

何时混合以及何时联接

欢迎观看这段对数据混合和跨数据库联接进行比较的视频。

您可以下载练习工作簿，以便在自己的 Tableau 软件副本中跟着操作。

简单地说，跨数据库联接在行级别对两个数据源进行整合，生成单个数据源。混合则会使用从每个数据源获得的独立的查询结果，在视图中对这些结果进行聚合。如需特定于跨数据库联接或数据混合的更多信息，请观看两个主题各自的视频。

一对多关系

首先，我们使用跨数据库联接视频中的“2016 产品”和“2016 销售额”数据源来设置一个混合。我们按市场对类别销售额进行快速分析。

- 从**产品**数据源，将“类别”移至“行”
- 然后切换到**销售额**数据源
 - 单击“产品 ID”旁边的链接符号，建立混合关系
 - 然后将“销售额”移至“列”
- 看起来不错，我们有了混合，可以了解每个类别的总体销售额
- 但如果将“市场”移至“颜色”，情况又会如何？现在，大部分条形变为灰色，也就是星号。为什么？

我们根据“产品 ID”进行混合，它与“市场”的关系是“一对多”的。淡青色条形显示仅在美国和加拿大市场销售的产品 ID，但所有灰色条形都在多个市场销售 -- 显示为聚合，用星号标记。

在进行混合时，来自每个数据源的结果被分别聚合，然后合并显示在视图中。因此，如果存在这种一对多的关系，我们就无法轻松地进行数据可视化。

在这个例子中，跨数据库联接优于混合。此数据源在行级别按照“产品 ID”进行联接。我们重新创建第一个视图

- “类别”拖至“行”
- 将“销售额”拖至“列”
- “市场”拖至“颜色”

现在这种做法正是我们需要的，按市场分解类别销售额。联接数据时，我们只需用产品信息来补充销售额信息，使来自“销售额”的原始记录数保持不变。

联接会增加行数

那么，跨数据库联接是否总是最佳选项呢？让我们切换到另一对数据集，来自数据混合视频的“办公城市”和“咖啡连锁店”。这是同一家母公司的两个零售连锁店。“办公城市”的销售点覆盖美国全部 50 个州，而“咖啡连锁店”的销售点分布在 20 个州。假设我们想查看两个组织在全部 50 个州的合并销售额。

我们有一个联接的数据源，联接依据为二者的共同字段“州/省”。

- 我们要显示来自此数据源中“办公城市”连接的“州/省”，因为其中包含所有 50 个州。
- 如果我们右键单击并编辑计算“添加的销售额”，可以看到，我们已经添加了“办公城市”和“咖啡连锁店”销售额，
 - 而且通过使用“Zero Null”函数，对于没有“咖啡连锁店”销售额的州，“Null”将作为零处理。
- 我们要将“添加的销售额”移至“列”。
- 哇。生意火爆 - 如果某些州的销售额真的达到了数亿元。这里面肯定有问题 — 让我们更加深入地调查一下。

此仪表板显示了 4 个州的“办公城市”销售额，以及这些州的“咖啡连锁店”销售额。合并销售额应该是我们在中间的方程中看到的销售额 - 将这些值简单地加起来

- 混合（黄色）的合并销售额给出了正确的答案，但联接数据源（紫色）让这些值大幅增加。

如果我们观察“办公城市”、“咖啡连锁店”和联接数据源的记录数，我们看到，数量明显变大了。

因为我们按照“州”进行联接，Tableau 为每个具有非重复信息的“州”合并项生成了一个行。“咖啡连锁店”中的“产品”和“日期”，以及“办公城市”中的“行 ID”。此交叉产品意味着我们对“销售额”进行了多次重复统计。

相反，混合提供了正确的值，因为它返回了“州/省”在一个数据源中的销售额，同一个州在另一个数据源中的销售额，我们可以直接将二者相加。在这个例子中，这是我们需要做的。

可以看到，我们用作联接或混合依据的字段对分析有巨大的影响。只有选择正确的字段，理解混合或联接的生成方式，我们才能知道应该使用哪种方式，以及在何时使用。

各自的优点

通过跨数据库联接创建集成数据源有很多好处。

- 联接在行级别进行，因此可以避免数据混合受到的“一对多”限制，也就是说，可以消除星号。
- 与数据混合不同，我们可以提取、保存、发布和与他人轻松共享单个数据源。与任何其他数据源一样

然而，如果联接字段的粒度不对，联接操作可能会人为地增大数据集，提供误导性信息。

数据混合向每个数据源发送单独的查询，并在视图中将它们聚合到需要的粒度级别，这可能正是分析需要的功能。

总结

感谢观看这段关于通过混合或跨数据库联接来合并数据源的视频。我们邀请您继续观看免费培训视频，进一步了解如何使用 Tableau。