SAP Lumira实验指导

SAP Lumira 是一款数据操作和可视化工具。用户可以连接到一个或多个数据源以创建数据集，从而使用多种图形统计图和表格在数据集中操作、清理和可视化数据。SAP Lumira 安装在本地计算机上，可以操作远程或本地数据。用户可以保存基于数据集构建的统计图，并且可以通过电子邮件将其发送出去。用户可以将数据集发布到 SAP HANA、SAP Explorer、SAP Lumira Cloud 以及 SAP Jam 活动。

本实验利用SAP Lumira Personal Edition（下载地址为：https://global.sap.com/campaign/ne/free\_trial/visual\_intelligence/index?kNtBzmUK9zU）对GBI公司的销售情况进行分析，简单介绍使用SAP Lumira操作、可视化数据的一般步骤。

1. 准备数据

（1）导入GBI销售数据。“文件”🡪“新建”，在弹出的“新建数据集”面板上选择“Microsoft Excel”作为数据源，如图1所示。同时可以从最近使用过的文件中进行快速选择。

单击下一步，在弹出的面板上单击“文件”一行右侧的“浏览”按钮，选择GBI的Excel文件，接着选择需要导入的工作表“GBI sales”，如图2所示。编辑数据集名称，同时可单击“高级选项”设置“显示隐藏行”和“显示隐藏列”。单击“创建”按钮完成数据导入。

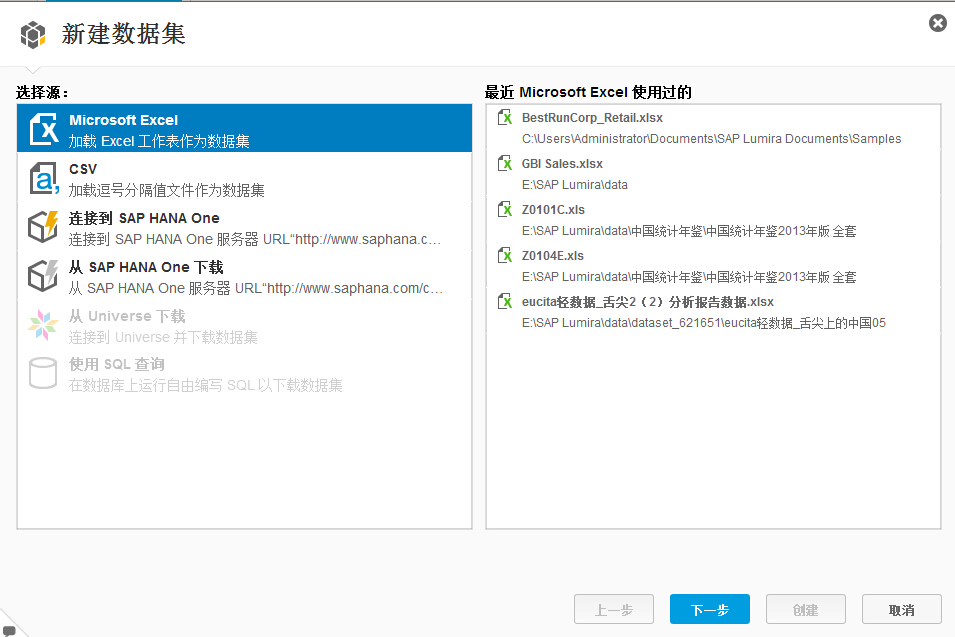


图1 创建新的数据集

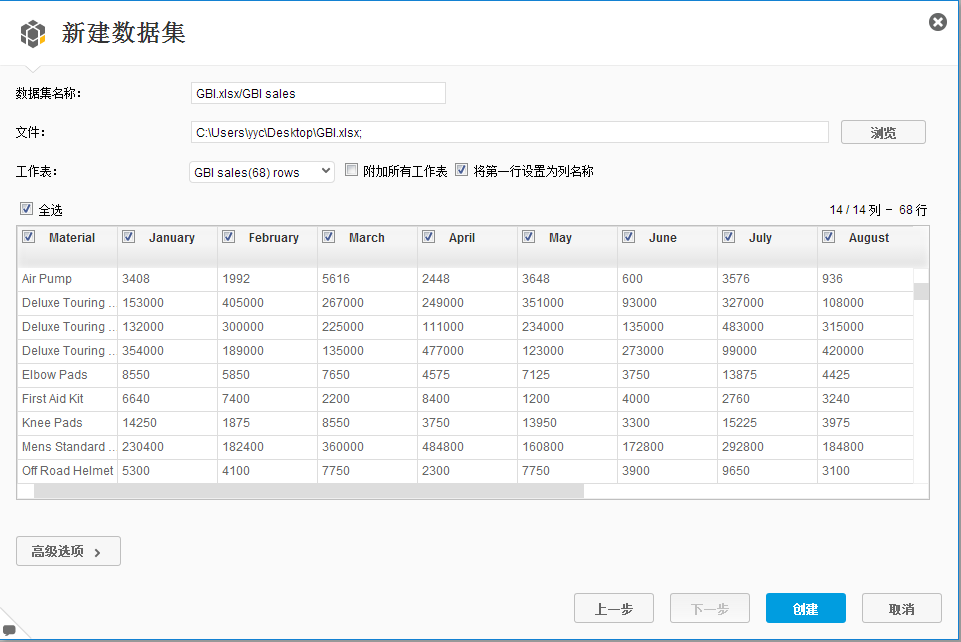
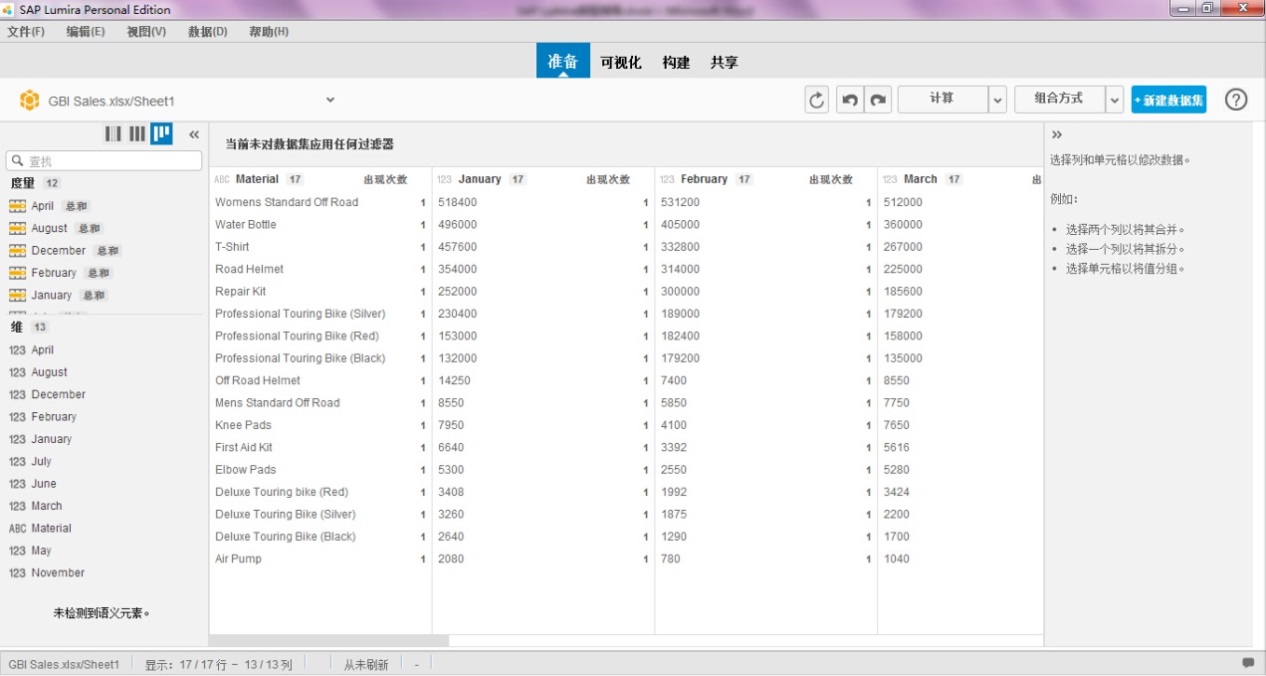


图2 导入数据

（2）编辑数据集。点击系统界面正上方的“准备”按钮，查看之前导入的数据。左侧为对象选择器，可以设置度量和维的各种属性，如显示格式、数据类型等；其中有三个矩形小图标 ，在此设置工作面板数据的显示格式，选择“构面”，系统自动统计各个列不同值的出现次数，如图3所示。中间为工作面板，可以在准备、可视化、构建和共享四个操作间进行转换；点击刷新按钮 ，同步更新数据集。右侧为操作工具栏，可对列和单元格进行一系列操作。



操作工具栏

对象选择器

图3 准备界面

（3）选择GBI sales数据集，添加全局过滤器。在“准备”选项下添加的过滤器对所有可视化对象均有过滤效果，以维“Material”为例。将鼠标移至该列，点击“出现次数”右侧出现的设置按钮 ，选择“过滤”，弹出图4所示的面板。在查找框中输入特定值随之显示搜索结果，点击记录“Elbow Pads”并选择“排除值”，单击“确定”，新增图5所示的全局过滤器。单击该过滤器的图标返回编辑面板，删除此图标即撤销此次过滤操作。

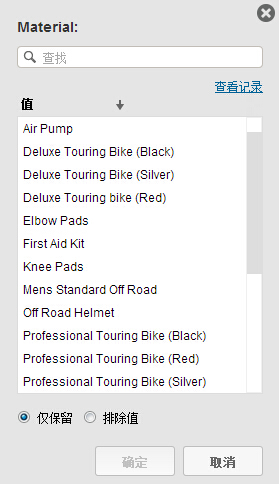


图4 编辑过滤器



图5 排除属性“Elbow Pads”

（4）添加新的数据。工具栏“数据”🡪“添加”，或者单击“新建数据集”按钮，导入GBI transactions工作表，如图6所示。重复操作，导入该文件的其他几个表格数据。

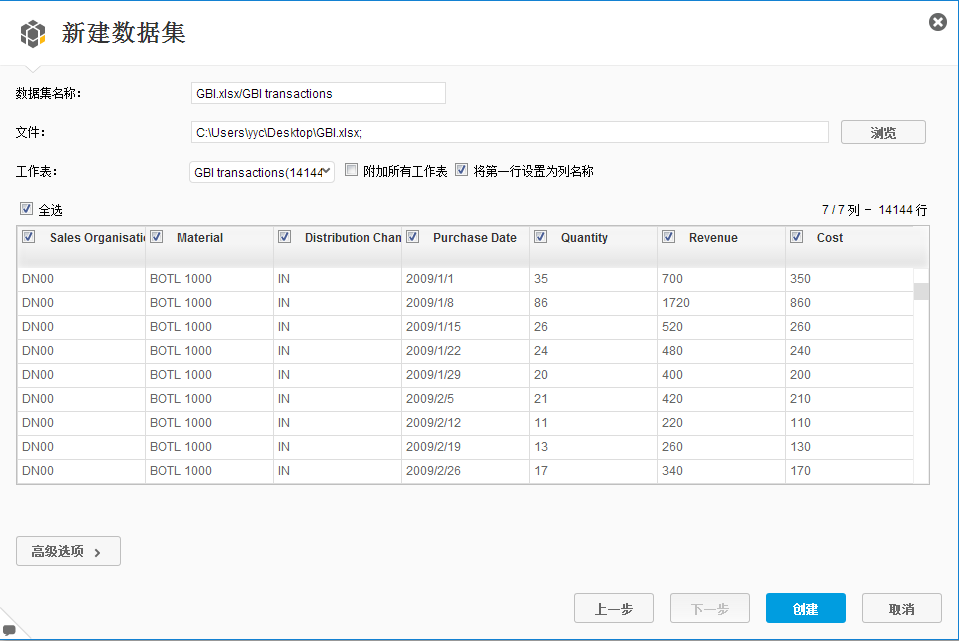


图6 添加GBI transactions数据集

（5）新建计算所得维。选择GBI sales数据集，“准备”🡪“计算”🡪“新建计算所得维”，或者在“对象选择器”中选择对应维，“选项”🡪“创建计算所得维”，弹出新建计算所得维面板如图7所示。

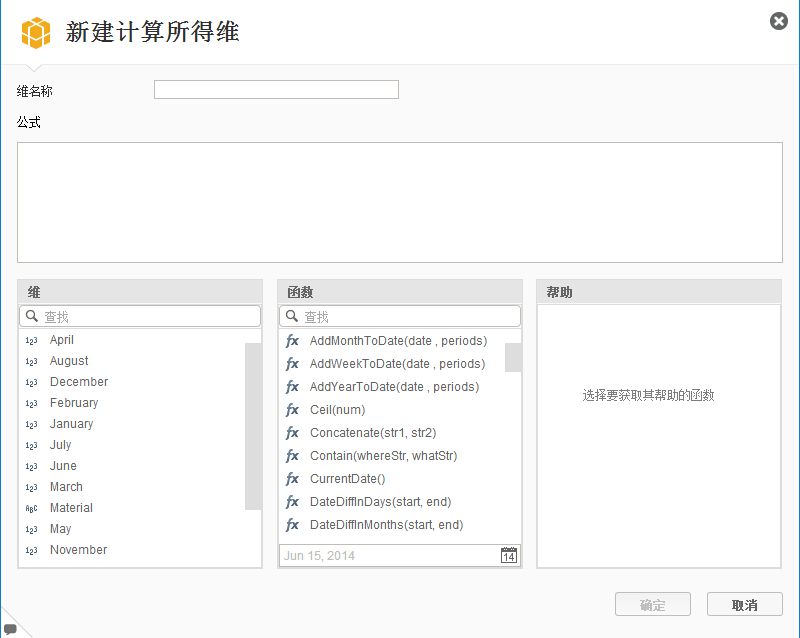


图7 新建计算所得维

Lumira提供公式语言创建公式并将其应用到列值。在列上创建公式时，系统将利用受该公式影响的值创建一个新列。在此选择函数GroupValues，在公式框中输入GroupValues( {Material} ,["Elbow Pads","Knee Pads"],"important")，维名称输入“备注”，操作结果如图8所示。新建了一个名为“备注”的列并满足条件：使得Material值为“Elbow Pads”或者“Knee Pads”的记录该项值为“important”。此外，Lumira还支持聚合函数、字符函数、日期和时间函数等操作。

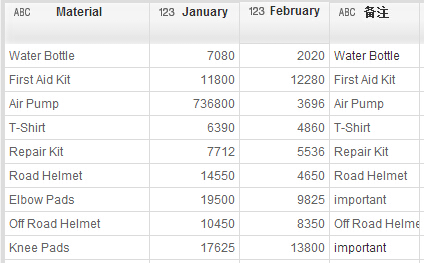


图8 使用GroupValues新建列

（6）新建计算所得度量。“准备”🡪“计算”🡪“新建计算所得度量”，或者在“对象选择器”中选择对应度量，“选项”🡪“创建计算所得度量”。同理可得满足相应函数条件的新度量。

（7）按名称创建地理层次结构。选择GBI xcelsius workshop数据集，对象选择器🡪“Sales Organisation”🡪“选项”🡪“创建地理层次结构”🡪“按名称”，弹出地理区域数据面板如图9所示。单击“确认”，可以看到按国家名称解析的地理层次结构，点击“编辑”按钮进行修改。



图9 创建地理层次结构

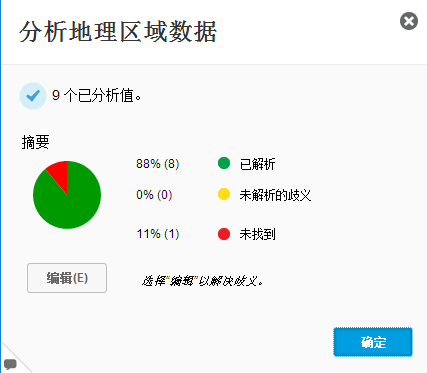


图10 分析地理区域数据

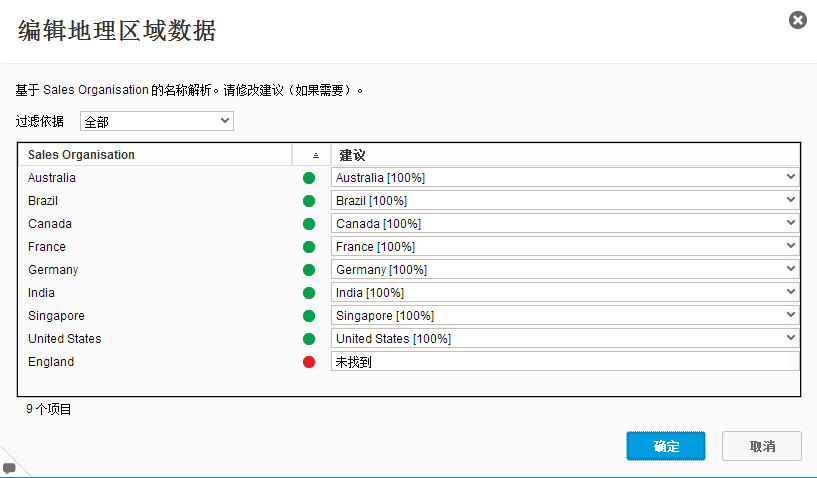


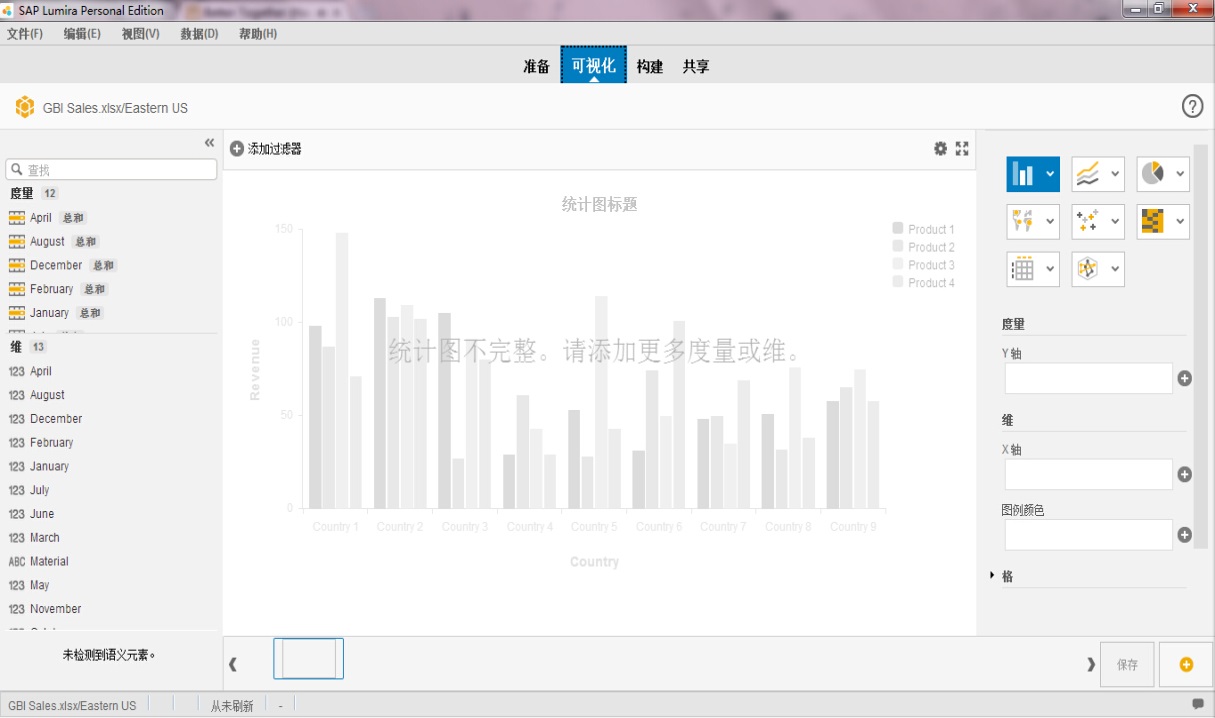
图11 创建地理层次结构

可以在“过滤依据”一栏对“已解析”“未解析的歧义”“未找到”三个属性进行筛选再编辑。点击“确定”返回之前的界面，地理层次结构创建成功。

（9）创建时间层次结构。与地理层次结构类似，单击对象选择器🡪对应时间的维🡪“选项”🡪“创建时间层次结构”即可。在此选择GBI transactions数据集的“Purchase Date”创建时间层次结构。

1. 可视化数据

可视化界面如图12所示，左侧为对象选择器，中间为工作面板。“可视化”选项下，右侧为图的属性控制区域，可以选择创建图表的类型、X轴和Y轴的属性等。



切换数据集

维

度量

图12 可视化操作界面

（1）查看各个值之间的差别以及度量的分类比较。选择GBI sales数据集，为了在第二季度及时补货并做出正确的销售策略，使用柱形图比较分析第一季度各类商品的销售情况。单击界面右侧按钮，选择“柱形图”。在此图标下方的面板中，点击Y轴属性框选择度量“January”“February”和“March”；点击X轴属性框选择维“Material”，或者将操作界面左侧相对应的度量和维直接拖入X、Y轴的属性框中，最后单击界面右下角的“保存”按钮。

由图13可知，Deluxe Touring Bike 系列、Professional Touring Bike系列、Standard Off Road系列在第一季度的销量远远超出其他商品，因此需要保证这几类商品货源充足，防止因缺货导致销售机会的流失。

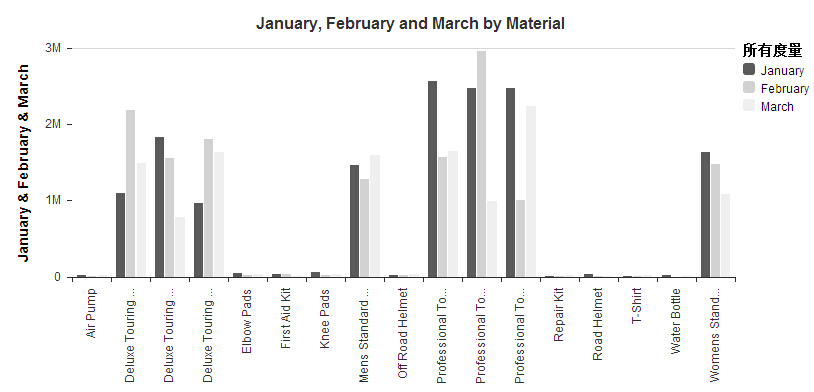


图13 柱形图a

通过数据过滤，抽取信息后再进行比较。单击图13中矩形框所示的维，选中后出现浮标，点击“过滤器”得到抽取后的信息（图14），注意此过滤器只用于筛选该柱形图的数据，不影响其他可视化对象。由图知Silver款的Deluxe Touring Bike销量处于下滑状态，需要加强Silver的营销策略并在存货方面做出相应调整。

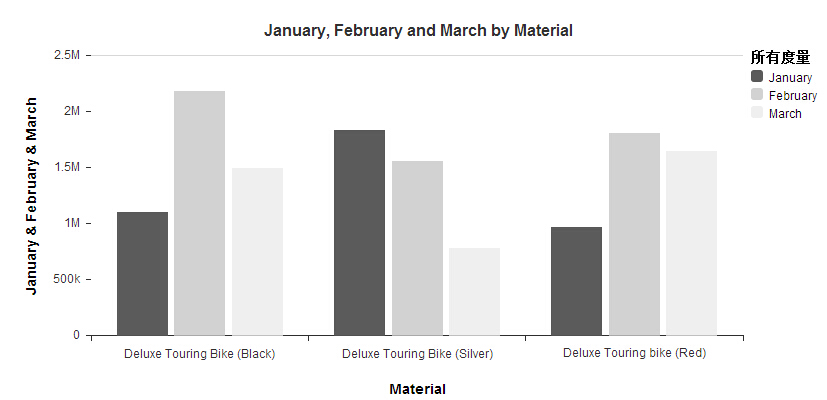


图14 柱形图b

雷达图同样用于各个维和度量间的比较。选择GBI sales(EUS)数据集，即Eastern US的销售数据，单击按钮 ，选择“雷达图”，Y轴设为“T-shirt”和“Water Bottom”，雷达图分支设为“月份”，结果如图15所示。一般来说，同心圆中最小的圆代表同行业平均水平的1/2值或最差的情况；中心圆代表同行业的平均水平或特定比较对象的水平，称为标准线（区）；大圆表示同行业平均水平的1.5倍或最佳状态。由图可知，T-shirt全年的销售情况中三月、七月和九月销量是最好的，高于平均水平；而Water Bottom 的销量则明显不及T-shirt，尤其是二月和八月销售情况很不理想，需要引起相关人员的重视。

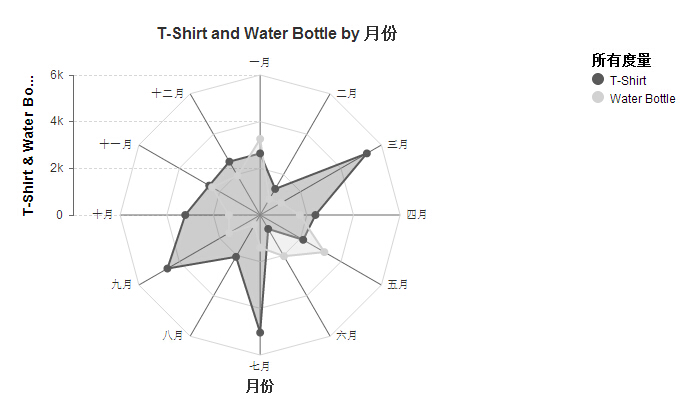


图15 雷达图

另外，标签云图可直观比较数据间的差异，如图16所示。选择GBI sales数据集，单词权重设为“June”，即单词大小代表六月各类商品销量的多寡；单词颜色设为“July”，即颜色深浅表示七月销量的高低。由图可知在六、七两个月份Professional Touring Bike系列红色和黑色款销量最好。



图16 标签云图

（2）显示数据值的趋势（尤其是基于时间的维），查看数据的级数以及可能具有的模式。选择GBI sales(EUS)数据集，使用折线图可视化Eastern US地区Standard Off Road系列产品近一年的销售走势。选中数据集，单击右侧的按钮并选择“具有双Y轴的折线图”。Y轴1设为“Mens Standard Off Road”，Y轴2设为“Womens Standard Off Road”，X轴选择“月份”，结果如图17所示。

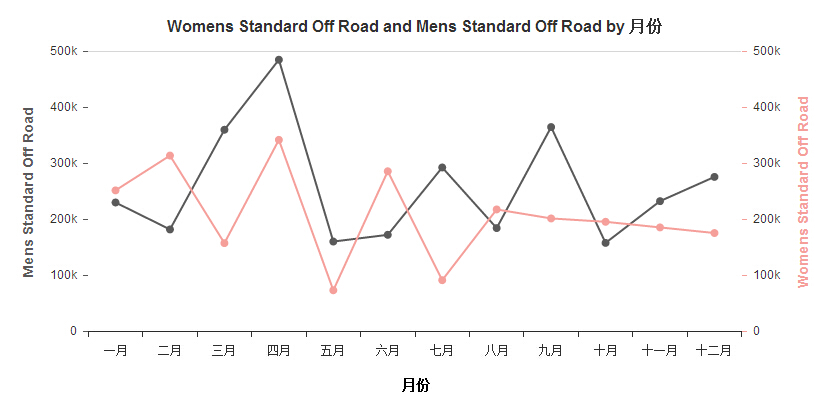


图17 具有双Y轴的折线图

由图可知Mens款二、三、四月销量持续较大幅度的上升，然而四月至五月Mens款和Womens款销量皆急剧下跌，尤其是Womens在五月销量跌至谷底，到了九月Mens款销量明显反弹。就整体而言，上半年两款商品的销量起伏波动较大，下半年相对来说更为平稳。

柱形折线组合图同样用于显示数据值的趋势。选取GBI transactions数据集，新建一个可视化对象，在之前折线图的选项下选取“具有双Y轴的柱形折线组合图”。Y轴1设为“Quantity”，Y轴2设为“Revenue”，X轴设为“月份”，并对Material添加过滤器。最终Standard Off Road系列产品的销量和收益发展趋势如图18所示。

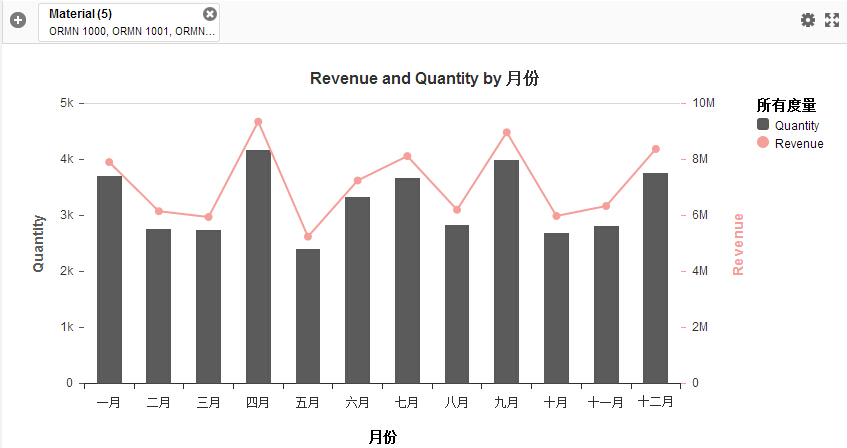


图18 具有双Y轴的柱形折线组合图

（3）使用图例显示百分比和总值。探究2010年不同类型单车收益所占比重情况，采用饼图做可视化分析。选取GBI xcelsius workshop数据集，单击右侧的按钮，选择“饼图”。饼图扇区选择“Revenue”，并设置“升序排序”，图例颜色选择“Product Group”，结果如图19所示。该图按照收益的大小依次显示了2010年GBI公司各类单车的销售收入及百分比，可以看出Mountain bike和Racing bike的销售收入占了总收入的2/3左右，Kids bike的收益最少。

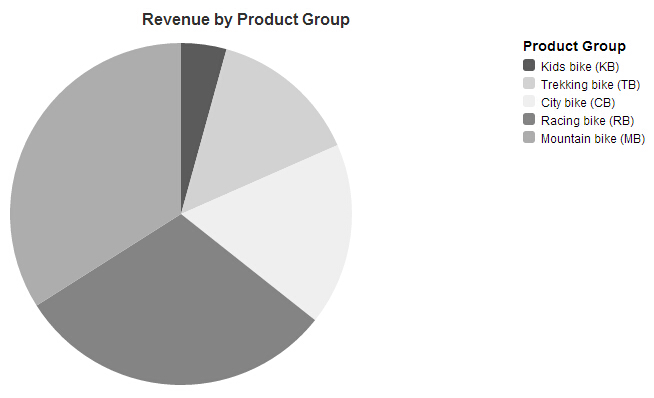


图19 普通饼图

为了更全面地反映各类单车的销售情况，新建一个“带纵深效果的饼图”。饼图扇区和图例颜色保持不变，新增的属性“饼图深度”选择“Quantity”（图20）。由图可知各扇形区域的高度差异非常小，观察数据可以发现收益相对于销量而言非常巨大，因此不同类型商品销售数量上的差异在该图中体现得并不明显。

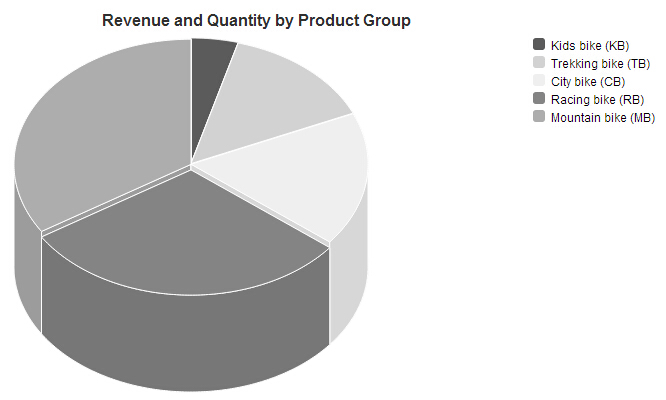


图20 带纵深效果的饼图

显示百分比和总值的可视化对象还有堆积柱形图（图21）。在此Y轴设为“Revenue”，X轴设为“Month”，图例颜色设为“Product Group”。由图可知各类单车在不同月份的收益分布情况，作为图19、20的细节信息补充。

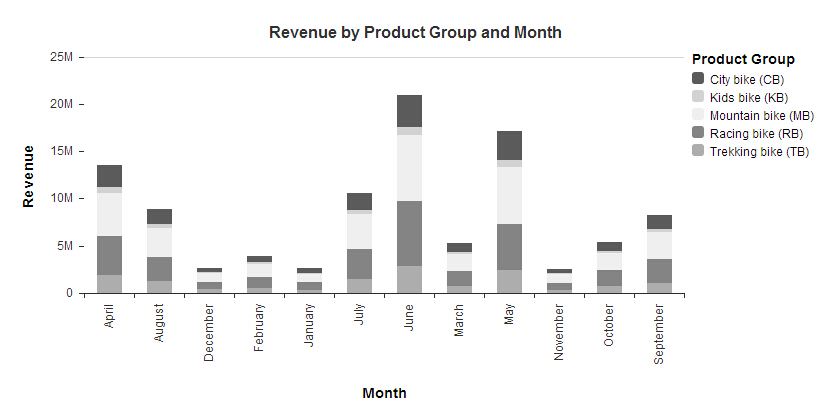


图21 堆积柱形图

树图不仅可以展示部分相对于整体的比重，还能很好展示不同关键词的热度。单击面板右侧 按钮，选择“树图”，“区域颜色”设为“Quantity”，即颜色越深销量越大；“区域权重”设为“Revenue”，即面积越大收益越高；“区域名称”设为“MONTH”。由图22可知三、四、六、七月是销售旺季，尤其是6月占有非常大的市场份额。

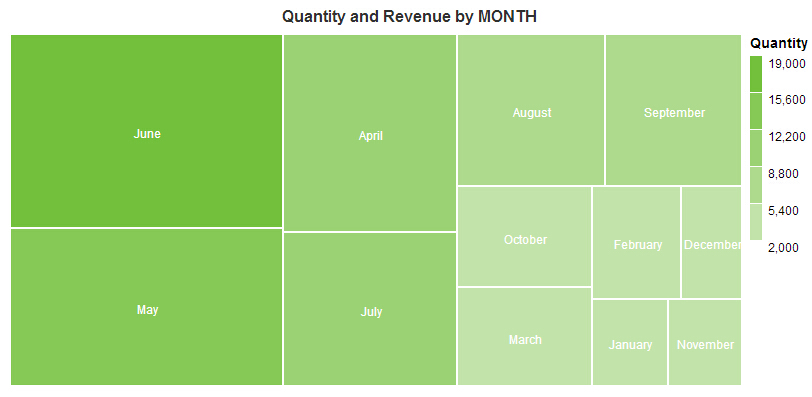


图22 树图

（4）查看两个度量之间的相关性，了解第一个度量对第二个度量的影响或者比较多个度量值。使用散点图探究收益与其他影响因素之间的关系：单击右侧按钮 ，选择“散点图”。Y轴1设为“Revenue”，表示营业额；Y轴2设为“Quantity”，表示数量；图例颜色选择“Product Group”和“Sales Organisation”，其中Product Group表示所销售的单车类型，Sales Organisation表示各个销售组织；图例形状选择“Distribution Channel”，有Wholesale和Retail两种类型，在图中分别用圆形和矩形标出。图23展示了随着销售数量的增加，不同类型的商品在各个销售区域的营业额分布情况。

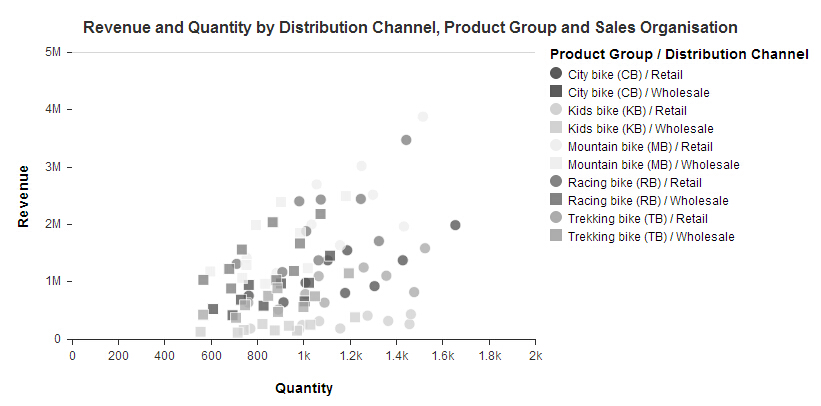
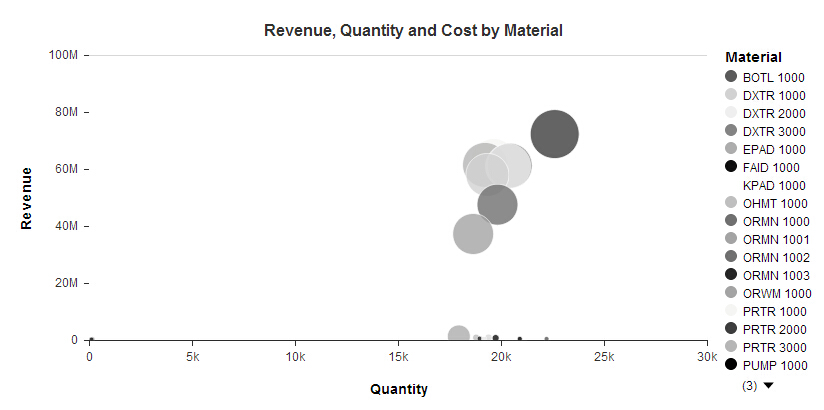


图23 散点图

使用气泡图观察不同商品间收益、销量与成本之间的关系。选择数据集GBI transactions，单击之前选择散点图的按钮，在出现的下拉选项中选择“气泡图”。X轴设为“Quantity”，Y轴设为“Revenue”，气泡宽度设为“Cost”，图例颜色设为“Material”，图24展示了各类商品的销量、收益与成本之间的关系。由图可知在相近的成本条件下PRTR2000即Professional Touring Bike (Silver)，其收益和销量都领先其他产品；EPAD1000即Elbow Pads在收益和销量相近的条件下，成本远高于其他产品。因此为了获得更高的利润，建议增加Professional Touring Bike (Silver)的存货量和促销力度，降低Elbow Pads的市场份额。



EPAD1000

PRTR2000

图24 气泡图a

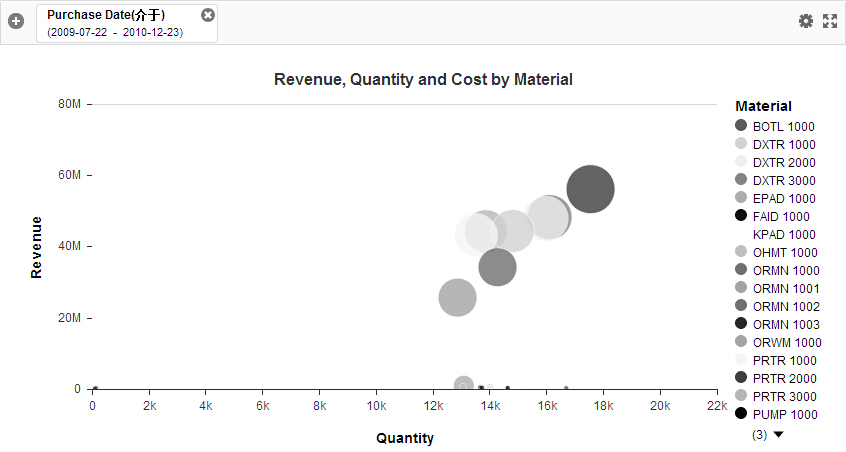
为了观察更细致的数据，使用切片和切块操作。切片与切块是指选定一部分维度后，关心度量数据在剩余维度上的分布。切片与切块的目的是通过对数据的过滤，使用户关注局部信息，Lumira通过对不同维度的数据添加过滤器实现多维数据的切片和切块。

对购买日期添加过滤器。单击“添加过滤器”按钮，选择维“Purchase Date”后弹出如图25所示面板，滑动操纵杆或者在日历中选择日期设定过滤区间，完成对购买日期的切块。在此选择下半年的销售区间以了解近期的销售状况。



图25 对Purchase Date添加过滤器

由图26可知，DXTR2000和DXTR3000，即Deluxe Touring Bike 的Silver款和Red款在下半年的销售中销量和收益都高于相近成本下的其他商品，仅次于Professional Touring Bike (Silver)；而ORWM1000即Womens Standard Off Road的销量和收益明显落后了，因此需要加大对该商品的促销力度，并查看存货和产品质量是否存在问题。



切块

DXTR2000/3000

ORWM1000

图26 气泡图b

对数据进一步过滤。单击“+”按钮，选择维“Cost”，筛选出成本介于83000-148500的商品，滑动操纵杆到这个范围的成本区间并单击“确定”，从而实现对数据的切片操作。结果如图27所示：



切片

图27 气泡图c

（5）使用地理区域图直观展示数据与地理信息之间的关系，我们已经提前创建好了地理层次结构。选择数据集GBI xcelsius workshop，单击右侧按钮，选择“地理区域分级统计图”，值选择“Quantity”，地理区域选择“国家/地区”，并且添加过滤器Product Group，仅保留City bike的值。City bike近一年销量的地理分布如图28所示，单击按钮“+”“-”或者滚动鼠标滚轮可以放大、缩小显示区域。

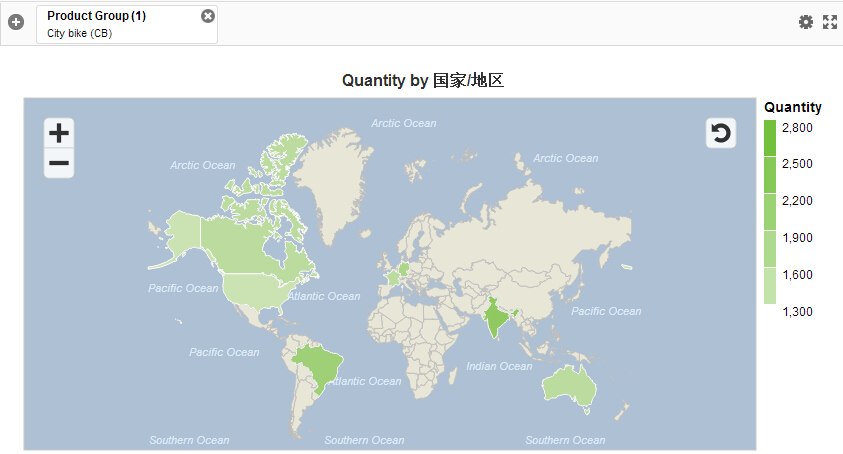


图28 地理区域分级统计图

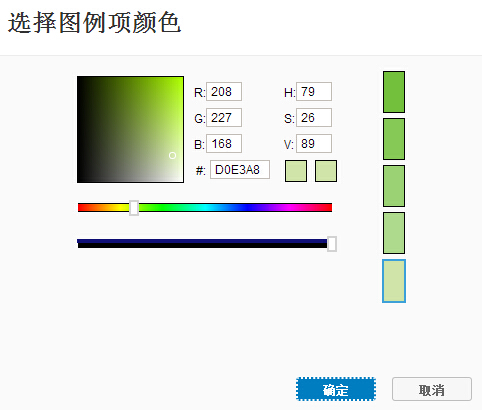


图29 选择图例项颜色

选择“添加过滤器”右侧的“设置”🡪“选择图例项颜色”，设置某一特定图例项的颜色，如图29所示。

1. 创建故事板

点击“构建”按钮，故事板的工作界面如图30所示。将想要展示的维和已经创建好的统计图直接拖入指定区域即可构建故事，在界面右侧的选项面板中设置故事名称、故事板标题、自定义字体、颜色和背景图片。



选项面板

图30 创建故事板工作界面

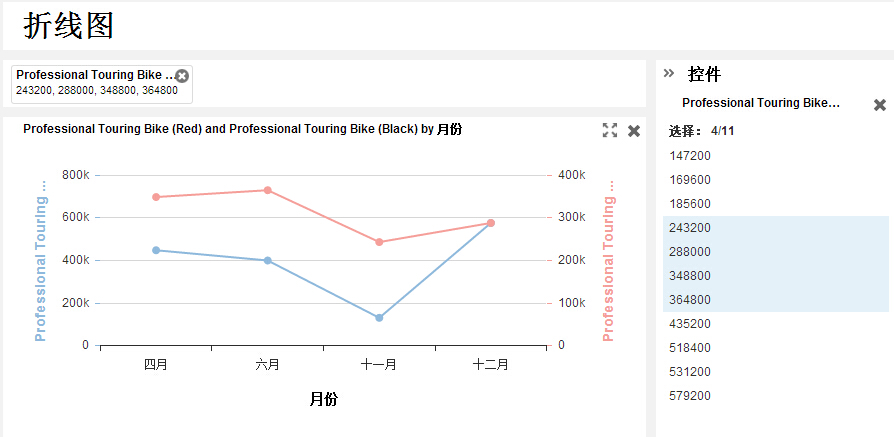


图31 筛选故事板中的数据

在“控件”一栏中，对展示在故事板中数据进行筛选。如图31线框所示部分，单击具体数值即可选择或者取消选择。

1. 共享

（1）选择Lumira中的数据集导出为CSV/Excel格式的文件。进入“共享”的操作界面，选择一个数据集，点击按钮“导出为文件”后弹出面板如图32所示。根据需要选择对应的选项，点击导出，输入文件名和选择保存类型保存即可。

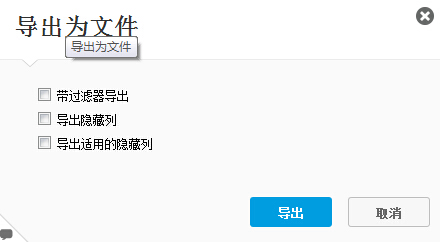


图32 导出文件面板

（2）发布到SAP Lumira Cloud和SAP Lumira Server。项目中的故事和可视化对象，如各种统计图，在联网的条件下可以发布到Cloud或者Server中保存。