

## 第五周

处理 Excel 文件以管理和分析数据，检查哪些同学没有提交用户名，对学生提交的数据进行整理和分析。用到的包括 openpyxl、pandas 和 polars。使用这些工具读取和处理 Excel 文件，并结合具体代码展示如何检查未提交用户名的同学。

### openpyxl

openpyxl 是一个专门用于读写 Excel 文件（.xlsx 和 .xlsm 格式）的 Python 库。它直接操作 Excel 文件的底层结构，适合处理不规范的 Excel 文件或需要精细控制文件格式的场景。可以通过 openpyxl 直接读取单元格内容、修改文件或创建新文件。

### pandas

pandas 是一个功能强大的数据分析库，提供了 read\_excel 函数来读取 Excel 文件，并将其转换为 DataFrame 格式，方便进行数据分析和处理。pandas 支持多种文件格式，包括 Excel (.xlsx) 和 Stata (.dta)。它依赖于底层的 openpyxl 或 xlrd 等库来解析 Excel 文件。

### polars

polars 是一个高性能的数据处理库，类似于 pandas，但在处理大数据时速度更快。它也提供了 read\_excel 函数，支持多种引擎（如 openpyxl 和 calamine）来读取 Excel 文件，并将数据转换为 DataFrame 格式。

## 代码实现

### 1. 使用 openpyxl 检查未提交用户名的同学

openpyxl 适合直接操作 Excel 文件的单元格，尤其在文件格式不规范时非常有用。以下代码展示了如何读取 Excel 文件并检查 Username 列为空的行。

```
environment.yml
1  import openpyxl
2
3  # 加载 Excel 文件
4  workbook = openpyxl.load_workbook('students.xlsx')
5  sheet = workbook.active # 获取活动工作表
6
7  # 存储未提交用户名的学号
8  missing_usernames = []
9
10 # 遍历行（假设第一行是标题行）
11 for row in sheet.iter_rows(min_row=2, values_only=True):
12     student_id, username = row[0], row[1]
13     if not username: # 检查用户名是否为空
14         missing_usernames.append(student_id)
15
16 # 输出结果
17 if missing_usernames:
18     print("以下同学未提交用户名：")
19     for student_id in missing_usernames:
20         print(f"学号: {student_id}")
21 else:
22     print("所有同学均已提交用户名！")
23
24 # 关闭工作簿
25 workbook.close()
26
```

## 2. 使用 pandas 检查未提交用户名的同学

pandas 的 read\_excel 函数可以将 Excel 文件直接读取为 DataFrame，适合快速进行数据筛选和分析。以下代码展示如何实现相同的功能。

```

main.py > ...
1  import pandas as pd
2
3  # 读取 Excel 文件
4  df = pd.read_excel('students.xlsx', engine='openpyxl')
5
6  # 筛选用户名为空的行
7  missing_usernames = df[df['Username'].isna()]
8
9  # 输出结果
10 if not missing_usernames.empty:
11     print("以下同学未提交用户名: ")
12     for index, row in missing_usernames.iterrows():
13         print(f"学号: {row['Student_ID']}")
14 else:
15     print("所有同学均已提交用户名! ")
16

```

### 3. 使用 polars 检查未提交用户名的同学

polars 是一个高性能替代方案, 适合处理大型数据集。以下代码展示了如何使用 polars 读取 Excel 文件并检查未提交用户名的同学。

```

main.py > ...
1  import polars as pl
2
3  # 读取 Excel 文件
4  df = pl.read_excel('students.xlsx', engine='openpyxl')
5
6  # 筛选用户名为空的行
7  missing_usernames = df.filter(pl.col('Username').is_null())
8
9  # 输出结果
10 if missing_usernames.height > 0:
11     print("以下同学未提交用户名: ")
12     for row in missing_usernames.iter_rows(named=True):
13         print(f"学号: {row['Student_ID']}")
14 else:
15     print("所有同学均已提交用户名! ")
16

```

通过以上代码, 我们实现了检查未提交用户名的功能。openpyxl 适合处理不规范的 Excel 文件, pandas 适合快速数据分析, polars 则在处理大型数据集时更高效。根据实际

需求选择合适的工具可以显著提高工作效率。