

**任务1.1，任务1.2：**

略

**任务2.1**

第一题：任务2.1第一题要求将会员按年龄划分为：老年、中年、青年，并绘制会员年龄占比饼图。

第二题：针对不同年龄的消费金额，对结果可视化

第三题：统计不同性别人数和不同性别消费金额比例，并对结果进行可视化

**一、完成思路：**

1. 第一题通过对第一个数据表的会员的出生日期进行分析，得到当前时间年份，提取出生日期中的年份，两者相减得到年龄，并以此划分人群和作图。
2. 第二题在第一题的基础上，将不同年龄段所对应的消费金额分别加总，画出可视化结果图。
3. 第三题，将性别列提取出来，对不同性别类型分类并计数，并提取不同性别的消费金额列，对其进行可视化。

**二、基本步骤：**

**第一题：**

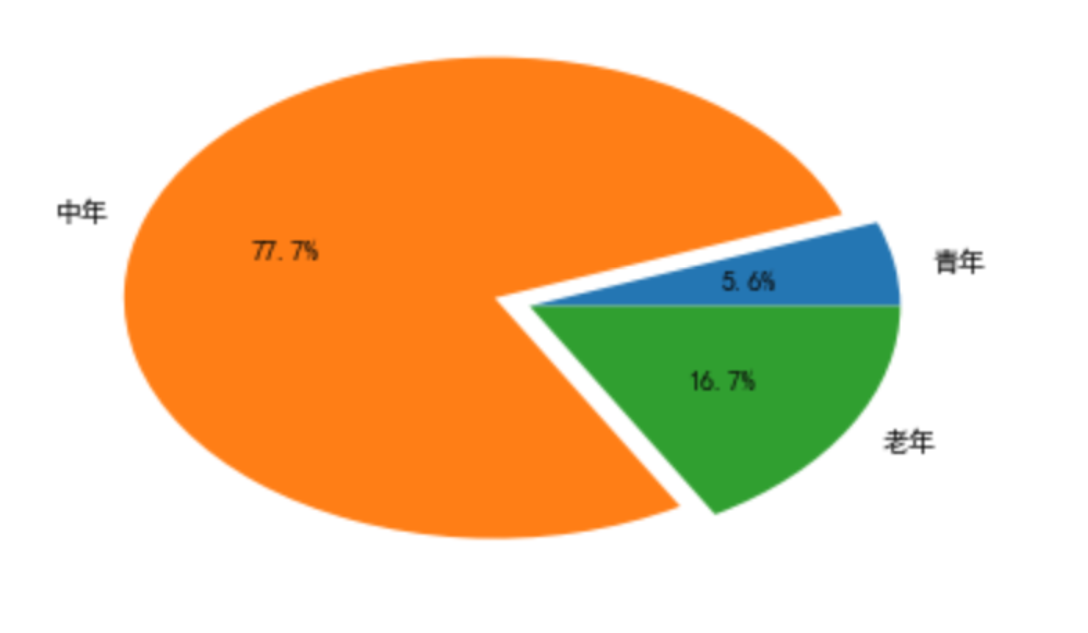
1、读取第一份关于会员基本信息的excel文件，将数据中出生日期的“csrq”一列按照时间格式读取，并去除其中出现异常值的出生时间。

2、创建空列表，用循环与append把出生日期中的年份放入列表，得到会员的年份列表。

3、获取现在的年份.datetime.now() ，利用datetime导出现在的时间的年份.year，创建空列表利用循环，把当前年份减去会员的出生年份的结果（即会员的年龄）放入一个列表。

4、再次利用循环与if-elif语句，把大于60岁的称为老年，介于30到60之间的称为中年，小于30的称为青年。分别创建三个空列表，按照上述划分把会员年龄放入不同的列表。

5、计算列表长度，从而得到各年龄层的人数总量，依此画出饼图。



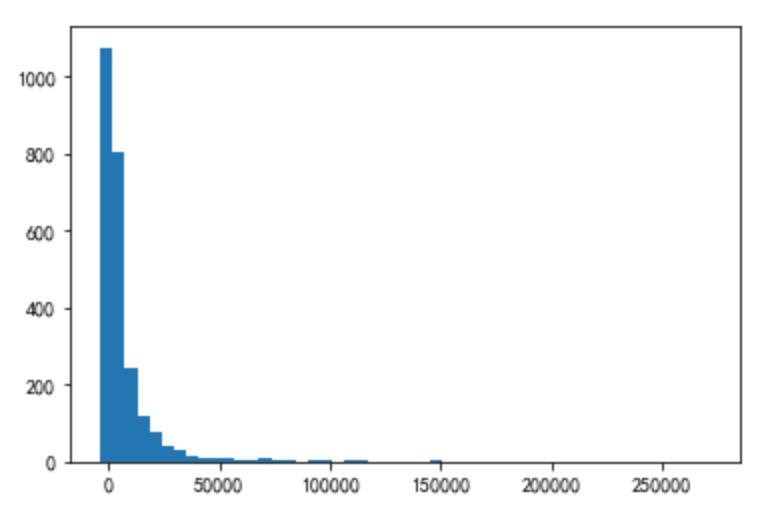
**第二题：**

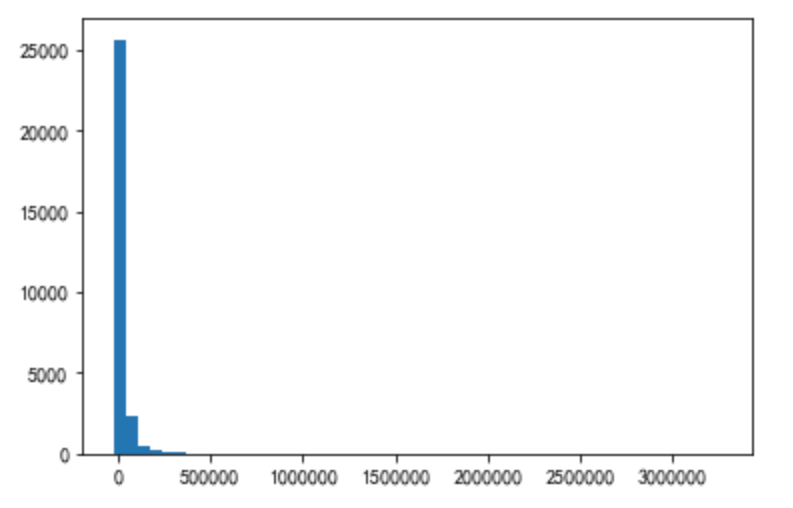
1、将年龄列表插入数据表中成为新的一列，并将两个数据表按照卡号为主键合并：data=pd.merge(data1,data2,on='kh') 。

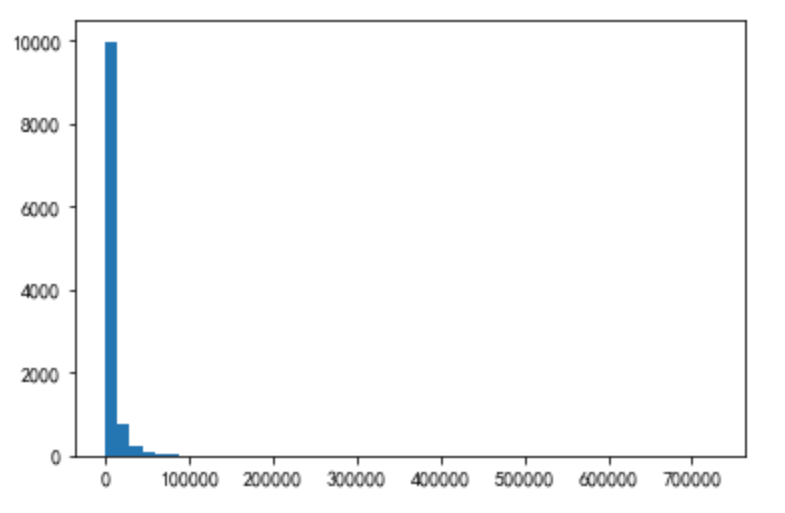
2、按照任务2.1的年龄分类，运用.loc[data['nl']<30,:]，将三个年龄段的数据分离开来变成三个数据表。

3、运用groupby和sum()对三个数据表对卡号“kh”进行聚类并求和.groupby('kh').sum()

4、提取聚类求和后的数据表的消费金额（即“je”列），运用hist画出直方图，由于数据分布极不均匀，所以设置bins=50，分别打印出三个年龄段的直方图。



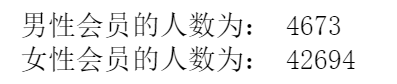




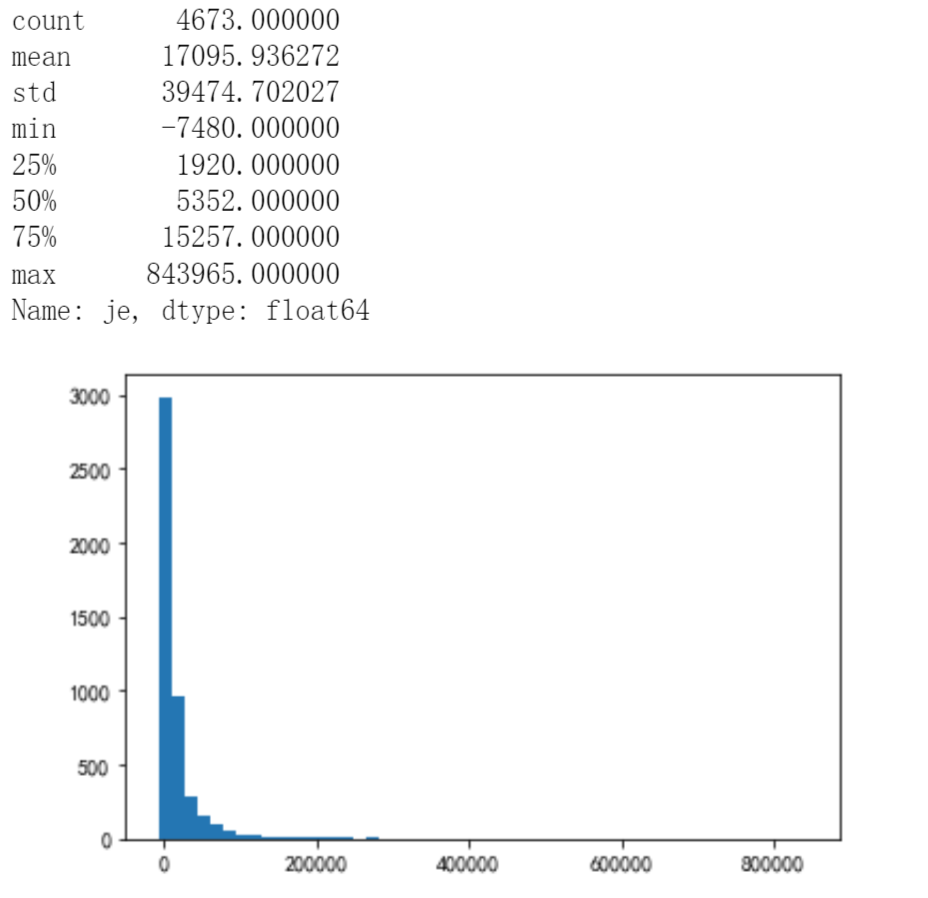
**第三题：**

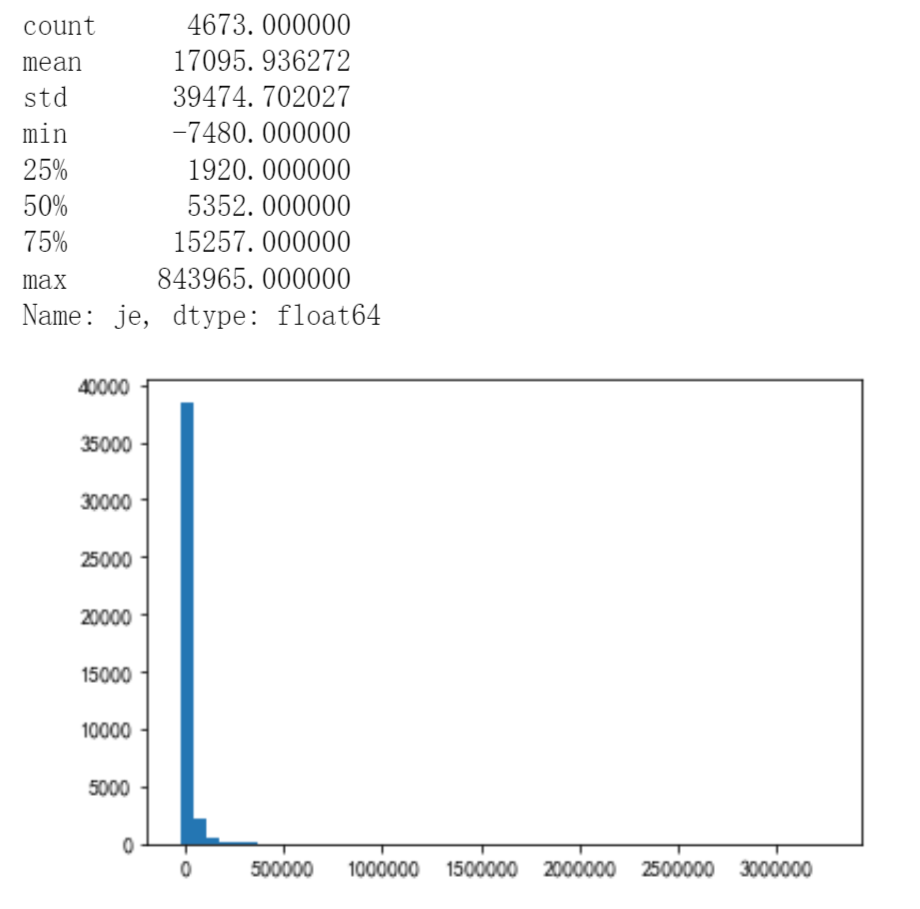
1、将已经主键合并后的数据表，按照“kh”列去重data.drop\_duplicates(['kh'])

2、提取性别“xb”一列转换为列表形式list(a['xb'])，创建两个空列表，将装有会员性别信息的列表利用循环分为男性和女性两个列表，分别计算两列表长度得到男性和女性的会员人数。



1. 将数据表运用data.loc[data['xb']==1,:] 划分为男性数据表和女性数据表，将男性女性数据表运用groupby对会员卡号“kh”聚类并加总data\_female.groupby('kh').sum()。
2. 提取聚类求和后的数据表的消费金额（即“je”列），运用hist画出直方图，由于数据分布极不均匀，所以设置bins=50，分别打印出男性和女性的消费直方图。





**任务2.2**

第一题：根据单据号字段，分别统计会员与非会员的订单数，并对结果进行可视化

第二题：根据消费金额字段，分别统计会员与非会员的消费金额，并可视

第三题：统计不同年份的会员消费金额，并对结果进行可视化

第四题：按不同年份统计会员各月份消费金额，并对结果进行可视化

1. **完成思路：**

一、将两数据表按卡号主键合并，提取合并后的数据表的单据号并对其计数，得到会员消费所有单据号数量，对第二张原始数据表，提取数据表的单据号并对其计数，得到总所有单据号数量。两者相减得到非会员的所有单据号数量，以此可视化。

二、对第二张原始数据表的金额一列进行加总，得到会员消费总金额，对主键合并后的数据表的金额一列进行加总，得到总消费金额，两者相减得到非会员的总消费量，由此可以分别得到会员的总消费量和非会员的总消费量，以此可视化

三、提取出生日期列，转换为时间格式，按照出生年份将其分为几个年龄段以便于可视化，再将不同年龄段的消费金额加总，对其可视化

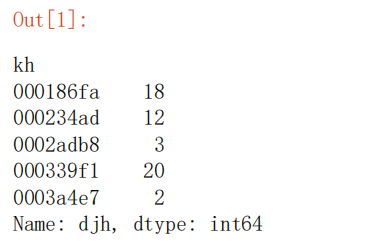
四、对上题分好的几个年龄段，逐个对各个年份的12个月的金额进行加总，比较各个年龄段客户的消费金额在12个月的占比。

**二、基本步骤：**

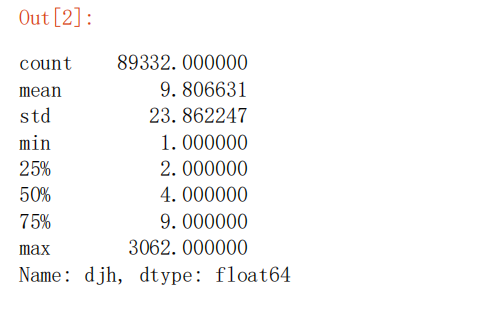
**第一题**

1、首先把两数据表以主键为会员卡号相连接，对卡号以groupby进行聚类，并使用count计数，hy\_djh=data2.groupby('kh').count()

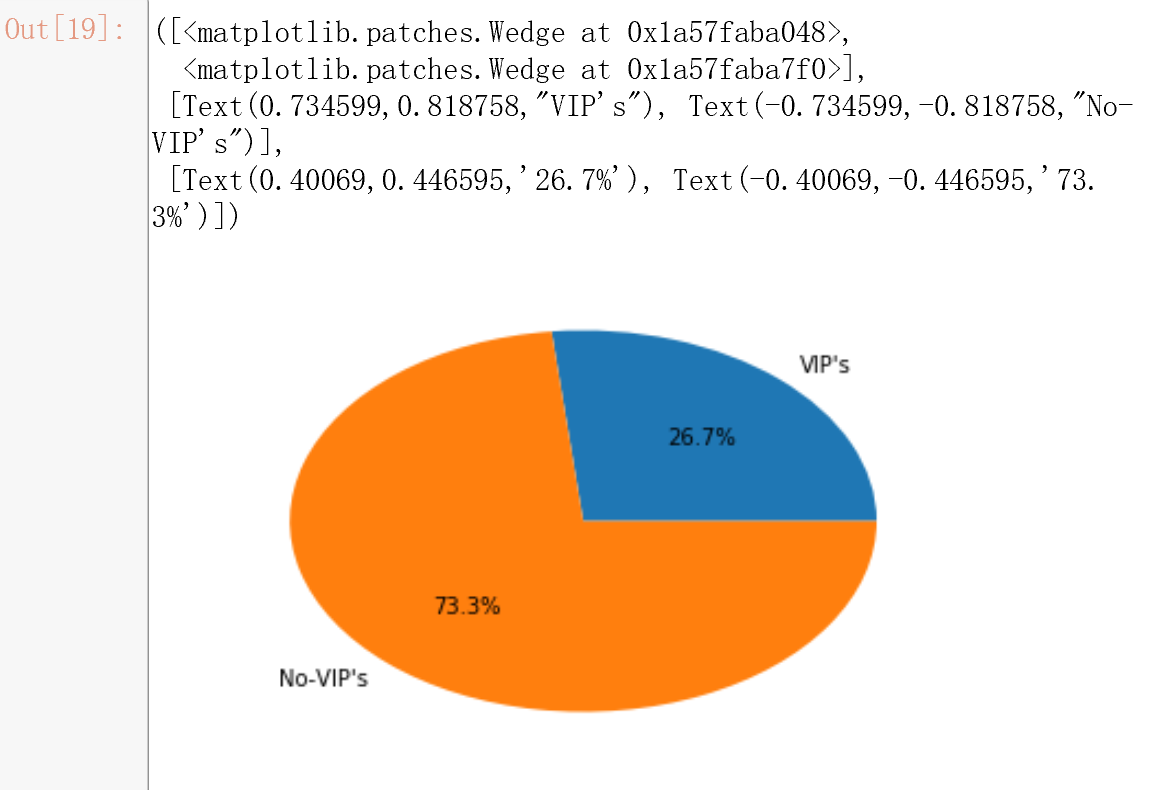
把单据号一列提取出来，使用head() 预览会员单据号数量



djh.describe() #了解会员单据号之和的数据情况



2、数据表二是所有的消费记录，对其单据号‘djh’计数可以得到所有单据号数量a，对主键连接后的数据表的单据号‘djh’进行计数，得到会员消费的所有单据号数量b，a-b为非会员消费的所有单据号数量，对这两个数量画出饼图。

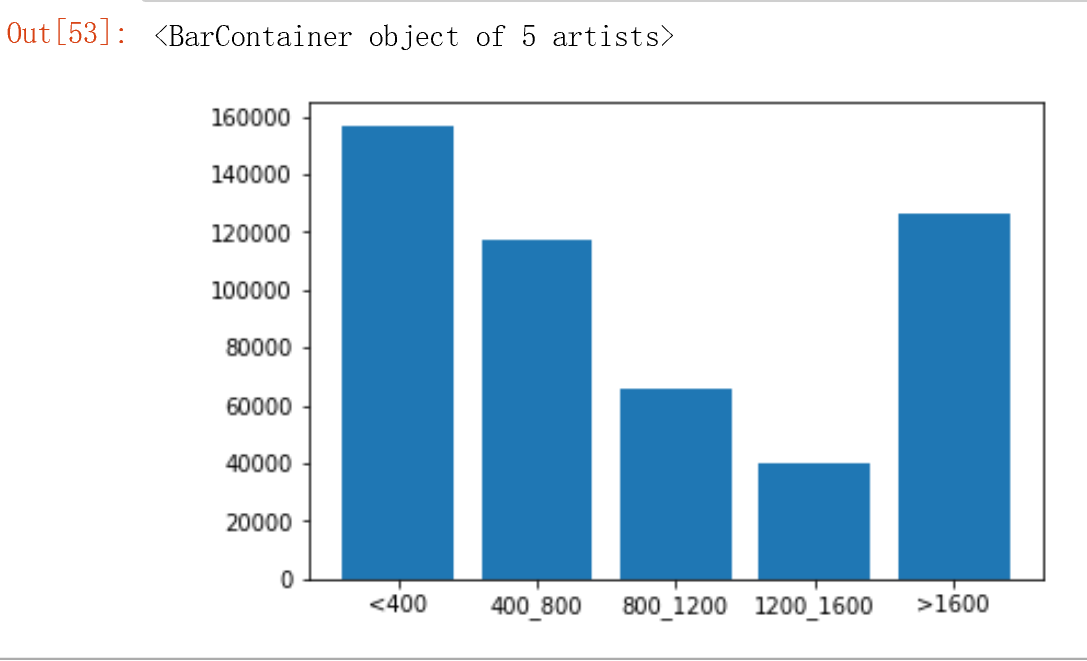


**第二题**

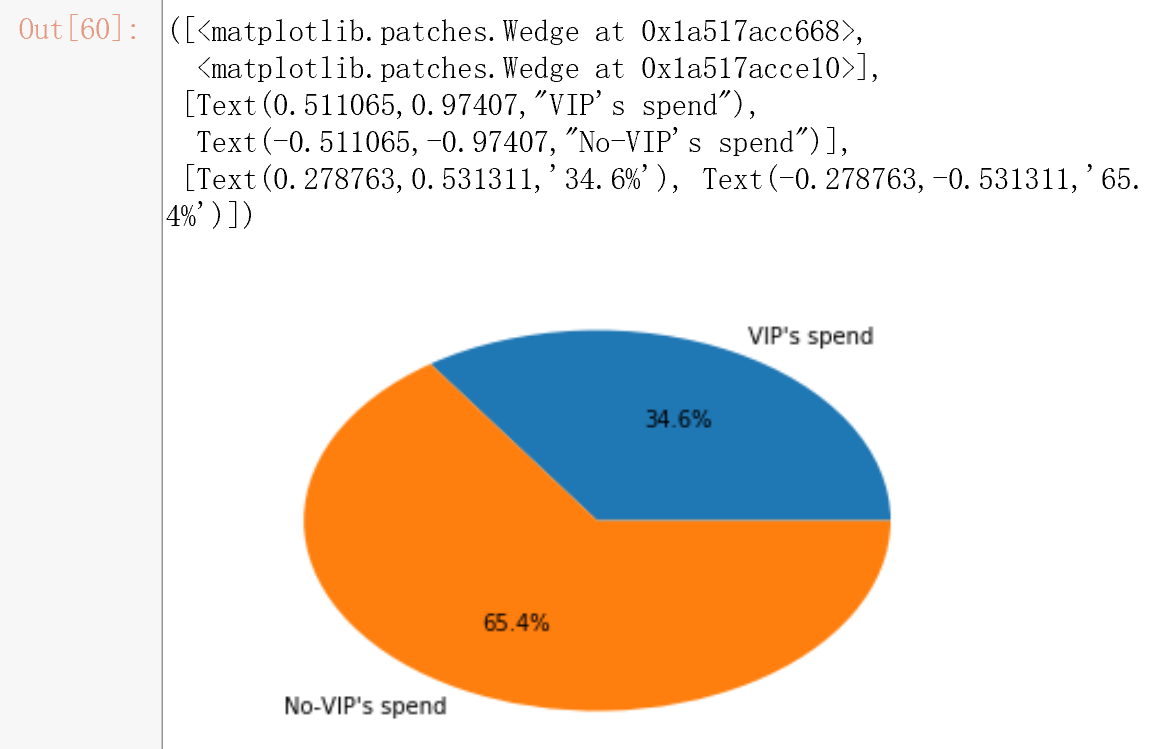
1、首先对会员金额‘je’一列进行简单分析，describe()，发现消费金额最小值出现了复数，因此对会员消费金额进行求绝对值，使之不出现负值

2、提取出求绝对值后的’je’一列，对会员消费金额进行分析，按照之前简单分析来看，设置五个区间段：小于400，400-800，800-1200，1200-1600，大于1600

3、创建空列表，使用循环语句和if-elif语句将不同会员金额用append放入各个列表，然后使用len()，求得各个列表长度，即各消费金额段的数量，按照这个数量画出柱状图：



4、对数据表二的 ’je’列加总，即求总的消费金额，再对主键连接的数据表求会员消费总金额，两者相减得到非会员的总消费金额，依次画出饼图：



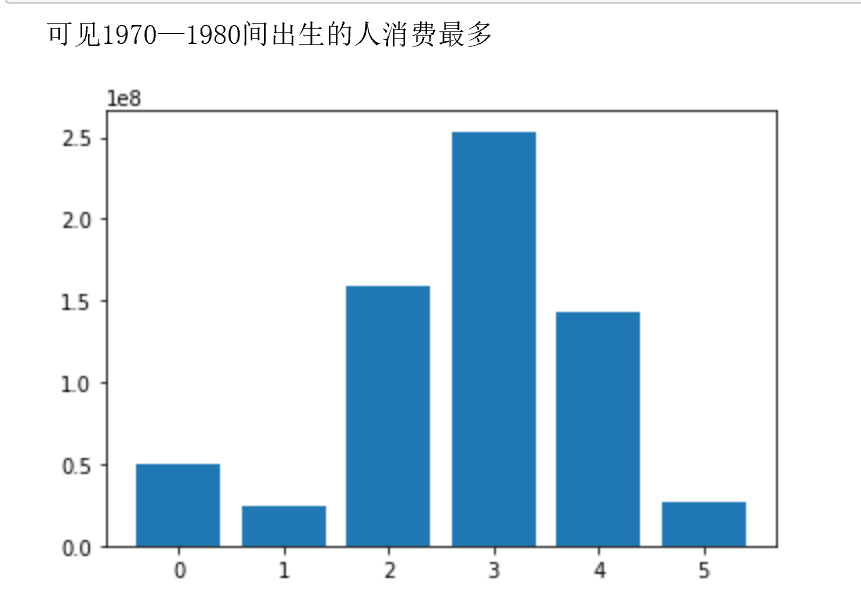
**第三题**

1、首先通过运用describe对主键连接后的数据表进行分析，可以发现某些出生时间‘csrq’出现明显错误，因此使用pd.to\_datetime(data['csrq'],errors='coerce') ，去除异常值的出生日期，并将其变成时间格式。

2、然后创建空列表，使用循环语句和csrq[i].year把会员的出生年份提取出来，并把出生年份’csnf’插入数据表中形成新的一列。

3、由于前置分析可以看出数据大多分布在中年，创建六个空列表，运用循环语句和if-eilf语句把’csnf’分别分成六个时间段：小于1920，1920-1960，1960-1970，1970-1980，1980-1990，大于1990，把属于不同年份段的索引序列放入六个空列表，

4、把数据表的金额列‘je’提取出来，按照上述年份段分好段的索引序列，再创建六个空列表，运用for循环语句对上述索引序列，按索引提取不同年份段的金额，把各年龄段的消费金额加总。从而得到不同年龄段的会员的总消费金额，以此画出条形图：

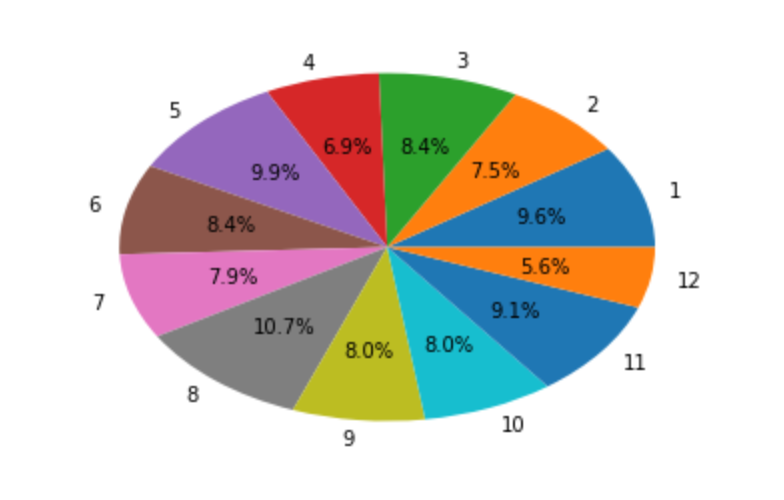
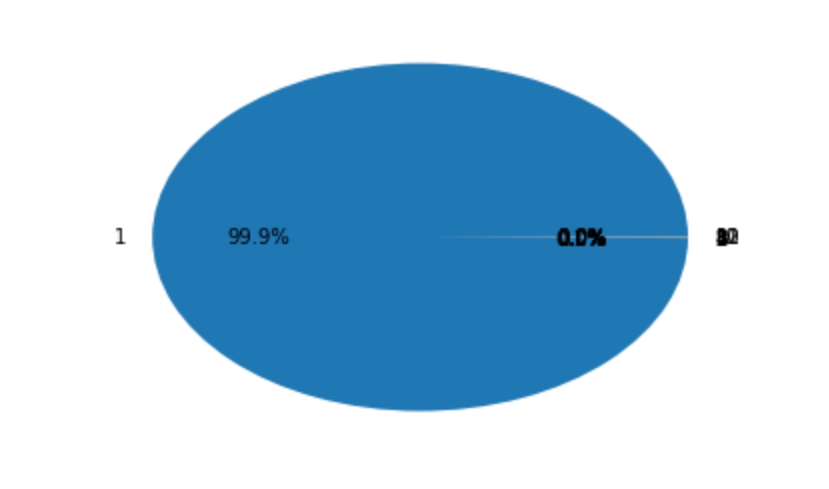


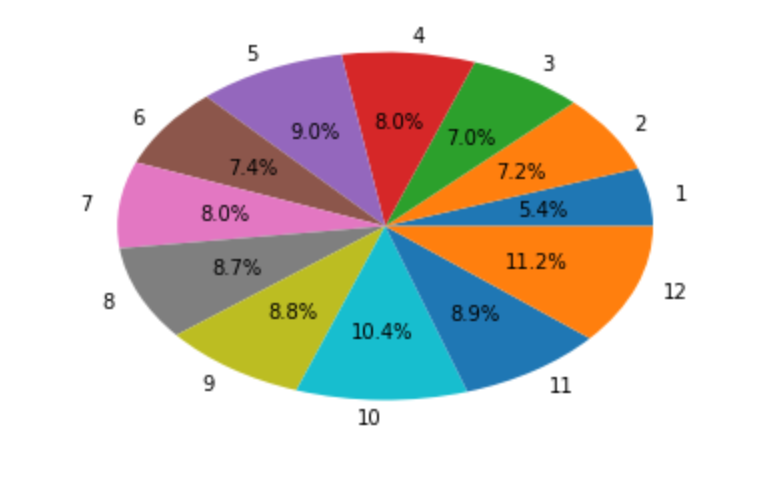
**第四题**

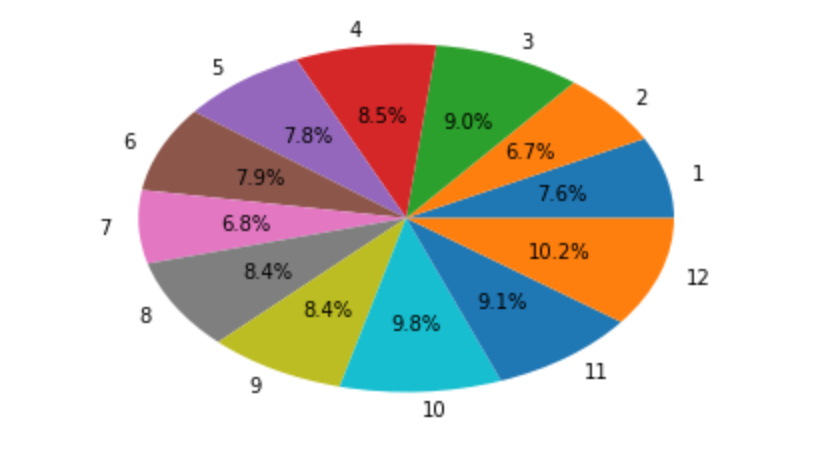
1、由上可知的用索引序列来求不同年龄段的总消费金额，首先求各个年份的索引序数，再次使用创建空列表和循环语句的方式，得到各个出生年份段的序数，并分别把他们放入六个列表中。

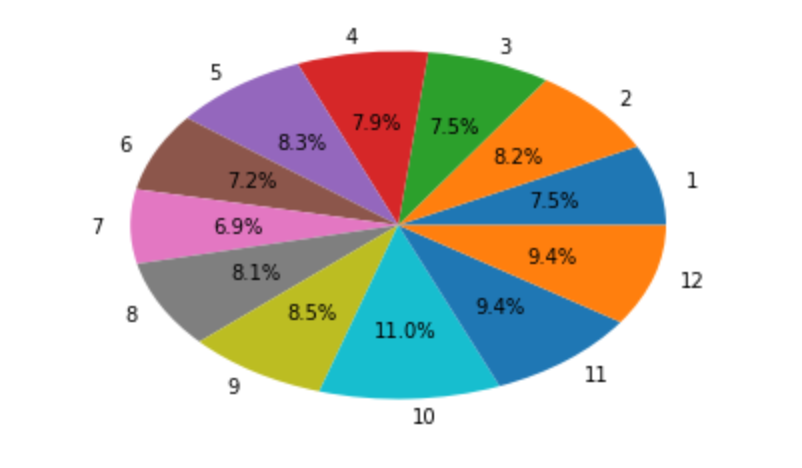
2、之后再对六个不同的年份段索引序数列表，分别创建12个空列表，使用for循环语句与if语句，把年份段里的12个月份的索引序数分别放入十二个空列表，即得到年份段中12个月的索引系数。

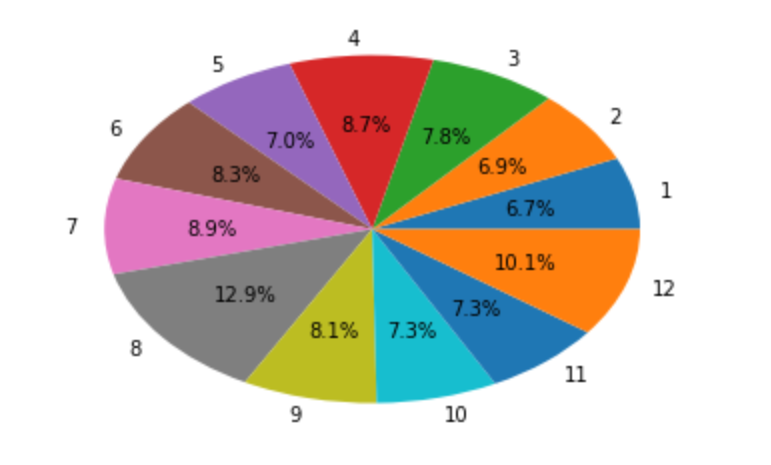
3、再按照索引系数对各个月份的消费金额进行加总，并依次画出饼图。











**任务2.3**

第一题根据会员的消费时间，将一天分为早上、中午、下午、晚上、凌晨五个时间段，统计每个时间段的会员人数，并对结果进行可视化

第二题根据会员消费时间，按年份统计每个季节的会员消费人数，对结果可视化

1. **完成思路：**

一、第一题首先获得会员消费信息表，提取消费时间列并将其转换为适当形式，按照时间将其分成不同时间段再统计人数

二、第二题提取消费时间列，将其先按照年份分成几个年份段，再根据不同年份段里会员的消费时间的月份信息，按照月份分为四个季节再分别统计人数

1. **基本步骤：**

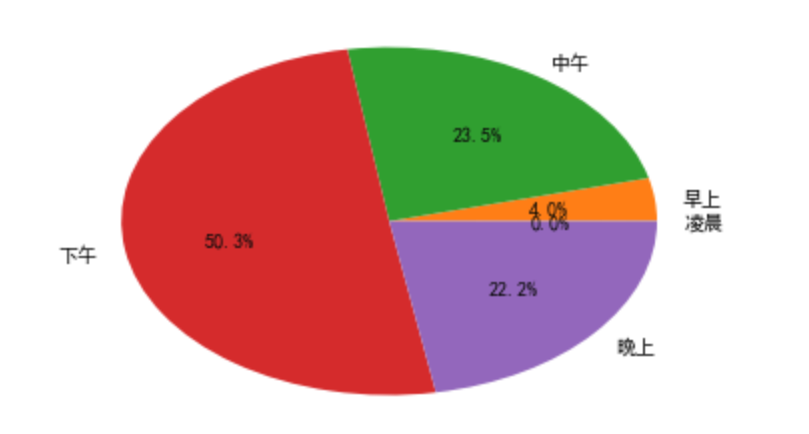
**第一题：**

1、首先将两个数据表按照卡号‘kh’进行主键连接，提取消费时间‘dtime’列，去除错误值，可以得出消费时间列中的元素并不是时间格式，使用for循环语句依次把各个消费时间转换为时间格式。

2、创建五个空列表，使用for循环与if-elif语句，以小时把时间分段，分别为凌晨：0-5，早晨：6-10，中午：11-13，下午：14-18，晚上：19-23，把各个索引序列分别放入相对应的五个空列表。

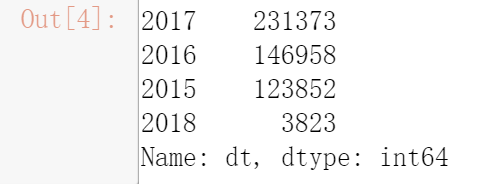
3、对五个列表测量长度，得到各个时间段的人数，并以此画出饼图。



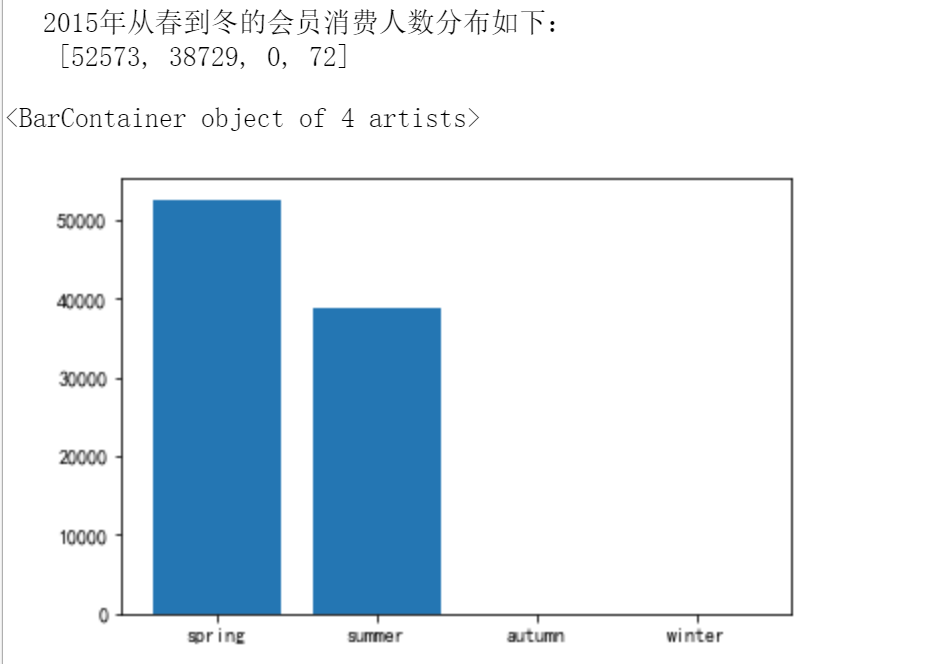


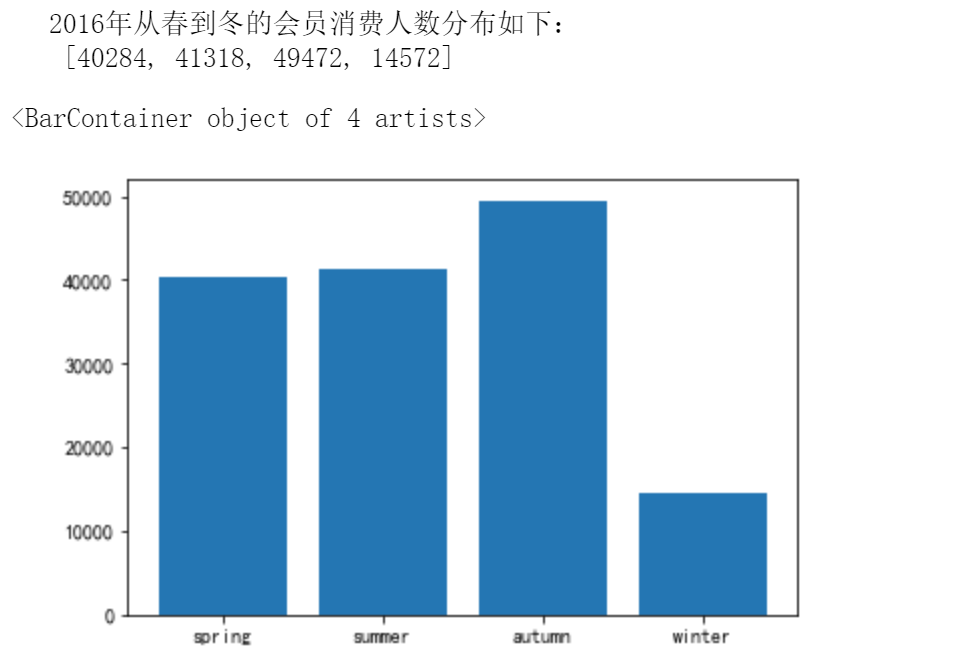
**第二题：**

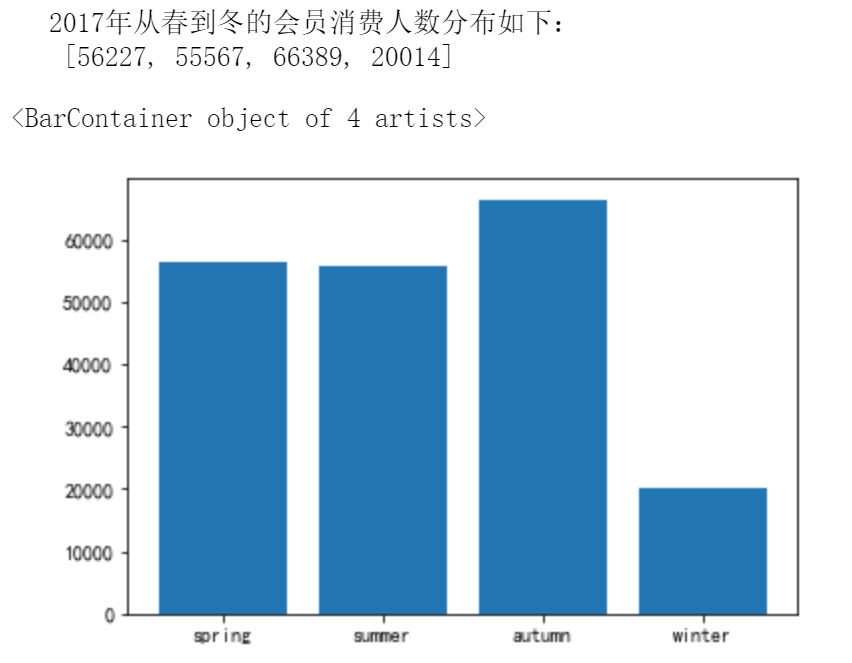
1. 按照年份统计，首先我们如任务2.2，将会员的消费时间列提取出来，再用for循环依次提取消费时间的年份，用value\_counts进行观测。结果发现消费时间只有2015-2018年。

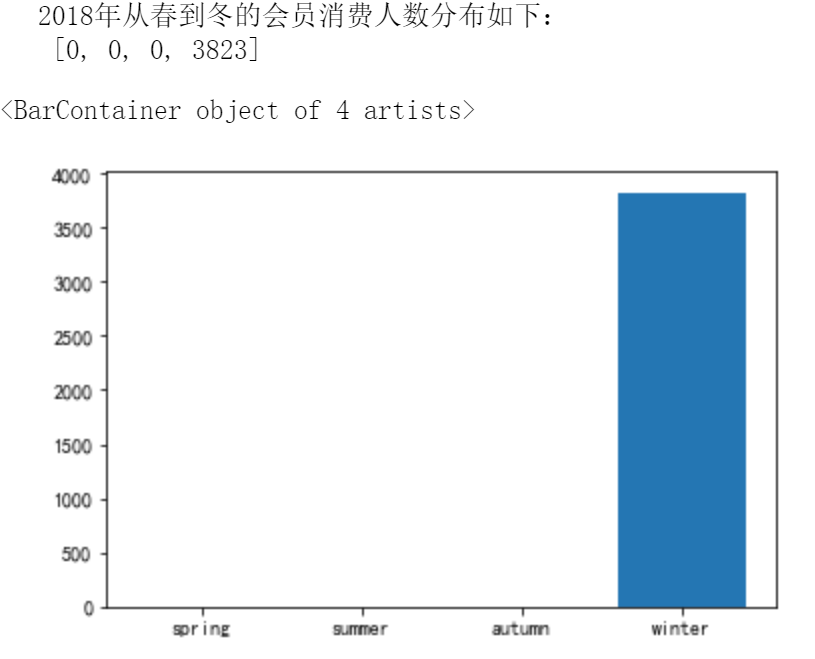


1. 设置4个空列表：a1,b1,c1,d1，代表2015-2018年对应索引序列，使用for循环和append将各个年份对应的索引序列装入列表。
2. 根据不同年份列表，再创建四个空列表，代表春夏秋冬四个季节，通过for循环和append将各个年份列表中的索引序列，分到四个空列表中。
3. len（）获取各个年份对应的四个列表的长度，即获得该年份对应的四个季节的消费人数，以此画出2015-2018年的可视化条形图。









**任务3.1**

确定会员的基本特征和构建用户标签，随机选取一位会员的数据，进行用户画像可视化

**一、完成思路：**

选取性别、年龄、出生日期为基本特征

选择购物季度分布，一天内的购物时间，购物上中下旬时间分布，购物次数，购买物品的次数，购买货柜构建用户标签，绘制可视化图，得到用户频率最高的选择，最后将其全部导入词云。

**二、基本步骤：**

1、首先将两初始数据表按照卡号‘kh’，对其进行主键合并。提取数据表中的卡号‘kh’列，并对其去重data['kh'].drop\_duplicates()

2、将去重后的卡号变为列表list(kh)，导入random模块，进行随机抽取random.choice(kh)

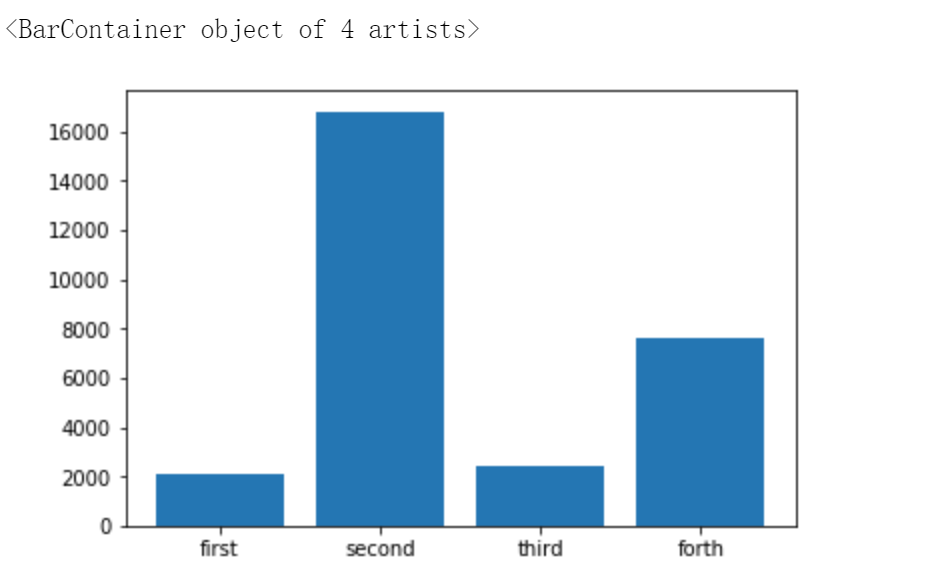
3、把随机卡号的所有行抽取出来data.loc[data['kh']=='f3bd3256',:]，从而获取各个基本信息

4、基本信息有出生年份，性别，年龄使用datetime模块得到当前的时间datetime.datetime.now()，将当前时间和出生日期的年份提取出来并相减，得到年龄，获取其基本特征。

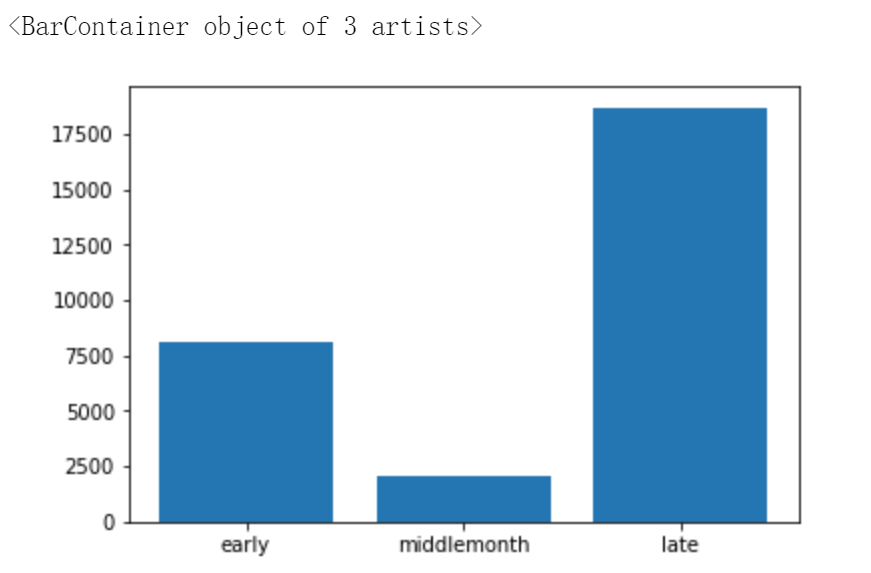
5、把选出的随机会员的销售日期提取出来，并转换成字符串类型，data.dtime.astype(str)，并采用切片的方式得到年份、月份，日期。

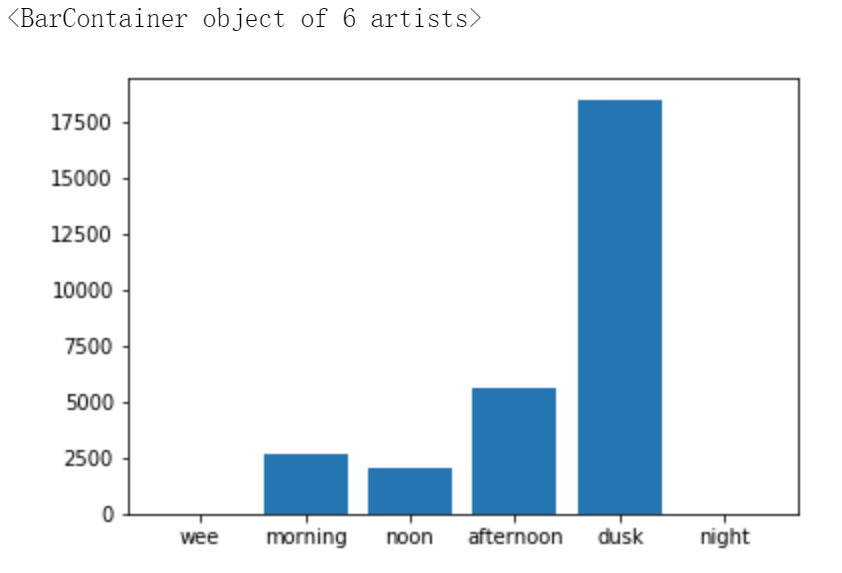
6、通过列表推导式，将字符串类型的月份依次转换为int类型，使用pd.cut将季度进行划分为：第一季度到第四季度，并通过groupby对季度分类聚类，并加总各季度的金额。rd.groupby(by=['季度划分'])['je'].sum()

以此画出条形图

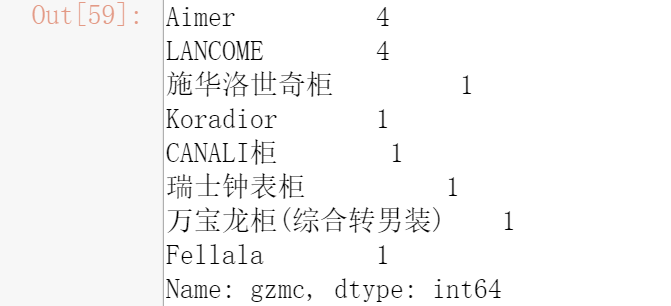


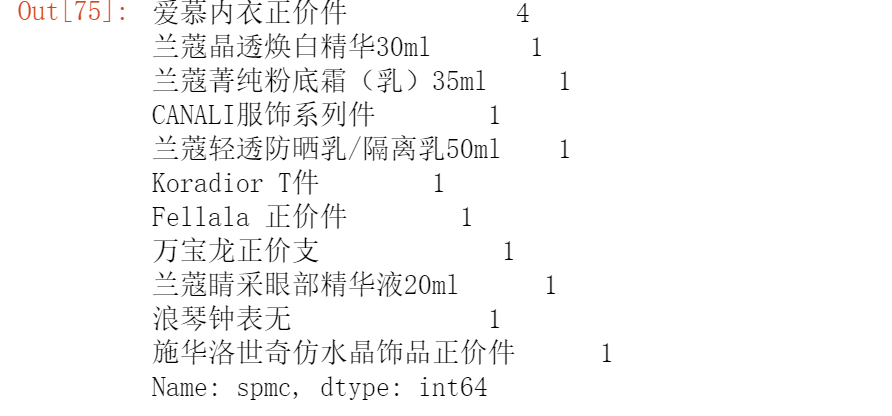
7、日期也通过使用pd.cut将日期划分成上旬、中旬、下旬，时间也通过pd.cut划分为凌晨、上午、中午、下午、傍晚、晚上，并通过groupby聚类加总。得到各个时间段的总消费金额，依次画出条形图





8、将数据表中的购买货柜与购买单件物品列提取出来，对货柜和购买单件进行计数，并观测选择结果，该用户最喜欢购买的货柜与单件





1. 导入词云模块，构建词云对象，调用词云对象的generate方法，将所需要文本输入，得到词云图



**任务4.1**

第一题：构建会员价值分析特征指标

第二题：使用K-means聚类算法对构建的会员价值分析特征进行聚类，并根据聚类中心绘制雷达图

第三题：针对不同群体制定相应的营销方案

**一、完成思路：**

第一题：选取各个会员的累计消费金额，累计消费数量，累计消费次数

第二题：选择购物季度分布，一天内的购物时间，购物上中下旬时间分布，购物次数，购买物品的次数构建用户标签

第三题：根据第二题的结果进行理性分析

**二、基本步骤：**

1、首先对两个原始数据表读取，对卡号‘kh’进行主键合并，再对数据进行去重，data=data.drop\_duplicates()

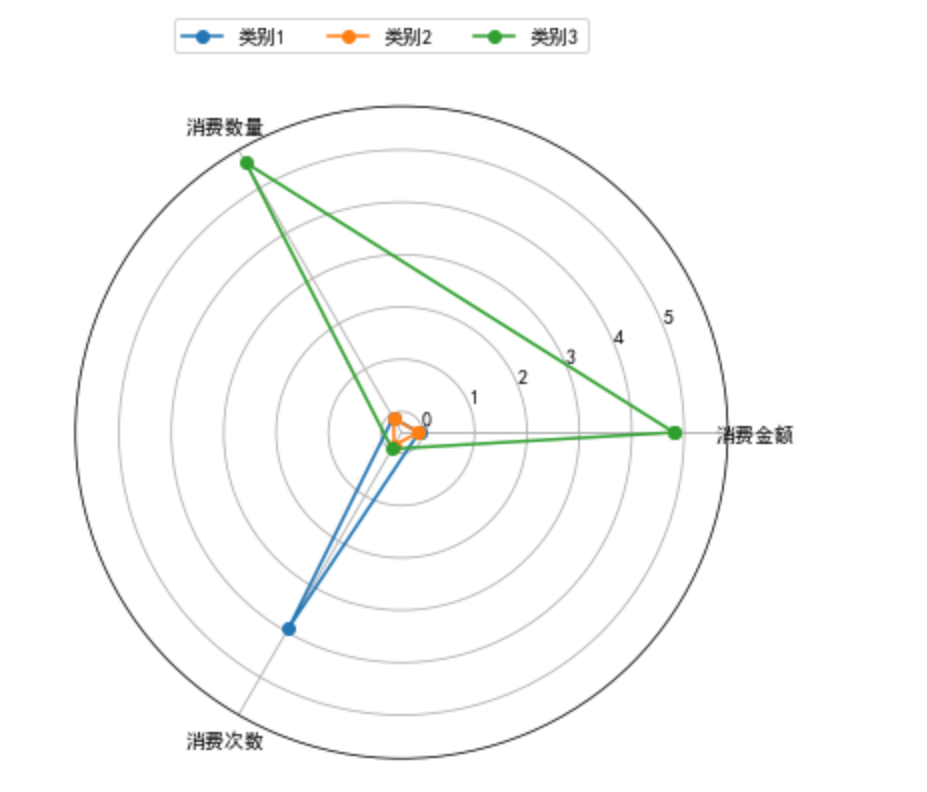
2、提取‘kh’列，并得出卡号出现的次数cs=kh.value\_counts() ，并转换为列表类型，即得到各个卡号对应的累计消费次数。

3、hy=data.groupby('kh').sum() 对每个卡号的各个类别加总，其中‘sl’和‘je’分别提取出来，并转换为列表类型，即得到累计消费数量，累计消费金额

4、使用dataframe将‘sl’、‘je’、‘cs’转换为数据表形式，并将其转置，对各列命名。

5、导入数据标准化模块、K-means聚类模块，将数据标准化，并构建聚类模块和训练聚类模块，K值为3

6、对各个会员的累计消费金额，累计消费商品数量，累计消费次数构建雷达图



**给出的建议：**

**类别一**：

累计消费次数多，但是消费数量和消费金额很少，可认为是高频的小量且小额消费客户，例如：经常只在超市买点文具这些小东西的小学生

**类别二：**

消费数量、消费金额、消费次数都很少，可认为是不活跃的潜在客户，例如：办了卡之后懒得去超市消费的中年大叔。

**类别三**：

消费数量和消费金额大，但是消费次数少。可认为是低频的大量且高消费客户。例如：一年回来一两次的土豪海归。

针对类别一：在广告投放上，着重推出便宜的小物品的广告，吸引其继续高频购买。

针对类别二：着重投放打折活动，满减活动，抽奖活动等等广告，调动此类客户的购物心。

针对类别三：可以综合投放广告，推出高品质高价格的商品，吸引客户单次购买中消费更多金额。