**任务五 实战演练：数据集成、数据可视化**

**一、数据集成**

从广义上来说，在企业中，由于开发时间或开发部门的不同往往有多个异构的、运行在不同的软硬件平台上的信息系统同时运行，这些系统的数据源彼此独立、相互封闭，这使得数据难以在系统之间交流、共享和融合，从而形成了“信息孤岛”随着信息化应用的不断深入，企业内部、企业与外部信息交互的需求日益强烈，急切需要对已有信息进行整合，联通“信息孤岛”，共享数据信息，这些信息数据整合的一系列方案被称为数据集成。

**任务目标：**按照目标需求进行数据源分析，要分别从省份和大区的维度统计销售额，能够将三张数据表关联，形成所需的全部分析数据。

**任务实现方式**：Python语言实现。

**任务流程：**1.进行数据源准备；2.使用分析云，进行销售数据关联；3.使用分析云，进行华扬联众利润表与资产负债表关联；4.使用分析云，将华扬联众与引力传媒利润表合并；5.使用Python代码编辑器进行数据关联及数据合并。

**任务指引：**数据资源下载>上传至分析云>新建数据关联或数据合并>执行>保存关联后的数据集

**任务准备：**完成任务所涉及的数据源可在资源下载处进行数据源下载。

**任务1:销售数据准备**

在数据清洗任务中，我们将销售数据表进行了数据清洗，现在需要将清洗后的销售数据表与资源下载处下载到的“城市表”、“省区表”、“销售数据\_清洗后”上传分析云。

**操作步骤：**

1.在资源下载处，下载“城市表”、“省区表”、“销售数据\_清洗后”文件；如图3-1、图3-2所示，资料见资料包-城市表、省区表、销售数据\_清洗后文件。

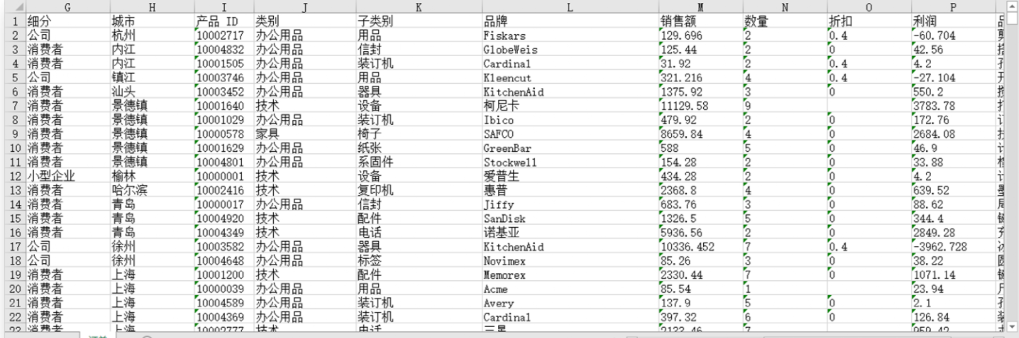


图3-1 销售数据表



图3-2 城市表和省区表

2.分别将三张表上传到分析云，如图3-3所示。

备注：销售数据\_清洗后数据表有很多sheet，本任务只需上传“订单”sheet即可。

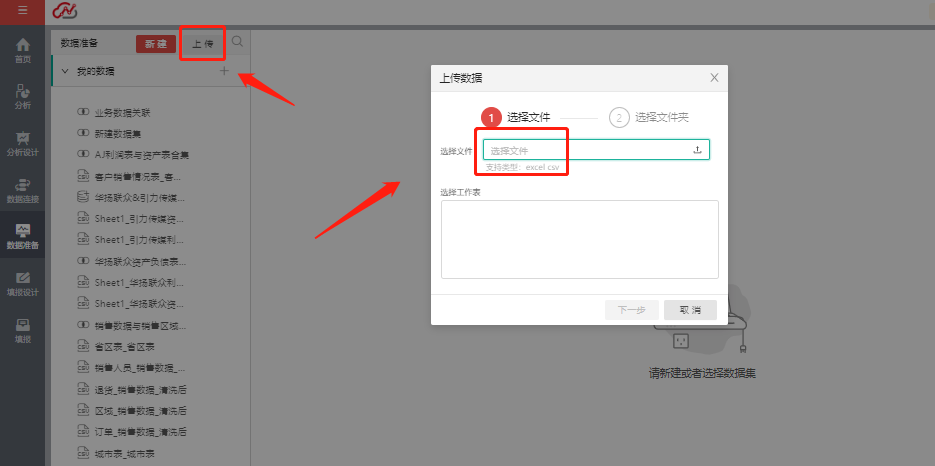


图3-3 上传数据源操作图

**任务2：销售数据关联**

在数据清洗任务中，我们将销售数据表进行了数据清洗，现在需要将清洗后的销售数据表与“城市表”、“省区表”在分析云中进行关联。

**操作步骤：**

1.新建数据集

（1） 在分析云的【数据准备】菜单下找到【新建】按钮，单击【新建】；

（2） 在弹出窗口中选择【关联数据集】；

（3） 填写关联后的数据集名称：业务数据关联；

（4） 选择数据集所在的文件夹位置：保存至“我的数据”；

（5） 单击【确定】；

2.添加关联数据表

（1） 将清洗后的销售数据表、城市表、省区表三张表分别拖动放置到关联面板内；

（2） 单击清洗后的销售数据表后再单击城市表，在连接条件里设置为连接方式为“左连接”，关联字段为“城市=城市”，如图3-4所示；

（3） 单击城市表后再单击省区表，在连接条件里设置为连接方式为“左连接”，关联字段为“省自治区=省自治区”；

3.检查数据集

（1） 单击右上角的【执行】；

（2） 查看数据结果预览，检查数据是否正确；

（3） 将数值型的字段单击字段名称前的小图标，切换成“123”类型（备注：123 代表数值型），如图3-5所示；

4.保存数据集

（1） 再次单击【执行】按钮；

（2） 单击【保存】。

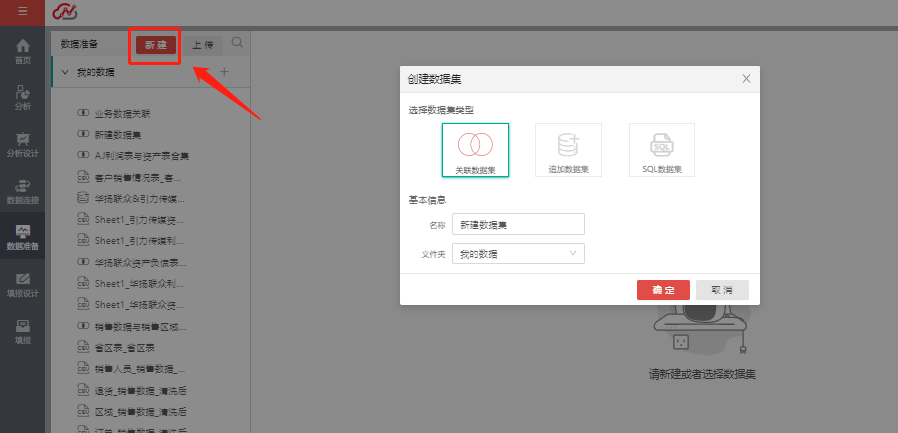


图3-4 添加关联数据表操作图

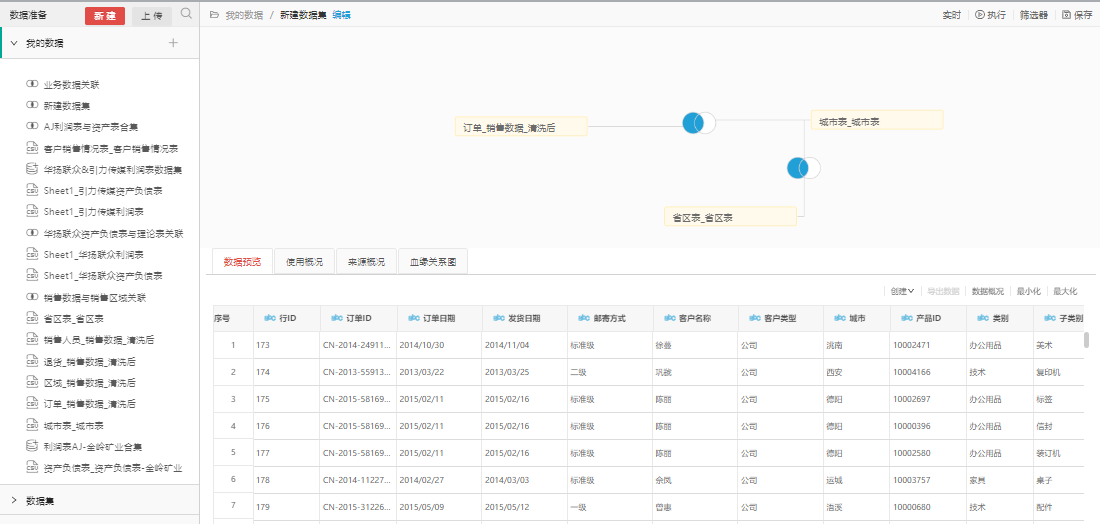


图3-5 检查数据集操作图

**任务3：华扬联众数据上传**

请将华扬联众财务报表上传分析云。

**操作步骤：**

（1） 在资源下载处，下载“华扬联众财报”；资料见资料包-华扬联众利润表、华扬联众资产负债表文件；

（2） 将财报上传到分析云。

**任务4：华扬联众数据表关联**

请将华扬联众利润表与资产负债表进行关联。

**操作步骤：**

1.新建数据集

（1） 单击【数据准备】；

（2） 单击【新建】；

（3） 选择数据类型为“关联数据集”，如图3-6所示；

（4） 将数据集名称命名为“华扬联众利润表与资产表合集”；

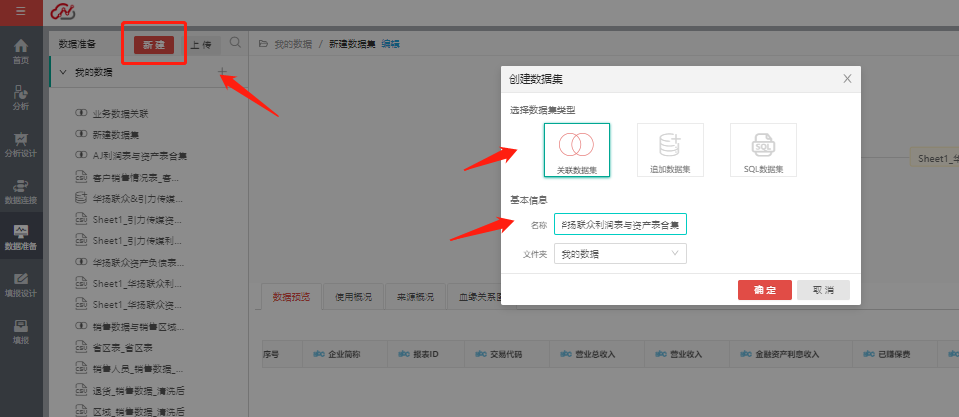


图3-6 新建数据集操作图

2.数据关联

（1） 分别拖拽资产负债表和利润表到数据预览区域；

（2） 单击两个需要关联的表进行连接，选择“内连接”；

（3） 关联条件：报表年份；

（4） 单击【确定】，如图3-7所示。

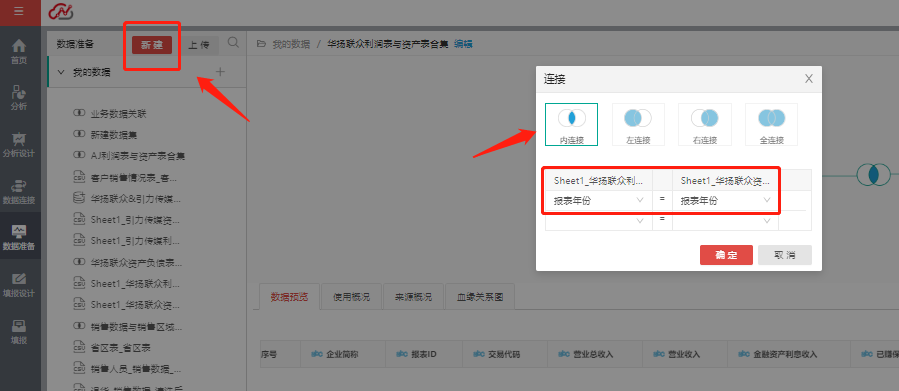


图3-7 数据关联操作图

3.数据集保存

（1）单击【执行】；

（2）单击“实时”中的数据物化，如图3-8所示；

（3）单击【保存】

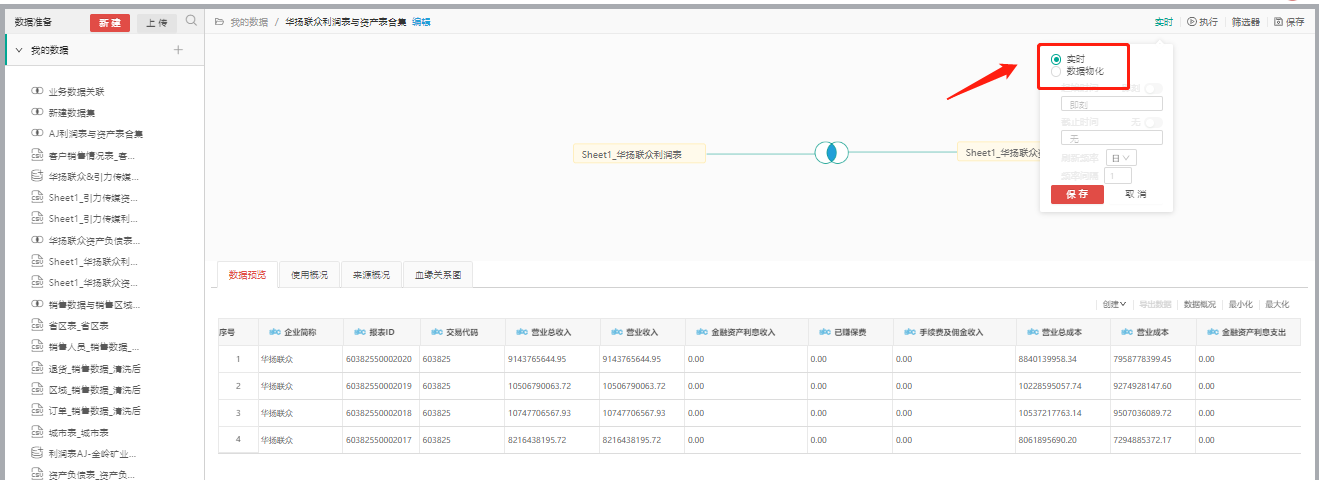


图3-8 数据集保存操作图

**任务5：引力传媒分析数据上传**

请上传引力传媒利润表与资产负债表。

**操作步骤：**在资源下载处，下载引力传媒公司的“利润表”和“资产负债表”；资料见资料包-引力传媒利润表、引力传媒资产负债表文件；分别将利润表和资产负债表上传到分析云。

**任务6：企业数据合并**

将华扬联众与引力传媒利润表合并。

**操作步骤：**

1.新建数据集

（1） 在分析云的【数据准备】菜单下找到【新建】按钮，单击【新建】；

（2） 在弹出窗口中选择【追加数据集】，如图3-9所示；

（3） 填写关联后的数据集名称：华扬联众&引力传媒利润表数据集；

（4） 选择数据集所在的文件夹位置：保存至“我的数据”；

（5） 单击【确定】；

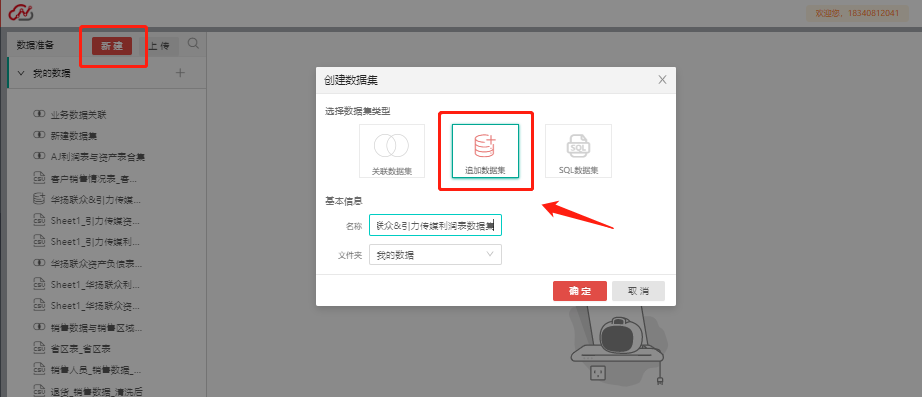


图3-9 新建追加数据集操作图

2.创建数据集

（1） 将华扬联众利润表托拉拽至右上方空白区域内；

（2） 选择所需字段，可以全选，也可以按需选择；

（3） 将引力传媒利润表托拉拽至右上方空白区域内；

（4） 选择所需字段，需要匹配华扬联众利润表所选字段进行匹配选择，如图3-10所示；



图3-10 创建追加数据集操作图

3.保存数据集

（1） 单击右上角的【执行】；

（2） 查看数据结果预览，检查数据是否正确；

（3） 单击【实时】，单击【数据物化】，将数据表固定；

（4） 单击【保存】；

**任务7：使用python工具进行数据关联**

请将销售数据与销售区域进行关联。

**操作步骤：**

1.导入pandas库文件

（1） 使用import函数导入pandas库文件；

（2） 给pandas库文件配置一个缩略名 pd；

（3） 将完整代码编写在下面横线处；

import pandas as pd

2.读取数据

（1） 找到需要处理文件的文件路径；

（2） 将“销售数据清洗结果”文件路径，复制并粘贴到变量df中；

（3） 将“城市表”文件路径，复制并粘贴到变量df1中；

（4） 将“省区表”文件路径，复制并粘贴到变量df2中；

（5） 复制粘贴时，注意要保留横线两侧的单引号；

df = pd.read\_excel('数据集成/数据关联/销售数据清洗结果.xlsx')

df1 = pd.read\_excel('数据集成/数据关联/城市表.xlsx')

df2 = pd.read\_excel('数据集成/数据关联/省区表.xlsx')

3.数据关联

（1） 使用merge函数进行数据关联，左连接用left表示，右连接用right表示，全连接用outer表示，内连接用inner表示；用how将两个数据表连接；on作为关联条件连接的字段；

（2） 将“销售数据清洗结果”与“城市表”进行左连接，关联条件是城市，关联后，新数据表将覆盖原有变量df；

（3） 将关联后df与“省区表”进行左连接，关联条件是省/自治区，关联后，新数据表将覆盖原有变量df；

df = pd.merge(df, df1, how='left', on='城市')

df = pd.merge(df, df2, how='left', on='省/自治区')

4.保存数据

（1） 将关联后的数据保存为excel类型（xlsx）；

（2） 将关联后的数据命名为“销售数据与销售区域关联结果”；

（3） 不保存行索引：False代表不保存，Ture代表保存；

（4） 保证中文无乱码，将数据表用utf-8-sig的方式进行编码；

df = pd.merge(df, df1, how='left', on='城市')

df = pd.merge(df, df2, how='left', on='省/自治区')

五、打印数据

（1） 打印关联后的df；

print(df)

Python代码运行如图3-11所示，运行结果见资料包-销售数据与销售区域关联结果文件：

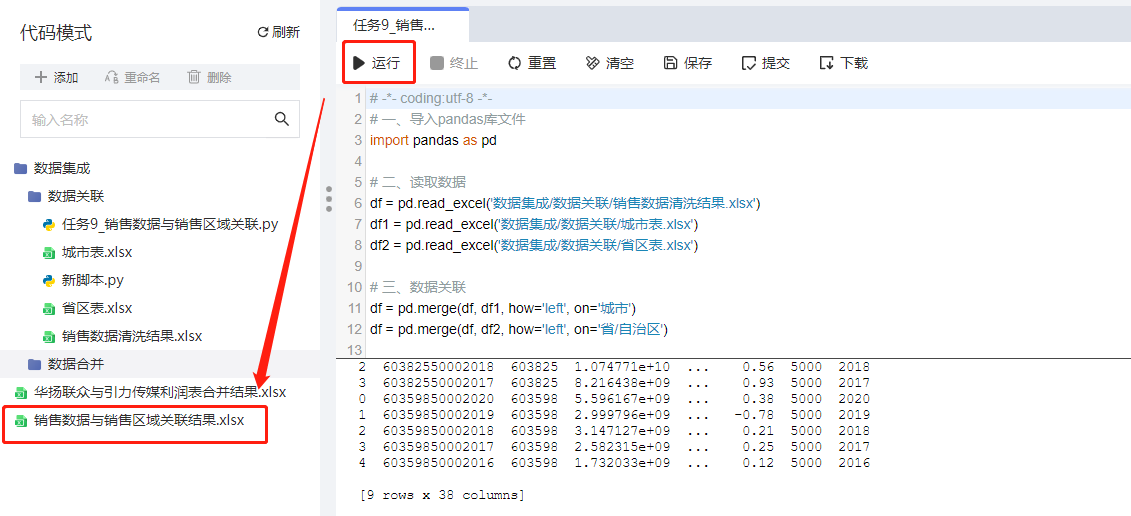


图3-11 python数据关联结果

**任务8：使用python工具进行数据合并**

请将华扬联众与引力传媒利润表表进行合并。

**操作步骤:**

1.导入pandas库文件

（1） 使用import函数导入pandas库文件；

（2） 给pandas库文件配置一个缩略名 pd；

（3） 将完整代码编写在下面横线处；

import pandas as pd

2.读取数据

（1） 找到需要处理文件的文件路径；

（2） 将“华扬联众利润表”文件路径，复制并粘贴到变量df中；

（3） 将“引力传媒利润表”文件路径，复制并粘贴到变量df1中；

df = pd.read\_excel('数据集成/数据合并/华扬联众\_利润表\_清洗后.xlsx')

df1 = pd.read\_excel('数据集成/数据合并/引力传媒\_利润表\_清洗后.xlsx')

3.数据合并

（1） 将“华扬联众利润表”与“引力传媒利润表”进行合并，合并后的新数据表将覆盖原有变量df；

（2） 使用append函数进行数据合并；

df = df.append(df1)

4.保存数据

（1） 将关联后的数据命名为“华扬联众与引力传媒利润表合并结果”；；

（2） 将关联后的数据保存为excel类型（xlsx）；

（3） 不保存行索引：False代表不保存，Ture代表保存；

（4） 保证中文无乱码，将数据表用utf-8-sig的方式进行编码；

df.to\_excel('华扬联众与引力传媒利润表合并结果.xlsx', index=False)

5.打印数据

（1） 打印合并后的df；

print(df)

Python代码运行如图3-12所示，运行结果见资料包-销售数据与销售区域关联结果文件。



图3-12 python数据合并结果

**二、数据可视化**

数据可视化主要是指将数据以图形图像形式展示，借助于图形化手段，清晰有效地传达与沟通信息。通过数据表，我们可以直观看到数据，但数据本身并不会说话，如果我们不知道如何观察数据和分析数据，那么数据就只是一堆冰冷、枯燥且没有意义的数字或符号。数据可视化是帮助我们观察数据的一种有效手段。借助数据可视化的图形化展示，人们可以清晰有效地传达信息和高效沟通。

**任务目标：**根据企业战略阶段化需求，完成可视化图表，设计可视化看板，直观分别反映公司的资产状况、公司营业收入、公司净利润及公司收入结构等情况，并形成财务分析决策报告。

**任务实现方式：**可视化工具（分析云）实现

**任务流程：**1.从资源处下载本项目数据，上传分析云；2.通过不同可视化效果，分别做出财务看板与经营看板可视化图表；3.利用分析云，从财务看板与经营看板中，挑选出六个可视化图形，组成一个可视化看板，用于管理层会议汇报。

**任务指引：**项目任务方案设计>数据源下载>数据集成>建立可视化看板>设计可视化图形>形成项目分析报告

**任务准备：**完成任务所涉及的数据源可在资源下载处进行数据源下载。

**任务1：建立表关联**

将AJ公司的资产负债表、利润表和客户销售情况表上传到分析云；资料见资料包-客户销售情况表、利润表-AJHXJL、资产负债表-AJHXJL文件；并将AJ公司的利润表和资产负债表建立关联。

**操作步骤：**

1.新建数据集

（1）单击【数据准备】；

（2）单击【新建】；

（3）选择数据类型为“关联数据集”；

（4）将数据集名称命名为“AJ利润表与资产表合集”；

2.数据关联

（1）分别拖拽资产负债表和利润表到数据预览区域；

（2）单击两个需要关联的表进行连接，选择“内连接”；

（3）关联条件：报表日期；

（4）单击【确定】

3.数据集保存

（1）单击【执行】；

（2）单击“实时”中的数据物化；

（3）单击【保存】

**任务2：营业入趋势图**

根据任务要求，针对营业收入趋势指标，做出适合的可视化图形。

**操作步骤：**

1.新建可视化

（1） 进入故事板设计页面，单击【可视化】；

（2） 单击【新建】；

（3） 将可视化命名为：营业收入趋势图；

2.选择数据源

（1） 可以选择自己在数据集成中完成的AJ公司利润表与资产表合集；

（2） 或选择系统内置数据（AJ利润表&资产负债表合集）

3.维度与指标的选择

（1） 维度选择：年\_报表日期

（2） 指标选择：营业收入；

4.排序

（1） 将维度“年\_报表日期”升序排列；

（2） 单击维度“年\_报表日期”标签的向下箭头，选择“升序”，选择“年\_报表日期”；

5.选择可视化图形

图形选择：折线图

6.修改数据格式（启用千分位，小数点保留2位）

（1） 单击指标“营业收入”；

（2） 单击“数据格式”；

（3） 千分位：启用；

（4） 小数点：2；

7.显示设置

（1） 维度轴设置：标题为“年份”；

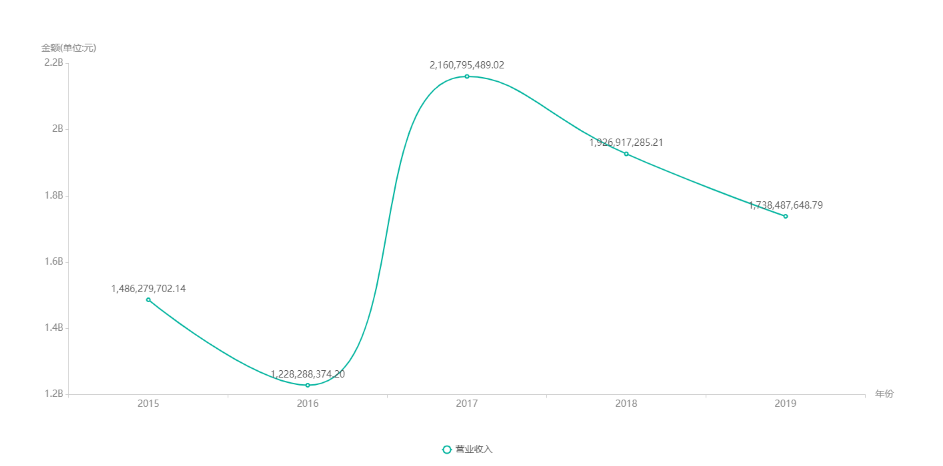
（2） 数值轴设置：标题为“金额（元）”

8.保存可视化

（1） 单击【保存】；

（2） 单击【退出】，回到故事版界面。

可视化设计生成营业收入趋势图如图3-13所示：



**图3-13 营业收入趋势图**

**任务3：营业收入平均值**

根据任务要求，针对营业收入平均值指标，做出适合的可视化图形。可复制上一任务可视化看板，在其基础上进行辅助线的设置，或按照下面步骤操作。

**操作步骤：**

1.新建可视化

（1） 进入故事板设计页面，单击【可视化】；

（2） 单击【新建】；

（3） 将可视化命名为：营业收入平均值；

2.选择数据源

（1） 可以选择自己在数据集成中完成的AJ公司利润表与资产表合集；

（2） 或选择系统内置数据（AJ利润表&资产负债表合集）

3.维度与指标的选择

（1） 维度选择：年\_报表日期

（2） 指标选择：营业收入；

4.排序

（1） 将维度“年\_报表日期”升序排列；

（2） 单击维度“年\_报表日期”标签的向下箭头，选择“升序”，选择“年\_报表日期”；

5.选择可视化图形

图形选择：折线图

6.修改数据格式（启用千分位，小数点保留2位）

（1） 单击指标“营业收入”；

（2） 单击“数据格式”；

（3） 千分位：启用；

（4） 小数点：2；

7.显示设置

（1） 维度轴设置：标题为“年份”；

（2） 数值轴设置：标题为“金额（元）”

8.设置辅助线（将营业收入的平均值作为辅助线显示）

（1） 单击“辅助线”；

（2） 将指标“营业收入”拖拽到辅助线下面，系统弹出“设置辅助线”窗口；

（3） 辅助线计算方式：计算线 选择 “平均值”；

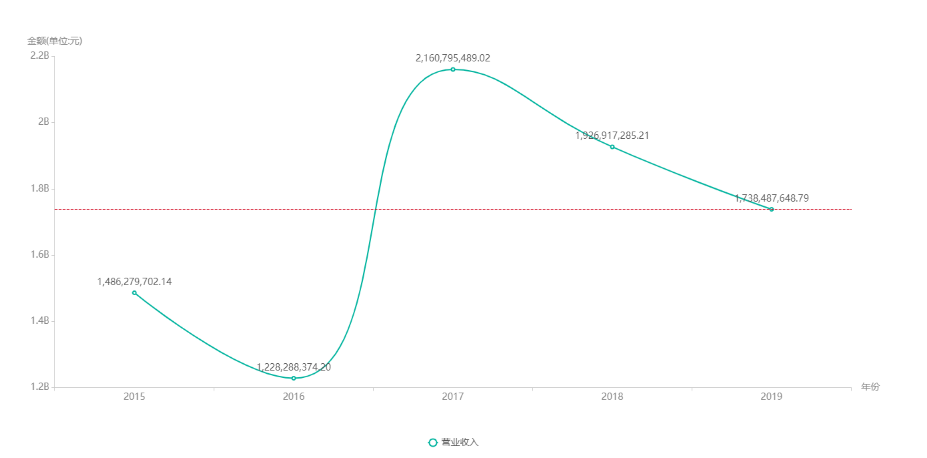
（4） 颜色设置为红色（颜色可任意选择）；

9.保存可视化

（1） 单击【保存】；

（2） 单击【退出】，回到故事版界面。

可视化设计生成的营业收入平均值如图3-14所示：



**图3-14 营业收入平均值**

**任务4：营业收入预警图**

根据任务要求，针对营业收入预警指标，做出适合的可视化图形。

**操作步骤：**

1.新建可视化

（1）进入故事板设计页面，单击【可视化】；

（2）单击【新建】；

（3）将可视化命名为：营业收入预警图；

2.选择数据源

（1）可以选择自己在数据集成中完成的AJ公司利润表与资产表合集；

（2）或选择系统内置数据（AJ利润表&资产负债表合集）

3.维度与指标的选择

（1）维度选择：年\_报表日期

（2）指标选择：营业收入；

4.排序

（1）将维度“年\_报表日期”升序排列；

（2）单击维度“年\_报表日期”标签的向下箭头，选择“升序”，选择“年\_报表日期”；

5.选择可视化图形

（1）图形选择：折线图

6.修改数据格式（启用千分位，小数点保留2位）

（1）单击指标“营业收入”；

（2）单击“数据格式”；

（3）千分位：启用；

（4）小数点：2；

7.显示设置

（1）维度轴设置：标题为“年份”；

（2）数值轴设置：标题为“金额（元）”

8.设置辅助线（将营业收入的平均值作为辅助线显示）

（1）单击“辅助线”；

（2）将指标“营业收入”拖拽到辅助线下面，系统弹出“设置辅助线”窗口；

（3）辅助线计算方式：计算线 选择 “平均值”；

（4）颜色设置为红色（颜色可任意选择）；

9.设置预警线

添加预警规则：

（1）指标聚合方式：求和

（2)预警指标满足：任一条件

（3）营业收入 小于 1800000000（18亿元）；

（4）单击【下一步】；

添加预警人员

（5）可以选择自己的手机号码，或选择BQ管理员；

（6）选中后单击 → 箭头；

（7）单击【下一步】；

添加预警设置

（8）预警级别：可根据需求添加（一般、重要、重大）；

（9）添加预警线颜色：黄色（颜色可任意选择）；

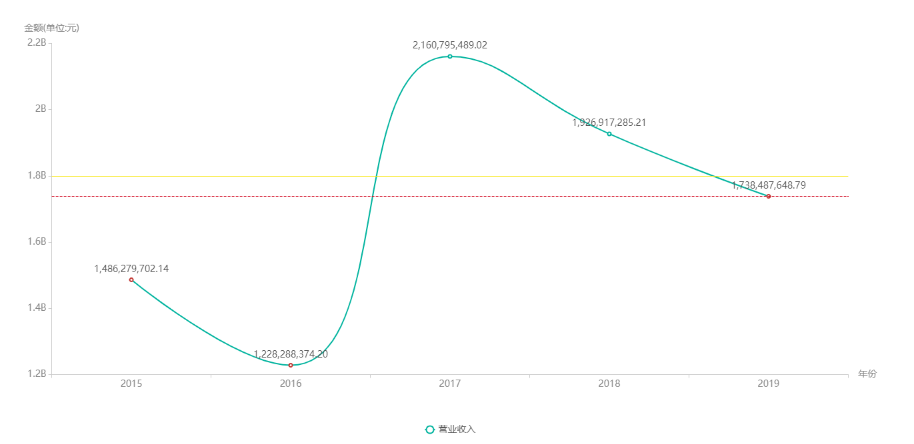
（10）单击【确认】；

10：保存可视化

（1）单击【保存】；

（2）单击【退出】，回到故事版界面。

可视化设计生成的营业收入预警图如图3-15所示：



**图3-15 营业收入预警图**

**任务5：收入结构占比**

根据任务要求，针对收入结构占比指标，做出适合的可视化图形。

**操作步骤：**

1.新建可视化

（1） 进入故事板设计页面，单击【可视化】；

（2） 单击【新建】；

（3） 将可视化命名为：收入结构占比图；

2.选择数据源

（1） 可以选择自己在数据集成中完成的AJ公司利润表与资产表合集，或

（2） 或选择系统内置数据（AJ利润表&资产负债表合集）

3.维度与指标的选择

（1） 维度选择：空（不需要任何维度）

（2） 指标选择：主营业务收入、其他业务收入、营业外收入、投资收益；

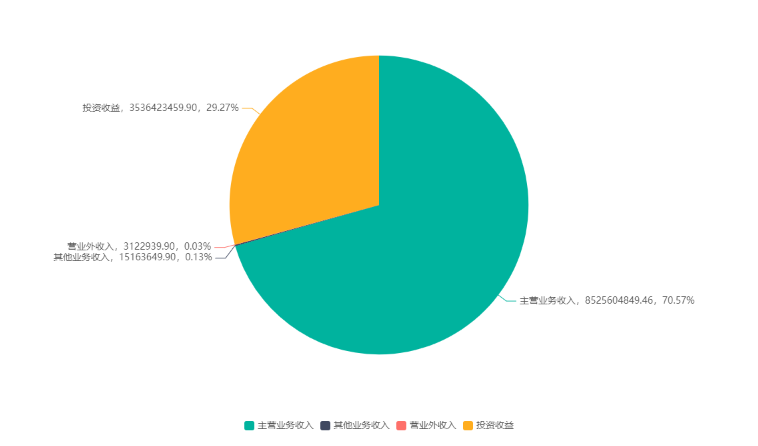
4.选择可视化图形

（1）图形选择：饼图或环形图

5.保存可视化

（1）单击【保存】。

可视化设计生成的收入结构占比如图3-16所示：



**图3-16 收入结构占比**

**任务6：净利润变动趋势图**

根据任务要求，针对净利润变动趋势指标，做出适合的可视化图形。

**操作步骤：**

1.新建可视化

（1） 进入故事板设计页面，单击【可视化】；

（2） 单击【新建】；

（3） 将可视化命名为：净利润变动趋势图；

2.选择数据源

（1） 可以选择自己在数据集成中完成的AJ公司利润表与资产表合集，或

（2） 或选择系统内置数据（AJ利润表&资产负债表合集）

3.维度与指标的选择

（1） 维度选择：年\_报表日期

（2） 指标选择：净利润；

4.排序

（1） 将维度“年\_报表日期”升序排列；

（2） 单击维度“年\_报表日期”标签的向下箭头，选择“升序”，选择“年\_报表日期”；

5.选择可视化图形

（1）图形选择：堆叠区域图

6.修改数据格式（启用千分位，小数点保留2位）

（1） 单击指标“净利润”；

（2） 单击“数据格式”；

（3） 千分位：启用；

（4） 小数点：2；

7.显示设置

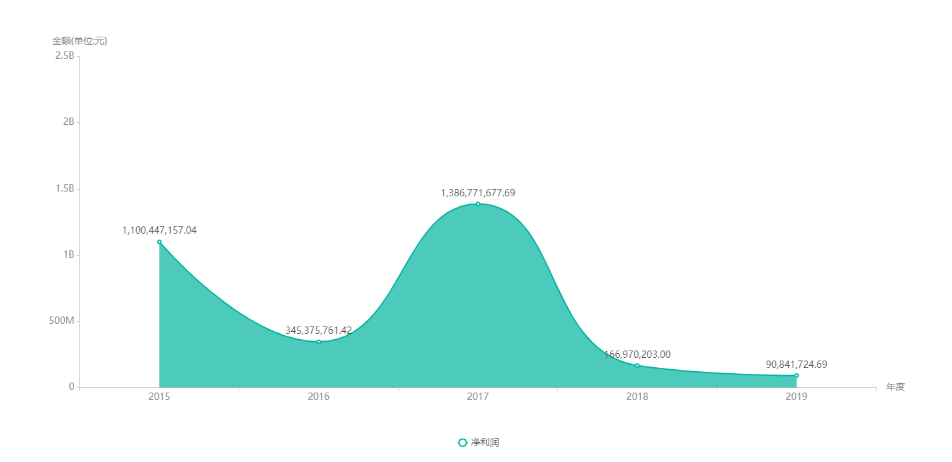
（1） 维度轴设置：标题为“年份”；

（2） 数值轴设置：标题为“金额（元）”

8.保存可视化

（1）单击【保存】。

可视化设计生成的净利润变动趋势图如图3-17所示：



**图3-17 净利润变动趋势图**

**任务7：资产总计**

根据任务要求，针对资产总计指标，做出适合的可视化图形。

**操作步骤：**

1.新建可视化

（1） 进入故事板设计页面，单击【可视化】；

（2） 单击【新建】；

（3） 将可视化命名为：资产总计；

2.选择数据源

（1） 可以选择自己在数据集成中完成的AJ公司利润表与资产表合集，或

（2） 或选择系统内置数据（AJ利润表&资产负债表合集）

3.维度与指标的选择

（1） 维度选择：年\_报表日期

（2） 指标选择：资产总计；

4.排序

（1） 将维度“年\_报表日期”升序排列；

（2） 单击维度“年\_报表日期”标签的向下箭头，选择“升序”，选择“年\_报表日期”；

5.选择可视化图形

图形选择：柱状图

6.修改数据格式（启用千分位，小数点保留2位）

（1） 单击指标“营业收入”；

（2） 单击“数据格式”；

（3） 千分位：启用；

（4） 小数点：2；

7.显示设置

（1） 维度轴设置：标题为“年份”；

（2） 数值轴设置：标题为“金额（元）”

8.保存可视化

（1）单击【保存】。

可视化设计生成的资产总计柱状图如图3-18所示：



**图3-18 资产总计柱状图**

**任务8：资产负债率**

根据任务要求，针对资产负债率指标，做出适合的可视化图形。

**操作步骤：**

1.新建可视化

（1） 进入故事板设计页面，单击【可视化】；

（2） 单击【新建】；

（3） 将可视化命名为：资产负债率；

2.选择数据源

（1） 可以选择自己在数据集成中完成的AJ公司利润表与资产表合集，或

（2） 或选择系统内置数据（AJ利润表&资产负债表合集）

3.维度与指标的选择

（1） 维度选择：年\_报表日期；

（2） 指标选择：资产总计；

4.新建指标

（1） 单击指标旁边【+】按钮；

（2） 单击“计算字段”；

（3） 添加字段，名称为“资产负债率”；

（4） 字段类型“数字”；

（5） 表达式：avg(负债合计x)/avg(资产总计x)，此处需要从“函数”中选择数学函数中的avg，并从“可选择字段”中选择负债合计与资产总计；

（6） 单击【确定】；

（7） 将资产负债率拖拽到指标字段；

5.排序

（1） 将维度“年\_报表日期”升序排列；

（2） 单击维度“年\_报表日期”标签的向下箭头，选择“升序”，选择“年\_报表日期”；

6.选择可视化图形

图形选择：双轴图

7.添加过滤条件

（1） 年\_报表日期 包含 2017、2018、2019

（2） 单击【确定】；

8.设置数据格式

（1） 单击指标“资产负债率”标签向下箭头；

（2） 单击“数据格式”；

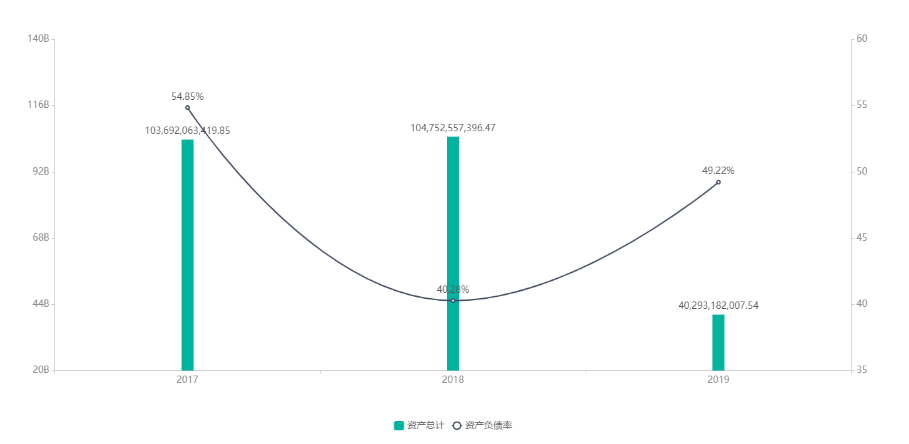
（3） 设置缩放率：0.01；

（4） 后导符：%；

9.保存可视化

单击【保存】。

可视化设计生成的资产负债率双轴图如图3-19所示：



**图3-19 资产负债率双轴图**