

第 5 章



数据建模

数据建模是指对现实世界中各类数据的抽象表述。在数据分析过程中,数据建模的目的是将现有的实际数据整合起来,能够表达出问题的本质,然后对现实世界进行分析、抽象并从中找出内在联系。在分析数据时,不可能总是对单个数据表进行分析,有时需要把多个数据表导入系统,对多个表中的数据及其关系执行一些复杂的数据分析任务。因此,为准确计算分析的结果,需要在数据建模中创建数据表之间的关系。在 Power BI 中,关系(Relationship)是指数据表之间的基数(Cardinality)和交叉筛选方向(Cross Filter Direction)。

5.1 基数和交叉筛选

数据关系指的是数据之间的逻辑关系,通过在不同表中的数据之间创建关系,可以增强数据分析的功能。

1. 基数关系

两个数据表进行关联,一般都是通过两个数据表之间的单个数据列进行关联,该数据列叫作查找列。两个数据表之间的基数关系是 $1:1$ 、 $1:N$ 或 $N:1$,基数关系表示的含义如下。

(1) 多对一或者一对多($N:1$ 或 $1:N$): 这是最常见的默认类型。这意味着一个表中的列可具有一个值的多个实例,而另一个相关表(常称为查找表)仅具有一个值的一个实例。

(2) 一对一($1:1$): 这意味着一个表中的列仅具有特定值的一个实例,而另一个相关表也是如此。

例如,表 A 和表 B 之间的基数关系是 1 : N,那么表 A 是表 B 的查找表,表 B 叫作引用表。在查找表中,查找列的值是唯一的,不允许存在重复值,而在引用表中,查找列的值不是唯一的。

有时在系统中,引用表会引用查找表中不存在的数据。默认情况下,系统会自动在查找表中增加一个查找值——空白列(Blank),所有不存在于查找表中的值都映射到空白列。

2. 交叉筛选方向

交叉筛选方向指的是筛选的流向,表示一个筛选条件对其他相关表进行筛选。例如,表 A 对表 B 筛选,其筛选方向可以是双向的,也可以是单向的。

(1) 单向:表示一个表只能对另外一个表进行筛选,而不能反向进行筛选操作。

(2) 双向:默认方向,指的是为了进行筛选,两个表均被视为同一个表。这非常适用于其周围具有多个查找表的单个表。

例如,在图 5-1 中的星形结构中,中间的表格是一个引用表,它的周围是 4 个查找表,引用表和查找表之间的筛选关系都是双向的。

通常情况下,双向筛选用于星形结构。但是,双向筛选不太适合如图 5-2 所示的模式。在图 5-2 中,筛选方向形成一个循环,在这类关系模式中,双向筛选会创建一组语义不明的关系。例如,求取表 A 中某个字段的总和,如果选择按照表 B 中的某个字段进行筛选,则不清楚筛选器应该如何流动,是通过顶部表还是底部表进行流动。

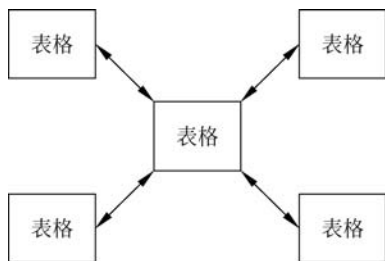


图 5-1 数据关系：星形结构例图

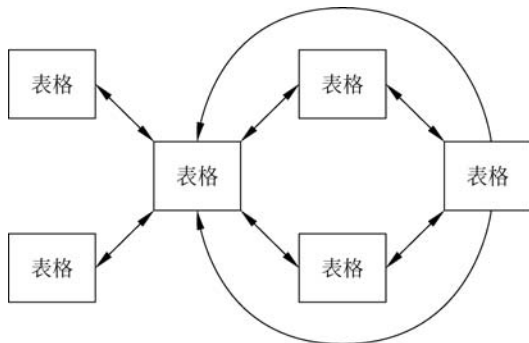


图 5-2 数据关系：星形结构加循环

如果双向筛选能够导致数据关系的多义性,建议可以导入数据表两次(在第二次使用该表时,将其修改成其他名字)以消除循环。

5.2 交互关系

在导入多个数据表时,很可能需要使用所提供的数据表中的数据进行关联来执行一些分析操作,为准确计算结果并在报表中显示正确信息,这些数据表之间建立的相互关系是必要的。下面介绍几种常见的交互关系。

5.2.1 单向交互关系

当关系的“交叉筛选方向”(Cross Filter Direction)属性设置为单向箭头时,即把交叉交互方向设置为单向箭头(Single)时,箭头由维度表指向事实表,维度表用于对事实表进行查询,一旦关系创建成功,就会按照维度表对事实表进行切片(聚合查询)。这种传统的数据模型和数据仓库的星形模型相同,其特点是:维度表包含属性,事实表包含度量。按照维度表的属性对事实表的度量进行切片/聚合查询,如图 5-3 所示。

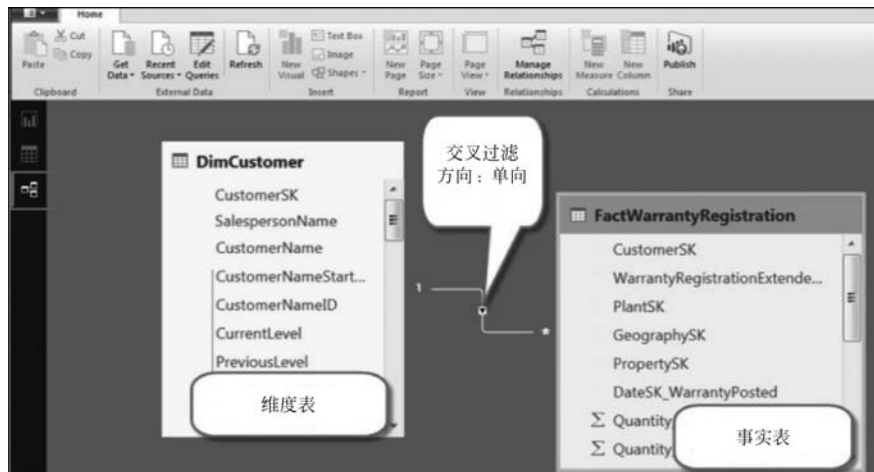


图 5-3 单向交互关系

5.2.2 双向交互关系

当关系的“交叉筛选方向”属性设置为双向箭头时,即把交叉交互方向设置为双向箭头(Both)时,为了实现数据的过滤,系统把这两个数据表展开成一个整体的大数据表,如图 5-4 所示。

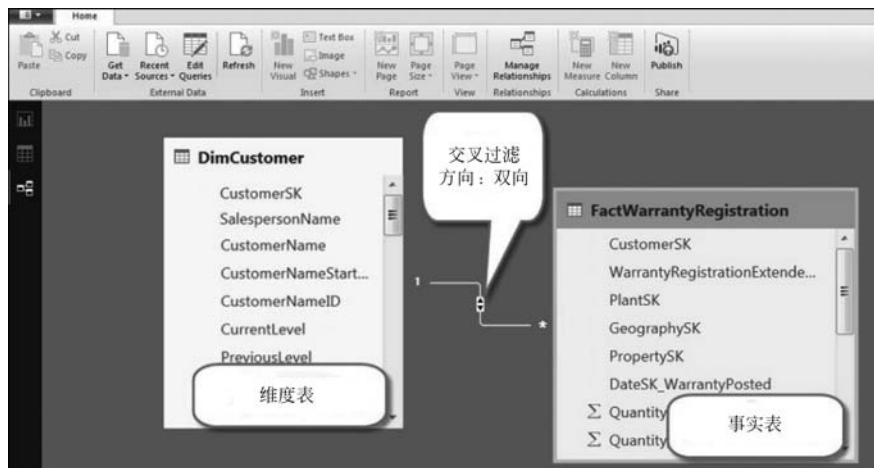


图 5-4 双向交互关系

双向交互方向会导致有些关系处于不活跃状态,当一个维度表和多个事实表有关系时,避免使用双向箭头方向,因为这样可能会导致部分关系失效,处于不活跃状态,所以一般情况下,单向筛选系统是默认设置。

5.2.3 关系的传递

在系统中,关系是可以传递的,这就意味着筛选条件是可以传递的。把筛选器看作是流水,箭头的指向是由上游指向下游(查找表处于上游,而数据表处于下游),筛选器由查找表流向数据表。一般情况下,按照查找表对数据表进行筛选,筛选器由查找表流向数据表,再流向其他关联的数据表;如果把交叉筛选的方向设置为双向筛选,那么系统可以按照数据表对查找表进行筛选,也就是说,筛选是由数据表逆流到查找表。双向交叉筛选使得查找表被筛选和切片,并能对查找表进行聚合查询,如图 5-5 所示。

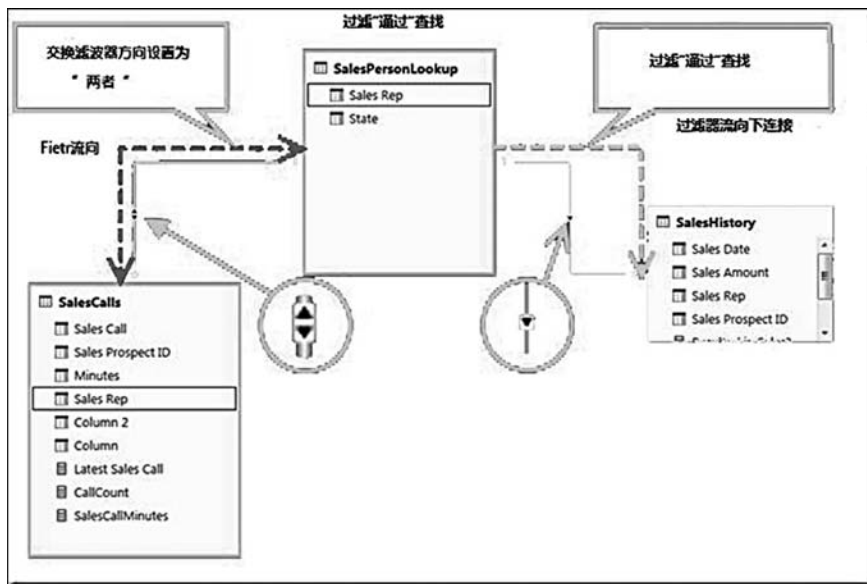


图 5-5 关系传递的效果示意

关系的传递有一个不太令人满意的地方,就是筛选的全选和不选有很大的不同:不选包含 Blank 值(空值),而全选不包含 Blank 值。在进行关系的传递时,数据行的缺失会导致下游数据出现空值。

例如,下面是 Courses、StudentCourse、Students、StudentSpeaker 4 个数据表,4 者之间的关系如图 5-6 所示,示例中的原始数据如图 5-7 所示。

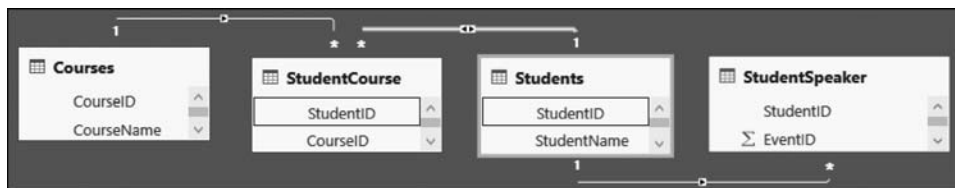


图 5-6 示例中数据表之间的关系

StudentID	EventID
1	101
4	102

StudentID	StudentName
1	stu_a
2	stu_b
3	stu_c
4	stu_d

StudentID	CourseID
1	1
2	2
3	2

CourseID	CourseName
1	English
2	Chinese

图 5-7 示例中的原始数据

把 CourseID 作为筛选器(切片可视化控件)时,下游数据可视化控件 Count(Distinct EventID)会出现空值(Blank),这是因为存在 StudentID=4 的数据行没有选择对应的 CourseID,结果如图 5-8(a)所示。在不选择任何筛选器时,Count of EventID 的值是 2,包含 Blank 对应的 EventID,结果如图 5-8(b)所示。当选择 CourseID=1 时,Count of EventID 的值是 1,结果如图 5-8(c)所示。当选择所有 Filter 时,Count of EventID 的值是 1,结果如图 5-8(d)所示。



图 5-8 示例设置及结果示意

5.2.4 间接关系

下面以示例的方式介绍间接关系的运用。

利用图 5-7 中的数据,在报表中实现以课程表的统计数据作为演讲者的用户数量。如果实现用户的需求,必须熟悉数据和数据之间的关系,在数据表 StudentCourse 中,共有 3 个用户选课,学号分别是 1、2 和 3,存在不选课的用户,而在数据表 StudentSpeaker 中,只有学号 1 的用户满足条件,因此,根据课程统计作为演讲者的用户数量的结果应该是:

- 选修 English 的用户数量是 0。
- 选修 Chinese 的用户数量是 1。
- 对所有课程做统计,用户数量是选修 English 和选修 Chinese 的数量之和。

1. 设置课程筛选器

数据表 Courses 是查找表,由于 StudentCourse 中的课程(CourseID)都存在于 Courses 表中,所以切片的图表中不存在 Blank 选项,如图 5-9 所示。



图 5-9 设置课程筛选器

2. 添加 Card 图表,显示统计数量

在页面中添加 Card 图表(Visualizations),在图表的 Fields 属性中,选择数据表 StudentSpeaker 的 StudentID 字段,属性值自动变成:聚合函数+of+字段值,如图 5-10 所示。

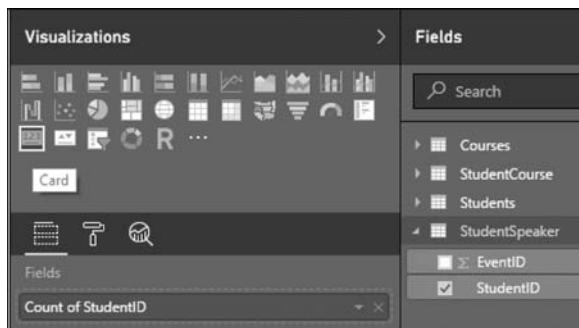


图 5-10 添加 Card 图表,显示统计数量

3. 设置聚合函数

由于一个用户可能在多个活动(Event)中担当演讲者,因此,必须对 StudentID 进行去重,对图表的 Fields 属性值 Count of StudentID 右击,在弹出的快捷菜单中选择聚合函数 Count(Distinct)命令,如图 5-11 所示。

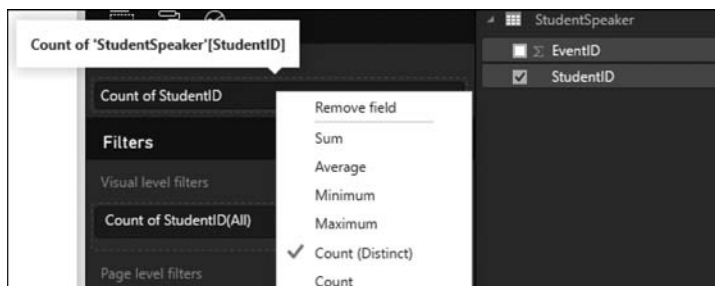


图 5-11 设置聚合函数

4. 设置图表的显示属性

切换到 图标,禁用 Category label,启用 Title,修改 Title Text、Font color、Alignment 和 Text Size,如图 5-12 所示。

5. 设置可视化控件及结果显示

如果课程选择 Chinese,数量是 Blank,结果如图 5-13(a)所示。如果课程选择 English,数量是 1,结果如图 5-13(b)所示。如果选择所有课程,数量是 1,结果如图 5-13(c)所示。

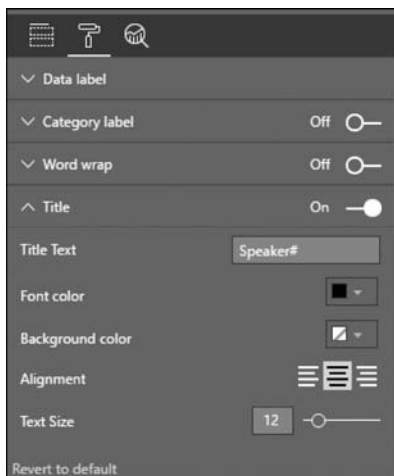


图 5-12 设置图表的显示属性

CourseName
☒ Chinese
☐ English

Speaker#
(Blank)

(a) 结果一

CourseName
☐ Chinese
☒ English

Speaker#
1

(b) 结果二

CourseName
☒ Chinese
☒ English

Speaker#
1

(c) 结果三

图 5-13 设置可视化控件及结果示意

6. 未设置筛选器

在默认情况下,图表不选择任何课程时,数量是 2,这个结果在逻辑上是错误的,对于没有选择任何选项的 Filter,系统不会做任何筛选关联,如图 5-14 所示。

导致错误的原因是数据表 StudentSpeaker 出现脏数据,出现没有选修任何课程的用户,本例是学号为 4 的用户出现在 StudentSpeaker 数据表中,如要修正查询的结果,必须清洗脏数据。

CourseName	Speaker#
☐ Chinese	2
☐ English	

图 5-14 未设置筛选器
结果示意

5.2.5 创建间接关系

数据表之间进行交互时,主要有两种关系:直接关系和间接关系。直接关系是指两个数据表相关联,而间接关系是指两个数据表不能直接相关联,而是通过中间数据表作为桥梁来建立关系,如图 5-15 所示。产品表和销售表之间的关系是直接关系,销售表和门店表之间的关系是直接关系,而产品表和门店表之间通过销售表建立的关系是间接关系。间接关系通过一系列有直接关系的数据表来实现数据的交互。值得说明的是,在数据建模中使用间接关系时,务必要注意系统对筛选器选项的全选和不选的处理是有区别的。

基数关系根据数据之间的关系创建,筛选方向根据筛选的逻辑来设置。默认情况下,系统会自动检查数据之间的关系,根据检查的结果(列名和列值的唯一性)自动创建关系。在“关系”视图中,关系是一条有方向的折线,折线的两端是数字,表示基数关系,折线中间



图 5-15 创建表之间的关系

的箭头表示筛选方向。

如果用户根据数据内在的关系来对系统自动创建的关系进行修正,可以采用以下示例的步骤来操作。例如,把销售表和门店表之间的关系修改为 1 : N 和双向筛选。首先双击关系(折线),弹出“编辑关系”对话框,如图 5-16 所示。

编辑关系

选择相互关联的表和列。

销售表

订单号	订单日期	店铺ID	产品ID	会员ID	购买数量
59752000006	Friday, January 01, 2016	111	5007	9356	4
59752000050	Sunday, January 03, 2016	106	5007	3497	4
59752000178	Thursday, January 07, 2016	101	5007	9713	4

门店表

店铺ID	店铺名
101	佳木斯市
102	哈尔滨市
103	北京市

基数: 多对一(*:1) 交叉筛选器方向: 单一

☒ 使此关系可用 ☐ 在两个方向上应用安全筛选器

☐ 假设引用完整性

确定 取消

图 5-16 “编辑关系”对话框

图 5-16 中每个表下方的面板中都会显示列名和示例数据。基数关系是“多对一”,其表达式为 * : 1,用于建立关系的数据列“店铺 ID”显示的是被选中状态。交叉筛选方向选择 Both,勾选“使这种关系可用”复选框,单击“确定”按钮即可完成关系的修正,如图 5-17 所示。



图 5-17 关系类型进行修正结果示意

5.2.6 关系的设计

在关系的设计中,把数据模型设计成维度表和事实表,维度表和事实表之间的关系是 $1:N$,交叉过滤方向由维度表指向事实表,避免使用双向交叉方向。由于系统不支持“多对多”关系类型,在处理这种数据时,通常有两种方法对“多对多”关系进行处理。

(1) 删除关系:把“多对多”的数据合并成一个数据表。

(2) 转换关系:把“多对多”的关系转换为两个“一对多”的关系。新建一个维度表,该维度表中只包含单列的唯一值,用这个新建的维度表来连接原“多对多”的两个表。

5.2.7 编辑交互行为

系统允许在不修改关系的情况下,编辑筛选条件(Filter)和度量值的交互行为,使报表中的不同图表选择性地响应或不响应筛选条件。

选择 Filter,切换到 Format 菜单,选择 Edit interactions 命令。编辑交互行为在默认情况下 Card 图表的 Filter 是选中的,将其切换到禁止状态,这样,选择 Course 筛选器中的任何一个选项都不会影响 Card 图表显示的数据值,如图 5-18 所示。

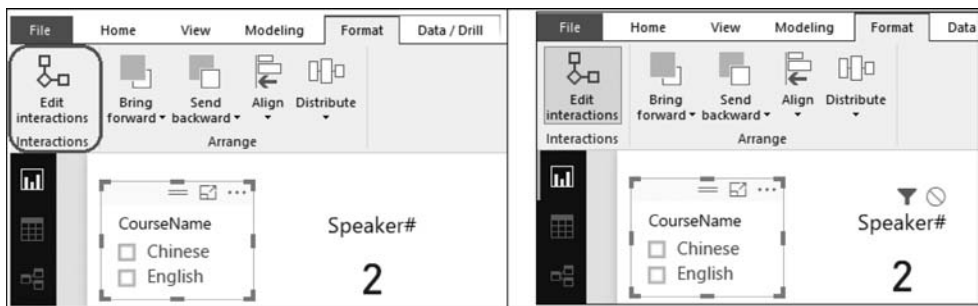


图 5-18 编辑交互关系

总之,在数据建模中要遵守一定的设计原则,用户不仅需要熟悉业务需求,而且需要熟悉数据及其数据表之间的关系,避免出现一些显而易见的错误。具体设计原则如下:

(1) 根据业务需求,设计报表的筛选条件和度量值。

(2) 筛选器是数据建模的出发点,根据筛选条件和数据之间内在的关系设计数据模型。

(3) 根据数据之间内在的关系来加载数据,保证数据表中不出现脏数据。

5.3 数据建模

下面介绍在系统中如何进行数据建模,以及如何在数据建模中创建计算列。此外,还介绍如何在系统数据建模中使用信息建模和在导航中创建计算表。系统数据建模是把扁平化、分散的信息综合到一张表上,用户可以使用来自多个数据源的多个表,只要定义了它们之间的连接,系统就会联合生成一个自定义表来计算,分配新指标以查看信息的特定表格,并在可视化中使用这些新指标进行简单的综合。

5.3.1 数学建模

要在系统中创建信息模型,用户希望以系统产生一个新报告来作为特征提供所有有用信息。在图 5-19 中使用了一个 XLS 文件来导入用户端信息和商品,选择要附加的信息供应,并单击“获取数据”按钮。在屏幕左侧的系统中,有 3 个选项卡:报告(Report)、数据(Data)和关系(Relationships)。当导航到“报告”选项卡时,用户能够看到仪表板和可视化组件。用户可以根据需要选择不同的图表类型。下面示例提供的可视化选择是一种表格排序,当单击“关系”选项卡时,用户能够根据信息来源概述的关系查看所有信息。

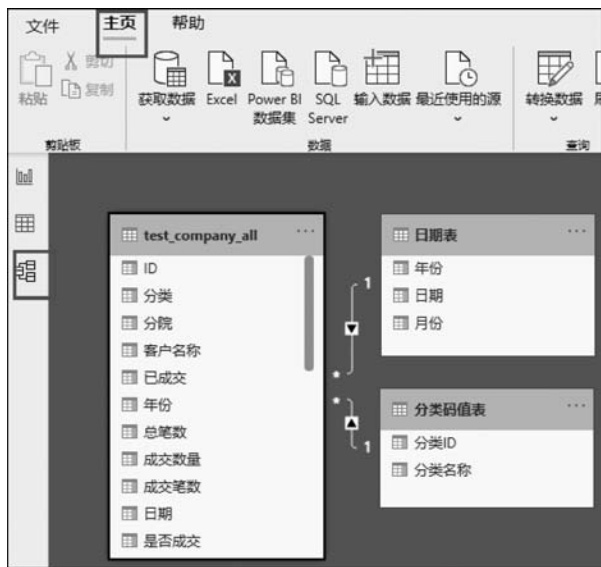


图 5-19 “获取数据”建模编辑界面

在“关系”选项卡中,用户能够看到信息源之间的连接。一旦多个信息源添加到系统可视化组件,该工具就会自动尝试发现列之间的连接。当导航到“关系”选项卡后,用户能够读取连接,并能够建立列之间的联合关系。“管理关系”对话框如图 5-20 所示。



图 5-20 “管理关系”对话框

用户也可以联合添加和消除信息视觉图像中的关系。如果要删除关系,则右击并在弹出的快捷菜单中选择“删除”命令。如果要修改或者替换关系,只需在信息源之间拖放关系箭头到连接关系的字段上,如图 5-21 所示。



图 5-21 删除或修改关系

用户可以联合使用读取连接来覆盖报告中的特定列,通过右击列名称,在弹出的快捷菜单中选择“隐藏在报表视图中”命令来完成。

5.3.2 创建计算列

用户要创建计算列,有以下 3 种方法。

- (1) 通过组合当前两个信息或多个组件,在系统中生成计算列。
- (2) 在关联的现有列上联合应用计算以替换度量标准或混合两列以形成一个新列。
- (3) 生成计算列,以确定表之间的关系。

要创建替换计算列,须导航到屏幕左侧的“数据”选项卡。单击“建模”菜单,如图 5-22 所示。



图 5-22 数据建模中创建计算列

然后单击“新建列”按钮,如图 5-23 所示。此时在数据表的最后增加了一列,列名自动命名为“列”,用户可以根据需要进行列名的修改。同时也可在屏幕上打开公式栏,用户根据需要输入 DAX 来执行计算。DAX 是一种强大的语言,用户可以通过数据计算公式或者函数来执行计算。



图 5-23 新建列效果示意

例如,在公式栏中输入公式:产品代码(Product_C),用于生成替换列,显示的结果来自“产品 ID”列的最后 3 个字符。然后,在新建列的公式栏中输入公式:产品生产日期 = DATE(2019,04,24),运行结果如图 5-24 所示。

5.3.3 创建计算表

用户可以在系统中的信息建模中联合生成替换计算表。要生成替换表,可单击“建模”菜单下的“新表”按钮,如图 5-25 所示。



图 5-24 使用公式替换新建列的名称和数据



图 5-25 创建计算表

使用 DAX 用于生成新表,在公式栏中等号和 DAX 等式的左侧输入替换表的名称,执行计算使该表正确。计算完成后,新表显示在模型的“字段”窗格中。

5.4 分析窗格

下面介绍何时创建系统 KPI(关键绩效指标)视觉效果列表,它是一个可视标志。KPI 取决于特定的度量,旨在使用户能够针对当前推荐和状态的特征化目标进行评估度量的标准。通过这种方式,KPI 视觉需要基础测量来评估客观测量。在使用中,KPI 数据集必须包含 KPI 的客观质量,如果用户的数据集不包含目标值,则可以通过在目标信息模型或 PBIX 记录中包含 Excel 表格来制定目标。下面介绍如何应用分析窗格的预测和限制。

借助分析窗格,可以为视觉效果添加动态参考线,并为重要模式或知识点提供中心。分析窗格位于系统的“可视化”窗格中,如图 5-26 所示。

注意,当用户在系统画布上选择可视化字段时,才会显示分析窗格。用户可以在分析表单中查找,如图 5-27 所示,选择“分析”表格时会显示外观框。

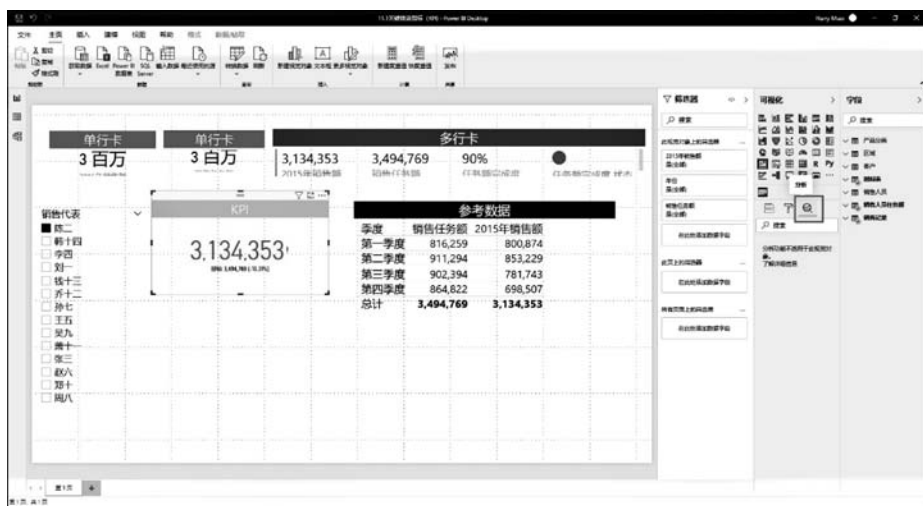


图 5-26 分析窗格界面

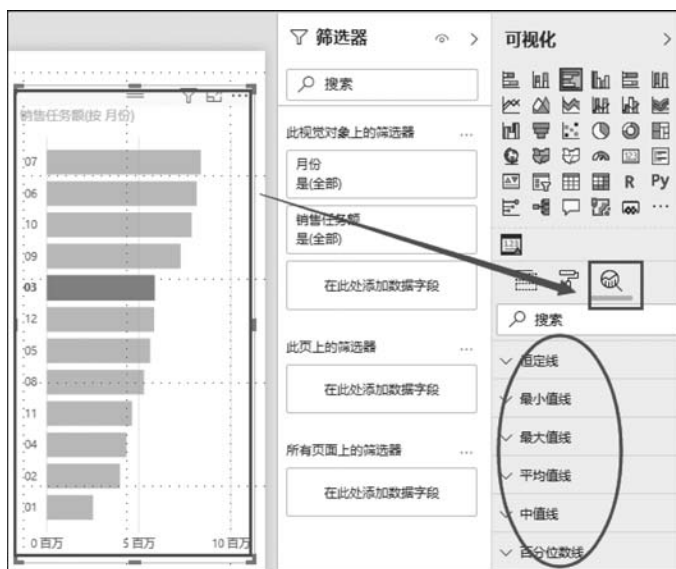


图 5-27 在分析表单中查找可视化字段

使用分析窗格,可以制作各种动态参考线,选择或制作视觉,此时从“可视化”窗格中选择“分析窗格”符号,要制作分析曲线,需要选择“+”图标来进行添加,如这里单击“平均值线”下的“+”图标。然后,用户可以通过下拉菜单的内容框来选定该曲线的属性参数,如图 5-28 所示。

还可以通过选择视图来进行预测,如果数据源中有时间数据,则可以使用“预测”功能。只需选择一个视觉对象,然后展开分析窗格的“预测”部分。可以指定多个输入以修改预测,例如预测长度或置信区间。在图 5-29 中显示了已应用预测的基线视觉对象。此时在系统中增加分析窗格的预测部分,用户可以在“预测”栏下进行设置,图中显示了连接测量的基本视线,用户可以来调整“置信区间”进行调整。

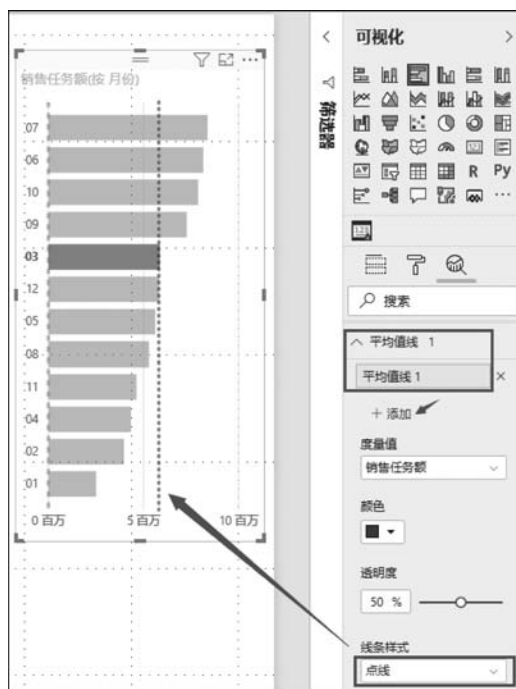


图 5-28 系统中使用分析窗格示意

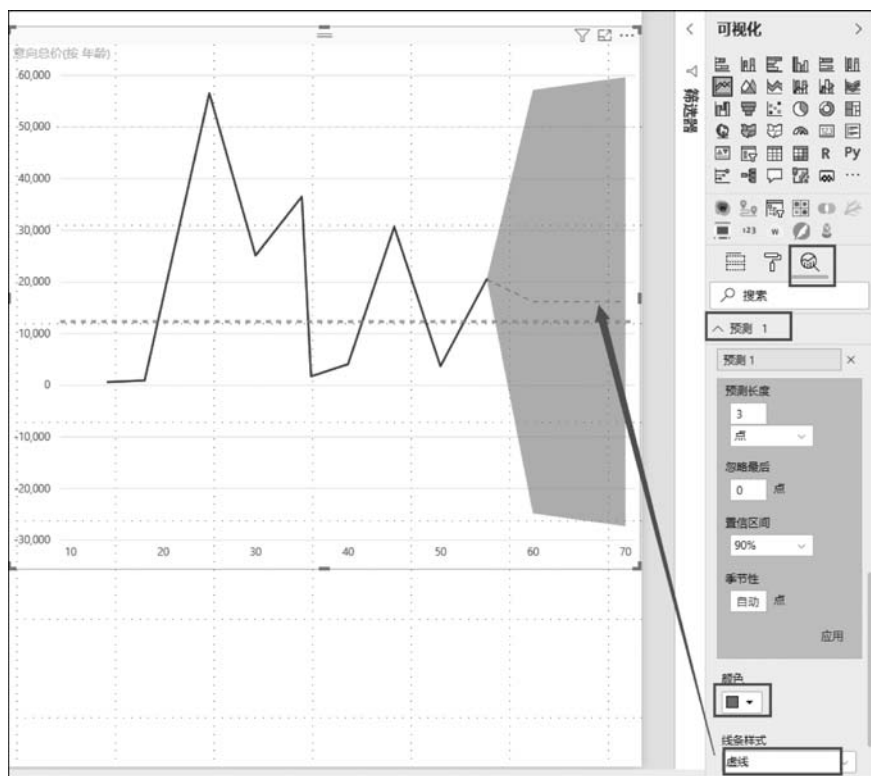


图 5-29 分析窗格的预测

动态参考曲线的呈现与否,取决于所使用的视觉对象的类型,用户可以在区域图、线形图、漫反射图、成束柱图、成束条形图视觉效果中使用动态曲线。

5.5 关系视图

导入多个表时,有时需要使用所有这些表中的数据来执行一些分析。为准确计算结果并在报表中正确显示信息,建立表格之间的关系是必需的。在大多数情况下可以使用自动检测功能来检测建立的关系。下面介绍系统桌面中关系的创建和编辑方式。

一般情况下,用户同时查询两个或多个表格时,或者在加载数据时,在加载期间系统会自动检测数据表之间的关系,并尝试为用户创建关系,自动设置基数、交叉筛选方向和活动属性。系统桌面查看表格中用户正在查询的列名,确定是否存在潜在关系。若存在,则自动创建这些关系。如果系统桌面无法确定是否存在匹配项,则不会自动创建关系。用户仍可使用“关系”对话框来创建或编辑关系。

在“主页”菜单中,单击“管理关系”图标,打开“管理关系”弹窗,单击“自动检测”按钮,如图 5-30 所示。



图 5-30 桌面菜单设置配置

默认情况下,系统桌面会自动配置新关系的基数、方向、交叉筛选器方向和活动属性。在必要时可以对其进行更改。但是,要求其中一列必须有唯一值,关系中至少有一个表必须具有连接字段值不同的唯一列。注意,如果关系选择的表均没有唯一值,会显示错误。

下面介绍如何在系统和基数中创建关系。

1. 基本概念

关系视图演示了模型中的大多数数据表、字段和表之间的连接。当模型在众多表之间具有复杂连接时,尤其有用。

(1) 关系视图符号:单击可以在关系视图中演示模型。

(2) 关系:用户可以将光标浮动到系统关系上以显示所使用的部分。双击关系将在“编辑关系”对话框中将其打开。

2. 基数

(1) 多对一($*$:1)。这是最常见的默认类型,意味着一个表中的列可具有一个值的多个实例,而另一个相关表(常称为查找表)仅具有一个值的一个实例。

(2) 一对一(1:1)。这意味着一个表中的列仅具有特定值的一个实例,而另一个相关表也是如此。

(3) 多对多关系 $*$: $*$ 。借助复合模型,可以在表之间建立多对多关系。这种方法删除了对表中唯一值的要求。

3. 在堆栈中自动检测

如果用户对两个或两个以上的表进行查询时,系统将查询表之间的关系并建立连接。系统会对表中的段名称进行判断,以确定是否存在任何潜在的连接。如果系统无法确定具有连接条件的匹配,那么自然不会产生系统关系。一般情况下用户也可以利用“管理关系”来建立或改变连接。

4. 建立关系

除了使用自动检测外,用户也可以实现手动操作来建立关系。在“主页”菜单中单击“管理关系”下的“新建”按钮,在主表的下拉列表中,选择一个表,然后选择需要在关系中使用字段。在第二个表的下拉列表中,选择系统的关系中需要的另一个表,此时选择用户需要使用的其他字段,最后单击“确定”按钮。

5. 改变关系视图

在“管理关系”中选择关系,然后单击“编辑”来改变关系视图。

6. 使关系动态化

当选中关系时,这意味着关系将作为系统中默认的动态关系填充。在两个表之间存在超过一个连接的情况下,动态关系为系统提供了一种方法,从而产生包含两个表的认知图。

5.6 钻取表中数据

钻取是指沿着层次结构(维度的层次)来查看数据。钻取可以实现变换分析数据的粒度。钻取分为上钻(Drill-up)和下钻(Drill-down)。上钻是沿着数据的维度结构向上聚合

数据,在更大的粒度上查看数据的统计信息;而下钻是沿着数据的维度向下,在更小的粒度上查看更详细的数据。例如,当前的粒度是月份,按照年份查看数据是上钻,而按照日期来查看数据是下钻,日期的数据是详细的数据,而每天的数据是高度聚合的数据。

在查看可视化图表的时候,可能想深入了解某个视觉对象的详细信息,或者进行更细粒度的分析。当图表中的数据存在层级结构时,可以在图表上直接下钻展示下一层级的数据,最常见的层级结构就是日期数据,从年度、季度、月份到日期,甚至到小时、分钟和秒,只要具体的日期数据的层次结构足够详细。例如看到 2019 年的总体数据,同时想知道每个季度甚至每个月的数据,通过 Power BI 的钻取功能,可以单击轻松实现。

下面介绍钻取功能以及如何钻取。为了能够直观地看到层次的变化,先用一个矩阵表来展现,把“时间表”中的“日期”拖入“行”框中,如图 5-31 所示。

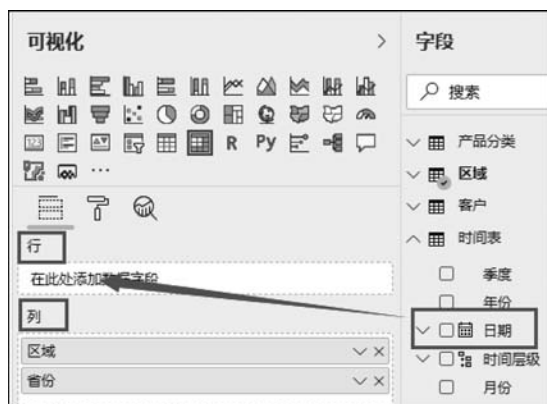


图 5-31 视图组件

Power BI 系统会自动添加时间层次结构,如果不要某个层次,单击右边的×按钮删除即可。把“销售额”也拖入值中,显示的是年度数据。

钻取到下级层级的数据有两种方式:

(1) 通过图表右上角的向下箭头“启用深化”。

(2) 使用顶部的“数据/钻取”菜单。选中图表,上方按钮区出现“数据/钻取”功能,如图 5-32 所示。



图 5-32 向下钻取数据

5.6.1 层次结构

钻取数据离不开层次结构,最常用的层次结构数据是日期维度,日期维度是自然层次结构,下层的结点只有一个父结点,如图 5-33 所示。

在系统报告视图中创建日期等级(Date Hierarchy),该层次结构由 3 个级别组成,从上至下依次是 YearKey、MonthKey 和 Date,如图 5-34 所示。

如果系统中没有内置层次结构切片,设计人员需从网上商城中下载自定义的等级切片器(Hierarchy Slicer),用于显示日期等级。显示结果如图 5-35 所示。

Date	DateKey	MonthKey	YearKey
2018-01-01	20180101	201801	2018
2018-01-02	20180102	201801	2018
2018-01-03	20180103	201801	2018

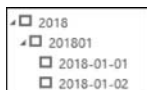
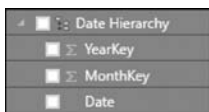


图 5-33 层次结构示例数据表 图 5-34 日期等级层次 图 5-35 日期等级层次显示结果

5.6.2 图表级别的钻取

在同一个图表上,通过钻取操作,可以方便地导航到不同的层次结构上查看数据,例如,图 5-36 所示是某数据仓库中包含的创建于 2018 年的所有 Post 的数据。

在“关系”视图中,通过创建的日期关键字和 Date 维度的 Date Key 关联起来。在“报告”视图中,通过 Line Chart 查看不同 Date 层次结构上 Post 的数量,Line Chart 的属性设置如图 5-37 所示。

PostID	PlatformID	CreatedDateKey
94006343	29	20180323
113091028	29	20180204
115847074	29	20180129

图 5-36 钻取数据表

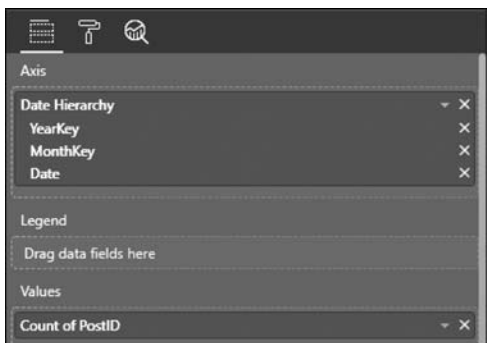


图 5-37 建立关联设置

默认的级别是顶层的 YearKey,在该级别上,Line 显示的数据是 2018 年的所有 Post 的总数,由于只有 2018 这一个年份,因此 Line Chart 只显示一个点,如图 5-38 所示。

通过“下钻”按钮(两个向下的箭头),导航到 MonthKey 级别查看 Post 的数据,在该级别上,可以看到 2018 年的所有月份的 Post 数量和增长趋势,如图 5-39 所示。

“上钻”按钮是一个向上的箭头,可以从 MonthKey 级别返回 YearKey 级别。上钻和下钻是按照层次结构逐层钻取的。用户也可以将钻取关联到其他图表上。钻取过滤默认是启用的,钻取会对其他的图表进行过滤,这就意味着,当在一个视图上进行钻取操作时,

其他视图上的数据也会被筛选。设置时需要在“Format”下,勾选 Drilling filters other visuals 复选框,如图 5-40 所示。



图 5-38 顶层数据(年)显示结果

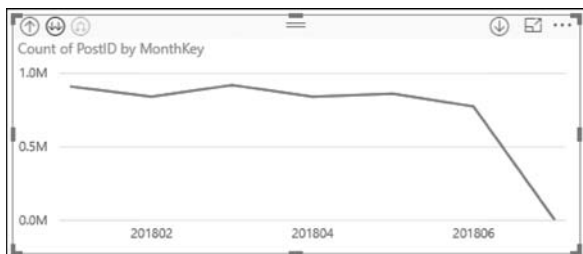


图 5-39 向下钻取数据(月)显示结果

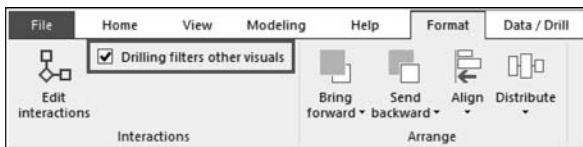


图 5-40 钻取关联其他图表设置

5.6.3 钻取

钻取(Drillthrough)允许用户在报表中创建一个页面,该页面(称作钻取页面)提供模型中单个实体的详细信息,然后在其他页面中引用该实体,通过使用数据点从当前页面导航到钻取画布上,并把筛选上下文传递到钻取画布上。

筛选上下文分为钻取筛选器和页面级筛选器,钻取筛选上下文是指拖放到钻取中的字段。当启用 Keep all filters 时,系统把当前页面中所有筛选器的上下文应用到钻取到的画布上;当禁用 Keep all filters 时,系统只把钻取筛选器的上下文应用到钻取到的画布上。Keep all filters 的默认设置是 Off,不把当前页面级筛选器上下文传递到钻取页面。当设置 Keep all filters 为 On,导航到钻取页面时,用户可以从钻取页面中查看传递到钻取页面的所有筛选上下文。

钻取是通过相同的字段实现的,在设计钻取时,需要在钻取画布上设置钻取筛选的字

段,源画布上的数据点也包含该字段。

1. 类别钻取

用户可以在 Fields 中设置钻取筛选器,在钻取画布上把 MonthKey 字段设置为钻取字段,用作类别,如图 5-41 所示。



图 5-41 类别钻取设置

选中一个数据点,该数据点的轴是 MonthKey(用作钻取的字段),右击,在弹出的快捷菜单中选择 Drillthrough 命令,导航到钻取页面,筛选器的上下文就会被引用到钻取页面,用户看到的实体详细信息是被筛选之后的数据,如图 5-42 所示。

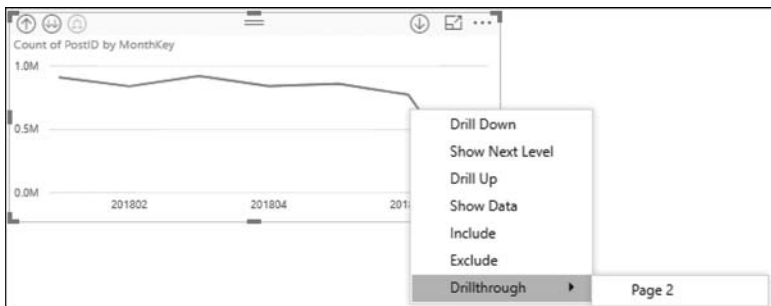


图 5-42 类别钻取被引用到钻取页面设置

2. 度量钻取

度量钻取(MeasureDrillthrough)是指把独立或汇总数字列传递到钻取页面中,之后把数字列用作类别或汇总时允许钻取。例如,对 PostID 进行钻取,设置当对 PostID 进行聚合时允许钻取,如图 5-43 所示。

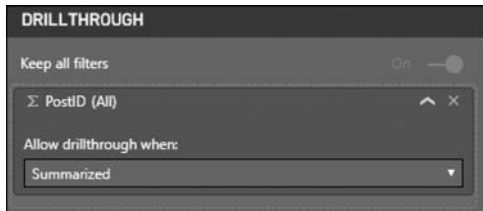


图 5-43 度量钻取设置

选中一个数据点,右击,在弹出的快捷菜单中选择 Drillthrough 命令,导航到钻取页面,从 DRILLTHROUGH 对话框显示的列表中查看所有传递到钻取页面的筛选上下文,如图 5-44 所示。



图 5-44 类别钻取被引用到钻取页面设置

小结

本章主要介绍了数据建模,重点掌握基数的概念,以及它建立的一对一、一对多的关系表。掌握交互关系以及它们关系的传递等设计与操作。学会数据建模,并掌握进行计算列、计算表等技能;了解桌面关系视图,并学会钻取表中的数据。

问答题

什么是多对多关系?如何在 Power BI 系统中解决这些问题?

实验

1. 打开一个报表,在表格间找出一对一、一对多关系的元素,更改它们的关系,查看其引起的后果。
2. 构建一个报表,并进行计算列、计算表的操作。