**HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**

**KHOA CÔNG NGHỆ THÔNG TIN**

**-----🙠🕮🙢-----**

**BÁO CÁO BÀI TẬP LỚN**

**Lập trình Python**

**Sinh viên thực hiện:**

**Họ và tên :**

**Mã SV:**

**Hà Nội, ngày 3/11/2023**

* Câu 1:
* Sử dụng thư viện url và beautifulsoup gửi yêu cầu truy cập đến web
* Sử dụng hàm find để tìm kiếm các đường link của các đội bóng trong tag tbody
* Ghi dữ liệu tên của các đội và đường link của các đội đó vào list teams \_data và file teams\_data.csv để tiện quan sát (Hình 1.2)

A computer screen shot of a program

Description automatically generated(Hình 1.1)

A screen shot of a computer

Description automatically generated

(Hình 1.2)

* Tiếp theo ta chia dữ liệu thành nhiều list để giảm độ khó trong việc xử lý (Hình 1.3), ta chia dữ liệu thanhf 10 list khác nhau

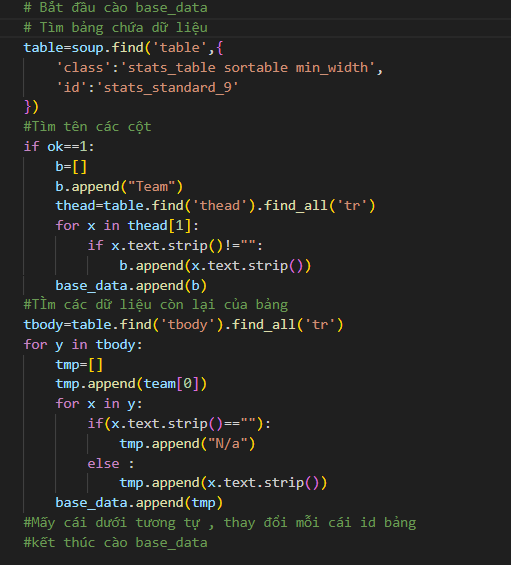
A screen shot of a computer program

Description automatically generated

(Hình 1.3)

* Tiếp theo ta duyệt list Data\_teams và tiến hành crawdata từng link đội bóng 1

+ Sau đây là minh họa code cào 1 list dữ liệu (cụ thể là list base\_data, những list khác cũng tương tự , chỉ thay đổi id của bảng) (Hình 1.4)



(HÌnh 1.4)

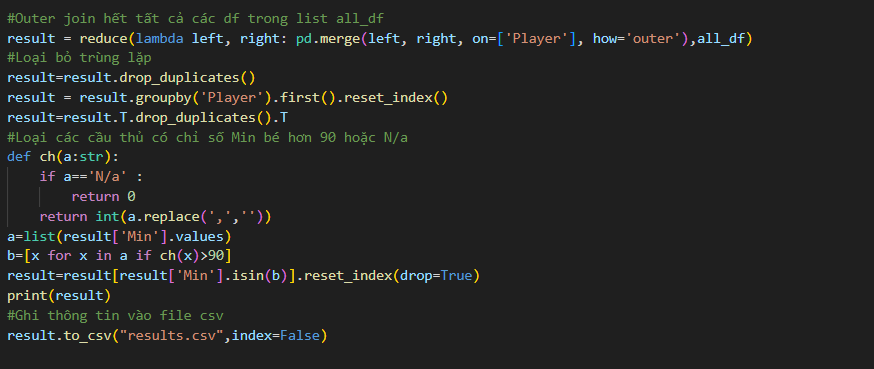
* Sau khi chạy xong , tất cả các dữ liệu cần thiết đã được them vào các list . TIếp theo ta chuyển các list thành các dataframe ,khử các cột không cần thiết , và đưa nó vào list chứa các dataframe để tiếp tục bước tiếp theo (Hình 1.5)

A screen shot of a computer program

Description automatically generated

(Hình 1.5)

* Cuối cùng ta outer join tất cả các dataframe ở trong cùng 1 list chứa tất cả các dataframe (df\_all), tiến hành lọc các cầu thủ có số phút chơi trên 90p và them vào file resuilt.csv (Hình 1.6)



(Hình 1.6)

* Câu 2 : Từ những thông tin từ file results.cvs , thực hiện các yêu cầu
* TÌm top 3 cầu thủ có điểm cao nhất và thấp nhất ở mỗi chỉ số (Hình 2.1)

+ Ý tưởng :

\_ Đầu tiên ta lọc ra các column có kiểu số

\_ TIếp đó tìm ra 3 cầu thủ có chỉ số cao nhất và thấp nhất ứng với tùng column và đưa vào list\_column

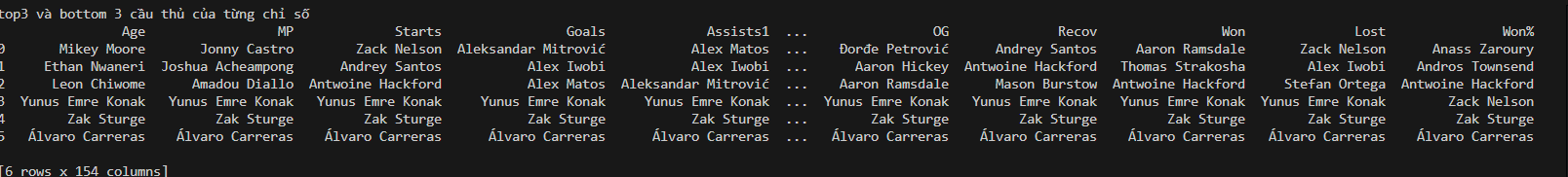
\_tạo 1 dataframe được tạo nên từ các list column

A computer screen with colorful text

Description automatically generated

(Hình 2.1)

+ Kết quả sau khi chạy code :



* Tìm trung vị của mỗi chỉ số. Tìm trung bình và độ lệch chuẩn của mỗi chỉ số cho các cầu thủ trong toàn giải và của mỗi đội

+ Ý tưởng :

\_ Đối với toàn bộ các cầu thủ thì ta có 3 bảng tìm về median,mean,std đối với mỗi chỉ số, sau khi tìm được thì t chỉ cần nối column của chúng lại tạo thành 1 dataframe1 mới (Hình 2.2)

A screen shot of a computer program

Description automatically generated

(Hình 2.2)

\_ Còn đối vs đội cũng vậy , ta cũng có 3 bảng về median,mean,std đối vs mỗi chỉ số được group theo “team” , từ đó ra nối column chúng lại tạo thành 1 dataframe2 mới (Hình 2.3)

A screen shot of a computer code

Description automatically generated

(Hình 2.3)

\_Cuối cùng ta nối df1 và df2 theo hang bằng hàm concat và them dữ liệu vào file result2.csv (Hình 2.4)

A black background with white text

Description automatically generated

(Hình 2.4)

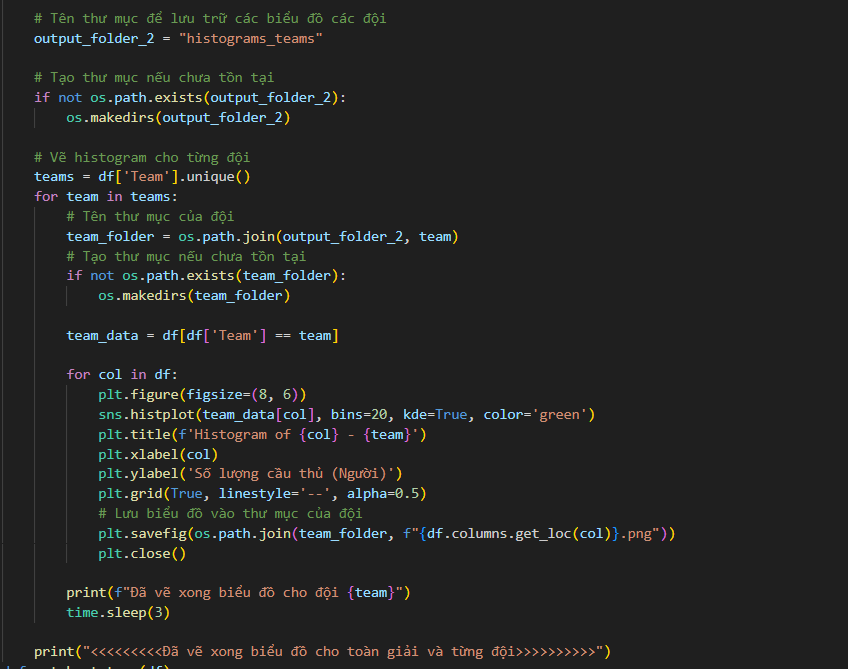
* Vẽ historgram phân bố của mỗi chỉ số của các cầu thủ trong toàn giải và mỗi đội (Hình 2.5)

+ Dầu tiên ta vẽ biểu đồ cho toàn giải (Hình 2.5.1) , sau đó ta tạo them thư mục để lưu trữ từng biểu đồ một (HÌnh 2.5.2)

A screen shot of a computer program

Description automatically generated

(Hình 2.5.1)



(HÌnh 2.5.2)

* Tìm các đội bóng có chỉ số cao nhất ở mỗi chỉ số

+ Ý tưởng t duyêt qua từng chỉ số của df1 (được tạo ra từ vc grouby ”team” với phương thức mean) và lưu lại vào trong list tên đội , chỉ số, giá trị max của chỉ số đó

A computer screen shot of text

Description automatically generated

(Hình 2.6)

A screenshot of a computer screen

Description automatically generated (HÌnh 2.7 : bảng df\_re sults)

* Từ hình 2.6 , ta thấy rằng việc tiếp theo là tìm tần xuất điểm của các đội bằng cách dung hàm couter để đếm tần xuất và sort để tìm giá trị lớn nhất

+ Kết quả của thu được sau khi chạy code : 

* Manchester City là đội có phong độ tốt nhất trong giải ngoại hạng Anh
* Trong hàm main , ta vt lại các hàm theo từng truy vấn (Hinh 2.8)

A screen shot of a computer program

Description automatically generated

(Hình 2.8)

* Câu 3:
* Sử dụng thuật toán K-means để phân loại các cầu thủ thành các nhóm có chỉ số giống nhau.
* Sử dụng thuật toán PCA, giảm số chiều dữ liệu xuống 2 chiều, vẽ hình phân cụm các điểm dữ liệu trên mặt 2D.

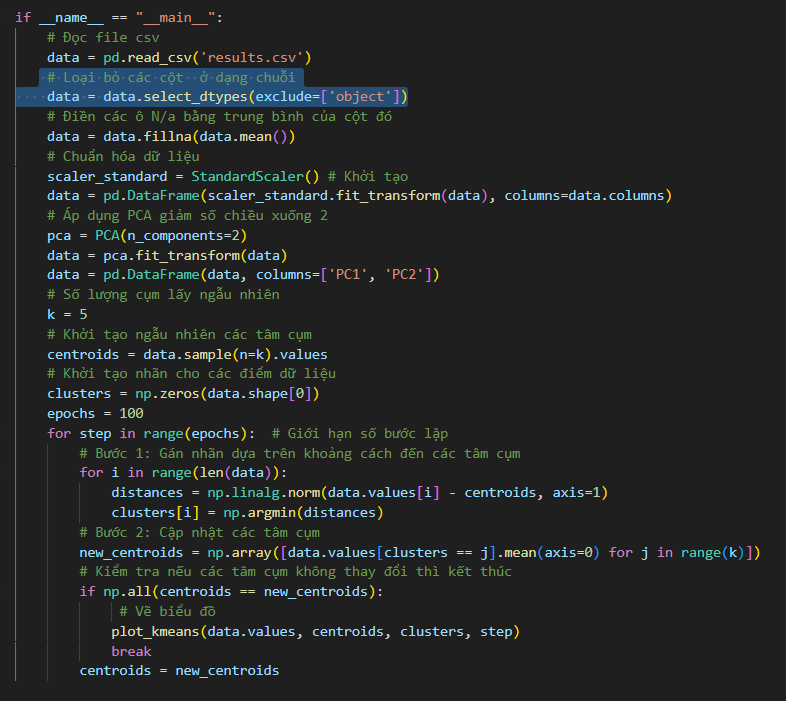
+ Hàm vẽ biểu đồ (HÌnh 3.1)

A computer screen with colorful text

Description automatically generated

(Hình 3.1)

+ Trong hàm main , ta khởi tạo các tham số cần thiệt cho hàm vẽ biểu đồ (Hình 3.1.2)



(Hình 3.1.2)

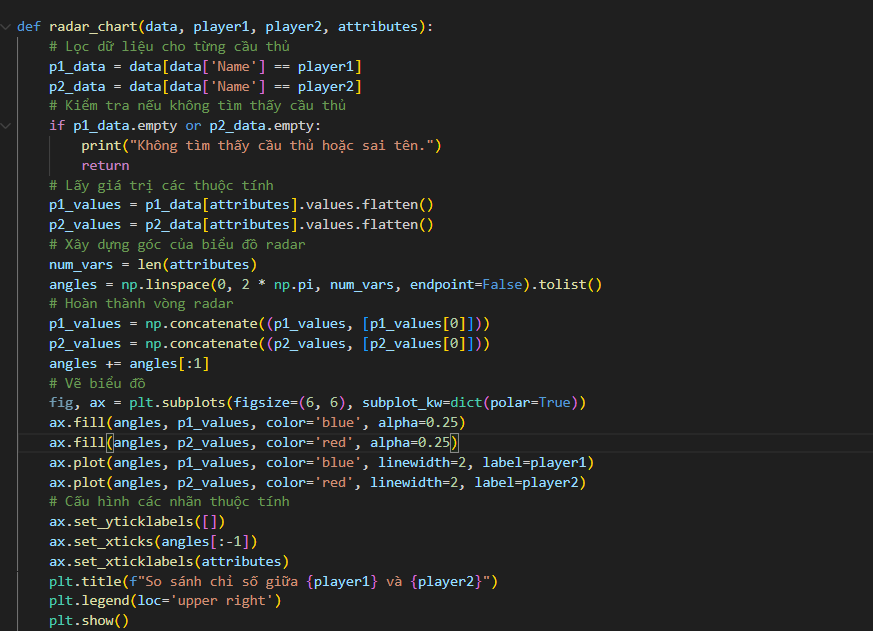
+ Kết quả sau khi chạy code (HÌnh 3.1.3):  
A diagram of a football player

Description automatically generated

(Hình 3.1.3)

* Viết chương trình python vẽ biểu đồ rada (radar chart) so sánh cầu thủ :

+ hàm vẽ biểu đồ (Hình 3.2.1)



(Hình 3.2.1)

+ Hàm main (Hình 3.2.2)

A black screen with colorful text

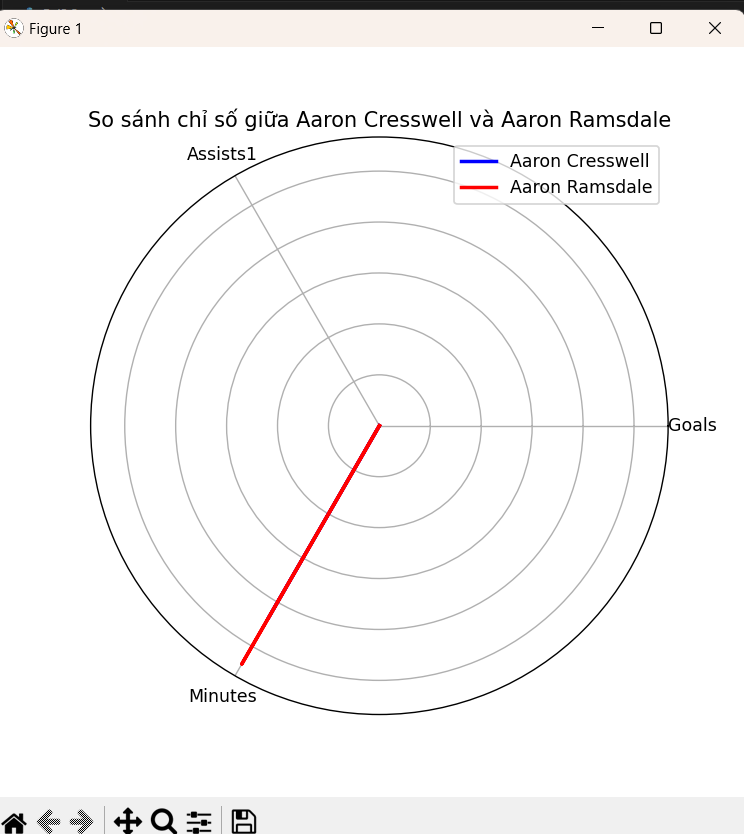
Description automatically generated

(Hình 3.2.2)

* Sau khi chạy lệnh ta dung lệnh <Đường dần file> <tên file> --p1 "Tên Cầu Thủ 1" --p2 "Tên Cầu Thủ 2" --Attribute "ChỉSố1,ChỉSố2,...,ChỉSốN"

\_ Cụ thể sau khi dung lệnh :

C:/Users/luong/AppData/Local/Programs/Python/Python312/python.exe Bai3,2.py --p1 "Aaron Cresswell" --p2 "Aaron Ramsdale" --Attribute "Goals, Assists1, Minutes" (đã được bôi xanh dòng cuối cùng) trên terminal (Hình 3.3.3)



(Hình 3.3.3)