**康雨馨—百货商场用户画像描绘与价值分析**

本项目通过大数据分析技术，对百货商场会员合理分类、挖掘客户价值以实现精准营销，能有效增加公司的盈利能力和成长水平。以下为整个项目的分析过程和研究报告。

项目的整体思路分为以下五个部分。

第一，读取数据并查看

第二，数据清洗

第三，数据合并

第四，数据可视化初步分析

第五，对用户进行用户画像分析

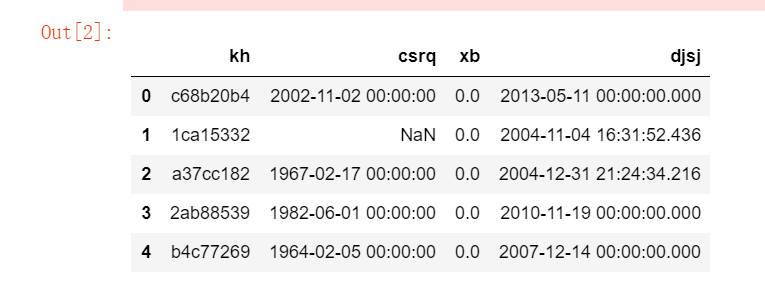
第六，客户价值K-means聚类分析

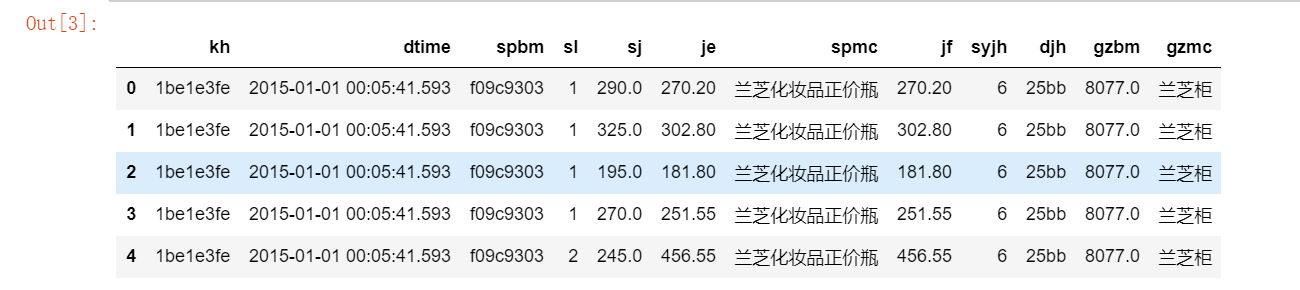
1. **数据读取**

首先导入两张表，分别查看两张表属性、大小、缺失值等信息

先用pd.head来查看表的基本结构

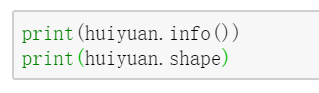


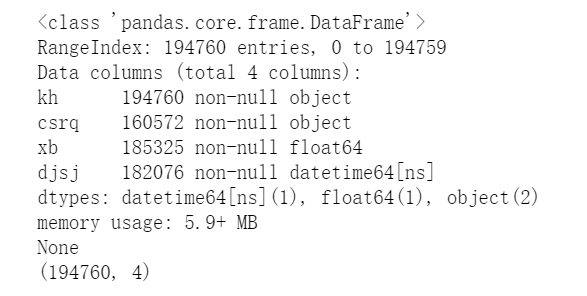




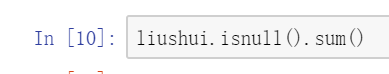
然后通过查重和缺失值查询，发现两张表中都有很多缺失值。

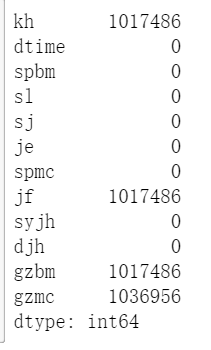
1.1查看会员表的空值：发现除卡号外的所有字段都有空值





1.2查看流水表的空值：发现卡号空值不是本商场的会员，积分空值很多。



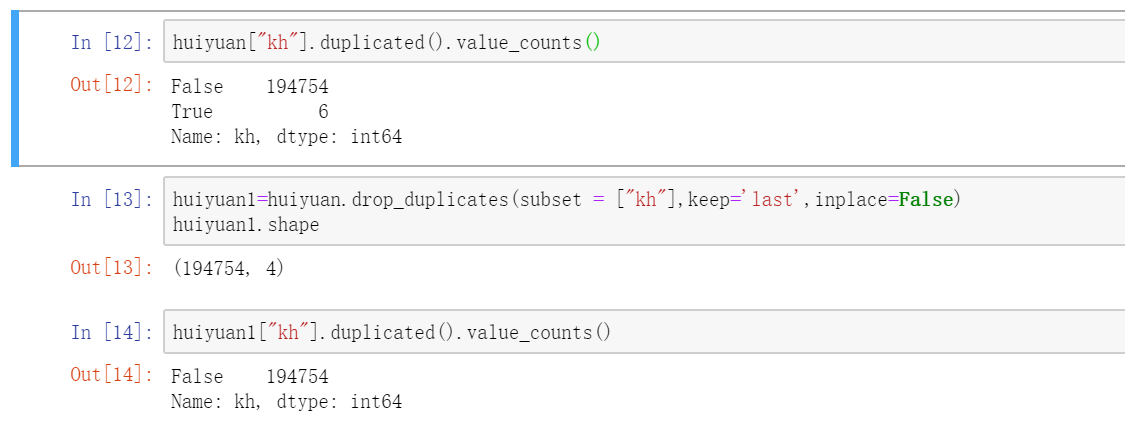


1. **数据清洗**

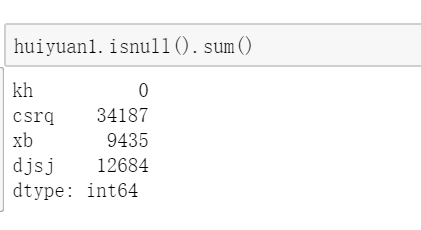
数据清洗分为三个部分：去重、缺失值处理、异常值处理

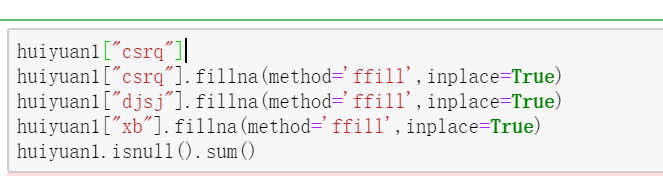
2.1去重

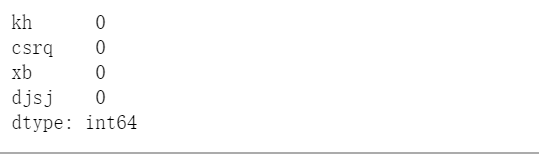
首先进行去重工作。会员表去掉卡号一样的行。



2.2 缺失值填充：由于上缺失数据较多，没有采用删除的方法，转而采用填充的办法。对会员表进行填充，方法是用前面填充后面。

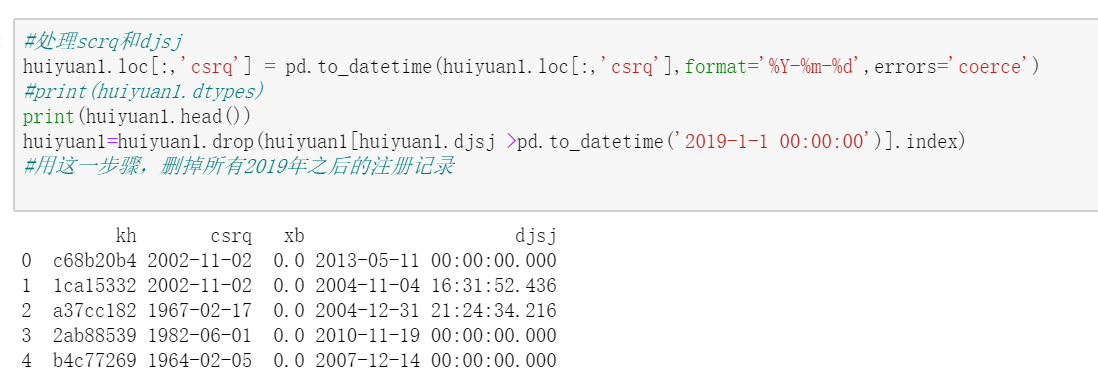






2.3 异常值处理

把会员表里面注册日期改为时间格式，去掉所有注册信息大于2019年的行。

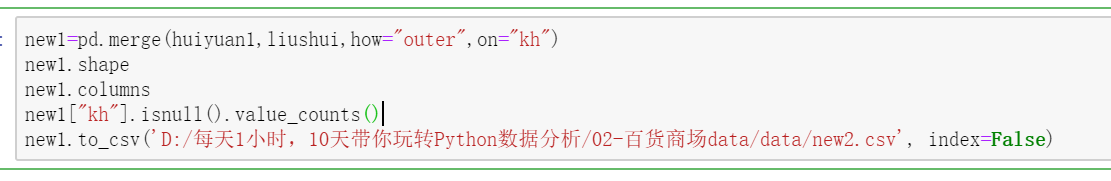


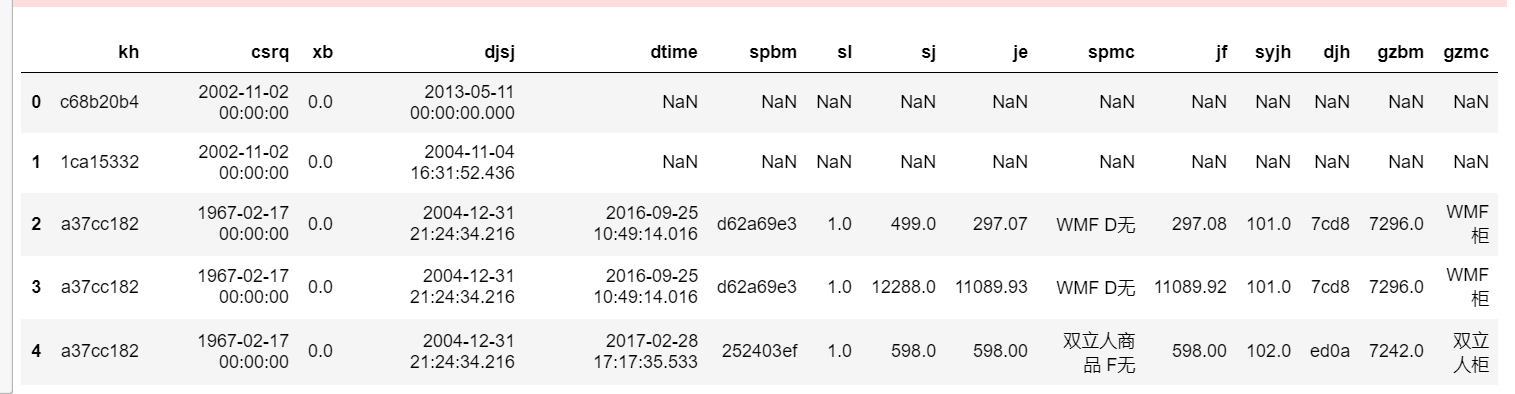
同时，流水表dtime也改为日期格式，对流水表去掉所有消费时间异常的行。



1. **数据合并**

合并两表，采用主键合并的方式，以“kh”为标准，实行外连接

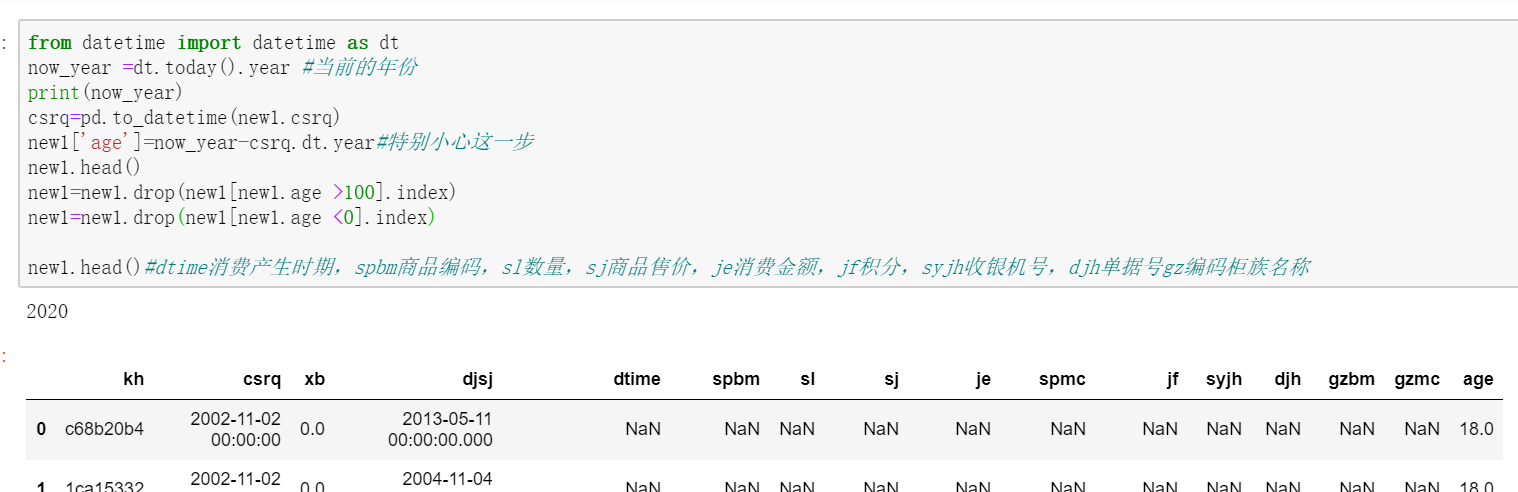




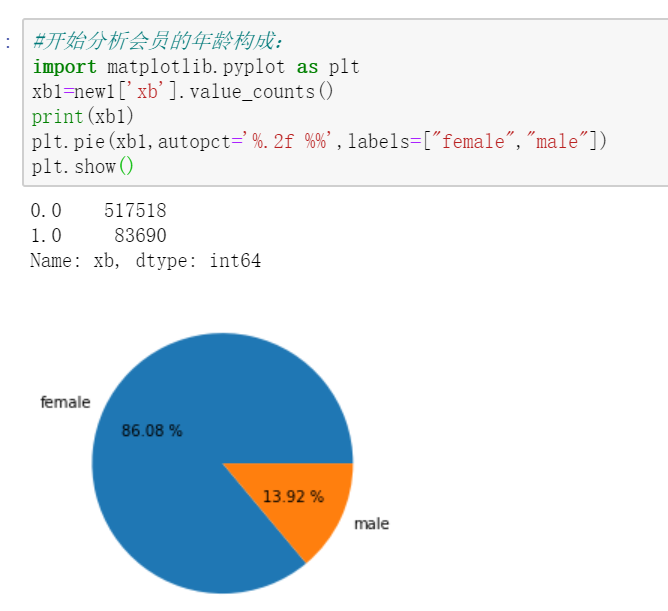
1. **数据可视化初步分析**

4.1 分析会员的性别构成

首先出生年月一栏格式转化为日期，生成年龄属性。再去掉年龄异常值。



利用valuecount统计并绘图



得出结论，女性消费者的占比85%以上，占绝大多数。

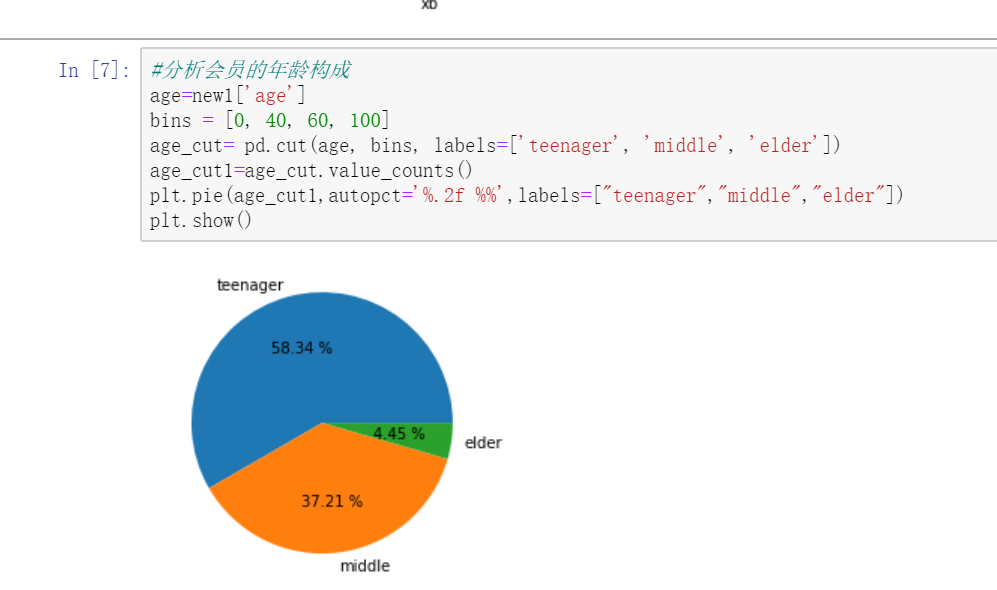
4.2 分析男性女性会员的消费金额



由图可以发现，女性消费金额远远超过男性。

**4.3分析消费者的年龄构成**

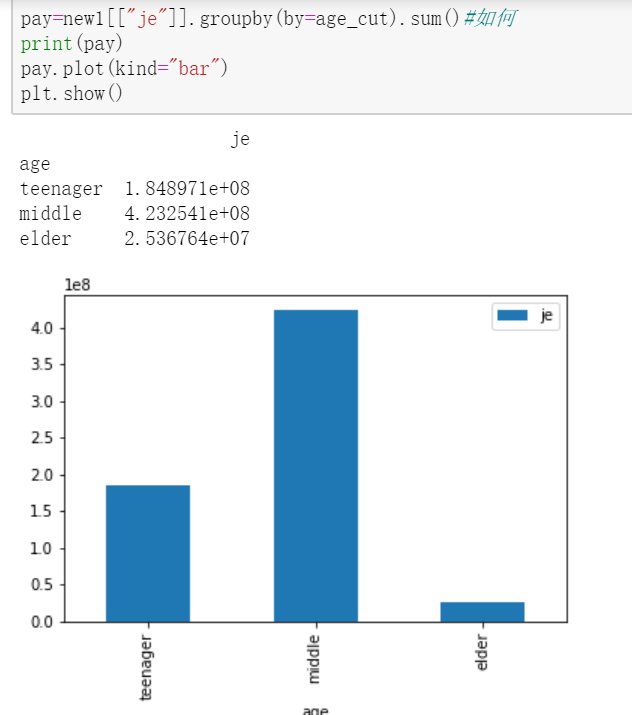
首先出生年月一栏格式转化为日期，生成年龄属性。再去掉年龄异常值。



可以发现，青年人是消费的主力军，老年人消费最少。

**4.4 分析不同年龄消费者的消费金额**

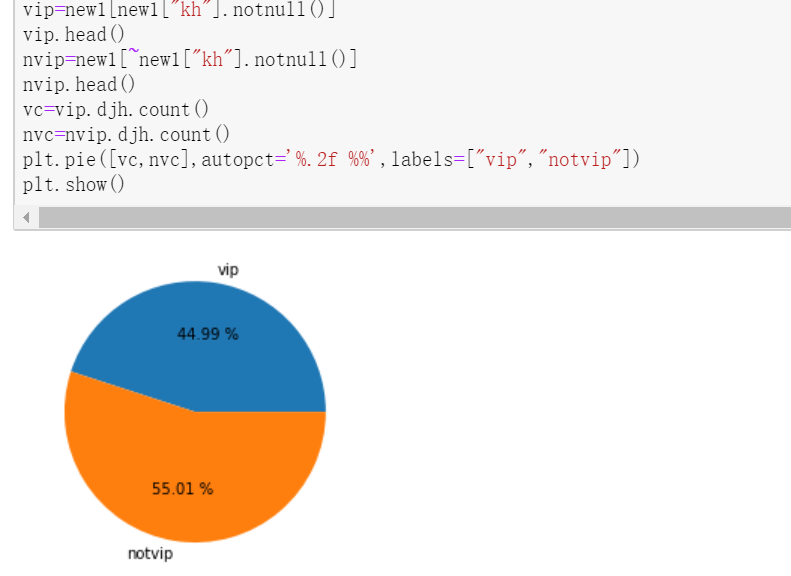
接着分析不同年龄群体的消费金额



由图可以看出，虽然中年人的消费人数不多，但它们却是贡献消费金额的主力军。

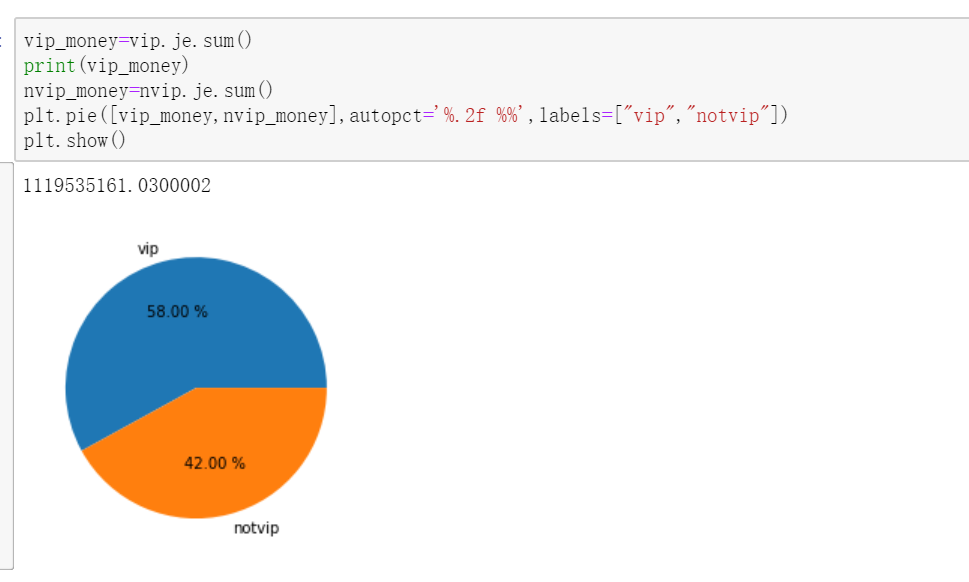
**4.5分析会员和非会员的人数占比**

通过分别提取会员和非会员的数量，绘制饼图



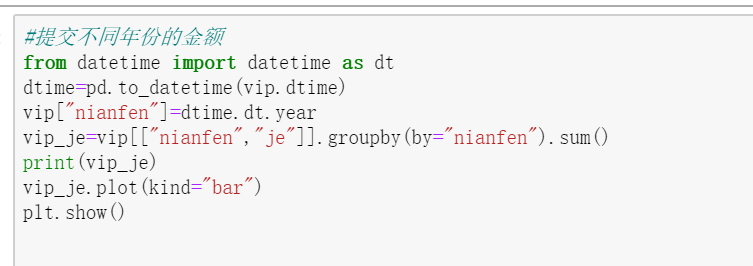
有分析可以发现，会员和非会员数量相差不多，但是会员数量占少数。

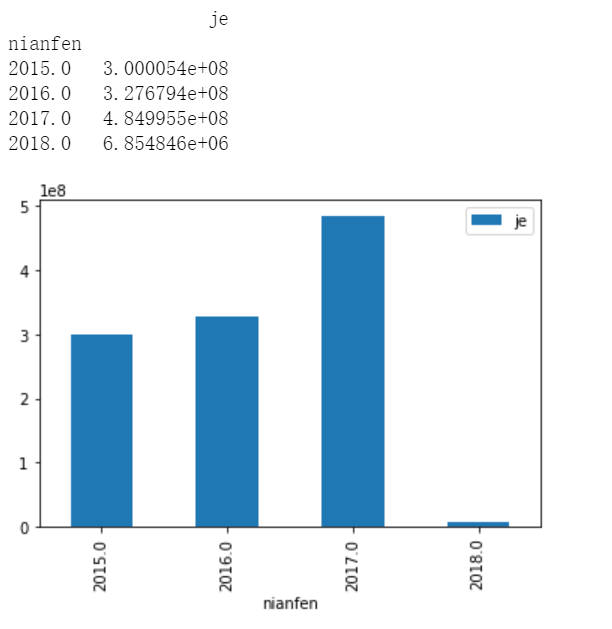
**4.6分析会员和非会员的消费金额**



根据饼图得出结论，会员的消费金额超过非会员的消费金额

**4.7统计不同年份的金额**

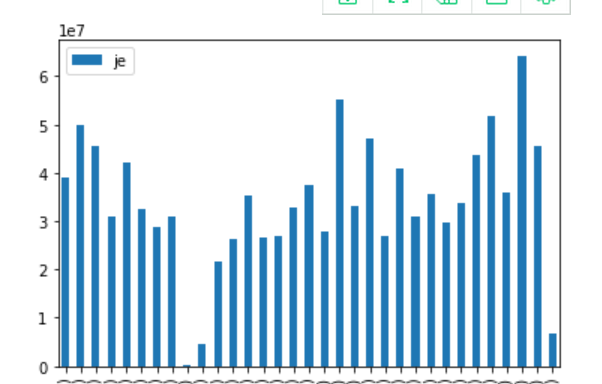




由图可以发现，2015-2017年消费金额不断上升，但是2018年消费金额较低。分析可能是统计金额不完善的问题或是2018年公司遭遇重创。

**4.8 分析每年里面不同月份的金额变化**

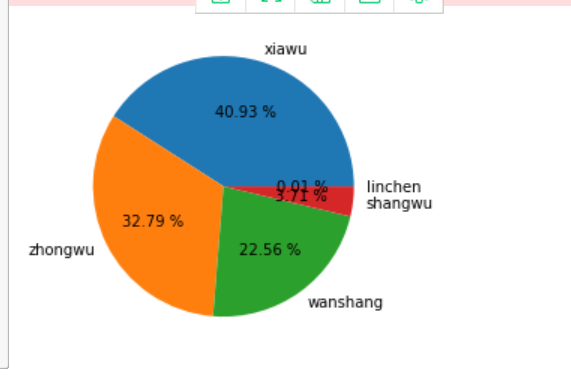




由图可以发现，销量呈现出波动变化的周期性形态。

**4.9 统计不同时间段的人数分布**

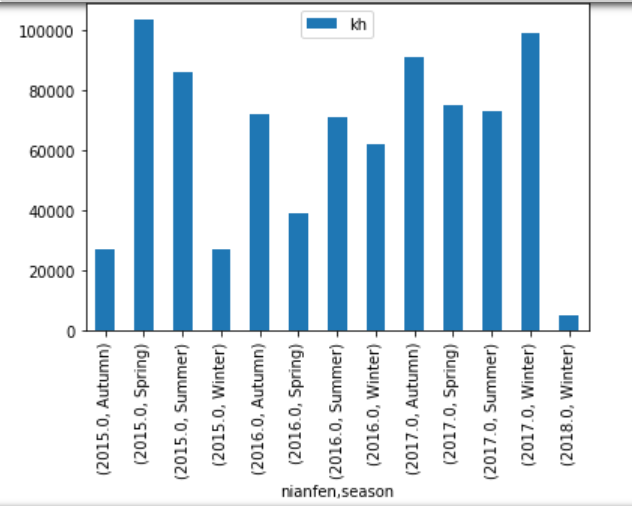




由饼图可以清楚地发现，下午逛商场的人数居多，凌晨逛商场的人数偏少。这也符合我们的认知。

**4.10 统计每年不同季度的人数分布**





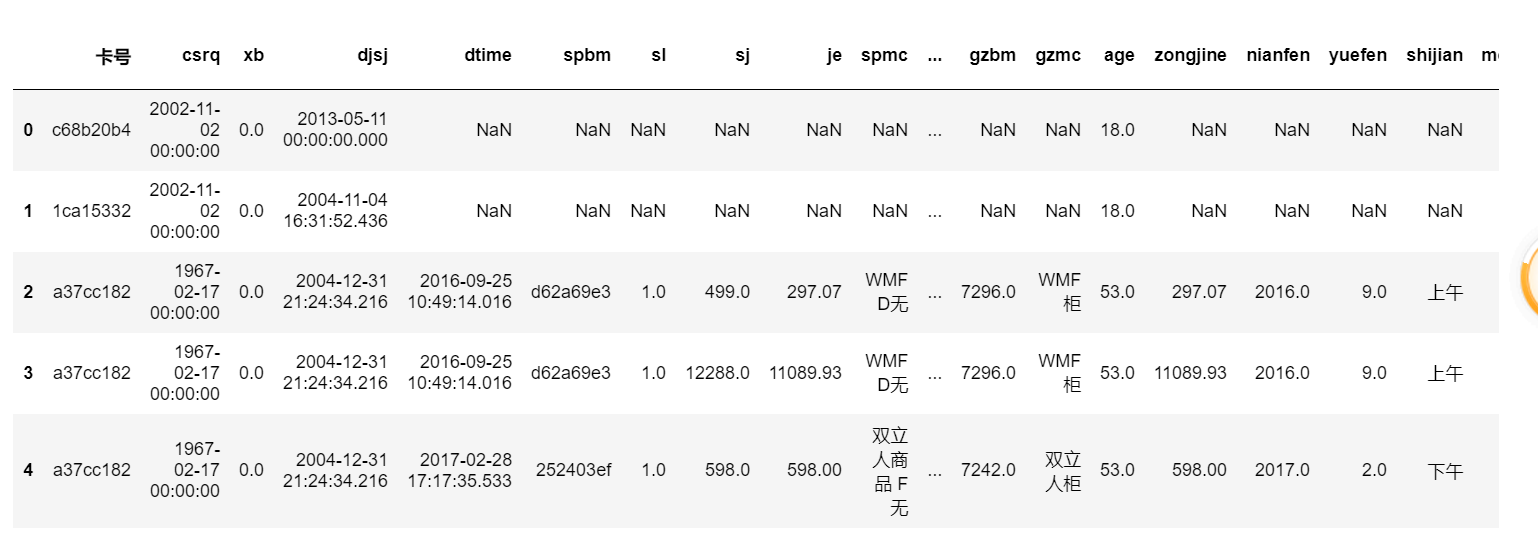
购物人数的起伏波动很大，似乎没有明显的规律。

1. **用户画像分析**

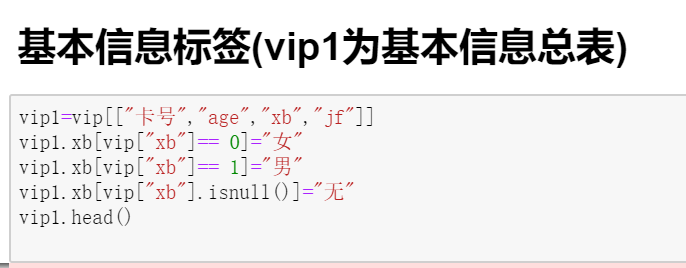
**这部分通过统计用户的基本信息、业务模式来对用户进行画像，采用词云进行可视化管理。**

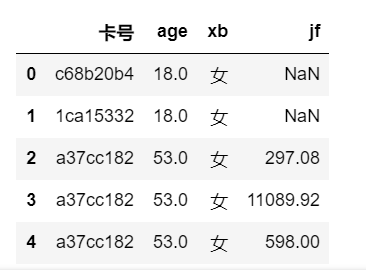
**首先我们导入数据。**



****

**5.1构建基本信息标签**





**5.2 构建用户业务信息标签**

**5.2.1消费频率标签构建**

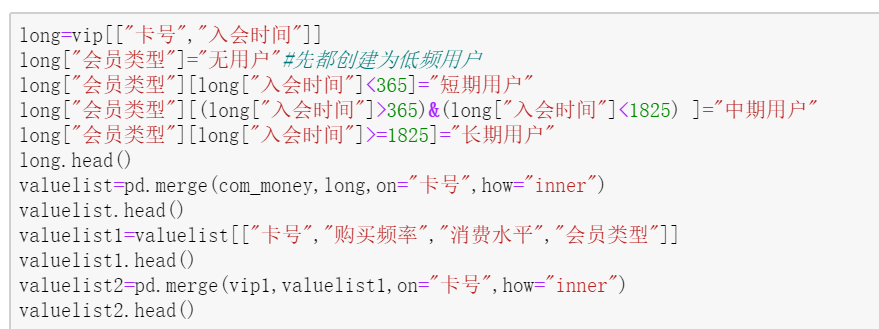




**5.2.2 消费水平标签构建**



**5.2.3 入会时间标签构建**





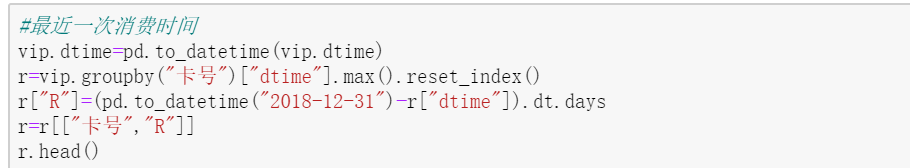
**5.2.4 词云构建，进行用户画像**

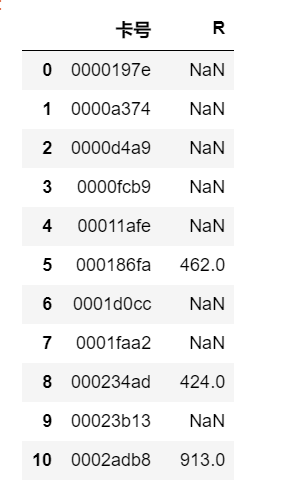


1. **用户价值分析**

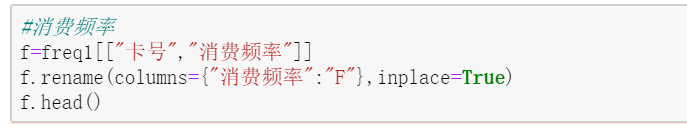
**用户价值分析采用RFM方法进行构建，分析出三类不同的客户群体，实行不同的精准营销策略。**

**1.构建R字段**





**2，构建F字段**

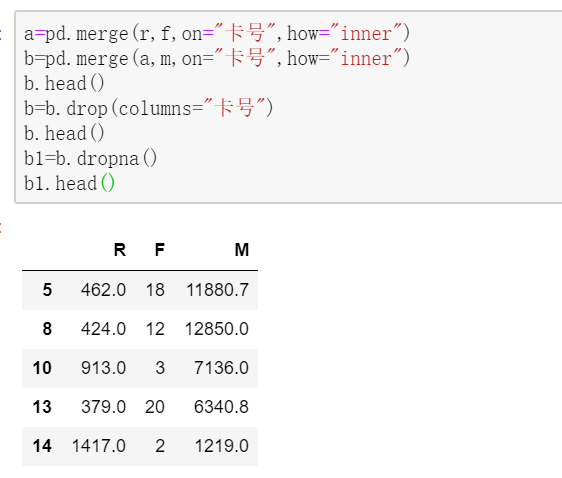




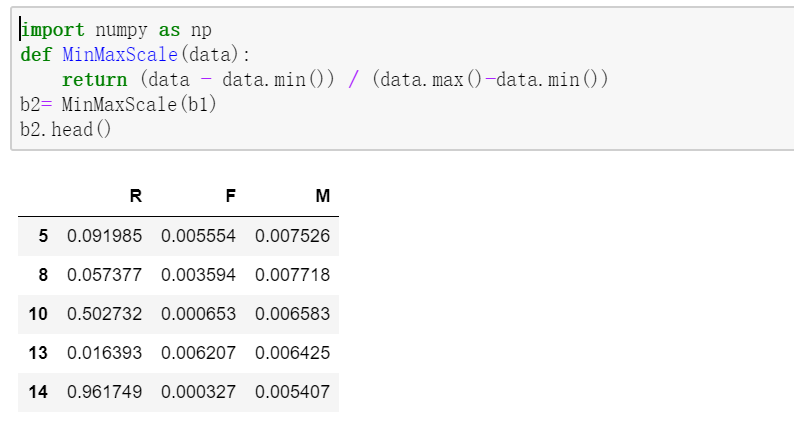
**3.构建M字段**



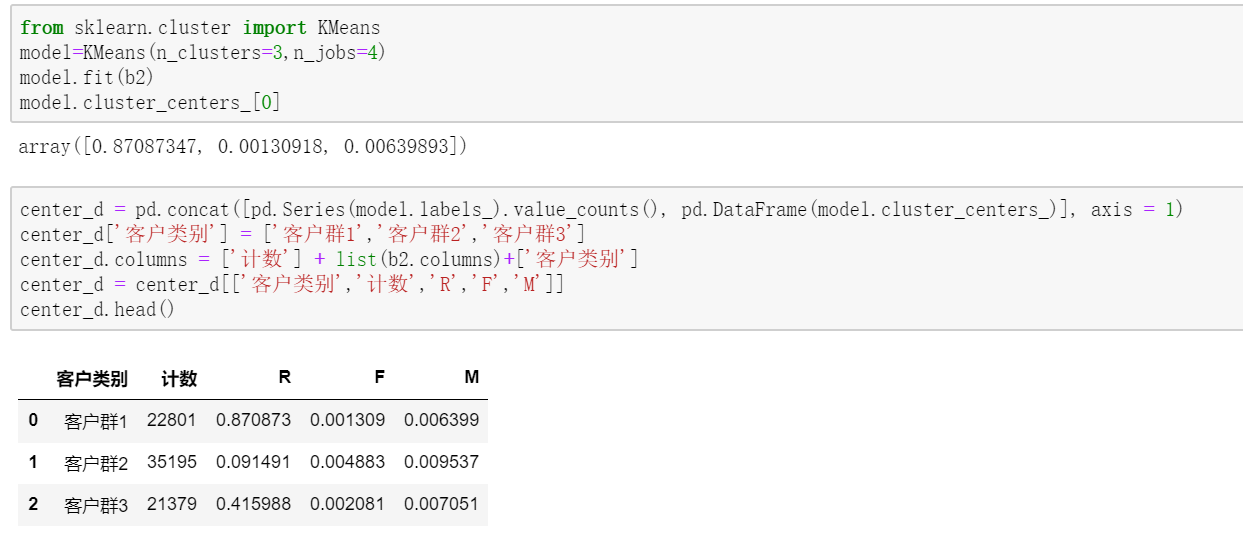
**4.合并三个字段，进行分析**



**5.对数据进行标准化处理**



**6.开始采用聚类分析**



**7.做雷达图分析客户状况**



