购物篮分析

- 背景描述
- 需求分析
- 现状分析
- 实施过程
    - 数据接入
    - 数据处理
    - 构建模型
        - 最小支持度
        - 最小置信度
    - 结果分析
- 总结

背景描述

购物篮分析是商业领域最前沿、最具挑战性的问题之一，也是许多企业研究重点问题。购物篮分析是通过发现顾客在一次购买行为中放入购物篮中不同商品之间的关联，研究客户的购买行为，从而达到辅助零售企业制定营销策略制定的一种数据分析方法。

需求分析

商场以获得最大的销售利润为目的。零售商都在考虑销售什么样的商品，采用什么样的促销策略，商品在货架上如何摆放，了解顾客的购买习惯和偏好。本案例对商场销售数据进行分析，从而得到顾客的购买行为特征，并根据发现的规律而采取有效的行动，制定商品摆放、商品定价、新商品采购计划，对商场增加销量有着重要意义。

本案例使用FP-Growth关联规则算法实现购物篮分析，发现超市不同商品之间的关联关系，并根据商品之间的关联规则制定销售策略。

根据购物篮案例数据实现以下目标：

1.分析零售企业商品销售现状。

2.了解某商品零售企业的基本数据情况。

3.熟悉购物篮分析的基本流程与步骤。

现状分析

现代商品种类繁多，顾客往往会由于需要购买的商品众多而变得疲于选择，且顾客并不会因为商品选择丰富而选择购买更多的商品。例如，货架上有可口可乐和百事可乐，若顾客需要选购可乐若干，或许会同时购买两种可乐，但是购买可乐的数量大多数情况下不会因为品牌数量增加而增加。

繁杂的选购过程往往会给顾客疲惫的购物体验。对于某些商品，顾客会选择同时购买，如面包与牛奶、薯片与可乐等，当面包与牛奶或者薯片与可乐分布在商场的两侧，且距离十分遥远时，顾客购买的欲望就会减少，在时间紧迫的情况下顾客甚至会放弃购买某些计划购买的商品。相反，把牛奶与面包摆放在相邻的位置，既给顾客提供便利，提升购物体验，又提高顾客购买的概率，达到了促销的目的。

其次，许多商场以打折方式作为主要促销手段，以更少的利润为代价获得更高的销量。打折往往会使顾客增加原计划购买商品的数量，对于原计划不打算购买且不必要的商品，打折的吸引力远远不足。而正确的商品摆放却能提醒顾客购买某些必需品，甚至吸引他们购买感兴趣的商品。

综合商品零售行业现状，本案例需要完成以下分析目标：

1.分析商品热销情况和商品结构。

2.分析商品之间的关联性。

3.根据分析结果给出销售建议。

实施过程

某商品零售企业共收集了300030条购物篮数据，以下是字段说明：

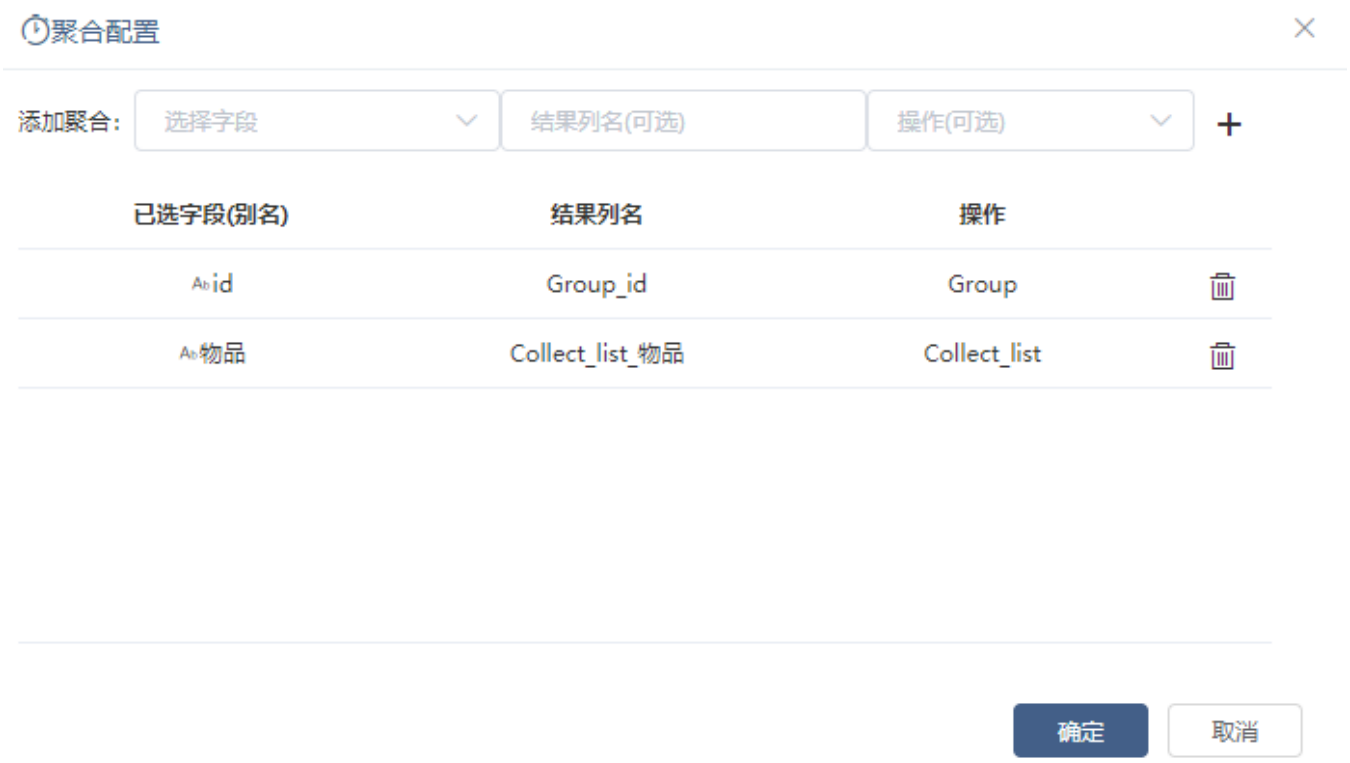
字段名称	类型	字段说明
id	浮点型	用户ID
卡号	浮点型	
消费额	浮点型	
付款方式	字符串	取值为{CHEQUE,CARD,CASH}
性别	字符串	取值为{F,M}
年龄	浮点型	
物品	字符串	

数据接入

在实验中添加 数据源 节点，将购物篮数据读取进来，部分数据如图所示：

数据处理

根据本案例分析的是商品之间的关联性，需要收集到用户购买的物品列表，通过 聚合 节点根据用户id分组收集物品列表，如图所示：



输出结果如图：

当前显示 100 条 / 总共有 940 条数据　提示:点击单元格可查看超出的内容

# Group_id	# Collect_list_物品
558.0	WrappedArray(蔬菜罐头, 肉罐头, 冻肉, 啤酒)
596.0	WrappedArray(蔬菜罐头, 白酒)
692.0	WrappedArray(冻肉, 啤酒, 鱼类, 糖果)
299.0	WrappedArray(水果)
305.0	WrappedArray(白酒, 糖果)
496.0	WrappedArray(肉罐头, 冻肉, 鱼类)
769.0	WrappedArray(鲜肉, 乳制品, 白酒)
934.0	WrappedArray(水果, 肉罐头, 冻肉)
576.0	WrappedArray(水果, 蔬菜罐头, 肉罐头, 糖果)
720.0	WrappedArray(蔬菜罐头, 冻肉, 啤酒, 糖果)
147.0	WrappedArray(鲜肉, 蔬菜罐头, 肉罐头, 冻肉, 啤酒, 白酒, 糖果)

注意：表头中◈表示特征列，*表示标签列　表头真名 表头别名

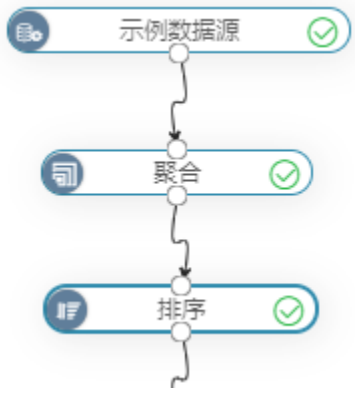
通过 排序 节点，根据用户id进行排序，输出结果如图：

当前显示 100 条 / 总共有 940 条数据　提示:点击单元格可查看超出的内容

# Group_id	# Collect_list_物品
1.0	WrappedArray(鲜肉, 乳制品, 糖果)
2.0	WrappedArray(鲜肉, 糖果)
3.0	WrappedArray(蔬菜罐头, 冻肉, 啤酒, 鱼类)
4.0	WrappedArray(乳制品, 白酒)
6.0	WrappedArray(鲜肉, 白酒, 鱼类)
7.0	WrappedArray(水果, 软饮料)
8.0	WrappedArray(啤酒)
9.0	WrappedArray(水果, 冻肉)
10.0	WrappedArray(水果, 鱼类)
11.0	WrappedArray(水果, 鲜肉, 乳制品, 蔬菜罐头, 白酒, 鱼类)
12.0	WrappedArray(水果, 鱼类)

注意：表头中◈表示特征列，*表示标签列　表头真名 表头别名

数据预处理流程图如图：



构建模型

经过特征选择物品列表一列，通过 拆分 节点按照7：3比例将数据集拆分为训练集和测试集。

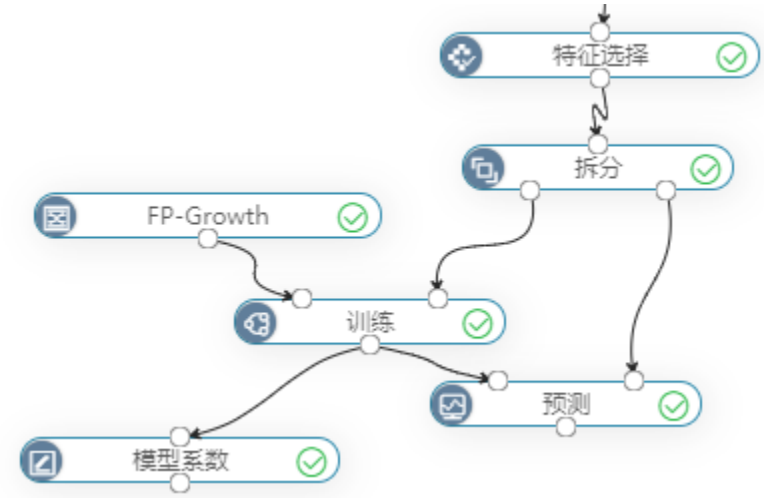
最小支持度

支持度，代表项集的频繁程度；最小支持度作为支持度的阈值，满足最小支持度的项集才会输出。

最小置信度

置信度，代表包含A事务中同时包含B事务的频繁程度；最小置信度作为置信度的阈值，满足最小置信度的项集才会输出。

我们使用一个关联规则算法，这里选用 FP-Growth 算法 。整体的实验流程如图：



算法参数配置如图：

结果分析

如实验流程图所示，通过 模型系数 节点输出训练模型，输出结果如下：

当前显示 18 条 / 总共有 18 条数据　提示:点击单元格可查看超出的内容

频繁项集	# 频率	# 支持度
[蔬菜罐头]	227	0.33880597014925373
[啤酒]	216	0.32238805970149254
[冻肉]	215	0.3208955223880597
[水果]	212	0.3164179104477612
[鱼类]	211	0.31492537313432833
[白酒]	203	0.3029850746268657
[糖果]	200	0.29850746268656714
[肉罐头]	146	0.21791044776119403
[软饮料]	134	0.2
[冻肉, 蔬菜罐头]	130	0.19402985074626866
[鲜肉]	126	0.1880597014925373

注意：表头中◇表示特征列，*表示标签列

表头真名 表头别名

预测结果如图：

当前显示 100 条 / 总共有 270 条数据　提示:点击单元格可查看超出的内容

# Group_id	# Collect_list_物品	# prediction
8.0	WrappedArray(啤酒)	WrappedArray(冻肉, 蔬菜罐头)
11.0	WrappedArray(水果, 鲜肉, 乳制品, 蔬菜罐头, 白酒, 鱼类)	WrappedArray(糖果, 啤酒, 冻肉)
13.0	WrappedArray(乳制品, 肉罐头, 冻肉, 鱼类)	WrappedArray(水果, 啤酒, 蔬菜罐头)
17.0	WrappedArray(水果, 蔬菜罐头, 鱼类)	WrappedArray(白酒, 啤酒, 冻肉)
19.0	WrappedArray(啤酒, 鱼类)	WrappedArray(冻肉, 蔬菜罐头, 水果)
22.0	WrappedArray(白酒, 糖果)	WrappedArray(蔬菜罐头)
24.0	WrappedArray(水果, 鲜肉, 乳制品, 蔬菜罐头, 肉罐头, 冻肉, 啤酒, 白酒)	WrappedArray(鱼类, 糖果)
30.0	WrappedArray(肉罐头)	WrappedArray()
31.0	WrappedArray(冻肉, 啤酒, 白酒, 鱼类)	WrappedArray(蔬菜罐头, 水果, 糖果)
33.0	WrappedArray(蔬菜罐头, 肉罐头, 啤酒, 白酒)	WrappedArray(冻肉, 糖果)
39.0	WrappedArray(鱼类, 糖果)	WrappedArray(水果, 白酒)

注意：表头中◇表示特征列，*表示标签列

表头真名 表头别名

通过模型的规则得出：顾客大多数比较喜欢购买蔬菜罐头、啤酒、冻肉等商品，并且在顾客购买啤酒的时候会同时购买冻肉、蔬菜罐头。

因此，商场应该根据实际情况将蔬菜罐头、啤酒等商品在顾客购买冻肉的必经之路，或者的显眼位置，方便顾客拿取。其他水果、鱼类、糖果等商品同时购买的概率较高，可以考虑捆绑销售，或者适当调整商场布置，将这些商品的距离尽量拉近，提升购物体验。

总结

本案例主要结合商品零售购物篮的案例，重点介绍了关联规则算法中的FP-Growth算法在商品零售购物篮分析案例中的应用。过程中详细的分析了商品零售的现状与问题，同时给出某商场的商品零售数据，分析了商品的热销程度，最后通过FP-Growth算法构建相应模型，并根据模型结果制定销售策略。