

# 超市销售数据的可视化分析

10215501412 彭一坤

## 一、实验介绍

项目基于DataEase可视化分析工具进行，以简单的超市销售明细数据做为数据源。

**DataEase** 是人人可用的开源数据可视化分析工具，帮助用户快速分析数据并洞察业务趋势，从而实现业务的改进与优化。DataEase 支持丰富的数据源连接，能够通过拖拉拽方式快速制作图表，并可以方便地与他人分享。

**DataEase**的主要功能包括：

- 图表展示: 支持 PC 端仪表盘及数据大屏；
- 图表制作: 支持丰富的图表类型、支持拖拉拽方式快速制作仪表盘；
- 数据引擎: 基于 Apache Calcite，实现统一的 SQL 解析、验证、优化和执行；
- 数据连接: 支持关系型数据库、数据文件、数据仓库等各种数据源。

**DataEase**具有以下优势：

- 开源开放：零门槛，线上快速获取和安装；快速获取用户反馈、按月发布新版本；
- 简单易用：极易上手，通过鼠标点击和拖拽即可完成分析；
- 全场景支持：多平台安装和多样化嵌入支持；
- 安全分享：支持多种数据分享方式，确保数据安全。

## 二、实验步骤

### 1.数据选择与介绍

数据大屏上所显示的数据，一般是某个特定领域中，逻辑关系比较密切的一系列指标和明细数据的组合，以不同的图表的方式进行可视化的目的是让核心数字一目了然、重要趋势印象深刻、明细数据体现活力、汇总层级转变灵活。

数据源应该具备以下特点：

- 与某个特定领域密切相关
- 能体现日常的例行化操作过程
- 基本连续的一段时间
- 维度信息比较丰富，比如时间、地理、组织、物品种类，等等
- 具备有价值的可衡量的数值，比如金额、物品数量、频率、重量、大小，等等。

经过对比筛选，最终选定了一份简单的零售超市的交易流水文件做为数据集（类似是收银台的明细）。

此数据集具有时间、商品类型、商品类别、销售方式、顾客等维度信息；也有销售数量、销售金额等可度量的数值信息，时间是基本连续的4个月。

此数据集的前20行示例数据如下表所示：

顾客编号	大类编码	大类名称	中类编码	中类名称	小类编码	小类名称	销售日期	销售月份	商品编码	规格型号	商品类型	单位	销售数量	销售金额	商品单价	是否促销
7	20	粮油	2001	袋装速食面	200102	猪肉口味	20150101	201501	DW-200102	82.6g	一般商品	袋	1	3	3	否
35	20	粮油	2013	调味酱	201301	番茄酱	20150101	201501	DW-201301	340g	一般商品	瓶	1	6.7	6.7	否
10	15	日配	1505	冷藏乳品	150503	冷藏果粒酸乳	20150101	201501	DW-150503	260g	一般商品	盒	1	5.9	8.9	是
26	21	冲调	2105	冲调食品	210501	麦片/粉	20150101	201501	DW-210501	700g	一般商品	袋	1	27.7	27.7	否
31	30	洗化	3006	洗衣用品	300604	洗衣皂	20150101	201501	DW-300604	226g	一般商品	块	1	4.9	4.9	否
31	30	洗化	3016	纸制品	301608	软抽纸巾	20150101	201501	DW-301608	200抽*3	一般商品	提	1	6.9	9.7	是
49	12	蔬果	1201	蔬菜	120104	花果	20150101	201501	DW-120104	散称	生鲜	千克	0.902	8.98	9.96	否
8	15	日配	1518	常温乳品	151801	利乐砖纯奶	20150101	201501	DW-151801	250ml*12	一般商品	提	1	49	65	是
15	31	家居	3118	打扫用品	311809	胶棉拖把	20150101	201501	DW-311809	MOJ-A	一般商品	把	1	70	70	否
50	22	休闲	2207	糖果	220703	软糖	20150101	201501	DW-220703	105g	一般商品	袋	1	10.9	10.9	否
1	12	蔬果	1201	蔬菜	120103	芽菜	20150101	201501	DW-120103	散称	生鲜	千克	0.456	1.27	2.78	否
51	30	洗化	3018	卫生巾	301802	夜用卫生巾	20150101	201501	DW-301802	8片	一般商品	包	1	8.9	8.9	否
51	30	洗化	3018	卫生巾	301804	卫生护垫	20150101	201501	DW-301804	12+2	一般商品	包	1	11.9	11.9	否
45	12	蔬果	1201	蔬菜	120105	鲜调味	20150101	201501	DW-120105	散称	生鲜	千克	0.24	2.3	9.6	否
7	20	粮油	2002	袋装速食面饼	200201	牛肉口味	20150101	201501	DW-200201	125g*5	一般商品	组	1	1	1	否
22	30	洗化	3016	纸制品	301603	无芯纸	20150101	201501	DW-301603	200抽	一般商品	提	1	18.9	24.9	是
9	22	休闲	2201	饼干	220111	趣味/休闲	20150101	201501	DW-220111	45g	一般商品	盒	1	6.9	6.9	否
46	12	蔬果	1203	水果	120302	苹果类	20150101	201501	DW-120302	散称	生鲜	KG	1.776	8.84	4.98	否
10	22	休闲	2202	糕点	220201	派	20150101	201501	DW-220201	198g	一般商品	盒	1	9.7	9.7	否

此数据集共有字段17个，记录42794条。最终设置维度列14个，包括顾客编号、大类编码、大类名称、中类编码、中类名称、小类编码、小类名称、销售日期、销售月份、商品编码、规格型号、商品类型、单位、是否促销；指标列3个，包括销售数量、销售金额、商品单价。维度与指标列的设置经过如下两图所示：

▼ 维度 (14)					
<input checked="" type="checkbox"/>	字段名称	原始名称	字段类型	字段原始类型	操作
<input checked="" type="checkbox"/>	顾客编号	顾客编号	T 文本	# 数值	转换为指标 复制
<input checked="" type="checkbox"/>	大类编码	大类编码	T 文本	# 数值	转换为指标 复制
<input checked="" type="checkbox"/>	大类名称	大类名称	T 文本	T 文本	转换为指标 复制
<input checked="" type="checkbox"/>	中类编码	中类编码	T 文本	# 数值	转换为指标 复制
<input checked="" type="checkbox"/>	中类名称	中类名称	T 文本	T 文本	转换为指标 复制
<input checked="" type="checkbox"/>	小类编码	小类编码	T 文本	# 数值	转换为指标 复制
<input checked="" type="checkbox"/>	小类名称	小类名称	T 文本	T 文本	转换为指标 复制
<input checked="" type="checkbox"/>	销售日期	销售日期	T 文本	# 数值	转换为指标 复制
<input checked="" type="checkbox"/>	销售月份	销售月份	T 文本	# 数值	转换为指标 复制
<input checked="" type="checkbox"/>	商品编码	商品编码	T 文本	T 文本	转换为指标 复制
<input checked="" type="checkbox"/>	规格型号	规格型号	T 文本	T 文本	转换为指标 复制
<input checked="" type="checkbox"/>	商品类型	商品类型	T 文本	T 文本	转换为指标 复制
<input checked="" type="checkbox"/>	单位	单位	T 文本	T 文本	转换为指标 复制
<input checked="" type="checkbox"/>	是否促销	是否促销	T 文本	T 文本	转换为指标 复制
▼ 指标 (3)					
<input checked="" type="checkbox"/>	字段名称	原始名称	字段类型	字段原始类型	操作
<input checked="" type="checkbox"/>	销售数量	销售数量	# 数值(小数)	# 数值(小数)	转换为维度 复制
<input checked="" type="checkbox"/>	销售金额	销售金额	# 数值(小数)	# 数值(小数)	转换为维度 复制
<input checked="" type="checkbox"/>	商品单价	商品单价	# 数值(小数)	# 数值(小数)	转换为维度 复制

## 2.模板选择与布局设计

可视化大屏的目标是信息醒目、印象深刻，所以选用的模板布局逻辑清晰、重要数字醒目、有趋势有排名。

其布局分为筛选条件（可交互）、概要统计区、文字提示区、趋势提示区和结构及排名区五部分，示例如下：



### 3.图表介绍

#### 销售月份（筛选条件）的设计和制作

- 1) 确定数据集，系统提示该数据集下可供选择的文本字段；
- 2) 选择“销售年月”做为文本筛选字段；
- 3) 修改此组件的“控制范围”，把需要受该组件影响和控制的视图名称选定好

关键的设计和制作界面如下图所示：



#### 概要统计区的设计和制作（以“月度总销售额”为例）

概要统计区都是按月度汇总的数值型指标，都是对一部分数据进行聚集的操作，包括求和、平均和计数。

关键的设计和制作界面如下图所示：



### 商品统计区的设计与制作（以“单月最高销量单品”为例）

商品统计区是对单月销量最多的商品或日期进行的统计，都是对“销售金额”进行汇总，但最终输出的都是商品类型、种类、日期等维度信息，所以此功能中用到了“视图字段”，以使用维度列的名字。

关键的设计和制作界面如下两图所示（第二图是对“视图字段”的使用）：





趋势统计区的设计与制作（以“销售额走势”为例）

趋势统计区是对整个时间段的关键指标的变化情况的统计，包括销量和销售额。展现方式包括趋势图、柱状图和趋势面积图。

此部分不与“筛选条件”相关，不参与交互，不与筛选条件联动。

关键的设计和制作界面如下图所示：



结构与排名区的设计与制作（以“品类占比”为例）

结构与排名区主要是对销售商品的品类进行统计，涉及分类汇总和对类别的排序和筛选。展现方式包括列表、饼图和横柱状图以及标签组件。

品类占比是以商品大类为分组条件，对销售数量进行汇总，并且需要对销售数量的汇总结果进行排序。

关键的设计和制作界面若下两图所示：

数据

样式

高级

Q 搜索

⚙️ ⇌

维度 ⓘ

T 顾客编号

T 大类编码

T 大类名称

T 中类编码

T 中类名称

T 小类编码

T 小类名称

指标

# 销售数量

# 销售金额

# 商品单价

# 记录数\*

图表类型

更改类型 ▾

结果展示

☒ 全部 ☐ 1000

刷新频率 ⓘ ☐ 开启刷新

5 分钟 ▾

扇区标签 / 维度

T 大类名称 ▾

扇区角度 / 指标

# 销售数量 求和 ▾

过滤器

过滤

钻取 / 维度 ⓘ

拖动字段至此处

数据

样式

高级

Q 搜索

⚙️ ⇌

维度 ⓘ

T 顾客编号

T 大类编码

T 大类名称

T 中类编码

T 中类名称

T 小类编码

T 小类名称

指标

# 销售数量

# 销售金额

# 商品单价

# 记录数\*

图表类型

更改类型 ▾

结果

■ 汇总方式 (求和) >

☒ 快速计算 (无) >

↕ 排序 (升序) >

🗑 过滤...

■ 数值格式...

☒ 显示名设置

🗑 删除

扇区

# 销售数量 求和 ▾

过滤器

过滤

钻取 / 维度 ⓘ

拖动字段至此处

## 4.跳转功能演示

跳转功能是提供给用户的交互体验。用户选择不同的“销售年月”，“概要统计区”、“商品统计区”和“结构与排名区”的数据（都是按月查询或汇总的统计结果）会随着发生变化。

分别选择不同的“销售年月”，比如“201502”、“201504”，各统计区的显示内容是联动的，示意图如下：



## 三、实验总结

### 1. 维度列与指标列的转换功能比较常用

如果数据文件中的列值为数字，DataEase会自动将其识别为指标列。比如本实验的数据文件中的“销售日期”、“销售年月”列，存储的都是8位或者4位的文本格式的数字（如20010101、200101），所以DataEase将其识别为指标列。后手动将其转换为维度列。说明DataEase的“维度转换为指标”、“指标转换为维度”功能是比较常用的。

### 2. 日期型字段的识别

“销售日期”、“销售年月”两列的内容均为文本格式的数字，DataEase并没有将其识别为日期型字段，在使用“时间过滤组件”时，系统无法提示出该数据集中与时间有关的字段，所以后来转用“文本过滤组件”。日期型数据比较常用，应该丰富DataEase对实质性的日期型字段的识别能力，应该支持定义“文本格式的日期”这种字段。

### 3. 排序功能的使用

必须注意对统计结果的排序，以保证显示效果的准确。包括直接对维度列的排序和基于汇总值而最终印象维度列的显示顺序的间接排序。

### 4. 钻取

钻取功能适用于“有层次”的维度列，比如时间维度的月季年、商品类别的小类中类大类。

如果数据集中，在每一条记录上都已经具备各层次的维度信息（各层次都表示为不同的列），实现起来相对比较容易。

附：作品访问方式是登录dataease平台，创建仪表板-导入 连锁饮品店销售看板-TEMPLATE.DET 文件，或可查看dataease平台上的分享消息。