**广 州 商 学 院**

**实验报告（第 1 次）**

实验名称 客户价值分析 实验时间 2022/3/23

同组同学 小组分工

**一、实验目的**

（1）熟练掌握Python常用数据结构

（2）熟练掌握数据分析相关库的运用

（3）熟练应用RFM模型，进行聚类分析

（4）熟练应用K-means聚类算法

（5）能用Python实现客户价值分析

**二、实验仪器设备或材料**

1.学生PC机 57 台

2.教师机 1 台

3.教师机安装有极域教师机端、学生PC机安装有极域学生端

4.教学机和学生机需要安装第三方模块 ：pandas、numpy、matplotlib、sklearn。

**三、实验原理**

1、教学机与学生机需要安装PyCharm和Anaconda3、数据分析常用的第三方库。

2、学生需要已经学习Python程序设计语言，熟悉Python的基本语法和结构。

3、学生在自己的电脑上要安装第三方模块 ：pandas、numpy、matplotlib、sklearn。

**四、实验内容与步骤**

**1、项目文件结构**

****

**2、源代码**

**data\_clean.py**

import pandas as pd

import numpy as np

aa =r'./data/TB201812.xls'

resultfile=r'./data/data.xls'

df = pd.DataFrame(pd.read\_excel(aa))

df1=df[['订单付款时间','买家会员名','买家实际支付金额','数据采集时间']]

#去除空值，订单付款时间非空值才保留

#去除买家实际支付金额为0的记录

df1=df1[df1['订单付款时间'].notnull() & df1['买家实际支付金额'] !=0]

#R值：R表示客户最近活跃时间与数据采集点时间距离，时间相减变为天：

#R代表客户最近的活跃时间距离数据采集点的时间距离，

#R越大，表示客户越久未发生交易，R越小，表示客户越近有交易发生。

#R越大则客户越可能会“沉睡”，流失的可能性越大。在这部分客户中，可能有些优质客户，值得通过一定的营销手段进行激活。

#F值：F代表客户过去某段时间内的活跃频率。F越大，则表示客户同本公司的交易越频繁，是非常忠诚的客户；

#F越小，则表示客户不够活跃，且可能是竞争对手的常客。针对F较小、且消费额较大的客户，需要推出一定的竞争策略，将这批客户从竞争对手中争取过来。

#M值：M表示客户每次消费金额的多少，可以用最近一次消费金额，也可以用过去的平均消费金额，根据分析的目的不同，可以有不同的标识方法。

#一般来讲，单次交易金额较大的客户，支付能力强，价格敏感度低，是较为优质的客户，而每次交易金额很小的客户，可能在支付能力和支付意愿上相对较低。

#R:最近消费时间距数据采集时间的间隔

#利用to\_datetime转换为时间格式：'yyyy-MM-dd HH:mm:ss'

df1['R'] = (pd.to\_datetime(df1['数据采集时间']) - pd.to\_datetime(df1['订单付款时间'])).values/np.timedelta64(1, 'D')

df1=df1[['订单付款时间','买家会员名','买家实际支付金额','R']]

df2=df1.groupby('买家会员名').agg({'R': 'min','买家实际支付金额':'sum'})

#F:消费频次，客户一定时间内的购买次数。

df2['F']=df1.groupby(["买家会员名"])['买家会员名'].size()

df2.to\_excel(resultfile) #导出结果

**data\_transform.py**

import pandas as pd

#标准化处理

datafile = r'./data/data.xls' #需要进行标准化的数据文件；

transformfile = r'./data/transformdata.xls' #标准化后的数据存储路径文件；

data = pd.read\_excel(datafile)

data=data[['R','F','买家实际支付金额']]

data = (data - data.mean(axis = 0))/(data.std(axis = 0)) #简洁的语句实现了标准化变换，类似地可以实现任何想要的变换。

data.columns=['R','F','M'] #表头重命名。

data.to\_excel(transformfile, index = False) #数据写入

**data\_kmeans.py**

#-\*- coding: utf-8 -\*-

import pandas as pd

import numpy as np

from pandas import to\_datetime

#引入sklearn框架，导入K均值聚类算法

from sklearn.cluster import KMeans

import matplotlib.pyplot as plt

#from sklearn.manifold import

inputfile = r'./data/transformdata.xls' #待聚类的数据文件

outputfile=r'./data/data\_type.xls'

#读取数据并进行聚类分析

data = pd.read\_excel(inputfile) #读取数据

#利用K-Means聚类算法对客户数据进行客户分群，聚成4类

k = 4 #需要进行的聚类类别数

iteration=500

kmodel = KMeans(n\_clusters = k,max\_iter=iteration)

kmodel.fit(data) #训练模型

r1=pd.Series(kmodel.labels\_).value\_counts()

r2=pd.DataFrame(kmodel.cluster\_centers\_)

r=pd.concat([r2,r1],axis=1)

r.columns=list(data.columns)+[u'聚类数量']

r3 = pd.Series(kmodel.labels\_,index=data.index)

r = pd.concat([data,r3], axis=1)

r.columns = list(data.columns)+[u'聚类类别']

r.to\_excel(outputfile)

kmodel.cluster\_centers\_

kmodel.labels\_

plt.rcParams['font.sans-serif']=['SimHei']

plt.rcParams['axes.unicode\_minus']=False

for i in range(k):

cls=data[r[u'聚类类别']==i]

cls.plot(kind='kde',linewidth=2,subplots=True,sharex=False)

plt.suptitle('客户群=%d;聚类数量=%d' %(i,r1[i]))

plt.legend()

plt.show()

**data\_view.py**

#-\*- coding: utf-8 -\*-

#对数据进行基本的探索

#返回缺失值个数以及最大最小值

import pandas as pd

datafile= r'./data/TB201812.xls' #原始数据,第一行为属性标签

resultfile = r'./data/view.xls' #数据探索结果表

data = pd.read\_excel(datafile, encoding = 'utf-8') #读取原始数据，指定UTF-8编码（需要用文本编辑器将数据装换为UTF-8编码）

data=data[['订单付款时间','买家会员名','买家实际支付金额','数据采集时间']]

view = data.describe(percentiles = [], include = 'all').T #包括对数据的基本描述，percentiles参数是指定计算多少的分位数表（如1/4分位数、中位数等）；T是转置，转置后更方便查阅

view['null'] = len(data)-view['count'] #describe()函数自动计算非空值数，需要手动计算空值数

view = view[['null', 'max', 'min']]

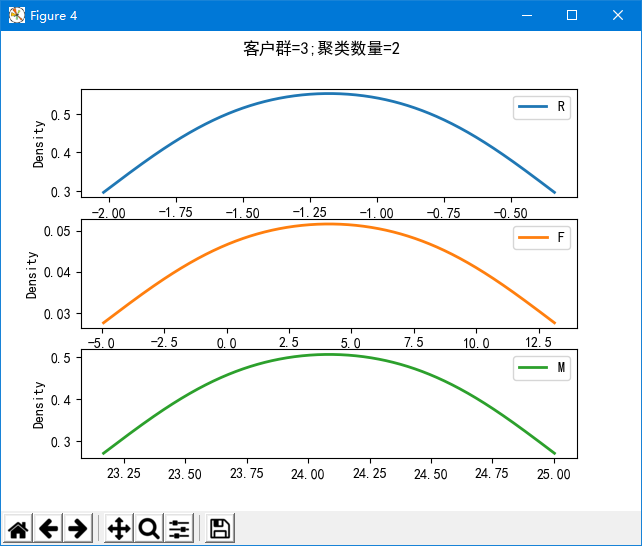
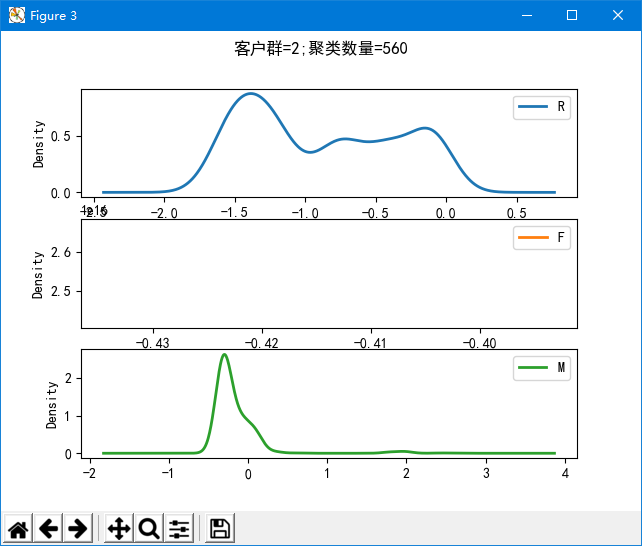
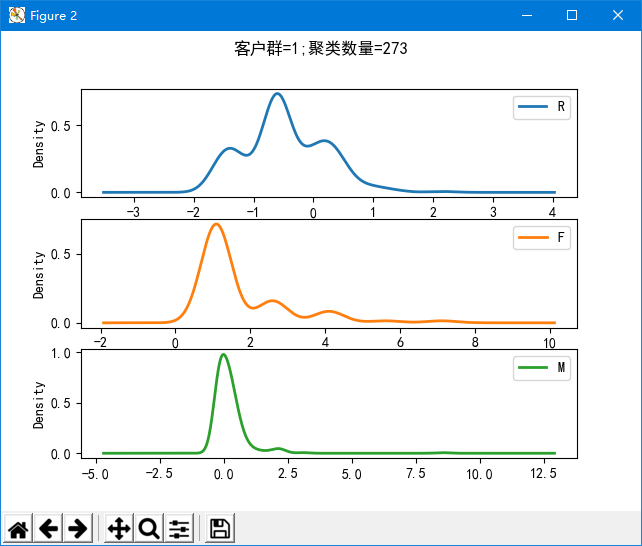
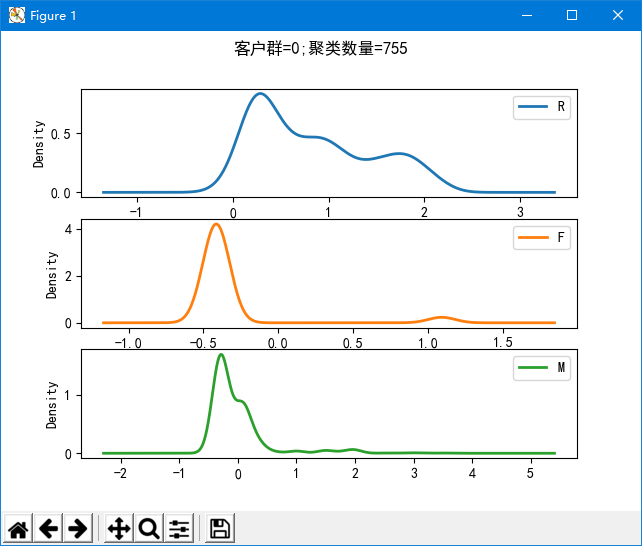
view.columns = [u'空值数', u'最大值', u'最小值'] #表头重命名

'''这里只选取部分探索结果。

describe()函数自动计算的字段有count（非空值数）、unique（唯一值数）、top（频数最高者）、freq（最高频数）、mean（平均值）、std（方差）、min（最小值）、50%（中位数）、max（最大值）'''

view.to\_excel(resultfile) #导出结果

**3、系统预览**



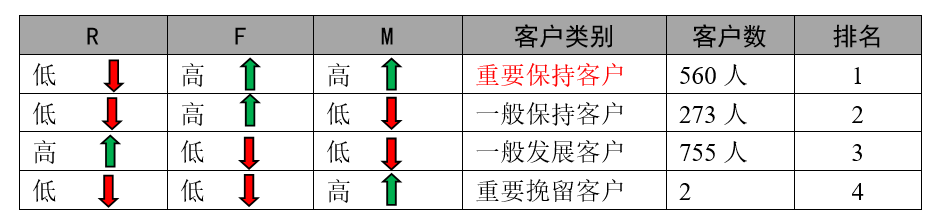
客户群0

客户群1

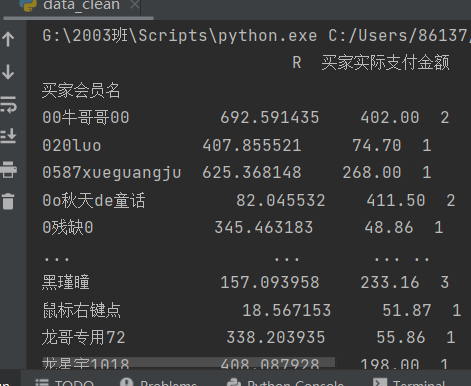
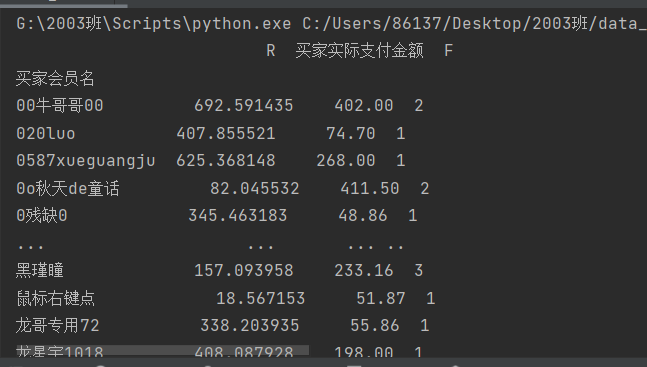
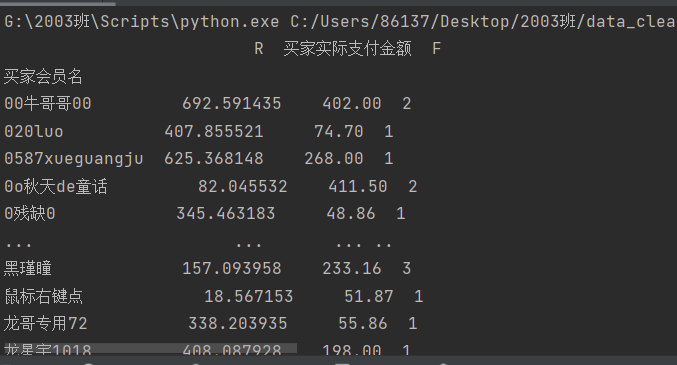
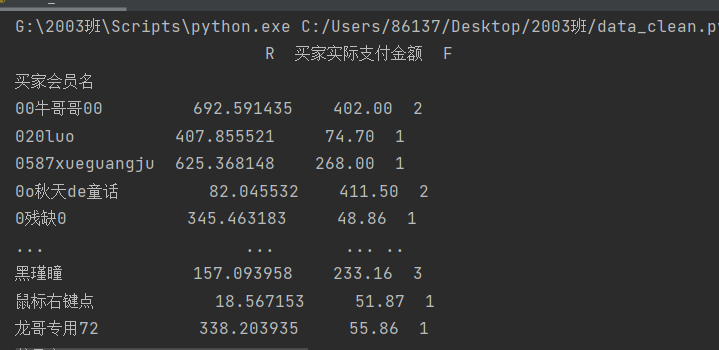
客户群2

客户群3

**4、客户价值分析结果解析**



**五、实验结果与分析**



**六、结论与体会**

结论与体会：本次实验学习了Python里数据分析库的运用和聚类分析，实验比较难理解，需要多练习。

**七、教师评语**