**广 州 商 学 院**

**实验报告（第 2 次）**

实验名称 销售收入分析与预测 实验时间 2022/4/10

同组同学 小组分工

1. **实验目的**

（1）明确做数据预处理的作用

（2）熟练掌握对数据重复值、缺失值、异常值的检测和处理方法

（3）能够对实际案例数据进行数据清洗工作

（4）熟练掌握线性回归与最小二乘法

1. **实验仪器设备或材料**

1、教学机与学生机需要安装第三方模块pandas、numpy、matplotlib、sklearn

2、教学机与学生机需要安装PyCharm和Anaconda3

1. **实验原理**

1.学生PC机 57 台

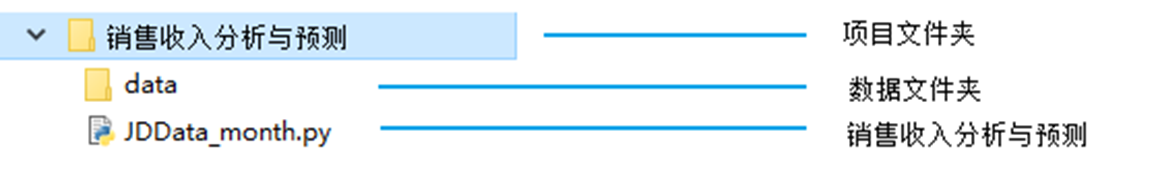
2.教师机 1 台

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 PyCharm和Anaconda3。

**四、实验内容与步骤**

**1、项目文件结构**

****

**2、源代码**

**JDData\_month.py**

import pandas as pd

import numpy as np

from sklearn import linear\_model

import matplotlib.pyplot as plt

aa =r'.\data\JDdata.xls'

bb=r'.\data\JDcar.xls'

resultfile1=r'result1.xls'

resultfile2=r'result2.xls'

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

dfbb=pd.DataFrame(pd.read\_excel(bb))

df1=dfaa[['业务日期','金额']]

df2=dfbb[['投放日期','支出']]

#去除空日期和金额为0的记录

df1=df1[df1['业务日期'].notnull() & df1['金额'] !=0]

df2=df2[df2['投放日期'].notnull() & df2['支出'] !=0]

df1['业务日期'] = pd.to\_datetime(df1['业务日期'])

df2['投放日期'] = pd.to\_datetime(df2['投放日期'])

dfData = df1.set\_index('业务日期',drop=True)

dfCar=df2.set\_index('投放日期',drop=True)

# 按月度统计并显示销售金额

dfData\_month=dfData.resample('M').sum().to\_period('M')

# 按月度统计并显示广告费支出金额

dfCar\_month=dfCar.resample('M').sum().to\_period('M')

dfData\_month.to\_excel(resultfile1) #导出结果

dfCar\_month.to\_excel(resultfile2) #导出结果

clf=linear\_model.LinearRegression(fit\_intercept=True,normalize=False)

#x为广告费用，y为销售收入

x=pd.DataFrame(dfCar\_month['支出'])

y=pd.DataFrame(dfData\_month['金额'])

# 图表字体为华文细黑，字号为10

plt.rc('font', family='SimHei', size=10)

plt.figure("销售收入分析")

plt.scatter(x, y, color='red') #真实值散点图

#plt.show()

#绘制拟合图

clf.fit(x,y) #拟合线性模型

k=clf.coef\_ #获取回归系数（斜率w1,w2,w3,...,wn）

b=clf.intercept\_ #获取截距w0

#7月预计投入60000元广告费（x0）

x0=60000

#预测7月销售收入（y0），y0=截距+X值\*斜率

y0=b+x0\*k

print(y0)

#使用线性模型进行预测y值

y\_pred =clf.predict(x)

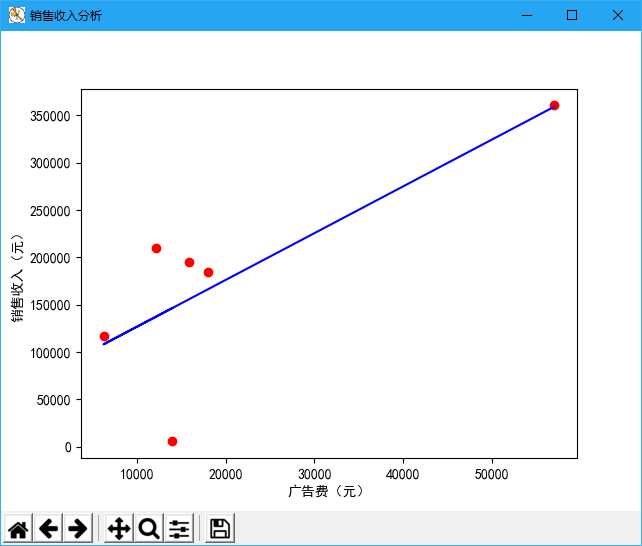
plt.plot(x,y\_pred, color='blue', linewidth=1.5) #预测回归线

plt.ylabel(u'销售收入（元）')

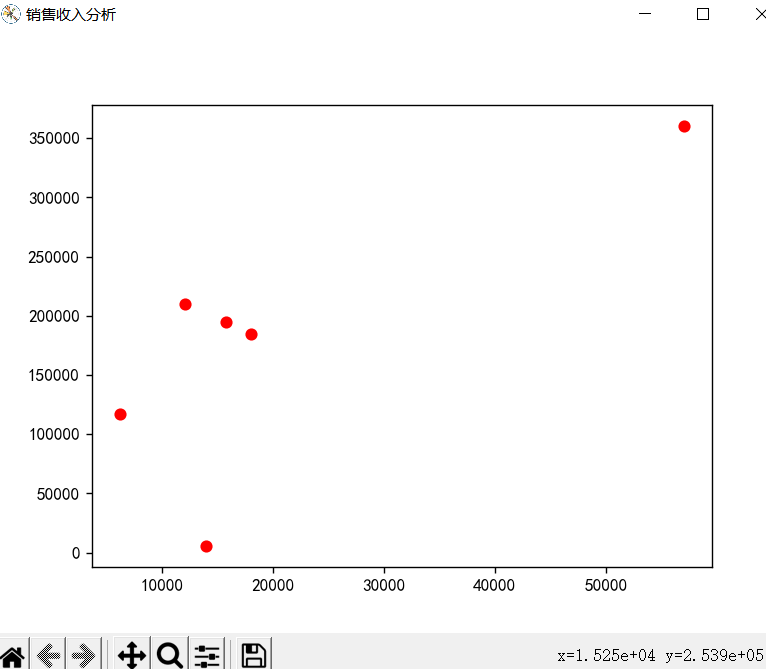
plt.xlabel(u'广告费（元）')

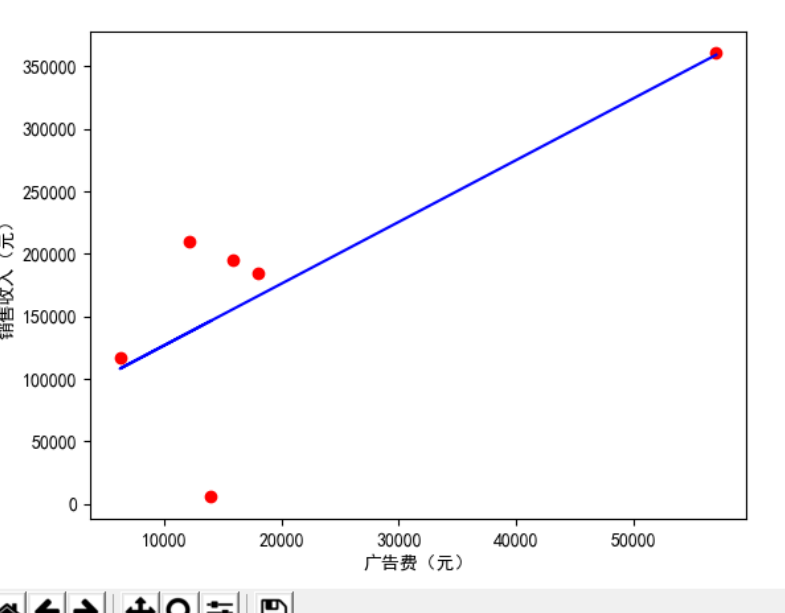
plt.show()

3、**系统预览**



1. **实验结果与分析**





**六、结论与体会**

结论与体会：本次实验学习了各种对数据的处理的方法，实验遇到了一些困难，通过上网查资料解决。

**七、教师评语**