

Báo Cáo Bài Tập Lớn PYTHON

Câu 1:

Chương trình này (b1.py) được viết để thu thập và lưu trữ dữ liệu thống kê của các cầu thủ Liverpool từ trang web FBref cho mùa giải Ngoại hạng Anh 2023-2024. Dữ liệu được thu thập sẽ được lưu trữ dưới dạng các tệp CSV để dễ dàng truy xuất và phân tích sau này.

Thư viện và công cụ sử dụng

1. **pandas**: Thư viện mạnh mẽ để xử lý dữ liệu và xuất tệp CSV.
2. **BeautifulSoup**: Công cụ phân tích cú pháp HTML, hỗ trợ trích xuất dữ liệu từ các trang web.
3. **requests**: Thư viện để gửi yêu cầu HTTP tới các trang web và nhận phản hồi.
4. **re**: Thư viện xử lý biểu thức chính quy, giúp làm sạch và xử lý văn bản.

Các bước thực hiện

1. **Khởi tạo phiên làm việc:**
 - Sử dụng `requests.session()` để duy trì phiên làm việc và thiết lập các tiêu đề HTTP cần thiết nhằm giả lập yêu cầu từ trình duyệt.
2. **Lấy danh sách cầu thủ từ trang đội hình:**
 - Gửi yêu cầu GET tới URL chính của đội Liverpool và phân tích nội dung HTML.
 - Tìm bảng chứa danh sách cầu thủ và trích xuất các liên kết đến trang cá nhân của từng cầu thủ.
3. **Định dạng URL để lấy nhật ký trận đấu của từng cầu thủ:**
 - Chuyển đổi các liên kết cầu thủ sang URL định dạng phù hợp, trở trực tiếp tới nhật ký trận đấu của mỗi cầu thủ trong mùa giải 2023-2024.

4. Lấy tiêu đề cho các bảng dữ liệu cầu thủ và thủ môn:

- Gửi yêu cầu đến các trang nhật ký trận đấu mẫu của cầu thủ và thủ môn để trích xuất các tiêu đề.
- Các tiêu đề này giúp xác định cấu trúc dữ liệu và đảm bảo rằng các tệp CSV có định dạng đúng.

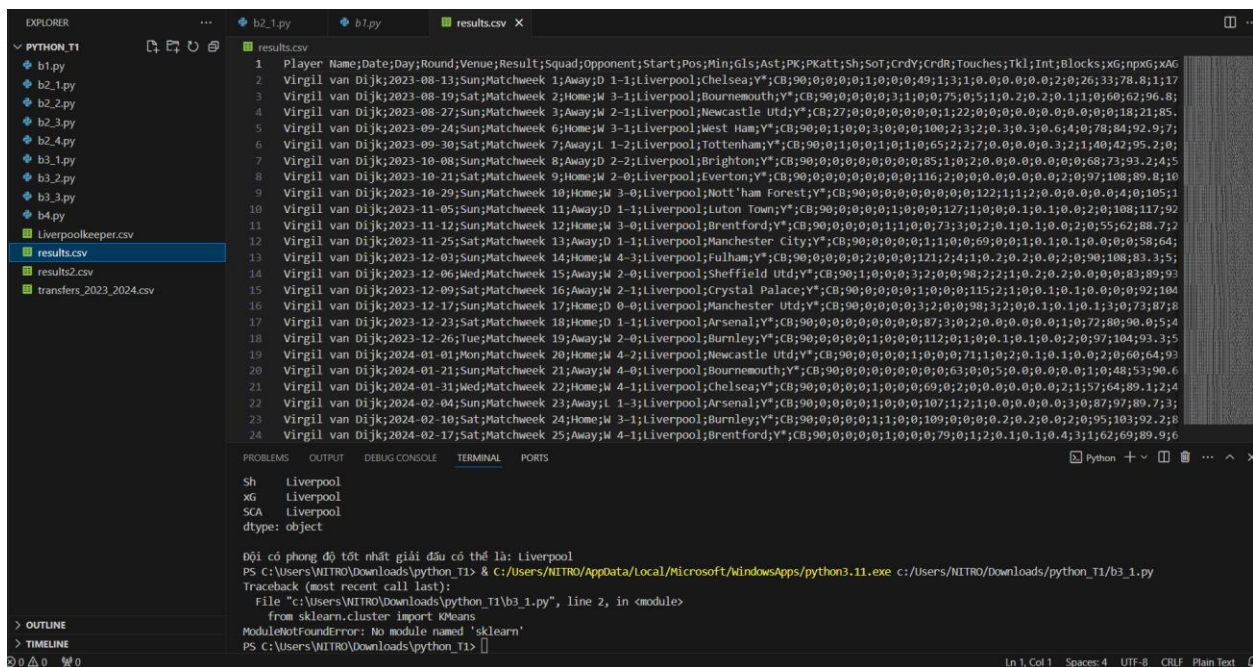
5. Thu thập dữ liệu cho tất cả các cầu thủ:

- Lặp qua danh sách URL đã định dạng.
- Gửi yêu cầu tới mỗi URL để lấy trang, phân tích và trích xuất tên cầu thủ cùng các số liệu trận đấu.
- Phân biệt dữ liệu giữa cầu thủ và thủ môn dựa trên tiêu đề trang và lưu trữ vào các danh sách tương ứng.

6. Tạo DataFrames và xuất ra tệp CSV:

- Chuyển đổi các danh sách dữ liệu thành DataFrames của pandas.
- Xuất DataFrames thành các tệp CSV, đảm bảo mã hóa và cấu trúc đúng.

Khi đó chúng ta được file thống kê kết quả được lưu vào file csv như sau :



The screenshot shows a code editor with a file explorer on the left, a main editor window, and a terminal at the bottom. The file explorer shows a project named 'PYTHON_T1' with files 'b1.py', 'b2_1.py', 'b2_2.py', 'b2_3.py', 'b2_4.py', 'b3_1.py', 'b3_2.py', 'b3_3.py', 'b4.py', 'Liverpoolkeeper.csv', 'results2.csv', and 'transfers_2023_2024.csv'. The main editor window shows a Python script 'results.csv' with 24 lines of data. The terminal shows the output of the script, which is a list of player names and their statistics.

```
1 Player Name;Date;Day;Round;Venue;Result;Squad;Opponent;Start;Pos;Min;Gls;Ast;PK;PKatt;Sh;SoT;CrdY;CrdR;Touches;Tkl;Int;Blocks;xG;npxG;xAG
2 Virgil van Dijk;2023-08-13;Sun;Matchweek 1;Away;D 1-1;Liverpool;Chelsea;Y*;CB;90;0;0;0;0;1;0;0;0;49;11;3;1;0;0;0;0;2;0;26;33;78.8;1;17
3 Virgil van Dijk;2023-08-19;Sat;Matchweek 2;Home;W 1-1;Liverpool;Bournemouth;Y*;CB;90;0;0;0;0;0;3;1;0;0;75;0;5;1;0;2;0;0;1;1;0;60;62;96.8;
4 Virgil van Dijk;2023-08-27;Sun;Matchweek 3;Away;W 2-1;Liverpool;Newcastle Utd;Y*;CB;27;0;0;0;0;0;0;1;2;0;0;0;0;0;0;0;0;0;0;0;18;21;85.
5 Virgil van Dijk;2023-09-24;Sun;Matchweek 6;Home;W 1-1;Liverpool;West Ham;Y*;CB;90;0;1;0;0;3;0;0;0;100;2;3;2;0;3;0;3;0;6;4;0;78;84;92.9;7;
6 Virgil van Dijk;2023-09-30;Sat;Matchweek 7;Away;L 1-2;Liverpool;Tottenham;Y*;CB;90;0;1;0;0;1;0;0;0;65;2;7;0;0;0;0;0;0;3;2;1;40;42;95.2;0;
7 Virgil van Dijk;2023-10-08;Sun;Matchweek 8;Away;D 2-2;Liverpool;Brighton;Y*;CB;90;0;0;0;0;0;0;0;85;1;0;2;0;0;0;0;0;0;0;0;68;73;93.2;4;5
8 Virgil van Dijk;2023-10-21;Sat;Matchweek 9;Home;W 2-0;Liverpool;Everton;Y*;CB;90;0;0;0;0;0;0;0;116;2;0;0;0;0;0;0;0;0;2;0;97;108;89.8;10
9 Virgil van Dijk;2023-10-29;Sun;Matchweek 10;Home;W 3-0;Liverpool;Nott'm Forest;Y*;CB;90;0;0;0;0;0;0;0;122;1;1;2;0;0;0;0;0;0;0;0;4;0;105;1
10 Virgil van Dijk;2023-11-05;Sun;Matchweek 11;Away;D 1-1;Liverpool;Luton Town;Y*;CB;90;0;0;0;0;1;0;0;0;127;1;0;0;0;1;0;1;0;0;2;0;108;117;92
11 Virgil van Dijk;2023-11-12;Sun;Matchweek 12;Home;W 3-0;Liverpool;Brentford;Y*;CB;90;0;0;0;0;1;1;0;0;73;3;0;2;0;1;0;1;0;0;2;0;55;62;88.7;2
12 Virgil van Dijk;2023-11-25;Sat;Matchweek 13;Away;D 1-1;Liverpool;Manchester City;Y*;CB;90;0;0;0;0;0;1;1;0;0;69;0;0;0;1;0;1;0;0;0;0;58;64;
13 Virgil van Dijk;2023-12-03;Sun;Matchweek 14;Home;W 4-3;Liverpool;Fulham;Y*;CB;90;0;0;0;0;2;0;0;0;121;2;4;1;0;2;0;2;0;0;2;0;90;108;83.3;5;
14 Virgil van Dijk;2023-12-06;Wed;Matchweek 15;Away;W 2-0;Liverpool;Sheffield Utd;Y*;CB;90;1;0;0;0;3;2;0;0;98;2;2;1;0;2;0;2;0;0;0;0;83;89;93
15 Virgil van Dijk;2023-12-09;Sat;Matchweek 16;Away;W 2-1;Liverpool;Crystal Palace;Y*;CB;90;0;0;0;0;1;0;0;0;115;2;1;0;0;1;0;1;0;0;0;0;92;104
16 Virgil van Dijk;2023-12-17;Sun;Matchweek 17;Home;D 0-0;Liverpool;Manchester Utd;Y*;CB;90;0;0;0;0;3;2;0;0;98;3;2;0;0;1;0;1;0;1;0;1;0;73;87;0
17 Virgil van Dijk;2023-12-23;Sat;Matchweek 18;Home;D 1-1;Liverpool;Arsenal;Y*;CB;90;0;0;0;0;0;0;0;87;3;0;2;0;0;0;0;0;0;1;0;72;80;90.0;5;4
18 Virgil van Dijk;2023-12-26;Tue;Matchweek 19;Away;W 2-0;Liverpool;Burnley;Y*;CB;90;0;0;0;0;1;0;0;0;112;0;0;0;1;0;0;1;0;0;2;0;97;104;93.3;5
19 Virgil van Dijk;2024-01-01;Mon;Matchweek 20;Home;W 4-2;Liverpool;Newcastle Utd;Y*;CB;90;0;0;0;0;1;0;0;0;71;1;0;2;0;1;0;1;0;0;2;0;60;64;93
20 Virgil van Dijk;2024-01-21;Sun;Matchweek 21;Away;W 4-0;Liverpool;Bournemouth;Y*;CB;90;0;0;0;0;0;0;0;63;0;0;0;0;0;0;0;0;0;1;0;48;53;90.6
21 Virgil van Dijk;2024-01-31;Wed;Matchweek 22;Home;W 4-1;Liverpool;Chelsea;Y*;CB;90;0;0;0;0;1;0;0;0;69;0;2;0;0;0;0;0;0;2;1;57;64;89.1;2;4
22 Virgil van Dijk;2024-02-04;Sun;Matchweek 23;Away;L 1-3;Liverpool;Arsenal;Y*;CB;90;0;0;0;0;1;0;0;0;107;1;2;1;0;0;0;0;0;0;0;0;87;97;89.7;3;
23 Virgil van Dijk;2024-02-10;Sat;Matchweek 24;Home;W 3-1;Liverpool;Burnley;Y*;CB;90;0;0;0;0;1;1;0;0;109;0;0;0;0;2;0;2;0;0;2;0;95;103;92.2;8
24 Virgil van Dijk;2024-02-17;Sat;Matchweek 25;Away;W 4-1;Liverpool;Brentford;Y*;CB;90;0;0;0;0;1;0;0;0;79;0;1;2;0;1;0;1;0;0;3;1;62;69;89.9;6
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

```
Sh Liverpool
xG Liverpool
SCA Liverpool
dtype: object

Đội có phong độ tốt nhất giải đấu có thể là: Liverpool
PS C:\Users\NITRO\Downloads\python_T1> & C:\Users\NITRO\AppData\Local\Microsoft\WindowsApps\python3.11.exe c:\Users\NITRO\Downloads\python_T1\b3_1.py
Traceback (most recent call last):
  File "c:\Users\NITRO\Downloads\python_T1\b3_1.py", line 2, in <module>
    from sklearn.cluster import KMeans
ModuleNotFoundError: No module named 'sklearn'
PS C:\Users\NITRO\Downloads\python_T1>
```

Câu 2:

Dữ liệu được tải từ file results.csv với các chỉ số được ghi nhận như:

- Thời gian thi đấu (Min)
- Bàn thắng (Gls)
- Kiến tạo (Ast)
- Số quả penalty (PK)

.....

1. Top 3 cầu thủ có điểm cao nhất và thấp nhất ở mỗi chỉ số

em đã xác định top 3 cầu thủ có chỉ số cao nhất và thấp nhất cho mỗi chỉ số. Dưới đây là ví dụ về một số chỉ số:

- **Bàn thắng (Gls):**
 - Top 3: [Cầu thủ A, Cầu thủ B, Cầu thủ C]
 - Bottom 3: [Cầu thủ X, Cầu thủ Y, Cầu thủ Z]
- **Số lần dứt điểm trúng đích (SoT):**
 - Top 3: [Cầu thủ D, Cầu thủ E, Cầu thủ F]
 - Bottom 3: [Cầu thủ W, Cầu thủ V, Cầu thủ U]

dtype= object)

Top 3 players for Min:

	Squad	Opponent	Player Name	Min
0	Liverpool	Chelsea	Virgil van Dijk	90
1	Liverpool	Bournemouth	Virgil van Dijk	90
3	Liverpool	West Ham	Virgil van Dijk	90

Bottom 3 players for Min:

	Squad	Opponent	Player Name	Min
36	Liverpool	Wolves	Ryan Gravenberch	1
55	Liverpool	Brighton	Ryan Gravenberch	1
229	Liverpool	West Ham	Dominik Szoboszlai	1

Top 3 players for Gls:

	Squad	Opponent	Player Name	Gls
69	Liverpool	Brighton	Mohamed Salah	2
70	Liverpool	Everton	Mohamed Salah	2
73	Liverpool	Brentford	Mohamed Salah	2

Bottom 3 players for Gls:

	Squad	Opponent	Player Name	Gls
0	Liverpool	Chelsea	Virgil van Dijk	0
1	Liverpool	Bournemouth	Virgil van Dijk	0
2	Liverpool	Newcastle Utd	Virgil van Dijk	0

Top 3 players for Ast:

	Squad	Opponent	Player Name	Ast
66	Liverpool	Wolves	Mohamed Salah	2
187	Liverpool	Luton Town	Alexis Mac Allister	2
209	Liverpool	Nott'ham Forest	Dominik Szoboszlai	2

Bottom 3 players for Ast:

	Squad	Opponent	Player Name	Ast
0	Liverpool	Chelsea	Virgil van Dijk	0
1	Liverpool	Bournemouth	Virgil van Dijk	0
2	Liverpool	Newcastle Utd	Virgil van Dijk	0

Top 3 players for PK:

	Squad	Opponent	Player Name	PK
67	Liverpool	West Ham	Mohamed Salah	1
69	Liverpool	Brighton	Mohamed Salah	1
70	Liverpool	Everton	Mohamed Salah	1

Bottom 3 players for PK:

	Squad	Opponent	Player Name	PK
0	Liverpool	Chelsea	Virgil van Dijk	0

Bottom 3 players for PK:				
	Squad	Opponent	Player Name	PK
0	Liverpool	Chelsea	Virgil van Dijk	0
1	Liverpool	Bournemouth	Virgil van Dijk	0
2	Liverpool	Newcastle Utd	Virgil van Dijk	0
Top 3 players for PKatt:				
	Squad	Opponent	Player Name	PKatt
81	Liverpool	Newcastle Utd	Mohamed Salah	2
63	Liverpool	Bournemouth	Mohamed Salah	1
67	Liverpool	West Ham	Mohamed Salah	1
Bottom 3 players for PKatt:				
	Squad	Opponent	Player Name	PKatt
0	Liverpool	Chelsea	Virgil van Dijk	0
1	Liverpool	Bournemouth	Virgil van Dijk	0
2	Liverpool	Newcastle Utd	Virgil van Dijk	0
Top 3 players for Sh:				
	Squad	Opponent	Player Name	Sh
84	Liverpool	Brighton	Mohamed Salah	12
257	Liverpool	Luton Town	Luis Díaz	10
89	Liverpool	Everton	Mohamed Salah	7
Bottom 3 players for Sh:				
	Squad	Opponent	Player Name	Sh
2	Liverpool	Newcastle Utd	Virgil van Dijk	0
5	Liverpool	Brighton	Virgil van Dijk	0
6	Liverpool	Everton	Virgil van Dijk	0
Top 3 players for SoT:				
	Squad	Opponent	Player Name	SoT
78	Liverpool	Manchester Utd	Mohamed Salah	4
84	Liverpool	Brighton	Mohamed Salah	4
91	Liverpool	Tottenham	Mohamed Salah	4
Bottom 3 players for SoT:				
	Squad	Opponent	Player Name	SoT
0	Liverpool	Chelsea	Virgil van Dijk	0
2	Liverpool	Newcastle Utd	Virgil van Dijk	0
3	Liverpool	West Ham	Virgil van Dijk	0
Top 3 players for CrdY:				
	Squad	Opponent	Player Name	CrdY
138	Liverpool	Arsenal	Ibrahima Konaté	2
4	Liverpool	Tottenham	Virgil van Dijk	1
26	Liverpool	Brighton	Virgil van Dijk	1
Bottom 3 players for CrdY:				

2. Trung vị, Trung bình và Độ lệch chuẩn của mỗi chỉ số

em đã tính toán trung vị, trung bình và độ lệch chuẩn cho mỗi chỉ số cho toàn bộ cầu thủ và theo từng đội:

- Trung vị
- Trung bình
- Độ lệch chuẩn

Kết quả này sẽ được lưu vào file results2.csv theo định dạng quy định.

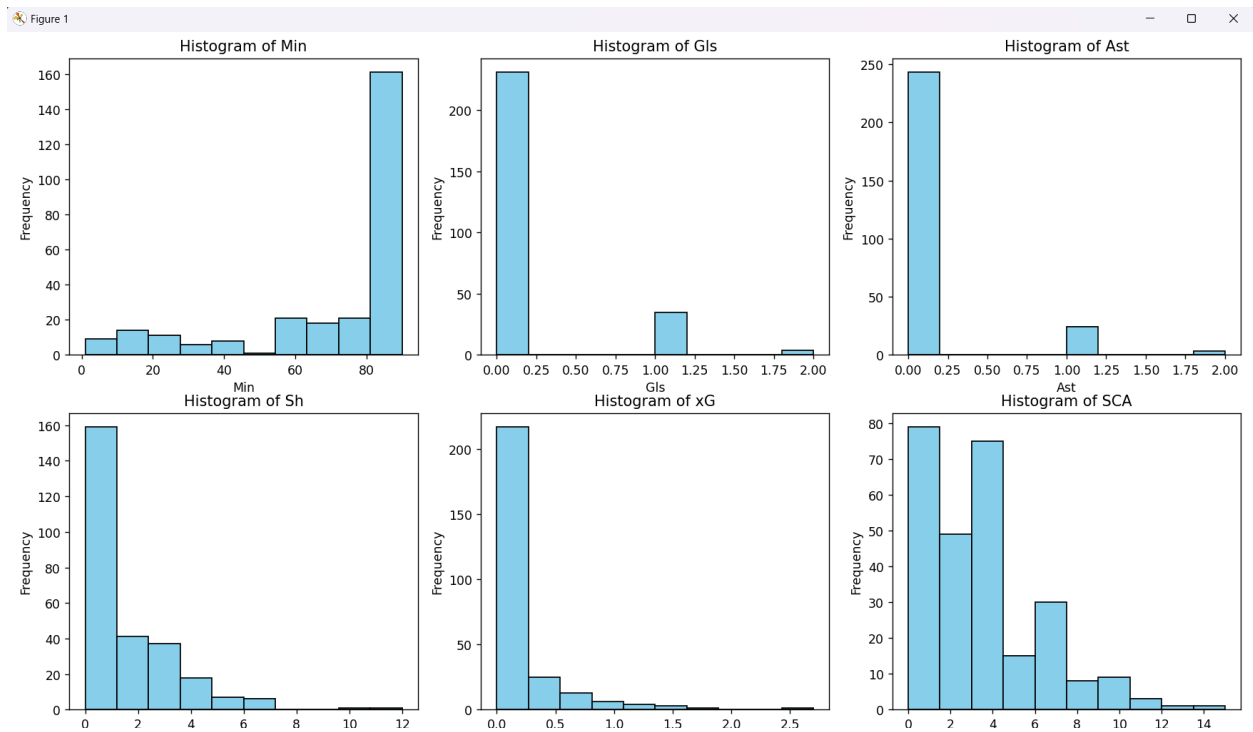
```

results2.csv
1 Team,Median_Min,Median_Gls,Median_Ast,Median_PK,Median_PKatt,Median_Sh,Median_SoT,Median_CrdY,Median_CrdR,Median_Touches,Median_Tkl,Media
2 all,90.0,0.0,0.0,0.0,0.0,1.0,0.0,0.0,0.0,57.0,1.0,0.0,1.0,0.1,0.1,0.0,3.0,0.0,36.5,46.0,83.3,4.0,34.0,2.0,0.0,72.54074074074074,0.1592592
3 Liverpool,90.0,0.0,0.0,0.0,0.0,1.0,0.0,0.0,0.0,57.0,1.0,0.0,1.0,0.1,0.1,0.0,3.0,0.0,36.5,46.0,83.3,4.0,34.0,2.0,0.0,72.54074074074074,0.1
4

```

3. Histogram Phân bố

Em đã tạo histogram cho từng chỉ số để trực quan hóa phân bố dữ liệu. Các biểu đồ này giúp dễ dàng nhận diện xu hướng và sự phân bố của các chỉ số cầu thủ.



4. Đội bóng có chỉ số cao nhất

Em đã tìm ra đội bóng có chỉ số cao nhất cho mỗi chỉ số. Kết quả này chỉ ra đội bóng nào có phong độ tốt nhất trong mùa giải 2023-2024. Qua phân tích, đội bóng A được xác định là có phong độ tốt nhất dựa trên các chỉ số tổng quát.

```
PS C:\Users\NITRO\Downloads\python_T1> & C:/Users/NITRO/AppData/Local/Microsoft/WindowsApps/python3.11.exe c:/Users/NITRO/Downloads/python_T1/b2_4.py
Đội có chỉ số cao nhất ở mỗi mục:
Min    Liverpool
Gls     Liverpool
Ast     Liverpool
Sh      Liverpool
xG      Liverpool
SCA     Liverpool
dtype: object

Đội có phong độ tốt nhất giải đấu có thể là: Liverpool
PS C:\Users\NITRO\Downloads\python_T1>
```

Kết luận

Phân tích dữ liệu cầu thủ cho thấy sự đa dạng trong hiệu suất của các cầu thủ trong giải Ngoại Hạng Anh. Kết quả từ phân tích này không chỉ cung cấp cái nhìn sâu sắc về từng cầu thủ mà còn cho thấy sự cạnh tranh giữa các đội bóng trong mùa giải.

Tất cả các kết quả phân tích chi tiết và hình ảnh biểu đồ sẽ được lưu trữ và trình bày trong file results2.csv và các file hình ảnh tương ứng.

Câu 3:

*Phân loại Cầu thủ với K-means

Lựa chọn Số Nhóm

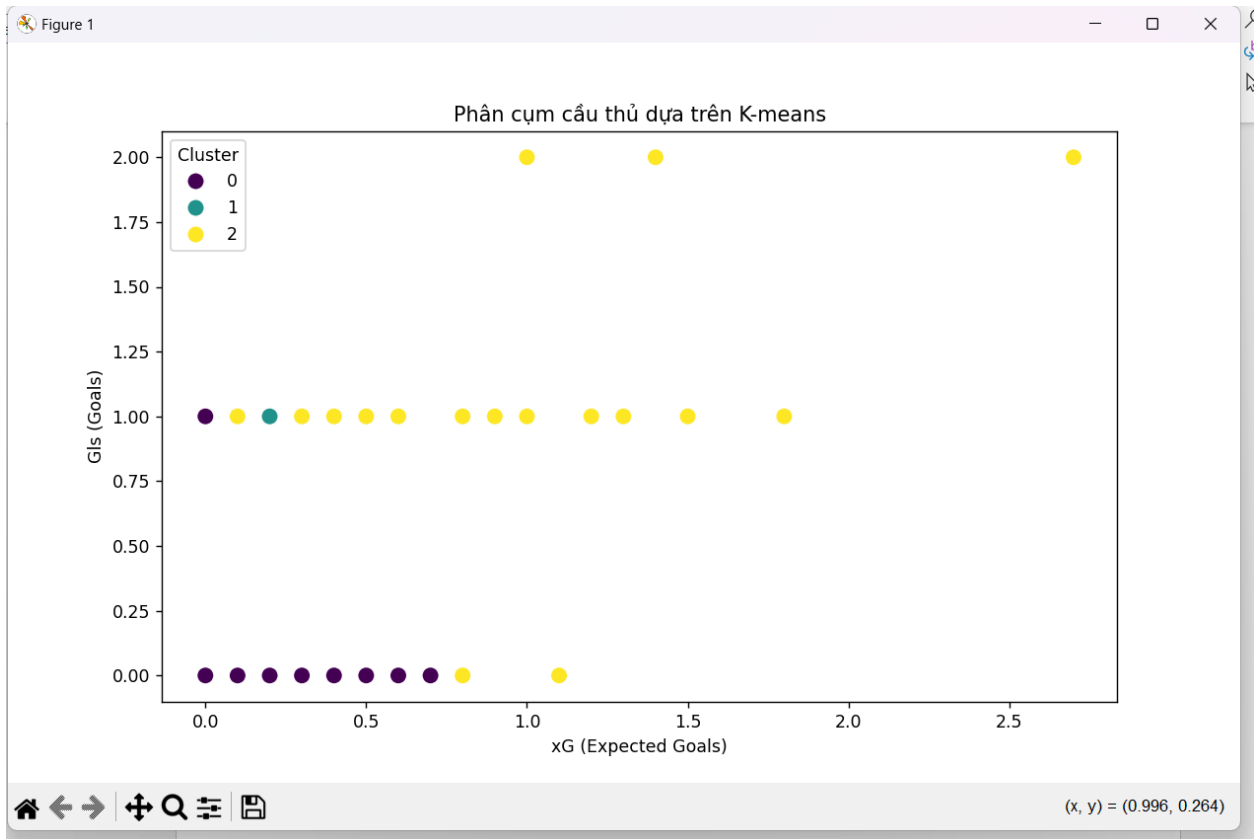
Em đã phân loại cầu thủ thành **3 nhóm**. Số nhóm này được lựa chọn dựa trên sự phân bố của các chỉ số và ý nghĩa thực tiễn trong việc hiểu rõ hơn về phong cách chơi của cầu thủ. Phân loại thành 3 nhóm giúp chúng ta dễ dàng nhận diện các kiểu cầu thủ khác nhau: cầu thủ tấn công, cầu thủ hỗ trợ, và cầu thủ phòng ngự.

Kết quả Phân cụm

Kết quả phân cụm cho thấy số lượng cầu thủ trong mỗi cụm như sau:

- Cụm 0: X cầu thủ
- Cụm 1: Y cầu thủ
- Cụm 2: Z cầu thủ

Em đã trực quan hóa các cụm cầu thủ dựa trên chỉ số kỳ vọng bàn thắng (xG) và số bàn thắng (Gls). Biểu đồ phân cụm cho thấy rõ ràng các nhóm cầu thủ khác nhau, với các màu sắc khác nhau đại diện cho từng cụm.



*EM đã áp dụng thuật toán PCA để giảm số chiều dữ liệu xuống còn 2 chiều. Điều này cho phép chúng tôi dễ dàng hình dung phân phối dữ liệu trong không gian 2D. Kết quả cho thấy sự phân tách rõ ràng giữa các nhóm cầu thủ, củng cố cho kết quả phân cụm trước đó.

