**实验12——Pandas数据分析库2**

|  |  |  |  |
| --- | --- | --- | --- |
| 学号 | 4201438 | 姓名 | 张雍汇 |
| **一、实验目的**  掌握pandas时间序列应用，并能够利用pandas对实际数据进行分析。 | | | |
| **二、实验准备**  连接网络的计算机一台 | | | |
| **三、实验步骤**  **1.独立完成以下程序的编写：**  股票每日的涨跌幅在区间[-10%, 10%]内，且只在周一到周五交易。先设置随机数种子为7，试生成模拟一只股票2019年全年交易日的涨跌幅数据。然后计算：  （1）该股票2019年的年涨跌幅。  （2）统计12个月中涨幅最大的月份和跌幅最大的月份。  2.利用tushare接口一次性下载所有股票基本面信息至stock.xlsx文件：  （1）按地区统计上市公司数量  （2）计算沪市（60开头）、深圳主板（00开头）、创业板（30开头）及科创板（68开头）各板块的pe值和股票数。  3.在studentscore.xlsx中存储了10个学生的成绩数据，请利用pandas库实现如下功能：  （1）利用pandas库中的read\_csv()函数读取studentscore.xlsx中的数据，并存入一个DataFrame对象df中。  （2）为df增加一列，列索引为“总分”，其值为相应学生的五门课程成绩之和。  （3）为df再增加一列，列索引为“均分”，其值为相应学生五门课程的均分。  （4）将更新过的df通过pandas库的to\_csv()函数写入文件studenscore.csv。  （5）输出五门课程平均分超过90分的学生学号和姓名。  （6）输出数学分数超过班级平均分的学生学号和姓名。  （7）按照总分降序输出学生的成绩单。   |  | | --- | | **四、实验过程及截图****[学生填写]**  （1）该股票2019年的年涨跌幅。      （2）统计12个月中涨幅最大的月份和跌幅最大的月份。      2.利用tushare接口一次性下载所有股票基本面信息至stock.xlsx文件：  （1）按地区统计上市公司数量    area  浙江 445  江苏 414  北京 332  广东 313  上海 297  深圳 290  山东 205  福建 136  四川 124  湖南 104  湖北 103  安徽 103  河南 79  辽宁 74  河北 57  新疆 55  陕西 52  天津 51  重庆 50  江西 42  吉林 41  广西 38  黑龙江 38  山西 37  云南 36  甘肃 32  海南 31  贵州 29  内蒙 25  西藏 18  宁夏 14  青海 12  Name: area, dtype: int64  （2）计算沪市（60开头）、深圳主板（00开头）、创业板（30开头）及科创板（68开头）各板块的pe值和股票数。      3.在studentscore.xlsx中存储了10个学生的成绩数据，请利用pandas库实现如下功能：  （1）利用pandas库中的read\_csv()函数读取studentscore.xlsx中的数据，并存入一个DataFrame对象df中。       1. 为df增加一列，列索引为“总分”，其值为相应学生的五门课程成绩之和。      1. 为df再增加一列，列索引为“均分”，其值为相应学生五门课程的均分。      1. 将更新过的df通过pandas库的to\_csv()函数写入文件studenscore.csv。      1. 输出五门课程平均分超过90分的学生学号和姓名。      1. 输出数学分数超过班级平均分的学生学号和姓名。     （7）按照总分降序输出学生的成绩单。    import pandas as pd  df = pd.read\_excel('studentscore.xlsx', dtype={'code': 'str'})  df["总分"]=df.语文+df.数学+df.英语+df.体育+df.美术  df["均分"]=df.总分/5  df.to\_csv("studenscore.csv")  df1=df[df.均分>90]  df1=df1[["学号","姓名"]]  print("平均成绩超过90分")  print("--------------")  print(df1)  print("--------------")  m=df.数学.mean()  print("数学成绩不足平均分")  df3=df[df.数学>m]  df3=df3[["学号","姓名"]]  print("--------------")  print(df3)  print("--------------")  df=df.sort\_values("总分",ascending=False)  print("--------------")  print(df)  print("--------------") | | **五、实验结论与分析讨论[学生填写]**  **（1）列出本次实验你认为重要的概念、术语。**  **索引，索引选取，groupby**  **（2）谈谈你认为本次实验需要注意的地方。**  **Groupby返回的为一个series**  **（3）实验过程中是否存在问题，如果有，你解决的办法是什么？**  **没有问题** | | | | |