机器学习实验报告

实验一：数据的获取

李家兴 202212063012

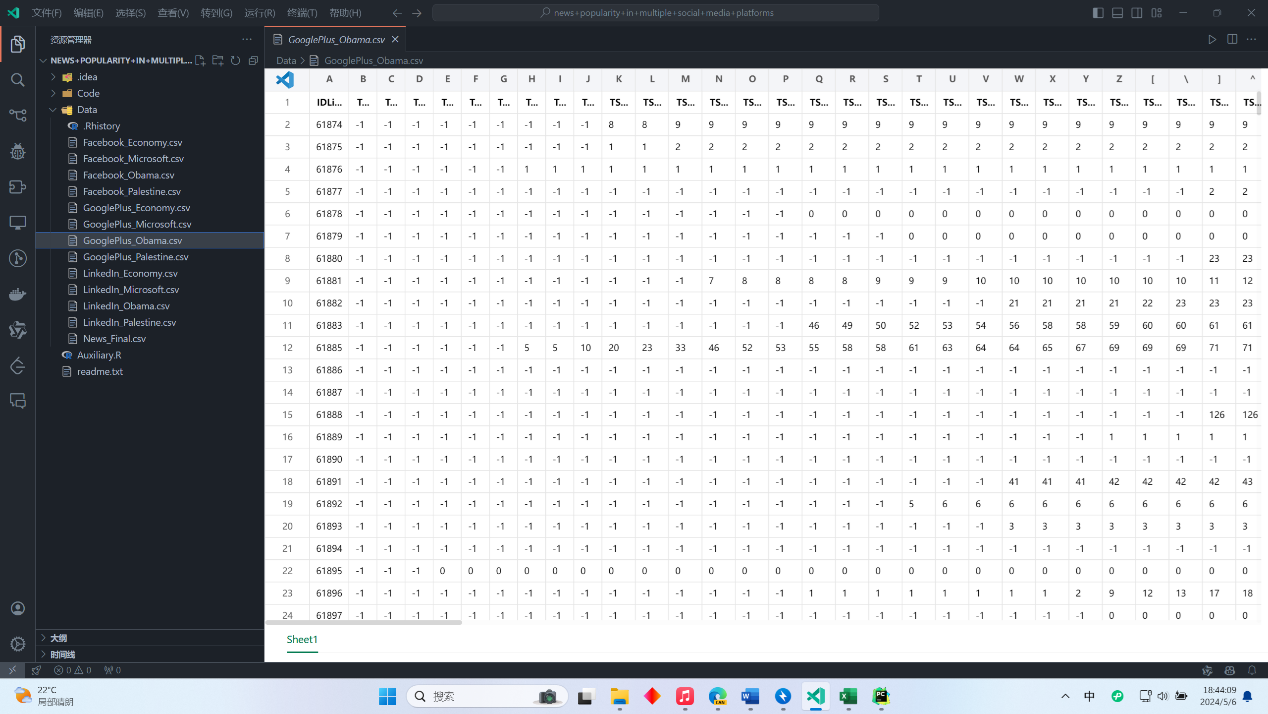
1. 实验目的

学习如何获取数据，特别是使用UCI数据库中的数据集作为示例，了解数据获取的整个流程，包括数据的来源、存储、清洗，以及后续的研究方向。

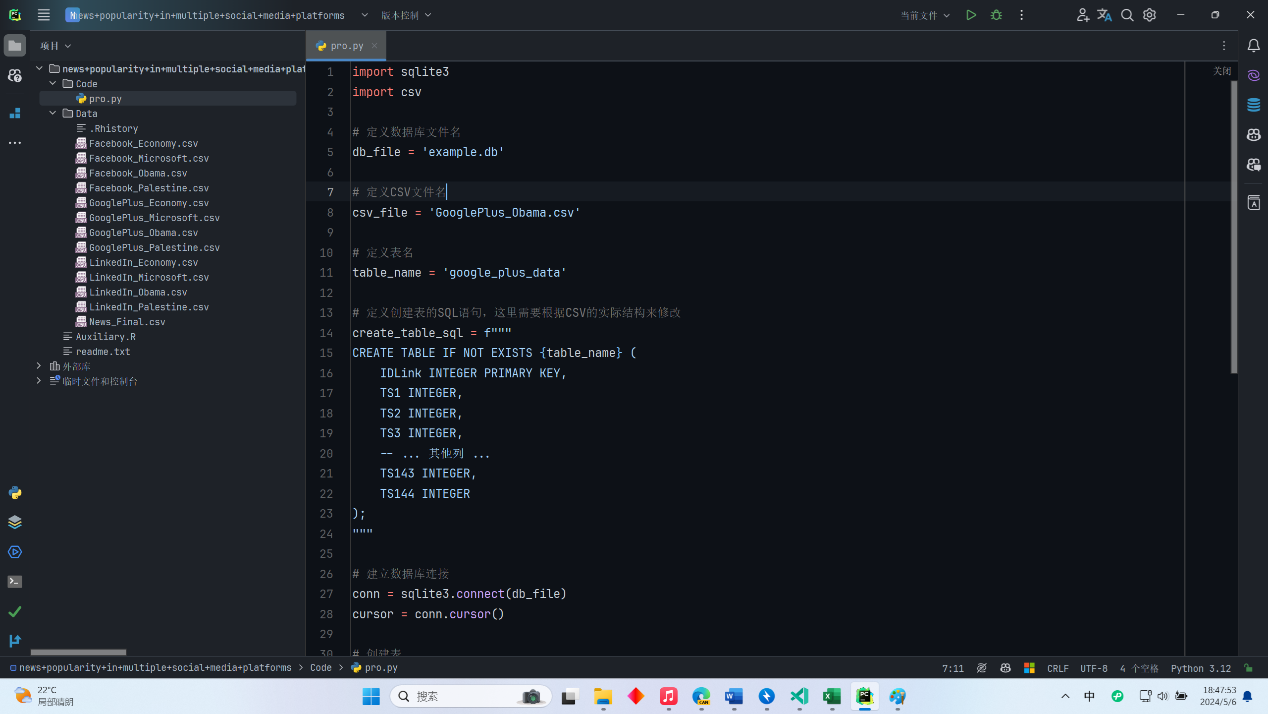
1. 实验内容
2. 数据的来源：使用搜索引擎找到UCI repository等可靠的数据来源。

这里选用了[News Popularity in Multiple Social Media Platforms](https://archive.ics.uci.edu/ml/datasets/News+Popularity+in+Multiple+Social+Media+Platforms): 这个数据集包含在多个平台（Facebook, Google+ 和 LinkedIn）上的新闻条目及其相应的社会反馈的大数据集。

1. 数据存储：将数据从CSV存储到数据库中

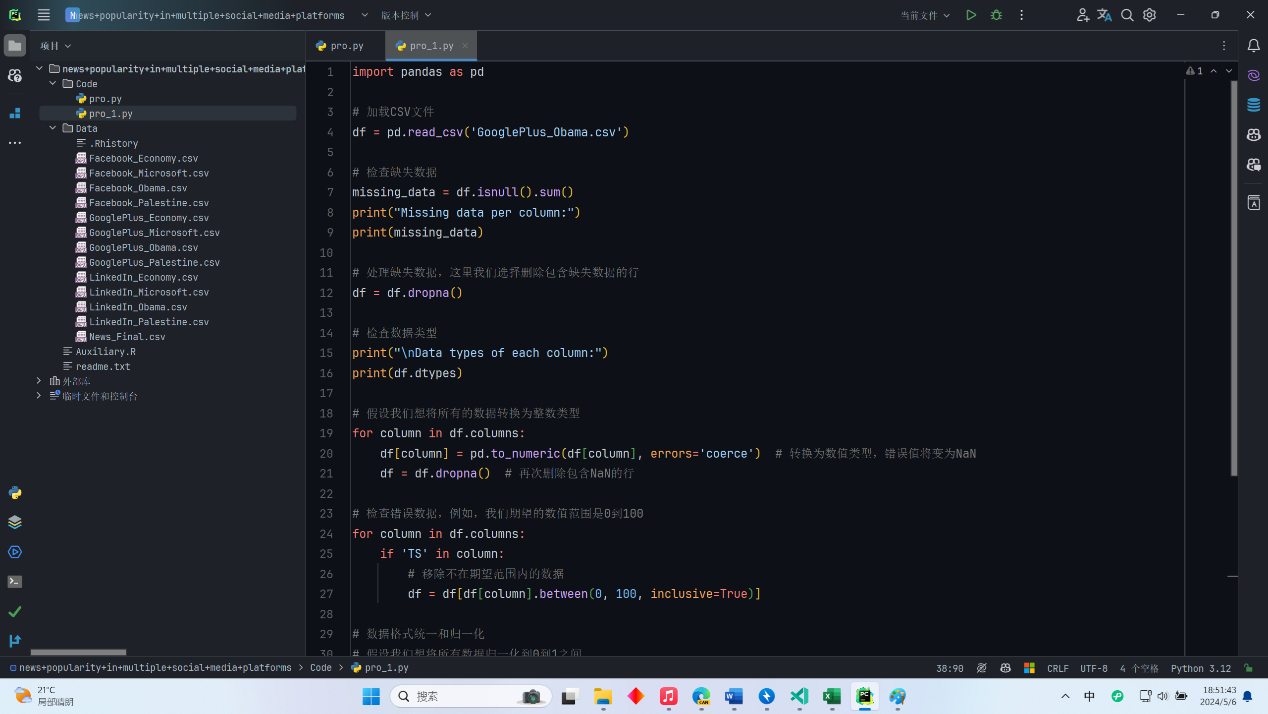


一个简单的将csv文件写入MySQL数据库的Python脚本如图：



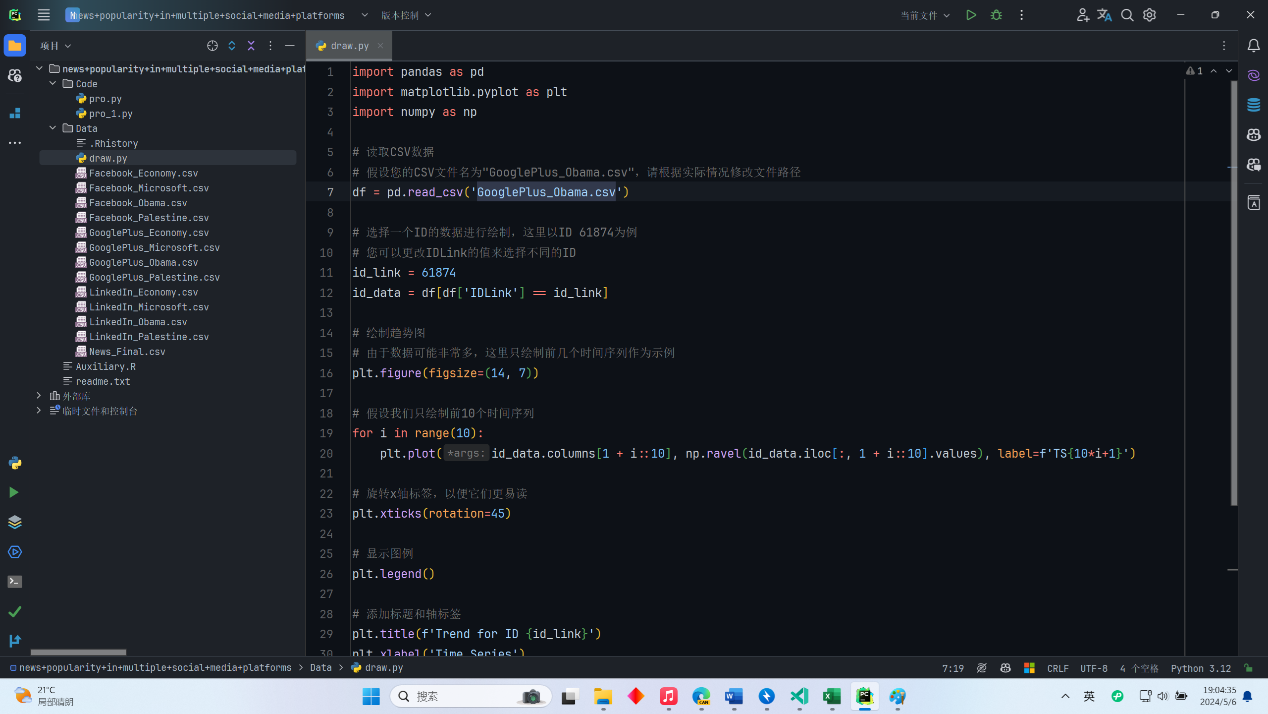
1. 数据清洗：包括检查缺失数据、错误数据，以及数据格式的统一和归一化。

一个简单的执行上述操作的Python脚本：

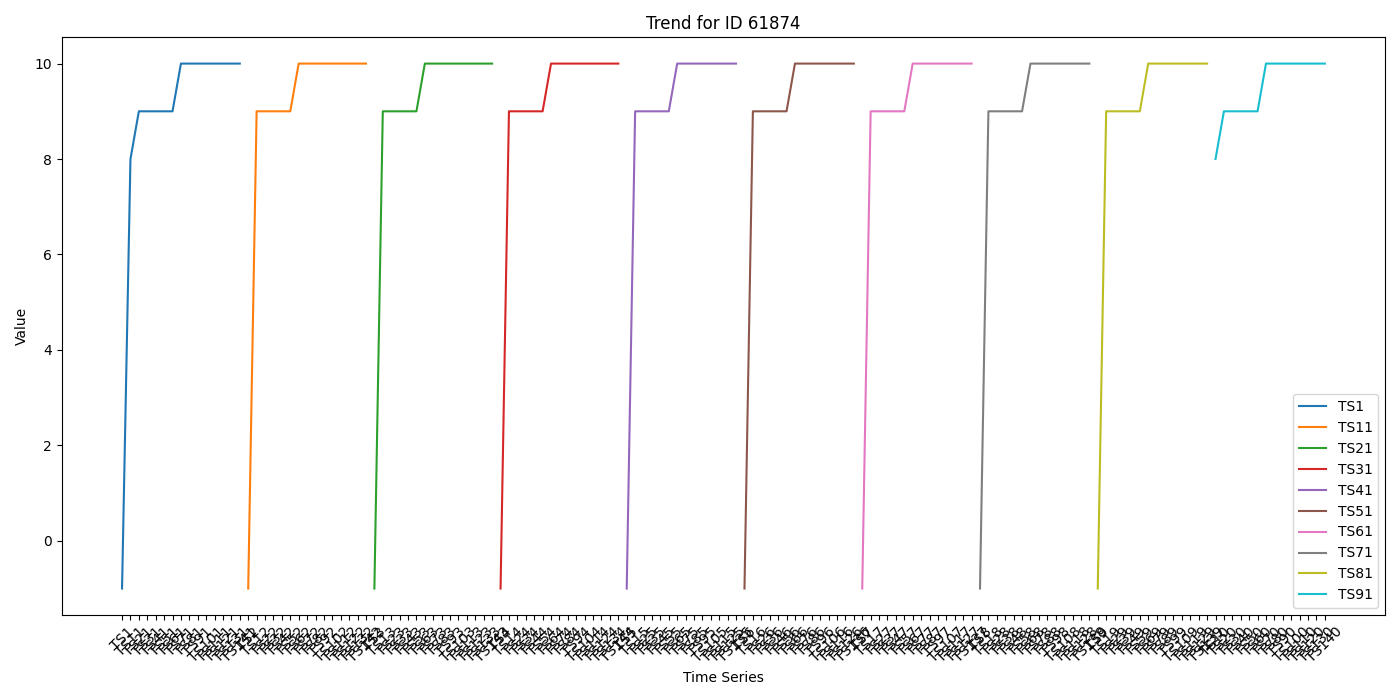


1. 研究方向：确定研究目标（如分类、聚类、回归、预测），选择方法，分配数据，确立模型，调整参数，训练模型，评价模型有效性，以及应用模型。
2. 趋势图的绘制：

代码如下：



图形绘制如下：



1. 实验结果

将数据集储存到了数据库中，并且绘制了趋势图。

1. 实验分析

了解了如何处理存储数据集，使用Python对数据进行图绘。

1. 实验结论

成功将数据集储存到了数据库中，并且绘制了趋势图。