**电商平台关注Necklace的潜在购买用户数据分析项目**

* 1. **数据理解及分析思路**

**a.数据理解**

**某电商用户消费数据.xlsx**

包含字段：

'客户编码',

'产品编码',

'行为时间',发生五种行为的时间

'行为类别',包含'PageView', 'Order', 'Comment', 'SavedCart', 'Follow'五种行为

'年龄分段',

'性别',

'客户注册日期',

'会员级别',

'会员城市级别',

'产品品牌',

'店铺编码',

'产品类别',包含Necklace在内的53种商品

'产品上市日期',

'商家编码',

'粉丝数',

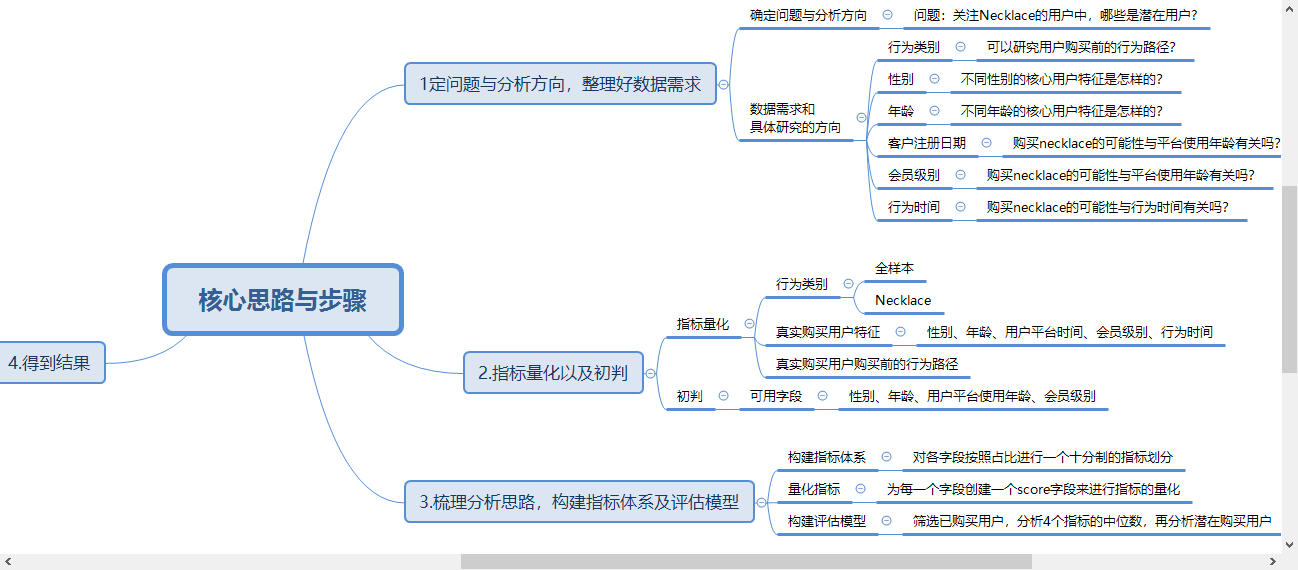
'会员数',

'开店时间',

'店铺主营',

'店铺打分'

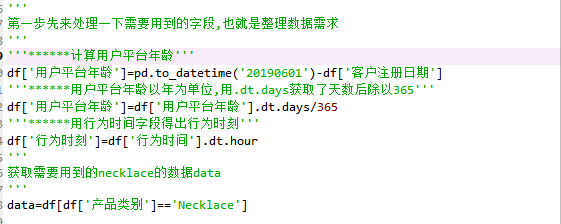
**b.分析思路**



* 1. **数据清洗**

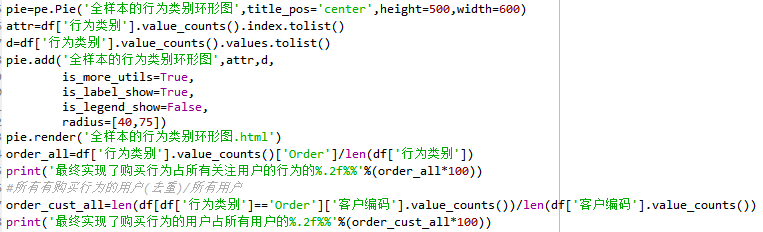
先计算可以直接计算出来的字段：用户平台年龄以及行为时刻

再把相关的数据提取出来

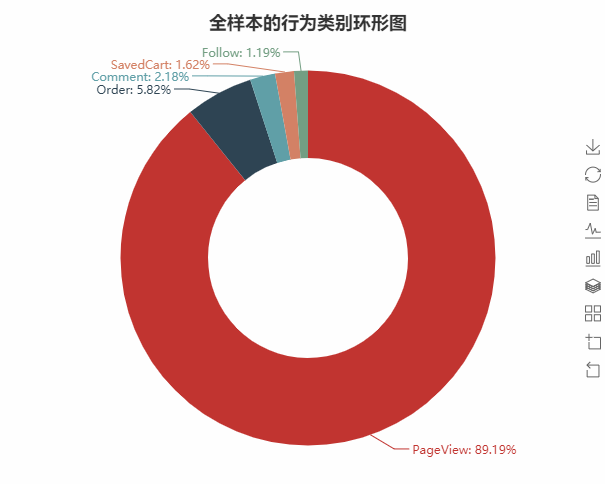


* 1. **指标的量化和初判**

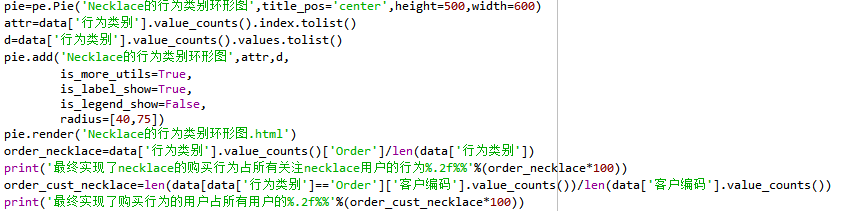
**a．先查看一下全样本的行为类别→计算购买率**



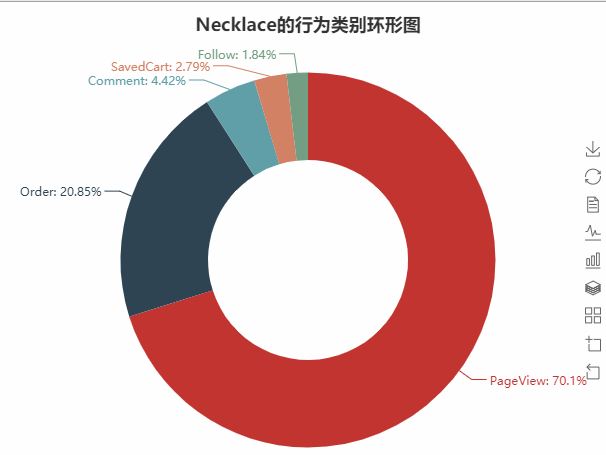




**b.再查看一下necklace的行为类别→计算购买率**







结论:

对于全样本：最终实现了购买行为占所有关注用户的行为的5.82%

最终实现了购买行为的用户占所有用户的7.57%

对于Necklace：

最终实现了necklace的购买行为占所有关注necklace用户的行为20.86%

最终实现了购买行为的用户占所有用户的21.16%

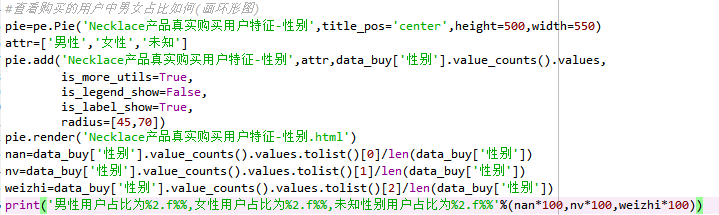
necklace两项数据均高于全样本数据,说明necklace引起用户的购买欲望高于平均水平，用户倾向于买necklace的可能性比较大,也就是说在这53个产品中，necklace是一个热门商品。

**c. 对有购买行为的用户的行为特征进行分析**

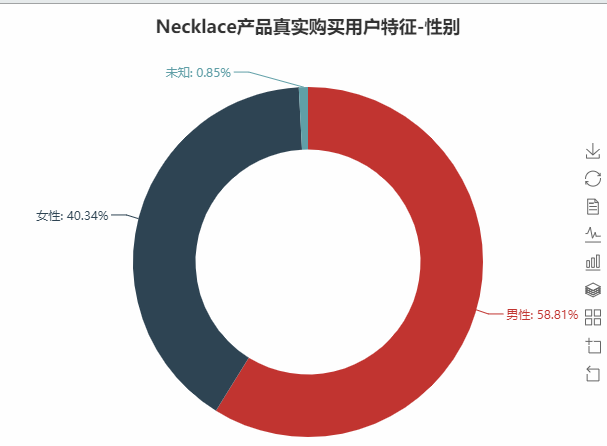
先获取一下有购买行为的用户数据



①查看性别占比



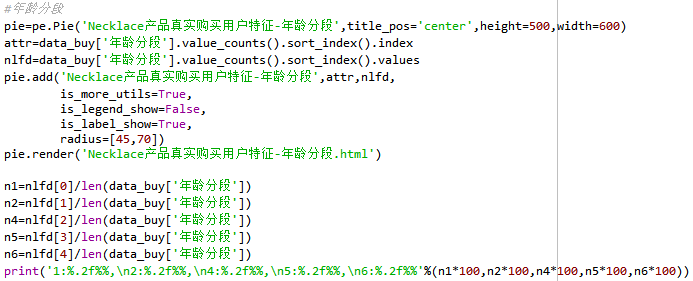


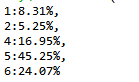


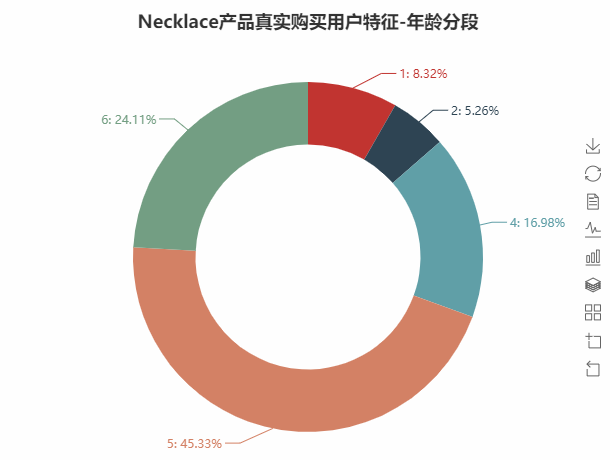
结论:

男性用户为59%,女性用户为40%,说明购买Necklace的男性用户比女性用户更多,按照常识,戴项链的女性比男性多,猜测这个结果是由于男性买项链送给女朋友老婆母亲所导致

②查看年龄分段占比



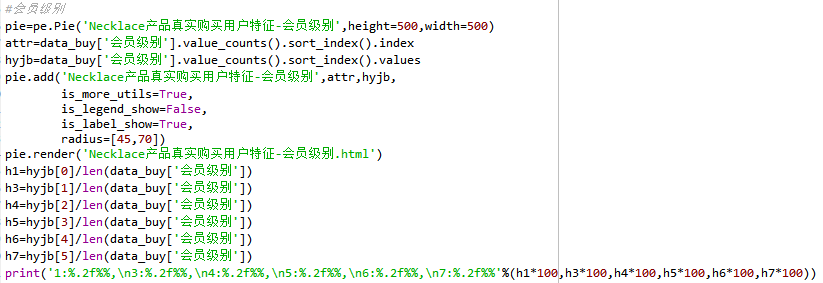




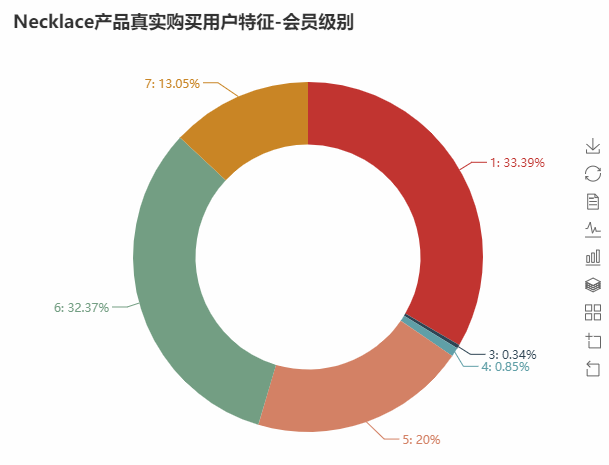
结论:

可以知道年龄段5购买的人数最多为45%,年龄段0购买的人数最少.所以这个Necklace在年龄段5中是比较受欢迎的

③查看会员级别占比



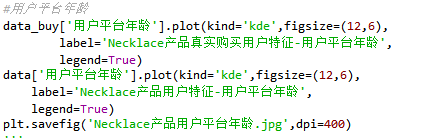


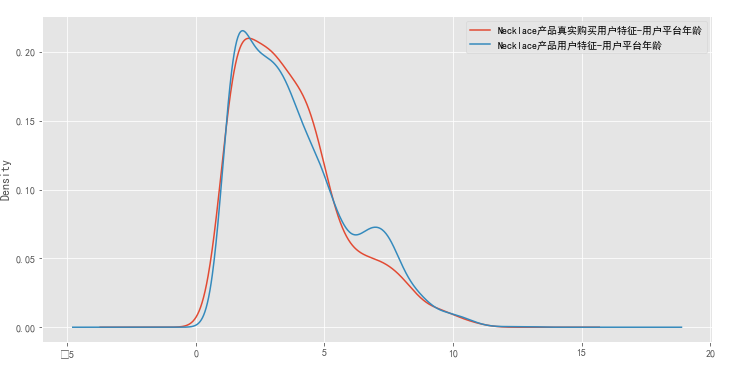


结论:

会员级别1跟6段购买最多34段购买最少

④查看用户平台年龄



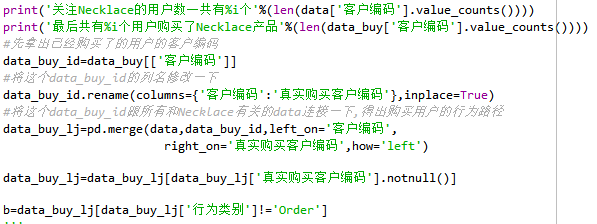


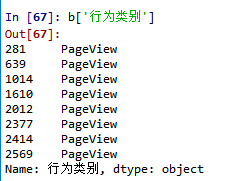
结论:

关注Necklace的用户平台使用年龄主要是2.5~7.5,其中2-3和6-7年用户有小波峰;

实际购买Necklace的用户平台使用年龄再2到-8年，可见老用户购买可能性低于新用户，或者说老用户更纠结

**d.真正购买用户购买前的购买路径分析**

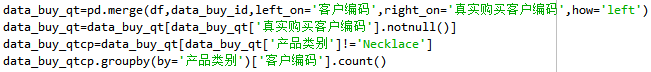


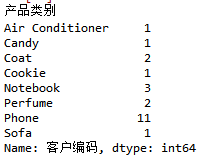


结论:

只有8个用户在购买前进行了浏览的行为,说明用户购买的目的还是挺明确的,也不怎么犹豫

**e.购买Necklace的用户还看过什么其他的产品**





结论:

购买Necklace的用户还有11个看了Phone,3个看了Notebook还有PerfumeCoat等等

**初判：**

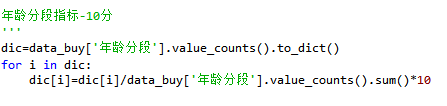
**性别、年龄分段、会员级别、平台使用年龄对购买Necklace的可能性有所影响**

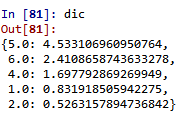
**而行为时间,购买前的路径分析等都没有太大影响,因此主要用这4个有效字段来进行分析**

* 1. **构建指标体系以及评估模型**

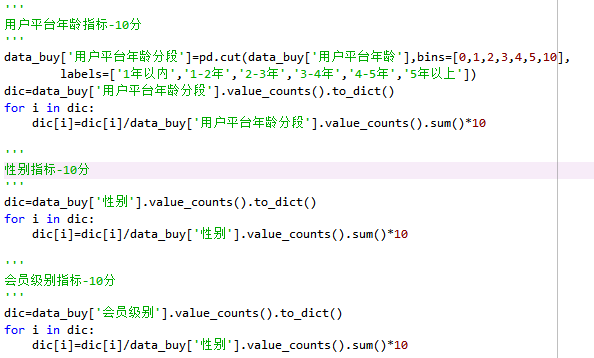
通过上一步基本确定了可用的字段为：性别、年龄、会员级别、用户平台使用年龄

对各字段按照占比进行一个十分制的指标划分

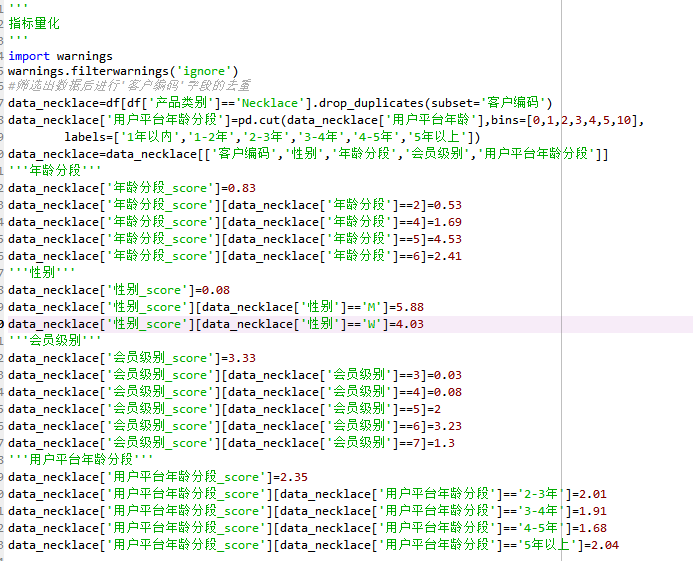




其他的字段也是类似

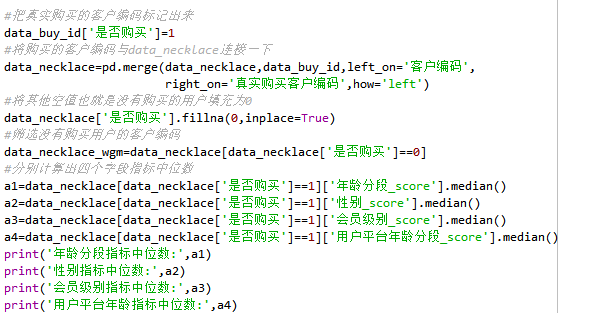


为每一个字段创建一个score字段来进行指标的量化

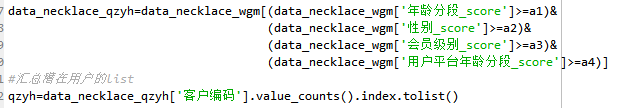


构建一个评估模型体系

筛选已购买用户，分析4个指标的中位数



对于未购买用户,如果四个指标值都大于上面得出的中位数,则认为是潜在用户(即购买可能性高)



最终得出270个购买可能性较高的潜在用户list