BÁO CÁO ASSIGNMENT 1

Bài 1:
Để thực hiện yêu cầu của bài, em sử dụng các thư viện:

```
from bs4 import BeautifulSoup
from bs4 import Comment
import pandas as pd
import requests
import os
```

Trong đó:

+BeautifulSoup: được sử dụng để phân tích cú pháp HTML và trích xuất dữ liệu từ các thẻ HTML
+pandas: dùng để xử lý dữ liệu, cho phép lưu trữ và phân tích dữ liệu
+requests: dùng để thực hiện các yêu cầu HTTP và nhận nội dung từ trang web
+os: dùng để tương tác với hệ điều hành.

Các chương trình con:

```
def crawler(url, stat_id, table_id):
```

-Chương trình dùng để nhận url của trang web và id để thu thập dữ liệu.
-Các bước chương trình thực thi:

+Đầu tiên, chương trình nhận URL của trang web, ID thống kê và số hiệu bảng để thu thập dữ liệu
+Sử dụng BeautifulSoup để phân tích HTML và tìm kiếm bảng thống kê với id theo stat_id được nhận
+Trích xuất dữ liệu từ các hàng tr(table row) của bảng và lưu vào 1 dictionary(res)
+Chuyển đổi dictionary thành DataFrame của pandas và lưu thành tệp csv.

```
def cleaner():
```

-Chương trình dùng để hợp nhất các bảng dữ liệu thu được thành 1 bảng result.csv
-Các bước chương trình thực thi:

+Đầu tiên, đọc bảng table1.csv làm bảng cơ sở
+Duyệt các bảng từ table3.csv đến table 10.csv, loại bỏ các cột trùng lặp và hợp nhất với bảng cơ sở
+Kết hợp dữ liệu với table2.csv dựa trên các cột chung
+Lọc ra các cầu thủ có số phút chơi từ 90 phút trở lên
+Sắp xếp kết quả theo tên cầu thủ và tuổi, lưu lại vào tệp result.csv.

```
def delete_tables():
```

-Xóa các tệp csv từ table1 đến table 10 để tránh gây trùng lặp cho lần chạy sau và để thư mục gọn gàng.

```
def delete_result():
```

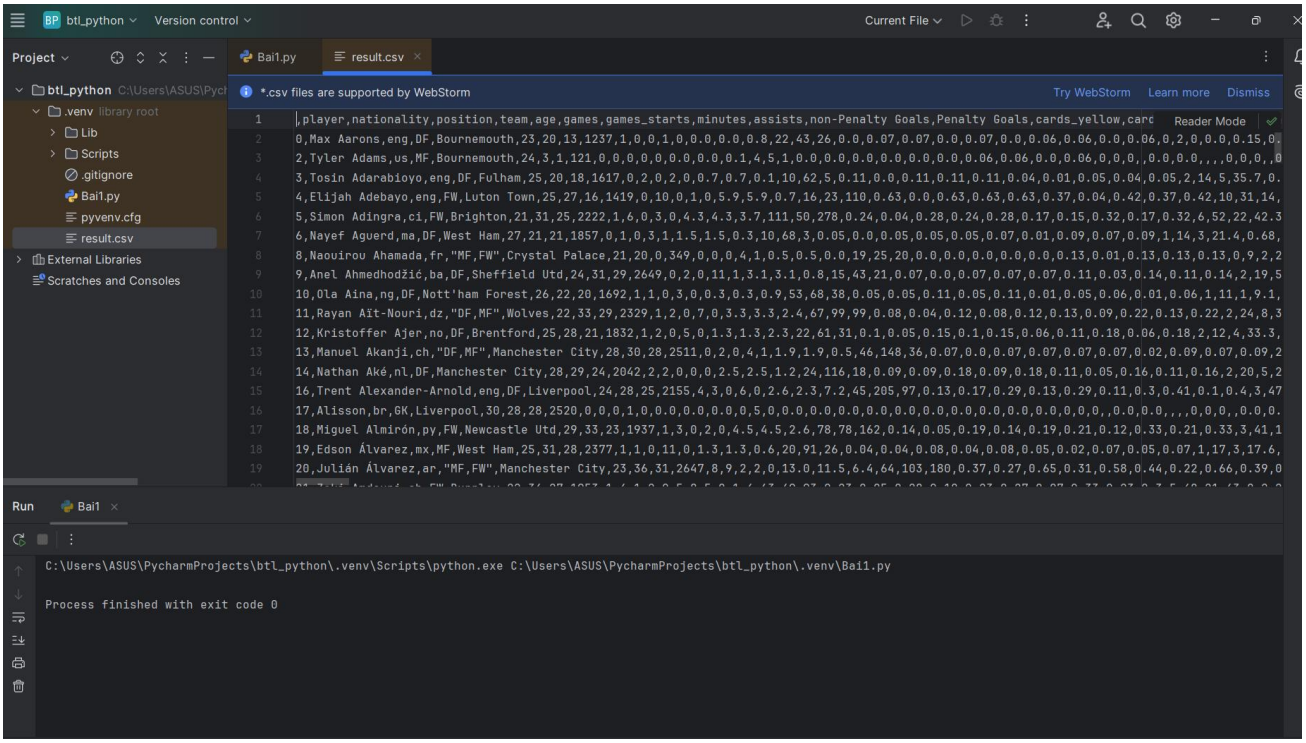
-Xóa tệp result.csv để tránh gây trùng lặp cho lần chạy sau.

```
if __name__ == '__main__':
```

-Trong hàm main:

+Đầu tiên, gọi delete_result để xóa tệp result.csv nếu có tồn tại.
+Tiếp theo, khai báo mảng các URL và ID cần thu thập dữ liệu
+Thực hiện request đến trang web để lấy thông tin cầu thủ và trích xuất dữ liệu vào table1.csv
+Gọi hàm crawler() cho từng URL trong mảng đã khai báo bên trên và lưu dữ liệu lần lượt vào các file từ table2.csv cho đến table10.csv
+Cuối cùng, gọi hàm cleaner() để xử lý và tổng hợp dữ liệu, và delete_tables() để xóa các bảng table.csv đã tạo.

-Ta nhận được được file result.csv khi chạy chương trình như trong hình:


btl_python
Version control
Current File
Project
btl_python C:\Users\ASUS\Pycl
.venv library root
Lib
Scripts
.gitignore
Bai1.py
pyvenv.cfg
result.csv
External Libraries
Scratches and Consoles
Bai1.py
result.csv
*.csv files are supported by WebStorm
Try WebStorm
Learn more
Dismiss
Reader Mode
,player,nationality,position,team,age,games,games_starts,minutes,assists,non-Penalty Goals,Penalty Goals,cards_yellow,card
0,Max Aarons,eng,DF,Bournemouth,23,20,13,1237,1,0,0,1,0,0.0,0.0,0.8,22,43,26,0.0,0.07,0.07,0.0,0.07,0.0,0.06,0.06,0.0,0.06,0,2,0,0.0,0.15,0.
2,Tyler Adams,us,MF,Bournemouth,24,3,1,121,0,0,0,0,0,0.0,0.0,0.1,4,5,1,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.06,0.06,0.0,0.06,0,0,0,,0.0,0.0,,,,0,0,0,,0
3,Tosin Adarabioyo,eng,DF,Fulham,25,20,18,1617,0,2,0,2,0,0.7,0.7,0.1,10,62,5,0.11,0.0,0.11,0.11,0.11,0.04,0.01,0.05,0.04,0.05,2,14,5,35.7,0.
4,Elijah Adebayo,eng,FW,Luton Town,25,27,16,1419,0,10,0,1,0,5.9,5.9,0.7,16,23,110,0.63,0.0,0.63,0.63,0.63,0.37,0.04,0.42,0.37,0.42,10,31,14,
5,Simon Adingra,ci,FW,Brighton,21,31,25,2222,1,6,0,3,0,4.3,4.3,3.7,111,50,278,0.24,0.04,0.28,0.24,0.28,0.17,0.15,0.32,0.17,0.32,6,52,22,42.3
6,Nayef Aguerd,ma,DF,West Ham,27,21,21,1857,0,1,0,3,1,1.5,1.5,0.3,10,68,3,0.05,0.0,0.05,0.05,0.05,0.07,0.01,0.09,0.07,0.09,1,14,3,21.4,0.68,
8,Naouirou Ahamada,fr,"MF,FW",Crystal Palace,21,20,0,349,0,0,0,4,1,0.5,0.5,0.0,19,25,20,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.13,0.01,0.13,0.13,0.13,0,9,2,2
9,Anel Ahmedhodžić,ba,DF,Sheffield Utd,24,31,29,2649,0,2,0,11,1,3.1,3.1,0.8,15,43,21,0.07,0.0,0.07,0.07,0.07,0.11,0.03,0.14,0.11,0.14,2,19,5
10,Ola Aina,ng,DF,Nott'ham Forest,26,22,20,1692,1,1,0,3,0,0.3,0.3,0.9,53,68,38,0.05,0.05,0.11,0.05,0.11,0.01,0.05,0.06,0.01,0.06,1,11,1,9.1,
11,Rayan Aït-Nouri,dz,"DF,MF",Wolves,22,33,29,2329,1,2,0,7,0,3.3,3.3,2.4,67,99,99,0.08,0.04,0.12,0.08,0.12,0.13,0.09,0.22,0.13,0.22,2,24,8,3
12,Kristoffer Ajer,no,DF,Brentford,25,28,21,1832,1,2,0,5,0,1.3,1.3,2.3,22,61,31,0.1,0.05,0.15,0.1,0.15,0.06,0.11,0.18,0.06,0.18,2,12,4,33.3,
13,Manuel Akanji,ch,"DF,MF",Manchester City,28,30,28,2511,0,2,0,4,1,1.9,1.9,0.5,46,148,36,0.07,0.0,0.07,0.07,0.07,0.07,0.02,0.09,0.07,0.09,2
14,Nathan Aké,nl,DF,Manchester City,28,29,24,2042,2,2,0,0,0,2.5,2.5,1.2,24,116,18,0.09,0.09,0.18,0.09,0.18,0.11,0.05,0.16,0.11,0.16,2,20,5,2
16,Trent Alexander-Arnold,eng,DF,Liverpool,24,28,25,2155,4,3,0,6,0,2.6,2.3,7.2,45,205,97,0.13,0.17,0.29,0.13,0.29,0.11,0.3,0.41,0.1,0.4,3,47
17,Alisson,br,GK,Liverpool,30,28,28,2520,0,0,0,1,0,0.0,0.0,0.0,0.0,0.0,5,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,,,0,0,0,,0.0,0.
18,Miguel Almirón,py,FW,Newcastle Utd,29,33,23,1937,1,3,0,2,0,4.5,4.5,2.6,78,78,162,0.14,0.05,0.19,0.14,0.19,0.21,0.12,0.33,0.21,0.33,3,41,1
19,Edson Álvarez,mx,MF,West Ham,25,31,28,2377,1,1,0,11,0,1.3,1.3,0.6,20,91,26,0.04,0.04,0.08,0.04,0.08,0.05,0.02,0.07,0.05,0.07,1,17,3,17.6,
20,Julián Álvarez,ar,"MF,FW",Manchester City,23,36,31,2647,8,9,2,2,0,13.0,11.5,6.4,64,103,180,0.37,0.27,0.65,0.31,0.58,0.44,0.22,0.66,0.39,0
Run
Bai1
C:\Users\ASUS\PycharmProjects\btl_python\.venv\Scripts\python.exe C:\Users\ASUS\PycharmProjects\btl_python\.venv\Bai1.py
Process finished with exit code 0