**Python与数据分析**

**实验报告**

|  |  |  |
| --- | --- | --- |
| **课程名称：**Python与数据分析 | **学期：**2022~2023学年上学期 | **成绩：** |
| **指导教师：**王灿 | **学生姓名：** | **学生学号：** |
| **实验名称：**pandas数据处理 | | |
| **实验编号：**11 | **实验日期：**10月24日 | **实验学时：**NA |
| **学院：**理学院 | **专业：**信息与计算科学 | **班级：**信计xxxx |

**一、实验目的**

1．熟悉pandas数据处理方法。

**二、实验内容**

1. 查阅pandas官方文档，完成下面的操作。
   1. 使用pd.read\_csv()函数读入“成绩表示例.csv”，用df变量来存储
   2. 对df使用.iloc[选取行,选取列]方法提取王十、冯廿、郑七的姓名和数学、语文、英语成绩（注意，4列的顺序必须是姓名、数学、语文、英语，可以用列表代替切片来实现非等差数列的选择），并把结果重新赋值为df\_part1
   3. 对df使用.loc[选取行,选取列]方法实现与1.2相同的提取效果，并分析loc和iloc在选取行的时候使用的切片有何区别（比如是否包括起止点），并把结果重新赋值为df\_part2
   4. 使用.equals()方法判断df\_part1和df\_part2是否相同，需要返回True
   5. 对df使用.set\_index()方法将'学号'和'班级'设置为index，并把结果重新赋值为df，即df = df.set\_index(填入正确参数实现要求)
   6. 对df使用.swaplevel()方法交换index中'学号'和'班级'的顺序，使'班级'在前，并把结果重新赋值为df
   7. 对df使用.sort\_index()方法按照'班级'和'学号'排序，并把结果重新赋值为df
   8. 对df使用.groupby()和.mean()，按照班级进行分组，并计算每个组每门课程的平均成绩（注意groupby参数使用level还是by）
   9. 对df使用.groupby()和.mean()，按照性别进行分组，并计算每个组每门课程的平均成绩（注意groupby参数使用level还是by）
   10. 对df使用.sort\_values()按照数学成绩降序排列
   11. 对df使用.sort\_values()对df按照语文、数学、英语成绩降序排列，即先按照语文排，语文相同时用数学排，数学还相同则用英语排

**三、实验环境**

Windows/Linux/Mac OS操作系统；conda创建的python3.10环境。

**四、实验过程和结果**

**1．实验步骤、结果和结论（包括必要的截图）**

**2．关键代码及其解释**

**3．调试过程（如果和后面的实验总结重复，可以只写实验总结）**

**五、实验总结（不能为空）**

**1．遇到的问题及解决过程**

**2．产生的错误及原因分析**

**3****．体会和收获**

**六、参考文献**

[1]柳毅等著，Python数据分析与实践，北京：清华大学出版社，2019.7.

**七、教师评语**