Python数据处理与分析

大作业要求2022

目录

[一、大作业提交要求 1](#_Toc101777391)

[二、实验平台 1](#_Toc101777392)

[三、大作业内容和要求（以下作业内容二选一） 1](#_Toc101777393)

[作业1．葡萄酒评论分析报告 1](#_Toc101777394)

[作业2. 大学排行榜分析 2](#_Toc101777395)

[四、实验报告（每个人需上交） 3](#_Toc101777396)

## 一、大作业提交要求

（1）本次作业有两个题目，要求二选一完成即可，第13周前每位同学提交本文档中第四部分【实验报告】及【py代码文档】。实验报告命名方式为班级+学号+姓名.docx，实验报告和代码一起打包，打包后命名学号+姓名，学委收齐后交给任课教师。

**（为降低难度，本次作业给出参考代码模板，同学们可以参考，但需补充缺失代码并进行完整调试。）**

（2）要求作业必须独立完成，每个步骤有描述有截图（为体现截图的独创性可以截取运行时完整桌面），截图和代码要有独创性，完全拷贝他人的截图和代码会扣除60分。

（3）本次大作业成绩计入期末考试总成绩。

## 二、实验平台

Python解释器版本：Python3.7.0及以上

开发工具：可用Pycharm或者Anaconda或者IDLE(不限制开发环境)。

## 三、大作业内容和要求（以下作业内容二选一）

## 作业1．****葡萄酒评论分析报告****

**任务描述:**

数据格式如下所示：

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| number | country | description | points | price | province ‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬ |  |
| 30 | France | "Red cherry fruit comes laced with...." | 86 | 15 | Beaujolais |  |
| 50 | Italy | "This blend of Nero Avola and Syrah...." | 86 | 15 | Sicily |  |
| 100 | US | "Fresh apple lemon and pear flavors...." |  | 88 | 18 | New York |

通过分析这些数据，用户可以根据产地、评份、价格等挑选适合自己的葡萄酒，商家可以分析消费者的购买行为习惯，可以更加准确地提供适合市场的产品，精准定位客户。

请读取文件中的数据，完成以下任务：

（1）输入'国家名列表'，统计文件中出现的葡萄酒生产国家，输出不重复的国家名列表，按字母表升序排序,若国家名数据缺失，略过该条数据，返回值中不包含空字符串元素

（2）输入'平均分'，计算每个国家的葡萄酒的平均得分( 保留最多2位小数)，返回值为国家名和得分的列表

（3）输入'平均分排序'，计算每个国家的葡萄酒的平均得分，返回值为国家名和得分的列表，按评分由高到低降序排列

（4）输入'评分最高'，输出评分最高的十款葡萄酒的编号、出产国、评分和价格，按评分降序输出

（5）输入'价格最高'，输出价格最高的二十款葡萄酒的编号、出产国、评分和价格，按价格降序输出

（6）输入'葡萄酒评分'，统计各个评分的葡萄酒数量是多少，输出包含评分和数量的列表

（7）输出拥有葡萄酒数量最多的评分和数量

（8）输出拥有葡萄酒数量最多的评分的葡萄酒的平均价格

以下示例仅为格式展示，与最终结果无关

**示例 1**；

|  |
| --- |
| 输入：国家名列表  输出：['Argentina', 'Armenia', ...... 'US', 'Ukraine', 'Uruguay'] |

**示例 2**

|  |
| --- |
| 输入：平均分  输出：[['Argentina', 86.72], ['Armenia', 87.0],...... ['Ukraine', 83.0], ['Uruguay', 88.0]] |

**示例 3**

|  |
| --- |
| 输入：葡萄酒评分  输出：[[80, 38], [81, 71], ...... [95, 140], [96, 50], [97, 26], [98, 8], [99, 3]][86, 1743] |

## 作业2.大学排行榜分析

任务描述：文件alumni.txt和soft.txt为按照不同评价体系给出的国内大学前100名排行，对比两个排行榜前10名的学校的上榜情况，分析不同排行榜排名的差异。

请读取文件中的数据，完成以下任务：

（1）输出在alumni.txt和soft.txt榜单中均排名前10名的大学

（2）分别输出在alumni.txt和soft.txt榜单中排名前10名的大学

（3）输出在alumni.txt中前10名但未出现在soft.txt榜单中前10名的大学

（4）输出没有同时出现在alumni.txt前10名和soft.txt榜单中前10名的大学

## 四、实验报告（每个人需上交）

|  |  |  |
| --- | --- | --- |
| Python数据处理与分析大作业实验报告 | | |
| 题目：《葡萄酒评论分析报告》或《大学排行榜分析》 | | 日期： |
| 班级： | 学号： | 姓名： |
| 实验环境：  自己填写 | | |
| 实验内容与完成情况：  （1）输入'国家名列表'，统计文件中出现的葡萄酒生产国家，输出不重复的国家名列表，按字母表升序排序,若国家名数据缺失，略过该条数据，返回值中不包含空字符串元素  【请插入第一步执行结果截图】  （2）输入'平均分'，计算每个国家的葡萄酒的平均得分( 保留最多2位小数)，返回值为国家名和得分的列表  【请插入第二步执行结果截图】  （3）输入'平均分排序'，计算每个国家的葡萄酒的平均得分，返回值为国家名和得分的列表，按评分由高到低降序排列  【请插入第三步执行结果截图】  ……….  ………  ……….  （8）输出拥有葡萄酒数量最多的评分的葡萄酒的平均价格  【请插入第8步执行结果截图】  (每一步需有截图) | | |
| 实验过程出现的问题：（1-3个问题）  问题1描述：【请插入问题1截图】  问题2描述：【请插入问题2截图】  问题3描述：【请插入问题3截图】  程序调试中出现问题需有截图 | | |
| 解决方案（列出遇到的问题和解决办法，列出没有解决的问题）：  对应上一步问题，在此注明解决方法，可以截图或者代码  问题1解决方案：  问题2解决方案：  问题3解决方案： | | |
| 实验总结：（总结本次实验用到的知识点及解决问题思路和自己收获至少100字）： | | |