

百货商场用户画像描绘和价值分析

姓名:许玉燕

时间：2020/2/10

**项目分析报告**

目录

[一、 背景与挖掘目标 1](#_Toc32487190)

[1. 项目背景 1](#_Toc32487191)

[2. 挖掘目标 1](#_Toc32487192)

[二、 分析过程与方法 1](#_Toc32487193)

[1. 数据预处理 1](#_Toc32487194)

[2. 数据分析 2](#_Toc32487195)

[1. 统计会员和非会员的订单数 2](#_Toc32487196)

[2. 统计会员和非会员的消费金额 3](#_Toc32487197)

[3. 会员年龄占比 5](#_Toc32487198)

[4. 统计不同年龄的消费金额 6](#_Toc32487199)

[5. 统计会员性别人数比例和不同性别消费金额比例 7](#_Toc32487200)

[6. 统计不同年份会员的消费金额 9](#_Toc32487201)

[7. 统计会员不同年份各月份消费金额 10](#_Toc32487202)

[8. 统计一天中各个时段的会员人数 11](#_Toc32487203)

[9. 统计不同年份每个季节会员消费人数 12](#_Toc32487204)

[3. 小结 13](#_Toc32487205)

[三、 构建用户画像 16](#_Toc32487206)

[1. 什么是用户画像 16](#_Toc32487207)

[2. 做用户画像的意义 16](#_Toc32487208)

[3. 用户画像的基本要素 16](#_Toc32487209)

[4. 构建用户画像的步骤 17](#_Toc32487210)

[5. 利用Python生成用户画像 18](#_Toc32487211)

[四、 客户价值分析 20](#_Toc32487212)

[1. 分析方法 20](#_Toc32487213)

[2. 数据挖掘 20](#_Toc32487214)

[3. 数据预处理 21](#_Toc32487215)

[4. 模型构建 21](#_Toc32487216)

[5. 客户价值分析 25](#_Toc32487217)

[五、 模型应用 26](#_Toc32487218)

[六、 小结 26](#_Toc32487219)

# 背景与挖掘目标

## 项目背景

1、互联网时代，电子商务进入迅猛发展阶段，零售业逐渐被淘汰，会员成为零售产业支撑下去的主要力量。

2、针对不同类型客户，进行精准营销，实现利润最大化；

3、建立合理的客户价值评估模型，进行客户分类，是解决问题的关键。

## 挖掘目标

1、借助某商场客户数据，对客户进行分类；

2、对不同的客户类别进行特征分析，比较不同类客户的客户价值；

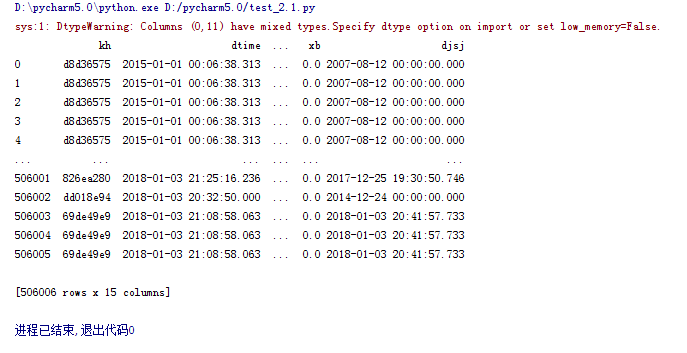
3、对不同价值的客户类别提供个性化服务，制定相应的营销策略。

# 分析过程与方法

## 数据预处理

主要包括数据合并、数据清洗、数据标准化、数据转化。本次商场提供的数据可用merge函数进行主键合并，如图。

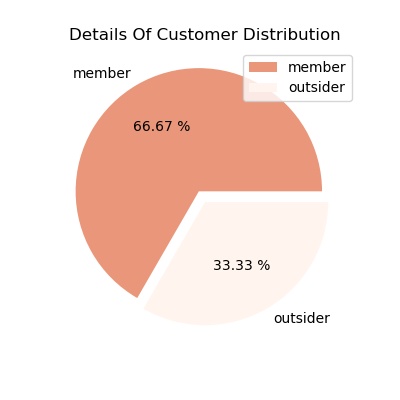
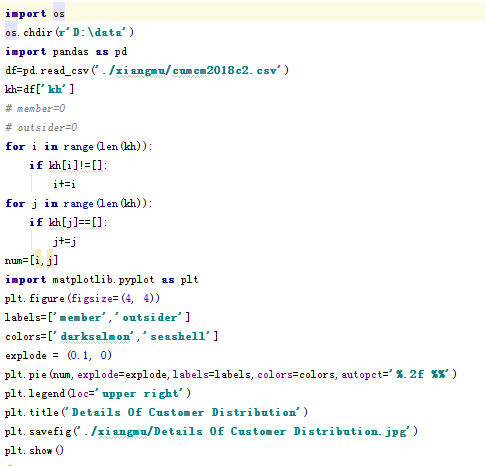




## 数据分析

### 统计会员和非会员的订单数

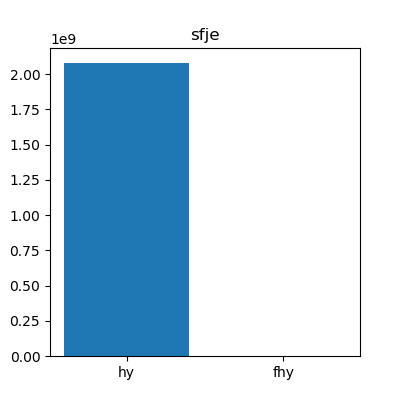
通过统计会员和非会员的订单数，探究该商场是否要将重点放在发展新会员上，还是需要保持和老会员的联系，来产生新的价值。下面用Python将结果进行可视化，如图。



### 统计会员和非会员的消费金额

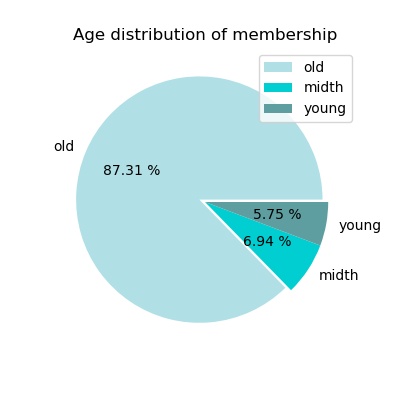
通过统计会员和非会员的消费金额，了解两者之间的购买力区别，哪一方能给商场带来更大的价值。下面用Python将结果进行可视化，如图。





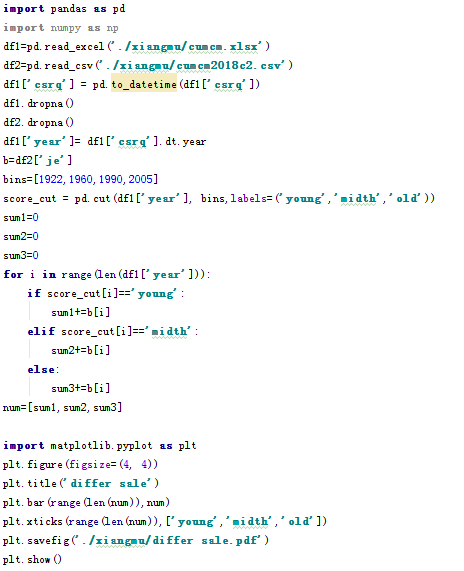
### 会员年龄占比

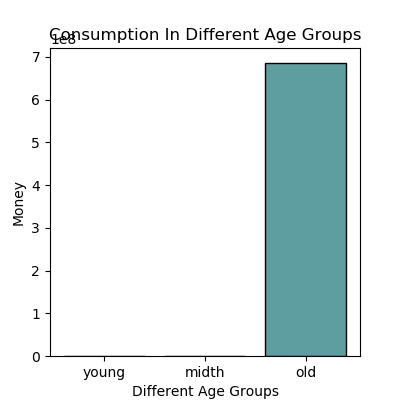
将会员年龄划分为老年、中年、青年，用年龄占比饼图实现。下面用Python将结果进行可视化，如图。



### 统计不同年龄的消费金额

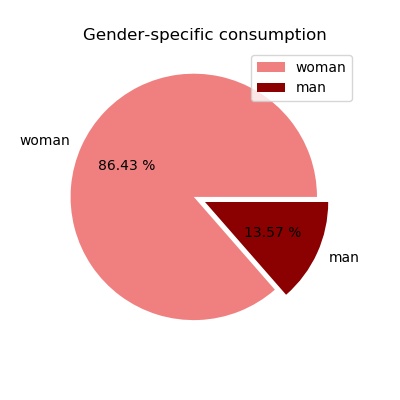
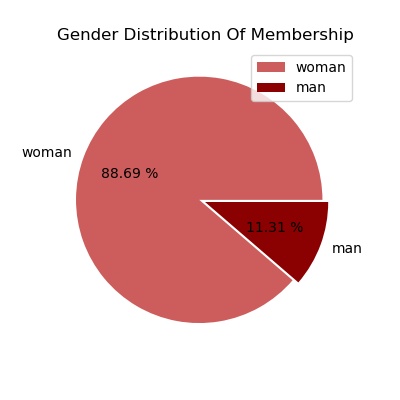
通过统计不同年龄的消费金额，了解不同年龄层的消费力情况，哪一个年龄层消费占比最大，重点照顾此年龄层的人，提供他们需要的产品。下面用Python将结果进行可视化，如图。





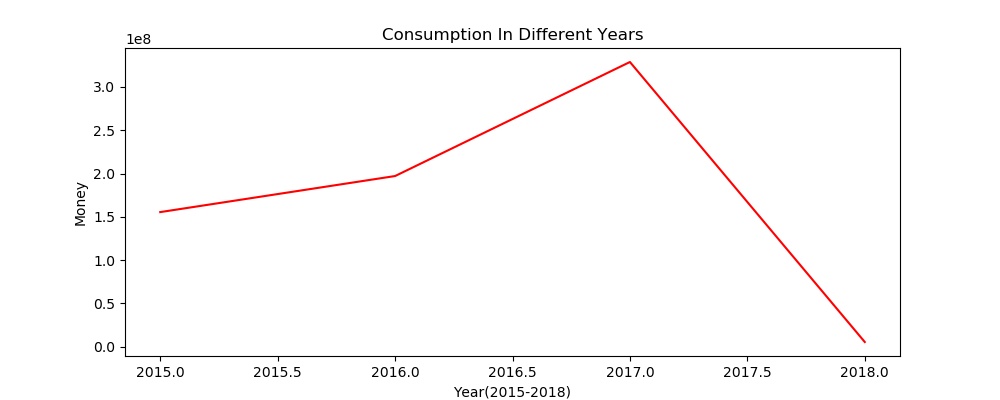
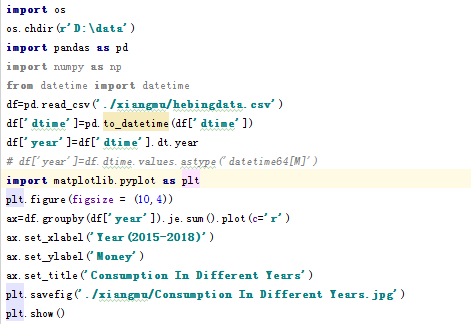
### 统计会员性别人数比例和不同性别消费金额比例

通过统计会员男女比例，了解哪一方占比大，商场需更多的满足某一方的需要。统计不同的性别的消费比例，了解谁的消费金额较少，可以提供一些折扣优惠，吸引其购物。下面用Python将结果进行可视化，如图1-1。



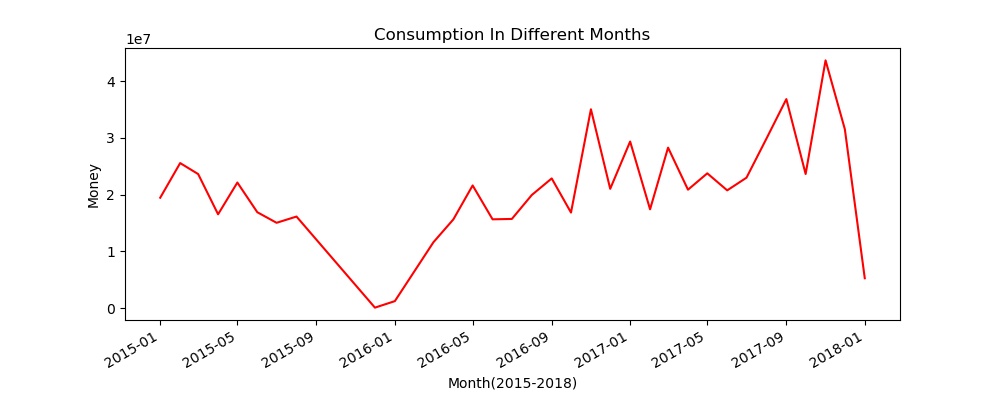
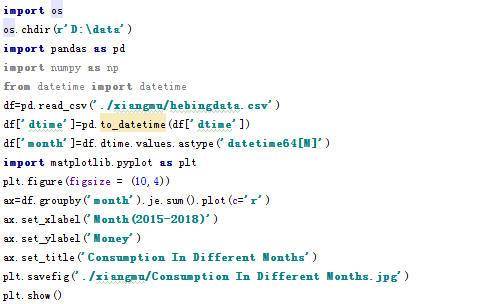
### 统计不同年份会员的消费金额

通过统计不同年份会员的消费金额，探究会员近年来的消费趋势，是否有所上升或下降。下面用Python将结果进行可视化，如图。



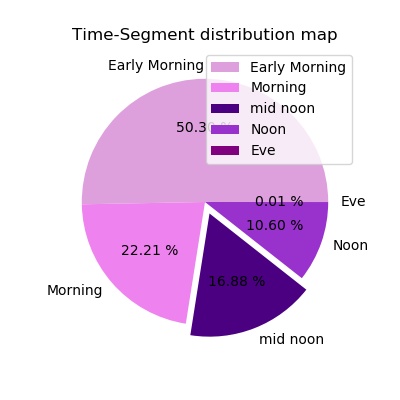
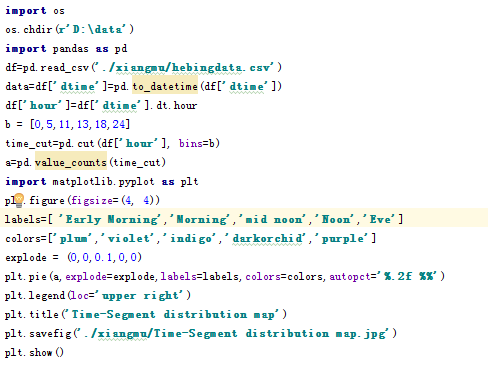
### 统计会员不同年份各月份消费金额

通过统计会员不同年份各月份消费金额，看不同月份消费额之间的差异，了解会员更喜欢在哪一些月份购物。下面用Python将结果进行可视化，如图。



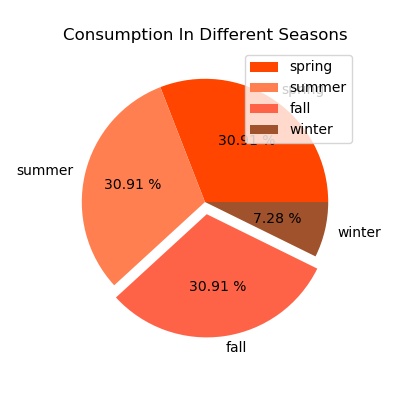
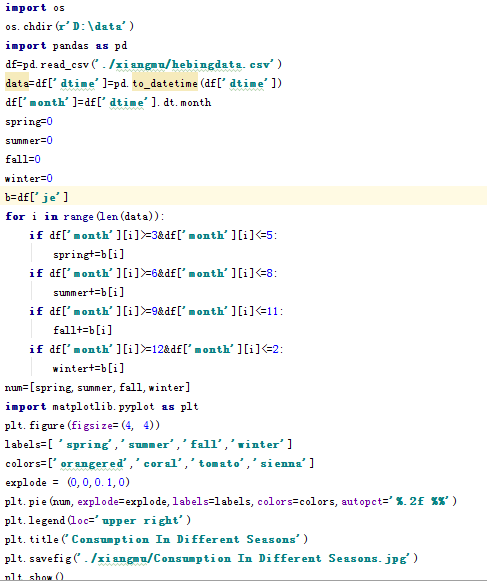
### 统计一天中各个时段的会员人数

将会员消费时间分为一天的早上、中午、下午、晚上、凌晨。了解会员更喜欢在哪一时段购物。下面用Python将结果进行可视化，如图。



### 统计不同年份每个季节会员消费人数

通过统计不同年份每个季节会员消费人数，了解会员更喜欢在哪个季节消费。下面用Python将结果进行可视化，如图1-1。



## 小结

* 通过[会员和非会员的订单数](#_统计会员和非会员的订单数)和[消费金额](#_统计会员和非会员的消费金额)的区别，我们可以看到会员的占比人数比非会员多出来一倍左右，且非会员的消费金额相比较于会员来说几乎接近0，说明该商场主要通过会员消费来获利，因此发展与会员之间的关系相当重要。



* 通过[会员年龄占比](#_会员年龄占比)和[不同年龄层的消费金额](#_统计不同年龄的消费金额)的区别，可以看到老年人是该商场的主要客户，并且老年人的消费也是最多的，但是一般老年人都是比较节省的，所以商场可以提供一些优惠活动，商场内也要多设置一些座椅之类的服务。



* 通过[会员性别占比](#_统计会员性别人数比例和不同性别消费金额比例)和[不同性别的消费金额](#_统计会员性别人数比例和不同性别消费金额比例)比例的区别，可以看出女性占比远大于男性，消费的金额也是，说明女性的老年人喜欢到该商场购物。再从[一天中各个时段的会员](#_统计一天中各个时段的会员人数)人数，可以看出凌晨和早晨的消费人数较多，说明奶奶们可能是去买菜，商场的生鲜蔬菜类供给应该要到位。



* 从[不同年份的消费金额](#_统计不同年份会员的消费金额)和[不同年份各月份的消费金额](#_统计会员不同年份各月份消费金额)，可以看出15，16年销售不是很好，而2017年的销售是比较好的，在2017年末达到消费高峰，说明将来会员在该商场消费的发展空间很大，商场需要加强与会员的联系。



# 构建用户画像

## 什么是用户画像

用户画像可以理解成是海量数据的标签，根据用户的目标、行为和观点的差异，将他们区分为不同的类型，然后每种类型中抽取出典型特征，赋予名字、照片、一些人口统计学要素、场景等描述，形成一个人物原型 。

## 做用户画像的意义

a.从公司战略层面来说，好的用户画像可以帮助企业进行市场洞察、预估市场规模，从而辅助制定阶段性目标，指导重大决策，提升ROI；更有助于避免同质化，进行个性化营销。

b.从产品本身角度来说，用户画像可以围绕产品进行人群细分，确定产品的核心人群，从而有助于确定产品定位，优化产品的功能点。

c.从数据管理角度来说，用户画像有助于建立数据资产，挖掘数据的价值，使数据分析更为精确，甚至可以进行数据交易，促进数据流通。

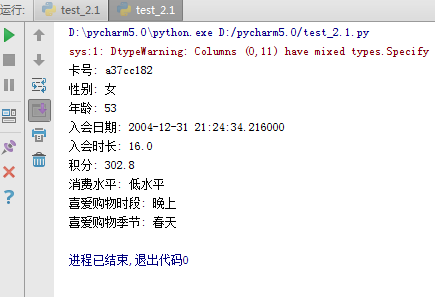
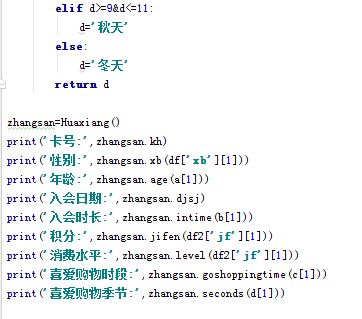
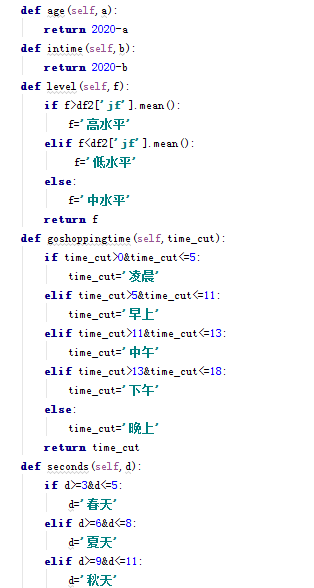
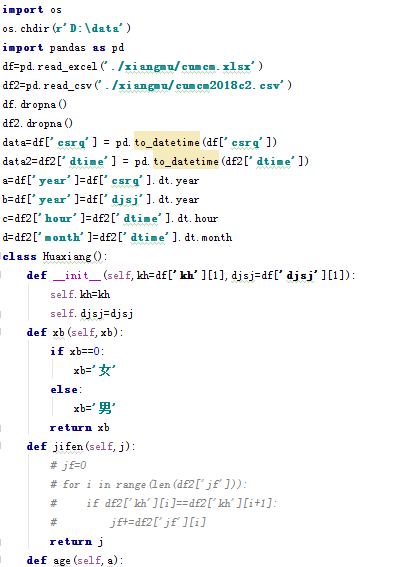
## 用户画像的基本要素

* 静态属性是勾勒用户画像的基础：性别、年龄、卡号、入会时间等。
* 动态属性是指用户商品偏好，直接反应了用户对你的商品是否感兴趣。
* 消费属性是指用户的消费水平，这反映的是用户对花钱有什么看法。到底是喜欢质量好的还是追求品质高的呢？到底是倾向于功能价值还是情感价值？

## 构建用户画像的步骤

* 数据清洗：原始数据源存在“脏数据”，包括数据空缺和噪声、不一致、重复、错误等问题，为了保证后期挖掘的准确性，避免对决策造成影响，须对原始数据进行预处理。
* 数据标准化：用户画像的建立需要有整合多源数据（跨屏跨媒体）的能力，例如一个实体可能使用多个设备，拥有网络世界的多个账号，则须把多个身份ID组合，建立统一的标准，才能完整标识实体的用户画像。
* 用户建模：通过算法模型来定义人群的用户画像，常见为分类模型和聚类模型，例如朴素贝叶斯、决策树、SVM，神经网络，k-means等。
* 数据可视化：即视觉呈现群体或个人的用户画像，包括柱状图/饼状图/表格等，可根据实际情况选择合适的可视化方式。

## 利用Python生成用户画像



# 客户价值分析

## 分析方法

1、首先，明确目标是客户价值识别；

2、识别客户价值，应用最广泛的客户细分模型是通过三个指标：最近消费时间间隔（Recency）、消费频率（Frequency）、消费金额（Monetary）；

3、以上模型简称 RFM 模型，作用是：识别出高价值的客户。

4、本案例以 LRFMC 五个指标为基础，采用聚类（K-Means）的方法，识别客户价值。

## 数据挖掘

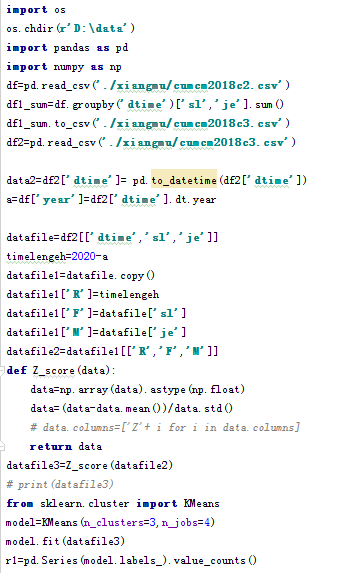
* 从数据源进行数据抽取，选择性抽取与新增数据抽取，形成历史数据和增量数据；
* 进行数据探索性分析和预处理，主要有缺失值与异常值的分析处理，数据的属性规约、清洗和变换；
* 利用处理后的数据作为建模数据，基于会员价值的RFM模型进行客户分群，对各个客户群再进行特征分析，识别有价值客户；
* 针对模型结果得到不同价值的客户，采用不同的营销手段，提供定制化的营销服务，或者针对性的优惠与关怀。（重点维护老客户）。

## 数据预处理

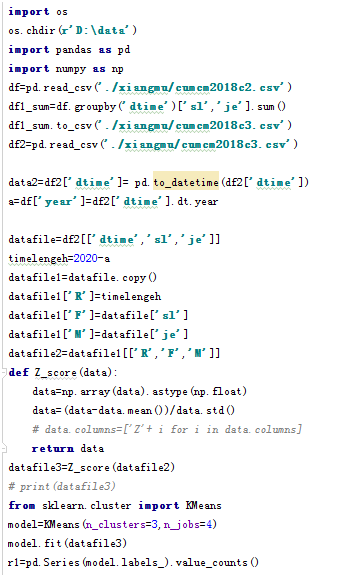
* 数据清洗：用dropna()函数丢弃会员卡号为空的记录；
* 属性规约：原始数据中属性太多，根据某商场客户价值RFM模型，选择与模型相关的三个属性。删除不相关、弱相关或冗余属性，如会员卡号、性别等。与模型相关的三个属性分别是：sl、je、dtime。

## 模型构建

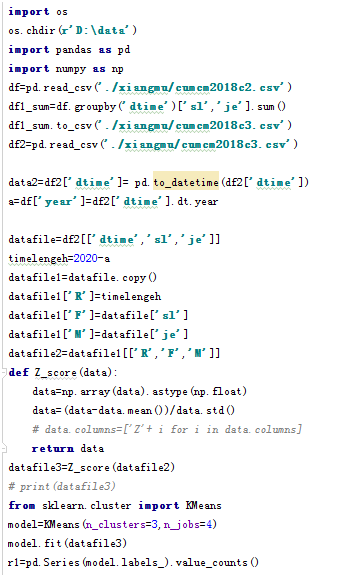
* 数据读取：读取原文件数据，按消费时间（dtime）来分类对销售量（sl）和消费金额（je）进行求和。再转化时间格式。



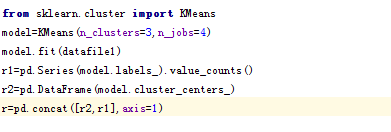
* 属性提取: 原始数据中属性太多，选择与模型相关的3个属性。与模型相关的六个属性分别是：dtime、sl、je。
* 数据变换：数据变换就是将数据转换成“适当”的格式，以适应挖掘任务及算法的需要；



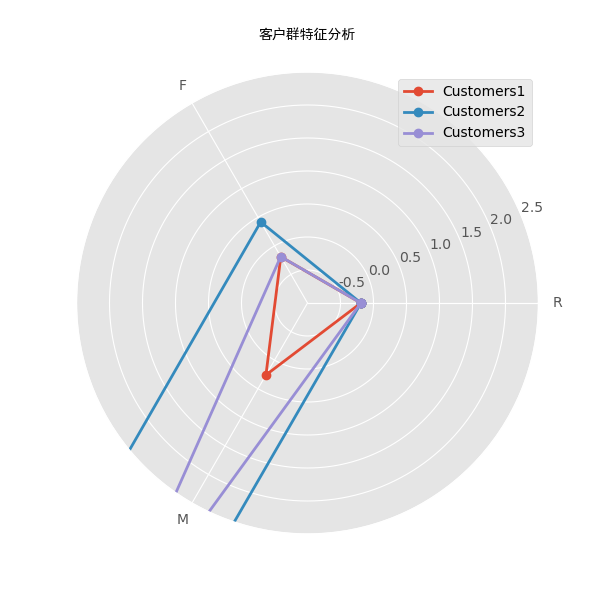
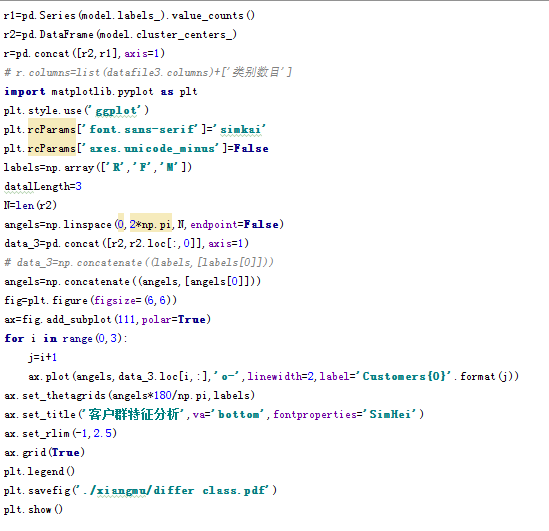
* 5个指标的取值范围差异较大，需进行标准化处理，消除数量级的影响；



* 客户聚类：采用K-Means算法对客户数据进行客户分群，聚成3类（需结合业务的理解与分析来确定客户的类别数量）



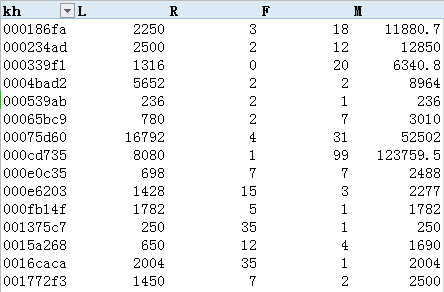
* 绘制客户群特征雷达图：

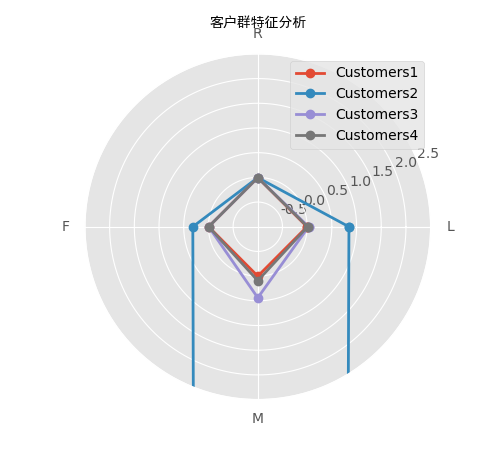
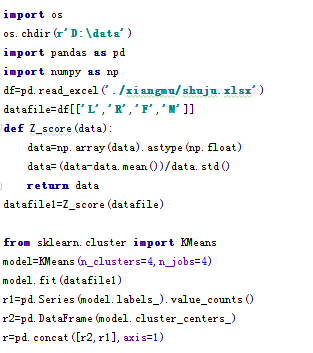


* **方案改进**

上面是直接用读取原来文件里的数据，进行计算，再画图的方法，计算方法是自定义的，如最后一次购买时间距离现在多久是用年份去计算，误差可能较大。

下面我在源数据上先做一个透视表，把RFM值筛选出来，再添加L（购买单笔最大金额），保存文件，修改一下代码，这样的数据准确度会高一些。





## 客户价值分析

针对“客户群特征分析图”所示的聚类结果。我们重点关注的是L（购买单笔最大金额）、F（消费频率）、M（消费金额）三个指标：

* 客户群1：在LRFM四个指标上的分布度不是很高，可定义为一般低价值客户，这里客户很少到商场购物，购物金额和频率也不高。
* 客户群2：R（最近一次购买时间间距）较低，但LFM三个指标都比其他几类客户高，可归类于重要保持客户，此类人群是上次获利的来源，需要加强联系。
* 客户群3：LRF三类指标都一般，但M（消费金额）偏高，说明此类客户消费频率不高，但单次购买金额高，有一定消费力，可定义为重要发展客户。
* 客户群4：此类客户和客户群1很相似，也是一般低价值客户。

# 模型应用

* 针对顾客中大部分是中老年女性，且在凌晨和早上的消费额较大，推断该人群可能是去购买食品，可对一些销量高的食品进行打折优惠，如买二送一，购满188减38，并在此时间段商场应该增加服务人员，给予顾客好的消费体验。
* 对于顾客群2、3，可对会员绑定的手机发一些新品推送，优惠特价活动。可在小程序上进行签到，每次签到可换取相应的积分，积分可抵扣消费金额，这样既有利于加强商场和顾客的联系，也可吸引尚未加入会员的人群。
* 对于顾客群1、4，此类人员消费较少，可进行挽留，激发消费，如在他们喜欢浏览的页面（抖音、…APP）发送一些本商场限时特价优惠活动，吸引其来抢购。

# 小结

* 获取新客成本远大于维持老客，应努力维系客户关系水平，保持优质客户。
* 差异化管理、精准营销，提高客户忠诚度、满意度，延长客户高水平消费。