

Wrangle report

此次数据分析经数据收集、评估、清理、分析和可视化等几大步骤进行。

数据收集：

文件一通过连接下载，导入 jupyter。

文件二通过 requests 模块进行下载，并利用编程方式保存至本地。

文件三通过链接下载 json 文件，并将文件转换为 dataframe。

确定自己想要研究的问题。

数据评估：

目测评估，加载三个文件，并对三个文件进行大体了解，如各列名含义、数据单位、是否存在缺失值、异常值、表结构是否合理等。

编程评估，对三个文件依次通过 .info() / .head() / .tail() / .sample() / .describe() / .value_counts() 等函数，来查看数据是否存在缺失值、异常值、数据类型是否错误等情况。

最终根据自己的研究需求，对发现的数据质量和整洁度问题进行汇总。

数据清理：

首先备份三个数据文件，再依次对问题通过定义，代码，测试等步骤来进行解决。

最终将清理后的完整文件保存。

数据分析和可视化：

对自己想要分析的有关数据进行相关分析和可视化，并根据可视化分析结果，得出相应结论。

提出本次分析的局限性。