

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



**BÁO CÁO BÀI TẬP LỚN**  
**LẬP TRÌNH PYTHON**

<b>Giảng viên hướng dẫn</b>	<b>: Kim Ngọc Bách</b>
<b>Sinh viên</b>	<b>: Phạm Văn Tuyến</b>
<b>Mã sinh viên</b>	<b>: B22DCCN773</b>
<b>Nhóm lớp</b>	<b>: 11</b>

**Hà Nội, 11/2024**

# Nội dung báo cáo bài tập assignment

## 1. Bài 1 :

Công nghệ sử dụng :

- Requests : Được sử dụng để lấy nội dung của trang web.
- BeautifulSoup : Được dùng để tìm kiếm và xử lý các phần tử HTML.
- Pandas : Được dùng để tạo DataFrame sau đó trích xuất sang tệp CSV.

Các bước thực hiện :

- Web scraping: Sử dụng requests và BeautifulSoup để thu thập và phân tích dữ liệu từ trang web.
- Lọc và thêm dữ liệu: Kiểm tra và cập nhật danh sách cầu thủ dựa trên điều kiện nhất định.
- Thao tác danh sách: Duyệt qua, loại bỏ cột không cần thiết, và sắp xếp danh sách.
- Xuất dữ liệu: Tạo DataFrame và xuất ra tệp CSV.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	Name	Nation	Position	Team	Age	Matches pl	Starts	Minutes	Assists	non-Penal	Penalty	Go Yellow	Car Red	Cards xG	npvG	xAG	PrgC	PrgP	PrgR	Gls	Ast	G+A	G-PK
2	Aaron Cresswell	eng	DF	West Ham	33	11	4	436	0	0	0	1	0	0	0	0.4	4	26	6	0	0	0	0
3	Aaron Ramsdale	eng	GK	Arsenal	25	6	6	540	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
4	Aaron Wanless	eng	DF	Manchester City	25	22	20	1780	2	0	0	4	0	0.1	0.1	1.5	30	77	54	0	0.1	0.1	0
5	Aaron Hickey	sct	SC	Brentford	21	9	9	713	0	0	0	5	0	0.2	0.2	0.1	9	21	13	0	0	0	0
6	Aaron Ramsdale	eng	MFW	Burnley	20	14	5	527	0	0	0	1	0	0.3	0.3	0.4	8	8	17	0	0	0	0
7	Abdoulaye Minty	MLI	FWMF	Everton	30	32	32	2629	1	7	0	7	0	8.8	8.8	2.9	55	97	140	0.24	0.03	0.27	0.24
8	Adam Lallana	eng	MFW	Brighton	35	25	13	850	1	0	0	2	0	0.8	0.8	1.7	9	39	55	0	0.11	0.11	0
9	Adam Smith	eng	DF	Bournemouth	32	28	25	2150	2	0	0	6	0	0.1	0.1	1.3	24	89	28	0	0.08	0.08	0
10	Adam Webster	eng	DF	Brighton	28	15	13	1144	0	0	0	2	0	0.4	0.4	0.1	16	80	2	0	0	0	0
11	Adam Whelan	eng	MF	Crystal Palace	19	16	15	1297	3	0	0	2	0	0.3	0.3	2.4	14	79	10	0	0.21	0.21	0
12	Adama Traoré	esp	FWMF	Fulham	27	17	1	377	3	2	0	2	0	1.5	1.5	0.7	29	16	48	0.48	0.72	1.19	0.48
13	Albert Sambi	bel	MF	Luton Town	23	17	16	1303	3	1	0	4	0	0.6	0.6	1.4	29	72	13	0.07	0.21	0.28	0.07
14	Alejandro Baraja	arg	FW	Manchester City	19	36	30	2565	4	7	0	4	0	8.4	8.3	5.1	178	62	281	0.25	0.14	0.39	0.25
15	Alex Iwobi	ng	MF	Everton	27	2	2	140	0	0	0	0	0	0.3	0.3	0.2	6	4	8	0	0	0	0
16	Alex Iwobi	ng	FWMF	Fulham	27	30	25	2192	2	5	0	2	0	5.3	5.3	4.7	109	147	175	0.21	0.08	0.29	0.21
17	Alex Scott	eng	MF	Bournemouth	19	23	11	1014	1	1	0	3	0	0.7	0.7	1.7	24	50	35	0.09	0.09	0.18	0.09
18	Alexander Seferovic	swe	FW	Newcastle	23	30	27	2255	2	16	5	1	0	20.3	15.6	3.7	68	71	129	0.84	0.08	0.92	0.64
19	Alexis Mac Allister	arg	MF	Liverpool	24	33	31	2599	5	4	1	7	1	3.7	2.9	3.6	44	209	48	0.17	0.17	0.35	0.14
20	Alfie Gilchrist	eng	DF	Luton Town	23	37	34	2925	8	2	0	5	0	1.3	1.3	6.3	94	106	164	0.06	0.25	0.31	0.06
21	Alfie Gilchrist	eng	DF	Chelsea	19	11	2	207	0	1	0	1	0	0.3	0.3	0	1	7	4	0.43	0	0.43	0.43
22	Alisson Becker	bra	GK	Liverpool	30	28	28	2520	0	0	0	1	0	0	0	0	0	5	0	0	0	0	0
23	Alphonse Areola	fra	GK	West Ham	30	31	31	2699	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0
24	Amad Diallo	ci	FWMF	Manchester City	21	9	3	390	1	1	0	1	0	0.5	0.5	1.1	11	10	31	0.23	0.23	0.46	0.23
25	Amadou Onana	bel	MF	Everton	21	30	23	2091	0	2	0	5	0	2	2	2.7	21	107	39	0.09	0	0.09	0.09
26	Amaro Beljima	jam	DF	Luton Town	29	21	21	1720	0	0	0	0	0	0.3	0.3	0.5	32	76	10	0	0	0	0
27	Ameen Al-Balushi	bel	DF	Burnley	21	13	12	1056	0	1	0	1	0	1.3	1.3	0.1	7	35	9	0.09	0	0.09	0.09
28	Anass Zarcori	mar	FWMF	Burnley	22	6	1	152	0	0	0	0	1	0.2	0.2	0	2	2	12	0	0	0	0
29	Andre Brocchi	eng	MDF	Sheffield United	19	20	12	938	0	0	0	1	0	0.4	0.4	0.1	12	18	22	0	0	0	0
30	Andreas Pini	bra	MF	Fulham	27	37	34	2613	7	3	0	6	0	4.6	4.6	7.1	56	138	125	0.1	0.24	0.34	0.1

Hình 1. Hình ảnh được trích từ file results.csv

## 2. Bài 2 :

Công nghệ chung được sử dụng :

- Matplotlib : Được dùng để tạo các biểu đồ và đồ thị.
- Pandas : Được dùng để tạo DataFrame sau đó trích xuất sang tệp CSV.

### a. Bài 2.1:

Các bước thực hiện :

- Đọc dữ liệu
- Chuyển đổi kiểu dữ liệu
- Tính toán các thống kê
- Tính thống kê tổng thể
- Tính thông kê theo nhóm
- Tạo DataFrame kết quả
- Định dạng DataFrame cuối cùng
- Xuất dữ liệu ra tệp CSV

Top 3 cầu thủ có chỉ số Won% cao nhất:	Top 3 cầu thủ có chỉ số Lost.1 cao nhất:
Łukasz Fabiański   100.0	Carlton Morris   169
Đorđe Petrović   100.0	Dominic Calvert-Lewin   144
Tim Iroegbunam   100.0	Dominic Solanke   135
Top 3 cầu thủ có chỉ số Won% thấp nhất:	Top 3 cầu thủ có chỉ số Lost.1 thấp nhất:
Bryan Gil   0.0	Aaron Ramsdale   0
Benson Manuel   0.0	Đorđe Petrović   0
Mahmoud Dahoud   0.0	Łukasz Fabiański   0

Hình 2. Hình ảnh mô tả kết quả Tìm kiếm Top 3 cầu thủ có chỉ số thấp nhất và cao nhất

b. Bài 2.2:

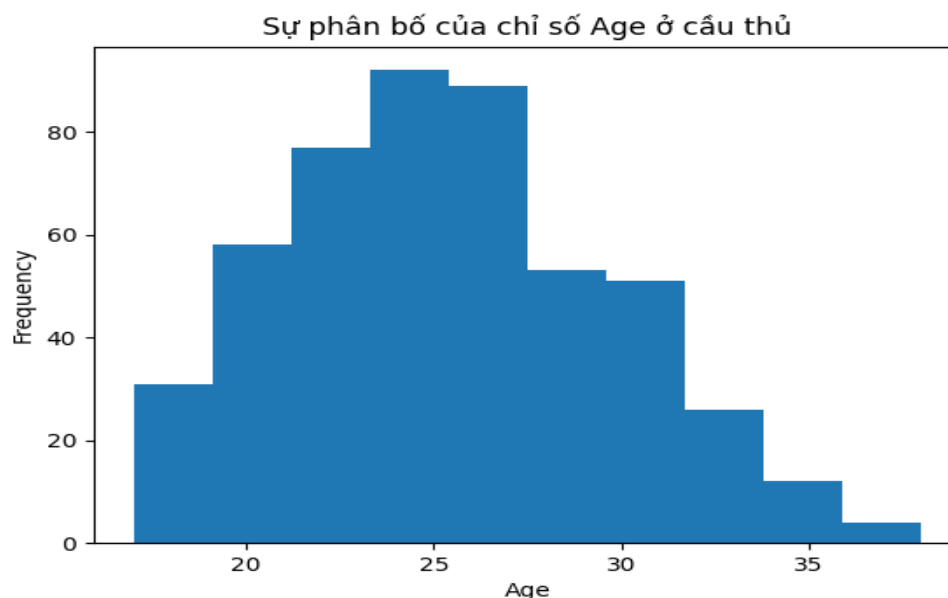
Name	Median of all	Median of West Ham	Median of Arsenal	Median of Manchester	Median of Brentford	Median of Burnley	Median of Everton	Median of Brighton	Median of Bournemouth	Median of Crystal Pal	Median of Fulham	Median of Luton Tow	Median of Newcastle	Median of Liverpool	Median of Chelsea	Median of Sheffield U	Median of Nott'ham F	Median of Tottenham	Median of Manchest	Median of Aston Villa	Median of Wolves
all	23	16	1419	1	1	0	3	0	1.1	1.1	1	17	44	30	0.07	0.05	0.15	0.07	0.15		
West Ham	23.5	21	1776	0	1	0	2.5	0	1.3	1.25	0.35	10	33.5	28	0.045	0	0.095	0.045	0.095		
Arsenal	27	18	1649	2	2	0	2	0	1.6	1.6	1.4	17	72	37	0.12	0.17	0.25	0.12	0.25		
Manchester	22	15	1351.5	1	1	0	2.5	0	0.85	0.85	0.9	14	52	32.5	0.07	0.07	0.155	0.07	0.155		
Brentford	26	15	1321	1	1	0	3	0	1.2	1.2	1.5	16	40	32	0.08	0.09	0.15	0.08	0.15		
Burnley	16	14	1213	1	1	0	2	0	0.55	0.55	0.8	15	34.5	25	0.035	0.055	0.09	0.035	0.09		
Everton	28	23	1884	0	1	0	2	0	1.5	1.5	1	14	38	39	0.09	0	0.09	0.09	0.09		
Brighton	20	15	1344.5	1	0	0	3	0	1	1	0.9	21	43	41	0	0.04	0.215	0	0.215		
Bournemouth	25.5	13	1317.5	1	0.5	0	3	0	1.15	1.15	1.1	22	43.5	27.5	0.015	0.08	0.15	0.015	0.15		
Crystal Pal	22.5	17.5	1587.5	1	0	0	2.5	0	0.6	0.6	1	14.5	40.5	17.5	0	0.045	0.08	0	0.08		
Fulham	29	18	1593	1	2	0	3	0	1.6	1.6	1.1	15	62	48	0.1	0.06	0.17	0.1	0.17		
Luton Tow	23	16	1304	0	1	0	3	0	1	1	0.7	15	37	21	0.07	0	0.18	0.07	0.18		
Newcastle	21	14.5	1296.5	1	1.5	0	2	0	1.35	1.35	1.05	21.5	52.5	57	0.135	0.07	0.21	0.135	0.21		
Liverpool	28	17	1671	2	2	0	2	0	2.3	2.15	1.65	41	91.5	80.5	0.125	0.1	0.24	0.125	0.24		
Chelsea	23	18	1576	1	1	0	3	0	1.1	1.1	1.2	22	46	35	0.09	0.08	0.2	0.08	0.16		
Sheffield U	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		
Nott'ham F	20	15	1439	1	0	0	2	0	0.7	0.7	0.65	10	31	14	0	0.05	0.075	0	0.075		
Tottenham	27.5	15.5	1401	1	1.5	0	3	0	1.65	1.65	0.85	23	71.5	33	0.085	0.095	0.185	0.085	0.185		
Manchest	29	24	2042	1	2	0	2	0	1.9	1.9	1.4	37	100	55	0.15	0.09	0.22	0.15	0.22		
Aston Villa	27	20	1652	1	2	0	4	0	1.4	1.4	1	32	55	33	0.11	0.08	0.21	0.11	0.21		
Wolves	25	11	1222	0	1	0	2	0	1.1	1.1	1	21	37	17	0.03	0	0.05	0.03	0.05		

Hình 3. Hình ảnh được trích từ file results2.csv

c. Bài 2.3:

Các bước thực hiện :

- Import thư viện cần thiết
- Đọc dữ liệu từ file CSV
- Vẽ histogram cho từng cột



Hình 4. Hình ảnh mô tả sự phân bố của chỉ số Age ở các cầu thủ

d. Bài 2.4:

Các bước thực hiện :

- Đọc dữ liệu
- Xác định các cột cần tính tổng
- Chuyển đổi dữ liệu sang kiểu số
- Tính tổng theo nhóm
- Tìm nhóm có tổng cao nhất
- Xuất kết quả

Matches played: Brighton	G+A: Newcastle Utd
Starts: Brentford	G-PK: Newcastle Utd
Minutes: Crystal Palace	G+A-PK: Newcastle Utd
Assists: Manchester City	xG.1: Bournemouth
non-Penalty Goals: Manchester City	xAG.1: Tottenham
Penalty Goals: Chelsea	xG + xAG: Bournemouth
Yellow Cards: Chelsea	npxG.1: Bournemouth
Red Cards: Burnley	npxG + xAG: Bournemouth
xG: Liverpool	GA: Sheffield Utd
npxG: Liverpool	GA90: Bournemouth
xAG: Liverpool	SoTA: Sheffield Utd
PrgC: Manchester City	Saves: West Ham
PrgP: Liverpool	Save%: Bournemouth
PrgR: Tottenham	W: Arsenal
	D: Brighton

Hình 5. Hình ảnh mô tả một phần đội bóng có số điểm cao nhất ở mỗi chỉ số

Dựa trên kết quả tìm được thì đội bóng Manchester City có phong độ tốt nhất giải ngoại hạng anh mùa 2023-2024

3. Bài 3 :

Công nghệ chung được sử dụng :

- Pandas : Được dùng để tạo DataFrame sau đó trích xuất sang tệp CSV.
- Numpy: Thư viện hỗ trợ tính toán số học và thao tác với mảng dữ liệu một cách hiệu quả.
- Matplotlib : Được dùng để tạo các biểu đồ và đồ thị.
- Sklearn : Để thực hiện phân cụm.
- Seaborn : Hỗ trợ tạo các biểu đồ trực quan, đẹp mắt và dễ dàng hơn. Được sử dụng để vẽ biểu đồ phân tán với các nhóm màu khác nhau.

a. Bài 3.1:

--- Cluster 0 ---				--- Cluster 2 ---			
	Name	Team	Cluster		Name	Team	Cluster
2	Aaron Wan-Bissaka	Manchester Utd	0	0	Aaron Cresswell	West Ham	2
24	Amari'i Bell	Luton Town	0	1	Aaron Ramsdale	Arsenal	2
29	Andrew Robertson	Liverpool	0	3	Aaron Hickey	Brentford	2
31	Andros Townsend	Luton Town	0	4	Aaron Ramsey	Burnley	2
43	Antony	Manchester Utd	0	10	Adama Traoré	Fulham	2
-----				-----			
--- Cluster 1 ---				--- Cluster 3 ---			
	Name	Team	Cluster		Name	Team	Cluster
5	Abdoulaye Doucouré	Everton	1	6	Adam Lallana	Brighton	3
12	Alejandro Garnacho	Manchester Utd	1	8	Adam Webster	Brighton	3
17	Alexis Mac Allister	Liverpool	1	9	Adam Wharton	Crystal Palace	3
18	Alfie Doughty	Luton Town	1	11	Albert Sambi Lokonga	Luton Town	3
21	Alphonse Areola	West Ham	1	15	Alex Scott	Bournemouth	3
-----				-----			

Hình 6. Hình ảnh mô tả kết quả sử dụng K-means để phân loại các cầu thủ có chỉ số giống nhau

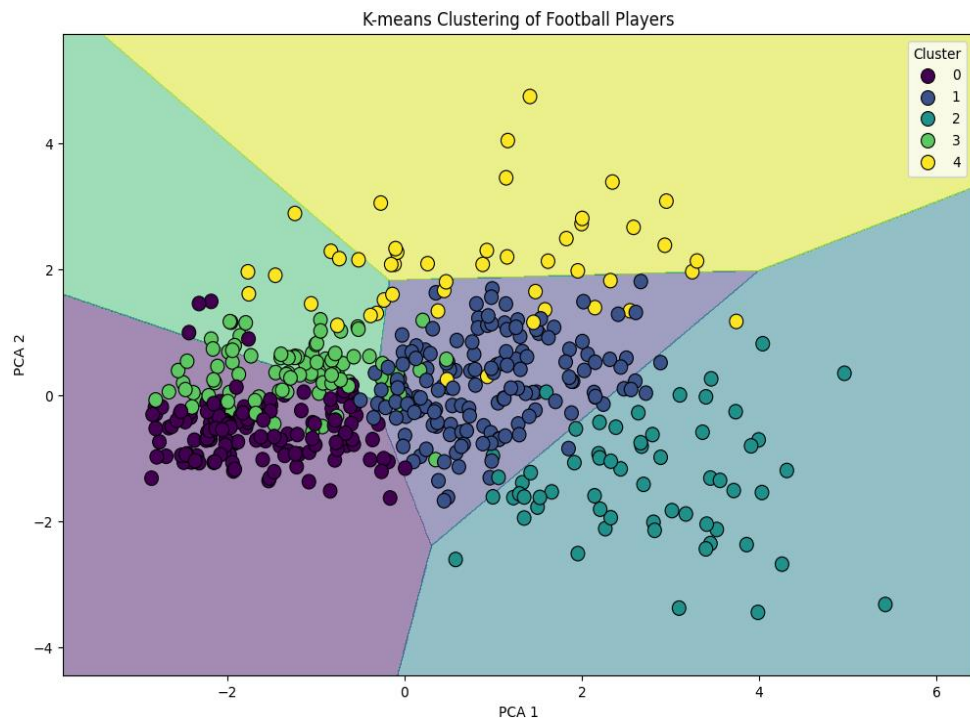
b. Bài 3.2:

Có thể chia cầu thủ thành 5 nhóm :

- Nhóm 1 là nhóm cầu thủ có thời gian thi đấu ít
- Nhóm 2 là nhóm cầu thủ có thời gian thi đấu trung bình nhưng chỉ số thấp
- Nhóm 3 là nhóm cầu thủ có thời gian thi đấu trung bình và chỉ số ở mức cao
- Nhóm 4 là nhóm cầu thủ có thời gian thi đấu nhiều và chỉ số ở mức cao
- Nhóm 5 là nhóm cầu thủ sắp kết thúc sự nghiệp

Các nhóm được chia dựa theo độ tuổi sự nghiệp, thời gian thi đấu và chỉ số cá nhân từ đó thấy được điểm mạnh, điểm yếu của cầu thủ; thấy được sự nghiệp đang ở mức nào.

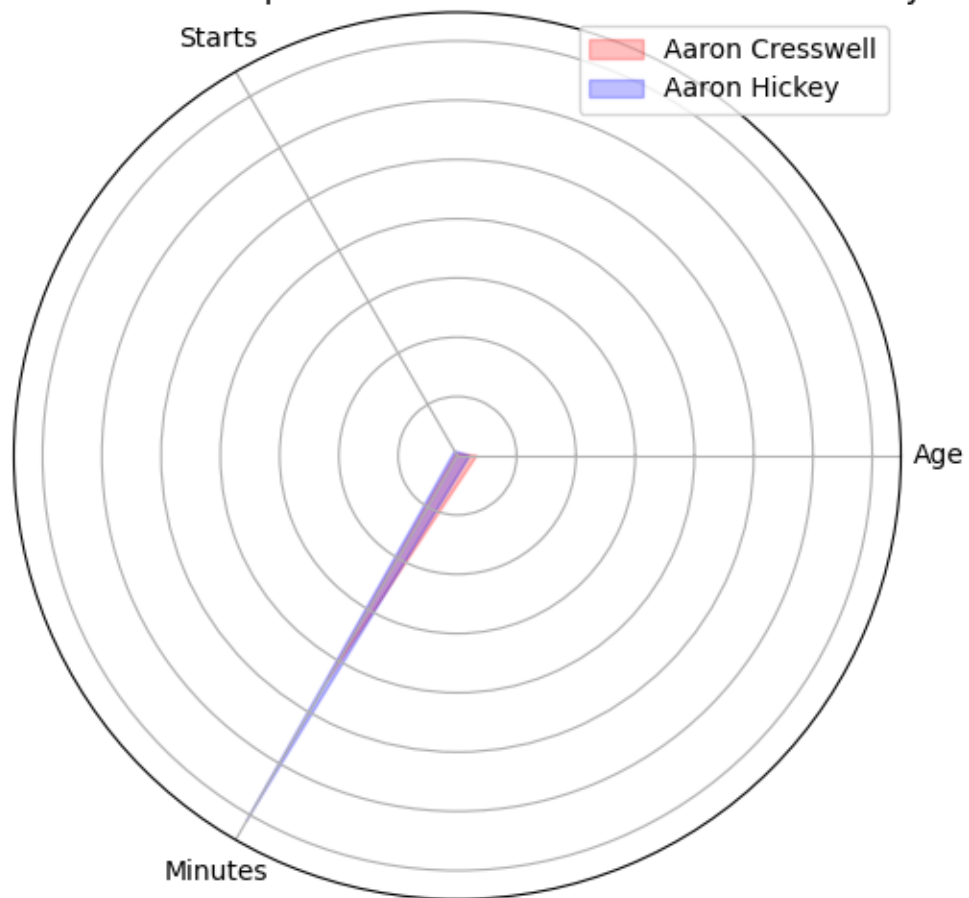
c. Bài 3.3:



Hình 7. Hình ảnh mô tả kết quả sử dụng thuật toán PCA để vẽ hình phân cụm

d. Bài 3.4:

Radar Chart Comparison: Aaron Cresswell vs Aaron Hickey



Hình 8 . Hình ảnh mô tả so sánh giữa cầu thủ Aaron Cresswell và Aaron Hickey về chỉ số Starts, Age và Minutes