2021-2022学年第2学期

**《Python程序设计》课程项目任务书**

1. **目的**
2. 了解用 Python 开发应用程序的一般流程和方法；
3. 练习用程序模块化和面向对象方法开发复杂程序的技能；
4. 培养利用计算机解决问题的计算思维能力。
5. **任务**

鼓励自己设计题目，最好是与专业相关的数据分析类题目。

1. **实施步骤**

按下面的顺序完成项目任务：

1. 选题

确定要编写的程序，包括名称和主要功能， 以及背景、 适用范围等。

1. 需求分析

确定程序需要完成哪些具体功能；

确定需要哪些数据和采集途径；

自己设计的题目要提前和老师讨论，并得到老师的认可。

1. 设计

确定程序包括哪些功能函数，以及函数的原型；

确定需要定义哪些类，以及类的成员。

1. 实现

使用了哪些标准库模块和/或扩展库；

编写代码实现上一步设计的类和/或功能函数。

1. 调试程序
2. 编写项目报告（模板见附2）
3. 提交成果
4. **完成方式**

2~3位同学组成一个小组完成，独自完成也可以。如果是小组做实验，则一起做分析和设计，分头编写代码，共同编写一份项目报告，写明每个组员的实际任务和完成情况，最后由一位成员提交全部成果。

1. **程序运行方式**

代码保存在1个或多个 .py 文件中，直接执行文件；或者在Jupyter Notebook中创建和执行程序。

**数据分析和可视化**

**（1）目的**

练习numpy、scipy、Pandas、matplotlib扩展库的基本使用方法。

练习用Pandas进行简单的数据分析。

练习用matplotlib绘图的基本技巧。

**（2）内容**

有8个数据集，共8个文件：

|  |  |  |  |
| --- | --- | --- | --- |
| **序号** | **数据集名称** | **文件名** | **说明** |
| 1 | IMDB电影评分数据 | imdb\_movies.csv |  |
| 2 | 个人年收入调查数据 | income.csv |  |
| 3 | 商店会员购物数据 | shopping.csv |  |
| 4 | 豆瓣电影数据集 | movie\_douban.zip | 爬取的豆瓣电影数据集，包含了 2011 — 2016 内地电影票房排行前25共计150部电影的基本信息、影评及影人的基本信息。 |
| 5 | 全球新冠肺炎COVID-19数据集 | covid-19.zip | 包含每日时间序列汇总表，包括确诊、死亡和治愈。所有数据来自每日病例报告。数据覆盖日期20/01/22-20/10/11 |
| 6 | 特朗普集会演讲数据集 | DonaldTrump.zip | 特朗普35次集会演讲数据集，以TXT文件的形式完整提供 |
| 7 | 安居客租房信息数据集 | 安居客租房信息.xlsx | 安居客（上海）的租赁信息，包括地点、楼层、价格、租赁方式、区域、小区、面积、户型等 |
| 8 | 电影评分数据 | ml-100.zip | MovieLens电影评分数据，详见数据集中的Readme文档 |

请尝试用Pandas读取其中1个文件中的数据，针对某个主题进行分析，然后用matplotlib绘制图形展示分析得到的数据。

以下是几个**参考主题**和难度级别：

（a）IMDB电影评分

imdb-1：评分最高的3部电影（简单）

imdb-2：最受关注的3类电影（适中）

imdb-3：收益最高的3类电影（适中）

imdb-4：参演电影类型最多的3位演员（难）

imdb-5：电影评分与票房的关系（适中）

（b）个人年收入调查

income-1：证明“受教育程度越高收入就越高”（适中）

（c）商店会员购物

shopping-1：哪个年龄段的女士最喜欢用现金支付（适中）

shopping-2：不同年龄性别人群的消费能力（适中）

shopping-3：用现金买水果的男士中哪个年龄段的最多（适中）

（d）MovieLens电影评分数据

Movielens-1：男性和女性评分差异分析（适中）

**（3）要求**

利用numpy、scipy、Pandas、matplotlib扩展库完成基本的数据**统计分析**和**可视化**功能，实现基本目标。

在Jupyter Notebook中完成，不要修改数据文件的名称和内容。

每个小组至少对8个数据集中的4个数据进行分析，每个数据至少选1个主题（可以使用上面举例的参考主题，例如imdb主题等）。鼓励设计新的分析主题。

附件2

2021-2022学年第2学期

**《Python程序设计》课程项目报告**

|  |  |  |  |
| --- | --- | --- | --- |
| 学 生1 | 学号： | 姓名： | 成绩： |
| 学 生2 | 学号： | 姓名： | 成绩： |
| 学 生3 | 学号： | 姓名： | 成绩： |
| 学 生4 | 学号： | 姓名： | 成绩： |
| 学 生5 | 学号： | 姓名： | 成绩： |
| 指导教师 |  | | |
| 完成日期 | 2022年 月 日 | | |

1. **选题**

（写明程序名称，还可以简要说明选题原因。）

1. **需求分析**

（假设程序应用背景，说明需要完成的功能。）

1. **设计**

（说明程序的主要模块；说明需要定义的类及其成员；说明需要定义的功能函数及其算法。鼓励增加图表。如果是小组，说明分工情况。）

1. **关键代码**

（粘贴主要类和功能函数的代码，并解释。）

1. **运行效果**

（在程序运行过程中截图，粘贴进来并加上文字解释。说明程序还存在的运行时错误。）

1. **总结**

（说明本次项目所完成的主要任务，遇到了什么问题，是否解决了问题，程序有什么特点，本次项目有什么收获，有哪些不足。）

1. **评语**

|  |  |
| --- | --- |
| 教师  评语 |  |