# 实验八 Python数据分析与可视化综合实验

## 一、实验目的

1、熟悉Python进行数据分析过程；

2、掌握numpy基本语法；

3、熟悉pandas基本使用方法；

4、掌握matplotlib绘图方法。

## 二、实验内容

1、（1）运行给定的程序，在当前文件夹中生成饭店营业额模拟数据文件data.csv。

（2）然后完成下面的任务：

1）使用 pandas 读取文件 data.csv 中的数据，创建 DataFrame 对象，并删除其中所有缺失值；

2）使用 matplotlib 生成折线图，反应该饭店每天的营业额情况，并把图形保存为本地文件 1.jpg；

3）按月份进行统计，使用 matplotlib 绘制柱状图显示每个月份的营业额，并把图形保存为本地文件2.jpg，使得生成的2.jpg中每个柱上的上部显示该柱对应的数据值；

4）按月份进行统计，找出相邻两个月最大涨幅，并把涨幅最大的月份写入文件maxMonth.txt；

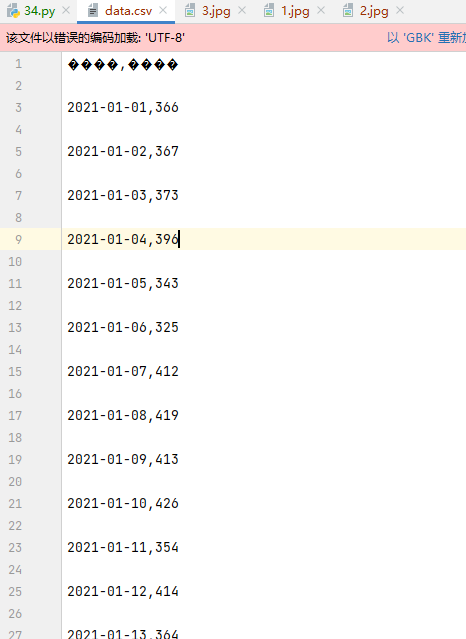
5）按季度统计该饭店 2021 年的营业额数据，使用 matplotlib 生成饼状图显示 2021年 4个季度的营业额分布情况，并把图形保存为本地文件 4.jpg，每个扇形要能够显示对应的比例。

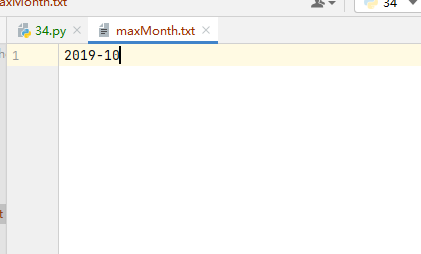
2、编写程序，用NumPy绘制学生成绩雷达图。

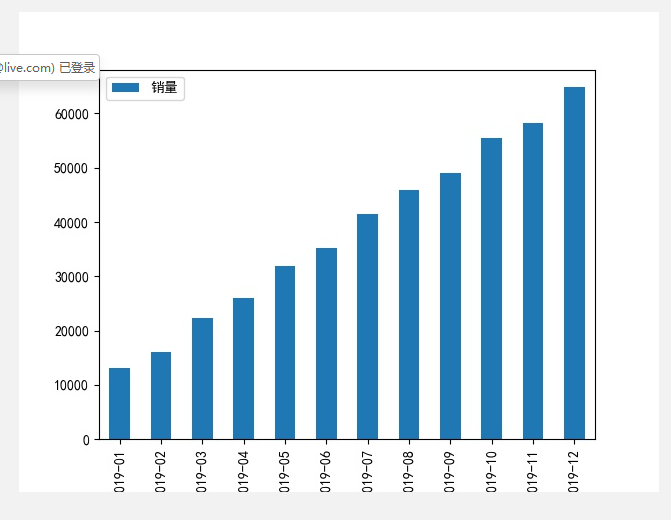
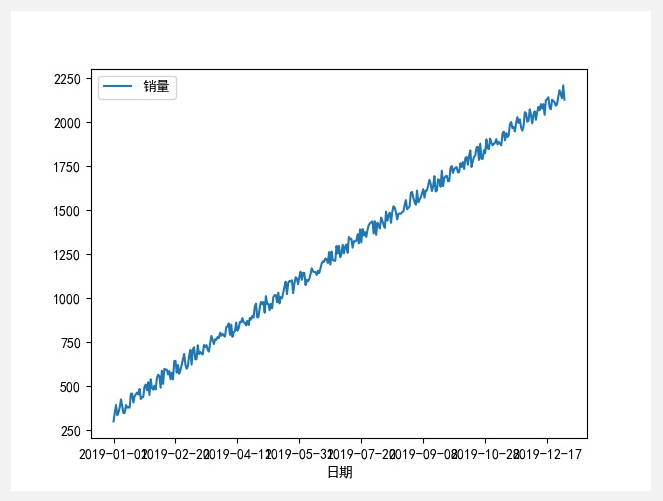
## 三、实验代码和运行结果截图

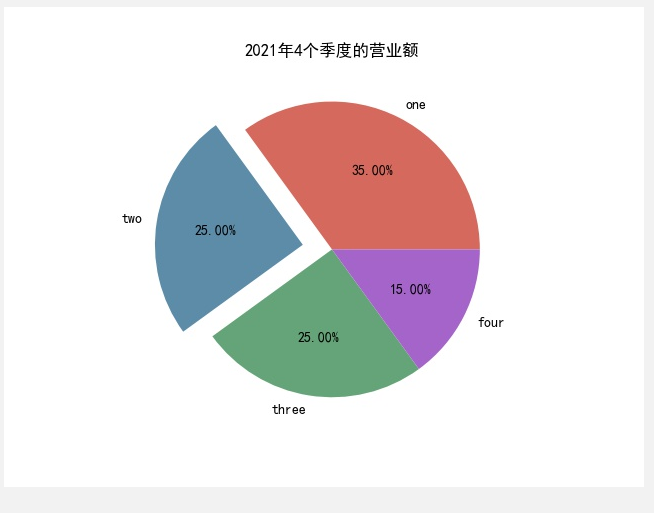
1、

*"""  
@Time ： 2021/12/7 14:09  
@Auth ： 021321752215舒洪凡  
@File ：34.py  
@IDE ：PyCharm  
"""  
  
#在当前文件夹中生成饭店营业额模拟数据文件data.csv***import** csv  
**import** random  
**import** datetime  
  
**import** numpy **as** np  
  
fn=**'D:/DEMO/python/Demo02/34/data.csv'  
with** open***(***fn,**'w'*)*as** fp:  
 wr=csv.writer***(***fp***)*** *#一行一行写入csv文件* wr.writerow***([*'日期'**,**'销量'*])*** *#写入表头* startDate=datetime.date***(***2021,1,1***)*** *#生成模拟时间* **for** i **in** range***(***365***)***:*#生成365个模拟数据，写入csv* amount=300+5\*i+random.randrange***(***100***)*** wr.writerow***([***str***(***startDate***)***,amount***])*** startDate=startDate+datetime.timedelta***(***days=1***)****#下一天  
  
#使用pandas读取数据文件，创建DataFrame对象，并删除其中所有缺失值；  
#使用matplotlib生成每天营业额折线图，保存图形为first.jpg  
#使用matplotlib绘制每个月营业额柱状图，保存为second.jpg  
#找出相邻两个月最大涨幅，并写入maxMonth.txt  
#使用matplotlib显示2018年4个季度的营业额分布饼状图，保存为third.jpg***import** pandas **as** pd  
**import** matplotlib.pyplot **as** plt  
**import** matplotlib  
matplotlib.rcParams***[*'font.sans-serif'*]***=***[*'SimHei'*]*** *#pyplot中文显示*df=pd.read\_csv***(*'data.csv'**,encoding=**'cp936'*)***df=df.dropna***()****#读取数据，丢弃缺失值  
#生成天营业额折线图*plt.figure***()***df.plot***(***x=**'日期'*)***plt.savefig***(*'D:/DEMO/python/Demo02/34/1.jpg'*)****#生成月营业额柱状图*plt.figure***()***df1=df***[***:***]*** *#str1.rindex(str2)返回子串 str2 在串str1中最后出现位置，如果没有匹配的字符串会报异常*df1***[*'month'*]***=df1***[*'日期'*]***.map***(*lambda** x:x***[***:x.rindex***(*'-'*)])****#提取出月份,新建了一个month列出来？？*df1=df1.groupby***(***by=**'month'**,as\_index=**False*)***.sum***()*** *#as\_inside=False不把month作为新的index*df1.plot***(***x=**'month'**,kind=**'bar'*)***plt.savefig***(*'D:/DEMO/python/Demo02/34/2.jpg'*)****#查找涨幅最大月份，写入文件*plt.figure***()***df2=df1.drop***(*'month'**,axis=1***)***.diff***()****#在销量列中每月跟上月相减的差集*m=df2***[*'销量'*]***.nlargest***(***1***)***.keys***()[***0***]****#查找销量列中差集最大的一个数所对应的索引***with** open***(*'D:/DEMO/python/Demo02/34/maxMonth.txt'**,**'w'*)*as** fp:  
 fp.write***(***df1.loc***[***m,**'month'*])****#索引m对应的month  
  
#生成季度营业额饼状图*plt.figure***()***one=df1***[***:3***][*'销量'*]***.sum***()***two=df1***[***3:6***][*'销量'*]***.sum***()***three=df1***[***6:9***][*'销量'*]***.sum***()***four=df1***[***9:12***][*'销量'*]***.sum***()***y = np.array***([***35, 25, 25, 15***])***plt.pie***(***y,labels=***[*'one'**,**'two'**,**'three'**,**'four'*]***,  
 colors=***[*"#d5695d"**, **"#5d8ca8"**, **"#65a479"**, **"#a564c9"*]***, *# 设置饼图颜色* explode=***(***0, 0.2, 0, 0***)***, *# 第二部分突出显示，值越大，距离中心越远* autopct=**'%.2f%%'**, *# 格式化输出百分比****)***;  
plt.title***(*"2021年4个季度的营业额"*)***plt.savefig***(*'D:/DEMO/python/Demo02/34/3.jpg'*)***



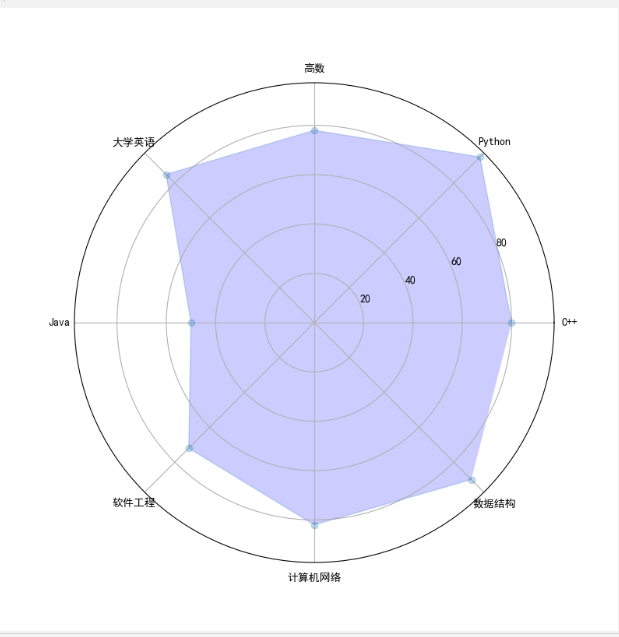




2、

*"""  
@Time ： 2021/12/7 14:56  
@Auth ： 021321752215舒洪凡  
@File ：35.py  
@IDE ：PyCharm  
"""***import** numpy **as** np  
**import** matplotlib.pyplot **as** plt  
**import** matplotlib  
*#coding:utf-8*plt.rcParams***[*'font.sans-serif'*]***=***[*'SimHei'*]*** *#用来正常显示中文标签,下一行的u也是为了正常显示中文*courses = ***[*u'C++'**, **u'Python'**, **u'高数'**, **u'大学英语'**, **u'Java'**, **u'软件工程'**, **u'计算机网络'**, **u'数据结构'*]***scores = ***[***80, 95, 78, 85, 50, 72, 82, 90***]****#把圆周分为datalength份*datalength = len***(***scores***)*** *#数据长度*angles = np.linspace***(***0, 2 \* np.pi, datalength, endpoint=**False*)****#设置为极坐标格式*fig, ax=plt.subplots***(***figsize=***(***8, 8***)***, subplot\_kw=dict***(***polar=**True*))****#修改网格的标签为courses列表*plt.thetagrids***(***angles \* 180 / np.pi, courses***)****#绘制雷达图外部轮廓*plt.plot***(***angles,scores,**'o-'**,linewidth=1,alpha=0.25***)****#填充内部*plt.fill***(***angles, scores, facecolor=**'b'**, alpha=0.2***)***plt.show***()***



## 四、实验小结

通过本次实验，我掌握了 python 在数据分析方面的基础知识，熟悉pandas基本使用方法（从各种文件格式中导入文件，并对文件中的数据进行运算操作）；并且使用matplotlib绘图（线图、散点图、等高线图、条形图、柱状图、3D 图形）。

存在的问题是 使用不熟练。