import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

%matplotlib inline

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

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

plt.style.use('ggplot')

#导入画图程序包

df=pd.read\_csv('data.csv')#导入附件一数据

df.head()

df.info()#查看数据类型

df['支付时间']=pd.to\_datetime(df.支付时间,format="%Y-%m-%d %H:%M:%S")#将时间转化为时间格式

df['month']=df.支付时间.values.astype('datetime64[M]')#提取支付时间的月份

df.info()

df.head()

df.describe()#对各数据进行描述性分析

df.isnull().sum() ##检查是否有缺失值

df[df.应付金额==df.实际金额]

ind\_A=(data4['地点']=='A')#任务1.1分别提取A、B、C、D、E售货机数据并保存成csv文件

data\_A=data4.loc[ind\_A,]

data\_A.to\_csv('F://pywork/task1\_1A.csv',encoding='gbk')

ind\_B=(data4['地点']=='B')

data\_B=data4.loc[ind\_B,]

data\_B.to\_csv('F://pywork/task1\_1B.csv',encoding='gbk')

ind\_C=(data4['地点']=='C')

data\_C=data4.loc[ind\_C,]

data\_C.to\_csv('F://pywork/task1\_1C.csv',encoding='gbk')

ind\_D=(data4['地点']=='D')

data\_D=data4.loc[ind\_D,]

data\_D.to\_csv('F://pywork/task1\_1D.csv',encoding='gbk')

ind\_E=(data4['地点']=='E')

data\_E=data4.loc[ind\_E,]

data\_E.to\_csv('F://pywork/task1\_1E.csv',encoding='gbk') data\_A.head()

data\_A.info()

data\_A['支付时间']=pd.to\_datetime(data\_A.支付时间,format="%Y-%m-%d %H:%M:%S")#将时间转化为时间格式

data\_A['month']=data\_A.支付时间.values.astype('datetime64[M]')#提取支付时间的月份

data\_A['day']=data\_A.支付时间.values.astype('datetime64[D]')#提取支付时间的日

data\_B['支付时间']=pd.to\_datetime(data\_B.支付时间,format="%Y-%m-%d %H:%M:%S")

data\_B['month']=data\_B.支付时间.values.astype('datetime64[M]')

data\_B['day']=data\_B.支付时间.values.astype('datetime64[D]')

data\_C['支付时间']=pd.to\_datetime(data\_C.支付时间,format="%Y-%m-%d %H:%M:%S")

data\_C['month']=data\_C.支付时间.values.astype('datetime64[M]')

data\_C['day']=data\_C.支付时间.values.astype('datetime64[D]')

data\_D['支付时间']=pd.to\_datetime(data\_D.支付时间,format="%Y-%m-%d %H:%M:%S")

data\_D['month']=data\_D.支付时间.values.astype('datetime64[M]')

data\_D['day']=data\_D.支付时间.values.astype('datetime64[D]')

data\_E['支付时间']=pd.to\_datetime(data\_E.支付时间,format="%Y-%m-%d %H:%M:%S")

data\_E['month']=data\_E.支付时间.values.astype('datetime64[M]')

data\_E['day']=data\_E.支付时间.values.astype('datetime64[D]')

data\_B.info()#再次检查数据类型

data\_A.head()

data\_A.pivot\_table(index='month',aggfunc={'实际金额':'sum','订单号':'count'}).rename(columns={'实际金额':'交易额','订单号':'订单量'})#创建数据透视表，计算A售货机每月的交易额和订单量

data\_B.pivot\_table(index='month',aggfunc={'实际金额':'sum','订单号':'count'}).rename(columns={'实际金额':'交易额','订单号':'订单量'})#创建数据透视表，计算B售货机每月的交易额和订单量

data\_C.pivot\_table(index='month',aggfunc={'实际金额':'sum','订单号':'count'}).rename(columns={'实际金额':'交易额','订单号':'订单量'})#创建数据透视表，计算C售货机每月的交易额和订单量

data\_C.pivot\_table(index='month',aggfunc={'实际金额':'sum','订单号':'count'}).rename(columns={'实际金额':'交易额','订单号':'订单量'})#创建数据透视表，计算C售货机每月的交易额和订单量

data\_C.pivot\_table(index='month',aggfunc={'实际金额':'sum','订单号':'count'}).rename(columns={'实际金额':'交易额','订单号':'订单量'})#创建数据透视表，计算C售货机每月的交易额和订单量

df.订单号.count()

df.实际金额.sum()

data\_A.pivot\_table(index='month',aggfunc={'实际金额':'mean'}).rename(columns={'实际金额':'每单平均交易额'})

data\_A\_pivot\_table.pivot\_table(index='month',aggfunc={'订单号':'mean'}).rename(columns={'订单号':'日均订单量'})

data\_B.pivot\_table(index='month',aggfunc={'实际金额':'mean'}).rename(columns={'实际金额':'每单平均交易额'})

data\_C.pivot\_table(index='month',aggfunc={'实际金额':'mean'}).rename(columns={'实际金额':'每单平均交易额'})

data\_D.pivot\_table(index='month',aggfunc={'实际金额':'mean'}).rename(columns={'实际金额':'每单平均交易额'})

data\_B\_pivot\_table=data\_B.pivot\_table(index=['month','day'],aggfunc={'订单号':'count'})

data\_B\_pivot\_table.pivot\_table(index='month',aggfunc={'订单号':'mean'}).rename(columns={'订单号':'日均订单量'})

data\_C\_pivot\_table=data\_C.pivot\_table(index=['month','day'],aggfunc={'订单号':'count'})

data\_C\_pivot\_table.pivot\_table(index='month',aggfunc={'订单号':'mean'}).rename(columns={'订单号':'日均订单量'})

data\_D\_pivot\_table=data\_D.pivot\_table(index=['month','day'],aggfunc={'订单号':'count'})

data\_D\_pivot\_table.pivot\_table(index='month',aggfunc={'订单号':'mean'}).rename(columns={'订单号':'日均订单量'})

data\_E\_pivot\_table=data\_E.pivot\_table(index=['month','day'],aggfunc={'订单号':'count'})

data\_E\_pivot\_table.pivot\_table(index='month',aggfunc={'订单号':'mean'}).rename(columns={'订单号':'日均订单量'})