计算机学院 大数据分析实践 课程实验报告

|  |  |  |
| --- | --- | --- |
| 实验题目：数据质量实践 | | 学号：202300130235 |
| 日期：2025.9.19 | 班级： 23数据 | 姓名： 危弘毅 |
| 实验目的：  本次实验主要围绕宝可梦数据集进行分析，考察在拿到数据后如何对现有的数据进行预处理清洗操作，建立起对于脏数据、缺失数据等异常情况的一套完整流程的认识。 | | |
| 实验软件和硬件环境：  python3.9，jupyter notebook | | |
| 实验步骤与内容：   1. 读入csv文件，删去最后4行无效数据     2.筛选出合法的type 2取值，过滤    3.去重    4.去除异常的attack值    5.还原被交换的Generation和Legendary列    6.过滤Generation和Legendary不合法的数据 | | |
| 结论分析与体会：  本次实验主要围绕宝可梦数据集进行分析，通过对现有的数据进行预处理清洗操作，我初步建立起对于脏数据、缺失数据等异常情况的一套完整流程的认识。数据清洗是数据预处理的核心环节，从杂乱原始数据到可用分析数据的转化，为后续分析利用数据提供了基础。 | | |