import pandas as pd

import numpy as np

import os

from pandas import DataFrame,Series

from sklearn.cluster import KMeans

from sklearn.cluster import Birch

os.chdir(r'G:\数据')

data= pd.read\_csv('task1\_X2.csv', engine = 'python', encoding = 'gbk')# 数据读取

index1=['第一食堂','第二食堂','第三食堂','第四食堂','第五食堂']

data=data.loc[data['Dept'].isin(index)]

group=data[['CardNo','Major','Sex','Money']].groupby(by=['CardNo','Major','Sex'])

ddd=group.count()

order=pd.DataFrame(ddd)

order.columns=['刷卡次数']

su=order.reset\_index()

ttt=group.sum()

sss=pd.DataFrame(ttt)

sss.columns=['本月消费总金额']

sd=sss.reset\_index()

Dept\_mapping = {

'第一食堂': 1,

'第二食堂': 2,

'第三食堂': 3,

'第四食堂':4,

'第五食堂':5}

data['Dept'] = data['Dept'].map(Dept\_mapping)

group1=data[['CardNo','Major','Sex','Dept']].groupby(by=['CardNo','Major','Sex'])

ggg=group1.mean()

order1=pd.DataFrame(ggg)

order1.columns=['最常去的饭堂']

mean=order1.reset\_index()

print(mean.head())

order\_merge1 = pd.merge(su, sd, left\_on=['CardNo','Major','Sex','刷卡次数'], right\_on =['CardNo','Major','Sex','刷卡次数'],how='inner')#数据合并

order\_merge2 = pd.merge(order\_merge1, mean, left\_on=['CardNo','Major','Sex'], right\_on =['CardNo','Major','Sex'],how='inner')#数据合并

print(order\_merge2)

print(order\_merge2.isnull().any())

from sklearn.preprocessing import StandardScaler #导入均值-方差规范化模块

X=order\_merge2.loc[:,['本月消费总金额','刷卡次数','最常去的饭堂']]

scaler = StandardScaler()

scaler.fit(X)

data\_prp1=scaler.transform(X)

from sklearn.cluster import KMeans

kmeans\_model = KMeans(n\_clusters= 4, max\_iter = 100)

kmeans\_model.fit(data\_prp1)

fit\_label= kmeans\_model.labels\_

center=kmeans\_model.cluster\_centers\_

Fs=pd.Series(fit\_label,index=order\_merge2['CardNo'])

print(Fs)

Fs = {'CardNo':Fs.index,'聚类结果':Fs.values}

fs= pd.DataFrame(Fs)

print(fs.head())

out= pd.merge(order\_merge2,fs, left\_on=['CardNo'], right\_on =['CardNo'],how='inner')#数据合并

out1=pd.DataFrame(out)

print(out1.head())

index2=['0']

out2=out1.loc[out1['聚类结果'].isin(index2)]

print(out2[['最常去的饭堂','刷卡次数','本月消费总金额']].mean())

from collections import Counter

num1=Counter(out2['Sex'])

print(num1)

index3=['1']

out3=out1.loc[out1['聚类结果'].isin(index3)]

print(out3[['最常去的饭堂','刷卡次数','本月消费总金额']].mean())

from collections import Counter

num2=Counter(out3['Sex'])

print(num2)

index4=['2']

out4=out1.loc[out1['聚类结果'].isin(index4)]

print(out4[['最常去的饭堂','刷卡次数','本月消费总金额']].mean())

from collections import Counter

num3=Counter(out4['Sex'])

print(num3)

index5=['3']

out5=out1.loc[out1['聚类结果'].isin(index5)]

print(out5[['最常去的饭堂','刷卡次数','本月消费总金额']].mean())

from collections import Counter

num4=Counter(out5['Sex'])

print(num4)