回结果。

只要配置了与 TabPy 服务器的连接，就可以在 Tableau Server 中运行包含脚本步骤的流程。当前不支持在 Tableau Cloud 中运行包含脚本步骤的流程。若要配置 Tableau Server，请参见**为 Tableau Server 配置 Tableau Python (TabPy) 服务器** 向下。

有关如何在 Tableau Server 上使用工作簿的分析扩展程序配置站点的信息，请参见配置与分析扩展程序的连接。

先决条件

若要在您的流程中包括 Python 脚本，请完成以下设置。当前不支持在 Tableau Cloud 中创建或运行包含脚本步骤的流程。

1. 下载并安装 Python。下载并安装适用于 Linux、Mac 或 Windows 的最新版本的 Python。
2. 下载并安装 Tableau Python 服务器 (TabPy)。按照用于安装 TabPy 的安装和配置说明进行操作。Tableau Prep Builder 使用 TabPy 以输入形式通过 TabPy 传递流程中的数据，应用您的脚本，然后将结果返回到流程。
3. 安装 Pandas。运行 `pip3 install pandas`。您必须在脚本中使用 pandas 数据框架来与 Tableau Prep Builder 集成。

为 Tableau Server 配置 Tableau Python (TabPy) 服务器

如果您计划在 Tableau Server 中发布、创建、编辑和运行包含脚本步骤的流程，您将需要配置 TabPy 服务器和 Tableau Server 之间的连接。

- **版本 2019.3 及更高版本**：您可以在 Tableau Server 中运行包含脚本步骤的已发布流程。
- **版本 2020.4.1 及更高版本**：您可以在 Tableau Server 中创建、编辑和运行包含脚本步骤的流程。
- **Tableau Cloud**：当前不支持创建或运行包含脚本步骤的流程。

1. 打开 TSM 命令行/shell。
2. 输入以下命令以设置主机地址、端口值和连接超时：

```
tsm security maestro-tabpy-ssl enable --connection-type {maestro-tabpy-secure/maestro-tabpy} --tabpy-host <TabPy IP address or host name> --tabpy-port <TabPy port> --tabpy-username <TabPy username> --tabpy-password <TabPy password> --tabpy-connect-timeout-ms <TabPy connect timeout>
```

   - 选择 `{maestro-tabpy-secure}` 以启用安全连接，或选择 `{maestro-tabpy}` 以启用不安全的连接。

- 如果选择 {maestro-tabpy-secure}，请在命令行中指定证书文件 -cf<certificate file path>。
- 以毫秒为单位指定 --tabpy-connect-timeout-ms <TabPy connect timeout>。例如，--tabpy-connect-timeout-ms 900000。

3. 若要禁用 TabPy 连接，请输入以下命令

```
tsm security maestro-tabpy-ssl disable
```

创建 Python 脚本

创建脚本时，请包括一个指定 pandas (pd.DataFrame) 作为函数参数的函数。这将从 Tableau Prep Builder 中调用您的数据。您还需要使用支持的数据类型在 pandas (pd.DataFrame) 中返回结果。

例如，若要向流程中的一组字段添加编码，可以编写以下脚本：

```
def encode(input):
  le = preprocessing.LabelEncoder()
  Return pd.DataFrame({
    'Opportunity Number' : input['Opportunity Number'],
    'Supplies Subgroup Encoded' : le.fit_transform(input['Supplies Subgroup']),
    'Region Encoded' : le.fit_transform(input['Region']),
    'Route To Market Encoded' : le.fit_transform(input['Route To Market']),
    'Opportunity Result Encoded' : le.fit_transform(input['Opportunity Result']),
    'Competitor Type Encoded' : le.fit_transform(input['Competitor Type']),
    'Supplies Group Encoded' : le.fit_transform(input['Supplies Group']),
})
```

支持以下数据类型：

Tableau Prep Builder 中的数据类型	Python 中的数据类型

字符串	标准 UTF-8 字符串
十进制	双精度
整数	整数
布尔值	布尔值
日期	ISO_DATE 格式“YYYY-MM-DD”的字符串，带有可选区域偏移。例如，“2011-12-03”是有效日期。
日期时间	ISO_DATE_TIME 格式“YYYY-MM-DDT:HH:mm:ss”的字符串，带有可选区域偏移。例如，“2011-12-03T10:15:30+01:00”是有效日期。

注意：日期和日期时间必须始终以有效字符串形式返回。

如果要返回不同于输入的字段，则需要在脚本中包含定义输出和数据类型的 get_output_schema 函数。否则，输出将使用输入数据中的字段，这些字段是从流程中脚本步骤之前的步骤中获取的。

在 get_output_schema 中指定字段的数据类型时，请使用以下语法：

Python 中的函数	**生成的数据类型**
prep_string()	字符串
prep_decimal()	十进制
prep_int()	整数
prep_bool()	布尔值
prep_date()	日期
prep_datetime()	日期时间

方括号可用于访问 DataFrame 中的行。单括号输出 Pandas Series，而双括号将输出 Pandas DataFrame。

```
def create_key(df):
return pd.DataFrame({
```

```
"Key": ['12345']
})
```

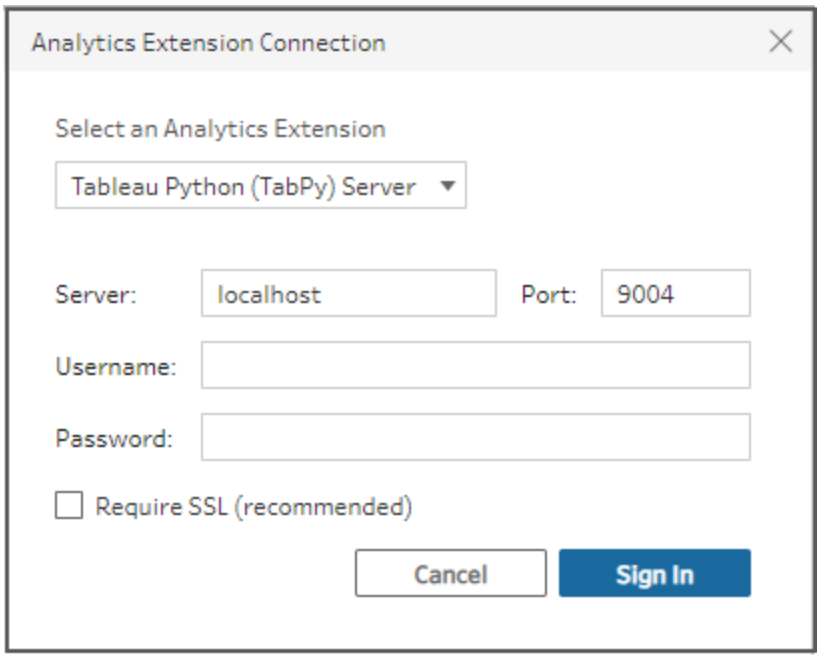
下面的示例演示添加到字段编码 Python 脚本的 get_output_schema 函数：

```
def get_output_schema():
  return pd.DataFrame({
    'Opportunity Number' : prep_int(),
    'Supplies Subgroup Encoded' : prep_int(),
    'Region Encoded' : prep_int(),
    'Route To Market Encoded' : prep_int(),
    'Opportunity Result Encoded' : prep_int(),
    'Competitor Type Encoded' : prep_int(),
    'Supplies Group Encoded' : prep_int()
})
```

连接到 Tableau Python (TabPy) 服务器

重要信息：从 Tableau Prep Builder 版本 2020.3.3 开始，您可以通过顶部**“帮助”**菜单配置一次服务器连接，而不是通过在“脚本”步骤中单击**“连接到 Tableau Python (TabPy) 服务器”**并输入连接详细信息，设置每个流程的连接。对于在版本 2020.3.3 中打开的 Tableau Prep Builder 旧版本中创建的任何流程，您需要使用此新菜单重新配置连接。

1. 选择**“帮助”>“设置和性能”>“管理分析扩展程序连接”**。
2. 在**“选择分析扩展程序”**下拉列表中，选择**“Tableau Python (TabPy) 服务器”**。



3. 输入您的凭据：
   - 端口 9004 是 TabPy 的默认端口。
   - 如果服务器需要凭据，请输入用户名和密码。
   - 如果服务器使用 SSL 加密，请选中“需要 SSL”复选框，然后单击**“未指定自定义配置文件....”**链接，为连接选择证书。这是您的 SSL 服务器证书文件。

     注意：Tableau Prep Builder 未提供测试连接的方式。如果连接出现错误，将会显示一条错误消息。

向流程中添加脚本

启动 TabPy 服务器，然后完成以下步骤：

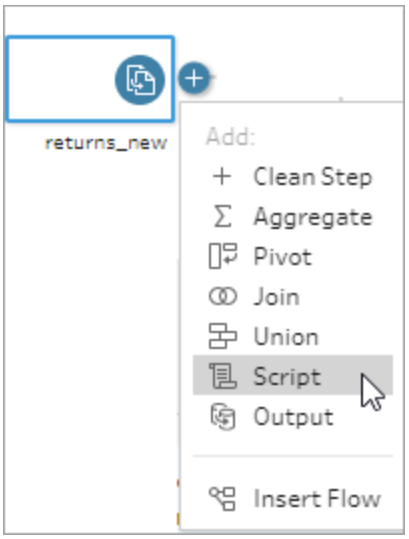
注意：TabPy 需要 tornado 程序包版本 5.1.1 才能运行。如果在尝试启动 TabPy 时遇到“'tornado.web' has no attribute 'asynchronous'”(“tornado.web”没有属性“asynchronous”) 错误，请从命令行中运行 `pip list`，检查已安装的 tornado 的版本。如果安装了其他版本，请下载 tornado 程序包版本 5.1.1。运行 `pip uninstall tornado` 卸载当前版本，然后运行 `pip install tornado==5.1.1` 安装所需的版本。

1. 打开 Tableau Prep Builder 并单击**“添加连接”**按钮。

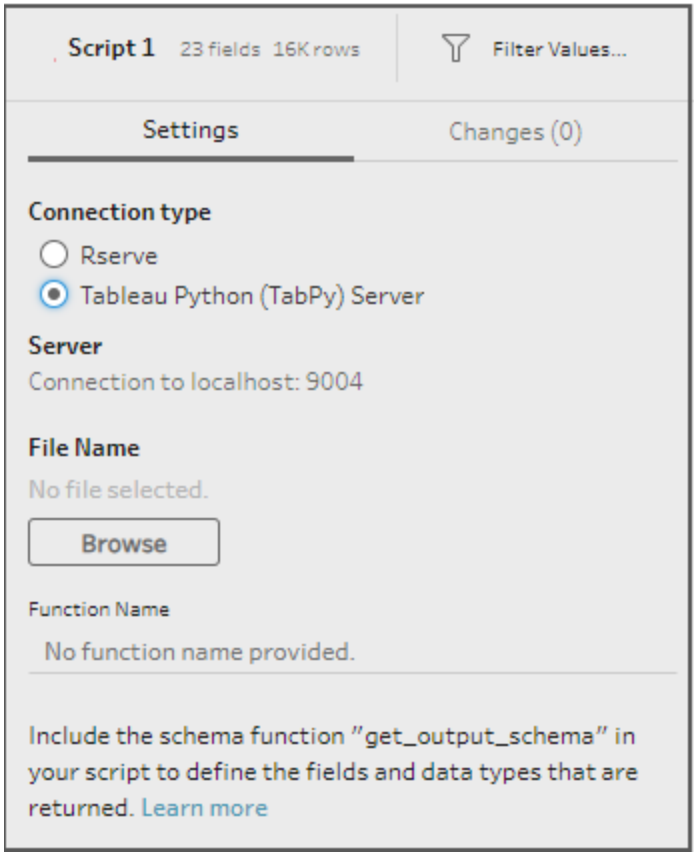
   在 Web 制作中，从**“主页”**中单击**“创建”>“流程”**，或从**“浏览”**页面中单击**“新建”>“流程”**。然后单击**“连接到数据”**。

2. 从连接器列表中，选择文件类型或托管您的数据的服务器。如果出现提示，请输入所需的信息以登录并访问数据。

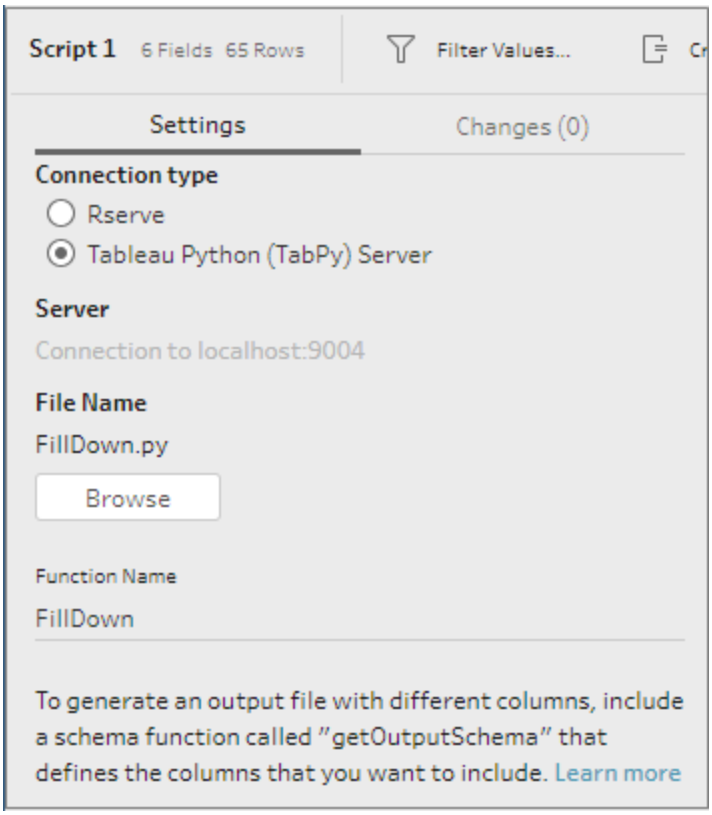
3. 单击加号图标，并从上下文菜单中选择**“添加脚本”**。



4. 在“脚本”窗格中的**“连接类型”**部分，选择**“Tableau Python (TabPy) 服务器”**。



5. 在**“文件名”**部分，单击**“浏览”**选择您的脚本文件。
6. 输入**“函数名称”**，然后按 **Enter** 运行脚本。


Script 1
6 Fields 65 Rows
Filter Values...
Settings
Changes (0)
Connection type
Rserve
Tableau Python (TabPy) Server
Server
Connection to localhost:9004
File Name
FillDown.py
Browse
Function Name
FillDown
To generate an output file with different columns, include a schema function called "getOutputSchema" that defines the columns that you want to include. Learn more

聚合、联接或合并数据

对数据进行聚合、联接或合并，以便将数据分组或合并来进行分析。

注意：从版本 2020.4.1 开始，您现在可以在 Tableau Server 和 Tableau Cloud 中创建和编辑流程。除非特别指出，本主题中的内容适用于所有平台。有关在 Web 上制作流程的详细信息，请参见 Tableau Server 和 Tableau Cloud 帮助中的 **Web 上的 Tableau Prep**。

对值进行聚合和分组

有时，您将需要调整某些数据的粒度，以便减少从流程中生成的数据量，或将数据与可能需要联接或合并在一起的其他数据对齐。例如，在将销售额表与客户表联接之前，您可能需要按客户聚合销售额数据。

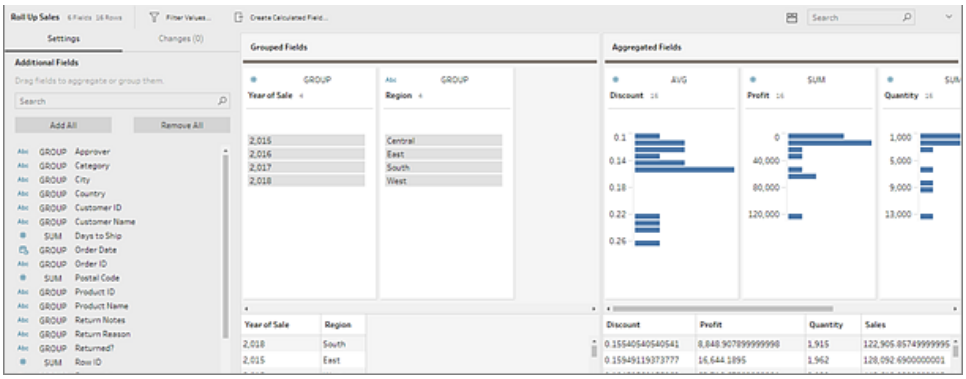
如果需要调整数据的粒度，请使用**“聚合”**选项创建用于分组和聚合数据的步骤。对数据进行聚合还是分组取决于数据类型(字符串、数字或日期)。

1. 在**“流程”**窗格中，单击加号 ⊕ 图标，并选择**“聚合”**。一个新的聚合步骤将显示在**“流程”**窗格中，并且**“配置”**窗格将更新以显示聚合和分组配置。
2. 将字段从左侧窗格拖动到**“分组字段”**窗格(构成行的字段)或拖动到**“聚合字段”**窗格(将在分组字段级别聚合和显示的数据)。

   您还可以：

   - 在两个窗格之间中拖放字段。
   - 在列表中搜索字段，并仅选择要包括在聚合中的字段。
   - 双击字段，将其添加到左侧或右侧窗格。
   - 更改字段的函数以将其自动添加到适当的窗格。
   - 单击**“全部添加”**或**“全部移除”**以批量应用或移除字段。
   - 将特定清理操作应用于字段。有关可以使用哪些清理选项的详细信息，请参见关于清理操作 在本页213。

   以下示例将显示利润和数量的总和，以及按地区和销售年份划分的平均折扣。



字段基于其数据类型分布在**“分组字段”**和**“聚合字段”**列之间。单击分组或聚合类型(例如，AVG 或 SUM)标题以更改分组或聚合类型。

在聚合和分组配置下方的数据网格中，您可以看到分组或聚合成员的示例。

系统会在**“更改”**窗格中跟踪对字段执行的所有清理操作。

联接数据

要分析的数据通常由与特定字段相关的表的集合组成。联接是一种在这些公共字段上合并相关数据的方法。使用联接合并数据后会产生一个通常通过添加数据字段横向扩展的表。

联接是一种可在流程中的任何地方执行的操作。在流程前期进行联接可帮助您了解您的数据集，并揭示出需要立即关注的区域。

Tableau Prep 支持以下联接类型：

联接类型	描述
左联接	对于每一行，包括左侧表的所有行，以及右侧表中任何对应的匹配项。当左侧表中的值在右侧表中没有对应匹配项时，您将在联接结果中看到 null 值。
内部联接	对于每一行，包括在两个表中具有匹配项的值。
右联接	对于每一行，包括右侧表中的所有值，以及左侧表中的对应的匹配项。

	当右侧表中的值在左侧表中没有对应匹配项时，您将在联接结果中看到 null 值。
仅左联接 	对于每一行，仅包括左侧表中与右侧表中的任何值不匹配的值。右侧表中的字段值在联接结果中显示为 null 值。
仅右联接 	对于每一行，仅包括右侧表中与左侧表中的任何值不匹配的值。左侧表中的字段值在联接结果中显示为 null 值。
非内部联接 	对于每一行，包括右侧和左侧表中不匹配的所有值。
完全联接 	对于每一行，包括两个表中的所有值。当任一表中的值在另一个表中没有匹配项时，您将在联接结果中看到 null 值。

若要创建联接，请执行以下操作：

1. 使用以下方法之一联接两个表：
   - 将至少两个表添加到**“流程”**窗格，则选择并将相关的表拖到其他表上，直至显示**“联接”**选项。

   - 单击 ⊕ 图标并从菜单中选择**“联接”**，然后将其他输入手动添加到联接，并添加联接子句。

> **注意**：如果连接到表关系已定义并包含相关字段的表，则可以选择**"联接"**，并从相关表列表中进行选择。Tableau Prep 基于组成两个表之间的关系的字段创建联接。
>
> 有关包含表关系的连接器的详细信息，请参见在"输入"步骤中联接数据 在本页133。

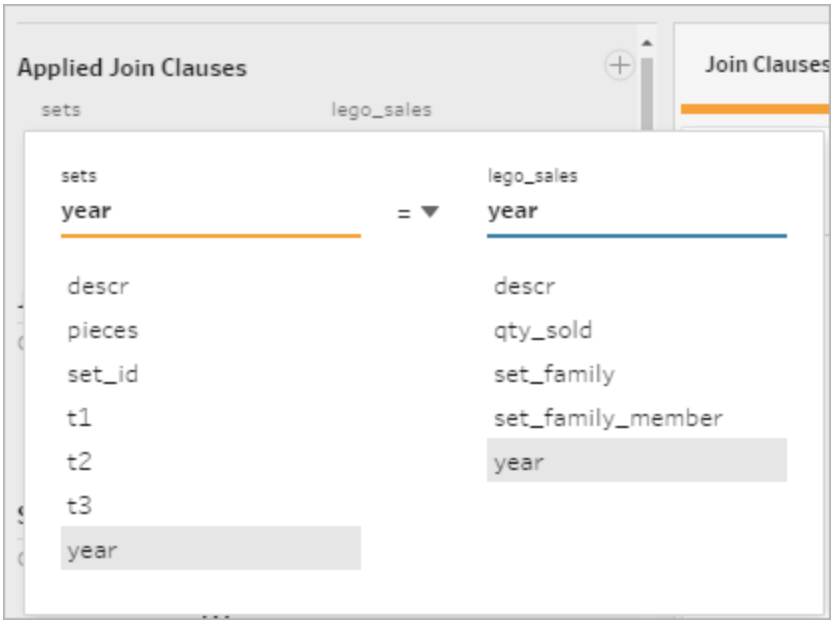
将会向流程中添加一个新联接步骤，并且配置窗格将更新以显示联接配置。

2. 若要查看和配置联接，请执行以下操作：

   a. 查看**"联接结果汇总"**，查看由于联接类型和联接条件的原因而包括和排除的字段数。

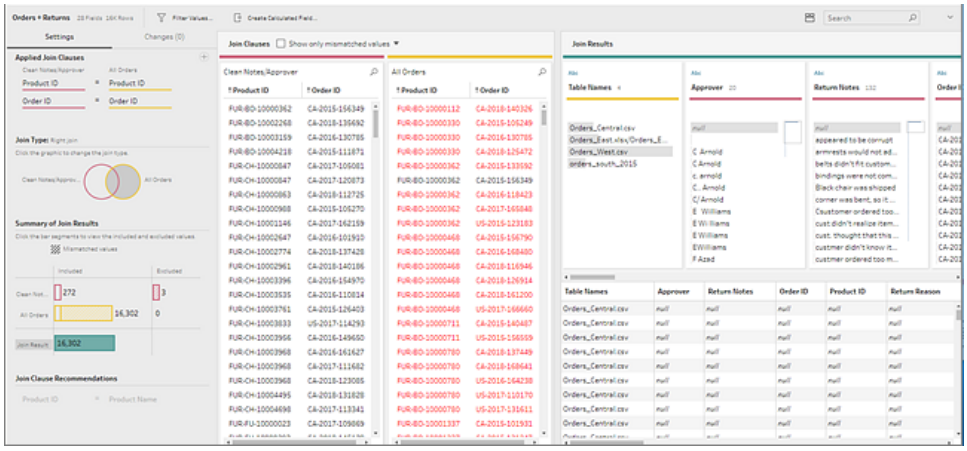
   b. 在**"联接类型"**下的维恩图中单击以指定所需联接的类型。

   c. 在**"已应用联接子句"**下，单击加号 ⊕ 图标，或在为默认联接条件选择的字段上指定或编辑联接子句。您在联接条件中指定的字段是联接中各表之间的共同字段。



   d. 您可以单击**"联接子句建议"**下显示的建议联接子句，将子句添加到已应用联

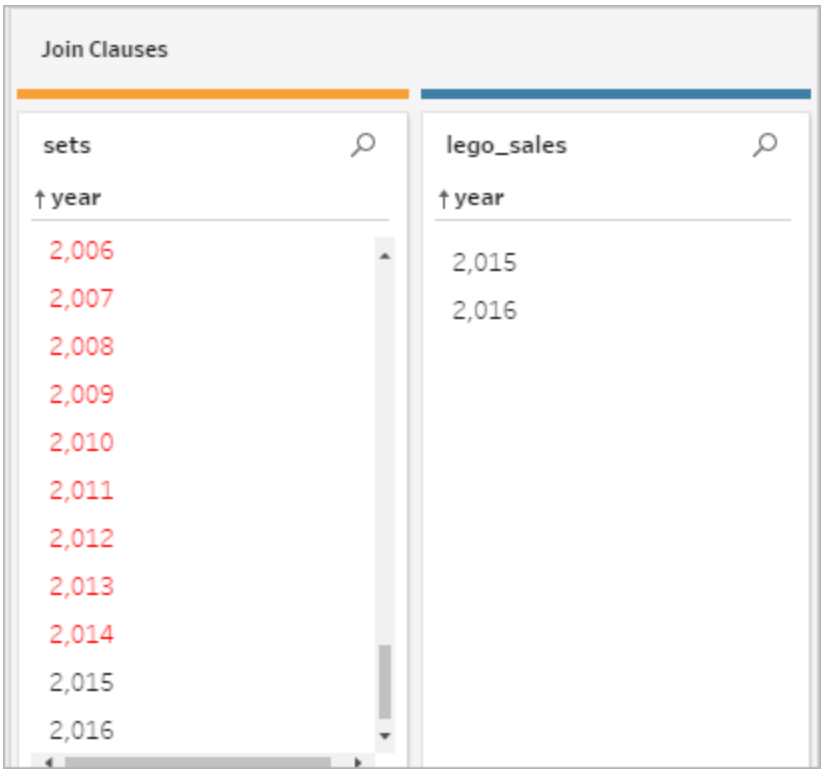
接子句的列表。



检查联接的结果

联接配置中的汇总显示有关联接的元数据，帮助您验证联接是否包括预期的数据。

- **已应用联接子句**：默认情况下，Tableau Prep 基于所联接的表中的共同字段名称定义第一个联接子句。根据需要添加或移除联接子句。
- **联接类型**：默认情况下，当您创建联接时，Tableau Prep 在表之间使用内部联接。根据您连接到的数据，您也许能够使用左联接、内部联接、右联接、仅左联接、仅右联接、非内部联接或完全联接。
- **联接结果汇总**："联接结果汇总"显示包括在联接的表中和从中排除的值的分布。
  - 单击每个**"已包括"**条形以隔离并查看包括在联接中的联接配置中的数据。
  - 单击每个**"已排除"**条形以隔离并查看从联接中排除的联接配置中的数据。
  - 单击**"已包括"**和**"已排除"**条形的任何组合以查看数据的累积性洞察。
- **联接子句建议**：单击建议联接子句旁边的加号 ⊕ 图标以将其添加到**"已应用联接子句"**列表。
- **"联接子句"**窗格：在**"联接子句"**窗格中，您可以看到联接子句中每个字段中的值。不符合联接子句条件的值将以红色文本显示。



- **“联接结果”**窗格：如果在**“联接结果”**窗格中看到要更改的值，您可以在此窗格中编辑值。

常见联接问题

如果在联接数据后未看到预期结果，您可能需要对字段值执行一些额外的清理操作。以下问题会导致 Tableau Prep 将值读取为非匹配值，并将其从联接中排除：

- 大小写不同：“My Sales”和“my sales”
- 拼写不同：“Hawaii”和“Hawai'i”
- 拼写错误或数据录入错误：“My Company Health”和“My Company Heath”
- 名称变化：“John Smith”和“John Smith Jr.”
- 缩写：“My Company Limited”和“My Company Ltd”
- 额外的分隔符：“Honolulu”和“Honolulu (Hawaii)”
- 额外的空格：这包括字符之间的额外空格、制表符空格或者额外的前导或尾随空格
- 句点的使用不一致：“Returned, not needed”和“Returned, not needed.”

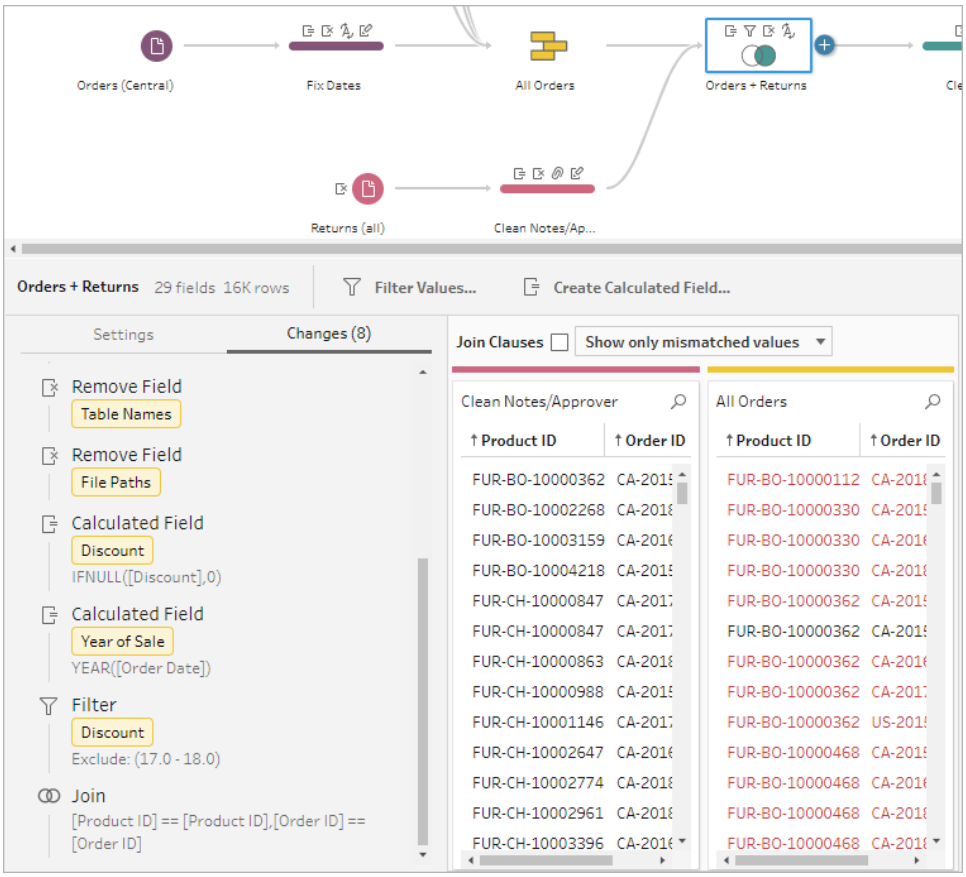
好消息是，如果您的字段值存在任何这些问题，您可以直接在**“联接子句”**中直接修复这些字段值，或者通过在**“联接结果汇总”**中单击**“已排除”**栏使用已排除的值，并在配置卡菜单中使用清理操作。

有关可在“联接”步骤中使用的不同清理选项的详细信息，请参见**关于清理操作** 在本页 213。

修复不匹配的字段以及更多操作

您可以直接在联接子句中修复不匹配的字段。双击或右键单击值，从要修复的字段上的上下文菜单中选择**“编辑值”**，并输入一个新值。系统会对您的数据更改进行跟踪，并将更改添加到“联接”步骤中的**“更改”**窗格。

您也可以在“联接子句”窗格中选择要保留、排除或筛选的多个值，或在“联接结果”窗格中应用其他清理操作。根据您更改的字段以及这些字段位于联接过程中的何处，将会在联接之前或之后应用您所做的更改，以便为您提供更正的结果。



有关清理字段的详细信息，请参见应用清理操作 在本页217。

合并数据

并集是一种通过将一个表的行附加于是另一个表来合并数据的方法。例如，您可能需要将一个表中的新事务添加到另一个表中的过去的事务列表。确保您合并的表具有相同的字段数、相同的字段名称，并且字段的数据类型相同。

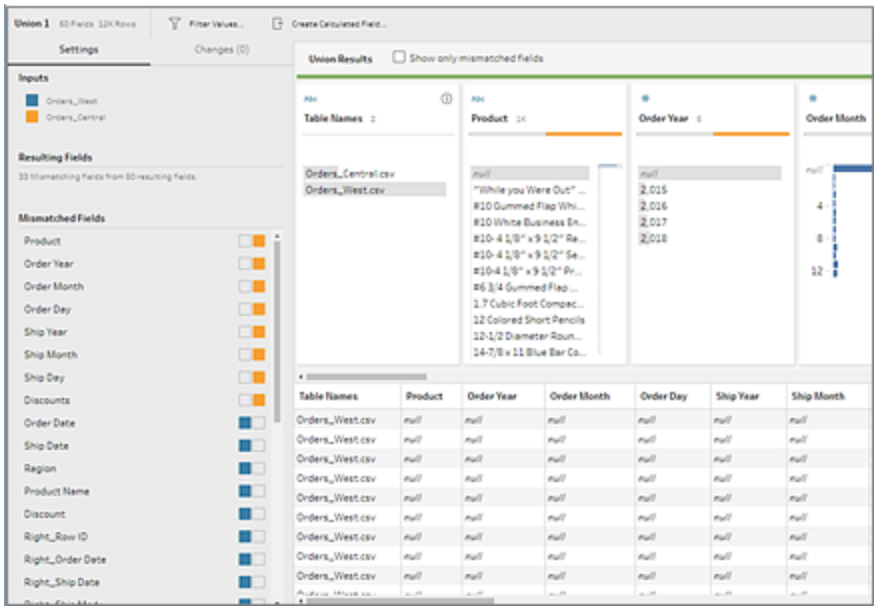
提示：为了最大程度地提升性能，一个并集最多可以有 10 个输入。如果需要合并超过 10 个文件或表，请尝试在输入步骤中合并文件。有关这种类型的合并的详细信息，请参见在输入步骤中合并文件和数据库表 在本页123。

与联接类似，您可以在流程中的任何地方使用合并操作。

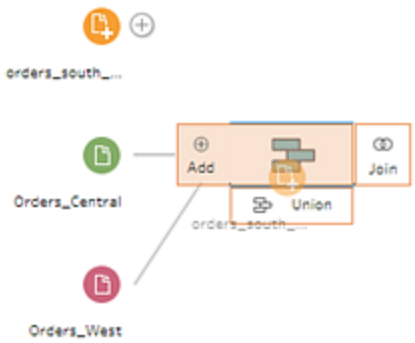
若要创建并集，请执行以下操作：

1. 将至少两个表添加到流程窗格之后，选择并将相关的表拖到其他表上，直至显示**“并集”**选项。您也可以单击 ⊕ 图标，并从菜单中选择**“并集”**。将会在**“流程”**窗格中

添加一个新的并集步骤，并且**“配置”**窗格将更新以显示并集配置。



2. 通过将表拖到合并的表，直至看到**“添加”**选项，向并集中添加其他表。

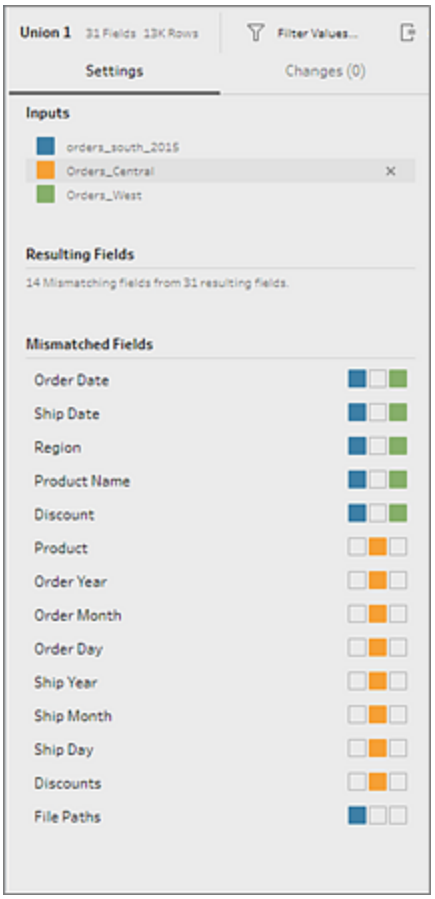


3. 在并集配置中，查看有关并集的元数据。您可以从并集中移除表，以及查看有关任何不匹配字段的详细信息。

检查并集的结果

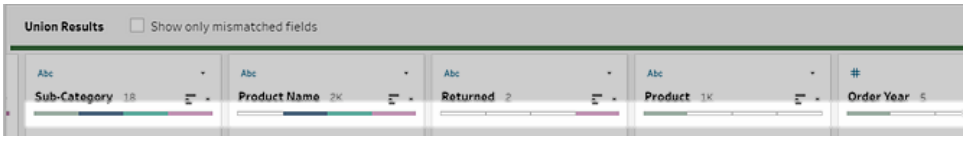
创建并集之后，请检查并集的结果以验证并集中的数据符合预期。若要验证合并数据，请检查以下区域：

- **查看并集元数据**：并集配置显示有关并集的一些元数据。在这里您可以看到组成并集的表、生成的字段数以及任何不匹配的字段。

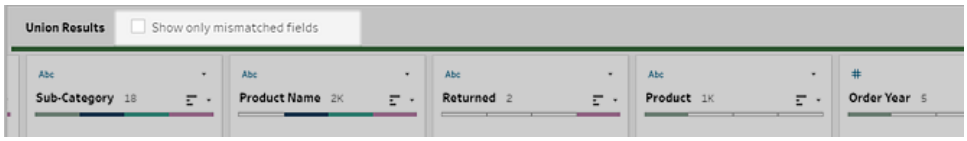


- **查看每个字段的颜色**:并集摘要中列出的每个字段的旁边以及并集配置中每个字段的上方有一组颜色。这些颜色与并集中的每个表对应。

  如果为该字段显示所有表颜色,则该字段的合并操作执行正确。缺少一种表颜色表明有不匹配的字段。



不匹配的字段是指可能包含类似数据但在某些方面有所不同的字段。您可以在并集摘要中看到不匹配字段的列表,以及它们所来自的表。如果想要更仔细地查看字段中的数据,请选中**"仅显示不匹配字段"**复选框来隔离并集配置中的不匹配字段。



若要修复这些字段，请按照下面**修复不匹配的字段** 向下部分中的建议之一进行操作。

修复不匹配的字段

当并集中的表不匹配时，合并操作将会生成额外的字段。这些额外字段是从其对应上下文中排除的有效数据。

为了解决字段不匹配问题，您必须将不匹配的字段合并在一起。

字段之所以可能不匹配，有很多原因。

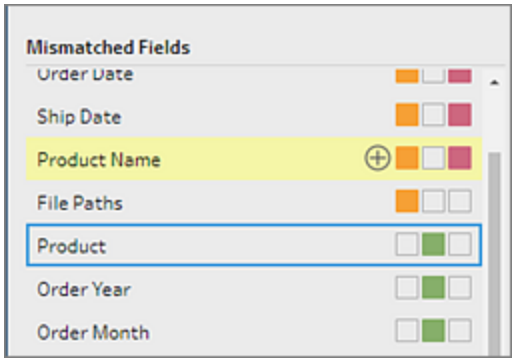
- **对应的字段名称不同**：如果表之间的对应字段的名称不同，您可以使用并集建议手动合并**"不匹配字段"**列表中的字段，或者重命名并集配置中的字段，将不匹配的字段合并在一起。

  若要使用并集建议，请执行以下操作：

  1. 在**"不匹配字段"**列表中，单击某个不匹配的字段。如果存在建议的匹配项，则匹配字段将以黄色突出显示。

     建议的匹配项基于具有类似数据类型和字段名的字段。

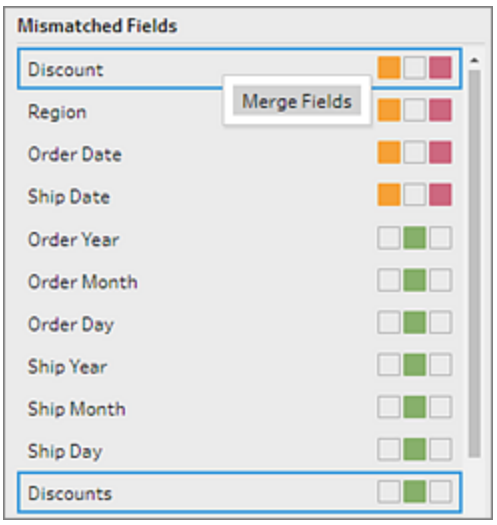
  2. 将光标悬停在突出显示的字段上，并单击加号图标来合并字段。



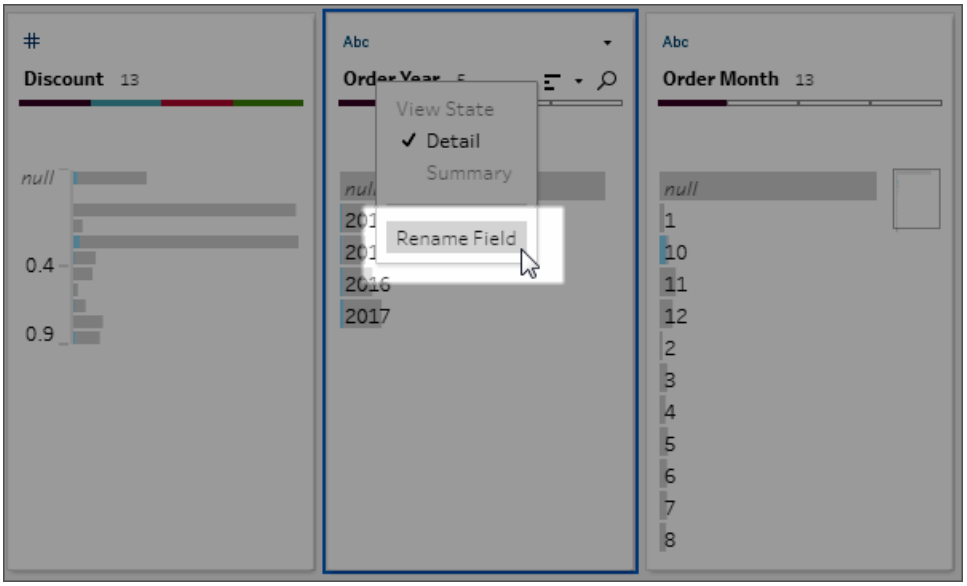
若要手动合并**"不匹配字段"**列表中的字段，请执行以下操作：

1. 在列表中选择一个或多个字段。
2. 右键单击或按住 Ctrl 单击 (MacOS) 所选字段，如果合并有效，**“合并字段”**菜单选项将出现。

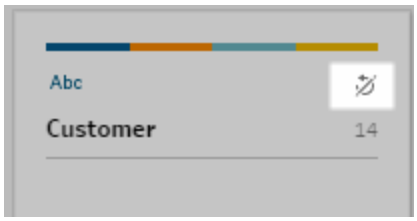
   如果在右键单击字段时看到**“无可用选项”**，这是因为字段不符合合并条件。例如，尝试合并来自同一输入的两个字段就不符合条件。
3. 单击**“合并字段”**来合并所选字段。



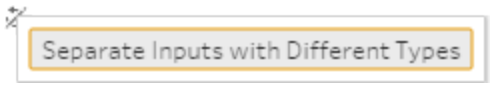
若要在并集配置窗格中重命名字段，请右键单击字段名，并单击**“重命名字段”**。



- **对应的字段名称相同，但类型不同：**默认情况下，如果对应字段的名称匹配，但字段的数据类型不匹配，Tableau Prep 将会更改其中一个字段的数据类型，使它们相互兼容。如果 Tableau Prep 进行此更改，则会在合并的字段上用一个“更改数据类型”图标加以注明。



某些情况下，Tableau Prep 可能无法选取正确的数据类型。如果发生这种情况，并且您想要撤消合并，请右键单击或按住 Ctrl 单击 (MacOS)**“更改数据类型”**图标，并选择**“分隔具有不同类型的输入”**。



然后，您可以通过以下方式再次合并字段：首先更改其中一个字段的数据类型，然后按照其他合并字段选项 向下中的建议进行操作。

- **对应的表的字段数不同：**为了合并表，并集中的每个表必须包含相同数量的字段。如果并集产生了额外的字段，请将字段合并到一个现有字段中。

其他合并字段选项

除了上面有关合并字段的部分中介绍的方法外，您还可以使用以下方法之一来合并字段。您可以在除“输出”步骤之外的任何步骤中合并字段。

有关如何合并同一文件中的字段的信息，请参见合并字段 在本页229。

若要合并字段，请执行以下操作之一：

- 将一个字段拖放到另一个字段上。将显示**“放置以合并字段”**指示符。
- 选择多个字段，并在所选范围内右键单击以打开上下文菜单，然后单击**“合并字段”**。
- 选择多个字段，然后在上下文相关工具栏上单击**“合并字段”**。

将 Einstein Discovery 预测添加到您的流程

在 Tableau Prep Builder 版本 2021.1.3 及更高版本以及 Web 版 Tableau Cloud 和 Tableau Server 版本 2021.2.0 及更高版本中受支持。

使用 Einstein Discovery 驱动的模型对流程中的数据预测进行批量评分。预测可以帮助您做出更明智的决策，并采取行动改善您的业务成果。

应用这些模型时，会自动将预测结果的新字段（以概率分数或估计平均值的形式）添加到您的流程中。通过在应用模型时选择这些选项，您还可以向流程数据添加顶级预测因子和顶级改进字段。顶级预测因子显示对预测贡献最大的因素。顶级改进显示建议采取的行动改善预测结果。

例如，为了预测员工保留率，您可以在 Einstein Discovery 中使用历史数据（其中您已经知道结果）构建模型，然后将该模型应用到流程中的数据集，并生成预测结果。预测结果在行级别应用，帮助您在 Tableau 中进行深入分析。

如果您需要将多个模型应用于数据集，则可以在流程中包含多个预测步骤。每个预测步骤将单个预测模型应用于流程。从版本 2021.2 开始，您可以在一个流程中登录多个 Einstein Discovery 服务器以选择您需要的模型。之前的版本将您限制为每个流程只能使用一个 Einstein Discovery 服务器。

> **注意**：您必须拥有配置为访问 Einstein Discovery 的 Salesforce 许可证和用户帐户才能使用此功能。有关详细信息，请参见**先决条件** 在对页。

什么是 Einstein Discovery?

Einstein Discovery 通过统计建模和监督式机器学习增强您的商业智能，以识别、显示和可视化对您的业务数据的洞察。它快速筛选数百万行数据，找出重要的相关性，预测结果，并提出改进这些预测结果的方法。

有关 Einstein Discovery 的详细信息，请参见 Salesforce 帮助中的发现入门和利用 Einstein Discovery 进行解释、预测并采取行动。您也可以通过 Trailhead 中的利用 Einstein Discovery 获得见解来扩展您的知识。

注意：Tableau 中的 Einstein Discovery 由 salesforce.com 提供技术支持。有关适用条款，请查阅与 salesforce.com 的协议。

先决条件

若要在流程中配置和使用 Einstein Discovery 预测，您需要 Salesforce 和 Tableau 中的某些许可证、访问权限和许可。

Salesforce 要求

要求	描述
Salesforce 许可证	以下许可证之一： • Tableau 中的 Einstein Discovery 许可证 • Tableau CRM Plus 许可证 • Einstein Predictions 许可证 这些许可证需要额外付费才能获得。
Salesforce 用户帐户	配置为访问 Einstein Discovery 的帐户。 如果您使用 Tableau 中的 Einstein Discovery 许可证，则必须为您的用户帐户分配“**通过 Connect API 查看 Einstein Discovery 建议**”系统权限。 如果您使用 Tableau CRM Plus 许可证或 Einstein Predictions 许可证： • 若要使用已部署的 Einstein Discovery 模型获取预测，该帐户必须分配有“**查看 Einstein Discovery 建议**”系统权限。 • 若要在 Einstein Discovery 中构建、部署和管理预测，必须为该帐户分配“**管理 Einstein Discovery**”权限。 若要配置用户帐户，请参见 Salesforce 帮助中的设置 Einstein Discovery。
管理员设置	Salesforce 管理员将需要： • **Tableau Prep 扩展程序**：配置 Salesforce，以便针对

要求	描述
	Tableau Server 创建已连接应用(基本)。仅对于 Tableau Server 是必需的。

Tableau Prep 要求

要求	描述
Tableau Prep 许可证和权限	“Creator”许可证 作为 Creator，您需要能够登录 Salesforce org 帐户以访问预测定义并将模型添加到您的流程。
Tableau 用户帐户	在 Tableau Server 和 Tableau Cloud 版本 2021.2 及更高版本中，用户可以保存 Salesforce 用户帐户凭据及其 Tableau 用户帐户。 有关连接到 Salesforce 数据的详细信息，请参见**连接到 Salesforce 数据** 在本页71。
管理员设置	Tableau Server 管理员需要将 Tableau Server 配置为与 Einstein Discovery for Tableau Prep 集成。有关详细信息，请参见 Tableau Server 帮助中的配置 Einstein Discovery 集成。

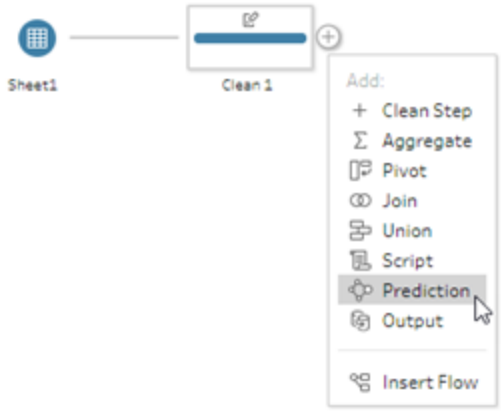
将预测数据添加到流程

注意：在版本 2021.1.4 及更低版本中，包含预测步骤的流程目前只能在 Tableau Prep Builder 中手动运行。

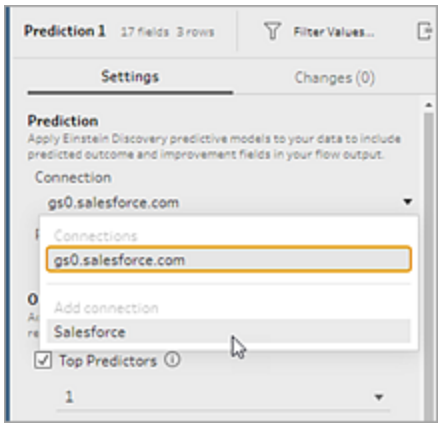
若要将 Einstein Discovery 预测应用到您的流程，您需要：

- 访问 Salesforce org。
- 访问 Tableau Prep Builder 版本 2021.1.3 或更高版本。
- 如果在 Web 上制作或运行流程，请访问已为 Einstein Discovery 预测启用的 Tableau Cloud 或 Tableau Server 版本 2021.2 或更高版本。
- 部署在 Salesforce 中的 Einstein Discovery 预测模型。
- Tableau Prep 中的源数据，其字段与 Einstein Discovery 预测模型所需的模型字段匹配。

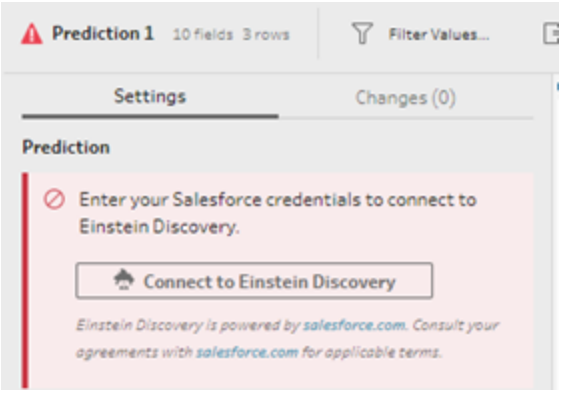
1. 打开 Tableau Prep 并连接到数据源。
2. 根据需要应用任何清理操作。
3. 单击加号 ⊕ 图标并从**“添加”**菜单中选择**“预测”**。



4. 在**“设置”**选项卡上的**“预测”**窗格中，根据您的版本执行以下操作之一：
   - **版本 2021.2 及更高版本**：在**“连接”**下拉列表中，连接到您的 Salesforce 服务器，或者，如果您已经建立了连接，则从列表中选择您的 Salesforce 服务器。

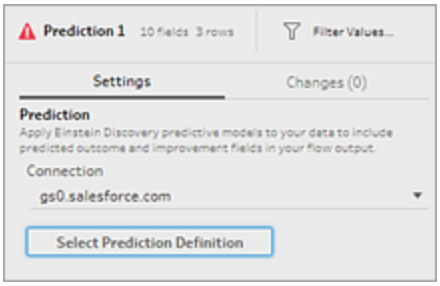


- **版本 2021.1.4 及更低版本**：单击**“连接到 Einstein Discovery”**。

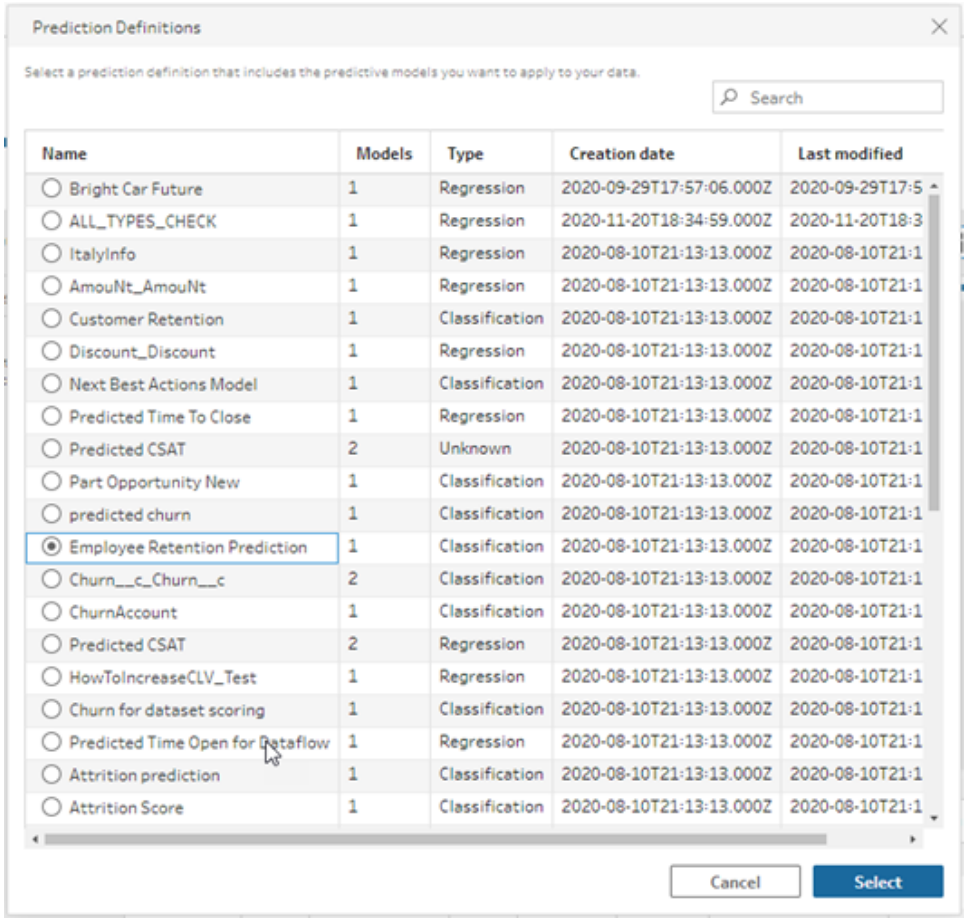


首次连接时，会打开一个网页，要求您使用您的 Salesforce 凭据登录 Salesforce 帐户。登录后，会打开一个网页，询问您是否想允许 Tableau 访问您的 Salesforce 数据。单击**“允许”**继续，然后关闭浏览器中生成的选项卡。

5. 单击**“选择预测定义”**。这将打开您有权访问的已部署模型列表。这些模型是使用 Einstein Discovery 在 Salesforce 中构建和部署的。有关预测模型的详细信息，请参见 Salesforce 中的关于模型。

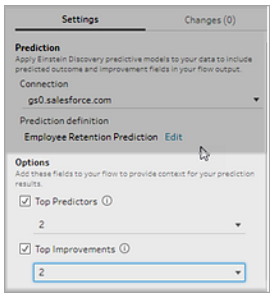


6. 在**“预测定义”**对话框中，选择映射到数据集的预测定义。若要使用流程数据生成预测结果，模型中的所有字段必须映射到对应的流程字段。



7. 在**“选项”**部分，选择最多 3 个顶级预测因子和改进，以包括在流程数据中。这是您可以添加到流程中的补充数据。

   - 顶级预测因子指明哪些因子对预测结果的贡献最大。
   - 顶级改进建议采取行动改善预测结果。



8. 在**“映射字段”**部分，将流程字段映射到模型字段。

- 所有模型字段必须映射到对应的流程字段。
- 完全匹配的字段名称会自动映射。
- 您无法将同一流程字段映射到多个模型字段。
- 模型和流程字段数据类型必须匹配。

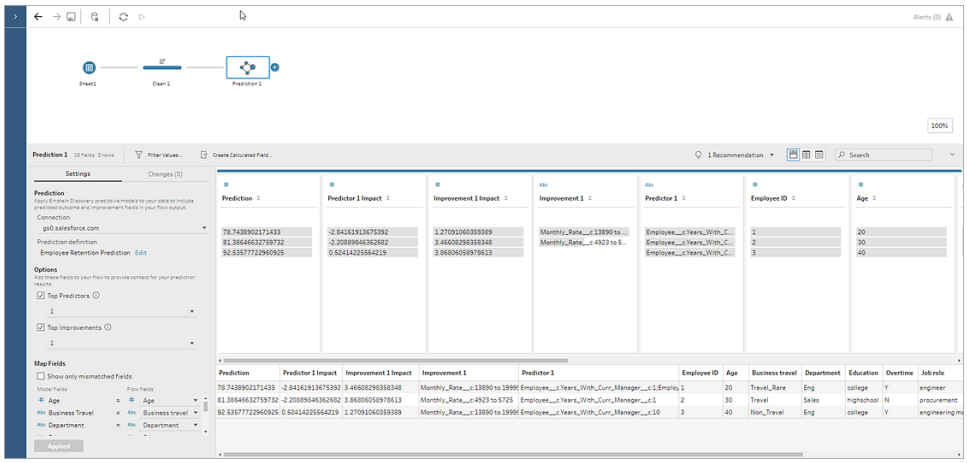
  如果为您的流程字段分配了不同的数据类型，则需要对其进行更改以与分配给模型字段的数据类型匹配。

  若要更改数据类型，只需在**“映射字段”**部分单击流程字段的数据类型，然后在菜单中选择新数据类型即可。您可以随后在后续清理步骤中将数据类型改回原来的类型。

  有关更改数据类型的信息，请参见**查看分配给数据的数据类型** 在本页153。

9. 若要应用您的设置并根据您的数据运行模型，请单击**“应用”**。预测结果显示在配置窗格和数据网络中。

   如果您更改了任何设置，则可以再次单击**“应用”**以使用您的更改重新运行模型。如果您在单击**“应用”**之前离开**“预测”**步骤，则模型不会运行，并且您的更改将丢失。

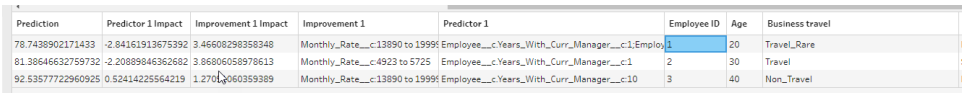


查看您的结果

将预测模型应用于流程数据后，您可以生成流程输出，并使用新数据源在 Tableau 的行级别分析预测结果。为了了解预测模型的结果，让我们来看一个例子。

在本主题中，我们将员工保留预测模型应用于 Tableau Prep 中的员工数据，以获得员工将留在公司的概率分数。

这给了我们以下结果：



Prediction	Predictor 1 Impact	Improvement 1 Impact	Improvement 1	Predictor 1	Employee ID	Age	Business travel
78.7438902171433	-2.84161913675392	3.46608298358348	Monthly_Rate__c:13890 to 1999(	Employee__c:Years_With_Curr_Manager__c:1;Employ	1	20	Travel_Rare
81.3864663275973	-2.20889846362682	3.86806058978613	Monthly_Rate__c:4923 to 5725	Employee__c:Years_With_Curr_Manager__c:1	2	30	Travel
92.5357772296092	0.52414225564219	1.27091060359389	Monthly_Rate__c:13890 to 1999(	Employee__c:Years_With_Curr_Manager__c:10	3	40	Non_Travel

对于员工 2，让我们来看看这些结果告诉了我们什么：

问题	预测	这是哪里？
该员工留任的可能性有多大？	Einstein Discovery 预测，他们有 81.38% 的机率将留任。	预测字段
影响此结果的因子是什么？	在现任经理工作的几年里，这名员工留任的可能性降低了 2.2%。	预测因子 1 字段（顶级预测因子） 预测因子 1 影响（顶级预测因子的影响百分比）
什么可以改善这个预测结果？	将员工的月薪从 4923 提高到 5725，员工留任的可能性会增加 3.86%。	改进 1 字段（顶级改进） 改进 1 影响（进行建议更改的影响百分比）

保存和共享工作

在流程的任何位置，您都可以手动保存工作，或者让 Tableau 在 Web 上创建或编辑流程时自动为您保存工作。在 Web 上处理流程时，存在一些差异。

有关在 Web 上制作流程的详细信息，请参见 Tableau Server 和 Tableau Cloud 帮助中的 **Web 上的 Tableau Prep**。

Tableau Prep Builder	Web 上的 Tableau Prep
• 在 Tableau Desktop 中查看流程中数据的预览。 • 在流程输入中包括直接文件连接或打包文件，然后将打包的流程发布到服务器。 • 将流程输出到文件、已发布数据源或数据库。	• 在 Web 上创建和编辑流程。 • 上载流程输入的文件并连接到各种数据源。 • 将流程输出到已发布的数据源或数据库。

若要使数据保持最新，您可以从 Tableau Prep Builder 中或通过命令行手动运行流程。您还可以运行在 Tableau Server 或 Tableau Cloud 上手动发布或按计划发布的流程。有关运行流程的详细信息，请参见**将流程发布到 Tableau Server 或 Tableau Cloud** 在本页 433。

保存您的流程

在 Tableau Prep Builder 中，您可以手动保存流程以在执行任何其他操作之前备份您的工作。流程以 Tableau Prep 流程 (.tfl) 文件格式保存。

您也可以随流程一起打包本地文件（Excel、文本文件和 Tableau 数据提取）以与其他人共享，就像在 Tableau Desktop 中打包工作簿用于共享一样。只能随流程一起打包本地文件。举例来说，不会包括来自数据库连接的数据。

在 Web 制作中，本地文件会自动随流程打包在一起。尚不支持直接文件连接。

在保存打包流程时，流程将保存为打包 Tableau 流程文件 (.tflx)。

- 若要手动保存流程，请从顶部菜单中选择**“文件”>“保存”**。
- 在 Tableau Prep Builder 中，若要随流程一起打包数据文件，请从顶部菜单中执行以下操作之一：

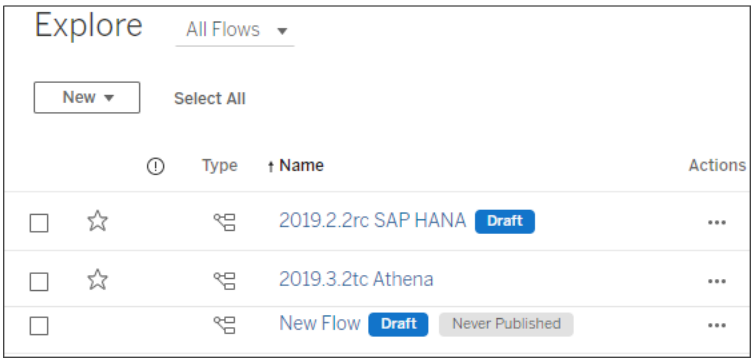
- 选择**“文件”>“导出打包流程”**
- 选择**“文件”>“另存为”**。然后，在**“另存为”**对话框中，从**“另存为类型”**下拉菜单中选择**“打包 Tableau 流程文件”**。

在 Web 上自动保存流程

如果在 Web 上创建或编辑流程，则在对流程进行更改(连接到数据源、添加步骤等)时，您的工作就会每隔几秒钟自动保存一次草稿，这样您就不会丢失工作。

您只能将流程保存到当前已登录的服务器。您不能在一台服务器上创建草稿流程，并尝试将其保存或发布到另一台服务器。如果要将流程发布到服务器上的其他项目，请使用**“文件”>“发布为”**菜单选项，然后从对话框中选择项目。

草稿流程只能由您查看，直到您发布它们，并使其可供有权访问您服务器上的项目的任何人使用。处于草稿状态的流程会使用**“草稿”**徽章进行标记，以便您可以轻松地找到正在进行中的流程。如果流程从未发布，则“草稿”徽章旁边将显示**“从未发布”**徽章。



发布流程以及编辑并重新发布流程后，将创建一个新版本。您可以在**“修订历史记录”**对话框中查看流程版本列表。从**“浏览”**页面中，单击 ··· **“操作”**菜单并选择**“修订历史记录”**。

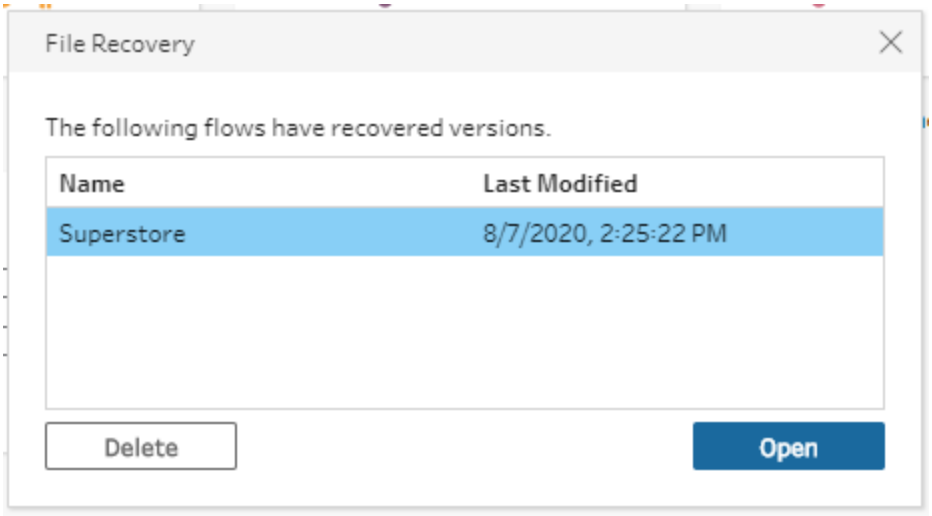
有关修订历史记录的详细信息，请参见 Tableau Desktop 帮助中的处理内容修订。

注意：“自动保存”功能默认情况下已启用。管理员可以(但不建议)在站点上禁用自动保存。若要关闭自动保存，请使用 Tableau Server REST API 方法“更新站点”，并将 `flowAutoSaveEnabled` 属性设置为 false。有关详细信息，请参见 Tableau Server REST API 站点方法：更新站点。

自动文件恢复

默认情况下，如果应用程序冻结或崩溃，Tableau Prep Builder 将自动保存任何已打开流程的草稿。草稿流程保存在**“我的 Tableau Prep 存储库”**中的**“已恢复流程”**文件夹中。下次打开应用程序时，将显示一个对话框，其中包含可供选择的已恢复流程的列表。您可以打开恢复的流程并在停止的位置继续，或者在不需要时删除恢复的流程文件。

注意：如果在**“已恢复流程”**文件夹中有已恢复流程，则每次打开应用程序时都会显示此对话框，直到该文件夹为空。



如果不希望启用此功能，作为管理员，您可以在安装期间或安装之后将其关闭。有关如何关闭此功能的详细信息，请参见 Tableau Desktop 和 Tableau Prep 部署指南中的关闭恢复。

在 Tableau Desktop 中查看流程输出

注意：此选项在 Web 上不可用。

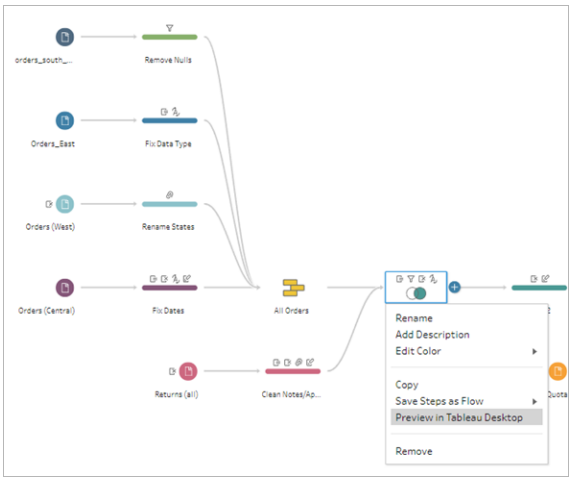
当您清理数据时，有时您可能希望通过在 Tableau Desktop 中查看数据来检查进度。当流程在 Tableau Desktop 中打开时，Tableau Prep Builder 会创建一个永久的 Tableau .hyper 文件和一个 Tableau 数据源 (.tds) 文件。这些文件保存在 Tableau 存储库的**“数据源”**文件中，以便您能够随时试验数据。

在 Tableau Desktop 中打开流程时，您可以看到流程中您正在处理的数据样本，其已应用到所选步骤为止的操作。

注意：尽管您可以体验数据，但 Tableau 只会向您显示数据示例，您将无法将工作簿另存为打包工作簿 (.twbx)。当您准备在 Tableau 处理自己的数据时，请在流程中创建一个输出步骤，并将输出保存到文件或另存为发布的数据源，然后在 Tableau 中连接到完整数据源。

若要在 Tableau Desktop 中查看数据样本，请执行以下操作：

1. 右键单击要在其中查看数据的步骤，并从上下文菜单中选择**“在 Tableau Desktop 中预览”**。



2. Tableau Desktop 将打开**“工作表”**选项卡。

创建数据提取文件和已发布数据源

若要创建流程输出，请运行流程。运行流程时，您所做的更改将应用于整个数据集。运行流程会生成 Tableau 数据源 (.tds) 和 Tableau 数据提取 (.hyper) 文件。

Tableau Prep Builder

您可以依据流程输出创建数据提取文件，以便在 Tableau Desktop 中使用，或与第三方共享数据。采用以下格式创建数据提取文件：

- **Hyper 数据提取 (.hyper)**：这是最新的 Tableau 数据提取文件类型。
- **逗号分隔值 (.csv)**：将数据提取保存到 .csv 文件以与第三方共享数据。导出的 CSV 文件的编码将为带 BOM 的 UTF-8。
- **Microsoft Excel (.xlsx)**：Microsoft Excel 电子表格。

Tableau Prep Builder 以及在 Web 上

将您的流程输出发布为已发布数据源或输出到数据库。

- 将您的流程输出作为数据源保存到 Tableau Server 或 Tableau Cloud，以共享您的数据并提供对您已清理、整形和组合的数据的集中访问。
- 将您的流程输出保存到数据库，以使用干净、准备好的流程数据创建、替换或附加表数据。有关详细信息，请参见**将流程输出数据保存到外部数据库** 在本页 370。

在运行流程时使用增量刷新，通过仅刷新新数据而不是完整数据集来节省时间和资源。有关如何使用增量刷新配置和运行流程的信息，请参见**使用增量刷新来刷新流程数据** 在本页 390。

> **注意**：若要将 Tableau Prep Builder 输出发布到 Tableau Server，必须启用 Tableau Server REST API。有关详细信息，请参见 Tableau REST API 帮助中的 REST API 要求。若要发布到使用安全套接字层 (SSL) 加密证书的服务器，还要在运行 Tableau Prep Builder 的计算机上执行一些额外配置步骤。有关详细信息，请参见 Tableau Desktop 和 Tableau Prep Builder 部署指南中的安装之前。

在流程输出中包含参数

从版本 2021.4 开始在 Tableau Prep Builder 和 Web 上受支持

在流程输出文件名、路径、表名或自定义 SQL 脚本(版本 2022.1.1 及更高版本)中包含参数值，以轻松针对不同数据集运行流程。有关详细信息，请参见**在流程中创建和使用参数** 在本页 192。

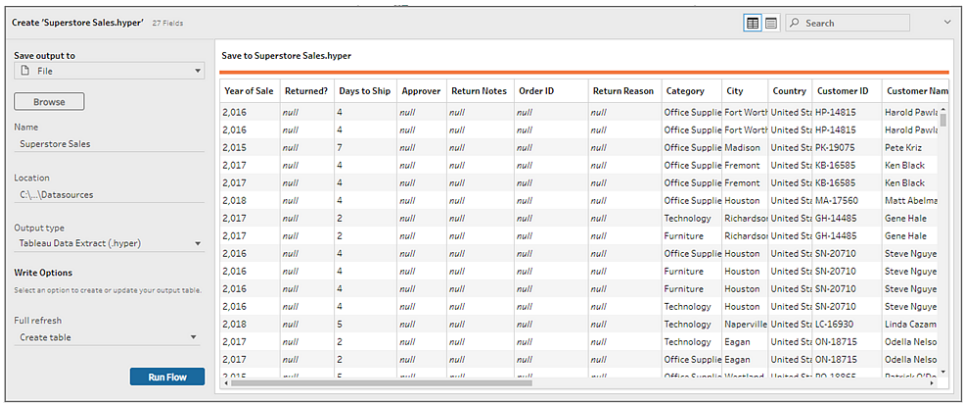
创建数据提取到文件

> **注意**：在 Web 上创建或编辑流程时，此输出选项不可用。

1. 单击步骤上的加号图标 ⊕，并选择**“添加输出”**。

   如果以前运行过流程，请单击输出步骤上的运行流程 ▷ 按钮。这将运行流程并更新输出。

   “输出”窗格将打开，并显示数据的快照。



2. 在左侧窗格中，从**“将输出保存到”**下拉列表中选择**“文件”**。在以前的版本中，选择**“保存到文件”**。
3. 单击**“浏览”**按钮，然后在**“将数据提取另存为”**对话框中输入文件的名称，并单击**“接受”**。
4. 在**“输出类型”**字段中，从以下输出类型中进行选择：
   - Tableau 数据提取 (.hyper)
   - 逗号分隔值 (.csv)
5. (Tableau Prep Builder) 在**“写入选项”**部分，查看用于将新数据写入文件并根据需要进行任何更改的默认写入选项。有关详细信息，请参见**配置写入选项** 在本页394。
   - **创建表**：此选项创建新表，或将现有表替换为新输出。
   - **追加到表**：此选项将新数据添加到现有表中。如果表不存在，将创建一个新表，后续运行会将新行添加到此表。

   > **注意**：对于 .csv 输出类型，不支持**“追加到表”**。有关支持的刷新组合的详细信息，请参见**流程刷新选项** 在本页391。
6. 单击**“运行流程”**以运行流程并生成数据提取文件。

创建数据提取到 Microsoft Excel 工作表

在 Tableau Prep Builder 版本 2021.1.2 及更高版本中受支持。在 Web 上创建或编辑流程时，此输出选项不可用。

当您流程数据输出到 Microsoft Excel 工作表时，您可以在现在工作表中附加或替换数据。以下条件适用：

- 仅支持 Microsoft Excel .xlsx 文件格式。
- 工作表行从单元格 A1 开始。
- 在附加或替换数据时，会假定第一行为标题。

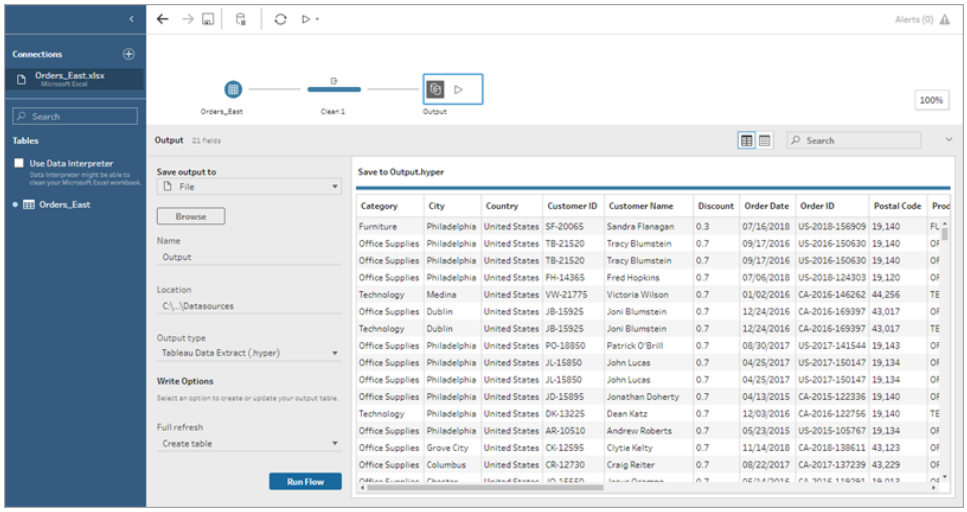
- 创建新工作表时会添加标题名称，但在将数据添加到现有工作表时不会添加。
- 现有工作表中的任何格式设置或公式不适用于流程输出。
- 目前不支持写入到指定的表或范围。
- 当前不支持增量刷新。

将流程数据输出到 Microsoft Excel 工作表文件

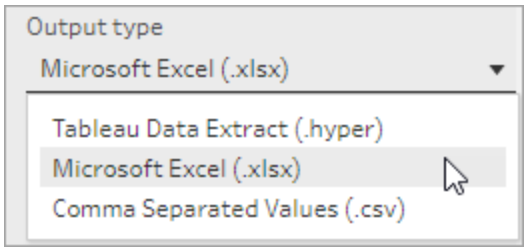
1. 单击步骤上的加号图标 ⊕，并选择**“添加输出”**。

   如果以前运行过流程，请单击输出步骤上的运行流程 ▷ 按钮。这将运行流程并更新输出。

   “输出”窗格将打开，并显示数据的快照。

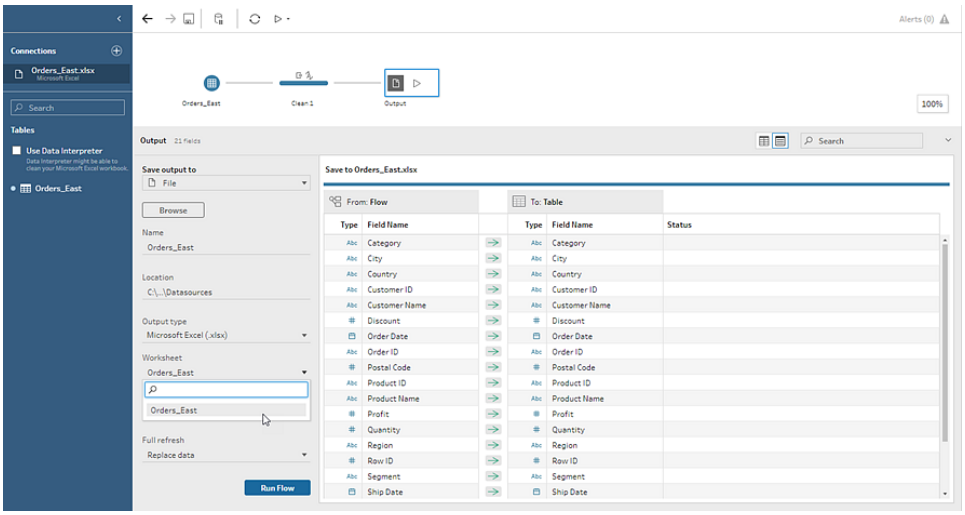


2. 在左侧窗格中，从**“将输出保存到”**下拉列表中选择**“文件”**。
3. 单击**“浏览”**按钮，然后在**“将数据提取另存为”**对话框中输入或选择文件名，并单击**“接受”**。
4. 在**“输出类型”**字段中，选择“ Microsoft Excel (.xlsx)”。



5. 在**“工作表”**字段中，选择要将结果写入的工作表，或改为在字段中输入新名称，然后单击**“创建新表”**。
6. 在**“写入选项”**部分中，选择以下写入选项之一：
   - **创建表：**使用流程数据创建或重新创建（如果文件已经存在）工作表。
   - **追加到表：**在现有工作表中添加新行。如果工作表不存在，则创建一个，并且后续流程运行会向该工作表添加行。
   - **替换数据：**用流程数据替换现有工作表中的第一行以外的所有现有数据。

     字段比较显示流程中与工作表中字段匹配的字段（如果工作表已存在）。如果工作表是新的，则显示一对一字段匹配。任何不匹配的字段都被忽略。



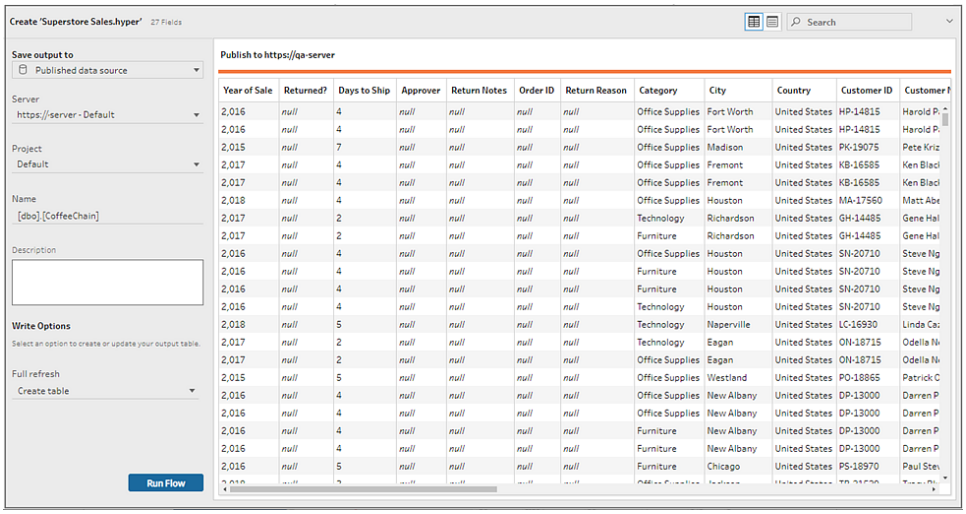
7. 单击**“运行流程”**以运行流程并生成 Microsoft Excel 数据提取文件。

创建发布的数据源

1. 单击步骤上的加号图标 ⊕，并选择**“添加输出”**。

> **注意：**Tableau Prep Builder 将刷新以前发布的数据源，并保留数据源中可能包含的任何数据建模（例如计算字段、数字格式设置等）。如果无法刷新数据源，则将替换数据源，包括数据建模。

2. “输出”窗格将打开，并显示数据的快照。



3. 从**“将输出保存到”**下拉列表中，选择**“已发布数据源”**(在以前的版本中选择**“发布为数据源”**)。填写以下字段：
   - **服务器**(仅限 Tableau Prep Builder)：选择要在其中发布数据源和数据提取的服务器。如果未登录到服务器，则会提示您登录。

     注意：从 Tableau Prep Builder 版本 2020.1.4 开始，在您登录到服务器后，Tableau Prep Builder 会在关闭应用程序时记住服务器名称和凭据。下次打开应用程序时，您已经登录服务器。

     在 Mac 上，可能会提示您提供对 Mac 密钥链的访问权限，以便 Tableau Prep Builder 能够安全地使用 SSL 证书来连接到 Tableau Server 或 Tableau Cloud 环境。

     如果要输出到 Tableau Cloud，请在“serverUrl”中包括您的站点托管于其上的 pod。例如，“https://eu-west-1a.online.tableau.com”，而不是“https://online.tableau.com”。
   - **项目**：选择要在其中加载数据源和数据提取的项目。
   - **名称**：输入文件名。
   - **说明**：输入数据源的说明。
4. (Tableau Prep Builder) 在**“写入选项”**部分，查看用于将新数据写入文件并根据需要进行任何更改的默认写入选项。有关详细信息，请参见**配置写入选项** 在本页394

- **创建表**：此选项创建新表，或将现有表替换为新输出。
- **追加到表**：此选项将新数据添加到现有表中。如果表不存在，将创建一个新表，后续运行会将新行添加到此表。

5. 单击**"运行流程"**以运行流程并发布数据源。

将流程输出数据保存到外部数据库

重要信息：此功能使您能够永久删除和替换外部数据库中的数据。请确保您具有写入数据库的权限。
为了防止数据丢失，您可以使用**"自定义 SQL"**选项复制表数据，并在将流程数据写入表之前运行它。

可以通过 Tableau Prep Builder 或 Web 支持的任何连接器连接到数据，并将数据输出到外部数据库。这使您能够在每次运行流程时使用干净、准备好的数据在数据库中添加或更新数据。此功能可用于增量刷新和完全刷新选项。有关如何配置增量刷新的详细信息，请参见使用增量刷新来刷新流程数据 在本页390。

将流程输出保存到外部数据库时，Tableau Prep 将执行以下操作：

1. 生成行并针对数据库运行任何 SQL 命令。
2. 将数据写入输出数据库中的临时表（如果输出到 Snowflake，则为暂存区域）。
3. 如果操作成功，数据将从临时表（或 Snowflake 的暂存区域）移动到目标表中。
4. 运行将数据写入数据库后要运行的任何 SQL 命令。

如果 SQL 脚本失败，流程将失败。但是，您的数据仍将加载到数据库表中。您可以尝试再次运行流程，也可以在数据库上手动运行 SQL 脚本以应用它。

输出选项

将数据写入数据库时，可以选择以下选项。如果表不存在，则在首次运行流程时创建该表。

- **追加到表**：此选项将数据添加到现有表中。如果表不存在，则在首次运行流程时创建该表，并在每次后续流程运行时将数据添加到该表中。
- **创建表**：此选项使用流程中的数据创建一个新表。如果该表已存在，则删除表和为表定义的任何现有数据结构或属性，并替换为使用流程数据结构的新表。流程中存在的任何字段都将添加到新的数据库表中。
- **替换数据**：此选项删除现有表中的数据并将其替换为流程中的数据，但保留数据库表的结构和属性。如果表不存在，则在首次运行流程时创建该表，并且在每次后续流程运行时，表数据将替换。

其他选项

除了写入选项之外，还可以包括自定义 SQL 脚本或向数据库添加新表。

- **自定义 SQL 脚本**：输入您的自定义 SQL 并选择是在将数据写入数据库表之前、之后还是之前和之后运行您的脚本。您可以使用这些脚本在将流程数据写入表之前创建数据库表的副本、添加索引、添加其他表属性等等。

  注意：从版本 2022.1.1 开始，您还可以在 SQL 脚本中插入参数。有关详细信息，请参见**将用户参数应用于输出步骤** 在本页202。

- **添加新表**：向数据库添加具有唯一名称的新表，而不是从现有表列表中选择一个表。如果要应用默认架构（Microsoft SQL Server 和 PostgreSQL）以外的架构，可以使用语法 `[schema name].[table name]` 指定该架构。

支持的数据库和数据库要求

Tableau Prep 支持将流程数据写入特定数量的数据库中的表。在 Tableau Cloud 中按计划运行的流程只有在云托管时才能写入这些数据库。

不支持专用网络连接和通过 Tableau Bridge 的本地连接。您可以在 Tableau Cloud 中创建使用 Bridge 刷新数据的已发布数据源，然后 Prep 可以连接到已发布的数据源。

某些数据库具有数据限制或要求。Tableau Prep 可能还会施加一些限制，以便在将数据写入支持的数据库时保持最佳性能。下表列出了可在其中保存流程数据的数据库，以及任何数据库限制或要求。不符合这些要求的数据在运行流程时可能会导致错误。

注意：尚不支持为字段设置字符限制。但是，您可以在数据库中创建包含字符限制约束的表，然后使用**“替换数据”**选项替换数据，但保留数据库中的表结构。

数据库	要求或限制
Amazon Redshift	• 不支持排序规则顺序。有关详细信息，请参见 Amazon Redshift 文档。 • 字段名将转换为全小写。 • 可为文本字段值最多写入 8192 个字符。较长的值会被截断。
Google BigQuery	• Tableau 最多可以将 2 GB 作为输出写入表。

Microsoft SQL Server	• 可为文本字段值最多写入 3072 个字符。较长的值会被截断。 • (版本:2022.3.1) 发布到 Tableau Server 的流程输出允许使用运行身份凭据对 Microsoft SQL Server 数据库进行写入访问。请参见 intsm configuration set 选项中的 `maestro.output.write_to_mssql_using_runas`。
MySQL	• 可为文本字段值最多写入 8192 个字符。较长的值会被截断。
Oracle	• 字段名和表名不能超过 30 个字符。 • 可为文本字段值最多写入 1000 个字符。较长的值会被截断。 • 字段名中的特殊字符可能会导致错误。
Pivotal Greenplum Database	• 可为文本字段值最多写入 8192 个字符。较长的值会被截断。
PostgreSQL	• 可为文本字段值最多写入 8192 个字符。较长的值会被截断。
SAP HANA	• 可为文本字段值最多写入 8192 个字符。较长的值会被截断。
Snowflake	• 可为文本字段值最多写入 8192 个字符。较长的值会被截断。 • 必须将仓库选项设置为自动恢复,以使 Tableau Prep 能够将数据写入数据库仓库。有关详细信息,请参见 Snowflake 文档中的自动暂停和自动恢复。
Teradata	• 可为文本字段值最多写入 1000 个字符。较长的值会被截断。
Vertica	• 可为文本字段值最多写入 8192 个字符。较长的值会被截断。

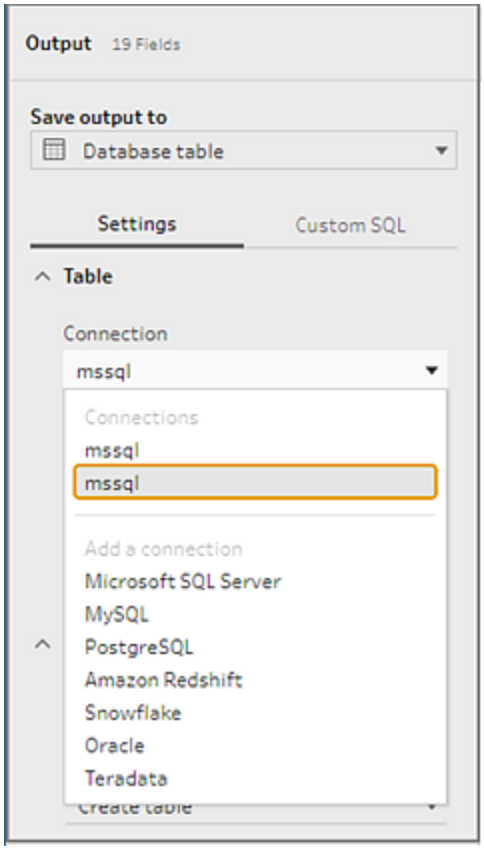
将流程数据保存到数据库

注意:不支持使用 Windows 身份验证将流程输出写入数据库。如果使用这种身份验证方法,则需要更改连接身份验证以使用用户名和密码。

您可以在发布流程时为数据库嵌入凭据。有关嵌入凭据的详细信息,请参见**从 Tableau Prep Builder 中发布流程** 在本页438中的**"数据库"**部分

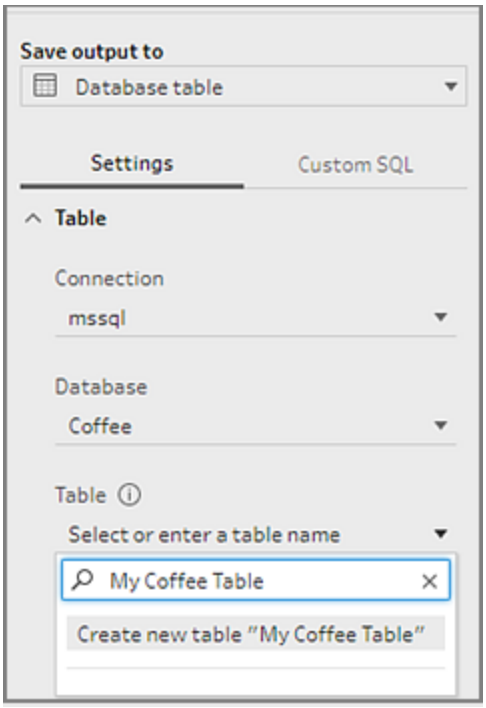
1. 单击步骤上的加号图标 ⊕，并选择**“添加输出”**。
2. 从**“将输出保存到”**下拉列表中，选择**“数据库表”**。
3. 在**“设置”**选项卡中，输入以下信息：
   - 在**“连接”**下拉列表中，选择要在其中写入流程输出的数据库连接器。仅显示受支持的连接器。这可以是用于流程输入的连接器的同一连接器或其他连接器。如果选择其他连接器，系统将提示您登录。

     重要信息：请确保您对选择的数据库具有写入权限。否则，流程可能只会部分处理数据。

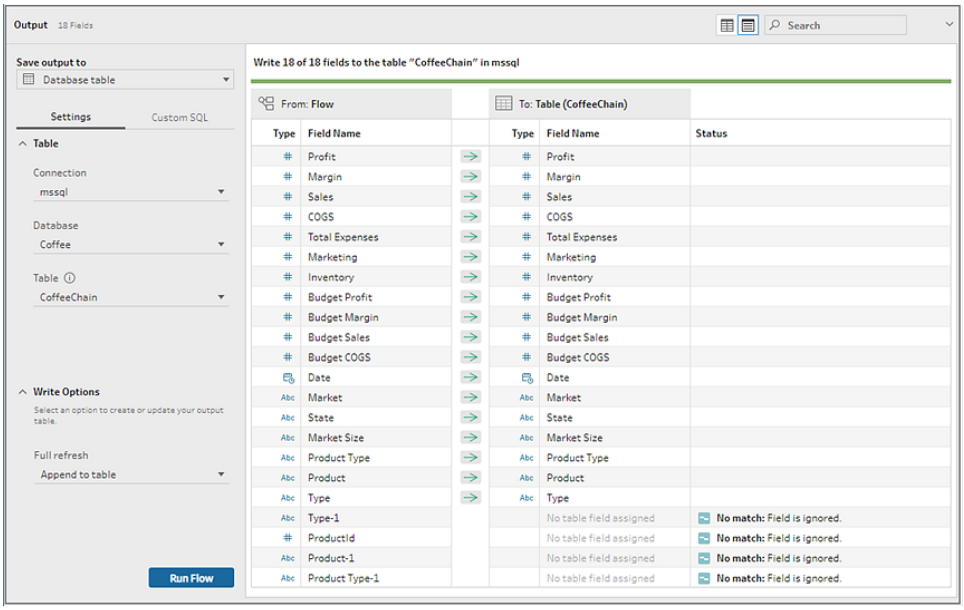


   - 在**“数据库”**下拉列表中，选择要在其中保存流程输出数据的数据库。
   - 在**“表”**下拉列表中，选择要在其中保存流程输出数据的表。根据所选的**“写入选项”**，将创建一个新表，流程数据将替换表中的任何现有数据，或者流程数据将添加到现有表中。

     若要在数据库中创建新表，请改为在字段中输入唯一的表名，然后单击**“创建新表”**。首次运行流程时，无论选择哪个写入选项，都将使用与流程相同的架构在数据库中创建表。



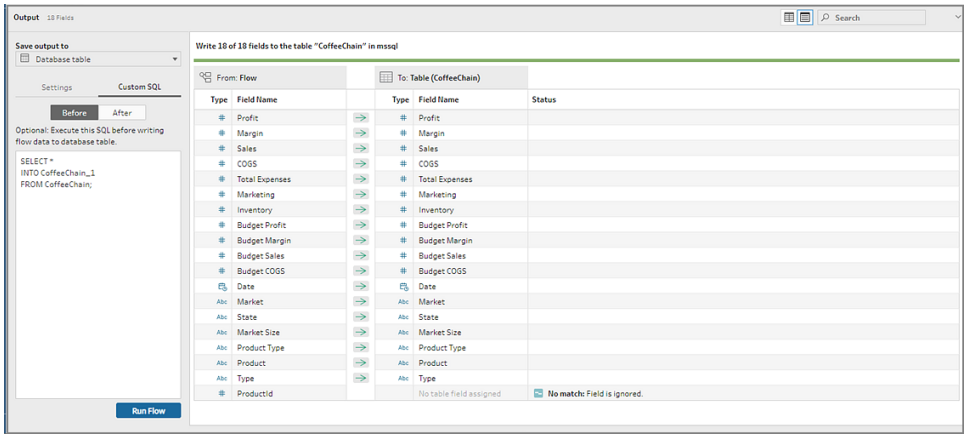
4. 输出窗格将显示数据的快照。字段比较显示流程中与表中字段匹配的字段(如果表已存在)。如果表是新的,则显示一对一字段匹配。



如果存在任何字段不匹配,状态说明会显示任何错误。

- **不匹配:字段将被忽略**:字段存在于流程中,但不在数据库中。除非选择**"创建表"**写入选项并执行完全刷新,否则不会将该字段添加到数据库表中。然后将流程字段添加到数据库表并使用流程输出架构。
- **不匹配:字段将包含 Null 值**:字段存在于数据库中,但不在流程中。流程将 Null 值传递到字段的数据库表。如果该字段确实存在流程中,但由于字段名不同而不匹配,您可以导航到清理步骤并编辑字段名以匹配数据库字段名。有关如何编辑字段名的详细信息,请参见**应用清理操作** 在本页222。
- **错误:字段数据类型不匹配**:分配给流程中字段和向其中写入输出的数据库表的数据类型必须匹配,否则流程将失败。您可以导航到清理步骤并编辑字段数据类型来修复此问题。有关更改数据类型的信息,请参见**查看分配给数据的数据类型** 在本页153。

5. 选择写入选项:您可以为完全刷新和增量刷新选择其他选项,并在选择流程运行方法时应用该选项。有关使用增量刷新运行流程的详细信息,请参见**使用增量刷新来刷新流程数据** 在本页390。
   - **追加到表**:此选项将数据添加到现有表中。如果表不存在,则在首次运行流程时创建该表,并在每次后续流程运行时将数据添加到该表中。
   - **创建表**:此选项将创建新表。如果具有相同名称的表已存在,则删除现有表并替换为新表。为表定义的任何现有数据结构或属性也将被删除,并替换为流程数据结构。流程中存在的任何字段都将添加到新的数据库表中。
   - **替换数据**:此选项删除现有表中的数据并将其替换为流程中的数据,但保留数据库表的结构和属性。
6. (可选)单击**"自定义 SQL"**选项卡并输入 SQL 脚本。您可以输入要在将数据写入表**之前**和**之后**运行的脚本。



7. 单击**"运行流程"**以运行流程并将数据写入所选的数据库。

将流程输出数据保存到 CRM Analytics 中的数据集

从版本 2022.3 开始在 Tableau Prep Builder 和 Web 上受支持。

注意：在整合外部来源的数据时，CRM Analytics 有几个要求和一些限制。为了确保您可以成功地将流程输出写入 CRM Analytics，请参见 Salesforce 帮助中的将数据集成到数据集之前的注意事项。

使用 Tableau Prep 清理数据，并在 CRM Analytics 中获得更好的预测结果。只需通过 Tableau Prep Builder 或 Web 上 Tableau Prep 支持的任何连接器连接到数据。然后，应用转换来清理您的数据，并将您的流程数据直接输出到您有权访问的 CRM Analytics 中的数据集。

无法使用命令行界面运行向 CRM Analytics 输出数据的流程。您可以使用 Tableau Prep Builder 手动运行流程，也可以使用 Tableau Prep Conductor 在 Web 上按计划运行流程。

先决条件

若要将流程数据输出到 CRM Analytics，请检查您在 Salesforce 和 Tableau 中是否拥有以下许可证、访问权限和许可。

Salesforce 要求

要求	描述
Salesforce 权限	您必须获得 **CRM Analytics Plus** 或 **CRM Analytics Growth** 许可证。 **CRM Analytics Plus** 许可证包括权限集： • **CRM Analytics Plus 管理员**：启用管理 CRM Analytics 平台和 Einstein Discovery 所需的所有权限，包括创建和管理 CRM Analytics 模板化应用和 Apps 的权限。 • **CRM Analytics Plus 用户**：启用使用 CRM Analytics 平台和 Einstein Discovery 以及 CRM Analytics 模板化应用程序以及应用程序所需的所有权限。

要求	描述
	CRM Analytics Growth 许可证包括权限集： • **CRM Analytic Growth 管理员**：启用管理 CRM Analytics 平台所需的所有权限，包括创建和管理 CRM Analytics 模板化应用和 Apps 的权限。 • **CRM Analytics Growth 用户**：启用使用 CRM Analytics 平台和 CRM Analytics 模板化应用程序以及应用程序所需的所有权限。 有关详细信息，请参见 Salesforce 帮助中的了解 CRM Analytics 许可证和权限集以及选择和分配用户权限集。
管理员设置	Salesforce 管理员将需要配置： • **Tableau Prep 扩展程序**：配置 Salesforce，以便针对 Tableau Server 创建已连接应用（基本）。仅对于 Tableau Server 是必需的。

Tableau Prep 要求

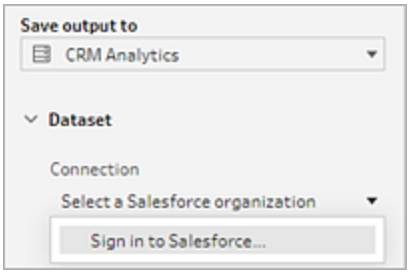
要求	描述
Tableau Prep 许可证和权限	“Creator”许可证 作为创建者，您需要登录您的 Salesforce org 帐户并进行身份验证，然后才能选择“应用和数据集”来输出您的流数据。
OAuth 数据连接	作为服务器管理员，在连接器上使用 Oauth 客户端 ID 和密码配置 Tableau Server。这是在 Tableau Server 上运行流程所必需的。 有关详细信息，请参见 Tableau Server 帮助中的针对 Salesforce.com Oauth 配置 Tableau Server。

将流程数据保存到 CRM Analytics

从 Tableau Prep Builder 保存到 CRM Analytic s时，以下 CRM Analytics 输入限制适用。

- 外部数据上传的最大文件大小：40 GB
- 连续 24 小时内所有外部数据上传的最大文件大小：50 GB

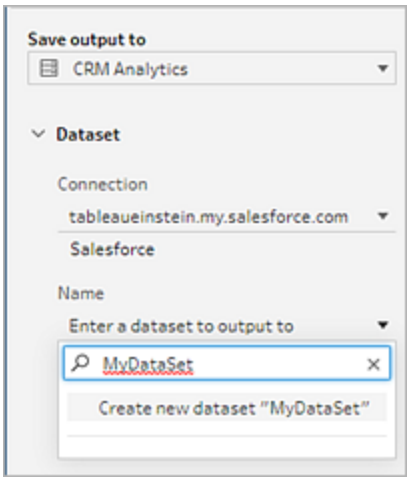
1. 单击步骤上的加号图标 ⊕，并选择**“添加输出”**。
2. 从**“将输出保存到”**下拉列表中，选择**“CRM Analytics”**。



3. 在**“数据集”**部分，连接到 Salesforce。

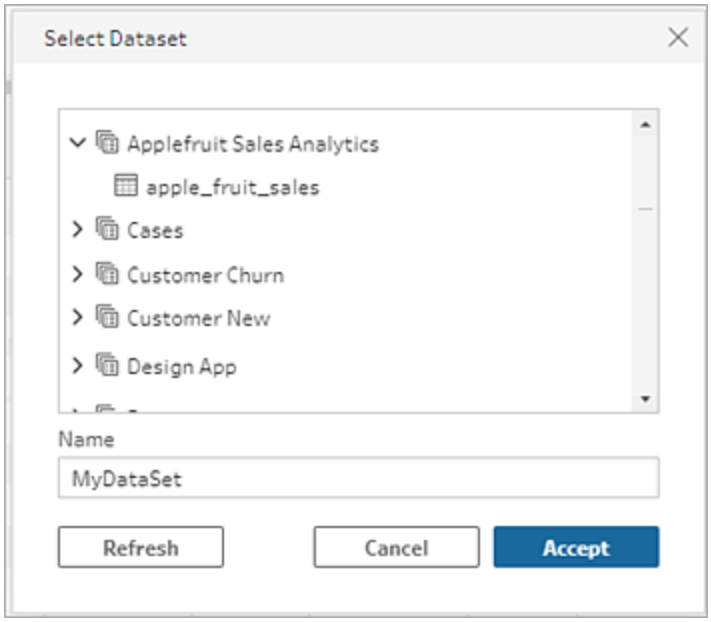
   登录 Salesforce 并单击**“Allow”（允许）**以授予 Tableau 访问 CRM Analytics 应用程序和数据集的权限，或选择现有的 Salesforce 连接
4. 在**“Name”（名称）**字段中，选择现有数据集名称。这将用您的流程输出覆盖和替换数据集。否则，键入新名称并单击**“Create new dataset”（创建新数据集）**，以在所选 CRM Analytics 应用软件创建新数据集。

   注意：数据集名称不能超过 80 个字符。



5. 在**“Name”(名称)**字段下，验证显示的应用程序是您有权写入的应用程序。

   若要更改应用程序，请单击**“Browse Datasets”(浏览数据集)**，然后从列表中选择应用程序，在**“Name”(名称)**字段中输入数据集名称，然后单击**“Accept”(接受)**。



6. 在**“Write Options”(写入选项)**部分，**“Full refresh”(完全刷新)**和**“Create table”(创建表)**是唯一受支持的选项。
7. 单击**“Run Flow”(运行流程)**以运行流程并将数据写入 CRM Analytics 数据集。

   如果您的流程运行成功，您可以在数据管理器的“Monitor”(监控)选项卡中验证 CRM Analytics 中的输出结果。有关此功能的详细信息，请参见 Salesforce 帮助中的监控外部数据上载。

将流程输出数据保存到 Data Cloud

从版本 2023.3 开始在 Tableau Prep Builder 和 Web 上受支持。

使用 Tableau Prep 准备数据，然后将数据与 Data Cloud 中的现有数据集关联。使用 Tableau Prep Builder 或 Web 上的 Tableau Prep 支持的任何连接器导入数据、清理和准备数据，然后使用摄取 API 将流程数据直接输出到 Data Cloud。

权限先决条件

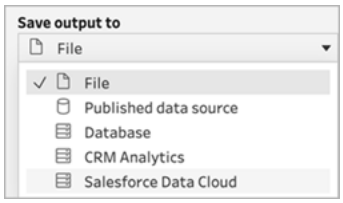
Salesforce 许可证	有关 Data Cloud 版本和附加许可证的信息，请参见 Salesforce 帮助中的 Data Cloud 标准版本和许可证。 另请参见 Data Cloud 限制和指南。
数据空间权限	您必须被分配到数据空间并被分配到 Data Cloud 中的以下权限集之一： • Customer Data Cloud 管理员 • 适用于营销管理员的 Data Cloud • 适用于营销数据感知专家的 Data Cloud 有关详细信息，请参见管理数据空间和使用旧版权限集管理数据空间。
摄取到 Data Cloud 权限	您必须被分配以下权限才能进行现场访问以摄取到 Data Cloud： • Data Cloud Salesforce 连接器 有关详细信息，请参见启用对象和字段权限。
Salesforce 配置文件	启用配置文件访问权限： • 管理权限：启用 API • 已连接应用访问权限：*NameOfConnectedApp*
Tableau Prep 许可证和权限	“Creator”许可证作为创建者，您必须登录您的 Salesforce org 帐户并进行身份验证，然后才能选择“应用和数据集”来输出您的流数据。

将流程数据保存到 Data Cloud

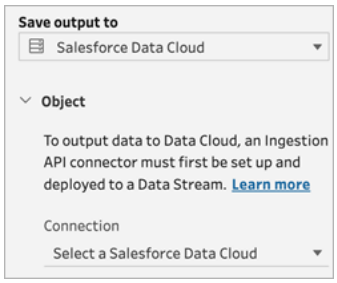
如果您已在使用摄取n API 并手动调用 API 将数据集保存到 Data Cloud，则可以使用 Tableau Prep 简化该工作流程。Tableau Prep 的先决条件配置相同。

如果这是您第一次将数据保存到 Data Cloud，请遵循**Data Cloud 设置先决条件** 在本页 383中的设置要求。

1. 单击步骤上的加号图标 ⊕，并选择**“添加输出”**。
2. 从“将输出保存到”下拉列表中，选择**“Salesforce Data Cloud”**。



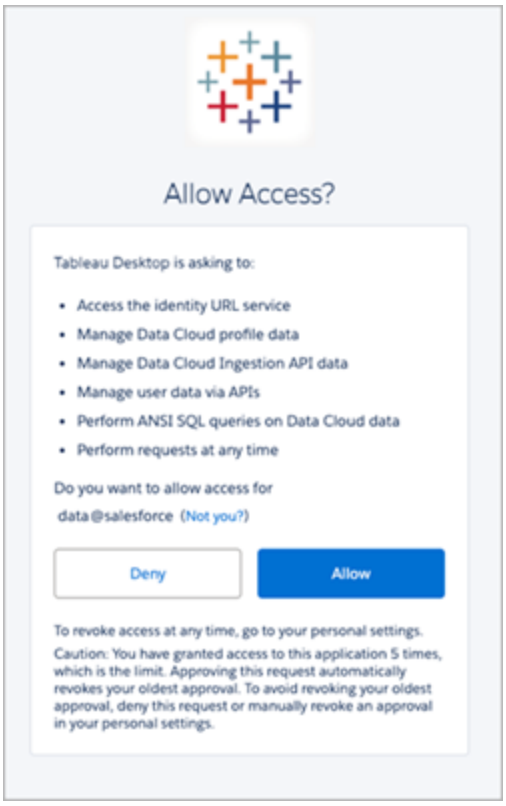
3. 从“对象”部分，选择要登录的 Salesforce Data Cloud 组织。



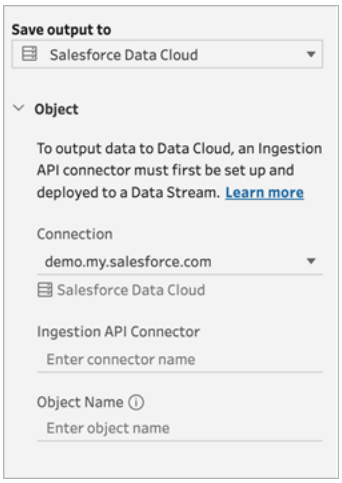
4. 从“Salesforce Data Cloud”菜单中，单击**“登录”**。

   浏览器窗口将打开并显示 *https://login.salesforce.com/*。

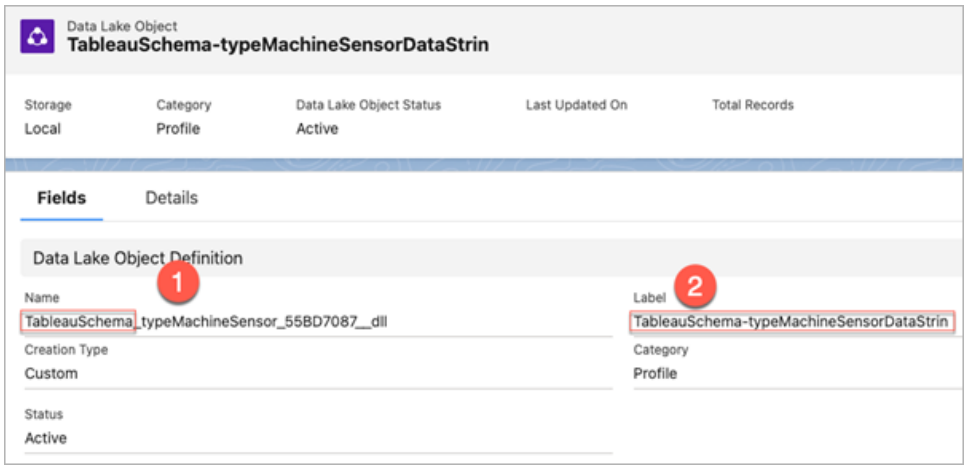
5. 使用您的用户名和密码登录 Data Cloud 组织。
6. 在“Allow Access”(允许访问)表单中，单击**“Allow”(允许)**。



7. 在“Save output to”(将输出保存到)部分输入摄取 API 连接器和对象名称。



Data Lake 对象名称是连接器和对象名称的组合：*IngestionAPIConnectorName-ObjectName*。在下面的示例中，连接器名称为 *TableauSchema* (1)。对象名称为 *typeMachineSensorDataString* (2)。



8. “Write Options”(写入选项)部分指示如果表中已存在指定值，则将更新现有行；如果指定值尚不存在，则将插入新行。
9. 单击**“Run Flow”(运行流程)**以运行流程并将数据写入 Data Cloud。
10. 通过查看数据流中的运行状态和数据资源管理器中的对象来验证 Data Cloud中的数据。

注意事项

- 您可以一次运行一个流程。必须先在 Data Cloud 中完成运行，然后才能运行另一个保存输出。
- 将流程保存到 Data Cloud 的等待时间可能需要一些时间才能完成。在 Data Cloud 中检查状态。
- 使用 Upsert 功能将数据保存到 Data Cloud。如果文件中的记录与现有记录匹配，则现有记录将使用数据中的值进行更新。如果未找到匹配项，则将记录创建为新实体。
- 对于 Prep Conductor，如果您计划自动运行相同的流程，则数据将不会更新。这是因为仅支持 Upsert。
- 在保存到 Data Cloud 过程中您无法中止作业。
- 不会对保存到 Data Cloud 的字段进行验证。验证 Data Cloud 中的数据。

Data Cloud 设置先决条件

这些步骤是将 Tableau Prep 流程保存到 Data Cloud 的先决条件。有关 Data Cloud 概念以及在 Tableau 数据源和 Data Cloud 之间映射数据的详细信息，请参见关于 Salesforce Data Cloud。

设置摄取 API 连接器

通过上载具有 .yaml 文件扩展名的 OpenAPI (OAS) 格式的架构文件，依据源对象创建摄取 API 数据流。架构文件描述了您网站上的数据的结构。有关详细信息，请参见**YMAL文件示例** 在本页387和摄取 API。

1. 单击**“设置”**齿轮图标，然后单击**“Data Cloud 设置”**。
2. 单击**“摄取 API”**。
3. 单击**“新建”**并提供连接器名称。
4. 在新连接器的详细信息页面上，上载 OpenAPI (OAS) 格式的架构文件，文件扩展名为 `.yaml`。架构文件描述了通过 API 传输的数据的结构。

   > **注意**：摄取 API 架构有设定的要求。在摄取之前，请参见**架构要求** 在本页386。

5. 单击“预览架构”表单上的**“保存”**。

创建数据流

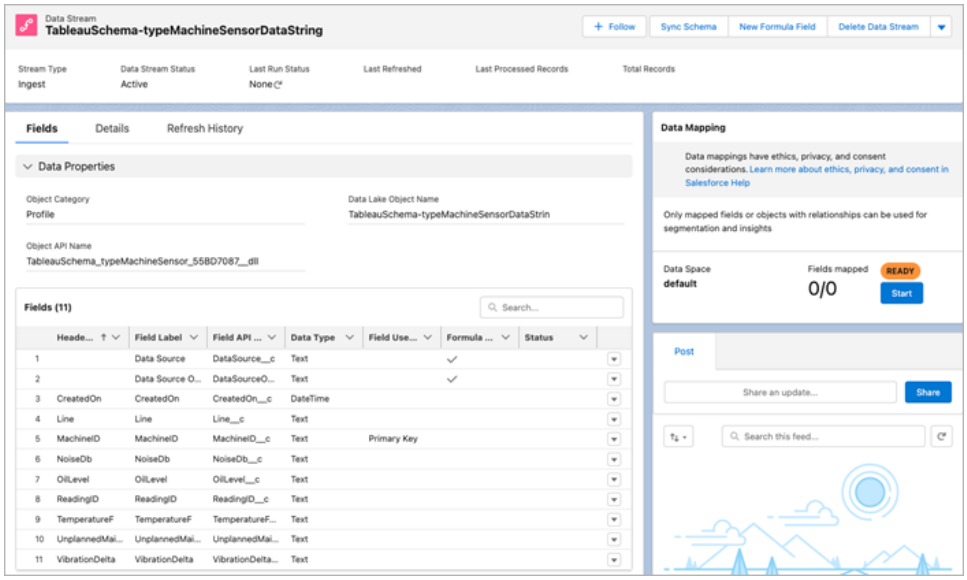
数据流是引入 Data Cloud 的数据源。它由摄取到 Data Cloud 中的连接和关联数据组成。

1. 转到应用程序启动器并选择**“Data Cloud”**。
2. 单击**“Data Streams”(数据流)**选项卡。
3. 单击“New”(新建)并选择**“Ingestion API”(摄取 API)**，然后单击**“Next”(下一步)**。
4. 选择摄取 API 和对象。
5. 选择“Data Space”(数据空间)、“Category”(类别)和“Primary Key”(主键)，然后单击**“Next”(下一步)**。

   必须为 Data Cloud 使用真正的主键。如果不存在，您需要为主键创建一个公式字段。

   对于“Category”(类别)，请在“Profile”(简档)、“Engagement”(参与)或“Other”(其他)之间进行选择。对于用于参与类别的对象，必须存在日期时间字段。简档或其他类型的对象不强加同样的要求。有关详细信息，请参见类别和主键。
6. 单击**“Deploy”(部署)**。

您现在拥有一个数据流和一个数据湖对象。您的数据流现在可以添加到数据空间。



将数据流添加到数据空间

当您将数据从任何来源引入 Data Cloud 时，您可以将数据湖对象 (DLO) 关联到相关数据空间（无论是否有筛选器）。

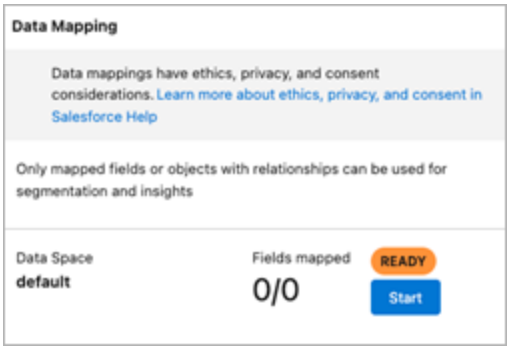
1. 单击**“Data Spaces”（数据空间）**选项卡。

2. 选择默认数据空间或您分配到的数据空间的名称。
3. 单击**“Add Data”（添加数据）**。
4. 选择您创建的数据湖对象，然后单击**“Next”（下一步）**。
5. （可选）为对象选择筛选器。
6. 单击**“Save”（保存）**。

将数据湖对象映射到 Salesforce 对象

数据映射将数据湖对象字段与数据模型对象 (DMO) 字段相关联。

1. 转到“Data Stream”(数据流)选项卡并选择您创建的数据流。



2. 从“Data Mapping”(数据映射)部分,单击**“Start”(开始)**。

   字段映射画布在左侧显示源 DLO,在右侧显示目标 DMO。有关详细信息,请参见映射数据模型对象。

为 Data Cloud 摄取 API 创建已连接应用

在使用摄取n API 将数据发送到 Data Cloud 之前,您必须配置已连接应用。有关更多详细信息,请参见为 API 集成启用 OAuth 设置和为 Data Cloud 摄取 API 创建已连接应用。

作为为摄取 API 设置的已连接应用的一部分,您必须选择以下 OAuth 范围:

- 访问和管理您的 Data Cloud 摄取n API 数据 (cdp_ingest_api)
- 管理 Data Cloud 简档数据 (cdp_profile_api)
- 对 Data Cloud 数据执行 ANSI SQL 查询 (cdp_query_api)
- 通过 API (api) 管理用户数据
- 随时代表您执行请求(refresh_token、offline_access)

架构要求

若要在 Data Cloud 中创建摄取 API 源,您上载的架构文件必须满足特定要求。请参见摄取 API 架构的要求。

- 上载的架构必须采用有效的 OpenAPI 格式,并带有 .yml 或 .yaml 扩展名。支持 OpenAPI 版本 3.0.x。
- 对象不能有嵌套对象。
- 每个架构至少必须有一个对象。每个对象至少必须有一个字段。
- 对象的字段不能超过 1000 个。
- 对象名称的长度不能超过 80 个字符。
- 对象名称只能包含 a-z、AZ-、0-9、_、-。没有 unicode 字符。

- 字段名称只能包含 a-z、A-Z、0-9、_、-。没有 unicode 字符。
- 字段名称不能是以下任何保留字：date_id、location_id、dat_account_currency、dat_exchange_rate、pacing_period、pacing_end_date、row_count、version。字段名称不能包含字符串 __。
- 字段名称不能超过 80 个字符。
- 字段满足以下类型和格式：
  - 对于文本或布尔类型：字符串
  - 对于数字类型：数字
  - 对于日期类型：字符串；格式：日期时间
- 对象名称不能重复；不区分大小写。
- 对象不能有重复的字段名称；不区分大小写。
- 有效负载中的日期时间数据类型字段必须采用 ISO 8601 UTC Zulu 格式，格式为 yyyy-MM-dd'T'HH:mm:ss.SSS'Z'。

更新架构时，请注意：

- 无法更改现有字段数据类型。
- 更新对象时，该对象的所有现有字段都必须存在。
- 更新后的架构文件仅包含更改的对象，因此您不必每次都提供完整的对象列表。
- 对于用于参与类别的对象，必须存在日期时间字段。`profile` 或 `other` 类型的对象不会强加同样的要求。

YMAL文件示例

```
openapi: 3.0.3
  components:
    schemas:
      owner:
        type: object
        required:
          - id
          - name
          - region
          - createddate
        properties:
          id:
            type: integer
            format: int64
```

```
    name:
      type: string
      maxLength: 50
    region:
      type: string
      maxLength: 50
    createddate:
      type: string
      format: date-time
car:
  type: object
  required:
    - car_id
    - color
    - createddate
  properties:
    car_id:
      type: integer
      format: int64
  color:
    type: string
    maxLength: 50
  createddate:
    type: string
    format: date-time
```

将流程输出数据保存到 Amazon S3

在 Tableau Prep Builder 2024.2 及更高版本以及 Web 制作和 Tableau Cloud 中可用。此功能尚未在 Tableau Server 中提供。

您可以从 Tableau Prep Builder 或 Web 支持的任何连接器连接到数据，并将流程输出作为 .parquet 或 .csv 文件保存到 Amazon S3。输出可以保存为新数据，也可以覆盖现有的 S3 数据。为了防止数据丢失，您可以使用“自定义 SQL”选项复制表数据，并在将流程数据保存到 S3 之前运行它。

保存流程输出和连接到 S3 连接器是彼此独立的。您不能重复使用用作 Tableau Prep 输入连接的现有 S3 连接。

您可以在 Amazon S3 中存储的数据总量和对象数量不受限制。单个 Amazon S3 对象的大小范围可以从最小 0 字节到最大 5 TB。一次 PUT 可上传的最大对象为 5 GB。对于大于 100 MB 的对象，客户应该考虑使用分段上传功能。请参见使用分段上传来上传和复制对象。

权限

若要写入 Amazon S3 存储桶，您需要您的存储桶区域、存储桶名称、访问密钥 ID 和密文访问密钥。若要获取这些密钥，您需要在 AWS 中创建一个身份和访问管理 (IAM) 用户。请参见管理 IAM 用户的访问密钥。

将流程数据保存到 Amazon S3

1. 单击步骤上的加号图标 ⊕，并选择**“添加输出”**。
2. 从“将输出保存到”下拉列表中，选择**“数据库和云存储”**。
3. 从“表 > 连接”部分中，选择**“Amazon S3 (仅输出)”**。
4. 在“Amazon S3(仅输出)”表单中，添加以下信息：
   - **访问密钥 ID**：您用于签署发送到 Amazon S3 的请求的密钥 ID。
   - **密文访问密钥**：用于验证您是否有权访问 AWS 资源的安全凭据 (密码、访问密钥)。
   - **存储桶区域**：Amazon S3 存储桶位置 (AWS 区域终端节点)。例如：us-east-2。
   - **存储桶名称**：您要写入流程输出的 S3 存储桶的名称。同一区域内任意两个 AWS 帐户的存储桶名称不能相同。

   注意：若要查找您的 S3 区域和存储桶名称，请登录您的 AWS S3 帐户并导航到 AWS S3 控制台。
5. 单击**“登录”**。
6. 在“S3 URI”字段中，输入 `.csv` 或 `.parquet` 文件的名称。默认情况下，该字段填充了以下内容：`s3://<your_bucket_name>`文件名必须包含扩展名 `.csv` 或 `.parquet`.

   您可以将流程输出保存为新的 S3 对象或覆盖现有的 S3 对象。
   - 对于新的 S3 对象，请键入 `.parquet` 或 `.csv` 文件的名称。URI 显示在预览文本中。例如：`s3://<bucket_name><name_file.csv>`。
   - 若要覆盖现有的 S3 对象，请键入 `.parquet` 或 `.csv` 文件的名称，或单击**“浏**

览”查找现有的 S3 .parquet 或 .csv 文件。

> **注意**：“浏览对象”窗口将仅显示以前登录 Amazon S3 时保存的文件。

7. 对于“写入选项”，将使用流程中的数据创建一个新的 S3 对象。如果数据已经存在，则删除为对象定义的任何现有数据结构或属性，并替换为现的流程数据。流程中存在的任何字段都将添加到新的 S3 对象。
8. 单击**“Run Flow”（运行流程）**以运行流程并将数据写入 S3。

   您可以通过登录您的 AWS S3 帐户并导航到 AWS S3 控制台来验证数据是否已保存到 S3。

使用增量刷新来刷新流程数据

> **注意**：从版本 2020.4.1 开始，您现在可以在 Tableau Server 和 Tableau Cloud 中创建和编辑流程。除非特别指出，本主题中的内容适用于所有平台。有关在 Web 上制作流程的详细信息，请参见 Tableau Server 帮助中的 Web 上的 Tableau Prep。

从 Tableau Prep Builder 版本 2020.2.1 开始以及在 Web 上，您可以将流程输入和输出配置为以增量方式刷新，以便在流程运行时仅检索和处理新行，从而节省时间和资源。

举例来说，如果流程包含每天更新的事务数据，则可以设置增量刷新以便每天仅检索和处理新事务，然后每周或每月运行一次完全刷新以刷新所有流程数据。

> **注意**：若要对使用 Salesforce 连接器的流程输入运行增量刷新，您必须使用 Tableau Prep Builder 版本 2021.1.2 或更高版本。将流程输出写入 Microsoft Excel 或 CRM Analytics 时，当前不支持增量刷新。

若要使用增量刷新来运行流程，Tableau Prep 需要以下信息：

- 用于检测输入表中的新行的字段。
- 用于将流程输出中最后一个处理的值与输入中的值进行比较以确定哪些行是新行的字段。有关详细信息，请参见使用 Append 进行增量刷新 在本页 394。
- 您希望如何将新数据写入表。您可以将新数据添加到现有表、使用新数据覆盖表数据，或者从 Tableau Prep Builder 版本 2020.3.1 开始以及在 Web 上，替换现有表中的数据。

流程刷新选项

Tableau Prep 使您能够选择如何刷新数据，以及如何使用流程输出更新表。下表描述了不同的选项及其优点。

刷新组合	处理的数据	表更新	好处
完全刷新 + 创建表	全部	使用完整数据集创建或覆盖现有表。	在每次流程运行时刷新所有数据。
完全刷新 + 追加到表	全部	将新行添加到现有表。	在每次流程运行时跟踪新数据和现有数据。**“追加到表”**不可用于 .csv 输出类型。
完全刷新 + 替换数据	全部	替换现有表中的行。	保留现有的表架构结构，但将所有数据替换为每个流程运行。
增量刷新 + 创建表	仅新行	仅使用新行创建或覆盖现有表。	创建一个仅使用新行作为完整数据集的新表。 如果流程运行时输出数据源不存在或无法连接，则流程将失败。需要完全刷新才能创建输出，然后才能以增量方式使用它。
增量刷新 + 追加到表	仅新行	将新行添加到现有表。	仅将新行添加到现有表。**“追加到表”**不可用于 .csv 输出类型。请参见使用 Append 进行增量刷新 在本页 394。
增量刷新 + 替换数据	仅新行	仅将现有表中的所有行替换为新行。	保留现有的表架构结构，但将所有数据替换为新行，使此数据集成为您的完整数据集。

配置增量刷新

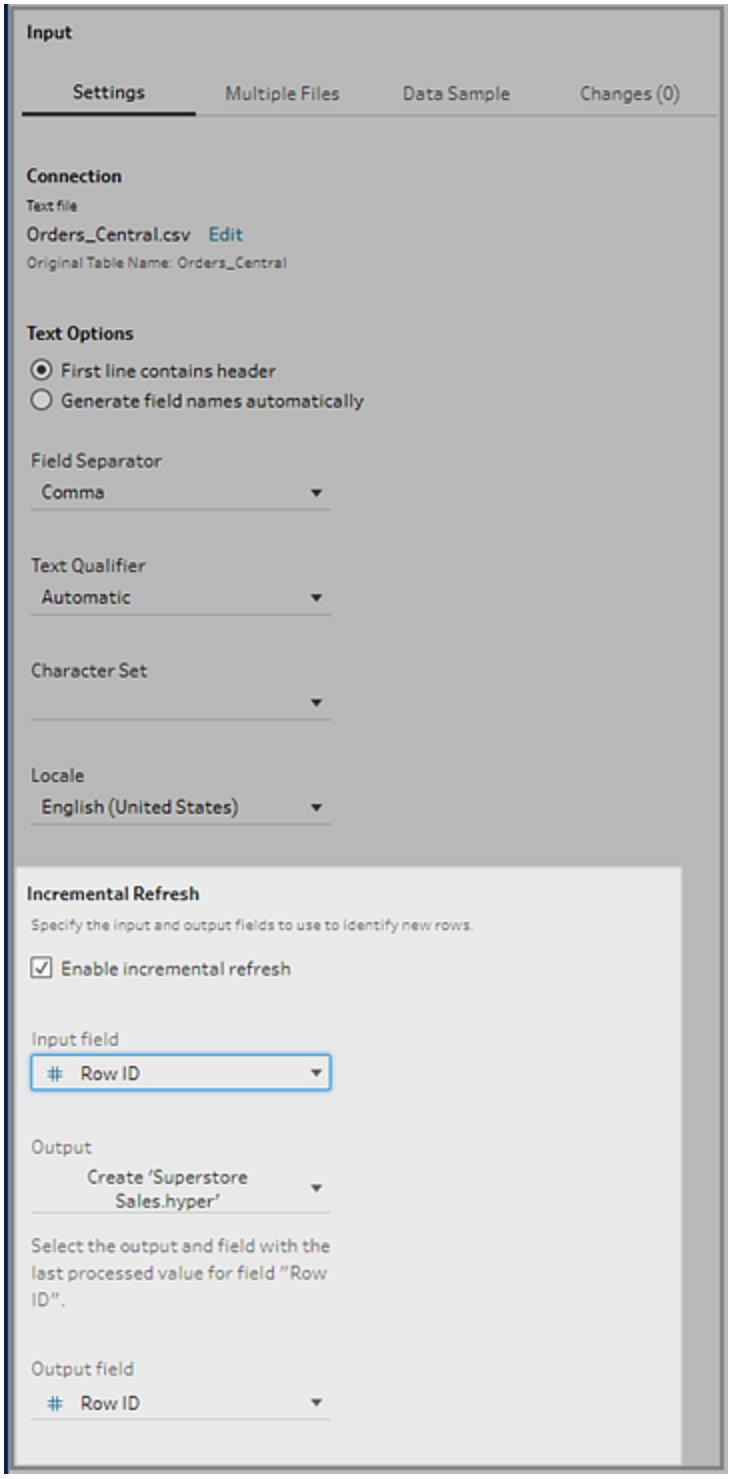
若要将流程配置为使用增量刷新，需要在要使用此选项的**“输入”**步骤和**“输出”**步骤上同时指定设置。在**“输入”**步骤中，指定 Tableau Prep 如何查找新行。在**“输出”**步骤中，指定如何将新行写入表。运行流程时，可以选择完全或增量刷新类型。

提示：为增量刷新配置输入和输出步骤后，可以保留配置并重复使用它们。复制并粘贴步骤以在当前流程的其他位置或 Tableau Prep Builder 中使用它们，或使用**“将步骤另存为流程”**将选定步骤保存到本地文件或服务器以在其他流程中重用这些步骤。有关复制、粘贴或重用步骤的详细信息，请参见**复制步骤、操作和字段** 在本页247。

1. 在流程窗格中，选择要为增量刷新配置的输入步骤。
2. 在**“设置”**选项卡的**“输入”**窗格中，在**“增量刷新”**（以前版本中为**“设置增量刷新”**部分）下设置以下选项：
   - 选择**“启用增量刷新”**（以前版本中为**“启用”**）。
   - **输入字段**（以前版本中为**“使用字段标识新行”**）：选择要在输入数据中刷新的字段。必须为此字段分配**“数字(整数)”**、**“日期”**或**“日期或时间”**数据类型。目前，您只能选择单个字段。

     注意：您可以稍后在流程中移除或重命名此字段，只要在**“输出字段”**（以前版本中为**“输出中的字段名称”**）中指定的字段可用于将此字段与最新输出进行比较来查找新行。
   - **输出**：选择与您的输入相关的输出，其中包括将用于比较行的字段。
   - **输出字段**（以前版本中为**“输出中的字段名称”**）：选择要用于将流程输出中最后处理的值与输入中的值进行比较来查找新行的字段。此字段的数据类型必须与在**“输入字段”**（以前版本中为**“使用字段标识新行”**）中指定的字段的数据

类型相同。


Input
Settings
Multiple Files
Data Sample
Changes (0)
Connection
Text file
Orders_Central.csv
Edit
Original Table Name: Orders_Central
Text Options
First line contains header
Generate field names automatically
Field Separator
Comma
Text Qualifier
Automatic
Character Set
Locale
English (United States)
Incremental Refresh
Specify the input and output fields to use to identify new rows.
Enable incremental refresh
Input field
Row ID
Output
Create 'Superstore Sales.hyper'
Select the output and field with the last processed value for field "Row ID".
Output field
Row ID

使用 Append 进行增量刷新

增量刷新首先在输出中搜索增量字段的现有最大值。然后，它会筛选输入中的行，以仅添加增量字段中值较大的行。例如：

现有表：

第 1 列	第 2 列
ID 5	第 5 行

根据第 1 列将新行追加到表中：

第 1 列	第 2 列
ID 1	新行1
ID 6	新行 6

- 未添加新行1。
- 添加了新行 6。

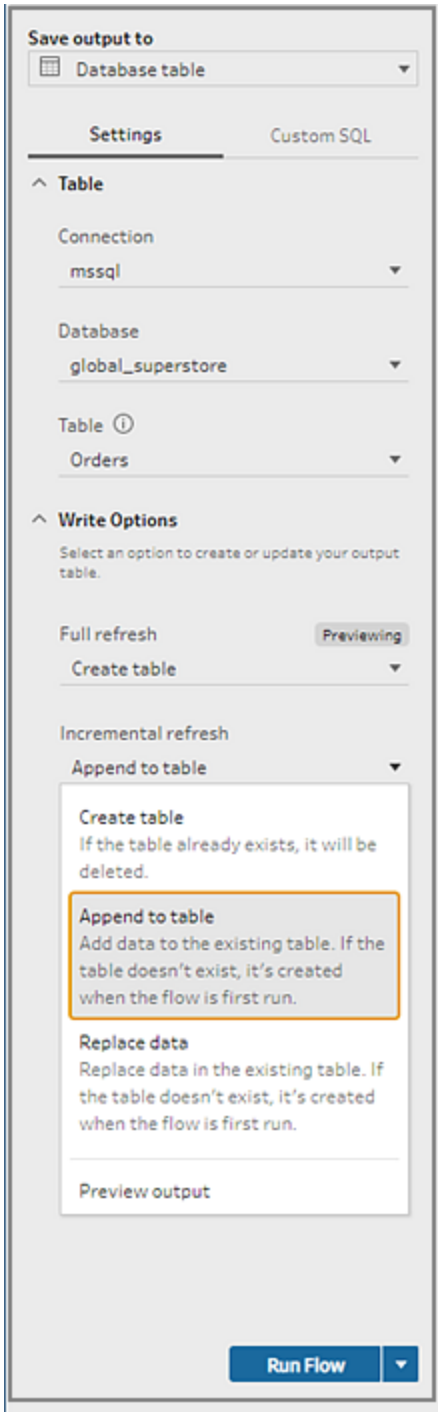
配置写入选项

若要完成增量刷新设置，请设置输出**“写入选项”**以指定如何将新行写入表。与配置的输入步骤相关的所有输出都选择了默认写入选项，但您可以将其更改为支持的选项。

您可以将行输出到文件（仅限 Tableau Prep Builder）、已发布的数据源或数据库。默认情况下，针对本地或已发布的 .hyper 数据提取的输出设置为 **“追加到表”**。针对 .csv 文件类型的输出设置为**“创建表”**。

1. 在流程窗格中，选择要为增量刷新配置的输出步骤。
2. 在“输出”窗格中的**“写入选项”**部分，查看默认写入选项并根据需要进行任何更改。
   - **创建表**：此选项创建新表，或将现有表替换为新输出。
   - **追加到表**：此选项将新数据添加到现有表中。如果表不存在，则在首次运行流程时创建一个新表，后续运行会将新行添加到此表。不可用于 .csv 输出类型。有关支持的刷新组合的详细信息，请参见流程刷新选项 在本页391。
   - **替换数据**（Tableau Prep Builder 版本 2020.3.1 及更高版本以及在 Web 上）：当您要将输出写回数据库中的现有表时，此选项可用。它将数据库表中的数据

替换为流程数据，但保留表架构结构。



运行流程

您可以在 Tableau Prep Builder 中、在 Web 上或通过命令行使用增量刷新来运行单个流程。有关通过命令行运行流程的信息，请参见**在启用增量刷新的情况下运行流程** 在本页 410。

如果启用了包含 Tableau Prep Conductor 的 Data Management，则可以通过在 Tableau Server 或 Tableau Cloud 上设置计划来使用增量刷新运行流程。

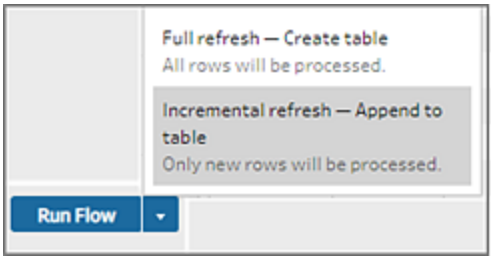
> **注意**：在以前的版本中，写入选项在 Tableau Prep Builder 中设置，在 Tableau Server 或 Tableau Cloud 中运行流程时，无法对其进行更改。从 Tableau Server 和 Tableau Cloud 版本 2020.4 开始，可以直接在 Web 中编辑流程。有关使用 Web 上的 Tableau Prep 的详细信息，请参见 Tableau Server 帮助中的 Web 上的 Tableau Prep。

如果找不到现有输出，无论您选择什么运行选项，Tableau Prep 都会对所有输出运行完全刷新。后续流程运行使用增量刷新过程，并仅检索和处理新行，除非缺少增量刷新配置数据，或者现有输出已移除。

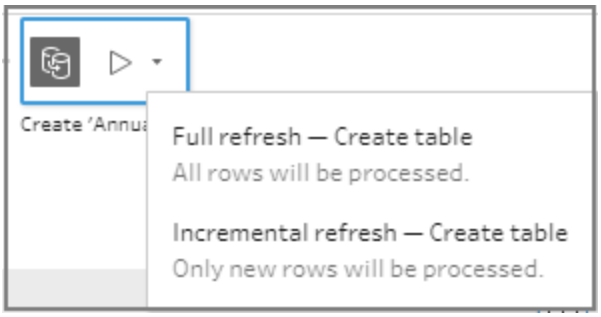
若要使用增量刷新在 Tableau Prep 中运行流程，请从以下位置之一选择**“增量刷新”**：

- 从顶部菜单中，单击**“运行”**按钮上的下拉选项。

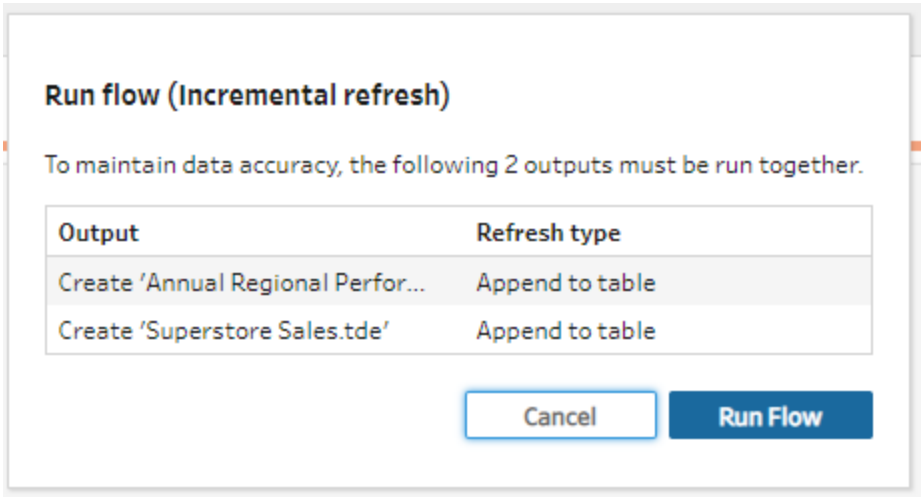
- 从“输出”窗格中，单击**“运行流程”**按钮上的下拉选项。



- 从“流程”窗格中，单击“输出”步骤旁边“运行”按钮上的下拉选项。



如果启用了增量刷新的一个输入与多个输出关联，这些输出必须一起运行，并且必须使用相同的刷新类型。在 Tableau Prep 中运行刷新时，将显示一个对话框，告知您必须同时运行这两个输出。



通过命令行刷新流程输出文件

仅 Tableau Prep Builder 中支持。

可以通过命令行运行流程以刷新流程输出，而不是从 Tableau Prep Builder 中运行流程。可以使用此方法一次运行一个流程。此选项在安装了 Tableau Prep Builder 的 Windows 和 Mac 计算机上都可用。

注意：如果您使用的是基于登录名的许可证管理 (LBLM)，请确保定期打开 Tableau Prep。否则，租约可能会过期，从而导致通过命令行运行的流程失败。您也可以联系您的管理员将您的租用期限更改为最大长度。有关详细信息，请参见基于登录名的许可证管理。

连接器限制：

- **JDBC 或 ODBC 连接器**：从版本 2019.2.3 开始，包含这些连接器的流程可以通过命令行运行。
- **云连接器**：包含云连接器（例如 Google BigQuery）的流程无法通过命令行运行。作为替代，请手动运行流程，或者在 Tableau Server 或 Tableau Cloud 中使用 Tableau Prep Conductor 按计划运行流程。有关详细信息，请参见**使流程数据保持最新** 在本页 428。
- **单点登录身份验证**：如果您使用单点登录身份验证，则不支持从命令行运行流程。您可以改为从 Tableau Prep Builder 中运行流程。
- **多重身份验证**：Tableau Prep 命令行界面 (CLI) 不支持具有多重身份验证 (MFA) 的 Tableau。有关详细信息，请参见 Tableau 知识库中的这篇文章。

对于 Windows 计算机，您也可以使用 Windows 任务计划程序来计划此过程。有关详细信息，请参见 Microsoft 联机帮助中的任务计划程序。

当您通过命令行运行流程时，Tableau Prep Builder 将使用 Tableau Prep Builder 中指定的输出步骤设置刷新流程的所有输出。有关如何指定输出位置的信息，请参见**创建数据提取文件和已发布数据源** 在本页 364。有关设置写入选项（版本 2020.2.1 及更高版本）的信息，请参见**配置写入选项** 在本页 394。

运行流程之前

若要通过命令行运行流程，您需要运行流程的计算机上的管理员权限，并且将需要以下信息：

- Tableau Prep Builder 的安装路径。
- 如果连接到数据库并将输出文件发布到服务器或数据库（版本 2020.3.1 及更高版本），则需要创建一个包含所有必需凭据的凭据 .json 文件。
- Tableau 流程 (.tfl) 文件所在的路径。

凭据 .json 文件要求

注意：如果流程连接到并输出至本地文件、网络共享上存储的文件或使用 Windows 身份验证 (SSPI) 的输入文件，则不需要凭据 .json 文件。有关 Windows 身份验证的详细信息，请参见 Microsoft 联机帮助中的 SSPI 模型。

如果使用远程连接，Tableau Prep Builder 将使用流程文件和凭据 .json 文件中的信息来运行流程。例如，远程连接的数据库名称和输出文件的项目名称来自于流程，服务器名称和登录凭据来自于凭据 .json 文件。

- 如果计划重用数据，请将其放在不会被 Tableau Prep Builder 安装过程覆盖的文件夹中。
- 如果运行的流程包含以下任何一项，则必须加入包含连接所需的凭据的 .json 文件。
  - 连接到数据库文件或已发布数据源。
  - 输出将发布到服务器或数据库（版本 2020.3.1 及更高版本）。
  - 流程包括 Rserve 或 TabPy 的脚本步骤。.json 文件必须包含连接到这些服务所需的凭据。有关详细信息，请参阅下面适用于您的版本的数组要求。
- 流程中指定的凭据和 .json 文件中包括的凭据必须匹配，否则流程将无法运行。
- 运行该过程时，主机名、端口和用户名用于在 Tableau 流程文件 (.tfl) 中查找匹配的连接，并在运行该过程之前更新。“端口 ID”和“站点 ID”是可选的（如果连接不需要此信息）。
- 如果连接到已发布数据源，请在输入连接中包括主机名、连接 URL 和端口（对于 http 为 80，对于 https 为 443）。需要主机名以在 Tableau 流程文件 (.tfl) 中查找匹配的连接，并且内容 URL 和端口用于建立与服务器的连接。
- 如果连接到 Tableau Cloud，请在要连接到的 pod 的输入连接中包含端口（80 或 443），并在服务器连接 URl 中确保包含相应的容器前缀以及 online.tableau.com。有关 Tableau Cloude 的详细信息，请参见 Tableau Cloud 帮助中的 Tableau Bridge 与 Tableau Cloud 的连接。
- （版本 2021.4.1 及更高版本）如果您在流程中包含参数，您可以在命令行中创建并包含参数覆盖 .json 文件，以更改当前默认值的参数值。有关详细信息，请参见**运行包含参数值的流程** 下一页。

根据您的 Tableau Prep Builder 版本，您的凭据信息的格式可能会不同。单击下面的选项卡可查看适用于您的 Tableau Prep Builder 版本的凭据格式。

版本 2020.3.1 及更高版本

根据您的连接，包括服务器或数据库凭据或者两者。当流程连接并输出到同一服务器或数据库时，您只需要在 .json 文件中包含一个块。如果连接到使用不同凭据的服务器或数据库，请使用逗号分隔数组。

服务器连接	**数据库连接**	**Rserver 或 Tableau Python 连接**
连接块名称：“tableauServerConnections”	连接块名称：“databaseConnection	只有在您的流程包含 R 或 TabPy 的脚本步骤时

在数组中包括以下数据： • serverUrl(服务器名称) 对于 Tableau Cloud，包括相应的 pod 前缀和 online.tableau.com。例如 “https://10az.online.tableau.com” • contentUrl(站点 ID) 。此信息出现在 Tableau Server 或 Tableau Cloud URL 中的 /site/ 之后。例如 “https://my.server/#/site/mysite”，设置 `"contentUrl": "mysite"`。)) • port(端口 ID) • username • password	s” 在数组中包括以下数据： • hostname(服务器名称) • port(端口 ID) • username • password	才包括此数组。 连接块名称：“extensions” 在数组中包括以下数据： • extensionName：指定 `"rSupport"` 或 `"pythonSupport"` • regular：包括 “host”和“port”。如果适用，您还可以包括 “username”或 “sslCertificate”(编码为 base64 字符串的公共 .pem 文件内容) 。 • sensitive：包括 “password”(如果使用) 。否则包括一个空白数组。

注意：服务器连接的 .json 文件中始终需要 **ContentUrl**。如果连接到默认站点(例如 “https://my.server/#/site/”，请将 ContentUrl 设置为空白。例如，`"contentUrl": ""`

运行包含参数值的流程

在 Tableau Prep Builder 版本 2021.4.1 及更高版本中受支持。

若要通过命令行运行包含参数值的流程，您可以创建一个包含要使用的参数值的参数覆盖 .json 文件。这些值会覆盖为参数定义的当前(默认) 值。

这是一个独立于您的 credentials.json 文件的文件，包含您的参数名称和值。

> **注意**：从版本 2022.1.1 开始，参数值不再需要用引号括起来。在以前的版本中，所有参数名称和值都必须包含引号。

示例：

```
{
    "Parameter 1": Value 1,
    "Number Parameter": 40,
    "Boolean Parameter": True
}
```

运行流程时，在命令行中包含 -p --parameters 和文件的名称。

示例：

Windows

```
 "\[Tableau Prep Builder install location]\Tableau Prep Builder
<version>\scripts"\tableau-prep-cli.bat -t "path\to\[your flow file
name].tfl" -p|--parameters parameters.override.json
```

Mac

```
/Applications/Tableau\ Prep\ Builder\ [Tableau Prep Builder
version].app/Contents/scripts/./tableau-prep-cli -t path/to/[your flow
file name].tfl -p|--parameters parameters.override.json
```

示例

此部分显示可以使用凭据 .json 要求创建的凭据文件的不同示例。

连接到服务器连接

此示例显示一个 .json 凭据文件，该文件连接并输出到使用相同凭据的服务器连接：

```
{
 "tableauServerConnections":[
   {
    "serverUrl":"https://my.server",
    "contentUrl": "mysite",
    "port":443,
```

```
    "username": "jsmith",
    "password": "passw0rd$"
   }
 ]
}
```

连接到服务器连接和数据库连接输出

此示例显示一个 .json 凭据文件，该文件连接到服务器连接并输出到数据库连接：

```
{
 "tableauServerConnections":[
   {
    "serverUrl":"https://my.server",
    "contentUrl": "mysite",
    "port":443,
    "username": "jsmith",
    "password": "passw0rd$"
   }
],
 "databaseConnections":[
   {
    "hostname":"example123.redshift.amazonaws.com",
    "port":"5439",
    "username":"jsmith",
    "password":"p@s$w0rd!"
   }
 ]
}
```

流程包括 Rserve 和 TabPy 脚本连接并输出到数据库连接

此示例显示一个 .json 凭据文件，其中包括 Rserve 和 Tabpy 凭据并输出到数据库连接：

```
{
 "extensions": [
    {
         "extensionName": "rSupport",
         "regular": {
```

```
                "host": "localhost",
                "port": "9000",
                "username": "jsmith"
        },
        "sensitive": {
          "password": "pwd"
        }
    },
    {
        "extensionName": "pythonSupport",
        "regular": {
          "host": "localhost",
          "port": "9000"
        },
        "sensitive": {
        }
    }
  ],
  "databaseConnections":[
    {
     "hostname":"example123.redshift.amazonaws.com",
     "port": "5439",
     "username": "jsmith",
     "password": "p@s$w0rd!"
    },
    {
     "hostname":"mysql.mydb.tsi.lan",
     "port": "3306",
     "username": "jsmith",
     "password": "mspa$$w0rd"
    }
  ]
}
```

连接并输出到不同的数据库连接

此示例显示一个 .json 凭据文件，该文件连接并输出到不同的数据库连接：

```
{
 "databaseConnections":[
{
  "hostname":"example123.redshift.amazonaws.com",
  "port": "5439",
  "username": "jsmith",
  "password": "p@s$w0rd!"
 },
 {
   "hostname":"mysql.mydb.tsi.lan",
   "port": "3306",
   "username": "jsmith",
   "password": "mspa$$w0rd"
  }
 ]
}
```

版本 2020.2.3 及更低版本

为输入和输出连接输入数组。

注意：如果使用 Tableau Prep Builder 版本 2018.2.2 至 2018.3.1，则始终包括“inputConnections”和“outputConnections”数组，即使流程不为输入或输出使用远程连接。只需将这些数组保留为空白。如果使用 Tableau Prep Builder 版本 2018.3.2 及更高版本，则无需包括空白数组。

输入连接	输出连接	Rserver 或 Tableau Python 连接
• hostname（服务器名称） • contentUrl（已发布数据源始终需要。有关	• serverUrl • contentUrl（站点 ID）。此信息出现在 Tableau Server 或 Tableau Cloud URL 中的 /site/ 之后。例如“https://my.server/#/site/mysite”，设置 `"contentUrl": "mysite"`。）	只有在您的流程包含需要密码的 R 或 TabPy 脚本步骤时才包括此数组。 • extensionName：指定 `"rSupport"` 或 `"pythonSupport"`

说明，请参见**“输出连接”**。 • port(端口 ID) • username • password	• username • password	• credentials：包括“password”。

示例

此部分显示可以使用凭据 .json 要求创建的凭据文件的两个不同示例。

连接到已发布数据源

此示例显示一个 .json 凭据文件，该文件连接到已发布数据源并将数据输出到包含站点 ID 的服务器

> **注意**：如果 inputConnection 或 outputConnection 使用“默认”站点，例如“https://my.server/#/site/”，请将 ContentUrl 设置为空白。例如，`"contentUrl": ""`

```
{
  "inputConnections":[
    {
     "hostname":"https://my.server",
     "contentUrl": "mysite",
     "port":443,
     "username": "jsmith",
     "password": "passw0rd$"
    }
 ],
  "outputConnections":[
    {
     "serverUrl":"https://my.server",
     "contentUrl":"mysite",
     "username":"jsmith",
     "password":"passw0rd$"
    }
```

```
 ]
}
```

连接到两个数据库

此示例显示一个 .json 凭据文件，该文件连接到 MySQL 和 Oracle，并将数据输出到包含站点 ID 的服务器。

```
{
  "inputConnections":[
    {
     "hostname":"mysql.example.lan",
     "port":1234,
     "username": "jsmith",
     "password": "passw0rd"
  },
    {
     "hostname":"Oracle.example.lan",
     "port":5678,
     "username": "jsmith",
     "password": "passw0rd"
  }
],
  "outputConnections":[
    {
     "serverUrl":"http://my.server",
     "contentUrl":"mysite",
     "username":"jsmith",
     "password":"passw0rd$"
    }
  ]
}
```

流程包括 Rserve 和 TabPy 的脚本步骤，并连接到数据库

此示例显示一个 .json 凭据文件，该文件包括 Rserve 和 TabPy 服务的密码，并连接到 MySQL。

```
{
  "inputConnections":[
    {
     "hostname":"mysql.example.lan",
     "port":1234,
     "username": "jsmith",
     "password": "passw0rd"
  }
],
  "extensions":[
    {
     "extensionName":"rSupport",
     "credentials":{
     "password":"pwd",
    }
  },
   {
       "extensionName" : "pythonSupport",
       "credentials": {
       "password": "pwd"
       }
   }
  ]
}
```

创建凭据文件的相关提示

为了避免在运行流程时出现错误，请确保凭据文件遵循以下准则：

- 如果使用 Tableau Prep Builder 版本 2018.2.2 至 2018.3.1，则始终包括“inputConnections”和“outputConnections”数组，即使流程不为输入或输出使用远程连接。只需将这些数组保留为空白。

  如果使用 Tableau Prep Builder 版本 2018.3.2 及更高版本，则无需包括空白数组。

  - 没有远程输入连接？在 .json 文件的顶部包括此语法

```
{
 "inputConnections":[
  ],
```

- 没有远程输出连接？在 .json 文件的底部包括此语法

```
"outputConnections":[
  ]
}
```

- 输入连接没有端口 ID，或者在服务器名称中指定了端口。

  如果连接没有端口 ID，请不要 .json 文件中包括 `"port":xxxx`，.引用，甚至不要包括 `"port": ""`。如果端口 ID 包含在服务器名称中，请在主机名中包括端口 ID。例如，`"hostname": "mssql.example.lan,1234"`

- 引用 `"serverUrl":` 时，不要在地址结尾包括“/”。例如，使用此 `"serverUrl": "http://server"`，而不是此 `"serverUrl": "http://server/"`。
- 如果有多个输入或输出连接，请为文件中的每个连接包括凭据。
- 如果连接到已发布的数据源，请确保在输入连接中包括 hostname 和 contentUrl。

运行流程

重要提示：下面的示例将“Tableau Prep”版本 2019.1.2 的名称更改为“Tableau Prep Builder”。如果您使用的是产品的更低版本，请改用“Tableau Prep”。

1. 以管理员身份打开命令提示符或终端命令提示符 (MacOS)。
2. 使用如下所示语法运行以下命令之一。

   - 流程连接到本地文件或网络共享上存储的文件，并发布至本地文件、网络共享上存储的文件，或使用 Windows 身份验证：

     > **注意：**如果连接到或输出至网络共享上存储的文件，请为路径使用 UNC 格式：\\服务器\路径\文件名。它不能通过密码保护。

     Windows

     ```
      "\[Tableau Prep Builder install location]\Tableau Prep
     Builder <version>\scripts"\tableau-prep-cli.bat -t "path\to\
     [your flow file name].tfl"
     ```

Mac

```
/Applications/Tableau\ Prep\ Builder\ [Tableau Prep Builder
version].app/Contents/scripts/./tableau-prep-cli -t path/to/
[your flow file name].tfl
```

- 流程连接到数据库或发布至服务器:

  Windows

  ```
  "\[Tableau Prep Builder install location]\Tableau Prep
  Builder <version>\scripts"\tableau-prep-cli.bat -c "path\to\
  [your credential file name].json" -t "path\to\[your flow file
  name].tfl"
  ```

  Mac

  ```
  /Applications/Tableau\ Prep\ Builder\ [Tableau Prep Builder
  version].app/Contents/scripts/./tableau-prep-cli -c path/to/
  [your credential file name].json -t path/to/[your flow file
  name].tfl
  ```

- 流程文件或凭据文件存储在网络共享上(为路径使用 UNC 格式:\\服务器\路径\文件名):

  Windows

  ```
  "\[Tableau Prep Builder install location]\Tableau Prep
  Builder <version>\scripts"\tableau-prep-cli.bat -c
  "\server\path\[your credential file name].json" -t
  "\server\path\[your flow file name].tfl"
  ```

  Mac:将网络共享映射到 Finder 中的 /Volumes 以使其保持持久,然后使用 `/Volumes/.../[your file]` 指定路径:

  ```
  /Applications/Tableau\ Prep\ Builder\ [Tableau Prep Builder
  version].app/Contents/scripts/./tableau-prep-cli -c
  /Volumes/.../[your credential file name].json -t path/to/
  [your flow file name].tfl
  ```

有关常见错误和解决方法,请参见**使用命令行运行流程时的常见错误** 在本页507。

在启用增量刷新的情况下运行流程

在 Tableau Prep Builder 版本 2020.2.1 及更高版本以及从版本 2020.4 开始的 Web 上受支持。将流程输出写入 Microsoft Excel 时，当前不支持增量刷新。

如果在服务器上未启用 Tableau Prep Conductor 来计划流程运行，则可以通过命令行中使用增量刷新来运行流程。只需在命令行中包括 `--incrementalRefresh` 参数，如下例所示。

Windows

```
 "\[Tableau Prep Builder install location]\Tableau Prep Builder
<version>\scripts"\tableau-prep-cli.bat --incrementalRefresh -t
"path\to\[your flow file name].tfl"
```

Mac

```
/Applications/Tableau\ Prep\ Builder\ [Tableau Prep Builder
version].app/Contents/scripts/./tableau-prep-cli --incrementalRefresh
-t path/to/[your flow file name].tfl
```

如果流程中的输入步骤启用了增量刷新，并且增量刷新参数配置正确，Tableau Prep Builder 将执行以下操作：

- 流程中启用了增量刷新的所有输入都将使用增量刷新运行所有相应的输出。
- 如果流程中没有输入启用了增量刷新，则将使用完全刷新运行所有输出。一条消息将显示刷新方法详细信息。
- 如果流程中的某些输入启用了增量刷新，则相应的输出将使用增量刷新运行。其他输出将使用完全刷新运行，一条消息将显示刷新方法详细信息。

有关配置流程以使用增量刷新的详细信息，请参见**使用增量刷新来刷新流程数据** 在本页 390

命令选项

如果要查看帮助选项，请在命令行中包括 `-h`。

命令选项	描述	说明
`-c, --connectio-ns <arg>`	凭据文	需要凭据文件所在的路径。

	件的连接路径。	
-d, --debug	调试流程进程。	包括此选项以查看更多信息，来帮助调试刷新流程的问题。日志文件存储位置:My Tableau Prep Builder Repository\Command Line Repository\Logs
-dsv, --disableSslValidation	禁用 SSL 验证 (MacOS)	使用 MacOS 上的命令行运行流程时，可能会显示一个对话框，询问钥匙串用户和密码。从 Tableau Prep Builder 版本 2019.3.2 开始，您可以传入此附加参数来禁用此钥匙串对话框。例如:/Applications/Tableau\ Prep\ Builder\ [Tableau Prep Builder version].app/Contents/scripts/./tableau-prep-cli -dsv -c path/to/[your credential file name].json -t path/to/[your flow file name].tfl
-h, --help	查看语法选项的帮助。	此帮助选项或语法错误显示以下信息: usage: tableau-prep-cli [-c <arg>] [-d] [-h] [-t <arg>] -c,--connections <arg> Path to a file with all connection information -d,--debug This option is for debugging -dsv,--disableSslValidation Disable SSL validation -h,--help Print usage message

		-inc,--incrementalRefresh Run incremental refresh for all outputs that are configured to support it -t,--tflFile <arg> The Tableau Prep Builder flow file
-inc,--incrementalRefresh	为配置为使用它的所有输出运行增量刷新。	包括此选项，以便为配置为使用它的所有输入运行增量刷新。增量刷新使 Tableau Prep Builder 能够仅检索和处理新行，而不是流程中的所有行。 输入步骤上的增量刷新配置设置确定哪些流程输出可以增量方式运行。所有其他输出将使用完全刷新运行，一条消息将显示刷新方法详细信息。 有关使用增量刷新运行流程的详细信息，请参见**使用增量刷新来刷新流程数据** 在本页390。
-t,--tflFile <arg>	.tfl 流程文件	需要 .tfl 流程文件所在的路径。
-p,--parameters	参数覆盖 .jso-	如果要覆盖应用于流程的当前(默认)参数值，请包含此文件。有关使用流程参数的详细信息，请参见**在流程中创建和使用参数** 在本页192。

	n 文件	

语法示例

下面的命令行显示使用以下条件运行流程的四个不同示例：

- **Tableau Prep Builder 版本**：2022.1.1

  重要信息：以下示例将 Tableau Prep 版本 2019.1.2 的名称更改为 Tableau Prep Builderr。如果您使用的是产品的更低版本，请改用“Tableau Prep”。

- **流程名称：**Flow1.tfl
- **流程位置：**C:\Users\jsmith\Documents\My Tableau Prep Builder Repository\Flows
- **凭据文件名：**Flow 1.json
- **凭据文件位置：**C:\Users\jsmith\Desktop\Flow credentials
- **存储在网络共享上的凭据文件位置：**\tsi.lan\files\Flow credentials

流程连接到并发布至本地文件

Windows

```
"\Program Files\Tableau\Tableau Prep Builder
2022.1.1\scripts"\tableau-prep-cli.bat -t
"\C:\Users\jsmith\Documents\My Tableau Prep Builder
Repository\Flows\Flow1.tfl"
```

Mac

```
/Applications/Tableau\ Prep\ Builder\
2022.1.1.app/Contents/scripts/./tableau-prep-cli -t
/Users/jsmith/Documents/My\ Tableau\ Prep\ Builder\
Repository/Flows.Flow1.tfl
```

流程连接并发布到本地文件，并为增量刷新使用短形式

Windows

```
"\Program Files\Tableau\Tableau Prep Builder
2022.1.1\scripts"\tableau-prep-cli.bat -inc -t
```

```
"\C:\Users\jsmith\Documents\My Tableau Prep Builder
Repository\Flows\Flow1.tfl"
```

Mac

```
/Applications/Tableau\ Prep\ Builder\
2022.1.1.app/Contents/scripts/./tableau-prep-cli -inc -t
/Users/jsmith/Documents/My\ Tableau\ Prep\ Builder\
Repository/Flows.Flow1.tfl
```

流程连接到数据库并发布至服务器

Windows

```
"\Program Files\Tableau\Tableau Prep Builder
2022.1.1\scripts"\tableau-prep-cli.bat -c
"\C:\Users\jsmith\Desktop\Flow credentials\Flow1.json" -t
"\C:\Users\jsmith\Documents\My Tableau Prep Builder
Repository\Flows\Flow1.tfl"
```

Mac

```
/Applications/Tableau\ Prep\ Builder\
2022.1.1.app/Contents/scripts/./tableau-prep-cli -c
/Users/jsmith/Desktop/Flow\ credentials/Flow1.json -t
/Users/jsmith/Documents/My\ Tableau\ Prep\ Builder\
Repository/Flows.Flow1.tfl
```

流程发布至服务器，并且凭据文件存储在网络共享上

Windows

```
"\Program Files\Tableau\Tableau Prep Builder
2022.1.1\scripts"\tableau-prep-cli.bat -c "\\tsi.lan\files\Flow
credentials\Flow1.json" -t "\C:\Users\jsmith\Documents\My Tableau
Prep Builder Repository\Flows\Flow1.tfl"
```

Mac

```
/Applications/Tableau\ Prep\ Builder\
2022.1.1.app/Contents/scripts/./tableau-prep-cli -c
/Volumes/files/Flow\ credentials/Flow1.json -t
```

```
/Users/jsmith/Documents/My\ Tableau\ Prep\ Builder\
Repository/Flows.Flow1.tfl
```

与 Tableau Prep 的版本兼容性

如果新版本的 Tableau Prep Builder 中引入了新功能或连接器，并且您使用的是旧版本，则在您尝试打开流程时可能会出现兼容性问题。

注意：从版本 2020.4 开始，可以直接在 Tableau Server 和 Tableau Cloud 上创建和编辑流程。在 Web 上创建的流程将始终与您使用的服务器版本兼容。有关在 Web 上制作流程的详细信息，请参见 Tableau Server 和 Tableau Cloud 帮助中的 **Web 上的 Tableau Prep**。

同样，如果将流程发布到 Tableau Server 或 Tableau Cloud 以将它们计划为使用 Tableau Prep Conductor 运行，并且流程包括的新功能或连接器在您的 Tableau Server 或 Tableau Cloud 版本中不受支持，则可能会遇到兼容性问题，使您无法在计划和运行流程。

版本号格式

从 Tableau Prep Builder 版本 2022.3 开始，发布版本编号方案现在与 Tableau Desktop 和 Tableau Server 保持一致。在以前的版本中，Tableau Desktop 和 Tableau Prep Builder 的版本号有不同的格式。例如：

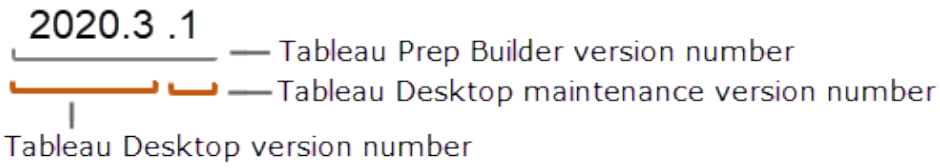


Tableau Desktop 和 Tableau Prep Builder 的维护版本不遵循相同的顺序。

	版本升级示例	第一个维护版本示例
Prep Builder	2022.1.1	2022.1.2
Desktop	2022.1	2022.1.1

查找您的版本

注意：若要下载特定版本的 Tableau Prep Builder，请打开下载页面，并从页面左侧的列表中选择**“Tableau Prep Builder”**。

Tableau Prep Builder

若要查找产品的发布版本，请打开 Tableau Prep Builder，然后在顶部菜单中执行以下操作之一：

- **Windows**：在顶部菜单中，单击**“帮助”>“关于 Tableau Prep Builder”**或**“关于 Tableau Prep”**，具体情况视您的版本而定。
- **Mac**：在顶部菜单中，单击**“Tableau Prep Builder”>“关于 Tableau Prep Builder”**或**“Tableau Prep”>“关于 Tableau Prep”**，具体取决于您的版本。

发布版本号显示在对话框的左下角。

Tableau Server

Tableau Prep Conductor 是作为 Tableau Server 版本 2019.1 的 Data Management 的一部分引入的。若要将流程计划为在 Tableau Server 上运行，您必须使用 Tableau Server 版本 2019.1 或更高版本，并且必须启用 Tableau Prep Conductor。

若要查找 Tableau Server 的版本，请在 Web 浏览器中打开 Tableau Server。在顶部菜单栏中，单击右上角的信息图标，并选择**“关于 Tableau Server”**。将打开一个对话框，告知您使用的 Tableau Server 的版本。有关如何启用 Tableau Prep Conductor 的信息，请参见 Tableau Server 帮助中的步骤 2：配置 Tableau Server 的流程设置。

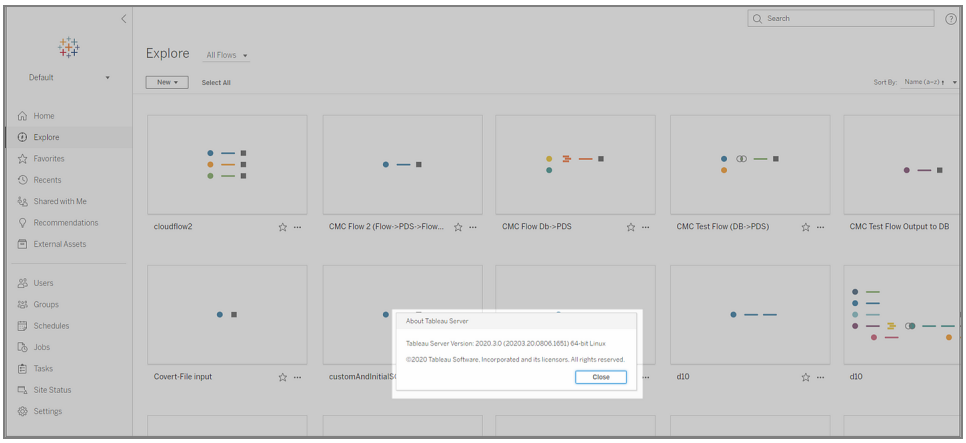


Tableau Cloud

Tableau Prep Conductor 是作为 Tableau Cloud 版本 2019.3 的 Data Management 的一部分引入的。若要将流程计划为在 Tableau Cloud 上运行，您必须使用 Tableau Cloud 版本 2019.3 或更高版本，并且必须启用 Tableau Prep Conductor。

若要查找您的版本，请在 Web 浏览器中打开 Tableau Cloud。在顶部菜单栏中，单击右上角的信息图标，并选择**“关于 Tableau Cloud”**。将打开一个对话框，告知您使用的 Tableau Cloud 的版本。有关启用 Tableau Prep Conductor 的信息，请参见 Tableau Cloud 帮助中的 Tableau Prep Conductor。

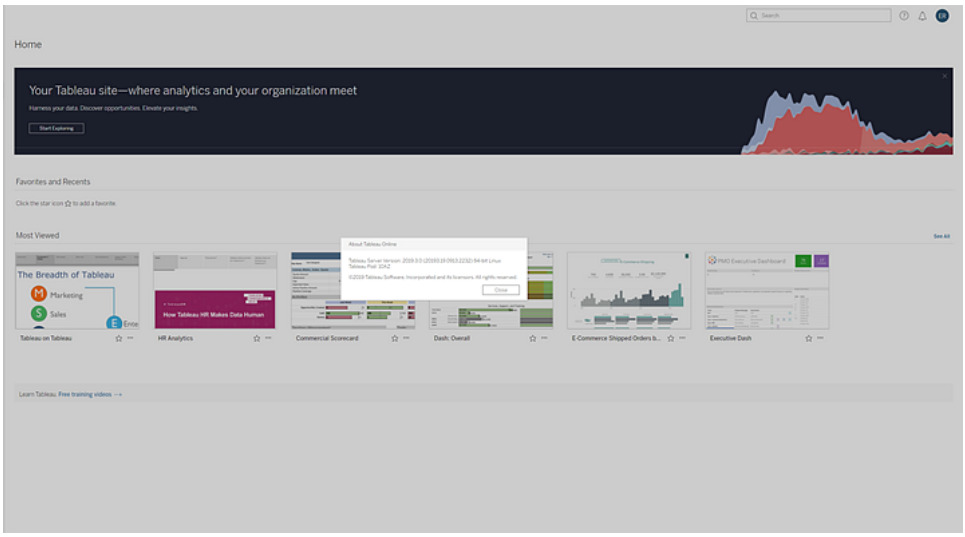


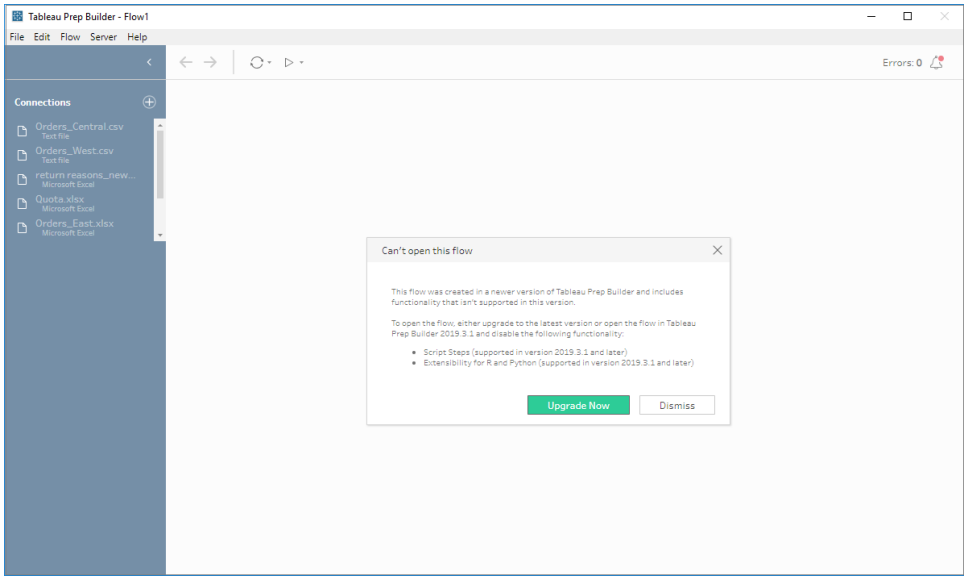
Tableau Prep Builder 不同版本之间的兼容性

通常，Tableau Prep Builder 的新版本可打开在旧版本中创建的流程。但是，当您尝试在新版本的 Tableau Prep Builder 中打开旧版本的流程，以及尝试在旧版本中打开新版本的流程时，或者甚至在使用不同的计算机在相同版本的 Tableau Prep Builder 中打开流程时，都可能会出现兼容性问题。

例如：

- 流程包括的输入连接器或功能在打开流程的版本中不受支持。
- 用于打开流程的计算机没有安装所需的输入连接器，或其具有的连接器驱动程序版本不兼容。Tableau Prep Builder 需要安装 64 位驱动程序才能处理流程输入连接器。

如果在尝试打开流程时出现兼容性问题，流程可能会打开，但包含错误，或者流程根本无法打开，并且您收到错误消息。在下面的示例中，流程不会打开，并且会显示一条错误消息，并列出不兼容的功能以及用于解决问题的选项。



修复 Tableau Prep Builder 的兼容性问题

若要修复兼容性问题，请尝试以下方法之一：

- 升级到 Tableau Prep Builder 的最新版本。

  单击“发现”窗格底部的更新按钮，下载产品的最新版本，并按照 Tableau Desktop 和 Tableau Prep Builder 部署指南中的说明安装 Tableau Prep Builder。如果您对“发现”窗

格上的更新按钮没有访问权限，则有关如何下载产品最新版本的说明包括在“**安装 Tableau Prep Builder**”主题中。
- 确保您的计算机与 Tableau Prep Builder 兼容。例如，确保您为流程使用的连接器安装 64 位驱动程序。若要安装驱动程序，请参见驱动程序下载页面。
- 打开移除了不兼容功能的流程副本。

不同版本的 Tableau Prep Builder 和 Tableau Server 之间的兼容性

从较新版本的 Tableau Prep Builder 发布到较旧版本的 Tableau Server 可能会导致兼容性问题。例如，Tableau Prep Builder 版本 2021.3.1 添加的新功能可能不兼容 Tableau Server 版本 2021.2，但将兼容 Tableau Server 版本 2021.4 和 Tableau Server 的任何更高主要版本，例如版本 2022.3。

在 Tableau Server 中，Tableau Prep Conductor 会在发布流程时检测流程中包括的功能。如果它发现不支持的功能，则流程仍然可发布到 Tableau Server，但无法运行、计划流程或将其添加到任务。Tableau Cloud 会定期自动更新，因此通常与 Tableau Prep Builder 的所有版本兼容。

如果有较旧版本的 Tableau Server，则仍可以在 Tableau Prep Builder 中手动运行或使用命令行运行不兼容的流程。有关使用此过程的详细信息，请参见通过命令行刷新流程输出文件。

检测不兼容的功能

根据所使用的 Tableau Prep Builder 版本，您可以通过不同方式找出不兼容的功能。

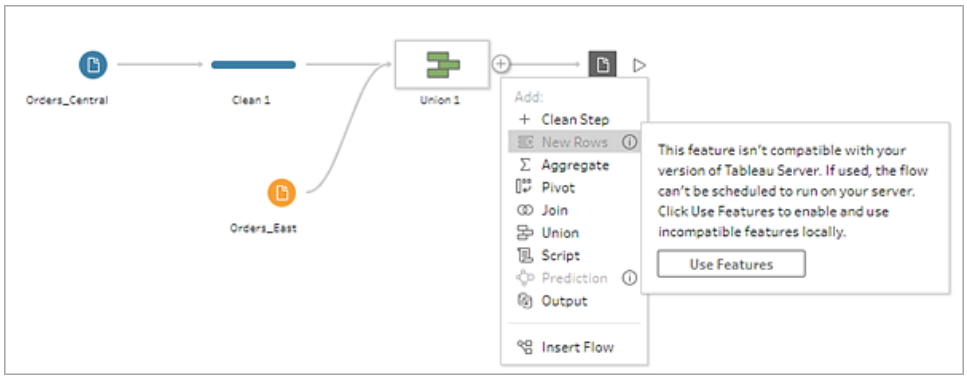
Tableau Prep Builder(版本 2020.1.1 及更高版本)

登录到 Tableau Server，Tableau Prep Builder 将检测并为您禁用不兼容的功能。任何不兼容的功能都将显示为灰色。如果仍要使用该功能并手动或通过命令行运行流程，您可以通过菜单启用它。

注意：从 Tableau Prep Builder 版本 2020.1.4 开始，一旦登录到服务器，Tableau Prep Builder 在关闭应用程序时会记住您的服务器名称和凭据，以便下次打开应用程序时您已经登录到服务器。

1. 将鼠标悬停在禁用的功能上以查看它是不是因为与服务器版本不兼容而禁用的，然后单击**“使用功能”**按钮。此选项可在“流程”窗格以及“配置”窗格、“结果”窗格和数据网格内的菜单中找到。

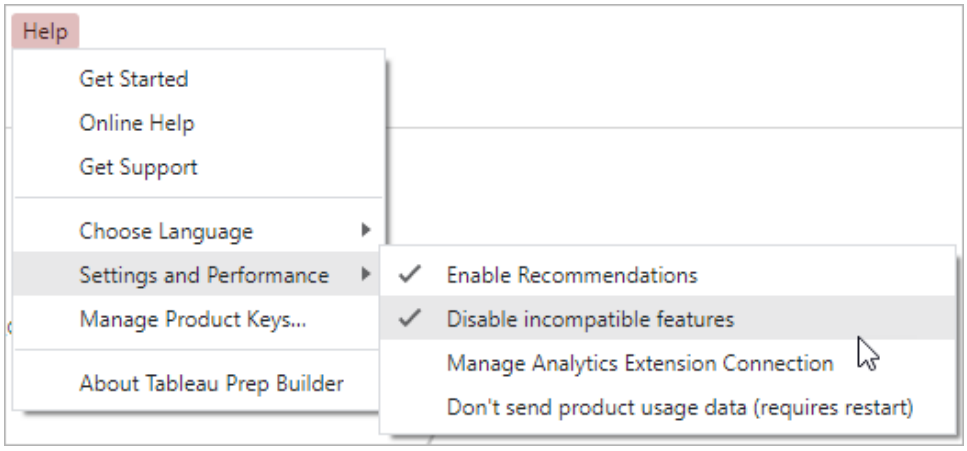
   注意：功能可能会由于其他原因而被禁用，例如数据更新被暂停，或者该选项对于特定步骤或数据类型不可用。



2. 将应用所选功能，并且所有不兼容的功能将启用并可供使用。不兼容的功能带有警告标志，这样，如果您希望使用您的 Tableau Server 版本中的计划运行流程，就能轻松地找到并移除它们。

若要完全禁用此功能并启用所有不兼容的功能，请执行以下操作：

1. 从顶部菜单中，选择**“帮助”>“设置和性能”>“禁用不兼容的功能”**。



2. 选择**“禁用不兼容的功能”**以清除此选项旁边的复选标记。若要再次启用该功能，请选择**“禁用不兼容的功能”**。默认情况此选项应已启用。

Tableau Prep Builder(版本 2019.3.1 及更高版本)

生成流程时，Tableau Prep Builder可以在添加不兼容功能时检测到它们，并使用通知图标标记这些功能。您必须登录到服务器才能看到这些通知。这种通知系统可帮助您快速识别流程中不兼容的功能，以便能够决定是将该功能保留在流程中，还是将其删除。

将鼠标悬停在“流程”窗格中的通知上以查看有关不兼容功能的信息，或使用通知中心来查看更多详细信息。在通知中心中，单击**“在流程中查看”**链接直接导航到触发警告的步骤、注释、字段或变更。

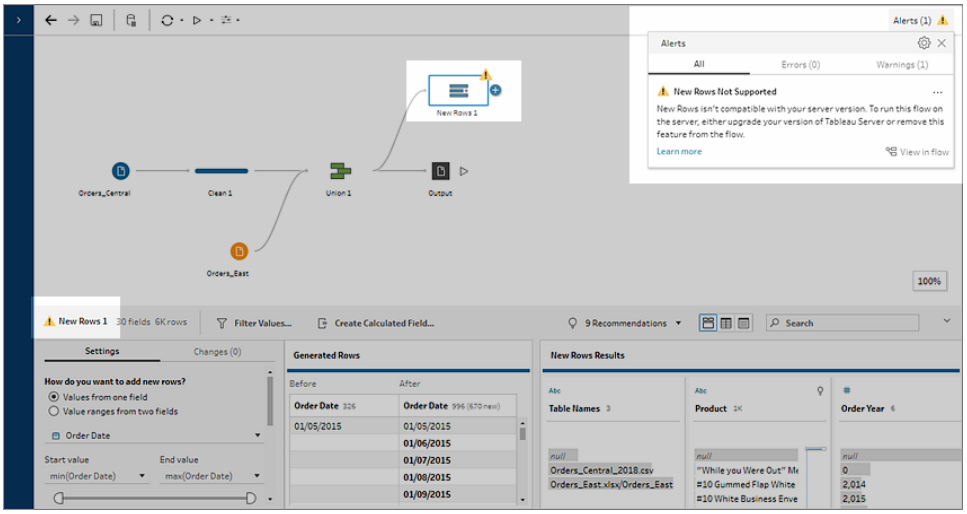
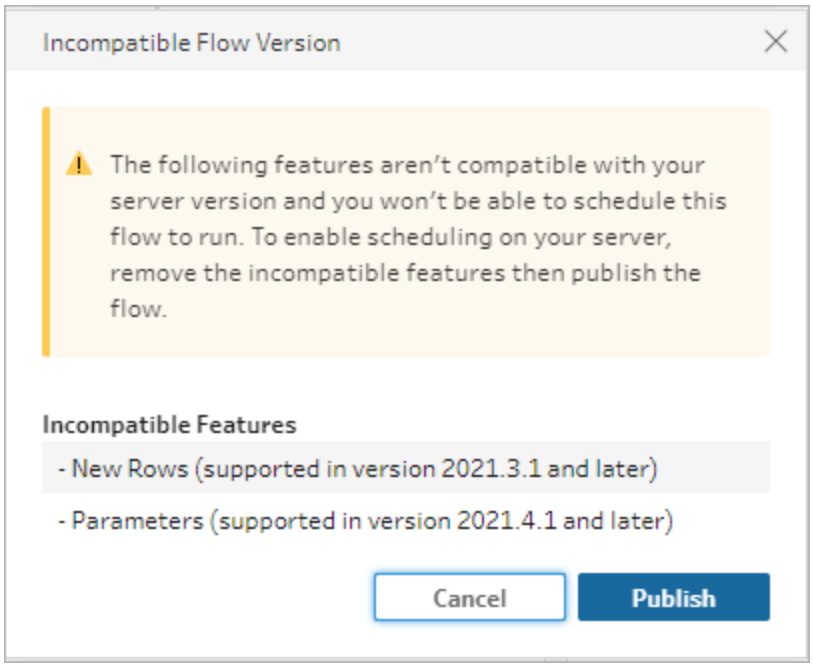


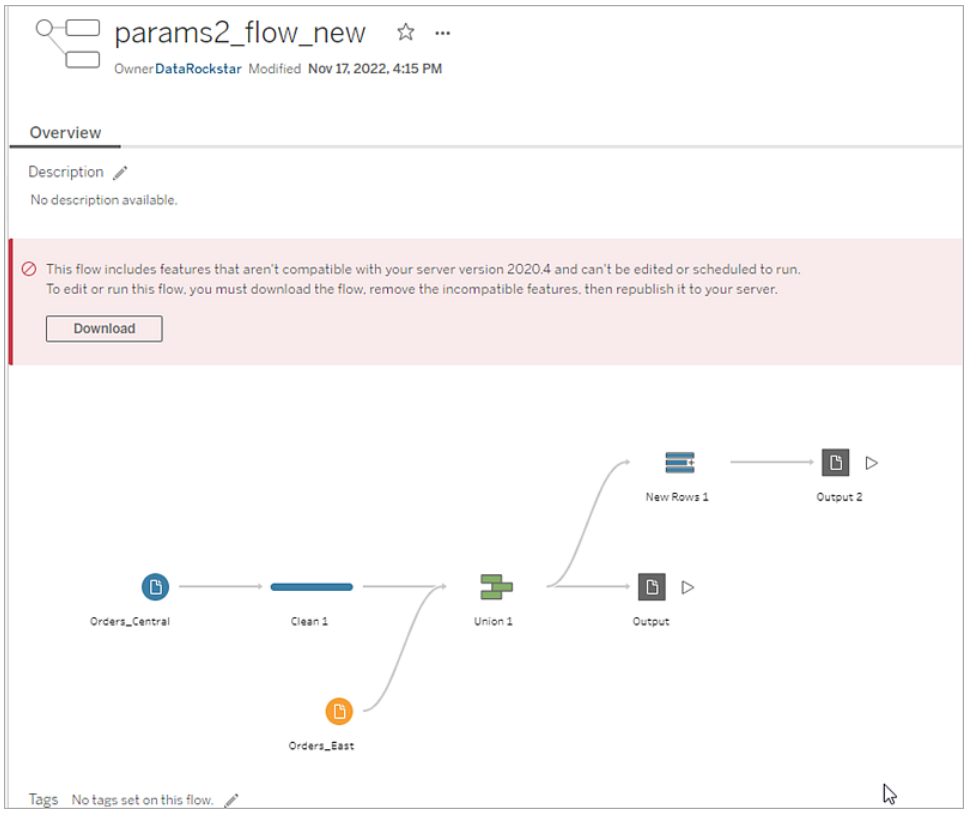
Tableau Prep Builder(所有版本)

如果发布包含不兼容功能的流程，则会显示以下消息，并列出您登录到的 Tableau Server 中不支持的功能。在 Tableau Prep Builder 版本 2019.2.3 及更低版本中，这是确定哪些功能在流程中不兼容的唯一方式。

注意：错误消息列出了引入该功能时的 Tableau Prep Builder 版本。Tableau Prep Builder 不会功能发布维护版本中的功能，因此，为了使功能兼容，或 Tableau Server 必须运行下一个主要发布版本。在下面的示例中，“复制字段”功能是在 Tableau Prep Builder 版本 2019.2.3 中引入的，因此它与 Tableau Server 的维护发布版本 2019.2.3 不兼容。相反，它将与 Tableau Server 的下一个主要版本 2019.3 兼容。



如果继续发布流程，发布将成功完成。但是，在 Tableau Server 或 Tableau Cloud 中打开流程时，您将看到以下消息：



若要在 Tableau Server 中计划并运行流程，您可以执行以下操作之一：

- 查找与您正在使用的 Tableau Prep Builder 版本兼容的 Tableau Server 的最新主要版本。例如，如果您使用 Tableau Prep Builder 版本 2019.2.3 中引入的功能，若要在 Tableau Server 中运行流程，则需要服务器版本为 2020.3 或更高版本。

  Tableau Cloud 会定期自动更新，通常每季度更新一次。在发布之前首先测试您的流程，确保其与当前版本的 Tableau Cloud 兼容。

- 在发布流程之前，从流程中移除不兼容的功能，然后发布流程。
- 如果已将流程发布到 Tableau Server，请尝试在服务器上直接编辑流程（版本 2019.4 及更高版本），下载流程并移除功能，或者在较旧版本的 Tableau Prep Builder 中创建只使用该版本中可用功能的流程。

  注意：若要下载特定版本的 Tableau Prep Builder，请打开下载页面，并从页面左侧的列表中选择**“Tableau Prep Builder”**。

修复兼容性问题

如果已将流程发布到 Tableau Server，请尝试下面的步骤，使用当前版本的 Tableau Prep Builder 移除不兼容的功能。移除了相关功能并且不再看到版本不兼容消息或警告后，请将流程重新发布到 Tableau Server 或 Tableau Cloud，并使用 Tableau Prep Conductor 来计划流程。

注意：Tableau Prep Conductor 是 Data Management 的一部分。必须在 Tableau Server 或 Tableau Cloud 中启用它才能使用计划功能运行流程。有关 Data Management 的详细信息，请参见 Tableau Data Management。有关在 Tableau Server 或 Tableau Cloud 中启用 Tableau Prep Conductor 的详细信息，请参见 Tableau Server 帮助中的步骤 2：配置 Tableau Server 的流程设置或 Tableau Cloud 帮助中的 Tableau Prep Conductor。

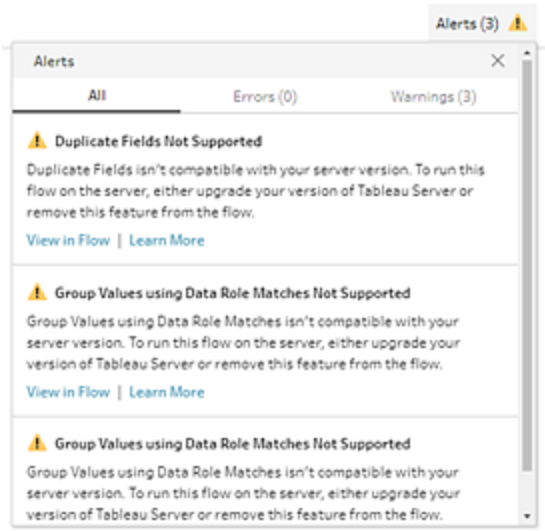
确定不兼容的功能

如果在 Tableau Server 中工作，它当前不会列出流程中不兼容的功能。若要确定要从流程中移除的功能的列表，您需要在 Tableau Prep Builder 中打开流程，然后在流程中查找并移除这些功能。

1. 打开流程。如果您在 Tableau Prep Conductor 中，请从**“更多操作”** ••• 菜单中单击**“下载”**，以下载并在 Tableau Prep Builder 中打开流程，或者只是在 Tableau Prep Builder 中打开流程。
2. 如果下载流程，请单击下载的流程将其打开。
3. 根据您的版本，执行以下操作之一：

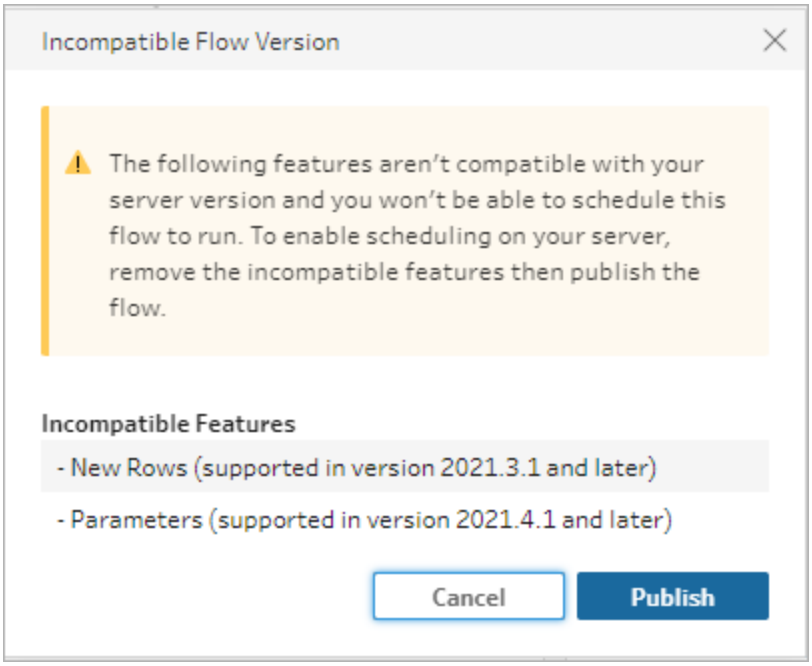
   - **版本 2019.3.1 及更高版本：**从顶部菜单中选择**“服务器”>“登录”**。确保选择与流程不兼容的同一服务器。任何不兼容的步骤、注释、字段或变更应标有通知图标。

     在流程窗格的右上角，单击**“通知”**以查看每个不兼容功能的详细信息。单击**“在流程中查看”**，导航到不兼容的功能以采取措施。

   - **版本 2019.2.3 及更低版本：**从顶部菜单中选择**“服务器”>“发布流程”**。如果需要再次登录服务器，请确保选择与流程不兼容的相同服务器。将打开一个警告对话框，其中列出与服务器版本不兼容的功能。记下这些功能，以便可以从

流程中识别并删除它们。然后单击**“取消”**关闭该对话框。



4. 从顶部菜单中，单击**“文件”>“另存为”**以保存流程的副本。使用以下部分中的选项从流程中移除不兼容的功能。

从流程中移除不兼容的功能

您可以使用各种方法来查找和移除流程中的功能。本部分介绍了一些选项来帮助您解决不兼容性错误。

不兼容的数据源

如果数据源不兼容，例如，添加了 Tableau Prep Conductor 中尚不支持的新连接器，则您需要连接到支持的数据源。

若要更改数据连接，请参见**替换您的数据源** 在本页 120。

不兼容的功能

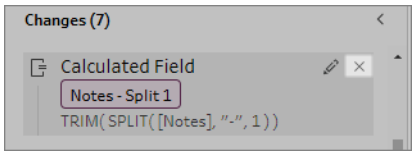
若要移除不兼容的功能，您需要查找使用了这些功能的步骤，并将其移除。您可以按照**确定不兼容的功能** 在本页 424中的说明来查找不兼容的功能。

1. 如果该功能是一种步骤类型，请在“流程”窗格中单击使用了该功能的步骤。右键单击或按住 Ctrl 并单击 (MacOS) 步骤，然后选择**“移除”**。

2. 如果该功能是清理操作，请在“流程”窗格中单击使用了该功能的步骤。您可以将指针悬停在**“流程”**窗格或者**“配置”**或**“结果”**窗格中的批注上，以查看更改的列表。

   注意：在 Tableau Prep Builder 版本 2019.1.3 及更高版本中，您可以将指针悬停在表示所查找的更改的图标上，或悬停在“流程”窗格或“配置”卡中的步骤上，然后从更改列表中选择批注。更改将在**“更改”**窗格、**“配置”**或**“结果”**窗格以及数据网格中突出显示。

3. 打开**“更改”**窗格（如果需要），并选择与需要移除的功能匹配的更改。单击更改将其选中，并单击**“移除”**将其从流程中移除。



4. 重复这些步骤以替换任何其他功能。然后保存您的流程并重新发布。

使流程数据保持最新

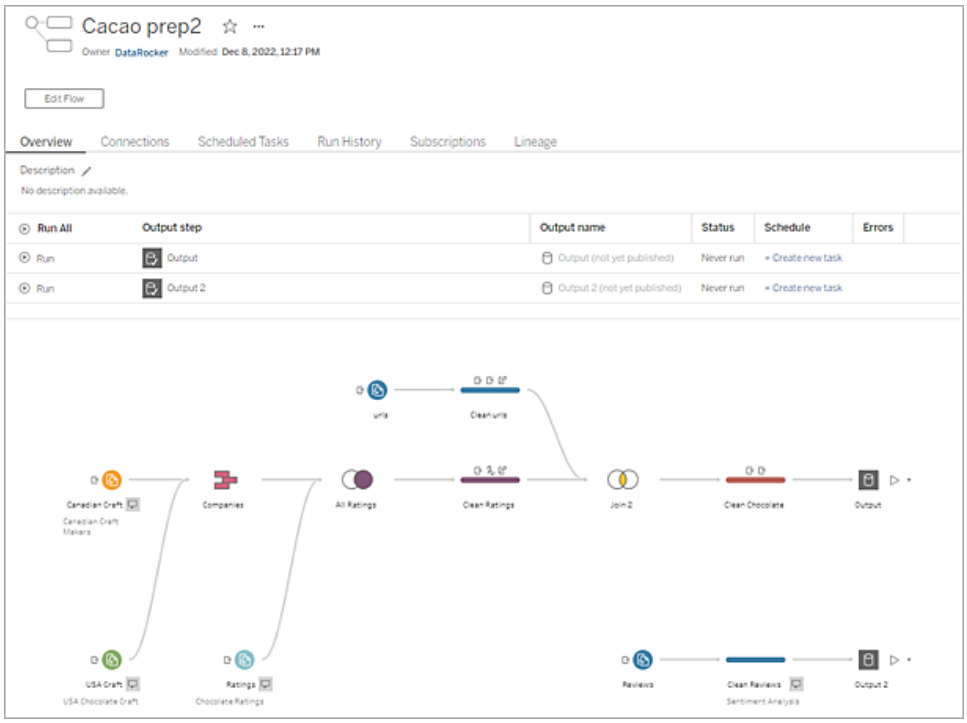
注意：本主题中的内容侧重于按计划运行流程，这需要启用包含 Tableau Prep Conductor 的 Data Management。从版本 2020.4.1 开始，不需要 Data Management 以在 Tableau Server 和 Tableau Cloud 中创建和编辑流程以及手动运行流程。

您已经构建了流程并清理了数据，但现在想要与其他人共享数据集，并希望保持使这些数据保持最新。您可以在 Tableau Prep Builder 中和 Web 上手动运行流程，并将数据提取发布到 Tableau Server，但现在有更好的方法。

使用 Tableau Server（从版本 2019.1 开始）和 Tableau Cloud 中提供的 Tableau Prep Conductor（Data Management 的一部分）。如果您将此选项添加到 Tableau Server 或 Tableau Cloud 安装，则可以使用 Tableau Prep Conductor 按计划运行流程以使流程数据保持最新。

有关如何配置 Tableau Prep Conductor 的信息，请参见Tableau Server 和 Tableau Cloud 帮助中的 Tableau Prep Conductor 内容。

从版本 2021.3 开始，您可以使用新的“链接任务”选项按计划逐个运行多达 20 个流程。有关使用链接任务运行流程的详细信息，请参见 Tableau Server 或 Tableau Cloud 帮助中的**“计划链接任务”**。



注意：如果已安装 Tableau Catalog，您可能还会看到有关流程输入数据的数据质量警告，并在新的**"世系"**选项卡中查看流程中字段的上游和下游影响。有关 Tableau Catalog 的详细信息，请参见 Tableau Server 帮助中的关于 Tableau Catalog。

使用 Tableau Prep Conductor，您可以执行以下操作：

- 将服务器或站点配置为使用 Tableau Prep Conductor
    - 为单独的站点启用或禁用 Tableau Prep Conductor
    - 为按需或使用计划运行的流程设置流程失败电子邮件通知
    - 配置流程超时设置
- 将流程从 Tableau Prep Builder 发布到 Tableau Server 或 Tableau Cloud。从版本 2020.4.1 开始，不需要 Data Management 将流程发布到 Web。
    - 上载数据文件或直接连接到文件（仅限 Tableau Prep Builder）或数据库。如果连接到数据库，您可以嵌入数据库凭据或要求用户提示。

      注意：如果通过直接连接方式连接到数据文件或将流程输出发布到文件共享，则文件需要位于 Tableau Server 可以访问的位置。此选项不适用

于在 Web 上创建的流程。有关详细信息，请参见 Tableau Server 帮助中的步骤 4：安全列表输入和输出位置。

  - 发布流程时从项目层次结构中选择
  - 输入标记和说明以帮助其他人查找您的流程
- 管理流程
  - 设置权限
  - 将流程移到其他项目
  - 更改流程所有者
  - 添加或编辑标记
  - 查看版本历史记录并从列表中选择以将流程还原到早期版本
  - 将流程标记为收藏夹并将其添加到收藏夹列表
  - 编辑输入连接和更新凭据
  - 查看通过流程创建的数据源并链接回创建它的流程
- 创建计划以运行您的流程或按需运行您的计划
  - 添加计划任务以运行流程并选择要更新的流程输出
  - 添加计划链接任务以逐个运行多个流程
- 在没有计划的情况下按需运行流程
- 监视流程
  - 设置电子邮件通知
  - 查看错误
  - 监视并重新启动已挂起的流程
  - 查看运行历史记录
  - 使用管理视图

运行流程

重要信息：从版本 2020.4.1 开始，不再需要 Data Management 在 Web 上手动运行流程。仅当您打算按计划运行流程时才需要（启用 Tableau Prep Conductor）。

若要生成流程输出，您需要运行流程。当您运行流程时，您的所有数据(不仅仅是您可能正在使用的数据样本)都将通过您的流程步骤运行。您的所有清理操作都将应用于您的完整数据集，从而形成一个整洁、干净的数据集，您现在可以使用它来分析您的数据。

注意：从版本 2021.4.1 开始，当您运行包含参数的流程时，系统会提示您输入参数值。您必须输入所需的参数值。您还可以输入任何可选参数值或接受参数的当前(默认)值。有关在流程中使用参数的详细信息，请参见**运行包含参数的流程** 在本页 210。

流程运行选项

通过命令行、使用 Tableau Server REST API 流程方法或使用计划手动运行您的流程。

- **手动**：在 Tableau Prep Builder 中和在 Web 上随时手动运行您的流程。不需要 Data Management。Web 上的流程必须先发布，然后才能运行。有关详细信息，请参见 Tableau Server 或 Tableau Cloud 帮助中的**“发布流程”**。
- **命令行界面**：如果您没有 Data Management，您可以使用命令行界面一次运行一个流程。有关详细信息，请参见**通过命令行刷新流程输出文件** 在本页 397。
- **REST API**：使用 Tableau Server 中的**“流程”**和**“流程任务”** REST API 方法来运行流程。需要 Data Management。有关详细信息，请参见 Tableau REST API 帮助中的流程方法。
- **使用计划**：在 Tableau Server 和 Tableau Cloud 中，您可以计划单个流程运行，或使用链接任务一个接一个地运行或运行多个流程。您的服务器必须包含启用了 Tableau Prep Conductor 的 Data Management。

  有关详细信息，请参见 Tableau Server 或 Tableau Cloud 帮助中的**“Tableau Prep Conductor”**。有关计划流程以自动运行的信息，请参见 Tableau Server 帮助中的计划流程任务。

手动运行流程

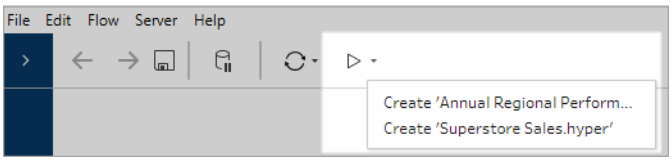
当您手动运行流程时，您可以一次运行一个流程。您可以运行整个流程或针对选定的输出运行流程。

如果在 Web 制作中运行流程(版本 2020.4 及更高版本)，则必须将流程发布到服务器才能运行它，并且在第一个流程完成之前，您不能运行另一个流程，即使从单独的选项卡运行也是如此。有关详细信息，请参见**将流程发布到 Tableau Server 或 Tableau Cloud** 在本页 433。

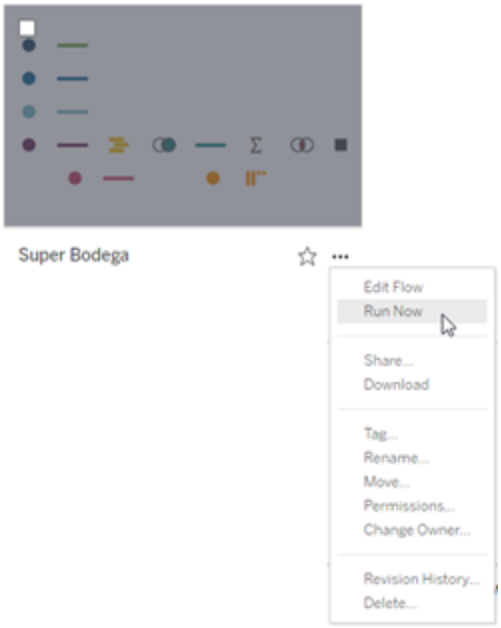
在 Tableau Cloud 中，您可以一天执行的流程运行次数也受站点管理员限制。有关详细信息，请参见 Tableau Cloud 帮助中的 Tableau Cloud 站点容量。

1. 在 Tableau Prep Builder 中或在您的服务器上，打开您的流程。
2. 执行以下操作之一：

   - 从顶部菜单中，单击“运行”▷ 运行整个流程，或单击下拉箭头在列表中选择流程输出。



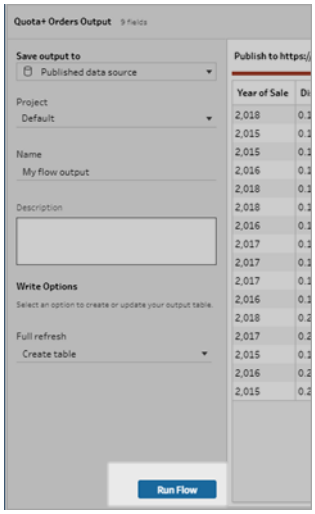
   - 在服务器上，从**“浏览”**页面中右键单击或按住 Cmd 单击 (MacOS)··· **“更多操作”**，然后从菜单中选择**“立即运行”**。这将运行您的整个流程。



   - 单击流程中的“输出”步骤，然后在“输出”窗格中单击**“运行流程”**。

     如果流程未在 Web 上打开，您需要单击**“编辑流程”**以在编辑模式下打开流程，然后单击**“发布”**以发布流程，或接受发布流程的提示，然后单击**“运行流**

程”。



将流程发布到 Tableau Server 或 Tableau Cloud

重要信息：从版本 2020.4.1 开始，不再需要 Data Management 将流程发布到 Tableau Server 或 Tableau Cloud，或在 Web 上手动运行流程。仅当您打算按计划运行流程时才需要（启用 Tableau Prep Conductor）。

将您的流程发布到 Tableau Server 或 Tableau Cloud 以与他人共享流程，或按计划自动运行流程并使用 Tableau Prep Conductor 刷新流程输出。您还可以在服务器上手动运行单独的流程。在 Web 上创建或编辑的流程（版本 2020.4.1 及更高版本）必须先发布，然后才能运行。

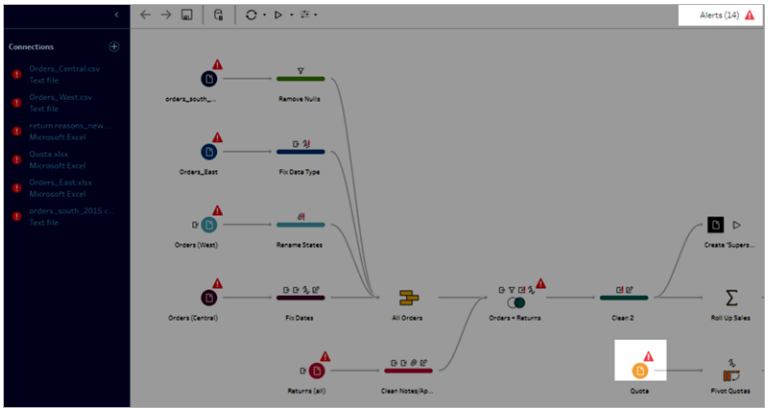
有关在 Web 上发布流程的信息，请参见 Tableau Server 或 Tableau Cloud 帮助中的**“发布流程”**。有关运行流程的信息，请参见**运行流程** 在本页430。

发布之前

为了确保可以运行流程，请检查以下各项：

1. 验证流程中是否没有错误。

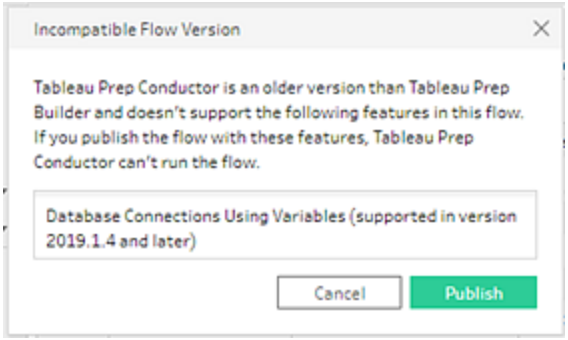
   当您尝试在 Tableau Server 或 Tableau Cloud 中运行流程时，包含错误的流程将失败。流程中的错误由红色感叹号以及画布右上角带有**错误**指示符的红点标识。



2. 验证流程未包含与您的 Tableau Server 版本不兼容的输入连接器或功能。在 Web 上创建的流程始终与创建流程的服务器版本兼容。

   您仍然可以从 Tableau Prep Builder 中发布流程，其中包括 Tableau Server 版本中尚不支持的连接器或功能，但无法安排它们运行。

   例如，Tableau Prep Builder 版本 2019.1.4 中引入了 SAP HANA 连接器，但在 Tableau Prep Conductor 的 Tableau Server 版本 2019.2 之前，此连接器不受支持。发布流程时，您会看到类似于如下的消息。



注意：若要将流程计划为在 Tableau Server 上运行，您必须使用 Tableau Server 版本 2019.1 或更高版本，并且必须启用 Tableau Prep Conductor。

为了在 Tableau Server 中运行流程，您需要采取适当的措施来使流程兼容。有关处理不兼容流程的详细信息，请参见**与 Tableau Prep 的版本兼容性** 在本页416。

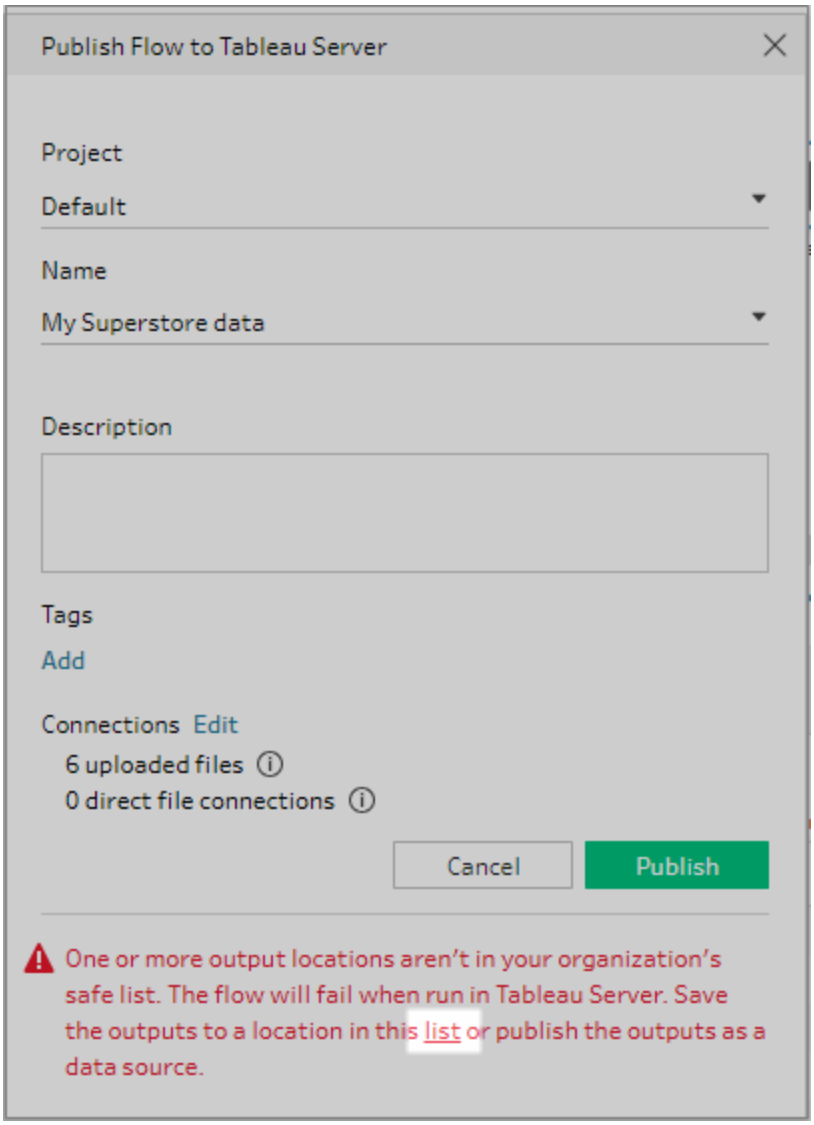
3. 包含与网络共享连接的输入或输出步骤的流程需要安全列表。Tableau Cloud 不支持此选项，并且必须在发布时将文件与流程打包在一起。

   注意：目前，在 Web 上创建的流程只能输出到已发布的数据源或数据库。

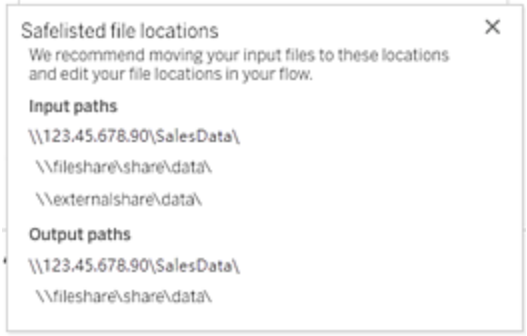
   Tableau Prep Builder

   除非服务器可以访问文件和路径，并且文件和路径包括在组织的安全列表中，否则不允许使用指向存储在网络共享（UNC 路径）中的文件的流程输入和输出步骤。如果您在未将文件位置添加到安全列表的情况下发布流程，则会发布流程，但当您尝试手动运行流程或使用 Tableau Server 中的计划时，您会收到错误消息。

   如果文件未存储在安全列出的位置，您将在发布流程时看到一条警告消息。



单击消息中的"列表"链接以查看允许位置的列表。将文件移动到列表中的某个位置，并确保流程指向这些新位置。



在 Tableau Server 中，若要配置允许的网络路径，请使用 Tableau Server 帮助中的步骤 4：安装列表输入和输出位置中描述的 tsm 命令选项。

如果不想将文件移到安全列出的位置，则需要将输入文件与流程打包在一起，并将流程输出发布到 Tableau Server 作为已发布数据源。有关设置这些选项的详细信息，请参见本主题中的**从 Tableau Prep Builder 中发布流程** 下一页。

4. (**仅限 Tableau Prep Builder**) 如果流程输出步骤设置为**“以数据源形式发布”**，则所有流输出步骤必须指向发布流程的同一服务器或站点。它们可以指向该服务器或站点上的不同项目，但只能选择一个服务器或站点。

   若要为输出步骤设置发布位置，请执行以下操作：

   a. 在“流程”窗格中，选择输出步骤。

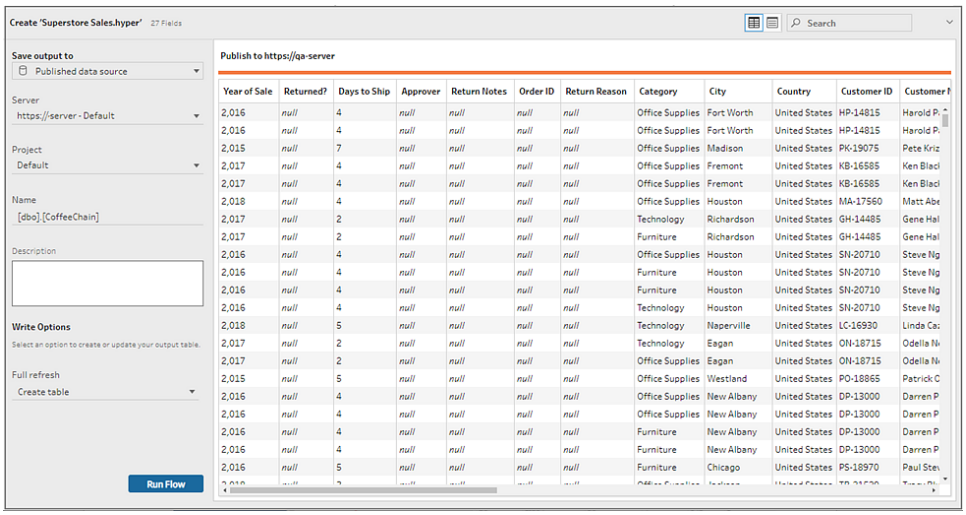
   b. 在发布窗格中，选择**“以数据源形式发布”**。

   c. 选择要在其中发布流程的服务器或站点以及项目。根据需要登录到服务器或站点。

   d. 输入每个输出的名称和说明。

      输出文件名应足够独特，以便运行流程的人员可以轻松确定要刷新的输出文件。在 Tableau Server 或 Tableau Cloud 中，文件名显示在**“概述”**和**“连接”**页面上。

   e. 保存您的流程。

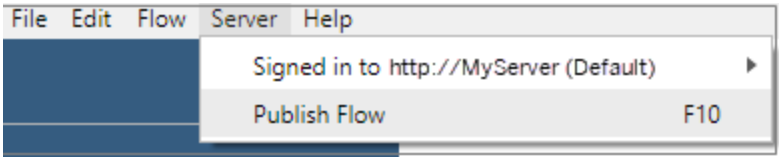


有关如何为发布配置输出步骤的详细信息，请参见**创建数据提取文件和已发布数据源** 在本页364。

从 Tableau Prep Builder 中发布流程

注意：发布流程时，系统会自动将您分配为默认流程所有者。如果流程连接到已发布数据源，则服务器将使用流程所有者连接到已发布数据源。只有站点或服务器管理员才能更改流程所有者，并且只能更改为自身。

1. 在 Tableau Prep Builder 中打开流程。
2. 从顶部菜单中选择“**服务器**”>“**发布流程**”。



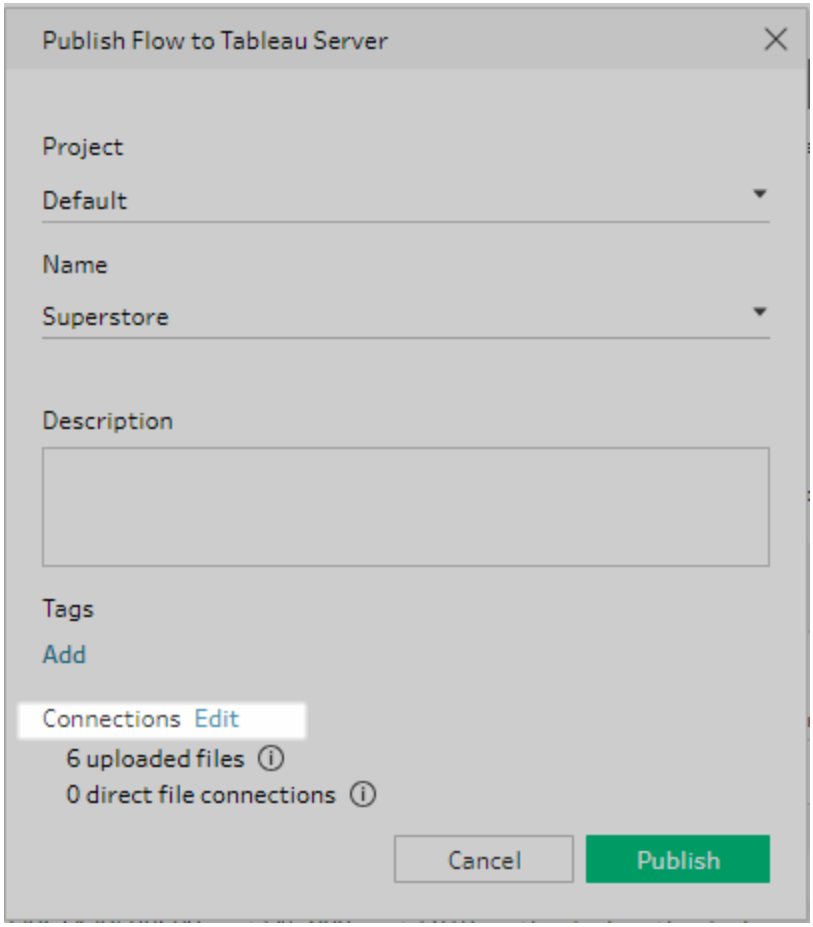
3. 针对您的平台填写字段。然后单击**“发布”**。Tableau Server 或 Tableau Cloud 会在默认浏览器中的流程“**概述**”页上自动打开。

Tableau Server

1. 在**“发布到 Tableau Server”**对话框中，填写以下字段：
   - **项目：**单击下拉选项可从项目分层结构中选择项目。此项目应与向其中发布输出文件的项目相同。

- **名称：**输入流程名称。此名称显示在服务器的“流程”页面上。如果要覆盖现有流程，请单击下拉选项从列表中选择一个名称。
- **说明（可选）：**输入流程的说明。
- **标记（可选）：**单击**“添加”**键入一个或多个标记来标识流程，以便用户可以轻松找到它。发布之后，还可以在 Tableau Server 的“流程”页面中添加标记。

2. 在“连接”部分单击**“编辑”**，以编辑连接设置或更改身份验证。

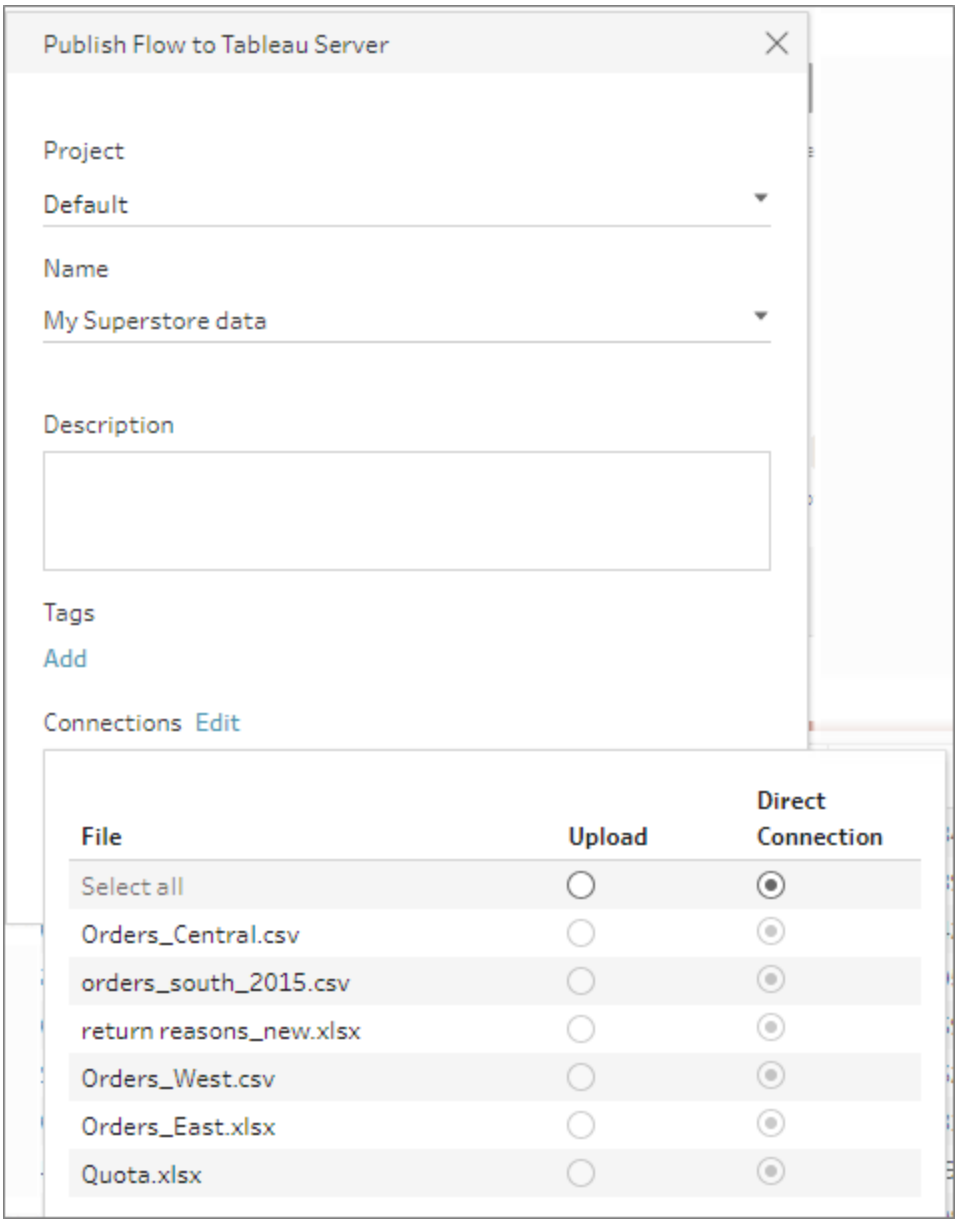


文件

默认情况下，文件输入连接与流程打包在一起。在 Tableau Server 中运行流程时，不会刷新打包的文件。所有文件必须具有相同设置，即**“上载”**或**“直接连接”**。

直接连接

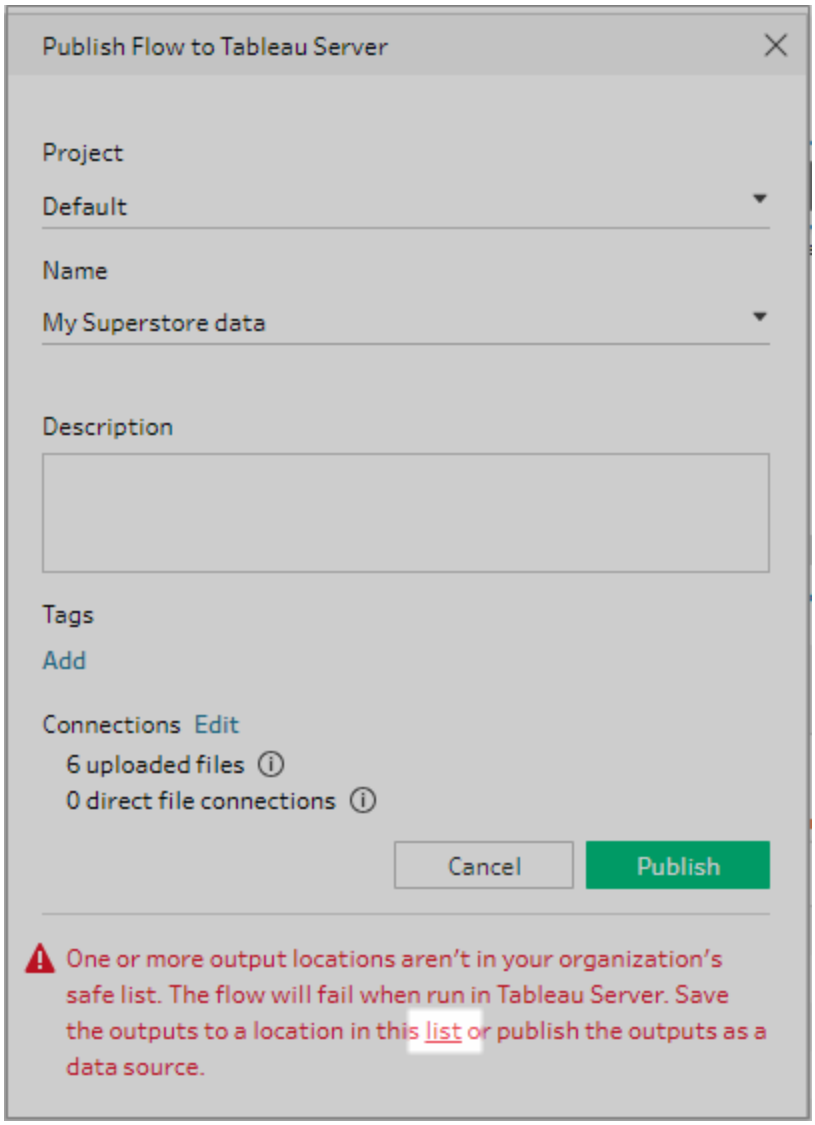
若要在刷新输出文件时检索最新数据，如果 Tableau Server 可连接到文件位置，并且该位置包括在组织的安全列表中，请选择**“直接连接”**。



存储在网络共享中的文件

如果输入或输出步骤指向存储在网络共享(UNC 路径)中的文件，并且位置未包括在组织的安全列表中，您将看到一条警告消息。单击消息中的链接以查看安全列出的位置的列表，移动文件并将输入和输出步骤指向新的文件位置。有关详细信息，请参见发布之前 在本页433中的“步骤 3”。

有关如何将位置添加到组织安全列表中的信息，请参见 Tableau Server 帮助中的步骤 4：列表输入和输出位置。



输入文件路径中的参数

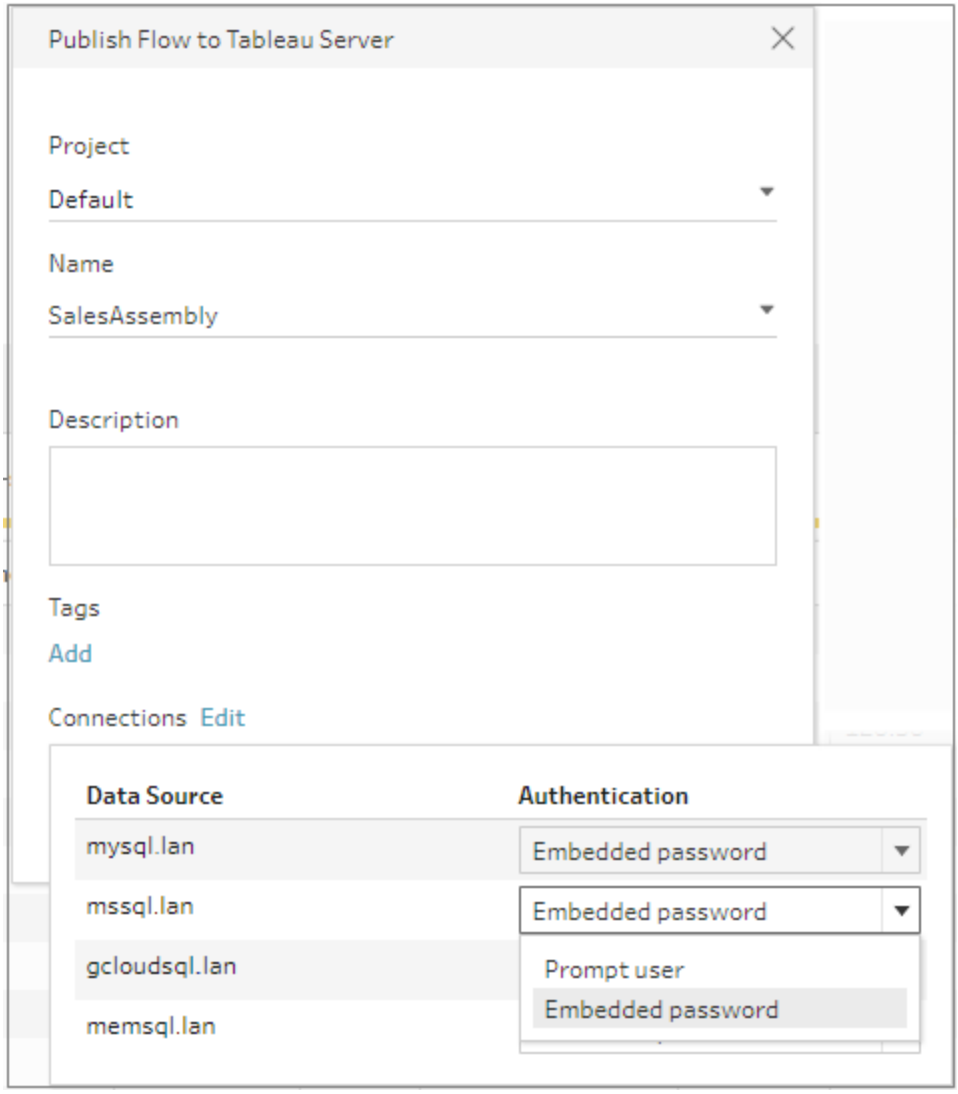
从版本 2022.1.1 开始，您可以在 Web 上计划和运行在输入文件路径中包含参数的流程。这需要直接文件连接。

如果您的文件与流程一起打包，或者您使用的是早期版本的 Tableau Prep，则文件路径中包含的所有参数都将更改为当前(默认)值，并且文件路径将变为静态。有关在流程中使用参数的详细信息，请参见**将参数应用于输入步骤** 在本页200。

数据库

如果您的流程连接到一个或多个数据库，请选择以下身份验证类型之一以用于连接到流程输入数据源。

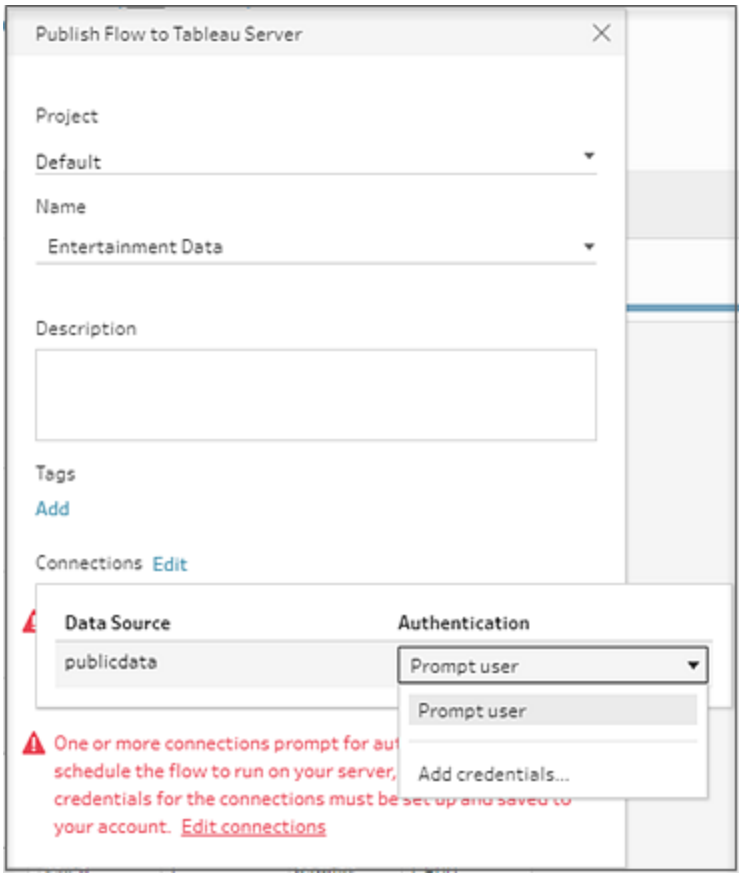
- **服务器运行身份帐户：**服务器的**“用户运行身份”**帐户将对所有用户进行身份验证。
- **提示用户：**在运行流程之前，您必须在 Tableau Server 中编辑连接并输入数据库凭据。
- **嵌入式密码：**用于连接到数据的凭据将随连接一起保存，并在流程按计划运行时使用。如果您打开流程进行编辑，您将需要重新输入您的凭据。



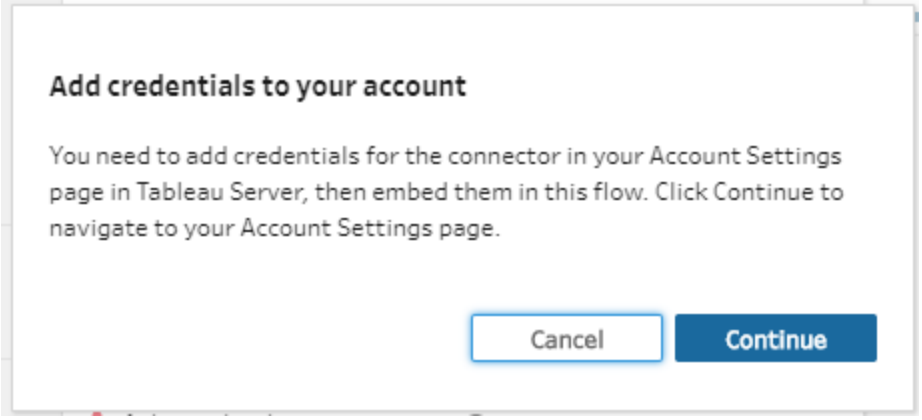
添加凭据(版本 2020.1.1 及更高版本)

如果您连接到云连接器,则可以直接从**"发布流程"**对话框添加您的凭据,以将它们嵌入流程中。

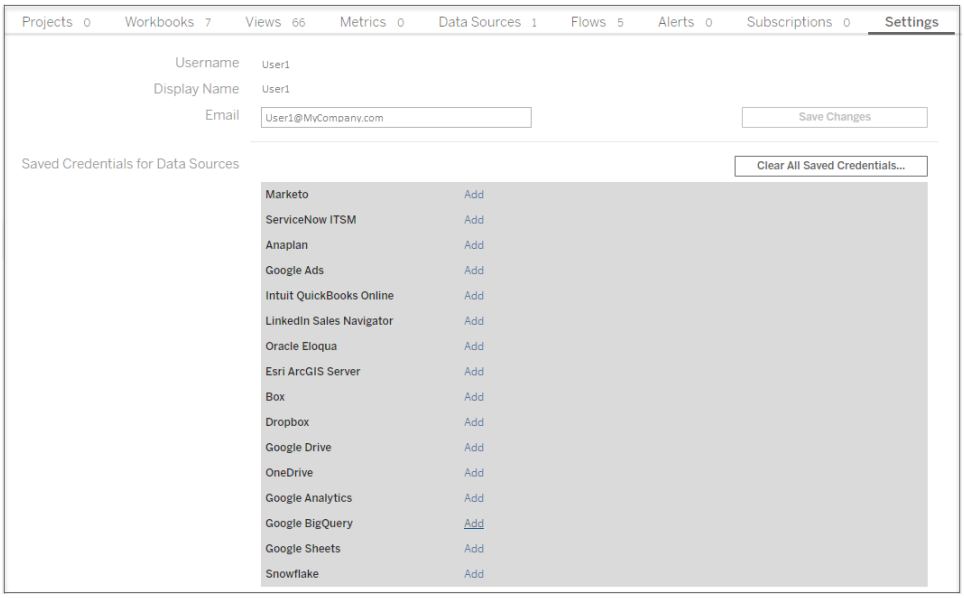
1. 单击"连接"部分中的**"编辑"**,或单击警告消息中的**"编辑凭据"**。然后从**"身份验证"**下拉菜单中单击**"添加凭据"**。



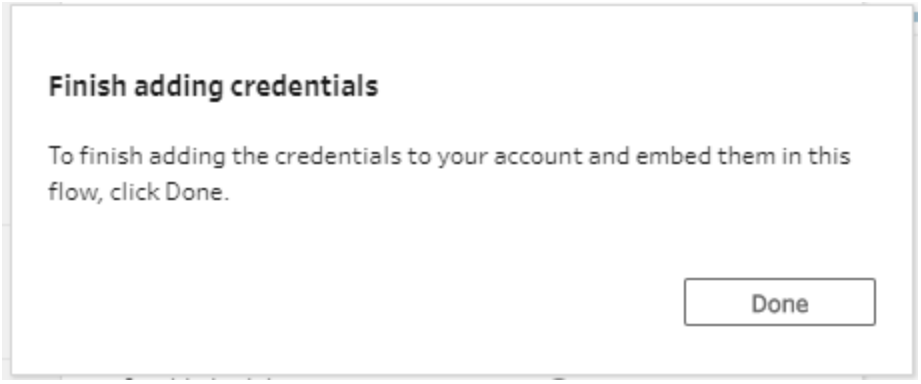
2. 在确认对话框中,单击**"继续"**。Tableau Prep Builder 会自动打开您登录的服务器的"帐户设置"页面。



3. 添加您的凭据，然后导航回 Tableau Prep Builder。



4. 在**“完成添加凭据”**话框中，单击**“完成”**。



5. 在**“连接”**部分单击**“编辑”**，并验证凭据是否已添加并嵌入到流程中。

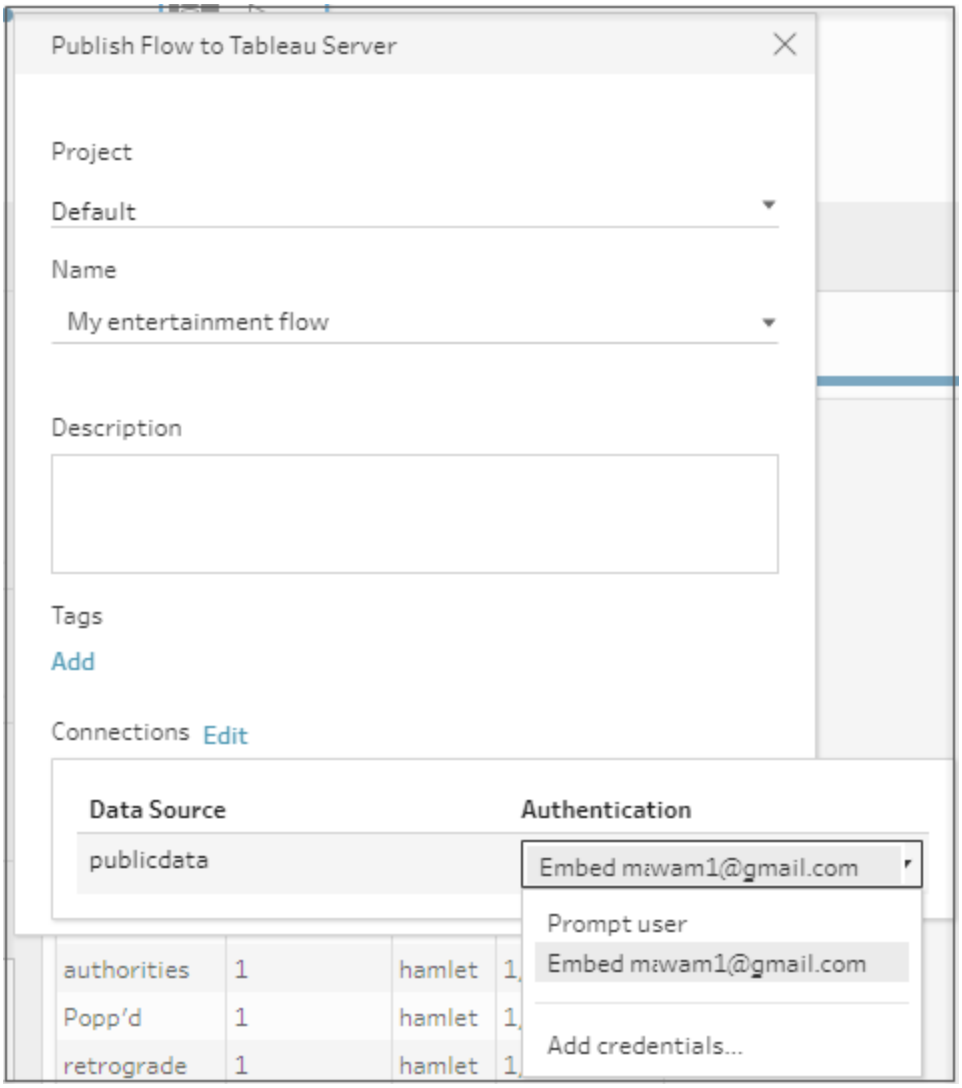
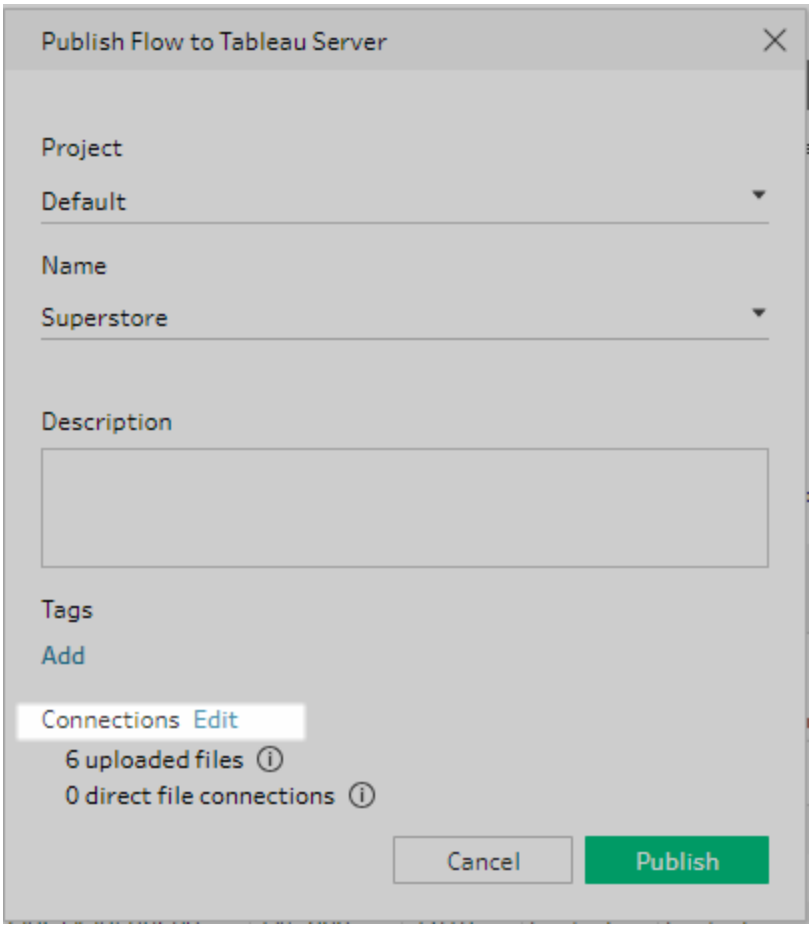


Tableau Cloud

1. 在**“发布到 Tableau Cloud”**对话框中，填写以下字段：
   - **项目：**单击下拉选项可从项目分层结构中选择项目。此项目应与向其中发布输出文件的项目相同。
   - **名称：**输入流程名称。此名称显示在服务器的“流程”页面上。如果要覆盖现有流程，请单击下拉选项从列表中选择一个名称。
   - **说明（可选）：**输入流程的说明。
   - **标记（可选）：**单击**“添加”**键入一个或多个标记来标识流程，以便用户可以轻松找到它。发布之后，还可以在 Tableau Server 的“流程”页面中添加标记。
2. 在“连接”部分单击**“编辑”**，以编辑连接设置或更改身份验证。



文件

Tableau Cloud 不支持输入步骤数据的直接文件连接，您必须将文件与流程打包在一起。在 Tableau Cloud 中运行流程时，不会刷新打包的文件。

> **注意**：Tableau Cloud 目前不支持计划和运行在输入文件路径中包含参数的流程，因为这需要直接文件连接。发布流程时，文件路径中包含的所有参数都将更改为当前(默认)值，并且文件路径变为静态。
>
> 作为替代方案，您可以在 Tableau Prep Builder 的文件路径中或使用命令行运行带有参数的流程。有关在流程中使用参数的详细信息，请参见**将参数应用于输入步骤** 在本页200。

数据库

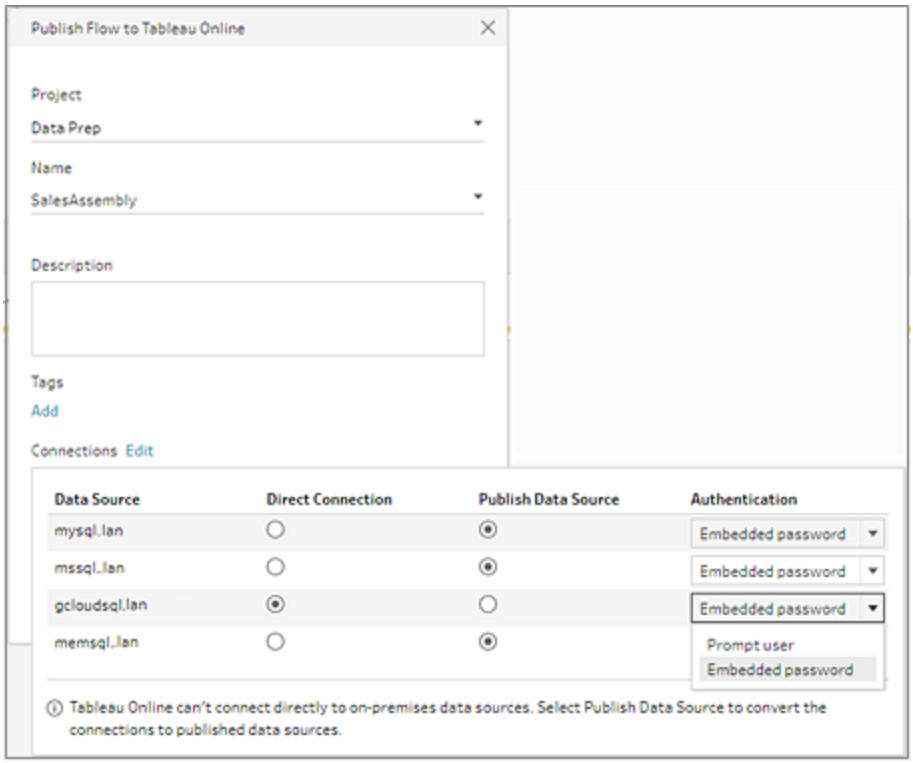
为了在将流程发布到 Tableau Cloud 时使数据保持最新，您只能直接连接到云托管数据源。在连接到本地数据源时，您必须将数据源转换为已发布数据源，并且，如果针对数据源配置了 Tableau Bridge，Tableau Cloud 可以使用 Tableau Bridge 客户端连接到数据。

有关 Tableau Cloud 支持的直接连接的详细信息，请参见允许直接连接到云平台上托管的数据。

有关使用 Tableau Bridge 的详细信息，请参见允许发布者维护本地数据的实时连接。

如果您的流程连接到支持直接连接的基于云的数据源，请选择下列身份验证类型之一以用于连接到流程输入数据源。

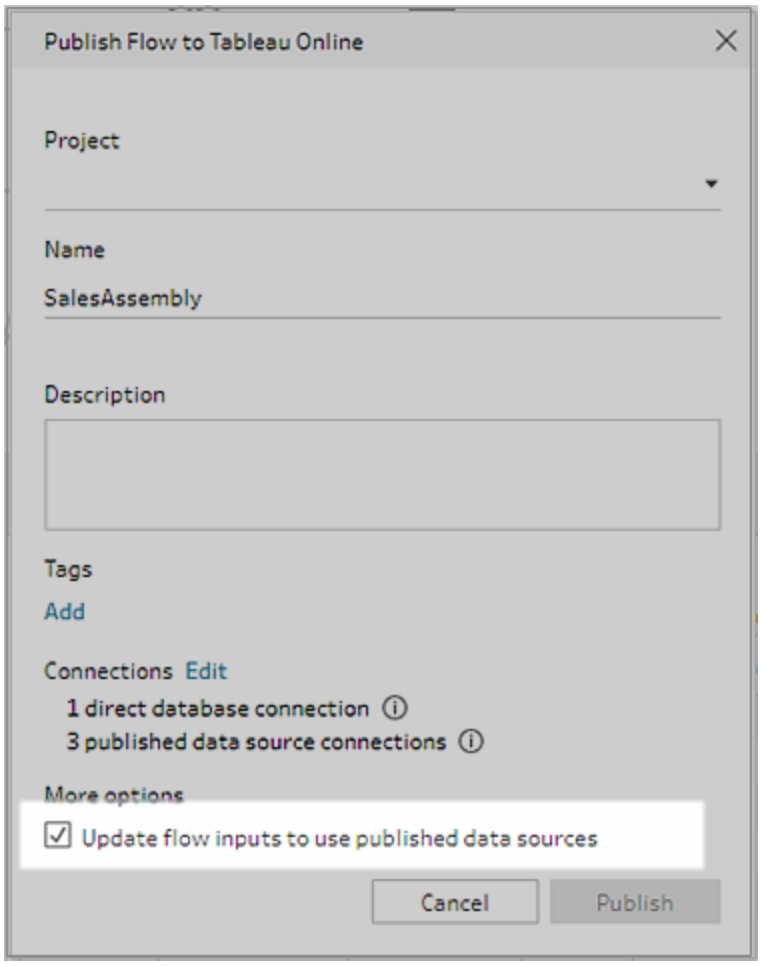
- **提示用户**：在运行流程之前，您必须在 Tableau Cloud 中编辑连接并输入数据库凭据。
- **嵌入式密码**：用于连接到数据的凭据将随连接一起保存，并在流程按计划运行时使用。如果您打开流程进行编辑，您将需要重新输入您的凭据。



- 为本地数据源选择**"发布数据源"**单选按钮。Tableau Cloud 无法直接连接到这些数据源来刷新数据。在将流程发布到 Tableau Cloud 时，选择此选项可将数据源输入连接转换为已发布数据源。

  如果为数据源配置了 Tableau Bridge，并且 Tableau Cloud 支持该数据源，则可以在运行流程时刷新数据。有关详细信息，请参见允许直接连接到云平台上托管的数据。
- 若要将 Tableau Prep Builder 中流程的本地数据源连接替换为已发布数据源，请在发布流程之前在**"更多选项"**部分中选择**"更新流程输入以使用发布的数据源"**。

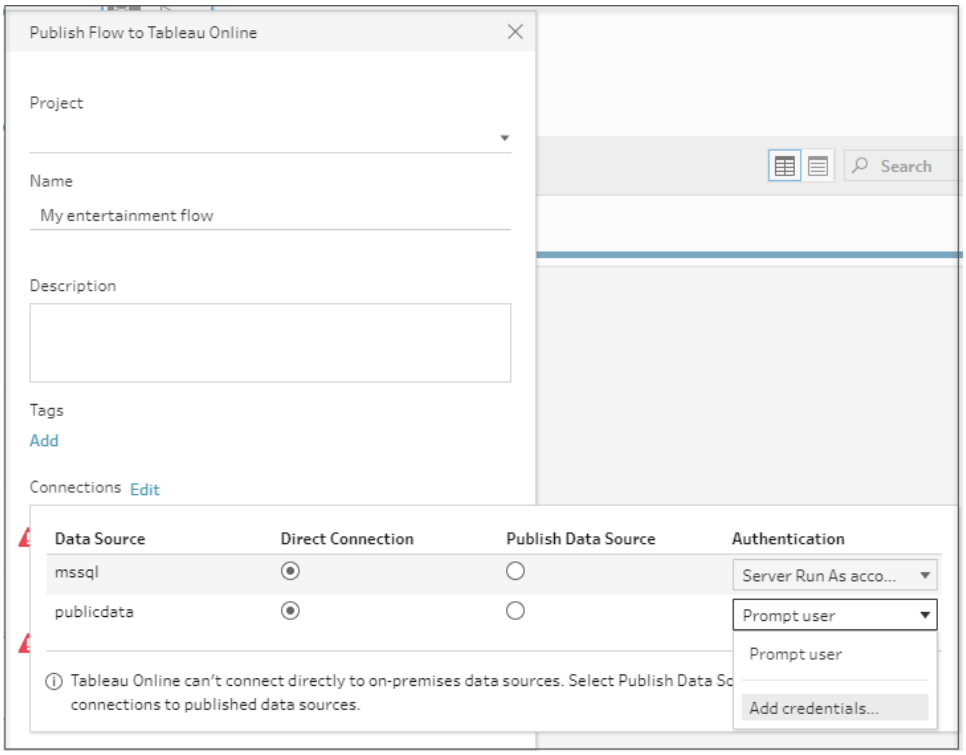
  如果未选中该复选框，则 Tableau Prep Builder 中的流程会一直连接到本地数据源，并且 Tableau Prep Builder 中的流程会变得与流程的已发布版本不同步。若要继续使用流程，您将需要从 Tableau Cloud 中下载流程以对其进行编辑，然后重新发布流程。



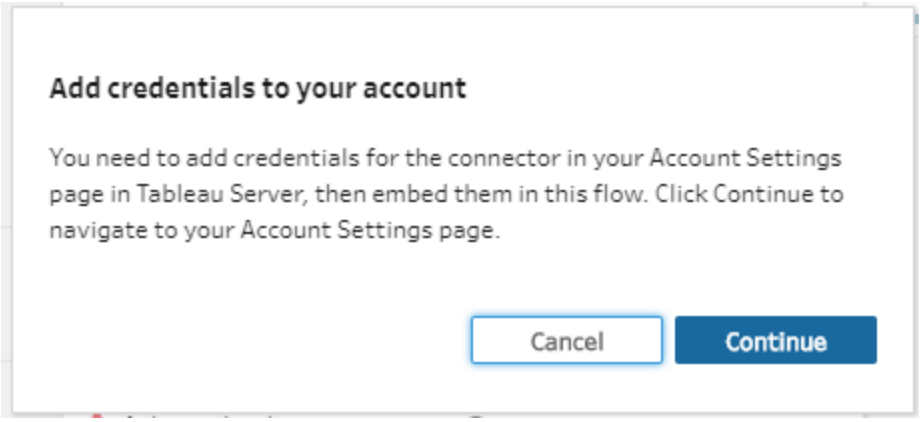
添加凭据(版本 2020.1.1 及更高版本)

如果您连接到云连接器，则可以直接从**"发布流程"**对话框添加您的凭据，以将它们嵌入流程中。

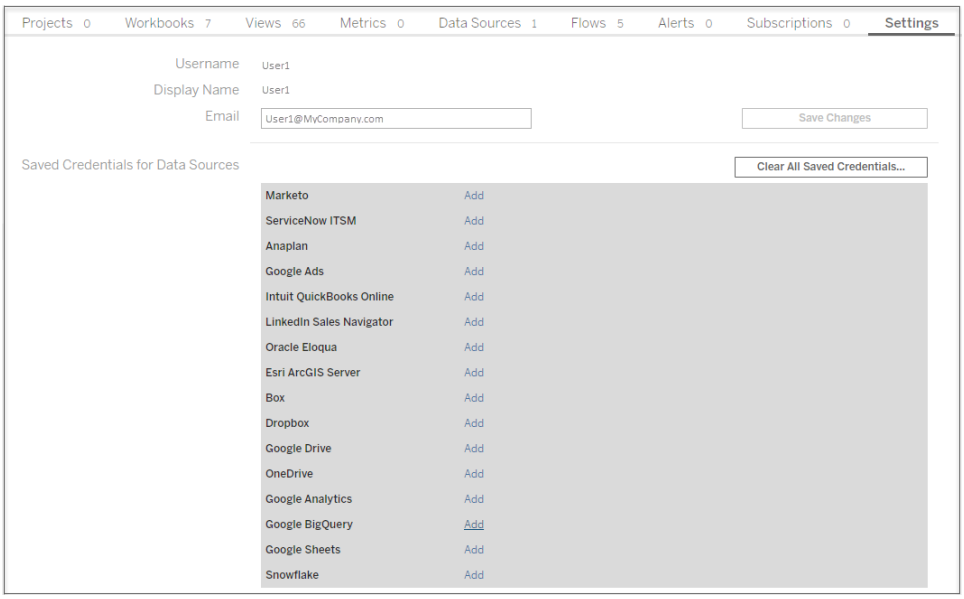
1. 单击"连接"部分中的**"编辑"**，或单击警告消息中的**"编辑凭据"**。然后从**"身份验证"**下拉菜单中单击**"添加凭据"**。



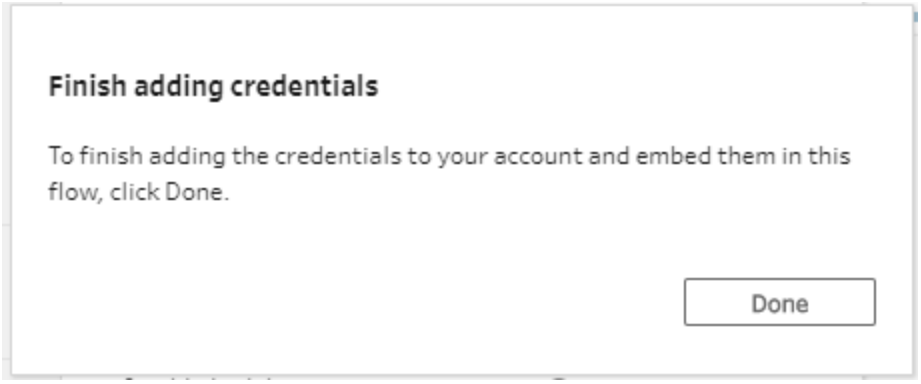
2. 在确认对话框中，单击**“继续”**。Tableau Prep Builder 会自动打开您登录的服务器的“帐户设置”页面。



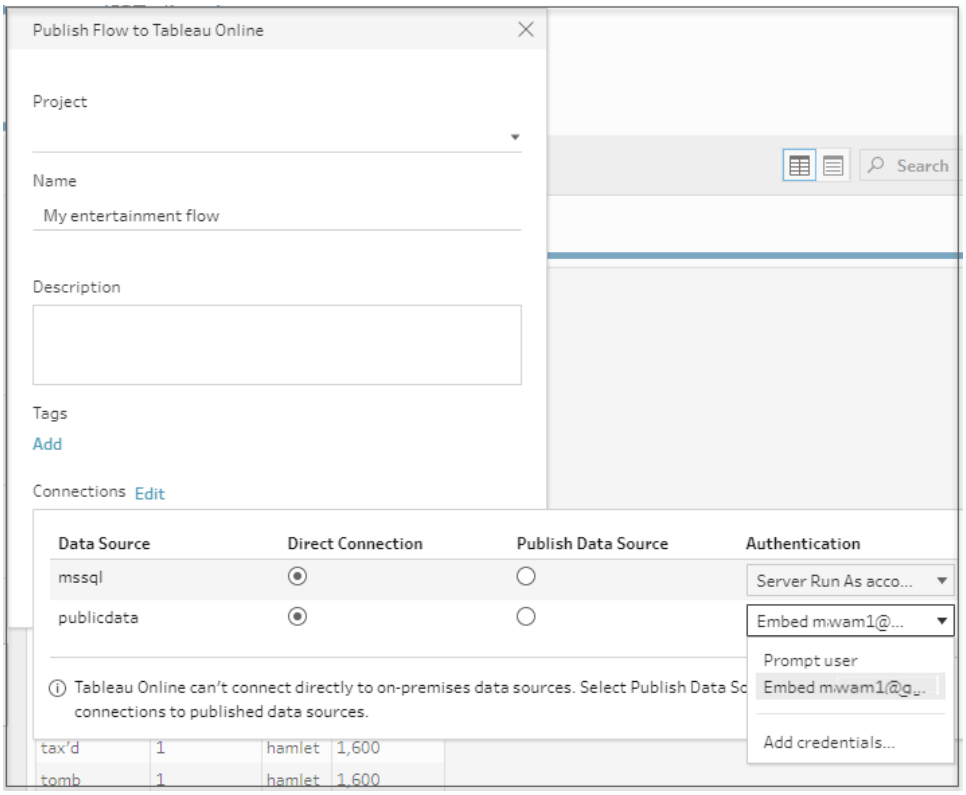
3. 添加您的凭据，然后导航回 Tableau Prep Builder。



4. 在**“完成添加凭据”**话框中，单击**“完成”**。



5. 在**“连接”**部分单击**“编辑”**，并验证凭据是否已添加并嵌入到流程中。



谁可以执行此操作

- 服务器管理员、站点管理员 Creator 和 Creator 允许完全连接和发布访问权限。
- Creator 可以执行 Web 制作任务。
- Explorer(可发布)

日常生活场景

调整数据的含义是什么？它会对可构建的可视化项以及可执行的分析产生怎样的影响？在下面的教程中，我们将探讨分析和可视化项的场景，确定对我们造成阻碍的数据限制，然后看看 Tableau Prep 如何能帮助我们调整数据来达到预期结果。

下载数据集，并使用 Tableau Prep 和 Tableau Desktop 执行这些日常生活场景。了解如何应用 Tableau Prep 中的特点和功能来准备好数据以在 Tableau Desktop 中进行分析。

为我们提供反馈。我们刚刚开始构建联机帮助的这一部分。如果有您想要在此处看到的特定场景，请不吝告知我们。使用页面顶部的反馈栏来告诉我们更多。

为了完成这些教程中的任务，您需要安装 Tableau Prep 和 Tableau Desktop，并将需要下载并将数据保存到您的计算机。

有关如何安装 Tableau Prep 和 Tableau Desktop 的信息，请参见 Tableau Desktop 和 Tableau Prep 部署指南中的通过用户界面安装 Tableau Desktop 和 Tableau Prep Builder。否则，您可以下载 Tableau Prep 和 Tableau Desktop 免费试用版。

使用 Tableau Prep 分析医院床位使用情况

医院满员会造成问题，但资源过多也同样会造成问题。务必要从床位即资源的角度来理解医院床位。但是，数据通常是从患者的角度进行存储的。我们如何才能获取相关数据，了解患者何时占用床位并确定床位使用情况？

注意：为了完成这些教程中的任务，您需要安装 Tableau Prep 并根据需要安装 Tableau Desktop：

若要安装 Tableau Prep 和 Tableau Desktop，请参见 Tableau Desktop 和 Tableau Prep 部署指南。否则，您可以下载 Tableau Prep 和 Tableau Desktop 免费试用版。

您还需要下载三个数据文件。建议将这些文件保存在 我的 Tableau Prep 存储库 > Datasources 文件夹中。
- Beds.xlsx
- Hours.xlsx
- Patient Beds.xlsx

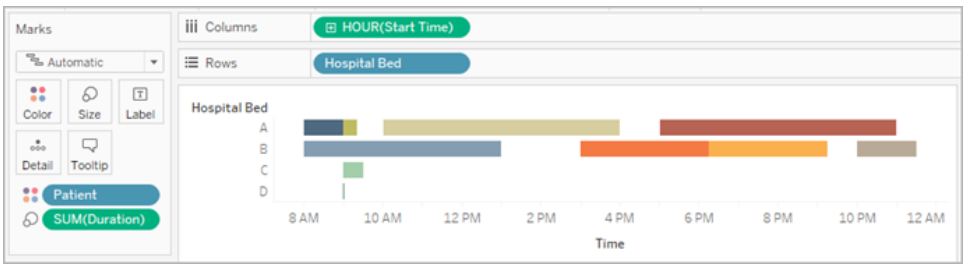
数据

对于我们的四个床位 A、B、C 和 D，我们将跟踪哪个患者占用了床位，以及他们占用床位的开始和结束时间。数据看起如下所示：

	A	B	C	D
1	**Hospital Bed**	**Patient**	**Start Time**	**End Time**
2	A	Person 1	1/1/2018 8:34	1/1/2018 9:34
3	A	Person 5	1/1/2018 9:55	1/1/2018 10:15
4	A	Person 9	1/1/2018 10:34	1/1/2018 16:34
5	A	Person 8	1/1/2018 17:00	1/1/2018 23:00
6	B	Person 2	1/1/2018 8:45	1/1/2018 13:45
7	B	Person 6	1/1/2018 15:13	1/1/2018 18:27
8	B	Person 7	1/1/2018 18:41	1/1/2018 21:56
9	B	Person 10	1/1/2018 22:13	1/1/2018 23:43
10	C	Person 3	1/1/2018 9:05	1/1/2018 9:35
11	D	Person 4	1/1/2018 9:30	

初步分析

如果我们将此数据引入 Tableau Desktop，我们可以创建一个甘特图来显示患者何时占用床位。



这是一种有用的视觉元素。我们可以看到，床位 A 和 B 的使用间隔时间很短，但床位 C 使用得非常少。床位 D 的患者没有结束时间，但我们将利用一些计算来解决该情况。这为我们提供了一个关于床位使用情况的视觉概览。

但是，如果我们想要计算床位空置的小时数，该怎么办？或者，如果我们要比较实施新政策前后的开放床位时间，该怎么办？通过采用当前结构方式的数据无法轻松地达到目的。

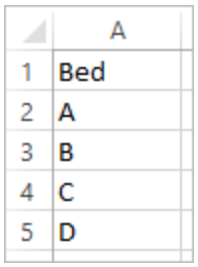
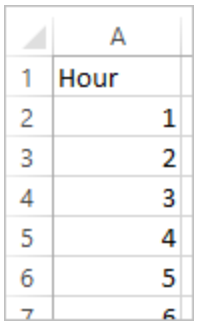
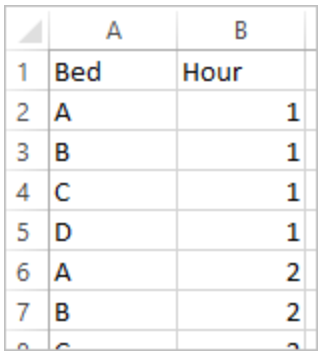
所需的数据结构

通过创建一些非常基本的数据集并在 Tableau Prep 中将其合并，我们可以将此数据集修改为允许我们执行更深入分析并创建更有用的可视化项的形式。

在进入 Tableau Prep 之前，让我们回顾一下，想一想我们需要创建什么才能回答以下问题：“每个床位空置了多少小时？”

我们需要能够查看每小时的每个床位，并了解是否有患者占用床位。目前，数据仅仅是患者何时占用床位；我们还未向 Tableau 提供有关*空置*小时数的信息。

为了创建包含所有床位和所有小时的完整矩阵，我们将创建两个新数据集。一个数据集是床位的列表（A、B、C、D），另一个数据集是小时的列表（1、2、3、...、23、24）。通过执行交叉联接（将一个数据集中的每一行与另一个数据集中的每一行联接），我们将最终获得床位和小时的每种可能组合。

Beds.xlsx 数据集看起来如下所示：	**Hours.xlsx** 数据集看起来如下所示：	而交叉联接的结果看起来如下所示：
		

	A
1	Bed
2	A
3	B
4	C
5	D

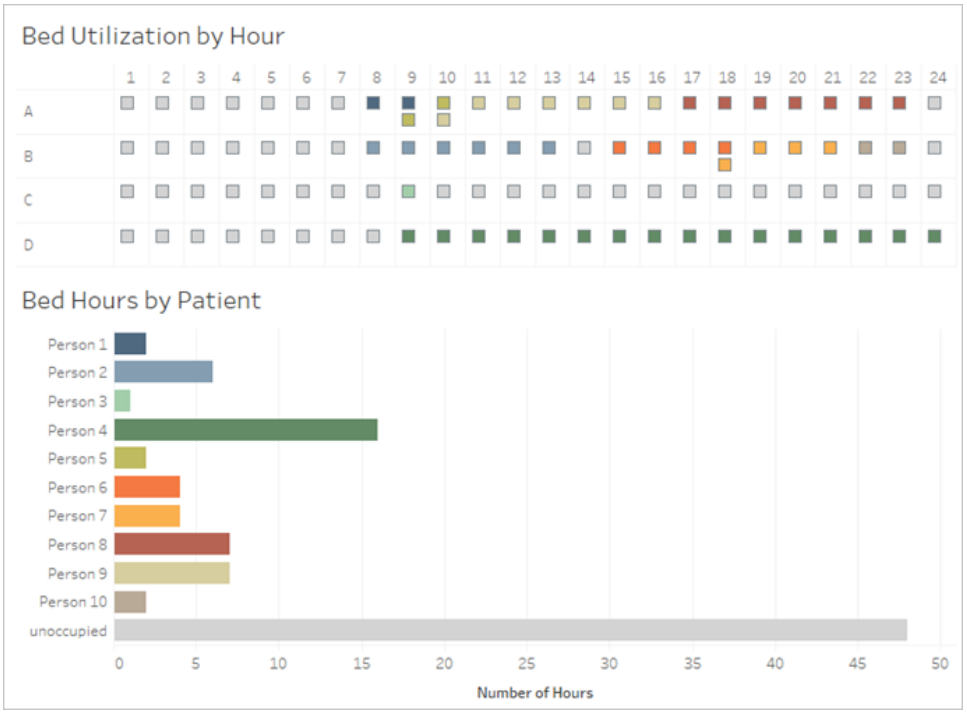
	A
1	Hour
2	1
3	2
4	3
5	4
6	5

	A	B
1	Bed	Hour
2	A	1
3	B	1
4	C	1
5	D	1
6	A	2
7	B	2

接着，我们将引入“**Patient Beds**”（患者床位）信息，以是否有特定患者的形式标记每个床位-小时组合。我们最终会得到一个数据集，该数据集中每个床位-小时组合占一行，并且，如果患者占用了床位，则包含患者的编号及开始和结束时间。Null 值指明床位未被占用。

	A	B	C	D	E
1	Bed	Hour	Patient	Start Time	End Time
29	D	7			
30	A	8	Person 1	1/1/2018 8:34	1/1/2018 9:34
31	B	8	Person 2	1/1/2018 8:45	1/1/2018 13:45
32	C	8			
33	D	8			
34	A	9	Person 5	1/1/2018 9:55	1/1/2018 10:15
35	A	9	Person 1	1/1/2018 8:34	1/1/2018 9:34
36	B	9	Person 2	1/1/2018 8:45	1/1/2018 13:45
37	C	9	Person 3	1/1/2018 9:05	1/1/2018 9:35
38	D	9	Person 4	1/1/2018 9:30	
39	A	10	Person 9	1/1/2018 10:34	1/1/2018 16:34
40	A	10	Person 5	1/1/2018 9:55	1/1/2018 10:15
41	B	10	Person 2	1/1/2018 8:45	1/1/2018 13:45
42	C	10			
43	D	10	Person 4	1/1/2018 9:30	
44	A	11	Person 9	1/1/2018 10:34	1/1/2018 16:34

利用此结构中的数据，我们可以执行像这样的分析，让我们能够像调查患者床位一样轻松地调查未占用的床位。



重构数据

那么如何使用 Tableau Prep 达到该目的？我们将分两个部分构建流程，首先构建“Bed Hours”(床位小时)矩阵，然后将其与“Patient Beds”(患者床位)数据合并。确保下载全部三个 Excel 文件(**Beds.xlsx**、**Hours.xlsx** 和 **Patient Beds.xlsx)** 以便继续操作。

床位小时矩阵

首先，我们将连接到 **Beds.xlsx** 文件。

1. 打开 Tableau Prep。
2. 从开始屏幕中，单击**“连接到数据”**。
3. 在**“连接”**窗格上，单击**“Microsoft Excel”**。导航到 **Beds.xlsx** 的保存位置，并单击**“打开”**。
4. **“Beds”**(床位)工作表应会自动显示在**“流程”**窗格中。

> **提示：**有关连接到数据的详细信息，请参见连接到数据 在本页69。

接着，我们需要创建一个可用于与“**Hours**”(小时)数据集执行交叉联接的字段。我们将添加一个仅仅为值 **1** 的计算。

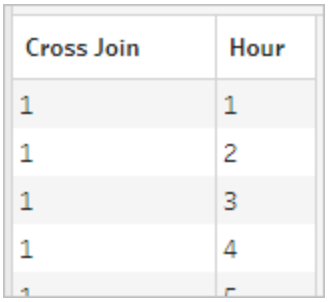
5. 在**“流程”**窗格中，选择**“Beds”**(床位)并单击建议的**“清理步骤”**。
6. 随我们刚刚添加的**“清理”**步骤一起，**“配置”**窗格将出现。单击工具栏中的**“创建计算字段”**。
7. 将字段命名为“**Cross Join**”(交叉联接)，并输入值 **1**。
8. **“数据”**网格应会更新为显示数据的当前状态。

Cross Join	Bed
1	A
1	B
1	C
1	D

现在，我们将使用“Hours”(小时)数据集重复该过程。

单击查看指引

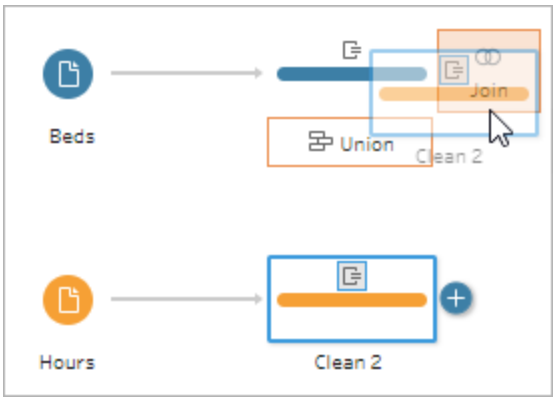
9. 在**“连接”**窗格上，单击“添加连接”按钮以添加另一个数据连接。
10. 选择**“Microsoft Excel”**，然后选择 **Hours.xlsx** 文件并单击**“打开”**。
11. 在**“流程”**窗格中，选择**“Hours”**（小时），并单击建议的**“清理步骤”**将其添加到流程。
12. 通过**“配置”**窗格中的工具栏创建一个名为**“Cross Join”**（交叉联接）的计算字段，并输入值 **1**。



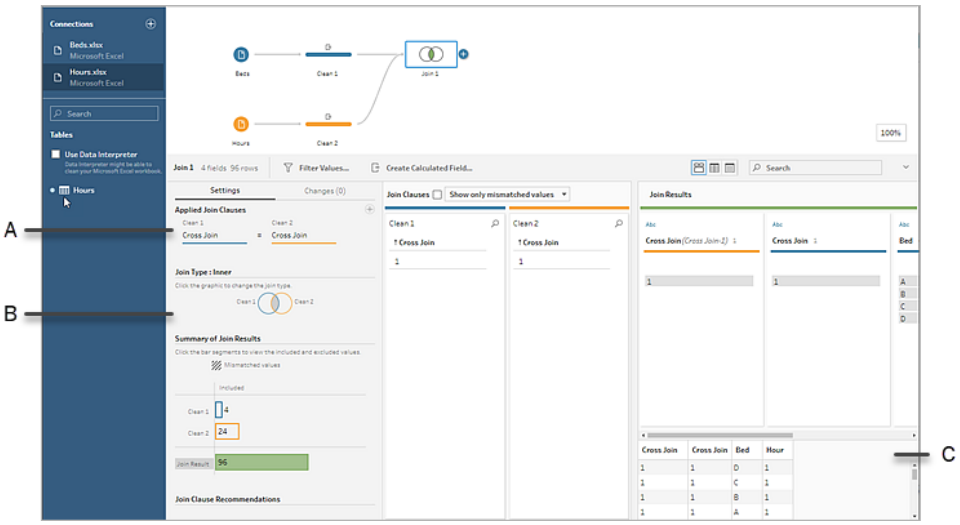
Cross Join	Hour
1	1
1	2
1	3
1	4

两个数据集现在有一个共同字段**“Cross Join”**（交叉联接），因此可以联接。

13. 通过将**“Clean 2”**（清理 2）拖到**“Clean 1”**（清理 1）并放在**“联接”**选项上来联接两个清理步骤。



14. 在下面的**“联接配置”**中，联接配置应会自动填充。
    - 由于我们将两个字段都命名为**“Cross Join”**（交叉联接），因此 Tableau Prep 自动将它们识别为共同字段，并创建适当的**“已应用联接子句”**。
    - 默认**“联接类型”**为内部联接，正是我们所需要的。
    - 此联接会将**“Beds”**（床位）中的所有行与**“Hours”**（小时）中的所有行匹配，如**“数据”**网格中所示。



A. 联接子句

B. 联接类型

C. 数据网格结果

提示：有关联接的详细信息，请参见**联接数据** 在本页340。

我们不再需要“**Cross Join**”(交叉联接)字段，因此可以将这些字段移除。

15. 在**“流程”**窗格中，选择“**Join** 1”(联接 1)，单击加号 ⊕ 图标，并选择**“添加清理步骤”**。
16. 选择字段“**Cross Join-1**”(交叉联接-1)和“**Cross Join**”(交叉联接)，并单击**“移除字段”**。
17. 双击“**Clean 3**”(清理 3)标签，并将该步骤重命名为“**Bed Hour Matrix**”(床位小时矩阵)。

我们现在有了“Bed Hour Matrix”(床位小时矩阵)数据集，其中包含所有床位和所有小时，并且我们已完成构建数据集的第一部分。

患者床位使用

第二部分是引入患者床位使用情况。首先，我们将连接到数据。

1. 在**“连接”**窗格上，单击“添加连接”按钮以添加另一个数据连接。
2. 选择**“Microsoft Excel”**，然后选择 **Patient Beds.xlsx** 文件，并单击**“打开”**。
3. 在**“流程”**窗格中，选择**“Patient Beds”**（患者床位），然后单击建议的**“清理步骤”**将其添加到流程。

由于“床位小时矩阵”文件基于*小时*，但“患者床位”基于*实际时间*，因此我们需要从“患者床位”开始和结束时间中提取小时。此外，对于结束时间，我们想要确保，如果患者在一天结束时（午夜，24 时）仍然占用床位，则我们指明床位被占用，即使数据集中没有结束时间也是如此。我们将在此新步骤中添加计算字段。

4. 在工具栏中，单击**“创建计算字段”**。
5. 将字段命名为“**Start Hour**”（开始小时）。对于计算，输入 `DATEPART('hour', [Start Time])`。

   这将获取开始时间的小时并将其提取出来。因此，“1/1/18 9:35 AM”将变为“9”。
6. 创建另一个名为“**End Hour**”（结束小时）的计算字段。对于计算，输入 `IFNULL (DATEPART('hour',[End Time]), 24)`。

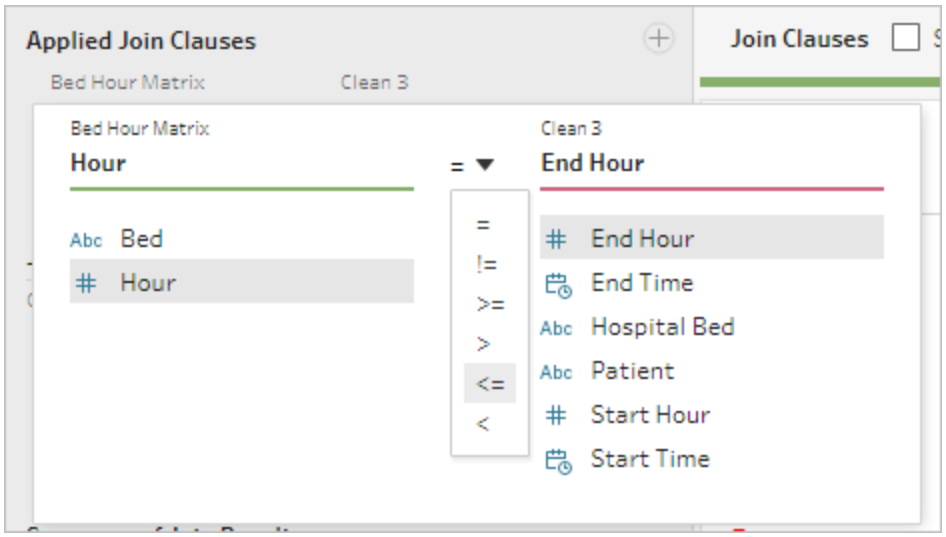
   `DATEPART` 部分获取结束时间的小时。`IFNULL` 部分将为任何缺少的结束时间分配结束时间 24 时（午夜）。

现在我们已准备好将患者床位使用情况联接到“**Bed Hour Matrix**”（床位小时矩阵）。与我们之前进行的操作相比，此联接操作更为复杂一些。内部联接将只会返回两个数据集中都存在的值。由于我们想要确保保留所有床位-小时槽位，而不管患者是否占用床位，因此我们需要执行左联接。此将生成许多 null 值，但这是适当的。

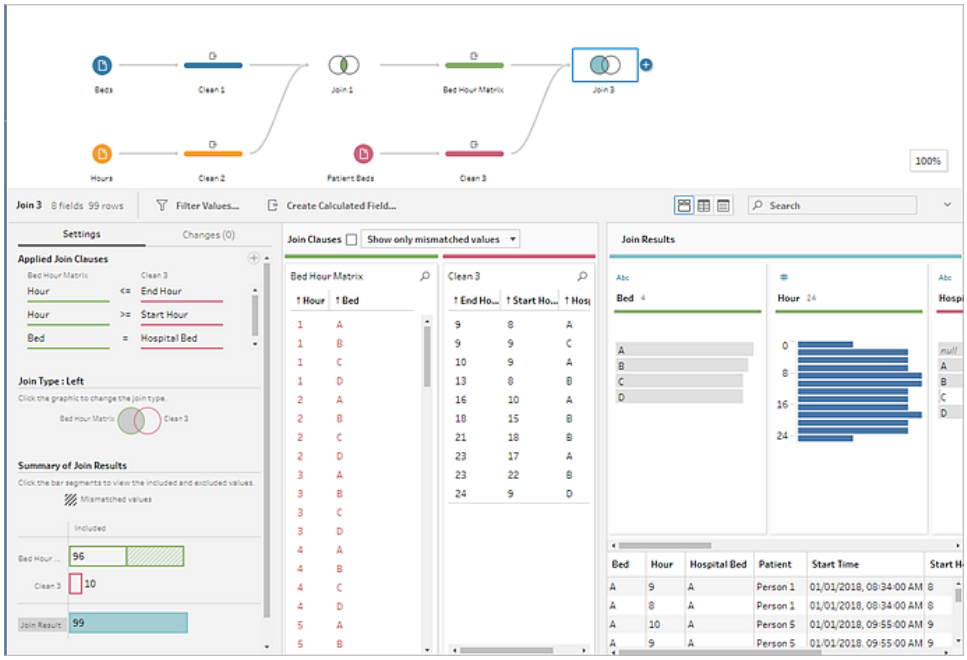
我们还需要匹配床位-小时槽位被一名（或多名）患者占用的时间。因此，除了匹配患者占用的床位外，我们还需要考虑时间。“Bed Hour Matrix”（床位小时矩阵）数据集只有“**Hour**”（小时）字段，而“**Patient Beds**”（患者床位）数据集有“**Start Hour**”（开始小时）和“**End Hour**”（结束小时）。我们使用一些基本逻辑来确定是否应将患者分配到给定床位-小时槽位：*如果其开始小时小于或等于 (<=) 床位-小时槽位，并且其结束小时大于或等于 (>=) 床位-小时槽位，则患者被视为占用床位。*

因此，需要三个联接子句来将这两个数据集相应地匹配在一起。

9. 将“**Clean 3**”（清理 3）步骤与“**Bed Hour Matrix**”（床位小时矩阵）步骤联接。
10. 在**“已应用联接子句”**区域中，默认值应为“**Hour = End Hour**”。单击联接子句，将运算符从“**=**”改为“**<=**”。



11. 单击**“已应用联接子句”**区域右上角的加号 ⊕ 按钮，以添加另一个联接子句。将其设置为**“Hour >= Start Hour”**
12. 为**“Bed = Hospital Bed”**添加第三个联接子句。
13. 在**“联接类型”**部分，单击**“Bed Hour Matrix”**（床位小时矩阵）旁边图形的无阴影区域，将联接类型更改为**“左”**联接。



注意：如果将“**Bed Hour Matrix**”(床位小时矩阵)拖到“**Clean 3**”(清理 3)，而不是相反，则通过使用右联接(而不是左联接)可以获得期望的结果。对于联接的方向，拖动步骤的顺序很重要。联接子句也将按相反顺序 — 一定要保持比较小时的正确逻辑。

数据现在已联接，但我们应从联接中清理一些构件，并确保字段保持整齐。我们不再需要“**Start Hour**”(开始小时)和“**End Hour**”(结束小时)。“**Hospital Bed**”(医院床位)和“**Bed**”(床位)也是多余的。最后，“**Patient**”(患者)字段中的值 null 实际上表示床位未被占用。

14. 在**“流程”**窗格中，添加一个清理步骤以便能对联接的数据进行整理。
15. 按住 Ctrl 单击(在 Mac 上按住 Command 单击)以多选方式选择字段“**End Hour**”(结束小时)、“**Start Hour**”(开始小时)和“**Hospital Bed**”(医院床位)，然后在工具栏中单击**“移除字段”**。
16. 在“**Patient**”(患者)字段配置卡上，双击“**null**”值并键入“**Unoccupied**”(未占用)。

我们现在有了每个床位-小时占一行的数据结构；如果在该小时期间有患者占用床位，则我们也有了患者信息。余下的工作就是添加输出步骤和生成数据集本身。

17. 在**“流程”**窗格中，选择“**Clean 4**”(清理 4)，单击加号 ⊕ 图标，并选择**“添加输出”**。
18. 在**“输出”**窗格中，将**“输出类型”**更改为 .csv，然后单击**“浏览”**。
19. 为名称输入“**Bed Hour Patient Matrix**”(床位小时患者矩阵)，并在单击**“接受”**保存之前选择所需的位置。
20. 单击窗格底部的**“运行流程”**按钮生成输出。在状态对话框中单击**“完成”**关闭对话框。

提示：有关输出和运行流程的详细信息，请参见保存和共享工作 在本页361。

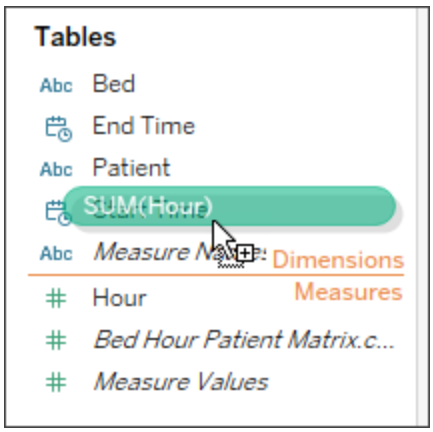
最终流程应如下所示：

Tableau Desktop 中的分析

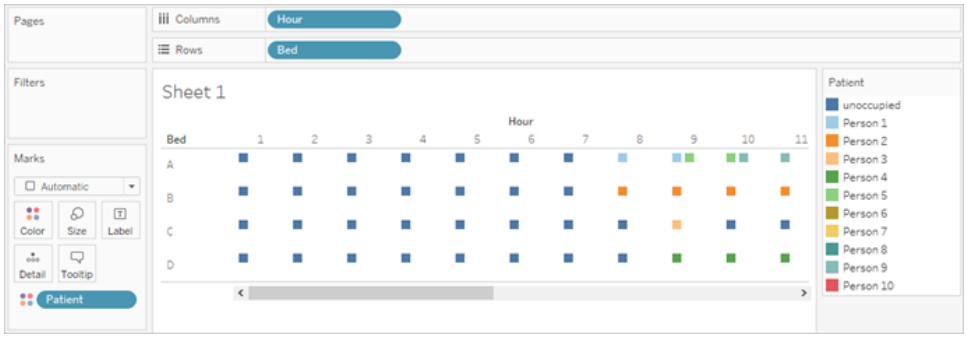
若要在继续执行本教程之前安装 Tableau Desktop，您可以下载免费试用版。

现在我们有了采用所需结构的数据集，我们可以执行比用原始数据更深入的分析。

1. 打开 Tableau Desktop。在**“连接”**窗格中，选择**“文本文件”**，导航到 **Bed Hour Patient Matrix.csv** 文件，并单击**“打开”**。
2. 在**“数据源”**选项卡上，默认情况下数据应出现在画布中。单击“**Sheet 1**”（工作表 1）。
3. 在**“数据”**窗格中，将“**Hour**”（小时）拖到分隔“度量”和“维度”的线上方，将其设为离散维度。

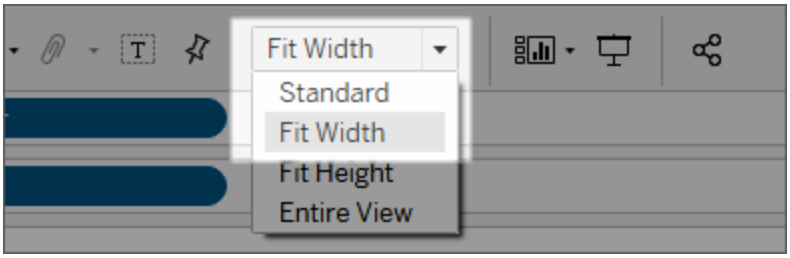


4. 将“**Bed**”（床位）拖到**“行”**功能区，并将“**Hour**”（小时）拖到**“列”**功能区。
5. 将“**Patient**”（患者）拖到**“颜色”**功能区。

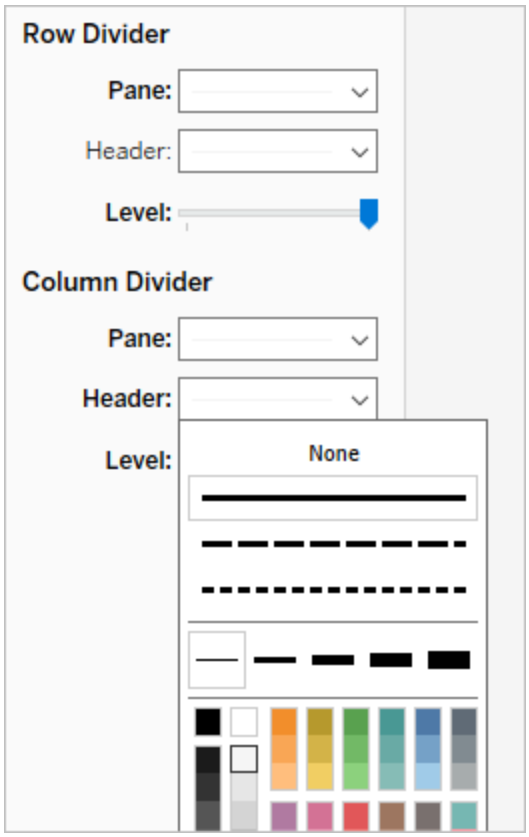


格式设置为可选，但可以帮助使图示更易于理解。

6. 单击**“颜色”**功能区，并选择**“编辑颜色”**。
7. 在左侧的区域中，选择“**Unoccupied**”(未占用)。从右侧的下拉列表中，选择**“西雅图灰”**调色板。
8. 选择第四个最浅的灰色，并单击**“确定”**。
9. 再次单击**“颜色”**功能区，然后单击**“边框”**下拉列表。选择最右侧的第二个灰色选项。
10. 在工具栏中，通过“大小”下拉列表将**“标准”**更改为**“适合宽度”**。

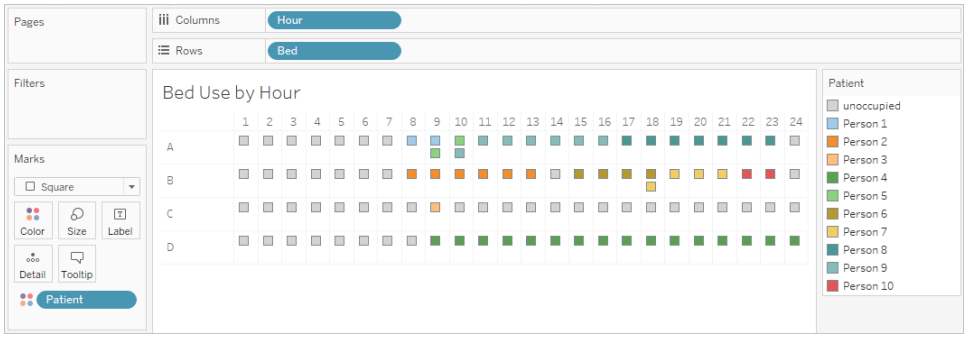


11. 单击**“设置格式”**菜单，然后单击**“边框”**。
12. 对于**“行分隔符”**，单击“区”下拉列表并选择一种非常浅的灰色。
13. 将**“级别”**滑块调整到第二个刻度标记。
14. 为**“列分隔符”**重复这些步骤。将**“区”**颜色设置为浅灰色，并将**“级别”**调整到第二个刻度标记。



15. 双击底部的工作表标签，并将其重命名为“**Bed Use by Hour**”(按小时列出的床位使用情况)。

此视图让我们能快速查看给定床位何时处于占用或空置状态。

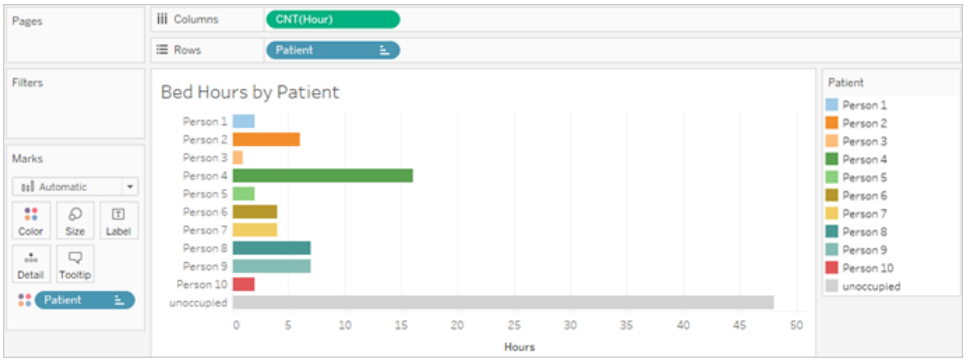


但我们可以更进一步，并计算每个床位未被占用的小时数。

16. 单击底部的新工作表标签 打开一个空白工作表。

17. 将“**Patient**”(患者)拖到**“行”**。

18. 将“**Hour**”(小时)拖到**“列”**。右键单击“Hour”(小时)胶囊打开菜单。选择**“度量”>“计数”**。
19. 将“**Patient**”(患者)字段的另一个副本从**“数据”**窗格拖到**“颜色”**功能区。
20. 右键单击轴,并选择**“编辑轴”**。将标题更改为“**Hours**”(时数)并关闭对话框。
21. 将工作表标签重命名为“**Bed Hours by Patient.”(患者占用床位时数)**。



此视图使我们能够确定我们有多少未占用床位时数,而这是使用原始数据集做不到的。您可以创建哪些其他图标或仪表板?既然您的数据结构正确,不妨试一下。

总结和资源

为了使用 Tableau Prep 构建此数据结构,我们需要执行以下操作:

1. 针对我们想要分析的各个方面(本例中为**床位**和**小时**)构建数据集。
2. 交叉联接这些数据集,创建一个包含床位和小时每种可能组合的“**Bed Hour Matrix**”(床位小时矩阵)数据集。
3. 将“**Bed Hour Matrix**”(床位小时矩阵)与“**Patient Bed**”(患者床位)数据联接,同时确保联接保留所有床位-槽位小时,并且联接子句恰当地将患者床位数据与床位-小时槽位匹配。

我们使用以下计算来创建可以联接的字段。第二个和第三个计算从原始日期时间字段中提取小时信息。

- **Cross Join**(交叉联接) = `1`
  - 此计算只是为每一行分配值 1
- **Start Hour**(开始小时) = `DATEPART('hour',[Start Time])`
  - 这将获取开始时间的小时并将其提取出来。因此,“1/1/18 9:35 AM”将变为“9”。

- **End Hour**(结束小时) = `IFNULL(DATEPART('hour',[End Time]), 24)`
  - 我们可以使用 `DATEPART('hour',[End Time])`，就像我们为“**Start Time**”(开始时间)所做的一样。这将获取结束时间的小时并将其提取出来。因此，“1/1/18 4:34 AM”将变为“4”。
  - 但我们想要指明仍被占用(没有结束时间)的患者床位在使用中，而不是空置。为此，我们将使用 `IFNULL` 函数为任何缺少的结束时间分配结束时间 24 时(午夜)。如果第一个参数 `DATEPART('hour',[End Time])` 为 null，则计算将改为返回“24”。

注意：想要检查您的工作？下载 Tableau Prep 打包流程文件 (Hospital Beds.tflx) 和 Tableau Desktop 打包工作簿文件 (Hospital Beds.twbx)。

资源：需要更多培训？参加现场培训课程。想知道我们涵盖的功能？请查看 Tableau Prep 联机帮助中的其他主题。在寻找其他资源？ Master Tableau Prep with this list of learning resources(利用此学习资源清单掌握 Tableau Prep) 博文正适合于您。

使用 Tableau Prep 查找第二个日期

分析中的一项常见需求是确定*第二个*事件(例如客户进行第二次购买并从而成为常客，或者司机第二次交通违章时)的日期。查找第一个事件的日期很容易，它只是最小日期。查找第二个日期则较为复杂。

在这个由两部分组成的教程中，我们将调整交通违章数据并回答以下问题：

1. 每个司机第一次和第二次违章相隔多长时间(以天数为单位)？
2. 比较第一次和第二次违章的罚款金额。这些金额是否相关？
3. 总体而言，哪个司机支付的罚款最多？谁支付的最少？
4. 有多少司机有过多种类型的违章？
5. 从未上过交通法规学习班的司机的平均罚款金额是多少？

在第一个阶段中，我们将使用 Tableau Prep Builder 针对分析重构数据。在第二个阶段 **Tableau Desktop 中包含第二个日期的分析** 在本页479中，我们将转而讲述 Tableau Desktop 中的分析。

本教程的目标是在一个真实场景的上下文中呈现各种概念并演练各个选项 — 而不是规定式地确定哪一个是最好的。最后，您应该会对数据结构对计算和分析的影响有更深的认识，并更加熟悉 Tableau Prep 的各项功能以及 Tableau Desktop 中的计算。

注意：为了完成本教程中的任务，您需要 Tableau Prep Builder（安装或通过浏览器访问），并下载相关数据。对于第二部分，您还需要安装 Tableau Desktop。

数据集为 Traffic Violations.xlsx。建议将其保存在 我的 Tableau Prep 存储库 > Datasources 文件夹中。

若要在继续学习本教程之前安装 Tableau Prep Builder 和 Tableau Desktop，请参见 Tableau Desktop 和 Tableau Prep 部署指南。否则，您可以下载 Tableau Prep 和 Tableau Desktop 免费试用版。

数据

对于此示例，我们将探讨交通违章数据。每条违章为一行。将记录司机、日期、违章类型、司机是否需要参加交通法规学习班以及罚款金额。

	A	B	C	D	E
1	Driver ID	Infraction Date	Infraction Type	Traffic School	Fine Amount
2	JO-151451402	1/8/2017	Speeding	Yes	115
3	CM-127151402	3/1/2017	Running a red light	No	55
4	AP-109151404	3/2/2017	Non-moving violation	No	95
5	SH-199751404	3/4/2017	Speeding	Yes	130
6	BT-114401404	3/20/2017	Non-moving violation	No	130
7	MO-175001406	5/30/2017	Speeding	Yes	118
8	RA-1988558	6/2/2017	Speeding	Yes	144
9	BT-1168027	6/5/2017	Speeding	Yes	128
10	MO-175001406	6/18/2017	Speeding	Yes	115
11	MP-174701406	6/19/2017	Speeding	No	125
12	AA-106451404	7/5/2017	Running a red light	No	60
13	RA-199151402	7/20/2017	Speeding	Yes	146
14	SC-202601404	8/31/2017	Running a red light	No	150
15	MO-175001406	9/7/2017	Non-moving violation	No	320

所需的数据结构

数据当前的构造方式是每条*违章*为一行。有多次违章的司机将出现在多行中，并且无法轻松分辨哪条违章是他们的第一次或第二次违章。

为了调查屡屡违章的司机，我们需要一个能够分出第一次和第二次违章日期的数据集、与其中每条违章关联的信息，并且需要每一行为一个*司机*。

	A	B	C	D	E	F	G	H	I
1	Driver ID	1st Infraction Date	1st Infraction Type	1st Traffic School	1st Fine Amount	2nd Infraction Date	2nd Infraction Type	2nd Traffic School	2nd Fine Amount
2	BD-117701406	12/25/2017	Speeding	Yes	140	2/7/2018	Speeding	Yes	125
3	JO-151451402	1/8/2017	Speeding	Yes	115	11/21/2018	Reckless driving	Yes	550
4	SN-207101402	12/27/2017	Speeding	Yes	280	4/26/2018	Speeding	Yes	130
5	CJ-120101402	11/26/2017	Speeding	Yes	122	3/28/2018	Speeding	Yes	116
6	JR-156701404	12/24/2017	Speeding	No	148	7/28/2018	Speeding	Yes	310
7	AP-109151404	3/2/2017	Non-moving violation	No	95	9/24/2018	Speeding	No	105
8	PC-187451406	11/11/2017	Speeding	Yes	220	12/30/2018	Non-moving violation	No	600
9	TS-214301406	9/13/2018	Speeding	Yes	115	11/10/2018	Non-moving violation	No	95
10	NP-187001404	12/11/2018	Non-moving violation	No	80	12/20/2018	Speeding	No	120
11	DB-129701402	5/13/2018	Running a red light	No	110	11/11/2018	Speeding	Yes	80
12	AJ-107951404	10/15/2017	Speeding	Yes	130	12/31/2017	Running a red light	No	85
13	BT-114401404	3/20/2017	Non-moving violation	No	130	11/13/2018	Speeding	Yes	96
14	AF-108851406	5/9/2018	Non-moving violation	No	200	9/2/2018	Speeding	No	130
15	SC-202601404	8/31/2017	Running a red light	No	150	11/10/2018	Speeding	Yes	50
16	KL-166451406	10/4/2017	Speeding	No	115	11/13/2017	Speeding	Yes	104
17	MO-175001406	5/30/2017	Speeding	Yes	118	6/18/2017	Speeding	Yes	115
18	CM-127151402	3/1/2017	Running a red light	No	55	8/1/2018	Running a red light	No	160
19	KT-164801402	5/31/2018	Non-moving violation	No	190	11/10/2018	Speeding	No	74
20	JB-160001402	11/18/2018	Speeding	Yes	220	12/5/2018	Non-moving violation	No	195
21	LH-170201404	5/6/2018	Running a red light	No	110	9/17/2018	Speeding	Yes	230
22	BG-1103555	12/25/2017	Speeding	Yes	195	12/8/2018	Speeding	Yes	315
23	MP-174701406	6/19/2017	Speeding	No	125	10/12/2017	Running a red light	No	175

重构数据

那么如何使用 Tableau Prep 达到该目的？我们将分阶段构建流程，首先抽取第一次违章日期，然后抽取第二次违章日期，接着根据需要调整最终数据。确保已下载了 Excel 文件 (Traffic Violations.xlsx) 以便继续操作。

第一次违章日期的初始聚合

首先，我们将连接到 **Traffic Violations.xlsx** 文件。

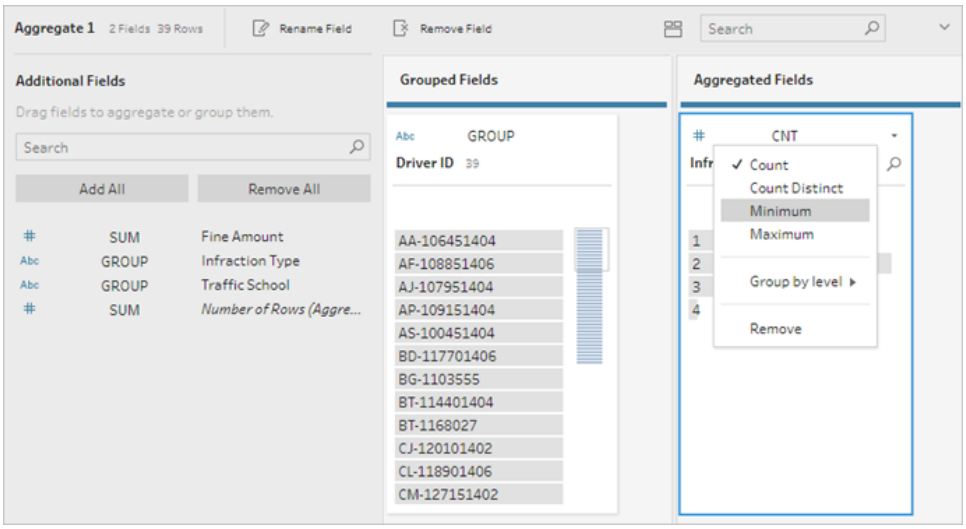
1. 打开 Tableau Prep Builder。
2. 从开始屏幕中，单击**“连接到数据”**。
3. 在**“连接”**窗格中，单击**“Microsoft Excel”**。导航到 **Traffic Violations.xlsx** 的保存位置，并单击**“打开”**。
4. “**Infractions**”(违章) 工作表应会自动显示在**“流程”**窗格中。

   有关连接到数据的详细信息，请参见**连接到数据** 在本页69。

接着，我们需要标识每个司机的第一次违章日期。我们将使用**“聚合”**步骤来执行此操作，同时创建一个包含“**Driver ID**”(司机 ID) 和“**Minimum Infraction Date**”(最早违章日期) 的微型数据集。

在 Tableau Prep 中使用“聚合”步骤时，任何应定义行组成内容的字段为**“分组字段”**。(对我们来说，该字段为“**Driver ID**”(司机 ID)。) 任何将聚合并显示在分组字段级别的字段为**“聚合字段”**。(对我们来说，该字段为“**Infraction Date**”(违章日期)。)

5. 在**“流程”**窗格中，选择“**Infractions**”（违章），单击加号 ⊕ 图标，并选择**“聚合”**。
6. 将“**Driver ID**”（司机 ID）拖到**“分组字段”**放置区域。
7. 将“**Infraction Date**”（违章日期）拖到**“聚合字段”**区域。默认聚合为 **CNT**（计数）。单击**“CNT”**，并将聚合更改为**“Minimum”**。

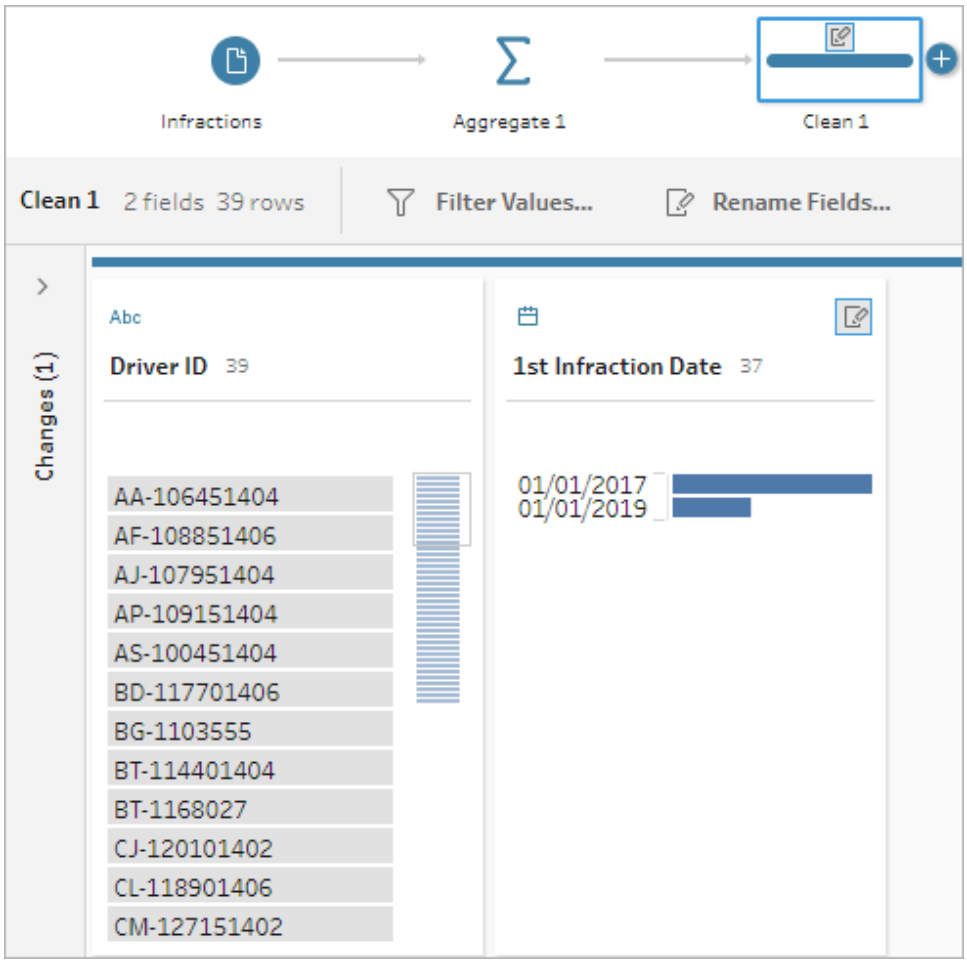


这将标识最小（最早）日期，即每个司机的第一次违章日期。

有关聚合的详细信息，请参见**清理和调整数据** 在本页213。

8. 在“流程”窗格中，选择**“聚合 1”**，单击加号 ⊕ 图标，并选择**“清理步骤”**，以便我们能够清理聚合的输出。
9. 在**“配置”**窗格中，双击字段名“**Infraction Date**”（违章日期），并将其更改为“**1st Infraction Date**”（第一次违章日期）。

在此阶段，流程和配置窗格应如下所示：

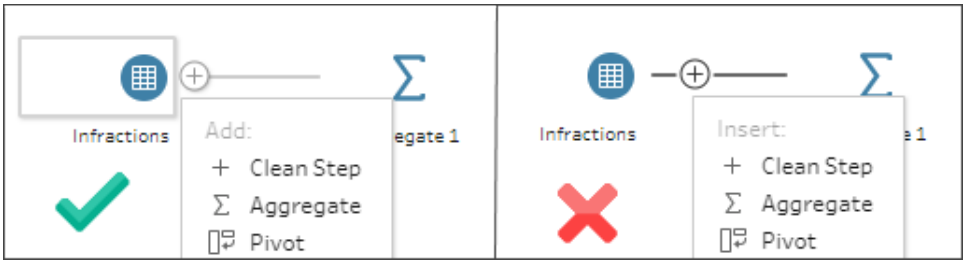


从此“清理”步骤的“配置”窗格中，我们可以看到数据现在包含 39 行，并且只包含 2 个字段。未用于分组或聚合的任何字段已丢失。但我们想要能够保留一些原始信息。我们可将这些字段添加到分组或聚合中(但这样做会更改详细级别或者需要对字段进行聚合)，或者将此微型数据集联接回原始数据集(本质上是为“**1st Infraction Date**”(第一次违章日期)向原始数据中添加一个新列)。让我们执行联接。

10. 在“流程”窗格中，选择“**Infractions**”(违章)，单击加号 ⊕ 图标，并选择**“清理步骤”**。

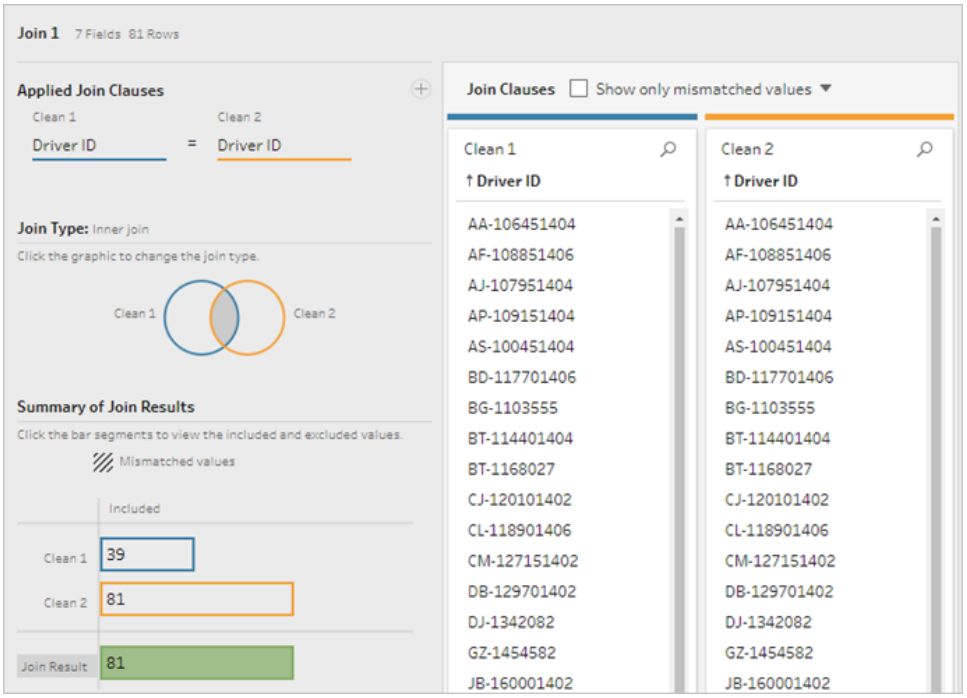
    请确保将鼠标直接悬停在违章步骤上，而不是它与聚合步骤之间的线上。如果在两者之间(而不是分支之间)插入新“清理”步骤，请使用工具栏中的“撤消”箭头，然后重

试。菜单应显示“*添加*”，而不是“*插入*”。



这会使用所有原始数据将您的流程分支。我们会将聚合的结果联接到完整数据的这一副本。通过依据“**Driver ID**”(司机 ID)进行联接，我们会将聚合中的最小日期添加到原始数据中。

11. 选择步骤“**清理 2**”，并将其拖到步骤“**清理 1**”上，放在**“联接”**上。
12. 默认联接配置应该是正确的：基于“**Driver ID = Driver ID**”的内部联接。

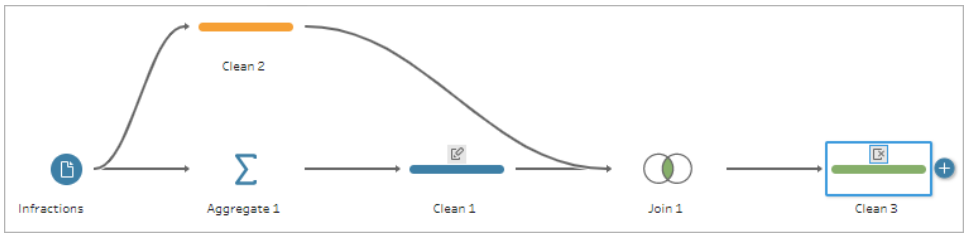


有关联接的详细信息，请参见**联接数据** 在本页340。

由于在联接过程中某些字段可能是重复的，例如联接子句中的字段，因此在执行联接之后清理无关的字段通常是个好主意。

13. 在“流程”窗格中，选择“**Join 1**”（联接 1），单击加号 ⊕ 图标，并选择**“清理步骤”**。
14. 在“配置”窗格中，右键单击或按住 Ctrl 单击 (MacOS)“**Driver ID-1**”（司机 ID-1）卡，然后选择**“移除”**。
15. 若要更改字段顺序，请将“**1st Infraction Date**”（第一次违章日期）卡拖到“**Driver ID**”（司机 ID）和“**Infraction Date**”（违章日期）之间您看到黑色行出现的位置。

在此阶段，流程应如下所示：



查看下面的数据网格，我们可以看到新的合并数据集。我们有了添加到数据集中每一行的每个司机的最小（即第一次）违章日期。

Driver ID	1st Infraction Date	Infraction Date	Infraction Type	Traffic School	Fine Amount
JO-151451402	01/08/2017	01/08/2017	Speeding	Yes	115
CM-127151402	03/01/2017	03/01/2017	Running a red light	No	55
AP-109151404	03/02/2017	03/02/2017	Non-moving violation	No	95
SH-199751404	03/04/2017	03/04/2017	Speeding	Yes	130
BT-114401404	03/20/2017	03/20/2017	Non-moving violation	No	130
MO-175001406	05/30/2017	05/30/2017	Speeding	Yes	118
RA-1988558	06/02/2017	06/02/2017	Speeding	Yes	144
BT-1168027	06/05/2017	06/05/2017	Speeding	Yes	128
MO-175001406	05/30/2017	06/18/2017	Speeding	Yes	115
MP-174701406	06/19/2017	06/19/2017	Speeding	No	125
AA-106451404	07/05/2017	07/05/2017	Running a red light	No	60
RA-199151402	07/20/2017	07/20/2017	Speeding	Yes	146
SC-202601404	08/31/2017	08/31/2017	Running a red light	No	150
MO-175001406	05/30/2017	09/07/2017	Non-moving violation	No	320
AS-100451404	09/26/2017	09/26/2017	Running a red light	No	50
SH-199751404	03/04/2017	09/27/2017	Speeding	Yes	225
AA-106451404	07/05/2017	09/28/2017	Running a red light	No	195

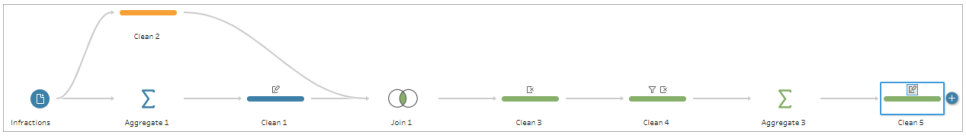
第二次违章日期的第二个聚合

我们还需要确定第二次违章日期。为此，我们需要筛选出违章日期等于最小日期的任何行—从而移除第一个日期。我们随后可以使用另一个聚合步骤获取剩余日期的最小值，剩下的就是第一次违章日期，为了清楚起见，我们将重命名该日期。

注意：由于我们将需要使用当前位于流程后面**“清理 3”**中的数据，因此我们将添加另一个**“清理”**步骤来获取第二次违章日期。这将在稍后使“清理 3”中数据的当前状态可用。

16. 在“流程”窗格中，选择**“Clean 3”**（清理 3），单击加号 ⊕ 图标，并选择**“清理步骤”**。
17. 在“配置”窗格上的工具栏中，选择**“筛选值”**。创建一个筛选器 `[Infraction Date] != [1st Infraction Date]`。
18. 移除字段**“1st Infraction Date”**（第一次违章日期）。
19. 在“流程”窗格中，选择**“Clean 4”**（清理 4），单击加号 ⊕ 图标，并选择**“聚合”**。
20. 将**“Driver ID”**（司机 ID）拖到**“分组字段”**放置区域。将**“Infraction Date”**（违章日期）拖到**“聚合字段”**区域，并将聚合更改为**“Minimum”**。
21. 在“流程”窗格中，选择**“Aggregate 2”**（聚合 2），单击加号 ⊕ 图标，并选择**“清理步骤”**。将**“Infraction Date”**（违章日期）重命名为**“2nd Infraction Date”**（第二次违章日期）。

在此阶段，流程应如下所示：



我们现在为每个司机标识了第二次违章日期。为了获取与每次违章关联的所有其他信息（类型、罚款、交通法规学习班），我们需要再次将此项联接回整个数据集。

22. 选择**“清理 5”**，并将其拖到步骤**“清理 3”**，放在**“联接”**上。
23. 同样，默认联接配置应该是正确的：基于**“Driver ID = Driver ID”**的内部联接。
24. 在**“流程”**窗格中，选择**“Join 2”**（联接 2），单击加号 ⊕ 图标，并选择**“清理步骤”**。删除字段**“Driver ID-1”**（司机 ID-1）和**“1st Infraction Date”**（第一次违章日期），因为不再需要这些字段。

在此阶段，流程应如下所示：

为第一次和第二次违章创建完整数据集

在执行任何进一步操作之前，让我们回顾一下，想一想我们已有的所有内容以及要如何将它们结合在一起。我们期望的最终状态是一个像这样的数据集：包含一个“**Driver ID**”（司机 ID）列，然后包含日期、类型、交通法规学习班列，以及第一次和第二次违章的罚款金额。

	A	B	C	D	E	F	G	H	I
1	Driver ID	1st Infraction Date	1st Infraction Type	1st Traffic School	1st Fine Amount	2nd Infraction Date	2nd Infraction Type	2nd Traffic School	2nd Fine Amount
2	BD-117701406	12/25/2017	Speeding	Yes	140	2/7/2018	Speeding	Yes	125
3	JO-151451402	1/8/2017	Speeding	Yes	115	11/21/2018	Reckless driving	Yes	550
4	SN-207101402	12/27/2017	Speeding	Yes	280	4/26/2018	Speeding	Yes	130
5	CJ-120101402	11/26/2017	Speeding	Yes	122	3/28/2018	Speeding	Yes	116
6	JR-156701404	12/24/2017	Speeding	No	148	7/28/2018	Speeding	Yes	310
7	AP-109151404	3/2/2017	Non-moving violation	No	95	9/24/2018	Speeding	No	105
8	PC-187451406	11/11/2017	Speeding	Yes	220	12/30/2018	Non-moving violation	No	600
9	TS-214301406	9/13/2018	Speeding	Yes	115	11/10/2018	Non-moving violation	No	95
10	NP-187001404	12/11/2018	Non-moving violation	No	80	12/20/2018	Speeding	No	120
11	DB-129701402	5/13/2018	Running a red light	No	110	11/11/2018	Speeding	Yes	80
12	AJ-107951404	10/15/2017	Speeding	Yes	130	12/31/2017	Running a red light	No	85
13	BT-114401404	3/20/2017	Non-moving violation	No	130	11/13/2018	Speeding	Yes	96
14	AF-108851406	5/9/2018	Non-moving violation	No	200	9/2/2018	Speeding	No	130
15	SC-202601404	8/31/2017	Running a red light	No	150	11/10/2018	Speeding	Yes	50
16	KL-166451406	10/4/2017	Speeding	No	115	11/13/2017	Speeding	Yes	104
17	MO-175001406	5/30/2017	Speeding	Yes	118	6/18/2017	Speeding	Yes	115
18	CM-127151402	3/1/2017	Running a red light	No	55	8/1/2018	Running a red light	No	160
19	KT-164801402	5/31/2018	Non-moving violation	No	190	11/10/2018	Speeding	No	74
20	JB-160001402	11/18/2018	Speeding	Yes	220	12/5/2018	Non-moving violation	No	195
21	LH-170201404	5/6/2018	Running a red light	No	110	9/17/2018	Speeding	Yes	230
22	BG-1103555	12/25/2017	Speeding	Yes	195	12/8/2018	Speeding	Yes	315
23	MP-174701406	6/19/2017	Speeding	No	125	10/12/2017	Running a red light	No	175

我们怎样从这里实现该目标？

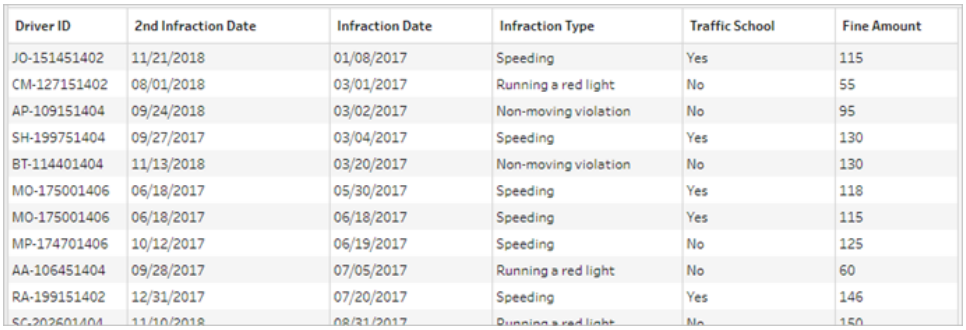
在步骤“**Clean 3**”（清理 3）中，我们有完整的数据集，其中有一列重复显示每个司机的第一次违章日期。

Driver ID	1st Infraction Date	Infraction Date	Infraction Type	Traffic School	Fine Amount
JO-151451402	01/08/2017	01/08/2017	Speeding	Yes	115
CM-127151402	03/01/2017	03/01/2017	Running a red light	No	55
AP-109151404	03/02/2017	03/02/2017	Non-moving violation	No	95
SH-199751404	03/04/2017	03/04/2017	Speeding	Yes	130
BT-114401404	03/20/2017	03/20/2017	Non-moving violation	No	130
MO-175001406	05/30/2017	05/30/2017	Speeding	Yes	118
RA-1988558	06/02/2017	06/02/2017	Speeding	Yes	144
BT-1168027	06/05/2017	06/05/2017	Speeding	Yes	128
MO-175001406	05/30/2017	06/18/2017	Speeding	Yes	115
MP-174701406	06/19/2017	06/19/2017	Speeding	No	125

我们想要清除不是第一次违章的司机的所有行，同时构建一个只包含第一次违章的数据集。也就是说，我们只想保留**“第一次违章日期”与“违章日期”相等**时给定司机的信息。筛

选为只保留第一次违章的行之后，我们可以移除“**Infraction Date**”(筛选日期)字段，并整理字段名称。

同样，在进行第二次聚合和联接之后，我们有了完整的数据集，其中有一列表示第二次违章日期。



Driver ID	2nd Infraction Date	Infraction Date	Infraction Type	Traffic School	Fine Amount
JO-151451402	11/21/2018	01/08/2017	Speeding	Yes	115
CM-127151402	08/01/2018	03/01/2017	Running a red light	No	55
AP-109151404	09/24/2018	03/02/2017	Non-moving violation	No	95
SH-199751404	09/27/2017	03/04/2017	Speeding	Yes	130
BT-114401404	11/13/2018	03/20/2017	Non-moving violation	No	130
MO-175001406	06/18/2017	05/30/2017	Speeding	Yes	118
MO-175001406	06/18/2017	06/18/2017	Speeding	Yes	115
MP-174701406	10/12/2017	06/19/2017	Speeding	No	125
AA-106451404	09/28/2017	07/05/2017	Running a red light	No	60
RA-199151402	12/31/2017	07/20/2017	Speeding	Yes	146
SC-202601404	11/10/2018	08/31/2017	Running a red light	No	150

我们可以执行类似的筛选器“**2nd Infraction Date = Infraction Date**”以仅保留每个司机的第二次违章的信息行。同样，我们也可以移除现在多余的“**Infraction Date**”(违章日期)并整理字段名称。

我们将从第一个违章数据集开始。

25. 在“流程”窗格中，选择“**Clean 3**”(清理 3)，单击加号 ⊕ 图标，并选择**“清理步骤”**。

    如上文步骤 10 所示，我们希望为新的清洁步骤添加一个分支，而不是将其插入“Clean 3”(清理 3)和“Clean 4”(清理 4) 之间。

26. 在这个新的**“清理”**步骤处于选定状态的情况下，在**“配置”**窗格的工具栏中单击**“筛选值”**。创建一个筛选器 `[1st Infraction Date] = [Infraction Date]`。

27. 移除字段“**Infraction Date**”(违章日期)。

28. 将“**Infraction Type**”(违章类型)、“**Traffic School**”(交通法规学习班)和“**Fine Amount**”(罚款金额)字段重命名为以“1st”开头。

29. 在**“流程”**窗格中的步骤下双击名称“**Clean 7**”(清理 7)，并将其重命名为“**Robust 1st**”。

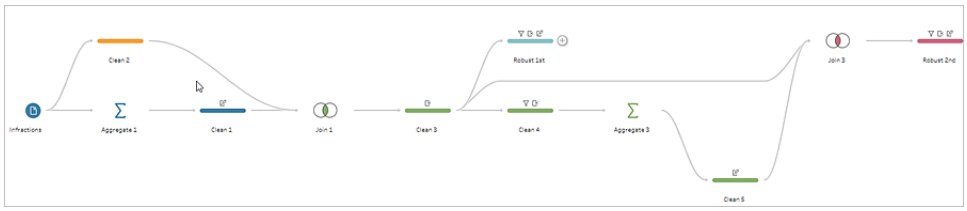
现在为第二个违章数据集执行操作。

30. 在“流程”窗格中的最后一个联接后选择“**Clean 6**”(清理 6)。

31. 在工具栏中单击**“筛选值”**。创建一个筛选器 `[2nd Infraction Date] = [Infraction Date]`。

32. 移除字段“**Infraction Date**”(违章日期)。
33. 将“**Infraction Type**”(违章类型)、“**Traffic School**”(交通法规学习班)和“**Fine Amount**”(罚款金额)字段重命名为以“2nd”开头。
34. 在“流程”窗格中的步骤下双击名称“**Clean 6**”(清理 6),并将其重命名为“**Robust 2nd**”。

在此阶段,流程应如下所示:



创建完整数据集

既然我们有了这两个包含各个司机第一次和第二次违章完整信息的两个整齐数据集,我们就可以基于“**Driver ID**”(司机 ID)将它们重新联接在一起,并最终生成我们期望的数据结构。

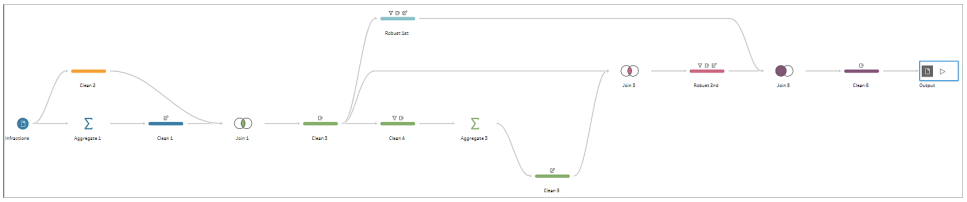
35. 选择“**Robust 2nd**”,并将其拖到“**Robust 1st**”,放在**“联接”**上。
36. 默认联接子句“**Driver ID = Driver ID**”应该是正确的。
37. 由于我们不想删除没有第二次违章的司机,我们需要将此联接设为左联接。在**“联接类型”**区域中,单击“**Robust 1st**”旁边图表的无阴影区域,将其转换为“**左**”联接。
38. 在“流程”窗格中,选择“**Join 3**”(联接 3),单击加号 ⊕ 图标,并选择**“清理步骤”**。移除重复字段“**Driver ID-1**”(司机 ID-1)。

数据处于期望状态,因此我们可以创建输出并继续进行分析。

39. 在“流程”窗格中,选择新添加的“**Clean 6**”(清理 6),单击加号 ⊕ 图标,并选择**“添加输出”**。
40. 在**“输出”**窗格中,将**“输出类型”**更改为 .csv,然后单击**“浏览”**。为名称输入“**Driver Infractions**”(司机违章),并在单击**“接受”**保存之前选择所需的位置。
41. 单击窗格底部的**“运行流程”** ▹ 按钮生成输出。在状态对话框中单击**“完成”**关闭对话框。

提示：有关输出和运行流程的详细信息，请参见保存和共享工作 在本页361。

最终流程应如下所示：



注意：您可以下载完成的流程文件来检查您的工作：Driver Infractions.tflx

总结

对于此教程的第一个阶段，我们的目标是获取原始数据集，并准备将其用于涉及第一次和第二次违章日期的分析。该过程由三个阶段组成：

确定第一次和第二次违章日期：

1. 创建用于保留“**Driver ID**”（司机 ID）和“MIN **Infraction Date**”（最小违章日期）的聚合。将此聚合与原始数据集联接，创建一个“中间数据集”，其中包含为每一行重复的第一次（最小）违章日期。
2. 在新步骤上，筛选出“**1st Infraction Date**”（第一次违章日期）与“**Infraction Date**”（违章日期）相同的所有行。从该筛选的数据集中，创建用于保留“**Driver ID**”（司机 ID）和“MIN **Infraction date**”（最小违章日期）的聚合。将此聚合与第一个步骤中的中间数据集联接。这将标识第二次违章日期。

为第一次和第二次违章构建干净的数据集：

3. 返回并依据中间数据集创建一个分支，并筛选为仅包含“**1st Infraction Date**”（第一次违章日期）与“**Infraction Date**”（违章日期）相同的行。这将只会为第一次违章构建数据集。通过移除任何不必要的字段对其进行整理，并重命名所有所需字段（“**Driver ID**”（司机 ID）除外），以指明它们用于第一次违章。这是“**Robust 1st**”数据集。
4. 针对第二次违章日期整理数据集。通过筛选为仅保留“**2nd Infraction Date**”（第二次违章日期）与“**Infraction Date**”（违章日期）相同的行，对步骤 2 中的联接结果进行清

理。移除任何不必要的字段，并重命名所有所需字段（“**Driver ID**”（司机 ID）除外），以指明它们用于第二次违章。这是“**Robust 2nd**”数据集。

将第一次和第二次违章数据合并为一个数据集：

5. 联接“**Robust 1st**”和“**Robust 2nd**”数据集，同时确保保留“**Robust 1st**”中的所有记录以防止丢失没有第二次违章的任何司机。

接着，我们想要探索如何能在 Tableau Desktop 中分析此数据。

继续阅读**Tableau Desktop 中包含第二个日期的分析** 向下。

注意：特别感谢 Ann Jackson 的 Workout Wednesday 主题 Do Customers Spend More on Their First or Second Purchase?（客户在第一次或第二次购买时是否花费更多？）以及 Andy Kriebel 的 Tableau Prep 技巧 Returning the First and Second Purchase Dates（返回第一次和第二次购买日期），这些文章为本教程提供了最初的灵感。单击这些链接会使您离开 Tableau 网站。Tableau 对外部提供商所维护的页面的准确性或新鲜度不担负任何责任。如果您对其内容有任何疑问，请与所有者联系。

Tableau Desktop 中包含第二个日期的分析

这是教程的第二个阶段，并且假定第一个阶段**使用 Tableau Prep 查找第二个日期** 在本页 467已完成。

在第一个阶段中，我们采用原始数据集并对其进行调整来回答以下问题：

1. 每个司机第一次和第二次违章相隔多长时间（以天数为单位）？
2. 比较第一次和第二次违章的罚款金额。这些金额是否相关？
3. 总体而言，哪个司机支付的罚款最多？谁支付的最少？
4. 有多少司机有过多种类型的违章？
5. 从未上过交通法规学习班的司机的平均罚款金额是多少？

我们现在探讨这些问题时，可以清晰地发现我们创建的第一个数据结构有一些优缺点。我们将返回到 Tableau Prep Builder 中并进行一些额外的重新调整，然后看看会对 Tableau Desktop 中同样的分析产生怎样的影响。最后，我们将探讨 Tableau Desktop 独有的将详细级别 (LOD) 表达式与原始数据结合使用进行分析的方法。

本教程的目标是在一个真实场景的上下文中呈现各种概念并演练各个选项 — 而不是规定式地确定哪一个是最好的。最后，您应该会对数据结构对计算和分析的影响有更深的认识，并更加熟悉 Tableau Prep 的各项功能以及 Tableau Desktop 中的计算。

注意：为了完成本教程中的任务，您需要安装 Tableau Prep Builder 并根据需要安装 Tableau Desktop，并下载相关数据。

若要在继续学习本教程之前安装 Tableau Prep 和 Tableau Desktop，请参见 Tableau Desktop 和 Tableau Prep 部署指南。否则，您可以下载 Tableau Prep 和 Tableau Desktop 免费试用版。

数据集是在第一个阶段中构建的 Driver Infractions.tflx 的输出。

Tableau Desktop 中的分析

既然我们已经配置了数据，我们就将其引入 Tableau Desktop。我们可以轻松地回答一些问题，但其他问题会涉及一些（或很多）计算。请尝试下面的问题；如果遇到困难，您可以展开每个问题来了解有关如何继续的基本信息。

注意：您可以下载工作簿 Driver Infractions.twbx，在上下文中探讨解决方案。请记住，可以通过一些替代方式来解读分析或寻找答案。

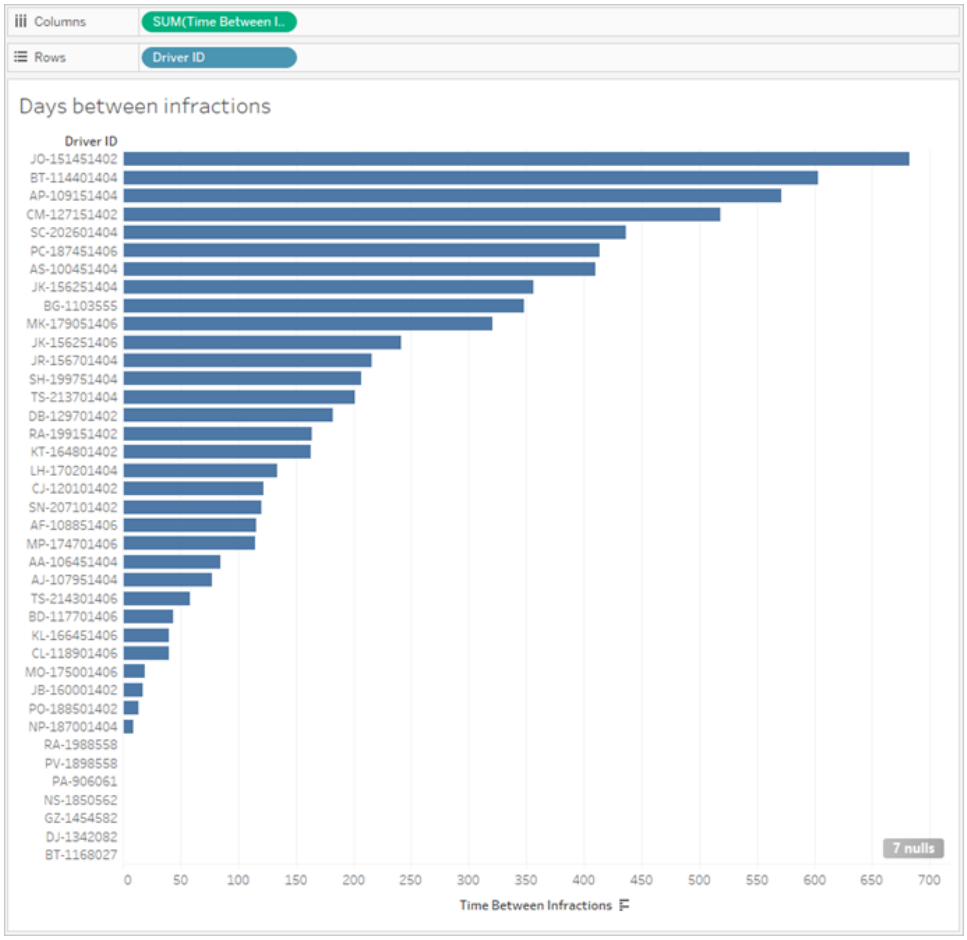
1.每个司机第一次和第二次违章相隔多长时间（以天数为单位）？

A. 为了在 Tableau Desktop 中回答此问题，我们将使用 DATEDIFF 函数。此函数采用三个参数 — 日期部分、开始日期和结束日期。由于我们想知道这些事件的间隔天数，因此将使用日期部分“day”。我们的开始日期和结束日期在数据集中为“**1st Infraction Date**”（第一次违章日期）和“**2nd Infraction Date**”（第二次违章日期）。

B. 计算为：

Time Between Infractions（违章间隔时间）= `DATEDIFF('day', [1st Infraction Date], [2nd Infraction Date])`

C. 我们依据“**Driver ID**”(司机 ID)将该计算绘制为条形图。请注意，7 名司机没有第二次违章，因此有 7 个 null 值。

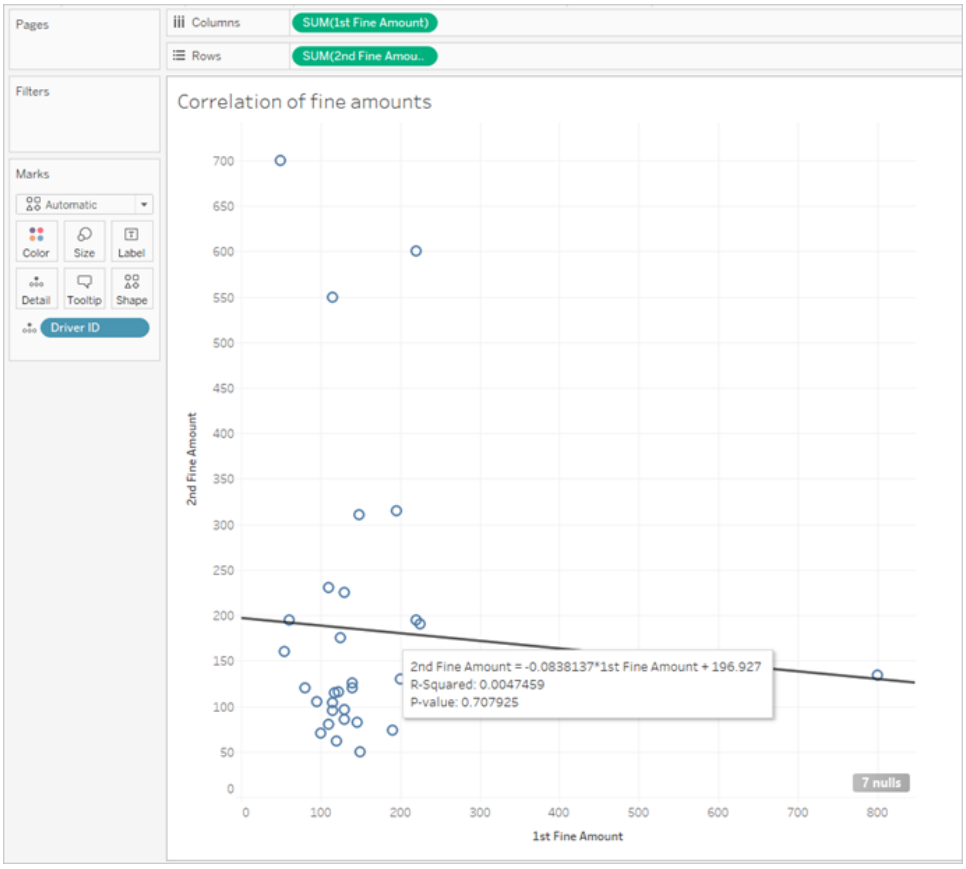


2.比较第一次和第二次违章的罚款金额。这些金额是否相关？

A. 为了在 Tableau Desktop 中回答此问题，我们将创建“**1st Fine Amount**”(第一次罚款金额)和“**2nd Fine Amount**”(第二次罚款金额)的散点图。通过将“**Driver ID**”(司机 ID)放到**“标记”**卡上的**“详细信息”**功能区，我们可以为每名司机创建一个标记。

B. 若要添加趋势线，请使用左侧窗格中的**“分析”**选项卡，并显示线性趋势线。将光标悬停在趋势线上，我们可以看到 R 平方值实际上为零，并且 p 值明显高于任何临界值。我们可以确定，第一次罚款金额和第二次罚款金额之间没有相关性。

如果我们要在仪表板中使用此散点图，则应移除趋势线。



3.总体而言，哪个司机支付的罚款最多？谁支付的最少？

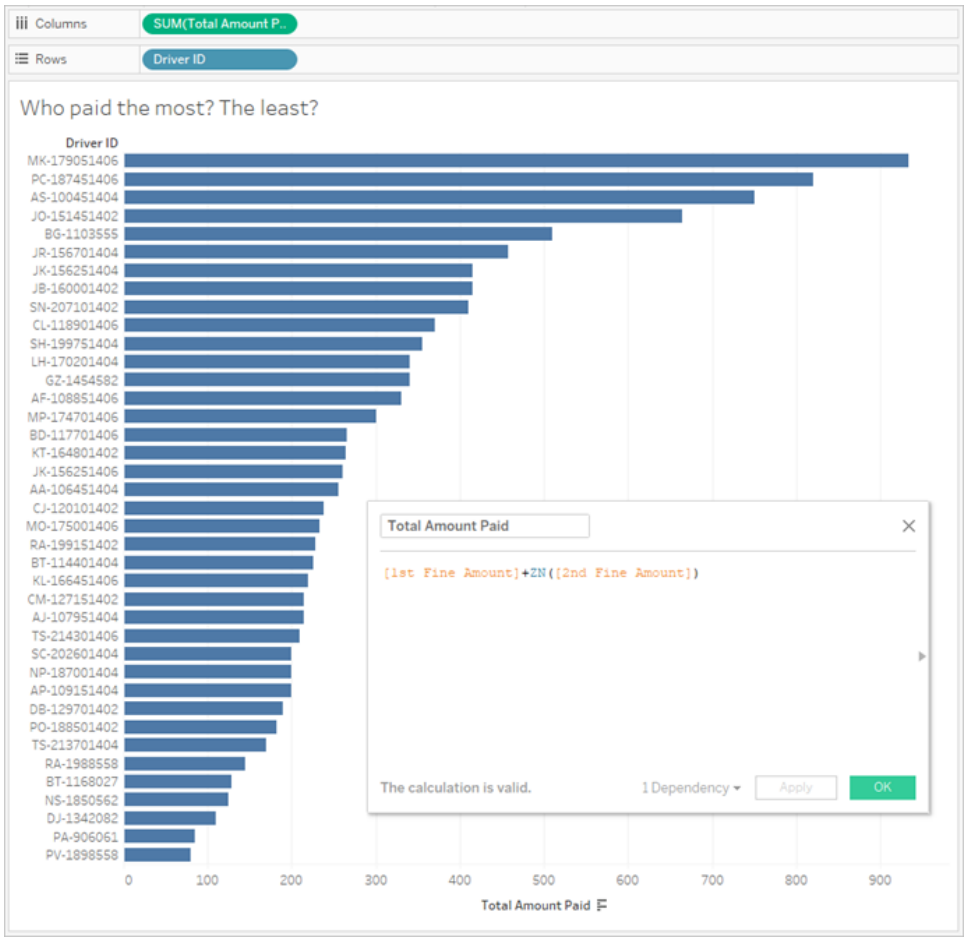
如果想要更深入地进行分析，我们可能需要创建一些计算。

A. 为了在 Tableau Desktop 中回答此问题，我们需要将两次违章的罚款添加到单一字段中。由于某些司机可能没有第二次违章，我们需要使用零 null ZN 函数将“**2nd Fine Amount**”(第二次罚款金额)的任何 null 值转换为零。如果没有第二次罚款，不这样做将导致 null。

B. 计算为：

Total Amount Paid(总支付金额) = `[1st Fine Amount] + ZN([2nd Fine Amount])`

C. 我们可以依据“**Driver ID**”(司机 ID)绘制“**Total Amount Paid**”(总支付金额)条形图并对其进行排序。



4.有多少司机有过多种类型的违章?

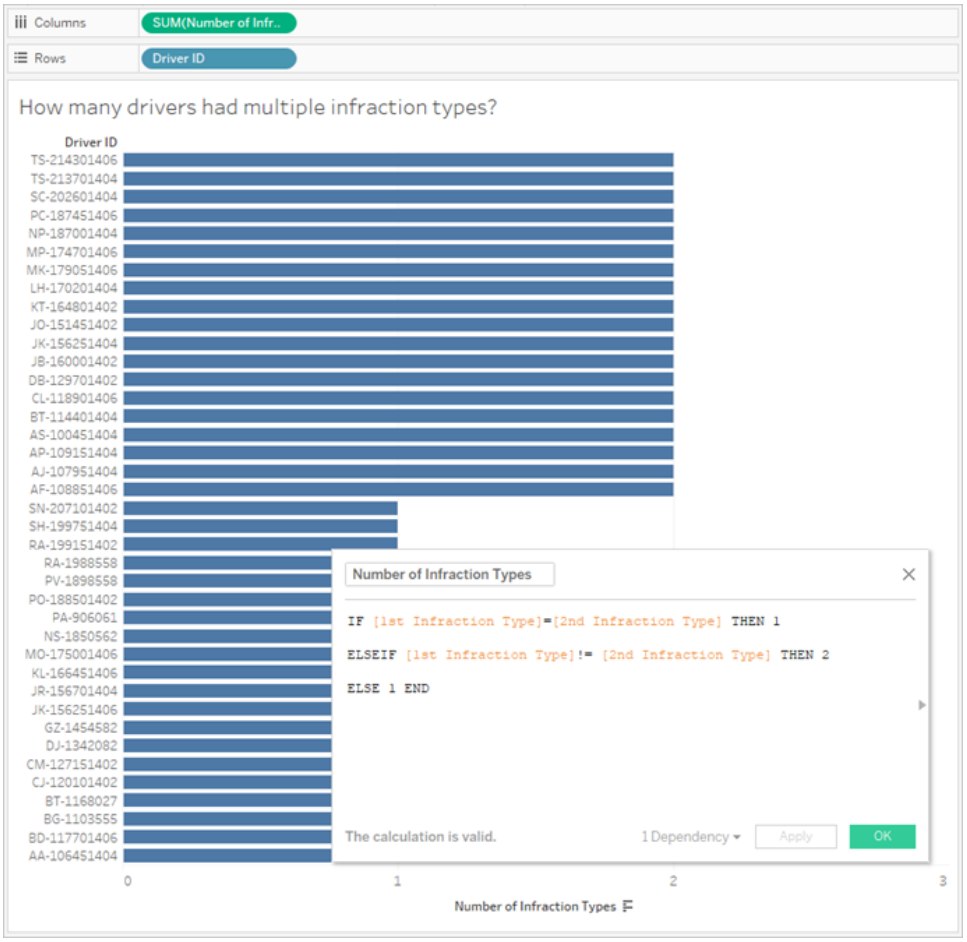
A. 为了在 Tableau Desktop 中回答此问题,我们需要执行一个更细致的 IF 计算,比较第一次和第二次违章类型是否相同。如果相同,我们想要分配值“1”。如果不同,我们将分配“2”。由于我们仅关心多个违章类型,因此任何其他结果(例如 null 第二次违章类型)将被分配“1”。

B. 计算为:

Number of Infraction Types(违章类型数量) =

```
IF [1st Infraction Type]=[2nd Infraction Type] THEN 1
ELSEIF [1st Infraction Type]!= [2nd Infraction Type] THEN 2
ELSE 1 END
```

C. 然后，我们可以依据“**Driver ID**”(司机 ID)绘制“**Number of Infraction Types**”(违章类型数量)条形图并对其进行排序。



5.从未上过交通法规学习班的司机的平均罚款金额是多少？

A. 为了在 Tableau Desktop 中回答此问题，我们无法简单地将总罚款金额除以二，因为某些司机只有一次违章。我们也无法计算每个司机的平均罚款并取这些值的平

均值，因为对平均值进行平均计算可能会导致不一致。作为替代，我们需要计算从未上过交通法规学习班的司机支付的总金额，然后除以与这些罚款关联的总违章次数。

1. 首先，我们需要确定是否每个司机都有第二次违章。如果没有第二次违章，则所有“2nd”字段中的信息都将为 null，我们可以利用这一事实并开始构建计算：

   ```
   IFNULL([2nd Infraction Type], 'no')
   ```

   这将返回违章类型（如果存在），或者如果没有第二次违章，则返回“no”。

2. 接着，我们需要将此信息转换为违章次数 1 或 2。如果 IFNULL 计算的结果为“no”，则应将司机标记为有一次罚款。对于任何其他结果，则应标记为有两次罚款。计算为：

   Number of Infractions（违章次数）=

   ```
   IF IFNULL([2nd Infraction Type], 'no') = 'no' THEN 1
   ELSE 2
   END
   ```

3. 现在，我们需要考虑总罚款金额。与上面的问题 3 类似，我们将添加第一次和第二次罚款金额，并将第二次罚款金额放在 ZN 函数内。但是，由于我们想在整个数据集级别计算此值，因此最佳做法是在计算本身中指定聚合 **SUM**。计算为：

   ```
   SUM([1st Fine Amount]) + SUM( ZN([2nd Fine Amount]) )
   ```

4. 总而言之，我们将利用这个总罚款金额，并将其除以新的“**Number of Infractions**”（违章次数）计算字段，从而确定平均罚款金额：

   Average Fine（平均罚款）= `( SUM([1st Fine Amount]) + SUM( ZN([2nd Fine Amount]) ) ) / SUM([Number of Infractions])`

B. 我们还需要筛选出曾经参加过交通法规学习班的司机 — 但该信息也跨两个字段存储。

1. Tableau 在进行数值计算时非常高效。我们将用尽可能多可帮助提高性能的数字表述这一点。为了合并这两个字段，我们将为每个字段创建一个显示为“Yes = 1”和“No = 0”的字段（对于没有第二次违章的司机，null 也应 = 0）。通过对这些计算的结果求和，总值为 0 的任何司机从未上过交通法规学习班（值 1 或 2 表示他们上过交通法规学习班的次数）。我们随后可以进行筛选，以仅保留值为 0 的司机。

2. 这一次，我们将使用 CASE 语句，而不是 IF。这些语句的功能非常相似，但具有不同的语法。计算的开头应如下所示：

```
CASE [1st Traffic School]
WHEN 'Yes' THEN 1
WHEN 'No' THEN
ELSE 0
END
```

3. 然后，我们将为“2nd Traffic School”执行相同的操作。通过将每个 CASE 语句放在括号内并在两者之间添加一个加号，我们可以在同一计算中添加这两段语句。移除一些换行符，它看起来如下所示：

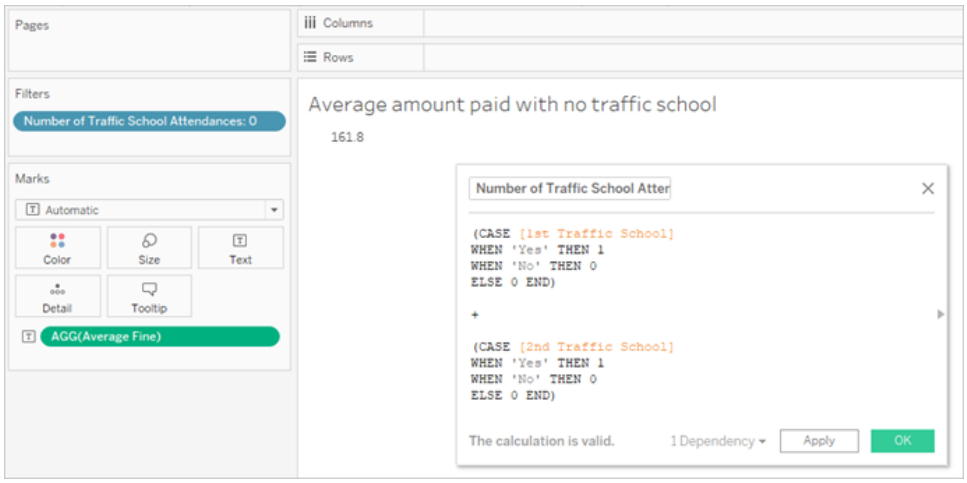
   Number of Traffic School Attendances(参加交通法规学习班次数) =

```
(CASE [1st Traffic School] WHEN 'Yes' THEN 1 WHEN 'No' THEN
0 ELSE 0 END)
+
(CASE [2nd Traffic School] WHEN 'Yes' THEN 1 WHEN 'No' THEN
0 ELSE 0 END)
```

4. 如果我们将“**Number of Traffic School Attendances**”(参加交通法规学习班次数) 拖到**“数据”**窗格的“维度”区域，则值 0–2 将变为离散。
5. 现在，如果我们对“**Number of Traffic School Attendances**”(参加交通法规学习班次数) 进行筛选，我们可以只选择 0，并且知道我们获得了从未参加过交通法规学习班的司机。

C. 为了回答原来的问题，我们只需将“**Average Fine**”(平均罚款) 放到**“标记”**卡上的**“文本”**功能区。

   由于我们在计算中构建了聚合，因此胶囊上的聚合将为 **AGG**，我们无法对其进行更改。这是预期行为。



更进一步 – 转置的数据

尽管我们一直在处理的数据结构良好，可以解决与第一次和第二次违章特定相关的问题，但它并不是我们建议用于 Tableau Desktop 的标准结构。我们的分析背离与违章日期相关的基本问题越远，用于将相关信息合并为可用形式的计算就会变得越复杂。

通常，如果为同一类型的数据使用多个列存储数据（例如两个列用于日期，两个列用于罚款金额等），并且唯一的信息存储在字段名中（例如是第一次还是第二次违章），这就表明应该要对数据进行转置。

在 Tableau Prep Builder 中执行多重转置可以很好地处理这一点。我们可以从上一教程**使用 Tableau Prep 查找第二个日期** 在本页467中创建的**“Driver Infraction”**（司机违章）Tableau Prep 流程的末尾开始工作。

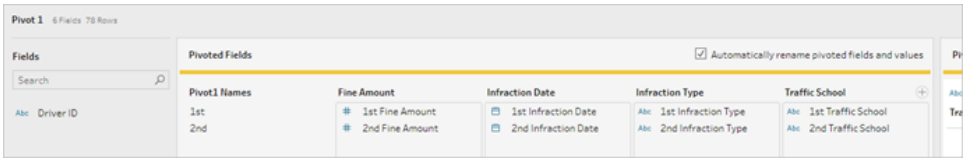
提示：请确保返回 Tableau Prep 来执行这些步骤。

1. 从最后的清理步骤中，添加一个**“Pivot”**（转置）步骤，该步骤将按每个重复的字段进行转置。使用**“转置的字段”**区域右上角的加号 ⊕ 图标来添加更多**“转置值”**。每一组字段（例如第一次和第二次罚款金额）应一起转置。

   有关转置的详细信息，请参见**清理和调整数据** 在本页213。

2. 在“转置的字段”区域中的**“Pivot1 Names”**（转置 1 名称）列下，双击每个值并将其重

命名为“1st”(第一)和“2nd”(第二)。



通过移除 null 日期以及重命名字段并对其进行重新排序，可以对结果进行整理。

3. 在转置后面添加清理步骤。在“**Infraction Date**”(违章日期)列中，右键单击 null 栏并选择**“排除”**。
4. 双击字段名“**Pivot1 Names**”(转置 1 名称)，并将其重命名“**Infraction Number**”(违章次数)。
5. 视情况而定拖动数据以按如下方式对其进行重新排序：

Driver ID	Infraction Number	Infraction Date	Infraction Type	Traffic School	Fine Amount
MO-175001406	1st	05/30/2017	Speeding	Yes	118
SH-199751404	1st	03/04/2017	Speeding	Yes	130
AA-106451404	1st	07/05/2017	Running a red light	No	60
MP-174701406	1st	06/19/2017	Speeding	No	125
PO-188501402	1st	10/30/2017	Speeding	Yes	120
KL-166451406	1st	10/04/2017	Speeding	No	115
RA-199151402	1st	07/20/2017	Speeding	Yes	146
AJ-107951404	1st	10/15/2017	Speeding	Yes	130
BD-117701406	1st	12/25/2017	Speeding	Yes	140
CJ-120101402	1st	11/26/2017	Speeding	Yes	122
SN-207101402	1st	12/27/2017	Speeding	Yes	280
TS-213701404	1st	10/23/2017	Speeding	Yes	100
JR-156701404	1st	12/24/2017	Speeding	No	148
CM-127151402	1st	03/01/2017	Running a red light	No	55
JK-156251406	1st	12/25/2017	Speeding	Yes	140

6. 依据新的转置的数据创建一个名为“**Pivoted Driver Infractions**”(转置的司机违章)的输出，并将其引入 Tableau Desktop。(在添加**“输出”**步骤之后，不要忘记运行流程。)

现在，我们可以用这个转置的数据结构再次探讨我们的五个问题；如果遇到困难，您可以展开每个问题来了解有关如何继续操作的基本信息。

注意：您可以下载完成的流程文件 Pivoted Driver Infractions.tflx 来检查您的工作，或者下载工作簿 Pivoted Driver Infractions.twbx 在上下文中查看解决方案。请记住，可以通过一些替代方式来解读分析或寻找答案。

1.每个司机第一次和第二次违章相隔多长时间(以天数为单位)?

A. 为了在 Tableau Desktop 中回答此问题,就像我们对第一个数据集所做的那样,我们将使用 DATEDIFF 函数。此函数需要一个开始日期和一个结束日期。此信息存在于我们的数据中,但全部在一个字段中。我们需要将其拉出到两个字段中。

1. 创建两个初始计算字段:

   1st Infraction Date(第一次违章日期) = `IF [Infraction Number] = "1st" THEN [Infraction Date] END`

   2nd Infraction Date(第二次违章日期) = `IF [Infraction Number] = "2nd" THEN [Infraction Date] END`

2. 由于我们想要确保这两个字段均可用于为每个司机比较,我们需要将它们固定到"**Driver ID**"(司机 ID)级别。

   注意:不相信我?尝试按原样对这两个字段进行 DATEDIFF 计算:**Time Between Infractions**(违章间隔时间) = `DATEDIFF('day', [1st Infraction Date], [2nd Infraction Date])`
   您将在所有地方都获得 null 结果,因为 Tableau 会尝试跨类似于如下的数据结构进行比较:

Driver ID	1st Infraction Date	2nd Infraction Date	Time between infractions
AA-106451404	Null	9/28/2017	Null
	7/5/2017	Null	Null
AF-108851406	Null	9/2/2018	Null
	5/9/2018	Null	Null
AJ-107951404	Null	12/31/2017	Null
	10/15/2017	Null	Null

   在这里,知道第一个日期的行不知道第二个日期,反之亦然。为了解决此问题,我们将使用"FIXED"详细级别表达式将第一个和第二个日期强制与"**Driver ID**"(司机 ID)相关。

   按如下方式编辑计算:

1st Infraction Date(第一次违章日期) = `{ FIXED [Driver ID] : MIN ( IF [Infraction Number] = "1st" THEN [Infraction Date] END ) }`

2nd Infraction Date(第二次违章日期) = `{ FIXED [Driver ID] : MIN ( IF [Infraction Number] = "2nd" THEN [Infraction Date] END ) }`

注意:如果嵌入在 LOD 表达式中,必须对原始 IF 计算进行聚合。我们可以使用任何将保留日期值的基本聚合(因此像 SUM、AVG 或 MIN 这样的聚合有用,而 CNT 或 CNTD 则不行)。

注意:这些计算也可以在 Tableau Prep Builder 中创建。有关 Prep 中 LOD 表达式的详细信息,请参见**创建详细级别、排名和分片计算** 在本页259。

3. 现在,我们可以按如下方式创建 DATEDIFF 计算:

   Time Between Infractions(违章间隔时间) = `DATEDIFF('day', [1st Infraction Date], [2nd Infraction Date])`

   - 如果我们想要查看周或月,只需修改日期部分(当前为 'day')。
   - 也可以通过将 FIXED 计算直接放在 DATEDIFF 中为整个事件创建单一计算:

     ```
     DATEDIFF ( 'day',
     { FIXED [Driver ID] : MIN ( IF [Infraction Number] = "1st" THEN [Infraction Date] END ) },
     { FIXED [Driver ID] : MIN ( IF [Infraction Number] = "2nd" THEN [Infraction Date] END ) }
     )
     ```

4. 在“列”上绘制“**Time Between Infractions**”(违章间隔时间),并在“行”上绘制“**Driver ID**”(司机 ID)。

结果将与使用未转置数据结构的结果相同。

2.比较第一次和第二次违章的罚款金额。这些金额是否相关?

A. 为了在 Tableau Desktop 中回答此问题,我们将使用与上一个问题非常相似的逻辑。我们将使用“**Infraction Number**”(违章次数)来确定给定行是第一次违章还是第二次违章,然后相应拉出罚款金额。

1. 如果我们只需要建立散点图,则可以跳过 LOD 部分,只使用 `IF` 计算:

   1st Fine Amount(第一次罚款金额) = `IF [Infraction Number] = "1st" THEN [Fine Amount] END`

   2nd Fine Amount(第二次罚款金额) = `IF [Infraction Number] = "2nd" THEN [Fine Amount] END`

2. 但是,如果我们想要进行比较并查看某个司机第一次和第二次罚款之间的金额差异,我们会遇到像日期一样的 null 问题。将这些计算封装在 `FIXED` LOD 内不会有什么坏处,因此从一开始这样做或许会不错:

   1st Fine Amount(第一次罚款金额) = `{ FIXED [Driver ID] : MIN ( IF [Infraction Number] = "1st" THEN [Fine Amount] END ) }`

   2nd Fine Amount(第二次罚款金额) = `{ FIXED [Driver ID] : MIN ( IF [Infraction Number] = "2nd" THEN [Fine Amount] END ) }`

   这些计算也可以在 Tableau Prep Builder 中创建。有关 Prep 中 LOD 表达式的详细信息,请参见**创建详细级别、排名和分片计算** 在本页259。

3. 在“列”上创建包含“**1st Fine Amount**”(第一次罚款金额)的散点图,在“行”上创建包含“**2nd Fine Amount**”(第二次罚款金额)的散点图,并像以前一样显示线性趋势线。

结果将与使用未转置数据结构的结果相同。

3.总体而言,哪个司机支付的罚款最多?谁支付的最少?

A. 为了在 Tableau Desktop 中回答此问题,转置的数据结构非常理想。我们只需在条形图中显示“**Driver ID**”(司机 ID)和“**Fine Amount**”(罚款金额)。默认聚合已经是 **SUM**,因此将自动绘制司机支付的总金额。

结果将与使用未转置数据结构的结果相同。

4.有多少司机有过多种类型的违章?

A. 为了在 Tableau Desktop 中回答此问题，转置的数据结构非常理想。我们只需以条形图形式显示“**Driver ID**”(司机 ID) 以及“**Infraction Type**”(违章类型) 的**“不重复计数”**，就能得到答案。

结果将与使用未转置数据结构的结果相同。

5.从未上过交通法规学习班的司机的平均罚款金额是多少？

A. 为了在 Tableau Desktop 中回答此问题，我们无法简单地将总罚款金额除以二，因为某些司机只有一次违章。我们也无法计算每个司机的平均罚款并取这些值的平均值，因为对平均值进行平均计算可能会导致不一致。作为替代，我们需要计算从未上过交通法规学习班的司机支付的总金额，然后除以与这些罚款关联的总违章次数。

1. 首先，我们需要确定是否每个司机都有第二次违章。如果没有第二次违章，则“**2nd Infraction Date**”(第二次违章日期) 将为 null，我们可以利用这一事实并开始构建计算：

   ```
   IFNULL(STR([2nd Infraction Date]), 'no')
   ```

   这将返回第二次违章的日期(如果存在)，或者如果没有第二次违章，则返回“no”。

   > **注意：**此计算的 STR 部分是必需的，因为 IFNULL 需要在其参数中保持数据类型的一致性。由于我们想要为 null 值返回字符串“no”，因此也要将日期转换为字符串。

2. 接着，我们需要将此信息转换为违章次数 1 或 2。如果 IFNULL 计算的结果为“no”，则应将司机标记为有一次罚款。对于任何其他结果，则应标记为有两次罚款。计算为：

   Number of Infractions(违章次数) =

   ```
   IF IFNULL(STR([2nd Infraction Date]), 'no')= 'no' THEN 1
   ELSE 2
   END
   ```

3. 现在，我们需要考虑平均罚款金额。我们已经有了一个“**Fine Amount**”（罚款金额）字段。我们只需将该字段除以新的“**Number of Infractions**”（违章次数）字段，同时将两者都放在 **SUM** 内：

   Average Fine（平均罚款）= `(SUM([Fine Amount]) / SUM([Number of Infractions])`

B. 我们还需要筛选出参加过交通法规学习班的司机。看起来我们可以使用**“Traffic School”（交通法规学习班）**字段并依据**“Traffic School = no”**进行筛选。但是，这将依据未与交通法规学习班关联的*违章行为*（而不是从示参加过交通法规学习班*司机*）进行筛选。如果一名司机因一项违章行为（而不是另一项违章行为）而参加交通法规学习班，我们不希望在这里考虑任何违章行为 - 该司机已经参加过交通法规学习班，因此不符合问题的参数。

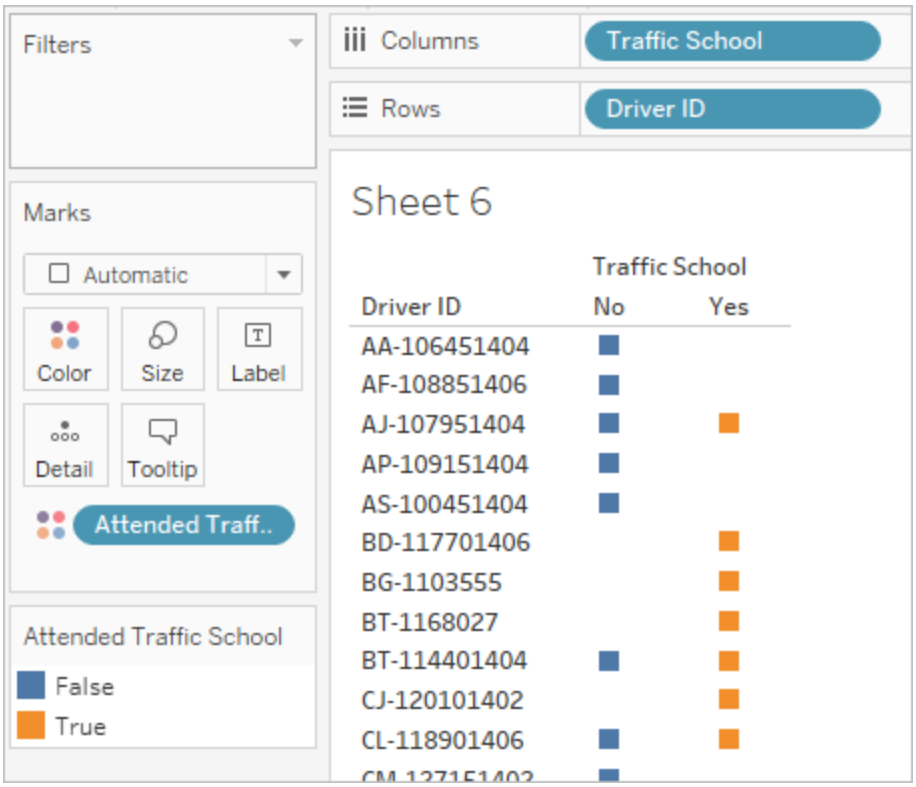
   我们需要做的是筛选出参加过交通法规学习班的任何司机。就数据而言，我们想要筛选出任何行上“**Traffic School**”（交通法规学习班）为“Yes”（是）的任何司机，而不管它与哪种违章行为关联。让我们分阶段构建计算，并使用一个简单的视图来帮助跟踪所发生的情况：

   1. 首先，我们想要知道司机的“Traffic School”（交通法规学习班）是否为“Yes”（是）。将“**Driver ID**”（司机 ID）拖到**“行”**，并将“**Traffic School**”（交通法规学习班）拖到**“列”**。我们将获得一个文本表，其中占位符“Abc”文本表示每个司机的相关值。

   2. 接着，我们想要构建一个计算，该计算将确定“**Traffic School**”（交通法规学习班）的值是否为“Yes”（是）。计算的第一个阶段是：

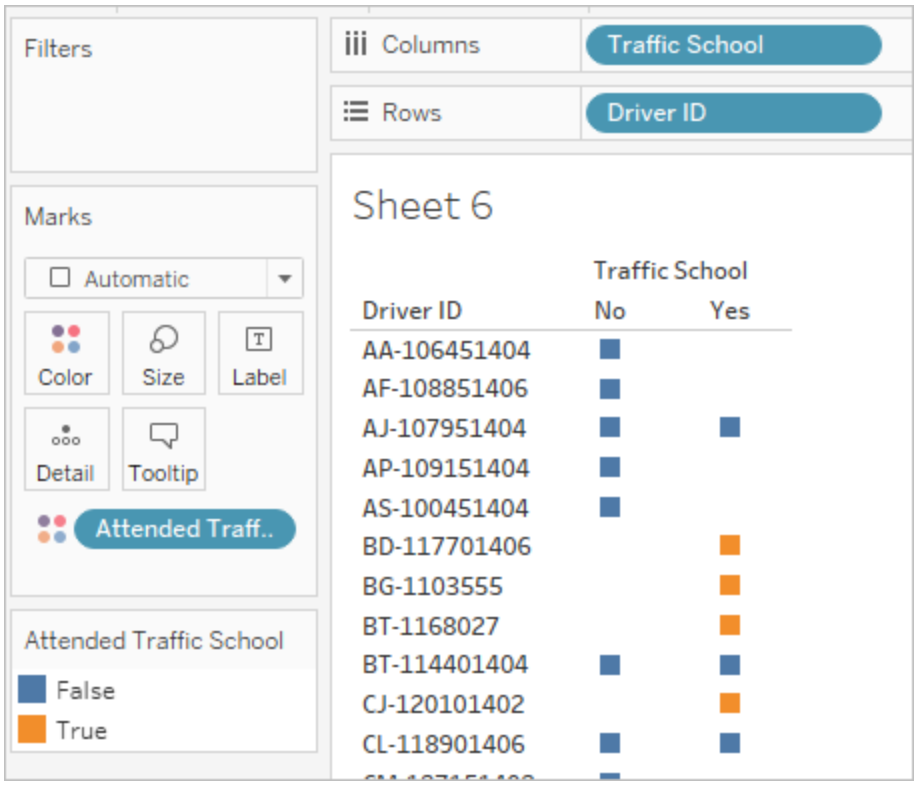
      Attended Traffic School（已参加交通法规学习班）= `CONTAINS([Traffic School), 'Yes')`

      如果我们将“**Attended Traffic School**”（已参加交通法规学习班）拖到**“标记”**卡上的**“颜色”**功能区，我们会看到它准确地针对“No”（否）列中的每个标记标为“False”，并针对“Yes”（是）列中的每个标记标为“True”。



3. 但是，我们真正需要的是位于*司机*（而不是*违章*）级别的此信息。尝试在与基本数据结构不同的详细级别计算结果时，LOD 表达式非常适合。我们将此表达式设为 `FIXED` LOD 表达式。但是，正如我们所知，必须对 LOD 的聚合表达式进行聚合。以前，我们使用过 **MIN**。它在这里是否有用？ 我们将计算修改为：

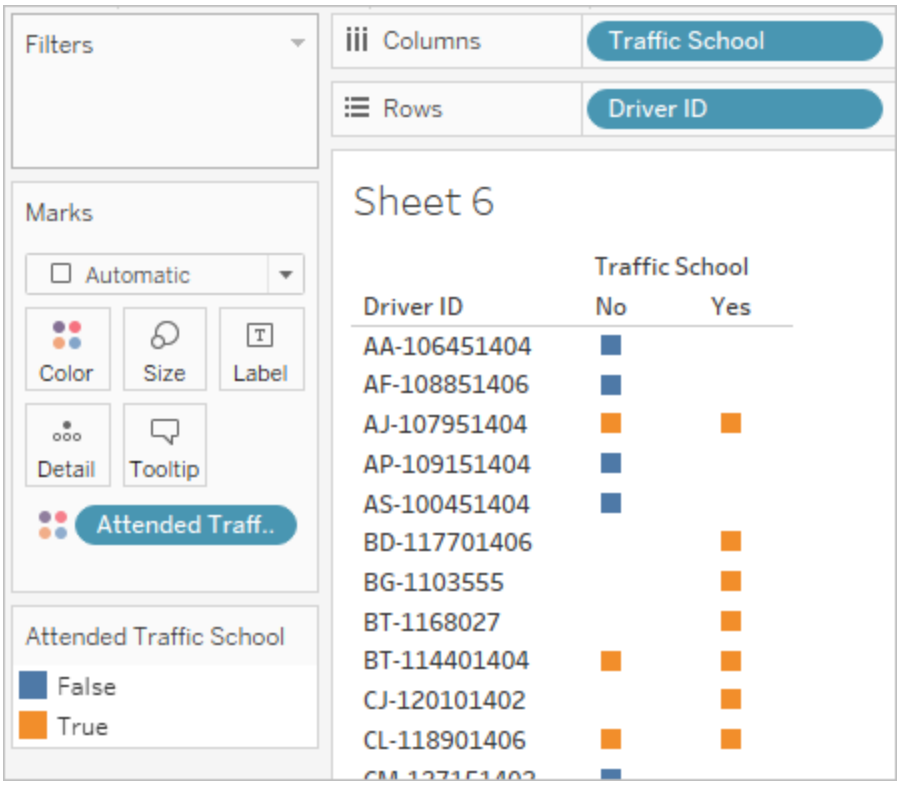
   Attended Traffic School（已参加交通法规学习班）= `{ FIXED [Driver ID] : MIN( CONTAINS([Traffic School], 'Yes'))}`



在视图中应用该更改之后，我们看到了与预期相反的结果。具有“No”值的任何司机全部被标记为“False”。相反，我们希望为该司机的每条记录将“Yes”标记为“True”。MIN 在这里做了什么？它按字母顺序选取第一个响应，即“No”。

4. 如果我们将其更改为 **MAX**，将会怎么样？该函数是否会按字母顺序获取最后一个响应？我们将计算修改为：

   Attended Traffic School（已参加交通法规学习班）= `{ FIXED [Driver ID] : MAX ( CONTAINS( [Traffic School], 'Yes') ) }`



我们得到了想要的结果：如果司机在数据中的任何位置有“Yes”，他们会因为参加过交通法规学习班而被标记为“True”，即使对于不涉及交通法规学习班的违章也是如此。

5. 如果我们将“**Attended Traffic School**”（已参加交通法规学习班）拖到**“筛选器”**功能区并仅选择“False”，将只会剩下从未参加过交通法规学习班的司机。

C. 为了回答原来的问题，在筛选器就位的情况下，我们只需将“**Average Fine**”（平均罚款）放到**“标记”**卡上的**“文本”**功能区。由于我们在计算中构建了聚合，因此字段上的聚合将为 **AGG**，我们无法对其进行更改。这是预期行为。

结果将与使用未转置数据结构的结果相同。

转置数据的优点

如果知道我们只需要回答可以使用教程中的原始数据结构轻松回答的问题，我们可以继续使用该结构。不过，转置的数据格式更加灵活。即使它需要一些计算，但在计算就位之后，生成的数据集将非常适合于回答更广泛的问题。

再进一步 – 只使用计算

如果您无法访问 Tableau Prep Builder，将会怎么样？ 如果被迫使用原始数据，您是否完全不走运？ 根本不会！

Tableau Desktop 和 LOD 表达式可以回答所有分析问题。如果我们连接到原始的 **Traffic Violations.xlsx**，它看起与已转置数据集非常相似 — 只是没有重要的“**Infraction Number**”(违章次数)字段。我们将需要通过 LOD 表达式模拟聚合步骤的输出。

> **注意：**您可以下载工作簿 LOD Driver Infractions.twbx，在上下文中探讨解决方案。请记住，可以通过一些替代方式来解读分析或寻找答案。

1.每个司机第一次和第二次违章相隔多长时间(以天数为单位)？

A. 为了在 Tableau Desktop 中回答此问题，我们将再次使用 DATEDIFF 函数。此函数需要一个开始日期和一个结束日期。此信息存在于我们的数据中，但全部在一个字段中。我们需要将其拉出到两个字段中。由于我们想要确保这两个字段均可用于为每个司机比较，我们需要将它们固定到“**Driver ID**”(司机 ID)级别。

1. 为了查找第一次违章日期，我们将使用计算：

   1st Infraction(第一次违章) = `{ FIXED [Driver ID] : MIN ( [Infraction Date] ) }`

2. 我们将分阶段执行第二次违章日期相关操作。

   a. 首先，我们需要只查看比第一个日期大的日期：

   ```
   IF [Infraction Date] > [1st Infraction] THEN [Infraction Date] END
   ```

   b. 但这将列出第一次违章之后*每次*违章，而我们只需要第二次违章。因此我们需要这些日期中最小的日期。将整个计算放在 **MIN** 内：

   ```
   MIN( IF [Infraction] : [1st Infraction] THEN [Infraction Date] END )
   ```

   c. 我们还想要为每个司机重新计算第二次违章日期。这就是 LOD 表达式的用武之地。我们会将此计算固定到“**Driver ID**”(司机 ID)级别：

2nd Infraction(第二次违章) = `{ FIXED [Driver ID] : MIN ( IF [Infraction Date] > [1st Infraction] THEN [Infraction Date] END ) }`

3. 现在我们可以创建 DATEDIFF 计算:

   Time Between Infractions(违章间隔时间) = `DATEDIFF('day', [1st Infraction], [2nd Infraction])`

结果将与使用其他两个数据结构的结果相同。

2.比较第一次和第二次违章的罚款金额。这些金额是否相关?

A. 为了在 Tableau Desktop 中回答此问题,我们将使用此问题的转置数据版本相似的逻辑。我们将使用为问题 I 创建的"**1st Infraction**"(第一次违章)和"**2nd Infraction**"(第二次违章)来确定给定行是第一次违章还是第二次违章,然后相应拉出罚款金额。

   1. 如果我们只需要建立散点图,则可以跳过 LOD 部分,只使用 IF 计算:

      1st Fine Amount(第一次罚款金额) = `IF [1st Infraction] = [Infraction Date] THEN [Fine Amount] END`

      2nd Fine Amount(第二次罚款金额) = `IF [2nd Infraction] = [Infraction Date] THEN [Fine Amount] END`

   2. 但是,如果我们想要进行比较并查看某个司机第一次和第二次罚款之间的金额差异,我们会遇到像第一个数据结构中一样的 null 问题。将这些计算封装在 FIXED LOD 内不会有什么坏处,因此从一开始这样做或许会不错:

      1st Fine Amount(第一次罚款金额) = `{ FIXED [Driver ID] : MIN ( IF [1st Infraction] = [Infraction Date] THEN [Fine Amount] END ) }`

      2nd Fine Amount(第二次罚款金额) = `{FIXED [Driver ID] : MIN( IF [2ndInfraction] = [Infraction Date] THEN [Fine Amount] END ) }`

结果将与使用其他两个数据结构的结果相同。

3.总体而言,哪个司机支付的罚款最多? 谁支付的最少?

A. 为了在 Tableau Desktop 中回答此问题，我们需要了解有关仅 LOD 方法的一些信息。使用 Tableau Prep 的两种方法都会为司机筛选出不是第一次或第二次违章的记录。Tableau Desktop 中的 LOD 方法会保留所有记录。这意味着，如果我们按“**Driver ID**”(司机 ID) 创建“**SUM(Amount Paid)**”的可视化项，仅 Tableau Desktop 版本将为违章超过两次的司机显示更高的金额。若要从与其他方法匹配的完整数据中获取“**Total Amount Paid**”(总支付金额) 值，而不是使用原始“**Fine Amount**”(罚款金额) 字段，则我们需要像使用第一个数据结构所做的那样将第一次和第二次罚款加总。

B. 使用我们为问题 2 创建的字段，我们将添加两个罚款金额。ZN 是必要的，以防止任何只有一次违章的司机的结果为 null。计算为：

   Total Amount Paid(总支付金额) = `[1st Fine Amount] + ZN([2nd Fine Amount])`

结果将与使用其他两个数据结构的结果相同。

4.有多少司机有过多种类型的违章？

A. 为了在 Tableau Desktop 中回答此问题，我们不能简单地显示“**Driver ID**”(司机 ID) 和“**Infraction Type**”(违章类型) 的**不重复计数**。由于此数据集有超过第二次的违章，因此某些司机可能有两种以上的违章类型。为了使结果与其他方法相符，我们需要将范围限制为仅包含前两次违章。

B. 我们可以拉出第一次和第二次违章类型，将这些类型放在 LOD 表达式中，使它们 FIXED (固定) 到司机级别，然后使用 IF 计算来对类型计数：

   1. **1st Infraction Type**(第一次违章类型) = `{ FIXED [Driver ID] : MIN ( IF [1st Infraction] = [Infraction Date] THEN [Infraction Type] END ) }`
   2. **2nd Infraction Type**(第二次违章类型) = `{ FIXED [Driver ID] : MIN ( IF [2nd Infraction] = [Infraction Date] THEN [Infraction Type] END ) }`
   3. **Number of Infraction Types**(违章类型数量) =

      ```
      IF [1st Infraction Type] = [2nd Infraction Type] THEN 1
      ELSEIF [1st Infraction Type] != [2nd Infraction Type] THEN 2
      ELSE 1 END
      ```

注意：也可以通过直接将初始计算嵌入在较大的计算中，以单一字段的形式创建多个这样的计算。此处的合并计算将如下所示：

```
IF
 {FIXED [Driver ID] : MIN(IF [1st Infraction]=
[Infraction Date] THEN [Infraction Type] END)}
 =
 {FIXED [Driver ID] : MIN(IF [2nd Infraction]=
[Infraction Date] THEN [Infraction Type] END)}
THEN 1

ELSEIF
 {FIXED [Driver ID] : MIN(IF [1st Infraction]=
[Infraction Date] THEN [Infraction Type] END)}
 !=
 {FIXED [Driver ID] : MIN(IF [2nd Infraction]=
[Infraction Date] THEN [Infraction Type] END)}
THEN 2

ELSE 1
END
```

理解起来有点困难，但如果愿意使用的话，它确实有用。（请注意，换行符和一些空格不会影响 Tableau 解读计算的方式。）

A. 然后，我们可以依据“**Driver ID**”（司机 ID）绘制“**Number of Infraction Types**”（违章类型数量）条形图并对其进行排序。

结果将与使用其他两个数据结构的结果相同。

5.从未上过交通法规学习班的司机的平均罚款金额是多少？

A. 为了在 Tableau Desktop 中回答此问题，我们无法简单地将总罚款金额除以二，因为某些司机只有一次违章。我们也无法计算每个司机的平均罚款并取这些值的平均值，因为对平均值进行平均计算可能会导致不一致。作为替代，我们需要计算从

未上过交通法规学习班的司机支付的总金额，然后除以与这些罚款关联的总违章次数。

1. 首先，我们需要确定是否每个司机都有第二次违章。如果没有第二次违章，则所有“2nd”字段中的信息都将为 null，我们可以利用这一事实并开始构建计算：

   ```
   IFNULL([2nd Infraction Type], 'no')
   ```

   这将返回违章类型（如果存在），或者如果没有第二次违章，则返回“no”。

2. 接着，我们需要将此信息转换为违章次数 1 或 2。如果 IFNULL 计算的结果为“no”，则应将司机标记为有一次罚款。对于任何其他结果，则应标记为有两次罚款。计算为：

   Number of Infractions（违章次数）=

   ```
   IF IFNULL([2nd Infraction Type], 'no') = 'no' THEN 1
   ELSE 2
   END
   ```

3. 对于“Total Amount Paid”（总支付金额），我们可以使用问题 3 中的计算。总而言之，我们将利用这个总罚款金额，并将其除以新的“**Number of Infractions**”（违章次数）计算字段，从而确定平均罚款金额：

   Average Fine（平均罚款）= `SUM([Total Amount Paid]) / SUM([Number of Infractions])`

B. 我们还需要筛选出参加过交通法规学习班的司机。由于此数据集包含一些有第三次或第四次违章的司机，因此我们无法使用与转置数据结构相同的方法。相反，我们将采用与未转置数据相同的方法，总结如下：

1. 首先，我们需要构建两个计算，确定第一次和第二次违章是否涉及交通法规学习班：

   1st Traffic School（第一次交通法规学习班）= `{ FIXED [Driver ID] : MIN (IF [1st Infraction] = [Infraction Date] THEN [Traffic School] END ) }`

   2nd Traffic School（第二次交通法规学习班）= `{ FIXED [Driver ID] : MIN (IF [2nd Infraction] = [Infraction Date] THEN [Traffic School] END ) }`

2. 然后，我们将添加这些值来获取参加交通法规学习班的总次数：

   Number of Traffic School Attendances（参加交通法规学习班次数）=

```
(CASE [1st Traffic School] WHEN 'Yes' THEN 1 WHEN 'No' THEN
0 ELSE 0 END)
+
(CASE [2nd Traffic School] WHEN 'Yes' THEN 1 WHEN 'No' THEN
0 ELSE 0 END)
```

3. 如果我们将“**Number of Traffic School Attendances**”(参加交通法规学习班次数)拖到**“数据”**窗格的**“维度”**区域，则值 0–2 将变为离散。
4. 现在，如果我们对“**Number of Traffic School Attendances**”(参加交通法规学习班次数)进行筛选，我们可以只选择 0，并且知道我们获得了从未参加过交通法规学习班的司机。

C. 为了回答原来的问题，我们只需将“**Average Fine**”(平均罚款)放到**“标记”**卡上的**“文本”**功能区。由于我们在计算中构建了聚合，因此字段上的聚合将为 **AGG**，我们无法对其进行更改。这是预期行为。

结果将与使用其他两个数据结构的结果相同。

请务必记住，此解决方案包含大量的嵌套计算和 LOD 表达式。视数据集的大小以及数据集的复杂性而定，可能会出现性能问题。

对各种方法的反思

那么您应该采用哪种方法呢？这完全取决于您，并且工具任由您随意使用。

- 如果您想要避开 LOD，则可以使用数据调整解决方案，尽管对于某些分析(**Tableau Desktop 中的分析** 在本页480)计算可能是必需的。
- 如果您能够调整数据并且熟悉计算(包括 LOD)，则中间型选项可以提供最好的灵活性(**更进一步 — 转置的数据** 在本页487)。
- 如果您熟悉 LOD、对性能的影响较小，并且/或者您无法访问 Tableau Prep，则单独使用 LOD 解决此问题是可行的选项(**再进一步 — 只使用计算** 在本页497)。

至少，它对于了解 Tableau Prep 中的聚合和 Tableau Desktop 中的详细级别表达式如何相互关联以及会对数据分析产生怎样的影响是有价值的。如同 Tableau 中的大多数功能一样，执行任何操作的方法都有多种。探索各种各样的选项可以帮助将概念联系在一起，让您能挑选最适合于自己的解决方案。

使用的计算：

Driver Infractions(司机违章)

- **Time Between Infractions**(违章间隔时间) = `DATEDIFF('day', [1st Infraction Date], [2nd Infraction Date])`
- **Total Amount Paid**(总支付金额) = `[1st Fine Amount] + ZN([2nd Fine Amount])`
- **Number of Infraction Types**(违章类型数量) = `IF [1st Infraction Type]=[2nd Infraction Type] THEN 1 ELSEIF [1st Infraction Type]!= [2nd Infraction Type] THEN 2 ELSE 1 END`
- **Number of Infractions**(违章次数) = `IF IFNULL([2nd Infraction Type], 'no') = 'no' THEN 1 ELSE 2 END`
- **Average Fine**(平均罚款) = `( SUM([1st Fine Amount]) + SUM( ZN([2nd Fine Amount]) ) ) / SUM([Number of Infractions])`
- **Number of Traffic School Attendances**(参加交通法规学习班次数) = `(CASE [1st Traffic School] WHEN 'Yes' THEN 1 WHEN 'No' THEN 0 ELSE 0 END) + (CASE [2nd Traffic School] WHEN 'Yes' THEN 1 WHEN 'No' THEN 0 ELSE 0 END)`

Pivoted Driver Infractions(转置的司机违章)

- **1st Infraction**(第一次违章) = `{FIXED [Driver ID] : MIN(IF [Infraction Number] = "1st" THEN [Infraction Date] END)}`
- **2nd Infraction**(第二次违章) = `{FIXED [Driver ID] : MIN(IF [Infraction Number] = "2nd" THEN [Infraction Date] END)}`
- **Time Between Infractions**(违章间隔时间) = `DATEDIFF('day', [1st Infraction], [2nd Infraction])`
- **1st Fine Amount**(第一次罚款金额) = `{FIXED [Driver ID] : MIN( IF [Infraction Number] = "1st" THEN [Fine Amount] END ) }`
- **Number of Infractions**(违章次数) = `IF IFNULL(STR([2nd Infraction]), 'no')= 'no' THEN 1 ELSE 2 END`
- **Average Fine**(平均罚款) = `SUM([Fine Amount])/SUM([Number of Infractions])`
- **Attended Traffic School**(已参加交通法规学习班) = `{ FIXED [Driver ID] : MAX( CONTAINS([Traffic School], 'Yes'))}`

LOD Driver Infractions(LOD 司机违章)

- **1st Infraction**(第一次违章) = `{FIXED [Driver ID] : MIN([Infraction Date])}`
- **2nd Infraction**(第二次违章) = `{ FIXED [Driver ID] : MIN( IF [Infraction Date] > [1st Infraction] THEN [Infraction Date] END ) }`
- **Time Between Infractions**(违章间隔时间) = `DATEDIFF('day', [1st Infraction], [2nd Infraction])`
- **1st Fine Amount**(第一次罚款金额) = `{FIXED [Driver ID] : MIN( IF [1st Infraction] = [Infraction Date] THEN [Fine Amount] END ) }`
- **2nd Fine Amount**(第二次罚款金额) = `{FIXED [Driver ID] : MIN( IF [2nd Infraction] = [Infraction Date] THEN [Fine Amount] END ) }`
- **Total Amount Paid**(总支付金额) = `[1st Fine Amount] + ZN([2nd Fine Amount])`
- **1st Infraction Type**(第一次违章类型) = `{FIXED [Driver ID] : MIN( IF [1st Infraction] = [Infraction Date] THEN [Infraction Type] END ) }`
- **2nd Infraction Type**(第二次违章类型) = `{FIXED [Driver ID] : MIN( IF [2nd Infraction] = [Infraction Date] THEN [Infraction Type] END ) }`
- **Number of Infraction Types**(违章类型数量) = `IF [1st Infraction Type]=[2nd Infraction Type] THEN 1 ELSEIF [1st Infraction Type]!= [2nd Infraction Type] THEN 2 ELSE 1 END`
- **Number of Infractions**(违章次数) = `IF IFNULL([2nd Infraction Type], 'no') = 'no' THEN 1 ELSE 2 END`
- **Average Fine**(平均罚款) = `SUM ([Total Amount Paid]) / SUM([Number of Infractions])`
- **1st Traffic School**(第一次交通法规学习班) = `{FIXED [Driver ID] : MIN (IF [1st Infraction] = [Infraction Date] THEN [Traffic School] END ) }`
- **2nd Traffic School**(第二次交通法规学习班) = `{FIXED [Driver ID] : MIN (IF [2nd Infraction] = [Infraction Date] THEN [Traffic School] END ) }`

- **Number of Traffic School Attendances**(参加交通法规学习班次数) = `(CASE [1st Traffic School] WHEN 'Yes' THEN 1 WHEN 'No' THEN 0 ELSE 0 END) + (CASE [2nd Traffic School] WHEN 'Yes' THEN 1 WHEN 'No' THEN 0 ELSE 0 END)`

注意：特别感谢 Ann Jackson 的 Workout Wednesday 主题 Do Customers Spend More on Their First or Second Purchase?(客户在第一次或第二次购买时是否花费更多？)以及 Andy Kriebel 的 Tableau Prep 技巧 Returning the First and Second Purchase Dates (返回第一次和第二次购买日期)，这些文章为本教程提供了最初的灵感。单击这些链接会使您离开 Tableau 网站。Tableau 对外部提供商所维护的页面的准确性或新鲜度不担负任何责任。如果您对其内容有任何疑问，请与所有者联系。

Tableau Prep Builder 疑难解答

本文列出了您在使用 Tableau Prep Builder 时可能遇到的问题以及有关如何解决这些问题的建议。

运行 LogShark

LogShark 是一个免费的开源命令行实用程序，您可以使用它从 Prep 日志文件中提取信息以进行故障排除并深入了解错误和使用情况。使用 LogShark `Prep.twbx` 插件，您可以生成带有错误和流程仪表板的工作簿，以帮助您分析和可视化 Prep 问题。

LogShark 要求您处理的 Prep 日志文件是压缩 (zipped) 文件。若要查找 Prep 日志文件，请导航到**“我的 Tableau Prep 存储库”**文件夹。位置为 `/Users/<username>/Documents/My Tableau Prep Repository`。

有关安装和运行 LogShark 的信息，请参见为您的计算机设置 LogShark 。

使用命令行运行流程时的常见错误

您可以从命令行中以编程方式运行流程来刷新输出文件，而不是打开 Tableau Prep Builder 以手动方式运行每个流程。尽管此过程可帮助提高流程进程的效率，但如果语法不正确，或者缺少连接或输出位置的凭据，则在运行此过程时会出现错误。

下表描述了常见错误以及如何解决这些错误。有关如何从命令行中运行流程的信息，请参见**通过命令行刷新流程输出文件** 在本页397。

错误	原因	如何修复
“Missing arguments”(缺少参数)	缺少必需的命令行参数之一。	使用“tableau-prep-cli -help”来查看命令行参数的列表。
“Unable to read the connections file.”(无法读取连接文件。)	输入连接的 credentials.json 文件中的语法或格式存在错误。	请检查 .json 文件中输入连接的语法。有关详细信息和示例，请参见**通过命令行**

		刷新流程输出文件 在本页397。
“There are errors in the flow.Unable to run the flow. Check that the credentials .json file includes all required credentials.Open the flow in Tableau Prep Builder to view error details.”(流程中存在错误。无法运行流程。请检查 credentials .json 文件是否包括所有必需的凭据。在 Tableau Prep Builder 中打开流程以查看错误详细信息。)	输入连接的 credentials.json 文件中缺少凭据，或者流程包含错误。	请检查 .json 文件是否具有所有连接的凭据，并在 Tableau Prep Builder 中打开流程文件，确定流程中是否存在任何错误。 如果流程有错误，您必须修复这些错误并将流程重新发布到 Tableau Server，然后再次尝试运行该过程。
“Could not find match for <hostname of inputConnections >”(找不到 <输入连接的主机名> 的匹配项)	credentials.json 文件缺少主机名 (服务器名称) 条目。	请确保 credentials.json 文件包括主机名 (服务器名称) 的正确凭据。 有关详细信息和示例，请参见**通过命令行刷新流程输出文件** 在本页397。
“We don't have credentials of all connections in tfl/tflx file.The following connection(s) were not found: <hostname of inputConnections>”(tfl/tflx 文件中没有所有连接的凭据。找不到以下连接 :<输入连接的主机名>)	credentials.json 文件缺失，或包含的凭据对于错误消息中显示的主机名 (服务器名称) 不正确。	请确保 credentials.json 文件包括错误消息中所列主机名 (服务器名称) 的正确凭据。

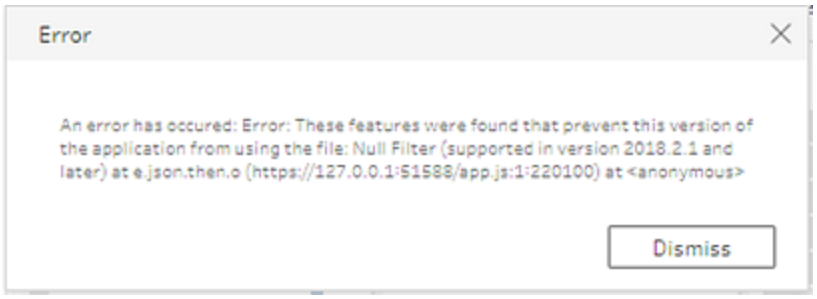
		有关详细信息和示例，请参见**通过命令行刷新流程输出文件** 在本页397。
“Error signing in server <serverUrl> as a user <userName>.Please check the credentials.”(以用户 <用户名> 身份登录到服务器 <服务器 URL> 时出错。请检查凭据。)	credentials.json 文件包含的凭据对于 Tableau Server 不正确。	请确保 credentials.json 文件包括适用于输出连接的所有正确凭据和元素。 有关详细信息和示例，请参见**通过命令行刷新流程输出文件** 在本页397。
“Could not sign in successfully as <userName> to server <serverUrl>(<contentUrl>)”(无法以 <用户名> 身份成功登录到服务器 <服务器 URL>(<内容 URL>))	credentials.json 文件包含的凭据对于 Tableau Server 不正确。	请确保 credentials.json 文件包括适用于输出连接的所有正确凭据和元素。 有关详细信息和示例，请参见**通过命令行刷新流程输出文件** 在本页397。
“We don't have credentials for Tableau Server to publish extract for one or more output nodes in tfl/tflx file.”(tfl/tflx 文件中没有用于为一个或多个输出节点发布数据提取的 Tableau Server 凭据。)	credentials.json 文件未作为命令行参数传入，或者其缺少输出连接的凭据。	请确保在命令行中包括 credentials.json 文件的路径，并验证 credentials.json

		文件是否包括适用于输出连接的所有正确凭据和元素。 有关详细信息和示例，请参见**通过命令行刷新流程输出文件** 在本页397。
“Loom rest api server not started”(Loom rest api 服务器未启动)	安装或环境设置不正确。	请确保正确安装了 Tableau Prep Builder，并且您以管理员身份运行命令。 有关如何安装 Tableau Prep Builder 的信息，请参见通过用户界面安装 Tableau Desktop 或 Tableau Prep Builder。
“Error.Flow file does not exist.”(错误。流程文件不存在。)	流程文件的路径不正确。	请确保在命令行中包括流程文件的正确路径。
“Error.Connections file does not exist.”(错误。连接文件不存在。)	credentials.json 文件的路径不正确。	请确保在命令行中包括 credentials.json 文件的正确路径。
“Could not find match for <mapr01:5181>,<mapr02:5181>,<mapr03:5181>”	使用 ZooKeeper 连接到 Apache	在命令行中包括为输入凭据指定

(找不到 <mapr01:5181>、<mapr02:5181>、<mapr03:5181> 的匹配项)	Drill 时，您必须指定具体端口 ID。	“端口”31010 的 credentials.json 文件。

错误：“These features were found that prevent this version of the application from using this file”(发现了使此版本的应用程序无法使用此文件的这些功能)

如果在较早版本的 Tableau Prep Builder 中打开在版本 2018.2.1 或更高版本中创建的流程，您可能会看到以下错误：



包括较早版本中不支持的功能的流程将导致此兼容性错误。为了解决错误，请在更高版本中打开流程，并保存没有所指明功能的流程的副本。在上面的示例中，请从应用了 null 筛选器的字段中将该筛选器移除。

然后在较早版本的 Tableau Prep Builder 中打开已移除该功能的副本。

使用 Tableau Prep 登录到启用 SSL 的 Tableau Server 时出现“You are using Server version: null...”(您正使用 Server 版本: null...)错误

从 Tableau Prep Builder 中登录到启用 SSL 的 Tableau Server 时，您必须在安装了 Tableau Prep Builder 的计算机上安装根证书。如果未安装证书，您可能会看到以下错误：

> **You are using Server version: null but the minimum compatible version is: 10.0. Please upgrade to a compatible version(您正使用 Server 版本: null，但兼容的最低版本为: 10.0。请升级到兼容版本)**

如果看到此错误，请与 IT 部门或系统管理员合作，在安装了 Tableau Prep Builder 的计算机上安装所需的根证书。有关详细信息，请参见 Tableau Desktop 和 Tableau Prep Builder 部署指南中的系统要求。

维护 Tableau Desktop 和 Tableau Prep 的许可证

Tableau Desktop 和 Tableau Prep Builder 可依据短期许可证模型获得许可。但是，当您购买新的 Tableau Server 或新的 Tableau Cloud 订阅时，将不再为 Tableau Desktop 或 Tableau Prep Builder 颁发产品密钥。作为替代，请使用 基于登录名的许可证管理 来激活并登录到 Tableau Server 或 Tableau Cloud。有关详细信息，请参见使用基于登录名的许可证管理激活 Tableau。

必须续订短期许可证并刷新产品密钥才能提供不中断的服务。您可以在每个指定的期间过期时连续续订短期许可证。如果未续订短期许可证并且期限过期，Tableau 将停止工作，您将不再能访问软件。有关续订许可证的详细信息，请参见如何预订 Tableau 许可证。

> **注意：**Tableau Desktop 或 Tableau Prep 的试用许可证会在一段设定的时间(通常为 14 天)后过期。试用期过期后，您将需要购买许可证才能继续使用产品。

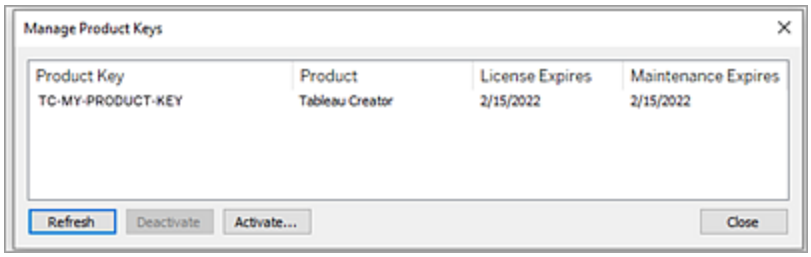
查看有关许可证的数据

安装 Tableau Desktop 或 Tableau Prep 之后，打开应用程序，然后从顶部菜单中导航到**“帮助”>“管理产品密钥”**，以查看有关您拥有的许可证类型及其何时过期的信息。

您也可以通过此对话框激活或停用产品密钥，或者，如果未使用“虚拟桌面 (ATR)”选项，则通过此对话框刷新维护产品密钥。

注意：Tableau 提供具备一定范围能力的短期许可证。您拥有的许可证的类型显示在**“产品”**字段中。有关不同类型基于用户的可用许可证的详细信息，请参见 Tableau Server 帮助中的基于用户的许可证。

现有 Tableau Desktop 用户可能拥有永久许可证。永久许可证不会过期，其**“管理产品密钥”**对话框中的**“许可证过期”**字段会显示“永久”。但是，若要访问产品更新和技术支持，您必须购买支持和维护服务。这些服务必须续订才能继续获得服务。不再为 Tableau Desktop 提供永久许可证。



使用以下按钮来对产品密钥执行操作：

- **刷新**（仅限非基于登录名的许可证管理或非虚拟桌面）：单击**“刷新”**按钮以刷新将要过期的维护许可证，然后关闭并重新启动 Tableau Desktop。如果**“维护过期”**日期未更新，请咨询许可证管理员，因为密钥或维护协议可能已更改。

  “许可证过期”值列为“永久”的产品密钥（如上面的“管理产品密钥”对话框所示）是旧产品密钥。只要 Tableau 客户门户中列出的维护结束日期迟于“Desktop 管理产品密钥”对话框中反映的日期，您就可以随时刷新“永久”产品密钥。

  如果产品密钥已达到过期日期（非永久产品密钥），则无法刷新产品密钥。访问 Tableau 客户门户以获取更新的订阅产品密钥并执行新的激活。如果产品密钥尚未达到过期日期，您可以刷新产品密钥。刷新尚未过期的产品密钥时，只有“许可证过期”值将更改，产品密钥不会更改。产品密钥将在达到过期日期时更改。

若要通过命令行刷新维护密钥，请参见 Tableau Desktop 和 Tableau Prep 部署指南中的刷新产品密钥。

> **注意：**如果 Tableau Desktop 处于脱机状态，则无法刷新产品密钥。如果在脱机模式下激活 Tableau Desktop，则必须从 Tableau 客户门户获取并激活新密钥。

- **停用**（仅限非基于登录名的许可证管理或非虚拟桌面）：在列表中选择一个产品密钥，然后单击**“停用”**以停用产品密钥。如果需要将产品密钥转移到另一台计算机或者此计算机上不再需要产品密钥，请将产品密钥停用。

  有关停用产品密钥的详细信息，请参见 Tableau Desktop 和 Tableau Prep 部署指南中的转移或停用产品密钥。

- **激活：**安装 Tableau Desktop 或 Tableau Prep 之后，单击**“激活”**打开激活对话框并输入您的产品密钥。如果遇到错误，并且无法使用产品密钥激活 Tableau Desktop 或 Tableau Prep，请与 Tableau 支持部门联系。

  有关激活产品密钥的详细信息，请参见 Tableau Desktop 和 Tableau Prep 部署指南中的激活和注册产品。

使用零停机许可自动刷新产品密钥

从 Tableau 版本 2021.1 开始，连接到 Internet 的 Tableau Desktop 和 Tableau Prep Builder 用户可能不必手动刷新产品密钥。如果用户已登录 Tableau Desktop 或 Tableau Prep Builder，则从订阅到期前 14 天开始，无需任何操作即可自动刷新长期许可证。永久产品密钥不会自动刷新，必须使用**“管理产品密钥”**菜单选项手动刷新。

Tableau Desktop 和 Tableau Prep Builder 将尝试以无提示方式刷新活动产品密钥，并将在许可证设置为过期前 14 天警告用户（如果无提示刷新不成功）。Tableau 将尝试刷新产品密钥三次（在许可证到期前 14 天、2 天和 1 天），以反映由于订阅续订而延长的许可证结束日期。产品密钥不会刷新，除非在这些期间 Tableau Desktop 用户登录到 Tableau Desktop。对于每天未登录 Tableau Desktop 的用户，您必须使用**“管理产品密钥”**菜单选项刷新他们的产品密钥。

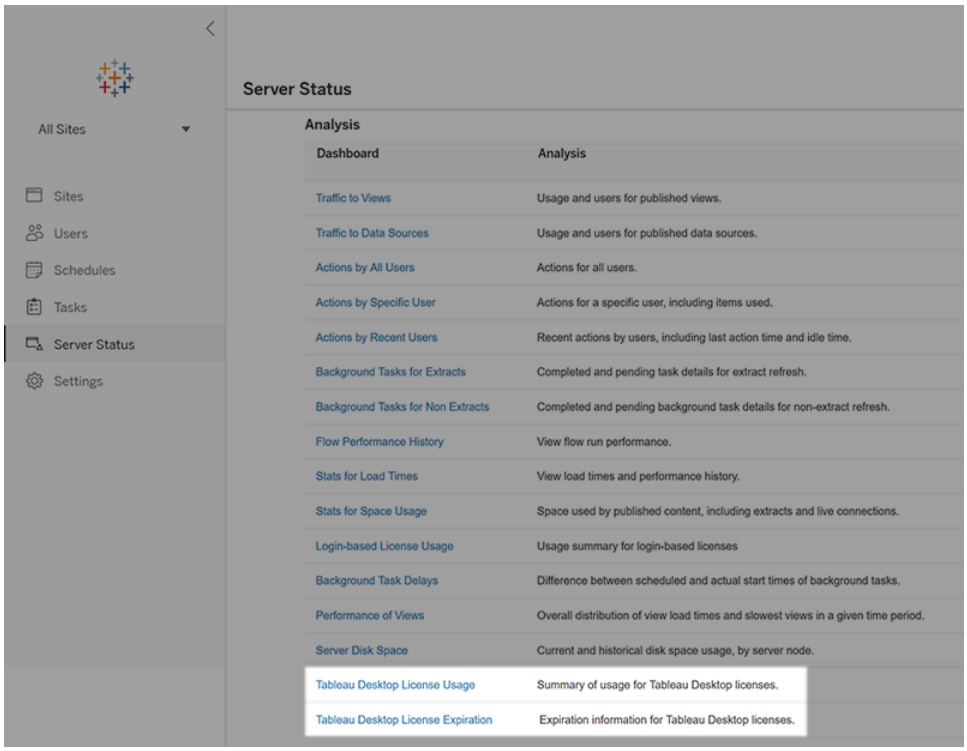
跟踪 Tableau Desktop 许可证使用情况和过期数据

如果在 Tableau Server 中要跟踪和查看 Tableau Desktop 的许可证使用情况和过期数据，您必须将 Tableau Desktop 配置为按设定的间隔将许可证数据发送到 Tableau Server，然后在 Tableau Server 上启用报告。

此样，服务器管理员将能够访问两个报告：

- **Desktop 许可证使用情况**：此报告让服务器管理员能够看到组织中 Tableau Desktop 许可证的使用情况数据。
- **Desktop 许可证过期**：此报告为服务器管理员提供有关组织中的哪些 Tableau Desktop 许可证已过期或需要维护续订的信息。

如果针对许可证报告配置了 Tableau Desktop 和 Tableau Server，在以管理员身份登录到 Tableau Server 时，您将看到列在**“分析”**部分的**“服务器状态”**页面上的这两个报告。



如果未看到列出的这些报告，则可能未针对 Tableau Desktop 使用情况报告配置 Tableau Desktop 和 Tableau Server。

有关如何针对使用情况报告配置 Tableau Desktop 和 Tableau Server 的信息，请参见 Tableau Desktop 和 Tableau Prep 部署指南中的管理 Tableau Desktop 许可证使用情况。

其他资源

有关管理许可证的详细信息，请参阅以下主题：

- 若要查找产品密钥并激活 Tableau Desktop 或 Tableau Prep Builder，请参见我的产品密钥在何处。
- 若要停用产品密钥或将其转移到另一台计算机，请参见转移或停用 Tableau Desktop。
- 若要详细了解非持久性虚拟桌面或定期重新映像的计算机的产品密钥，请参见配置虚拟桌面支持。
- 若要详细了解 Tableau Server 或 Tableau Cloud 的产品密钥管理，请参见“许可概述”(Linux | Windows)
- 若要详细了解许可证续订过程或续订许可证，请参见如何续订 Tableau 许可证。

Tableau Prep 函数参考

Tableau Prep 计算函数允许您使用计算字段，通过数据源中已存在的数据来创建新数据。

数字函数

语法	描述
ABS(number)	返回给定数字的绝对值。 示例： ABS(-7) = 7 ABS([Budget Variance]) 返回“Budget Variance”(预算差异)字段中包含的所有数字的绝对值。
ACOS(number)	返回给定数字的反余弦。结果以弧度表示。 示例： ACOS(-1) = 3.14159265358979
ASIN(number)	返回给定数字的反正弦。结果以弧度表示。 示例： ASIN(1) = 1.5707963267949
ATAN(number)	返回给定数字的反正切。结果以弧度表示。 示例： ATAN(180) = 1.5652408283942
ATAN2(y number, x number)	返回两个给定数字(x 和 y)的反正切。结果以弧度表示。 示例： ATAN2(2, 1) = 1.10714871779409
CEILING(number)	将数字舍入为值相等或更大的最近整数。

	示例： CEILING(3.1415) = 4
COS(number)	返回角度的余弦。以弧度为单位指定角度。 示例： COS(PI( ) /4) = 0.707106781186548
COT(angle)	返回角度的余切。以弧度为单位指定角度。 示例： COT(PI( ) /4) = 1
DIV(integer1, integer2)	返回将 integer1 除以 integer2 的除法运算的整数部分。 示例： DIV(11,2) = 5
EXP(number)	返回 e 的给定数字次幂。 示例： EXP(2) = 7.389 EXP(-[Growth Rate]*[Time])
FLOOR(number)	将数字舍入为值相等或更小的最近整数。 示例： FLOOR(3.1415) = 3
HEXBINX(number, number)	将 x、y 坐标映射到最接近的六边形数据桶的 x 坐标。数据桶的边长为 1，因此，可能需要相应地缩放输入。 示例： HEXBINX([Longitude], [Latitude])
HEXBINY(number, number)	将 x、y 坐标映射到最接近的六边形数据桶的 y 坐标。数据桶的边长为 1，因此，可能需要相应地缩放输入。

	示例： `HEXBINY([Longitude], [Latitude])`
`LN(number)`	返回数字的自然对数。如果数字小于或等于 0，则返回 Null。 示例： `LN(EXP(5)) = 5`
`LOG(number [, base])`	返回数字以给定底数为底的对数。如果省略了底数值，则使用底数 10。 示例： `LOG(256, 2) = 8`

`MAX(number, number)`	返回两个参数（必须为相同类型）中的较大值。如果任一参数为 Null，则返回 Null。MAX 也可应用于聚合计算中的单个字段。
	示例：
	`MAX(4,7)`
	`MAX(Sales,Profit)`
	`MAX([First Name],[Last Name])`

`MIN(number, number)`	返回两个参数（必须为相同类型）中的较小值。如果任一参数为 Null，则返回 Null。MIN 也可应用于聚合计算中的单个字段。
	示例：
	`MIN(4,7)`
	`MIN(Sales,Profit)`
	`MIN([First Name],[Last Name])`

`PI( )`	返回数字常量 pi：3.14159。 示例： `PI() = 3.14159265358979`

POWER(number, power)	计算数字的指定次幂。 示例： POWER(5,2) = 52 = 25 POWER (Temperature, 2) 也可以使用 ^ 符号： 5^2 = POWER(5,2) = 25
RADIANS(number)	将给定数字从度数转换为弧度。 示例： RADIANS(180) = 3.14159
ROUND(number, [decimals])	将数字舍入为指定位数。decimals 参数指定要在最终结果中包含的小数位数精度。如果省略 decimals，则 number 舍入为最接近的整数。 示例： 此示例将每个 Sales 值舍入为整数： ROUND(Sales)
SIGN(number)	返回数字的符号：可能的返回值为：在数字为负时为 -1，在数字为零时为 0，在数字为正时为 1。 示例： 如果 profit 字段的平均值为负值，则 SIGN(AVG(Profit)) = -1
SIN(number)	返回角度的正弦。以弧度为单位指定角度。 示例： SIN(0) = 1.0 SIN(PI( )/4) = 0.707106781186548
SQRT(number)	返回数字的平方根。

	示例： `SQRT(25) = 5`
`SQUARE(number)`	返回数字的平方。 示例： `SQUARE(5) = 25`
`TAN(number)`	返回角度的正切。以弧度为单位指定角度。 示例： `TAN(PI ( )/4) = 1.0`
`ZN(expression)`	如果表达式不为 Null，则返回该表达式，否则返回零。使用此函数可使用零值而不是 Null 值。 示例： `ZN([Profit]) = [Profit]`

字符串函数

语法	描述
`ASCII(string)`	返回字符串的第一个字符的 ASCII 代码。 示例： `ASCII('A') = 65`
`CHAR(number)`	返回通过 ASCII 代码编码的字符。 示例： `CHAR(65) = 'A'`
`CONTAINS(string, substring)`	如果给定字符串包含指定子字符串，则返回 true。 示例： `CONTAINS("Calculation", "alcu") =`

	true
ENDSWITH(string, substring)	如果给定字符串以指定 substring 结尾，则返回 true。会忽略尾随空格。 示例： ENDSWITH(“Tableau”, “leau”) = true
FIND(string, substring, [start])	返回 substring 在字符串中的索引位置，如果未找到子字符串，则返回 0。如果添加了可选参数 start，则函数会忽略在索引位置 start 之前出现的任何 substring 实例。字符串中第一个字符的位置为 1。 示例： FIND("Calculation", "alcu") = 2
FINDNTH(string, substring, occurrence)	返回指定字符串内的第 n 个子字符串的位置，其中 n 由 occurrence 参数定义。 示例： FINDNTH("Calculation", "a", 2) =7
ISDATE(string)	如果给定字符串为有效日期，则返回 true。 示例： ISDATE("2004-04-15") = True
LEFT(string, number)	返回字符串最左侧一定数量的字符。 示例： LEFT("Matador", 4) = "Mata"
LEN(string)	返回字符串长度。 示例： LEN("Matador") = 7
LOWER(string)	返回字符串，其所有字符为小写。 示例：

	`LOWER("ProductVersion") = "productversion"`
`LTRIM(string)`	返回移除了所有前导空格的字符串。 示例： `Example: LTRIM(" Sales") = "Sales"`

`(MID(string, start, [length])`	返回从索引位置 start 开始的字符串。字符串中第一个字符的位置为 1。如果添加了可选参数 length，则返回的字符串仅包含该数量的字符。
	示例：
	`MID("Calculation", 2) = "alculation"`
	`MID("Calculation", 2, 5) ="alcul"`

`PROPER(string)`	转换文本字符串，使每个单词的第一个字母大写，其余字母小写。空格和非字母数字字符（如标点符号）也可用作分隔符。 示例： `PROPER("PRODUCT name") = "Product Name"`
`REGEXP_EXTRACT(string, pattern)`	返回与正则表达式模式中的捕获组匹配的给定字符串的子字符串。正则表达式模式只需要一个捕获组 示例： `REGEXP_EXTRACT('abc 123', '[a-z]+\\s+(\\d+)') = '123'`
`REGEXP_EXTRACT_NTH(string, pattern, index)`	使用正则表达式模式返回给定字符串的子字符串。子字符串匹配到第 n 个捕获组，其中 n 是给定的索引。 示例：

	`REGEXP_EXTRACT_NTH('abc 123', '([a-z]+)\\s+(\\d+)', 2) = '123'`
`REGEXP_MATCH(string, pattern)`	如果提供的字符串的子字符串匹配正则表达式模式，则返回 true。 示例： `REGEXP_MATCH('-([1234].[ The.Market])-', '\\[\\s*(\\w*\\.)(\\w*\\s*\\])') = true`
`REGEXP_REPLACE(string, pattern, replacement)`	返回给定字符串的副本，其中匹配模式被替换字符串取代。 示例： `REGEXP_REPLACE('abc 123', '\\s', '-') = 'abc--123'`
`REPLACE(string, substring, replacement)`	在字符串中搜索子字符串，并将其替换为替代子字符串。如果未找到子字符串，则字符串保持不变。 示例： `REPLACE("Calculation", "ion", "ed") = "Calculated"`
`RIGHT(string, number)`	返回字符串最右侧一定数量的字符。 示例： `RIGHT("Calculation", 4) = "tion"`
`RTRIM(string)`	返回移除了所有尾随空格的字符串。 示例： `Example: RTRIM("Market ") = "Market"`
`SPACE(number)`	返回由指定数量的重复空格组成的字符串。 示例：

	SPACE(1) = " "
SPLIT(string, delimiter, token number)	返回字符串中的子字符串，由从字符串开头或结尾提取字符的分隔符确定。 示例： SPLIT('a-b-c-d', '-', 2) = 'b'
STARTSWITH(string, substring)	如果字符串以 substring 开头，则返回 true。会忽略前导空格。 示例： STARTSWITH(“Joker”, “Jo”) = true
TRIM(string)	返回移除了前导和尾随空格的字符串。 示例： TRIM(" Calculation ") = "Calculation"
UPPER(string)	返回字符串，其所有字符为大写。 示例： UPPER("Calculation") = "CALCULATION"

聚合函数

语法	描述
AVG(expression)	返回表达式中所有值的平均值。AVG 只能用于数字字段。会忽略 Null 值 示例； AVG([Profit])
COUNT(expression)	返回组中的项目数。不对 Null 值计数。 示例：

	`COUNT([Customer ID])`
`COUNTD(expression)`	返回组中不同项目的数量。不对 NULL 值计数。每个唯一值仅计算一次。 示例： `COUNTD([Region])`
`MEDIAN(expression)`	返回单个表达式的中位数。MEDIAN 只能用于数字字段。会忽略 Null 值。 示例： `MEDIAN([Profit])`
`PERCENTILE(expression, number)`	从给定表达式中返回与指定数字相对应的百分点值的聚合计算。数字的有效值为 0 到 1。PERCENTILE([表达式], 0.50) 将始终返回表达式中的中值。 示例： `PERCENTILE([Sales], 0.90)`
`STDEV(expression)`	返回表达式的样本标准差。 示例： `STDEV([Profit])`
`STDEVP(expression)`	返回表达式的总体标准差。 示例： `STDEVP([Profit])`
`SUM(expression)`	返回表达式中所有值的总计。SUM 只能用于数字字段。会忽略 Null 值。 示例： `SUM([Profit])`
`VAR(expression)`	基于群体样本返回给定表达式中所有值的统计方差。

	示例： VAR([Height])
VARP(expression)	对整个群体返回给定表达式中所有值的统计方差。 示例： VARP([Test Score])
ZN(expression)	如果表达式不为 Null，则返回该表达式，否则返回零。使用此函数可使用零值而不是 Null 值。 示例： ZN([Profit]) = [Profit]

类型转换函数

语法	描述
DATE	在给定数字、字符串或日期表达式的情况下返回日期。 示例： DATE("2006-06-15 14:52") = 2006-06-15 请注意，引号是必需的。
DATETIME(expression)	在给定数字、字符串或日期表达式的情况下返回日期时间。 示例： DATETIME("April 15, 2004 07:59:00") 请注意，引号是必需的。
FLOAT(expression)	给定任何类型的表达式，返回一个浮点数。此函数需要未格式化的数字，这意味着排除逗

	号和其他符号。 示例： `FLOAT("3")= 3.000`
`INT(expression)`	给定表达式，返回一个整数。此函数将结果截断为最接近于 0 的整数。 示例： `INT(8.0/3.0) = 2 or INT(-9.7) = -9`
`MAKEDATE(year, month, day)`	返回一个由年、月和日构造的日期值。 示例： `MAKEDATE(2014, 3, 18)`
`MAKEDATETIME(date, time)`	在给定日期表达式和时间表达式的情况下返回日期和时间值。 示例： `MAKEDATETIME(#2012-11-12#, #07:59:00#)`
`MAKETIME(hour, minute, second)`	返回一个由小时数、分钟数和秒数构造的时间值。 示例： `MAKETIME(14, 52, 40)`
`STR(expression)`	给定表达式，返回字符串。 示例： `STR([Age])` 以字符串形式返回 Age 度量的所有值。

日期函数

语法	描述

DATEADD(date_part, interval, date)	向指定日期添加增量并返回新日期。增量由间隔和 date_part 定义。 示例： DATEADD('month', 3, #2004-04-15#) = 2004-07-15 12:00:00 AM
DATEDIFF(date_part, start_date, end_date, [start_of_week])	返回从 end_date 中减去 start_date 后的两个日期之差。差以 date_part 为单位表示。如果省略 start_of_week，则周开始日由为数据源配置的开始日确定。 示例： DATEDIFF('month', #2004-07-15#, #2004-04-03#, 'sunday') = -3
DATENAME(date_part, date, [start_of_week])	以字符串形式返回给定日期的一部分，其中该部分由 date_part 定义。如果省略 start_of_week，则周开始日由为数据源配置的开始日确定。 示例： DATENAME('month', #2004-04-15#) = "April"
DATEPARSE(format, string)	将字符串转换为指定格式的日期。 示例： DATEPARSE ("dd.MMMM.yyyy", "15.April.2004") = 2004-04-15 12:00:00 AM
DATEPARTT(date_part, date, [start_of_week])	以整数形式返回给定日期的一部分，其中该部分由 date_part 定义。如果省略 start_of_week，则周开始日由为数据源配置的开始日确定。 示例： DATEPART('month', #2004-04-15#) =

	4
DATETRUNC(date_part, date, [start_of_week])	按 date_part 指定的准确度截断指定日期，并返回新日期。如果省略 start_of_week，则周开始日由为数据源配置的开始日确定。 示例： DATETRUNC('quarter', #2004-08-15#) = 2004-07-01 12:00:00 AM
DAY(date)	以整数的形式返回给定日期的天。 示例： DAY(#2004-04-12#) = 12
ISDATE(string)	如果给定字符串为有效日期，则返回 true。 示例： ISDATE("2004-04-15") = True
MONTH(date)	以整数的形式返回给定日期的月份。 示例： MONTH(#2004-04-12#) = 4
NOW()	返回当前日期和时间。 示例： NOW() = 2004-05-12 1:08:21 PM
TODAY()	返回当前日期。 示例： TODAY() = 2004-05-12
YEAR(date)	以整数的形式返回给定日期的年份。 示例： YEAR(#2004-04-12#) = 2004

逻辑函数

语法	描述
`IFNULL(expr1, expr2)`	如果 <expr1> 不为 null，则返回该表达式，否则返回 <expr2>。 示例： `IFNULL([Profit], 0)`
`IIF(test, then, else, [unknown])`	检查某个条件是否得到满足，如果为 TRUE 则返回一个值，如果为 FALSE 则返回另一个值，如果未知，则返回可选的第三个值或 NULL。 示例： `IIF([Profit] > 0, 'Profit', 'Loss')`
`ISDATE(string)`	如果给定字符串为有效日期，则返回 true。 示例： `ISDATE("2004-04-15") = True`
`ISNULL(expression)`	如果表达式未包含有效数据 (`Null`)，则返回 true。 示例： `ISNULL([Profit])`
`ZN(expression)`	如果表达式不为 Null，则返回该表达式，否则返回零。使用此函数可使用零值而不是 Null 值。 示例： `ZN([Profit]) = [Profit]`

分析函数

语法	描述

LAST_VALUE(expression, [return_last_non-null_value])	返回给定表达式从分区中第一行到当前行的最后一个值。接受可选的第二个布尔参数以返回最后一个非 null 值。 示例： { ORDERBY [Row ID] ASC: LAST_VALUE ([Category], true) } = returns the last non-null value of [Category]
LOOKUP(expression, [offset])	返回目标行(指定为与当前行的相对偏移)中给定表达式的值。如果无法确定目标行，则返回 NULL。 LOOKUP 函数的 [offset] 参数只能接受字面整数，例如 1。 示例： {ORDERBY [DATE] ASC : LOOKUP ([SALES], -1)} 返回上一行中的 SALES 值。
NTILE(number)	将选定分区中的行分配到指定数量的组或图块中。使用此函数，分配给 3 个图块的值集 (6,9,9,14) 将按升序分配给图块 (1,2,2,3)。 示例： {PARTITION [Customer]: {ORDERBY [Sales] ASC: NTILE(3) }}
RANK()	返回分区中当前行的标准竞争排名。为相同的值分配相同的排名。 示例： {PARTITION [Customer]: {ORDERBY [Sales] ASC: RANK() }}
RANK_DENSE()	返回分区中当前行的密集排名。为相同的值分配相同的排名，但不会向数字序列中插入间距。

	示例： `{PARTITION [Customer]: {ORDERBY [Sales] ASC: RANK_DENSE() }}`
`RANK_MODIFIED()`	返回分区中当前行的调整后竞争排名。为相同的值分配相同的排名。 示例： `{PARTITION [Customer]: {ORDERBY [Sales] ASC: RANK_MODIFIED() }}`
`RANK_PERCENTILE()`	返回分区中当前行的百分位排名。 示例： `{PARTITION [Customer]: {ORDERBY [Sales] ASC: RANK_PERCENTILE() }}`
`ROW_NUMBER()`	为每一行返回唯一的连续行 ID。 示例： `{PARTITION [Customer]: {ORDERBY [Sales] ASC: ROW_NUMBER() }}`
`RUNNING_AVG(expression)`	返回给定表达式从分区中第一行到当前行的运行平均值。 示例： `{ PARTITION [Category] : { ORDERBY [Row ID] ASC : RUNNING_AVG([Sales]) } } = running average of Sales, per Category.`
`RUNNING_SUM(expression)`	返回给定表达式从分区中第一行到当前行的运行总计。 示例： `{ ORDERBY [Row ID] ASC : RUNNING_SUM([Sales]) } = running sum of all Sales`

关键字

函数	描述
AND	对两个表达式执行逻辑合取运算 示例： `IF (ATTR([Market]) = "New Business" AND SUM([Sales]) > [Emerging Threshold] )THEN "Well Performing"`
ASC	为 ORDERBY 运算定义升序。 示例 `{ORDERBY [Sales] ASC:RANK()}`
CASE	查找第一个与 <expr> 匹配的 <value>，并返回对应的 <return>。 示例： `CASE [RomanNumeral] WHEN "|" THEN 1 WHEN "II" THEN 2 ELSE 3 END`
DESC	为 ORDERBY 运算定义降序。 示例： `{ORDERBY [Sales] DESC:RANK()}`
ELSE	测试一系列表达式，同时为第一个为 true 的 <expr> 返回 <then> 值。 示例： `IF [Profit] > 0 THEN 'Profitable' ELSE 'Loss' END`
ELSEIF	测试一系列表达式，同时为第一个为 true 的 <expr> 返回 <then> 值。相当于在 ELSE 运算中放置一个 if 语句。 示例：

	`IF [Profit] > 0 THEN 'Profitable' ELSEIF [Profit] = 0 THEN 'Breakeven' ELSE 'Loss' END`
`END`	终止由相应指令引入的运算。 示例： `IF [Profit] > 0 THEN 'Profitable' ELSEIF [Profit] = 0 THEN 'Breakeven' ELSE 'Loss' END`
`{ FIXED }`	FIXED 详细级别表达式使用指定的维度计算值而不引用视图详细级别。 示例： `{FIXED[Customer ID]:MIN([Order Date])}`
`IF`	创建条件语句(`IF` 语句)，并且允许您仅在条件为 True 时执行。 示例： `IF [Profit] > 0 THEN 'Profitable' ELSEIF [Profit] = 0 THEN 'Breakeven' ELSE 'Loss' END`
`NOT`	对一个表达式执行逻辑非运算。 示例： `IF NOT [Profit] > 0 THEN "Unprofitable" END`
`OR`	对两个表达式执行逻辑析取运算。 示例： `IF [Profit] < 0 OR [Profit] = 0 THEN "Needs Improvement" END`
`{ ORDERBY }`	定义应用分析函数的顺序。使用可选的 asc |

	desc 参数指定每个字段的升序或降序。 示例： `{ORDERBY [Customer Name] ASC, [Sales] DESC: RANK() }`
`{ PARTITION }`	将组定义为分析函数。ORDERBY 必须位于分区内。 示例： `{PARTITION [Customer]: {ORDERBY [Sales]: RANK()}}`
`THEN`	测试一系列表达式，同时为第一个为 true 的 <expr> 返回 <then> 值。 示例： `IF [Profit] > 0 THEN 'Profitable' ELSEIF [Profit] = 0 THEN 'Break even' ELSE 'unprofitable' END`
`WHEN`	查找第一个与 expr> 匹配的 <value>，并返回对应的 <return>。 示例： `CASE [RomanNumberal] WHEN 'I' THEN 1 WHEN 'II' THEN 2 ELSE 3 END`