

**BỘ THÔNG TIN VÀ TRUYỀN THÔNG**  
**HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**

---



**BÀI TẬP LỚN**  
**LẬP TRÌNH VỚI PYTHON**

<b>Giảng viên hướng dẫn</b>	<b>: Kim Ngọc Bách</b>
<b>Nhóm</b>	<b>: 11</b>
<b>Hệ</b>	<b>: Đại học chính quy</b>
<b>Sinh viên</b>	<b>: Lê Văn Đô – B22DCCN214</b>

**Hà Nội – 2024**

## MỤC LỤC

<b>Bài 1.1</b> .....	2
<b>Bài 2.1</b> .....	3
<b>Bài 2.2</b> .....	4
<b>Bài 2.3</b> .....	5
<b>Bài 2.4</b> .....	6
<b>Bài 3.1</b> .....	7
<b>Bài 3.2</b> .....	8
<b>Bài 3.3</b> .....	9
<b>Bài 3.4</b> .....	10

# Bài 1

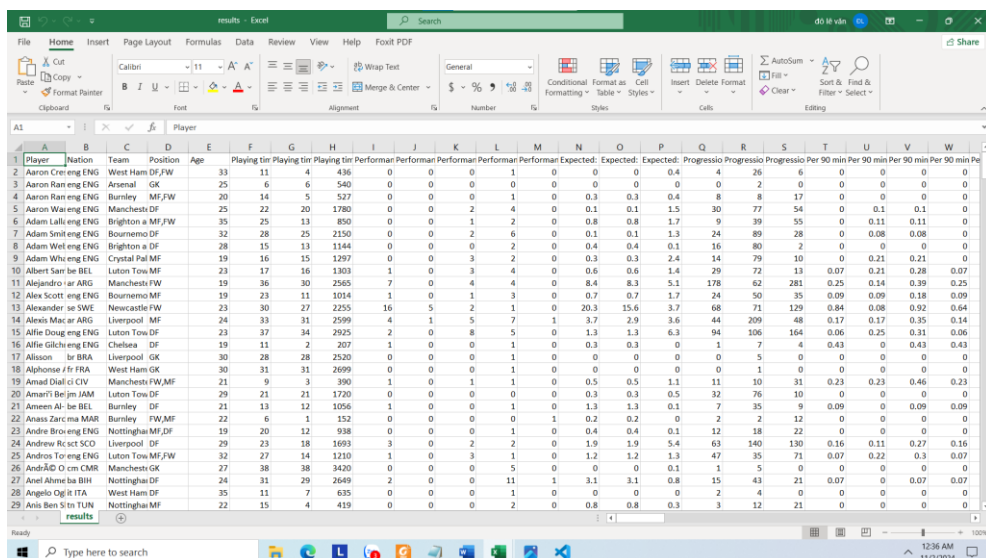
## Công nghệ:

- Requests: Dùng để lấy dữ liệu từ trang web.
- BeautifulSoup4: Dùng để phân tích cấu trúc HTML của trang web.
- pandas: Dùng để xử lý và phân tích dữ liệu, tạo bảng dữ liệu.

## Thuật toán:

- Bước 1: Gửi yêu cầu HTTP đến danh sách URL để lấy nội dung trang web.
- Sử dụng requests.get(url) để tải HTML từ các trang web chứa dữ liệu thống kê cầu thủ bóng đá. Mỗi URL trong danh sách sẽ được xử lý tuần tự.
- Bước 2: Phân tích cú pháp HTML để tìm và trích xuất dữ liệu từ các bảng trên trang web.
- Dùng BeautifulSoup(html\_content, 'html.parser') để phân tích nội dung HTML. Sau đó, tìm bảng dữ liệu cầu thủ bằng cách dùng find() hoặc find\_all() để tìm các thẻ <table>, và trích xuất từng dòng cầu thủ với thẻ <tr>. Mỗi cột trong bảng chứa các chỉ số như "Tuổi", "Fls", "Aerial Duels", v.v., sẽ được lưu vào từ điển.
- Bước 3: Tạo bảng dữ liệu (DataFrame) từ danh sách các từ điển chứa dữ liệu cầu thủ, sắp xếp và xuất ra file CSV.

Dùng pandas.DataFrame() để chuyển danh sách các cầu thủ thành DataFrame. Dữ liệu sau đó được sắp xếp theo tên cầu thủ và tuổi. Những giá trị bị thiếu (NaN) được xử lý bằng fillna('N/a'). Cuối cùng, bảng dữ liệu được lưu vào file results.csv bằng to\_csv().



Player	Nation	Team	Position	Age	Playing time	Goals	Assists	Expected Goals	Expected Assists	Progression	Progression Per 90 min	Per 90 min	Per 90 min	Per 90 min	Per 90 min	Per 90 min	Per 90 min	Per 90 min	Per 90 min
Aaron Cresswell	ENG	West Ham	DF	33	11	4	436	0	0	0	0	0	0	0	0	0	0	0	0
Aaron Ramsdale	ENG	Arsenal	GK	25	6	6	540	0	0	0	0	0	0	0	0	0	0	0	0
Aaron Ramsdale	ENG	Burnley	MF/FW	20	14	5	527	0	0	0	0	0	0	0	0	0	0	0	0
Aaron Wan	ENG	Mancheste	DF	25	22	20	1780	0	0	2	4	0	0.1	0.1	1.5	30	77	54	0
Adam Lall	ENG	Brighton	MF/FW	35	25	13	850	0	0	1	2	0	0.8	0.8	1.7	9	39	55	0
Adam Smith	ENG	Bournemouth	DF	32	28	25	2150	0	0	2	6	0	0.1	0.1	1.3	24	89	28	0
Adam West	ENG	Brighton	DF	28	15	13	1144	0	0	0	2	0	0.4	0.4	0.1	16	80	2	0
Adam Wh	ENG	Crystal Pal	MF	19	16	15	1297	0	0	3	2	0	0.3	0.3	2.4	14	79	10	0
Albert San	BEL	Luton Tow	MF	23	17	16	1303	1	0	3	4	0	0.6	0.6	1.4	29	72	13	0
Alexandro	ARG	Mancheste	FW	19	36	30	2565	7	0	4	4	0	8.4	8.3	5.1	178	62	281	0.25
Alex Scott	ENG	Bournemouth	MF	19	23	11	1014	1	0	1	3	0	0.7	0.7	1.7	24	50	35	0.09
Alexander	SWE	Newcastle	FW	23	30	27	2255	16	5	2	1	0	20.3	15.6	3.7	68	71	129	0.84
Alexis Mac	ARG	Liverpool	MF	24	33	31	2599	4	1	5	7	1	3.7	2.9	3.6	44	209	48	0.17
Alfie Donag	ENG	Luton Tow	DF	23	37	34	2925	2	0	8	5	0	1.3	1.3	6.3	94	106	164	0.06
Alfie Gilch	ENG	Chelsea	DF	19	11	2	207	1	0	0	1	0	0.3	0.3	0	1	7	4	0.43
Alisson	BRA	Liverpool	GK	30	28	28	2520	0	0	0	1	0	0	0	0	0	5	0	0
Alphonse	FRA	West Ham	GK	30	31	31	2699	0	0	0	1	0	0	0	0	0	1	0	0
Amad Dial	GUI	Mancheste	FW/MF	21	9	3	390	1	0	1	1	0	0.5	0.5	1.1	11	10	31	0.23
Amarin Bel	JAM	Luton Tow	DF	29	21	21	1720	0	0	0	0	0	0.3	0.3	0.5	32	76	10	0
Ameen Al	BEL	Burnley	DF	21	13	12	1056	1	0	0	1	0	1.3	1.3	0.1	7	35	9	0.09
Anass Zar	MAR	Burnley	FW/MF	22	6	1	152	0	0	0	0	0	0.2	0.2	0	2	2	12	0
Andre Bro	ENG	Nottingham	MF	19	20	12	938	0	0	0	1	0	0.4	0.4	0.1	12	22	0	0
Andrew R	SCO	Liverpool	DF	29	23	18	1693	3	0	2	2	0	1.9	1.9	5.4	63	140	130	0.16
Andros Tor	ENG	Luton Tow	MF/FW	32	27	14	1210	1	0	3	1	0	1.2	1.2	1.3	47	35	71	0.07
Andre-O	CMR	Mancheste	GK	27	38	38	3420	0	0	0	5	0	0	0	0	0.1	1	5	0
Ariel Ahme	BH	Nottingham	DF	24	31	29	2649	2	0	0	11	1	3.1	3.1	0.8	15	43	21	0.07
Angelo Og	ITA	West Ham	DF	35	11	7	635	0	0	0	1	0	0	0	0	2	4	0	0
Aris Ben	TUN	Nottingham	MF	22	15	4	419	0	0	0	2	0	0.8	0.8	0.3	3	12	21	0

Tuy nhiên vì gửi yêu cầu đến nhiều URL một nên có thể xuất hiện lỗi không thể lấy được dữ liệu

## Bài 2

### Phần 1:

Công nghệ:

Pandas: Xử lý và phân tích dữ liệu dạng bảng.

Thuật toán:

- Đọc Dữ Liệu
- Xử Lý Dữ Liệu
- Lặp Qua Các Cột
- Sắp Xếp và Lọc Dữ Liệu
- Hiện Thị Kết Quả

Ta được bảng sau( hình ảnh một phần nhỏ của bảng):

3 cầu thủ có Expected: KP cao nhất		
	Player	Expected: KP
50	Bruno Fernandes	114
291	Pascal Groß	103
245	Martin Ødegaard	102
3 cầu thủ có Expected: KP thấp nhất		
	Player	Expected: KP
1	Aaron Ramsdale	0
368	Łukasz Fabiański	0
360	Willy Kambwala	0

3 cầu thủ có Aerial Duels: Won cao nhất		
	Player	Aerial Duels: Won
352	Virgil van Dijk	140
60	Carlton Morris	133
86	Dara O'Shea	127
3 cầu thủ có Aerial Duels: Won thấp nhất		
	Player	Aerial Duels: Won
1	Aaron Ramsdale	0
45	Benson Manuel	0
52	Bryan Gil	0

## Phần 2:

Công nghệ:

## Pandas: Xử lý và phân tích dữ liệu dạng bảng.

Thuật toán:

- Đọc dữ liệu: Tải dữ liệu từ results.csv bằng pd.read\_csv().
- Tiền xử lý: Chuyển đổi các cột số thành kiểu dữ liệu số, thay thế giá trị không hợp lệ bằng 0.
- Tính toán thống kê: Tính Median, Mean, Std cho toàn giải và từng đội.
- Kết hợp dữ liệu: Tạo DataFrame từ kết quả thống kê.
- Xuất dữ liệu: Lưu kết quả vào results2.csv.

Ta được file resulst2.csv như sau:

The screenshot displays the Microsoft Excel application window. The title bar indicates the file is 'results2 - Excel'. The ribbon is set to the 'Home' tab, showing options for Font, Paragraph, Styles, Cells, and Editing. The data table is visible, with columns labeled A through W. The first column (A) contains an 'Index' from 1 to 29. The second column (B) lists football clubs. Columns C through W contain numerical data, likely representing performance metrics. The status bar at the bottom shows 'Ready', a search bar, and the date '11/3/2024'.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
Index	Name	Median	Median	Median	Median	Median	Median	Median	Median	Median	Median	Median	Median	Median	Median	Median	Median	Median	Median	Median	Median	Median
1	all	23	15	1367	1	0	1	2	0	1.1	1.1	1	19	46	30	0.07	0.06	0.16	0.07	0.16	0.18	0.18
2	1 West Ham	23.5	21	1776	1	0	0	2.5	0	1.3	1.25	0.35	10	33.5	28	0.045	0	0.095	0.045	0.095	0.115	0.115
3	2 Arsenal	27	18	1649	2	0	2	2	0	1.6	1.6	1.4	17	72	37	0.12	0.17	0.25	0.12	0.25	0.24	0.24
4	3 Burnley	16	14	1213	1	0	1	2	0	0.55	0.55	0.8	15	34.5	25	0.035	0.055	0.09	0.035	0.09	0.125	0.125
5	4 Manchest	22	15	1351.5	1	0	1	2.5	0	0.85	0.85	0.9	14	52	32.5	0.07	0.07	0.155	0.07	0.155	0.19	0.19
6	5 Brighton a	20	15	1344.5	0	0	1	3	0	1	1	0.9	21	43	41	0	0.04	0.215	0	0.215	0.23	0.23
7	6 Bournemo	25.5	13	1317.5	0.5	0	1	3	0	1.15	1.15	1.1	22	43.5	27.5	0.015	0.08	0.15	0.015	0.15	0.15	0.15
8	7 Crystal Pal	22.5	17.5	1587.5	0	0	1	2.5	0	0.6	0.6	1	14.5	40.5	17.5	0	0.045	0.18	0	0.08	0.125	0.125
9	8 Luton Tow	23	16	1304	1	0	0	3	0	1	1	0.7	15	37	21	0.07	0	0.18	0.07	0.18	0.14	0.14
10	9 Newcastle	21	14.5	1296.5	1.5	0	1	2	0	1.35	1.35	1.05	21.5	52.5	57	0.135	0.07	0.21	0.135	0.21	0.295	0.295
11	10 Liverpool	28	17	1671	2	0	2	2	0	2.3	2.15	1.65	41	91.5	80.5	0.125	0.1	0.24	0.125	0.24	0.25	0.24
12	11 Chelsea	23	18	1576	1	0	1	3	0	1.1	1.1	1.2	22	46	35	0.09	0.08	0.2	0.08	0.16	0.27	0.27
13	12 Nottingha	14.5	11	940.5	0	0	0	2	0	0.45	0.4	0.45	7	16.5	18	0	0	0	0	0	0.11	0.105
14	13 Tottenharr	27.5	15.5	1401	1.5	0	1	3	0	1.65	1.65	0.85	23	71.5	33	0.085	0.095	0.185	0.085	0.185	0.24	0.24
15	14 Manchest	29	24	2042	2	0	1	2	0	1.9	1.9	1.4	37	100	55	0.15	0.09	0.22	0.15	0.22	0.19	0.19
16	15 Aston Villa	27	20	1652	2	0	1	4	0	1.4	1.4	1	32	55	33	0.11	0.08	0.21	0.11	0.21	0.21	0.21

### **Phần 3:**

Công nghệ:

- Pandas: Xử lý và phân tích dữ liệu dạng bảng.
- Matplotlib: Tạo đồ thị và biểu đồ trực quan.

Thuật toán:

- Đọc dữ liệu: Sử dụng `pd.read_csv()` để tải dữ liệu vào DataFrame.
- Lặp qua các cột: Xử lý từng cột bằng cách sao chép DataFrame và loại bỏ giá trị NaN.
- Vẽ biểu đồ: Sử dụng `plt.hist()` để tạo histogram và hiển thị tần suất của các giá trị trong mỗi cột.

Biểu đồ: Hình 2.3 trong Folder Hình ảnh

## Phần 4:

Công nghệ:

Pandas: Xử lý và phân tích dữ liệu dạng bảng.

Thuật toán:

- Đọc tệp CSV: Dữ liệu được tải vào DataFrame.
- Chọn cột cần tính tổng: Các cột từ chỉ số thứ 6 trở đi được chọn để tính tổng.
- Chuyển đổi kiểu dữ liệu: Các cột này được chuyển đổi thành kiểu số, với giá trị không hợp lệ được thay bằng NaN.
- Tính tổng theo đội: Dữ liệu được nhóm theo tên đội và tính tổng các chỉ số.
- Xác định đội mạnh nhất: Sử dụng idxmax() để tìm đội có chỉ số cao nhất cho từng chỉ số và in ra kết quả.

Dựa trên dữ liệu phân tích và ý kiến cá nhân đội Manchester City đang là đội có thể gọi là mạnh nhất trong mùa giải này khi đứng đầu ở rất nhiều chỉ số quan trọng

Một phần của bảng:

```
'Playing time: Matches Played': Brighton and Hove Albion
'Playing time: Starts': Burnley
'Playing time: Minutes': Crystal Palace
'Performances: Non-Penalty Goals': Manchester City
'Performances: Penalty Goals': Chelsea
'Performances: Assists': Manchester City
'Performances: Yellow Cards': Chelsea
'Performances: Red Cards': Burnley
'Expected: xG': Liverpool
'Expected: npxG': Liverpool
'Expected: xAG': Liverpool
'Progression: PrgC': Manchester City
'Progression: PrgP': Liverpool
'Progression: PrgR': Tottenham Hotspur
'Per 90 minutes: Gls': Newcastle United
'Per 90 minutes: Ast': Tottenham Hotspur
'Per 90 minutes: G+A': Newcastle United
'Per 90 minutes: G-PK': Newcastle United
'Per 90 minutes: G+A-PK': Newcastle United
'Per 90 minutes: xG + xAG': Bournemouth
'Per 90 minutes: npxG + xAG': Bournemouth
'Performances: GA': Nottingham Forest
'Performances: GA90': Bournemouth
'Performances: SoTA': Nottingham Forest
'Performances: Saves': West Ham United
'Performances: Save%': Bournemouth
'Performances: W': Arsenal
'Performances: D': Brighton and Hove Albion
'Performances: L': Nottingham Forest
'Performances: CS': Arsenal
'Performances: CS%': Manchester City
'Penalty Kicks: PKatt': West Ham United
'Penalty Kicks: PKA': West Ham United
'Penalty Kicks: PKsv': West Ham United
'Penalty Kicks: PKm': Nottingham Forest
'Penalty Kicks: PK Save%': Aston Villa
'Standard: Gls': Manchester City
```

## Bài 3:

### Phần 1:

Công nghệ:

- Pandas: Xử lý và phân tích dữ liệu dạng bảng.
- Matplotlib: Thư viện này dùng để vẽ biểu đồ. Trong đoạn mã, nó được sử dụng để hiển thị biểu đồ phân cụm.
- Scikit-learn: Thư viện Machine Learning

Thuật toán:

- Đọc dữ liệu: Dữ liệu được đọc từ file CSV và các cột chỉ số được chọn.
- Xử lý dữ liệu: Thay thế các giá trị không hợp lệ và chuyển đổi các cột chỉ số sang định dạng số.
- Chuẩn hóa dữ liệu: Sử dụng StandardScaler để chuẩn hóa dữ liệu.
- Phân cụm: Áp dụng thuật toán KMeans để phân nhóm cầu thủ thành  $n\_clusters$ .
- Gán nhãn: Gán nhãn cụm cho mỗi cầu thủ dựa trên kết quả phân cụm.
- Trực quan hóa: Hiển thị biểu đồ phân cụm để trực quan hóa kết quả.

Bảng:

Cluster 0:			
	Player	Team	Cluster
0	Aaron Cresswell	West Ham United	0
1	Aaron Ramsdale	Arsenal	0
2	Aaron Ramsey	Burnley	0
14	Alfie Gilchrist	Chelsea	0
17	Amad Diallo	Manchester United	0
Cluster 1:			
	Player	Team	Cluster
9	Alejandro Garnacho	Manchester United	1
12	Alexis Mac Allister	Liverpool	1
13	Alfie Doughty	Luton Town	1
16	Alphonse Areola	West Ham United	1
24	André Onana	Manchester United	1
Cluster 2:			
	Player	Team	Cluster
3	Aaron Wan-Bissaka	Manchester United	2
18	Amari'i Bell	Luton Town	2
22	Andrew Robertson	Liverpool	2
23	Andros Townsend	Luton Town	2
32	Antony	Manchester United	2
Cluster 3:			
	Player	Team	Cluster
4	Adam Lallana	Brighton and Hove Albion	3
6	Adam Webster	Brighton and Hove Albion	3
7	Adam Wharton	Crystal Palace	3
8	Albert Sambi Lokonga	Luton Town	3
10	Alex Scott	Bournemouth	3
Cluster 4:			
	Player	Team	Cluster
5	Adam Smith	Bournemouth	4
11	Alexander Isak	Newcastle United	4
15	Alisson	Liverpool	4
25	Anel Ahmedhodžić	Nottingham Forest	4
31	Antoine Semenyo	Bournemouth	4

Biểu đồ: hình 3.1 trong Folder Hình ảnh



## **Phần 2:**

Ta có thể chia cầu thủ thành 4 nhóm là:

- Nhóm cầu thủ trẻ và có ít thời gian thi đấu: nhiều tiềm năng
- Nhóm cầu thủ có thời gian thi đấu ở mức giữa nhưng không nổi bật : đóng vai trò hỗ trợ
- Nhóm các cầu thủ trụ cột: mọi thứ ở mức tốt
- Nhóm cầu thủ sắp kết thúc sự nghiệp: đóng góp tốt với thời gian thi đấu hạn chế

Ta có thể phân thành 4 nhóm này vì có sự khác biệt rất rõ về vai trò giúp ta có cái nhìn trực quan hơn về các điểm mạnh và yếu của các nhóm

### Phần 3:

Công nghệ:

- Pandas: Xử lý và phân tích dữ liệu, đọc file CSV.
- Matplotlib: Trực quan hóa dữ liệu 2D, tạo biểu đồ.
- Seaborn: Tạo biểu đồ thống kê hấp dẫn.
- Scikit-learn: Bao gồm phân nhóm, chuẩn hóa và giảm chiều dữ liệu.

Thuật toán:

- Đọc dữ liệu từ CSV.
- Chuẩn hóa dữ liệu.
- Áp dụng K-means để phân nhóm.
- Giảm chiều dữ liệu với PCA.
- Hiển thị kết quả và vẽ biểu đồ phân cụm.

Ta được bảng :

Cluster 0:			
	Player	Team	Cluster
0	Aaron Cresswell	West Ham United	0
1	Aaron Ramsdale	Arsenal	0
2	Aaron Ramsey	Burnley	0
4	Adam Lallana	Brighton and Hove Albion	0
6	Adam Webster	Brighton and Hove Albion	0
Cluster 1:			
	Player	Team	Cluster
12	Alexis Mac Allister	Liverpool	1
13	Alfie Doughty	Luton Town	1
22	Andrew Robertson	Liverpool	1
31	Antoine Semenyo	Bournemouth	1
32	Antony	Manchester United	1
Cluster 2:			
	Player	Team	Cluster
11	Alexander Isak	Newcastle United	2
50	Bruno Fernandes	Manchester United	2
53	Bukayo Saka	Arsenal	2
60	Carlton Morris	Luton Town	2
74	Cole Palmer	Chelsea	2
Cluster 3:			
	Player	Team	Cluster
9	Alejandro Garnacho	Manchester United	3
29	Anthony Gordon	Newcastle United	3
46	Bernardo Silva	Manchester City	3
49	Brennan Johnson	Tottenham Hotspur	3
55	Callum Wilson	Newcastle United	3
Cluster 4:			
	Player	Team	Cluster
3	Aaron Wan-Bissaka	Manchester United	4
5	Adam Smith	Bournemouth	4
15	Alisson	Liverpool	4
16	Alphonse Areola	West Ham United	4
18	Amari'i Bell	Luton Town	4

Biểu đồ: hình 3.3 trong Folder Hình ảnh

## Phần 4:

Công nghệ:

- Pandas: Dùng để đọc và xử lý dữ liệu từ tệp CSV, lưu trữ trong DataFrame.
- NumPy: Hỗ trợ thao tác với mảng và tạo góc cho biểu đồ radar.
- Matplotlib: Vẽ biểu đồ radar để so sánh chỉ số giữa hai cầu thủ.
- Argparse: Quản lý tham số đầu vào từ dòng lệnh, cho phép người dùng nhập tên cầu thủ và thuộc tính cần so sánh

Thuật toán:

- Đọc Dữ liệu: Sử dụng Pandas để tải dữ liệu cầu thủ từ tệp CSV.
- Nhận Tham số: Sử dụng Argparse để nhận tên cầu thủ và thuộc tính từ dòng lệnh.
- Lấy Dữ liệu: Trích xuất dữ liệu cho từng cầu thủ dựa trên tên và thuộc tính.
- Kiểm tra Dữ liệu: Đảm bảo dữ liệu không rỗng cho cả hai cầu thủ.
- Vẽ Biểu đồ Radar: Sử dụng Matplotlib để tạo biểu đồ radar hiển thị so sánh chỉ số giữa hai cầu thủ.

Biểu đồ: 3.4 trong Folder Hình vẽ