

2022 春《Python 玩转大数据》课程大作业

作业内容：

- 1、撰写一份用 Python 语言进行数据处理以解决相关问题的**设计方案**。

设计方案应首先阐明要解决的问题，对要解决的问题抽象出与之相关数据（内容与要求），给出使用 Python 工具从网上获取这些数据源的方法（也可以使用自己专业领域存在的数据源），对于获取的数据源，构建一个数据处理模型，表明使用 Python 的哪些内置库及第三方库对这些数据进行处理，处理的结果应以图形、图表或图像等形式表示，再对结果分析、解释，并最终给出一个结论。

- 2、编写从网上爬取与问题相关数据、保存为数据文件，及对数据进行处理及结果表示的 Python 的程序编码。
- 3、获取的与问题相关的源数据以文件形式保存。

作业要求：

- 1、掌握 Python 语言的编程环境，熟悉与所解决问题相关的 Python 内置库及第三方库的使用。
- 2、与问题相关的源数据可以通过 Python 工具从网上获取，并保存为文件，或阐明自己专业领域相关的大数据。
- 3、对数据处理的结果应以图形、图表或图像形式展现。
- 4、对处理的结果，要从自己理解的角度进行分析，并最终给出分析的结论。

作业形式：

上交的作业包括：

- 1、设计方案一份(word 文档)
- 2、Python 语言程序代码文件(.py, ipynb)
- 3、从网上获取的与问题相关的源数据所保存的文件(.csv, .json, .xlsx)及处理结果的保存文件，图表应以图像文件或网页文件保存。
- 4、数据处理结果分析报告一份(word 文档)。

欢迎大家设计一个用 Python 对**本专业相关课程中的大数据**进行处理、数据分析及可视化大作业。

本次大作业上交**截止时间：2022 年 6 月 12 日**

本次作业以小组为单位，每个小组由组长负责，完成的作业放在以“班级+小组”为名的文件夹中，上传到交作业的服务器（FTP://10.111.100.207）上的对应文件夹中。

希望大作业**成绩达优**的小组同学，请于 6 月 5 日前将**大作业设计方案**通过 QQ 发给我，提前报名，待大作业交后，**经审核**，再确定是否可以参加**申优答辩**。申优答辩时间将安排在 6 月 20 日之后，具体时间和地点，见 QQ 群通知。