计算机学院 大数据分析实践 课程实验报告

|  |  |  |
| --- | --- | --- |
| 实验题目：数据采样实践 | | 学号：202300130235 |
| 日期：2025.9.19 | 班级： 23数据 | 姓名： 危弘毅 |
| 实验目的：  利用Pandas库实现多种数据采样和过滤的方法 | | |
| 实验软件和硬件环境：  python3.9，jupyter notebook | | |
| 实验步骤与内容：   1. 库的导入与数据的读入       2. 采用dropna方法并指定参数为any删除多余的空行      过滤得到traffic不等于0且from\_level=一般节点的数据      3.对数据抽样  （1）加权采样：to\_level的值为一般节点与网络核心的权重之比为 1 : 5      （2）随机抽样      （3）分层抽样，根据比例一般节点抽17个，网络核心抽33个 | | |
| 结论分析与体会：  通过本次实验，我了解了数据采样的各种方法，学会了使用pandas库进行基本的数据操作 | | |