# 实验报告

姓名：张嘉铭

学号：1829210092

## 实验名称：Colab\_2: 频繁项挖掘

### 任务概述：

使用spark的频繁项挖掘算法，分析3 Million Instacart Orders数据集中顾客的购物行为，找出顾客购物行为中的频繁项集和关联规则。

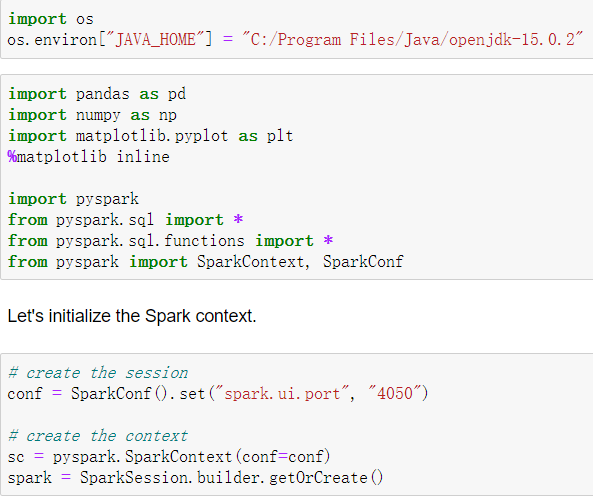
### 使用数据

products.csv: 产品数据，包含产品ID、产品名、通道ID、部门ID。

order\_products\_\_train.csv: 订单数据，包含订单ID、产品ID、购买顺序、reordered。

### **实**验过程及结果

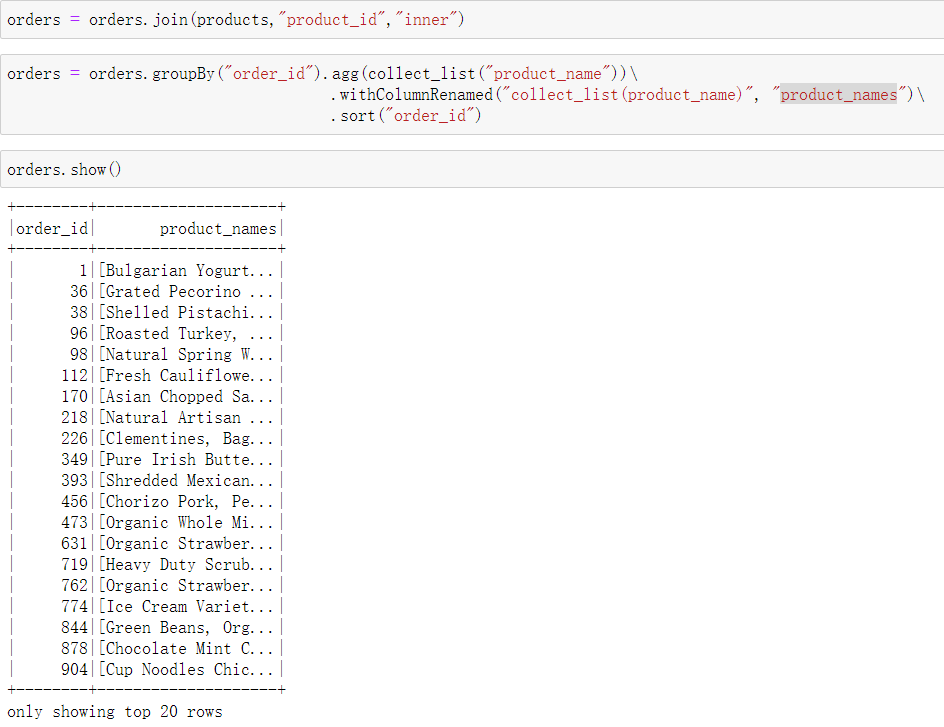
1. 环境配置



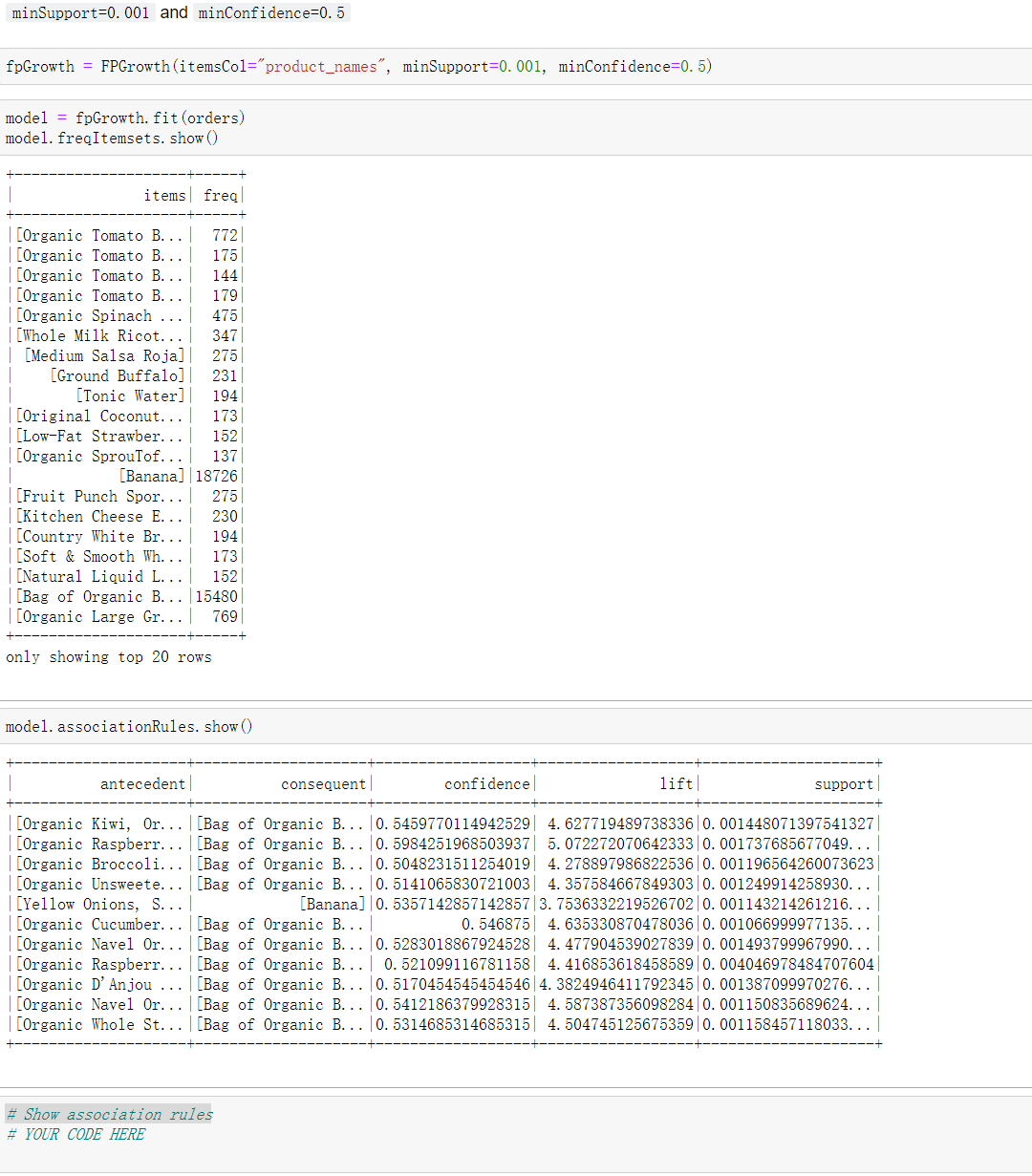
1. 读入数据



1. 连接两张表并通过order\_id分组来得到每个订单所购买的物品信息。



1. 用[FP-Growth](https://spark.apache.org/docs/latest/ml-frequent-pattern-mining.html" \l "fp-growth" \t "http://localhost:8888/notebooks/Downloads/_blank)算法来频繁项挖掘



### 实验中的问题及解决方法

无。