



---

## **Đồ án kết thúc học phần Trực quan hóa dữ liệu**

---

### **Project 1 - Sports Data Analysis**

**Giảng viên: TS. Tô Đức Khánh**

**Nhóm thực hiện: 13**

Nguyễn Hoàng Linh	21110333
Trịnh Hoàng Phú	21110370
Trần Anh Quân	21110374
Lê Thanh Tiến	21110409
Đặng Anh Tuấn	21110434
Trần Phước Vinh	21110450

Ngày 15 tháng 1 năm 2025

## Mục lục

<b>1</b>	<b>Giới thiệu đề tài</b>	<b>1</b>
1.1	Tổng quan về bóng đá . . . . .	1
1.1.1	Lịch sử bóng đá . . . . .	1
1.1.2	Luật chơi cơ bản . . . . .	1
1.1.3	Các giải đấu nổi bật . . . . .	1
1.1.4	Ý nghĩa và ảnh hưởng . . . . .	1
1.1.5	Một số câu thủ huyền thoại . . . . .	2
1.1.6	Bóng đá tại Việt Nam . . . . .	2
1.2	Lý do chọn dữ liệu . . . . .	2
1.3	Mục tiêu dự án . . . . .	2
1.3.1	Mục tiêu tổng quát . . . . .	2
1.3.2	Mục tiêu chi tiết . . . . .	2
<b>2</b>	<b>Tiền xử lý dữ liệu</b>	<b>3</b>
2.1	Kiểm tra cấu trúc dữ liệu . . . . .	3
2.2	Chuẩn hóa định dạng dữ liệu . . . . .	3
<b>3</b>	<b>Lựa chọn và sử dụng biểu đồ để trực quan</b>	<b>5</b>
<b>4</b>	<b>Đánh giá</b>	<b>29</b>
<b>5</b>	<b>Tài liệu tham khảo</b>	<b>30</b>

# 1 Giới thiệu đề tài

## 1.1 Tổng quan về bóng đá

Bóng đá, còn được gọi là “*môn thể thao vua*”, là môn thể thao phổ biến nhất trên thế giới, thu hút hàng tỷ người hâm mộ. Trò chơi được chơi giữa hai đội, mỗi đội có 11 cầu thủ, bao gồm một thủ môn. Mục tiêu chính là đưa bóng vào lưới đối phương để ghi bàn.

### 1.1.1 Lịch sử bóng đá

- **Nguồn gốc:** Bóng đá hiện đại có nguồn gốc từ Anh vào thế kỷ 19.
- **Quốc tế hóa:** Năm 1904, Liên đoàn bóng đá quốc tế (FIFA) được thành lập, thúc đẩy bóng đá phát triển toàn cầu. Các giải đấu lớn như World Cup bắt đầu từ năm 1930.

### 1.1.2 Luật chơi cơ bản

- **Thời gian:** Một trận đấu kéo dài 90 phút, chia thành hai hiệp, mỗi hiệp 45 phút.
- **Sân bóng:** Sân cỏ hình chữ nhật, với khung thành ở hai đầu.
- **Luật cơ bản:**
  - Không được dùng tay chơi bóng (trừ thủ môn).
  - Khi phạm lỗi nghiêm trọng, cầu thủ có thể bị phạt thẻ vàng hoặc thẻ đỏ.
  - Luật việt vị được áp dụng để đảm bảo tính công bằng.

### 1.1.3 Các giải đấu nổi bật

- **Cấp câu lạc bộ:** UEFA Champions League, English Premier League, La Liga, Serie A, Bundesliga, v.v.
- **Cấp đội tuyển quốc gia:** FIFA World Cup, Euro, Copa América, Asian Cup.

### 1.1.4 Ý nghĩa và ảnh hưởng

- **Văn hóa:** Bóng đá là cầu nối văn hóa, đoàn kết cộng đồng.
- **Kinh tế:** Tạo ra doanh thu từ vé, quảng cáo, truyền hình và tài trợ.

### 1.1.5 Một số cầu thủ huyền thoại

- **Thế giới:** Pelé, Diego Maradona, Lionel Messi, Cristiano Ronaldo.
- **Việt Nam:** Lê Công Vinh, Nguyễn Quang Hải, Đoàn Văn Hậu.

### 1.1.6 Bóng đá tại Việt Nam

- **Yêu thích:** Môn bóng đá được yêu thích tại Việt Nam.
- **Hệ thống giải:** V-League, Cúp Quốc gia, hạng Nhất.

## 1.2 Lý do chọn dữ liệu

- Bộ dữ liệu phong phú về thông tin cầu thủ từ nhiều quốc gia và vị trí thi đấu khác nhau sẽ giúp phân tích sâu sắc về chiến thuật. Ví dụ: So sánh hiệu suất cầu thủ ở từng vị trí (tiền đạo, tiền vệ, hậu vệ, thủ môn) có thể hỗ trợ việc xây dựng đội hình tối ưu.
- Đưa ra các chiến thuật phù hợp dựa trên năng lực của từng cầu thủ, từ đó giúp đội bóng cải thiện hiệu suất thi đấu trên sân.
- Việc đánh giá, phân tích dữ liệu đã trở thành yếu tố quan trọng giúp các đội bóng, huấn luyện viên, và nhà đầu tư đưa ra quyết định về tuyển dụng, chiến thuật, và quản lý đội hình. Hơn hết, nó còn giúp phân tích các chỉ số cầu thủ như hiệu suất, giá trị tài chính, và các kỹ năng cá nhân để đưa ra các nhận định hữu ích.

## 1.3 Mục tiêu dự án

### 1.3.1 Mục tiêu tổng quát

Phân tích và trực quan hóa dữ liệu cầu thủ FIFA để khám phá các đặc điểm nổi bật, mối tương quan giữa các chỉ số và đưa ra những nhận định hữu ích.

### 1.3.2 Mục tiêu chi tiết

- **Tổng quan:** Cung cấp cái nhìn tổng quan về dữ liệu cầu thủ bóng đá toàn cầu.
- **Thống kê:** Phân tích chuyên sâu khả năng của cầu thủ theo vị trí, bao gồm:
  - Chân thuận theo vị trí.
  - Mối tương quan giữa lương, giá trị và danh tiếng.
  - Mức độ tiềm năng theo độ tuổi.

- Độ năng nổ trong tấn công và phòng ngự theo từng vị trí.
- **Câu lạc bộ:** Phân tích giá trị và tính chất của các câu lạc bộ thông qua:
  - Giá trị hợp đồng và cầu thủ.
  - Số cầu thủ gia nhập sau mỗi năm.
  - Quốc tịch của các cầu thủ trong câu lạc bộ.
  - Giá trị của câu lạc bộ.
  - Số lượng cầu thủ trong từng khoảng tuổi ở từng câu lạc bộ.
- **So sánh câu lạc bộ:** Đặc tính nổi bật và khác nhau giữa các câu lạc bộ, bao gồm:
  - Sự tương quan giữa quỹ lương và giá trị của từng câu lạc bộ.
  - Độ tuổi trung bình của từng câu lạc bộ.
  - Chiều cao, cân nặng ảnh hưởng đến khả năng đánh đầu và khả năng tranh chấp bóng.
- **Chỉ số cầu thủ:** Kỹ năng nổi bật của cầu thủ như:
  - Kỹ năng thăng bằng.
  - Kỹ năng chơi bóng, kiểm soát bóng của từng cầu thủ.
  - Kỹ năng cản bóng của từng thủ môn.
- **Giá trị cầu thủ:** Thấy được giá trị của các cầu thủ dựa vào lương, tài năng và giá trị mà cầu thủ mang lại.

## 2 Tiền xử lý dữ liệu

### 2.1 Kiểm tra cấu trúc dữ liệu

- Đảm bảo rằng các cột quan trọng như *Age*, *Overall*, *Value*, *Wage* có định dạng phù hợp.
- Phát hiện các giá trị bị thiếu hoặc không hợp lệ.
- Kiểm tra các giá trị bất thường, ví dụ: cầu thủ có tuổi quá nhỏ hoặc lương không hợp lý.

### 2.2 Chuẩn hóa định dạng dữ liệu

- Loại bỏ ký tự không cần thiết (như euro, M, K) và chuyển đổi sang kiểu số.

## Sử dụng cho Radar chart

Radar Chart đầu tiên trực quan 6 chỉ số của 1 hoặc nhiều cầu thủ bất kỳ (không phải thủ môn). Với 6 chỉ số được tính theo chuẩn công thức của tựa game FIFA19 như sau:

- **SHO (Shooting):**

$$\begin{aligned} \text{SHO} = & ([\text{Finishing}] \times 0.45 + [\text{ShotPower}] \times 0.20 \\ & + [\text{LongShots}] \times 0.20 + [\text{Volleys}] \times 0.10 \\ & + [\text{Penalties}] \times 0.05) \end{aligned}$$

- **PAS (Passing):**

$$\begin{aligned} \text{PAS} = & ([\text{ShortPassing}] \times 0.35 + [\text{Vision}] \times 0.30 \\ & + [\text{Crossing}] \times 0.20 + [\text{LongPassing}] \times 0.10 \\ & + [\text{Curve}] \times 0.05) \end{aligned}$$

- **PAC (Pace):**

$$\text{PAC} = ([\text{Acceleration}] \times 0.55 + [\text{SprintSpeed}] \times 0.45)$$

- **PHY (Physicality):**

$$\begin{aligned} \text{PHY} = & ([\text{Strength}] \times 0.35 + [\text{Stamina}] \times 0.30 \\ & + [\text{Jumping}] \times 0.20 + [\text{Aggression}] \times 0.15) \end{aligned}$$

- **DEF (Defense):**

$$\begin{aligned} \text{DEF} = & ([\text{Interceptions}] \times 0.30 + [\text{StandingTackle}] \times 0.30 \\ & + [\text{SlidingTackle}] \times 0.20 + [\text{Marking}] \times 0.15 \\ & + [\text{Aggression}] \times 0.05) \end{aligned}$$

- **DRI (Dribbling):**

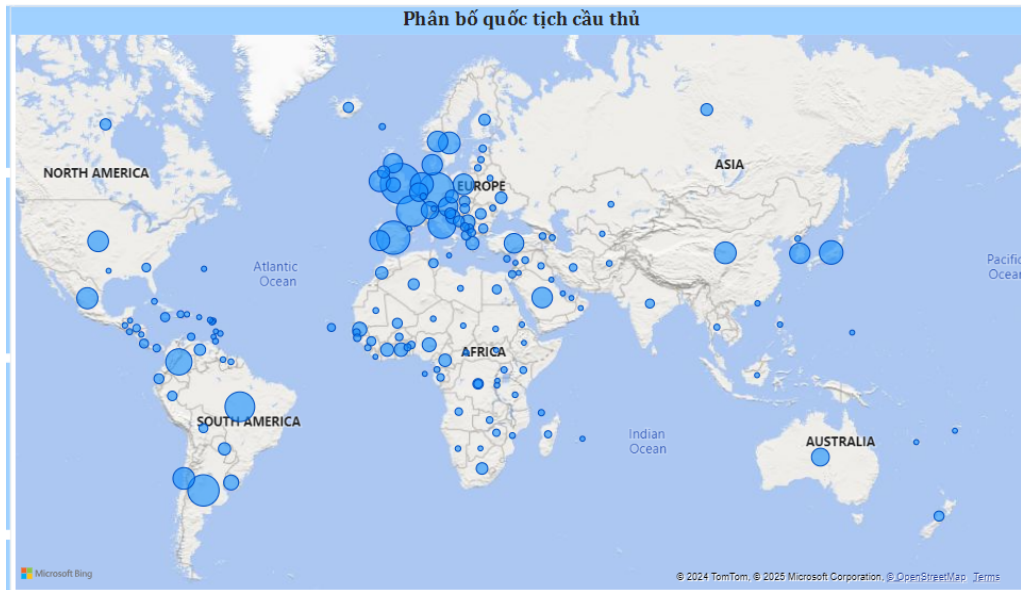
$$\begin{aligned} \text{DRI} = & ([\text{Dribbling}] \times 0.35 + [\text{BallControl}] \times 0.30 \\ & + [\text{Agility}] \times 0.20 + [\text{Balance}] \times 0.10 \\ & + [\text{Reactions}] \times 0.05) \end{aligned}$$

Radar Chart thứ 2 tương tự như biểu đồ đầu tiên nhưng chỉ dùng để trực quan hóa thủ môn. Biểu đồ cũng dùng 6 chỉ số bao gồm 5 chỉ số GK có sẵn và 1 chỉ số được tính theo công thức game:

$$\text{GK\_Speed} = \frac{[\text{Acceleration}] + [\text{SprintSpeed}]}{2}$$

### 3 Lựa chọn và sử dụng biểu đồ để trực quan

#### 1. Biểu đồ Map: Biểu đồ phân bố quốc tịch theo cầu thủ



Hình 1: Biểu đồ Map: Phân bố quốc tịch theo cầu thủ

#### Lý do chọn Map:

- **Hiển thị sự phân bố địa lý:** Trực quan hóa sự phân bố quốc tịch của các cầu thủ.
- **Tương quan với vị trí địa lý:** Xác định mối liên hệ giữa quốc gia và số lượng cầu thủ.
- **Sự phổ biến và ảnh hưởng của bóng đá:** Thể hiện sự phổ biến bóng đá ở các quốc gia.
- **Phân tích nhân khẩu học:** Hiểu rõ hơn về độ đa dạng và sự phong phú của các cầu thủ.

#### Nhận xét:

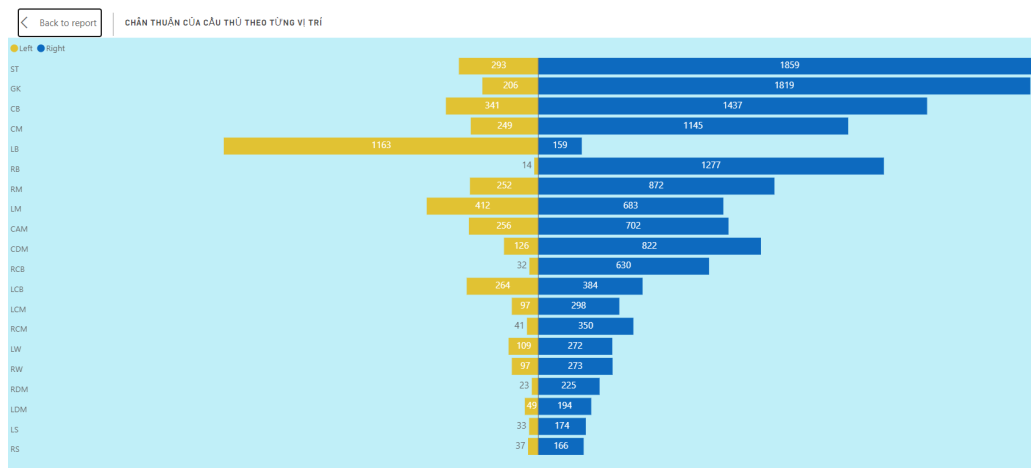
- **Châu Âu:** Khu vực tập trung nhiều cầu thủ nhất với các cụm lớn ở Anh, Đức, Tây Ban Nha, và Ý.

- **Nam Mỹ:** Đáng chú ý ở Brazil và Argentina.
- **Châu Phi, Châu Á và Bắc Mỹ:** Ít cụm hơn so với Châu Âu và Nam Mỹ.

### Kết luận:

- **Tập trung phát triển bóng đá:** Các khu vực như Châu Á, Châu Phi và Bắc Mỹ cần tập trung phát triển bóng đá hơn.

## 2. Biểu đồ thanh ngang (Horizontal Bar Chart): Chân thuận của các cầu thủ theo từng vị trí



Hình 2: Biểu đồ thanh ngang: Chân thuận của các cầu thủ theo từng vị trí

### Lý do chọn biểu đồ:

- **Hiển thị rõ ràng sự khác biệt:** Giúp người xem dễ dàng nhận ra các xu hướng hoặc mẫu.
- **Tính tương tác và linh hoạt:** Giúp người dùng dễ dàng tìm hiểu và khám phá dữ liệu.

### Nhận xét:

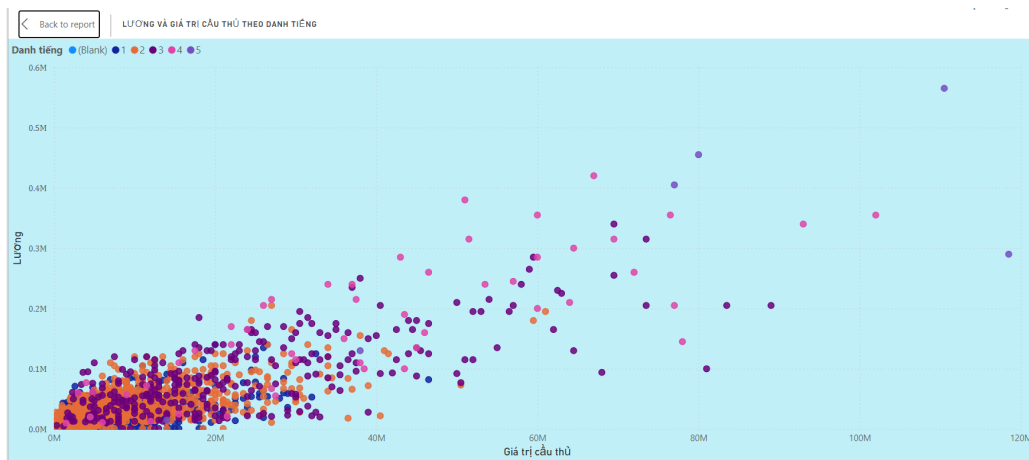
- **Cầu thủ thuận chân phải (Right):** Áp đảo ở hầu hết các vị trí.
- **Cầu thủ thuận chân trái (Left):** Nổi bật nhất ở vị trí LB (Hậu vệ cánh trái).



### Kết luận:

- **Chiến thuật, đào tạo, và tuyển chọn cầu thủ:** Số lượng cầu thủ thuận chân phải áp đảo so với chân trái, phản ánh xu hướng phổ biến trong bóng đá.

### 3. Biểu đồ phân tán (Scatter Plot): Biểu thị mối quan hệ giữa Lương, Giá trị, Danh tiếng



Hình 3: Biểu đồ phân tán: Mối quan hệ giữa Lương, Giá trị, Danh tiếng

### Lý do chọn biểu đồ:

- **Hiển thị mối quan hệ giữa hai biến số:** Dễ dàng nhận thấy mối quan hệ giữa "Giá trị" và "Lương trung bình".
- **Xác định xu hướng và phát hiện ngoại lệ:** Giúp xác định xu hướng tổng quát của dữ liệu và phát hiện các điểm dữ liệu ngoại lệ.
- **Phân loại dữ liệu:** Các điểm dữ liệu được mã hóa màu sắc theo "Danh tiếng quốc tế".

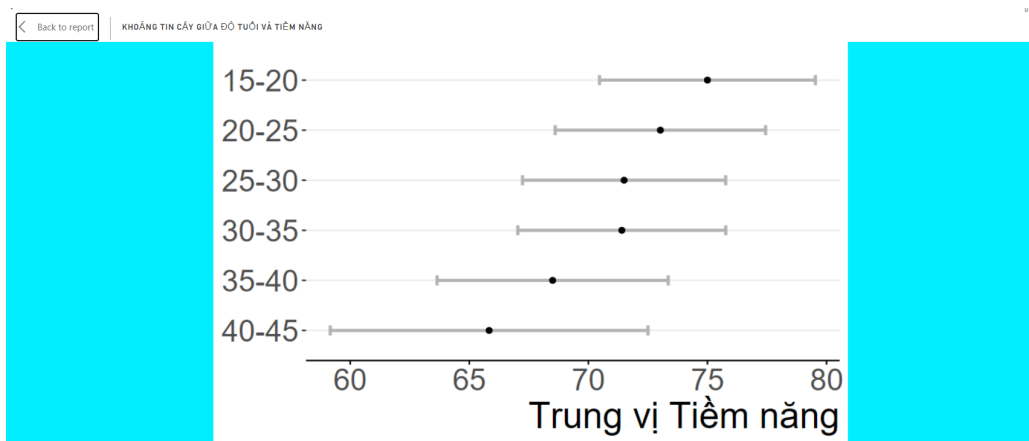
### Nhận xét:

- **Mối quan hệ giữa giá trị và lương:** Giá trị của cầu thủ càng cao thì mức lương cũng càng cao. Tuy nhiên, mối quan hệ này không hoàn toàn tuyến tính.

### Kết luận:

- **Giá trị và lượng của câu thủ:** Có mối liên hệ chặt chẽ nhưng không tuyệt đối, phụ thuộc vào nhiều yếu tố khác.

## 4. Biểu đồ khoảng tin cậy (Confidence Interval)



Hình 4: Biểu đồ khoảng tin cậy: Phân tích tiềm năng theo độ tuổi

### Lý do chọn biểu đồ:

- **Hiển thị sự biến động và xác định độ tin cậy của dữ liệu:** Giúp dễ dàng nhận thấy sự biến động của giá trị tiềm năng giữa các nhóm tuổi khác nhau.
- **Phát hiện xu hướng và so sánh trực quan:** Giúp xác định xu hướng tổng quát của tiềm năng theo độ tuổi.

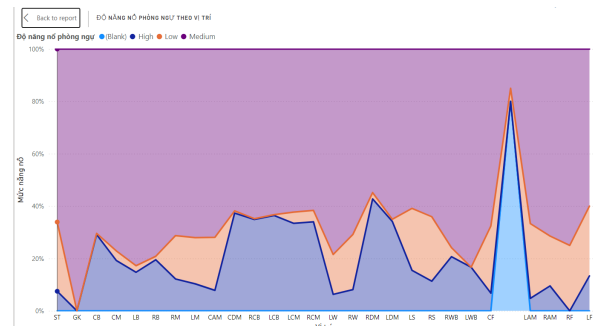
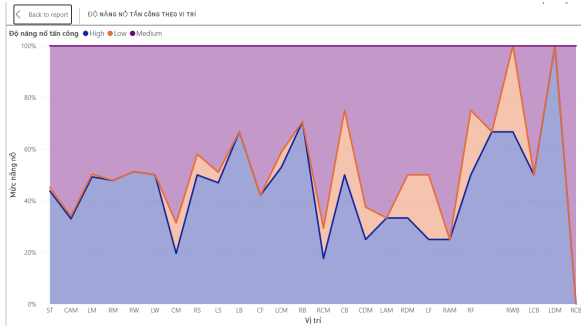
### Nhận xét:

- **Tiềm năng trung bình của các nhóm tuổi:** Khá ổn định, không có sự chênh lệch quá rõ rệt về tiềm năng giữa các nhóm tuổi.

### Kết luận:

- **Tuổi và tiềm năng:** Không có mối liên hệ rõ ràng giữa tuổi và tiềm năng, tiềm năng của một người không hoàn toàn phụ thuộc vào tuổi tác.

## 5. Biểu đồ vùng xếp chồng (Stacked Area Chart) Work rate (Tấn công và Phòng thủ) theo từng vị trí



Hình 5: Biểu đồ vùng xếp chồng: Work rate (Tấn công và Phòng thủ) theo từng vị trí

### Lý do chọn biểu đồ:

- **So sánh dễ dàng:** Giúp so sánh số lượng giữa các nhóm hoặc danh mục một cách dễ dàng và trực quan.
- **Dễ đọc và dễ hiểu:** Hiển thị rõ ràng và không bị chồng chéo, đặc biệt khi có nhiều danh mục hoặc tên danh mục dài.
- **Hiển thị rõ ràng sự khác biệt:** Giúp người xem nhận ra các xu hướng hoặc mẫu một cách rõ ràng.
- **Tính tương tác và linh hoạt:** Thường có tính tương tác cao, giúp người dùng dễ dàng khám phá dữ liệu.

### 5.1. Attack of Work Rate theo từng vị trí (bên trái):

#### Nhận xét:

- **Tiền đạo (ST), Tiền đạo cánh (LW, RW):** Hầu hết có chỉ số công thủ ở mức trung bình hoặc cao, thường được yêu cầu tham gia tấn công nhiều hơn là phòng thủ.
- **Thủ môn (GK):** Tất cả đều có chỉ số công thủ ở mức thấp, vì nhiệm vụ chính là phòng thủ.
- **Hậu vệ (CB, LB, RB):** Thường có chỉ số công thủ ở mức trung bình hoặc thấp, đặc biệt là trung vệ, ưu tiên phòng thủ hơn tấn công.

- **Tiền vệ (CM, CDM, LM, RM):** Có sự phân hóa rõ rệt về chỉ số công thủ, tiền vệ trung tâm (CM) ở mức trung bình, tiền vệ phòng ngự (CDM) thiên về phòng thủ, tiền vệ cánh (LM, RM) thiên về tấn công.

#### **Kết luận:**

- **Chỉ số công thủ phụ thuộc vào vị trí:** Mỗi vị trí trên sân có yêu cầu khác nhau về chỉ số công thủ, phản ánh trong sự đa dạng của các cầu thủ.
- **Sự đa dạng trong mỗi vị trí:** Các cầu thủ có những phẩm chất khác nhau để đáp ứng yêu cầu chiến thuật của đội bóng.
- **Quan hệ giữa tấn công và phòng thủ:** Biểu đồ cho thấy sự cân bằng giữa tấn công và phòng thủ ở các vị trí khác nhau, các đội bóng thường xây dựng đội hình với sự kết hợp hài hòa giữa các cầu thủ có khả năng tấn công và phòng thủ tốt.

#### **5.2. Defense of Work Rate theo từng vị trí (bên phải):**

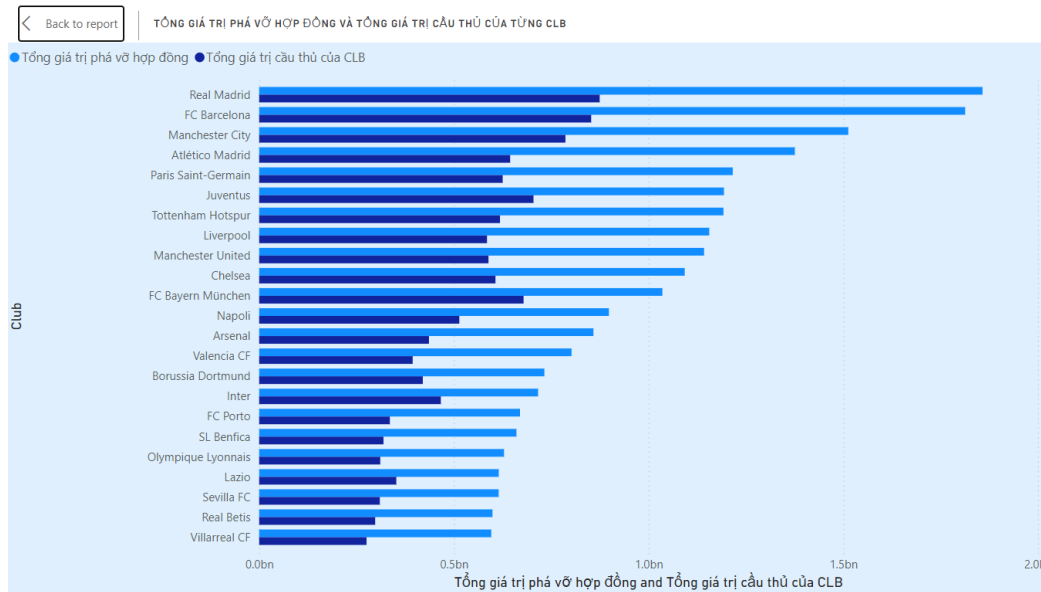
##### **Nhận xét:**

- **Thủ môn (GK):** Tất cả có chỉ số phòng ngự ở mức thấp nhất, nhiệm vụ chính là ngăn chặn đối phương ghi bàn.
- **Hậu vệ (CB, LB, RB):** Hầu hết có chỉ số phòng ngự ở mức trung bình hoặc cao, nhiệm vụ chính là phòng ngự và bảo vệ khung thành.
- **Tiền vệ (CM, CDM, LM, RM):** Có sự phân hóa rõ rệt về khả năng phòng ngự, tiền vệ phòng ngự (CDM) có chỉ số phòng ngự cao hơn so với các tiền vệ khác.
- **Tiền đạo (ST, LW, RW):** Thường có chỉ số phòng ngự thấp nhất, nhiệm vụ chính là ghi bàn, nên khả năng tấn công được ưu tiên hơn.

#### **Kết luận:**

- **Khả năng phòng ngự phụ thuộc vào vị trí:** Mỗi vị trí trên sân có yêu cầu khác nhau về khả năng phòng ngự, phản ánh trong sự bố trí cầu thủ.
- **Sự đa dạng trong mỗi vị trí:** Các cầu thủ có những phẩm chất khác nhau để đáp ứng yêu cầu chiến thuật của đội bóng.
- **Cân bằng giữa tấn công và phòng ngự:** Biểu đồ cho thấy sự cân bằng giữa tấn công và phòng ngự ở các vị trí khác nhau, các đội bóng thường xây dựng đội hình với sự kết hợp hài hòa giữa các cầu thủ có khả năng tấn công và phòng ngự tốt.

## 6. Biểu đồ thanh cụm (Clustered Bar Chart): Tổng giá trị phá vỡ hợp đồng và tổng giá trị cầu thủ theo từng CLB



Hình 6: Biểu đồ thanh cụm: Tổng giá trị phá vỡ hợp đồng và tổng giá trị cầu thủ theo từng CLB

### Lý do chọn biểu đồ:

- **So sánh dễ dàng giữa các nhóm:** Biểu đồ thanh cụm cho phép so sánh dễ dàng và trực quan giữa các nhóm và các hạng mục trong từng nhóm.
- **Hiển thị nhiều thông tin trong một biểu đồ:** Giúp hiển thị nhiều biến số và hạng mục trong một biểu đồ duy nhất.
- **Phân tích chi tiết:** Cho phép phân tích chi tiết từng nhóm và so sánh chúng với các nhóm khác.
- **Trực quan và dễ hiểu:** Biểu đồ thanh cụm giúp người xem dễ dàng nhận ra các mẫu và xu hướng dữ liệu.

### Nhận xét:

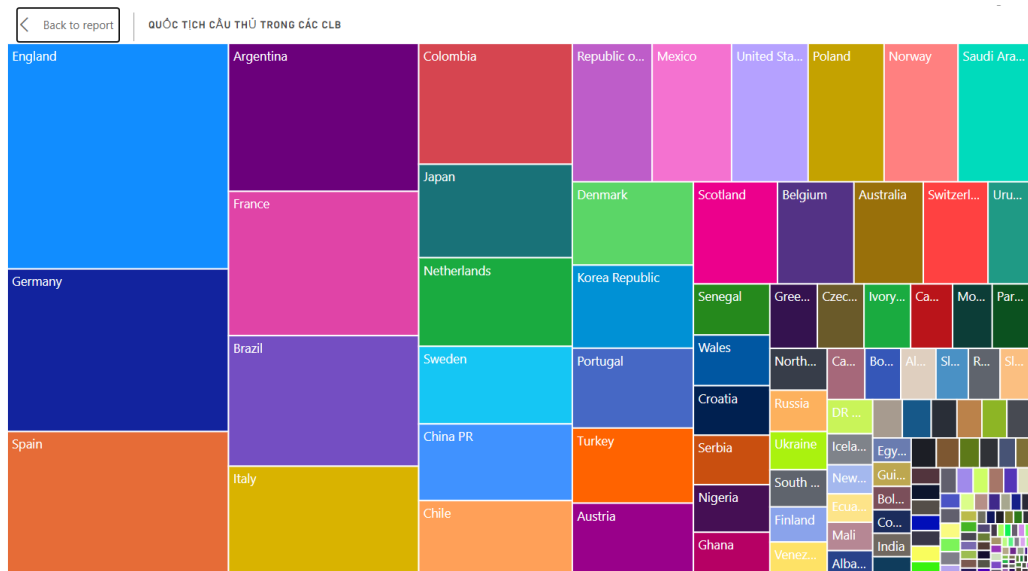
- **Dẫn đầu:** Real Madrid và FC Barcelona dẫn đầu về cả tổng giá trị phá vỡ hợp đồng và tổng giá trị cầu thủ.
- **Sự chênh lệch:** Juventus và Manchester United có sự chênh lệch khá lớn giữa hai chỉ số.

- **Các CLB nhỏ hơn:** Valencia CF, Sevilla FC, và Villarreal CF có cả hai chỉ số ở mức thấp hơn so với nhóm dẫn đầu.

### Kết luận:

- **Chiến lược tài chính và mức độ rủi ro:** Sự khác biệt giữa tổng giá trị cầu thủ và giá trị phá vỡ hợp đồng có thể phản ánh chiến lược quản lý tài chính và mức độ rủi ro của từng CLB.
- **CLB nhỏ hoặc trung bình:** Các CLB nhỏ hoặc trung bình như Sevilla, Villarreal, hay Real Betis không có tổng giá trị tài sản cao như các đội bóng lớn.

## 7. Treemap: Quốc tịch của các cầu thủ trong các CLB



Hình 7: Treemap: Quốc tịch của các cầu thủ trong các CLB

### Lý do chọn biểu đồ:

- **Hiển thị dữ liệu phân cấp:** Treemap giúp hiển thị dữ liệu phân cấp một cách trực quan.
- **Sử dụng không gian hiệu quả:** Treemap tận dụng tối đa không gian có sẵn.
- **So sánh trực quan:** Cho phép so sánh các phần tử trong cùng một cấp độ phân cấp.
- **Hiển thị nhiều thông tin:** Treemap có thể hiển thị đồng thời nhiều biến số bằng cách sử dụng màu sắc và kích thước ô.

- **Phát hiện xu hướng và mẫu:** Giúp người xem nhận biết các xu hướng và mẫu dữ liệu nhanh chóng.
- **Tương tác và khám phá:** Treemap thường được sử dụng trong các ứng dụng tương tác.

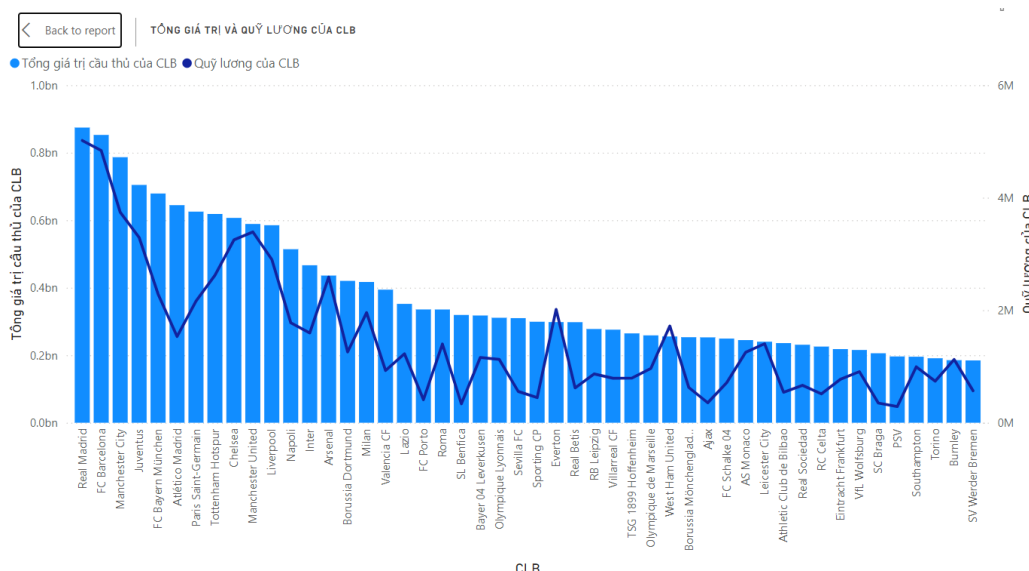
#### Nhận xét:

- **Quốc gia dẫn đầu:** Anh có số lượng cầu thủ lớn nhất, tiếp theo là Pháp, Đức, Tây Ban Nha, Brazil, và Ý.
- **Châu Âu chiếm tỷ lệ lớn:** Phản ánh vai trò chủ đạo của châu Âu trong bóng đá chuyên nghiệp.
- **Nam Mỹ có sự hiện diện đáng kể:** Brazil và Argentina là nguồn cung cầu thủ chất lượng cao.
- **Châu Phi và châu Á:** Có sự góp mặt nhưng với số lượng thấp hơn.

#### Kết luận:

- **Nguồn cung cầu thủ toàn cầu:** Phân tích này giúp hiểu rõ hơn về nguồn cầu thủ toàn cầu, có thể hỗ trợ trong việc xây dựng chiến lược tìm kiếm và phát triển tài năng bóng đá.

## 8. Biểu đồ cột và đường (Line and Clustered Column Chart): Tổng giá trị và quỹ lương của CLB



Hình 8: Biểu đồ cột và đường: Tổng giá trị và quỹ lương của CLB

### Lý do chọn biểu đồ:

- **Kết hợp thông tin chi tiết:** Biểu đồ kết hợp giữa cột cụm và đường giúp hiển thị hai loại dữ liệu khác nhau trên cùng một biểu đồ.
- **So sánh dễ dàng:** Dễ dàng so sánh tổng giá trị và quỹ lương của các CLB khác nhau.
- **Hiển thị xu hướng:** Biểu đồ đường giúp hiển thị sự thay đổi hoặc xu hướng của một trong hai chỉ số.
- **Phân tích đa chiều:** Cung cấp khả năng phân tích đa chiều, giúp hiểu rõ hơn mối quan hệ giữa tổng giá trị và quỹ lương của các CLB.
- **Dễ đọc và trực quan:** Biểu đồ cột cụm giúp phân biệt và so sánh các chỉ số rõ ràng.

### Nhận xét:

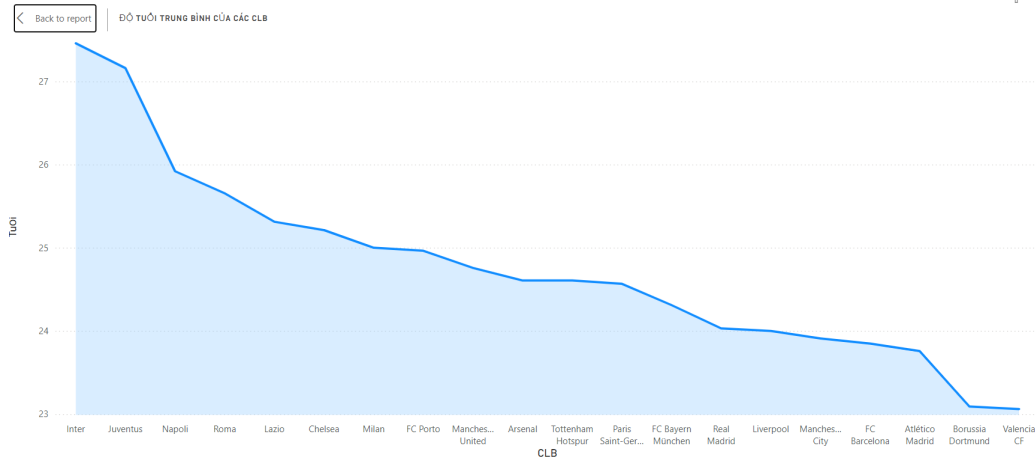
- **Quỹ lương:** Paris Saint-Germain (PSG) có quỹ lương cao nhưng tổng giá trị cầu thủ không vượt trội so với Real Madrid hay Barcelona.
- **Atlético Madrid:** Có tổng giá trị cầu thủ cao nhưng quỹ lương tương đối thấp.
- **Nhóm giữa:** Tottenham, Liverpool, Arsenal, và Inter Milan có tổng giá trị cầu thủ từ 0.3 đến 0.5 tỷ euro với quỹ lương tương đương.
- **CLB nhỏ:** Real Sociedad, Athletic Bilbao, và Leicester City có tổng giá trị cầu thủ thấp hơn đáng kể nhưng một số CLB nhỏ hơn như RB Leipzig, Sporting CP lại có quỹ lương vượt trội.

### Kết luận:

- **Chiến lược tài chính:** Real Madrid, Barcelona, và Manchester City không chỉ có đội hình chất lượng cao mà còn phản ánh sức mạnh tài chính vượt trội.
- **Quỹ lương và giá trị đội hình:** Không hoàn toàn tỷ lệ thuận, thể hiện chiến lược quản lý tài chính và chi trả cho cầu thủ của từng CLB.
- **Sự chênh lệch:** CLB hàng đầu châu Âu tập trung nguồn lực lớn, tạo ra sự chênh lệch đáng kể so với nhóm CLB nhỏ hơn.
- **PSG:** Quỹ lương cao do sở hữu nhiều ngôi sao như Messi, Neymar, và Mbappe.
- **CLB nhỏ hơn:** Real Sociedad, Athletic Bilbao, và Leicester City có tổng giá trị cầu thủ thấp hơn nhưng một số CLB nhỏ hơn như RB Leipzig, Sporting CP lại có quỹ lương vượt trội.



## 9. Biểu đồ khu vực (Area Chart): Độ tuổi trung bình của các CLB



Hình 9: Biểu đồ khu vực: Độ tuổi trung bình của các CLB

### Lý do chọn biểu đồ:

- **Hiển thị xu hướng theo thời gian:** Giúp nhận biết sự thay đổi của độ tuổi trung bình của các CLB theo thời gian hoặc theo từng CLB.
- **So sánh trực quan:** Dễ dàng so sánh độ tuổi trung bình giữa các CLB khác nhau.
- **Tích lũy thông tin:** Thể hiện sự tích lũy của dữ liệu, giúp người xem hiểu rõ hơn về sự tăng trưởng hoặc giảm sút của độ tuổi trung bình.
- **Phát hiện xu hướng và mẫu:** Giúp phát hiện các xu hướng dài hạn và mẫu dữ liệu.
- **Dễ đọc và trực quan:** Giúp người xem dễ dàng nhận biết sự biến động của dữ liệu.

### Nhận xét:

- **Độ tuổi cao nhất:** CLB Inter có độ tuổi trung bình cao nhất, trên 27 tuổi.
- **Độ tuổi thấp nhất:** Valencia CF có độ tuổi trung bình thấp nhất, chỉ hơn 23 tuổi.
- **Xu hướng giảm dần:** Từ Inter đến Valencia CF, độ tuổi trung bình của các cầu thủ có xu hướng giảm dần.
- **CLB nổi bật:** Borussia Dortmund, Atlético Madrid, FC Barcelona và Manchester City có độ tuổi trung bình thấp hơn.

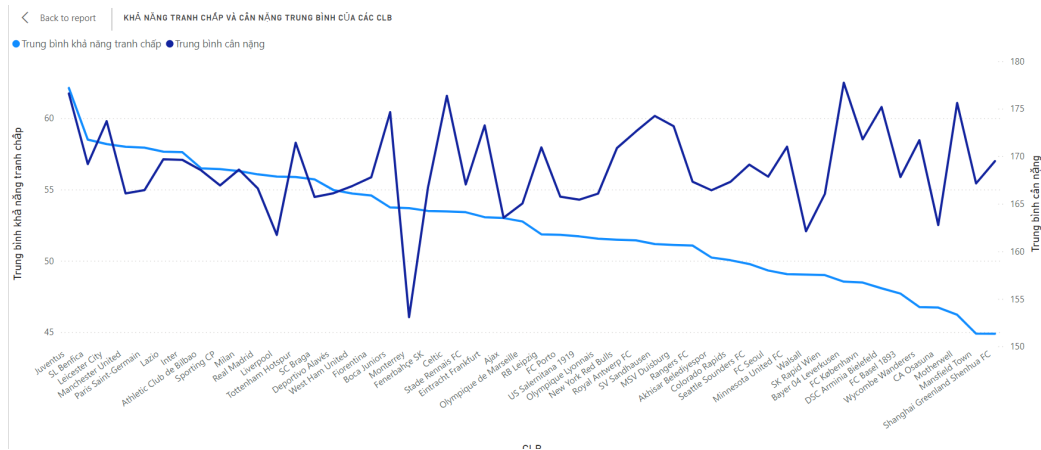
- [illegible]

- **Mối quan hệ giữa chiều cao và khả năng đánh đầu:** CLB có chiều cao trung bình cao hơn thường có khả năng đánh đầu tốt hơn, nhưng có một số ngoại lệ.

### Kết luận:

- **Chiều cao và khả năng đánh đầu:** Những CLB có chiều cao trung bình lớn hơn thường có khả năng đánh đầu tốt hơn, điều này có thể phản ánh sự lựa chọn chiến thuật và ưu tiên trong việc tuyển dụng cầu thủ. Tuy nhiên, cũng có những ngoại lệ cần được xem xét kỹ lưỡng hơn.

## 11. Biểu đồ đường (Line Chart): Biểu diễn khả năng tranh chấp và cân nặng giữa các CLB



Hình 11: Biểu đồ đường: Biểu diễn khả năng tranh chấp và cân nặng giữa các CLB

### Lý do chọn biểu đồ:

- **Hiển thị xu hướng theo thời gian:** Hiệu quả trong việc hiển thị dữ liệu biến đổi theo thời gian, giúp nhận biết xu hướng dài hạn hoặc ngắn hạn.
- **So sánh giá trị dữ liệu:** Dễ dàng so sánh các giá trị dữ liệu tại các thời điểm khác nhau.
- **Phát hiện biến động:** Giúp phát hiện các biến động hoặc dao động trong dữ liệu.
- **Phân tích nhiều chuỗi dữ liệu:** Cho phép hiển thị nhiều chuỗi dữ liệu trên cùng một biểu đồ.
- **Dễ đọc và trực quan:** Hiển thị các thông tin một cách rõ ràng và dễ hiểu.

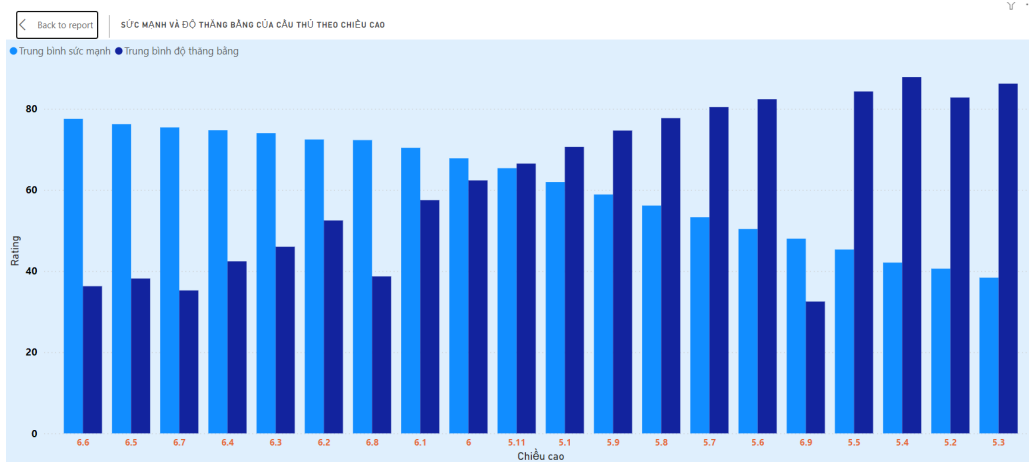
### Nhận xét:

- **Khả năng tranh chấp cao và cân nặng:** Các CLB như Juventus, Manchester United và Paris Saint-Germain có chỉ số khả năng tranh chấp trung bình cao nhất và cân nặng trung bình cao hơn so với các CLB khác.
- **Mối quan hệ giữa cân nặng và khả năng tranh chấp:** Có sự tương quan giữa cân nặng trung bình và khả năng tranh chấp của các cầu thủ. CLB có cân nặng trung bình cao thường có khả năng tranh chấp tốt hơn.
- **Ngoại lệ:** Một số CLB như Borussia Dortmund và Real Sociedad có chỉ số khả năng tranh chấp thấp hơn dù cân nặng trung bình của các cầu thủ không quá thấp.

### Kết luận:

- **Mối quan hệ giữa cân nặng và khả năng tranh chấp:** Cân nặng trung bình của các cầu thủ có liên quan đến khả năng tranh chấp trên sân. CLB có cân nặng trung bình cao hơn thường có khả năng tranh chấp tốt hơn. Điều này có thể phản ánh sự lựa chọn chiến thuật của các đội bóng. Tuy nhiên, cũng có những ngoại lệ và cần các nghiên cứu sâu hơn để hiểu rõ mối quan hệ này.

## 12. Biểu đồ thanh cụm (Clustered Column Chart): Sức mạnh và độ thăng bằng của các cầu thủ theo chiều cao



Hình 12: Biểu đồ thanh cụm: Sức mạnh và độ thăng bằng của các cầu thủ theo chiều cao

### Lý do chọn biểu đồ:

- **Hiển thị sự khác biệt giữa các nhóm:** Giúp so sánh trực tiếp các chỉ số sức mạnh và độ thăng bằng theo chiều cao.
- **Dễ hiểu và trực quan:** Hiển thị thông tin một cách rõ ràng và dễ hiểu.
- **So sánh trực quan:** Giúp nhận diện nhanh chóng sự khác biệt và xu hướng giữa các chỉ số.

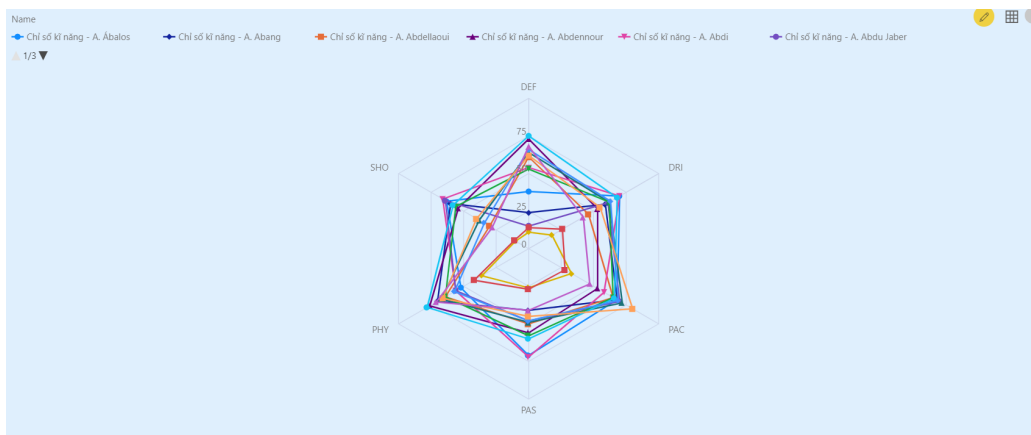
### Nhận xét:

- **Sức mạnh và độ thăng bằng:** Các cầu thủ cao hơn thường có chỉ số sức mạnh cao hơn.
- **Chiều cao và khả năng cân bằng:** Các cầu thủ có chiều cao trung bình (khoảng 5.9 feet) có chỉ số độ thăng bằng cao nhất.

### Kết luận:

- **Mối quan hệ giữa chiều cao, sức mạnh và độ thăng bằng:** Các cầu thủ cao thường có sức mạnh tốt hơn, trong khi các cầu thủ thấp hơn có độ thăng bằng tốt hơn. Thông tin này rất hữu ích trong việc phân tích và lựa chọn cầu thủ phù hợp cho các vị trí và vai trò chiến thuật khác nhau trong đội hình.

## 13. Biểu đồ radar: Các chỉ số kỹ năng của các cầu thủ



Hình 13: Biểu đồ radar: Các chỉ số kỹ năng của các cầu thủ

### Lý do chọn biểu đồ:

- **Hiển thị nhiều biến số cùng lúc:** Giúp dễ dàng so sánh các chỉ số kỹ năng của từng cầu thủ.
- **So sánh trực quan:** Hữu ích để so sánh trực tiếp các cầu thủ với nhau.
- **Nhận diện điểm mạnh và yếu:** Giúp nhận diện nhanh chóng điểm mạnh và điểm yếu của từng cầu thủ.
- **Dễ hiểu và dễ đọc:** Hiển thị thông tin phức tạp một cách trực quan và dễ hiểu.
- **Hiển thị mối quan hệ giữa các biến số:** Giúp hiểu rõ hơn về dữ liệu.

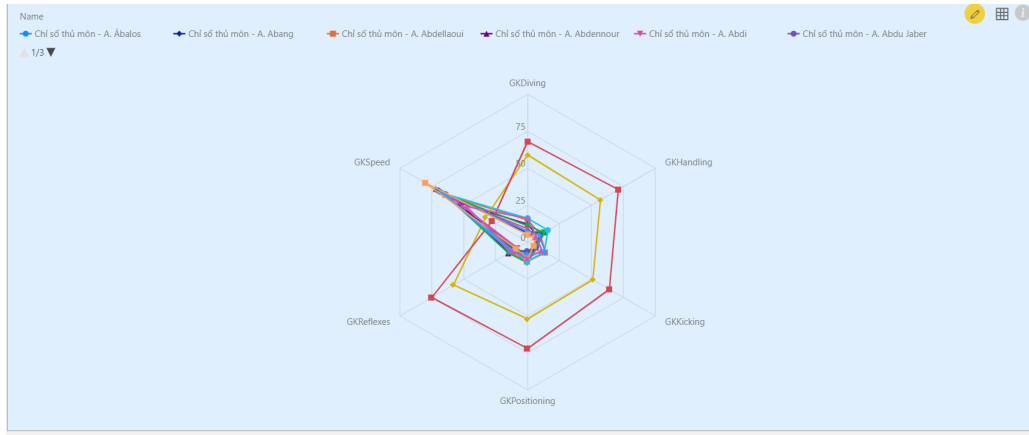
### Nhận xét:

- Biểu đồ radar với 6 chỉ số cung cấp cái nhìn tổng quan toàn diện, dễ dàng nhận biết cầu thủ có thiên hướng phòng ngự, tấn công, hay toàn diện.
- Hình dạng của biểu đồ (cân đối hay lệch) phản ánh rõ ràng phong cách chơi hoặc đặc điểm nổi bật của cầu thủ. Nếu biểu đồ lệch hẳn về một hoặc hai chỉ số: Điều này cho thấy cầu thủ có chuyên môn cao trong một số kỹ năng cụ thể, ví dụ, một tiền đạo có chỉ số "SHO" và "PAC" vượt trội nhưng chỉ số DEF thấp.
- Nếu biểu đồ cân đối: Đây là một cầu thủ toàn diện, có thể đảm nhận nhiều vai trò trên sân.

### Kết luận:

- **Sự khác biệt và tương quan giữa các chỉ số kỹ năng:** Biểu đồ radar này cung cấp một cái nhìn trực quan về sự khác biệt và tương quan giữa các chỉ số kỹ năng của các cầu thủ. Nó cho phép đánh giá và so sánh một cách toàn diện các khía cạnh kỹ thuật khác nhau của cầu thủ.

## 14. Biểu đồ radar: Kỹ năng thủ môn của các cầu thủ



Hình 14: Biểu đồ radar: Kỹ năng thủ môn của các cầu thủ

### Lý do chọn biểu đồ:

- Biểu đồ radar của thủ môn thể hiện các khía cạnh chuyên môn cần thiết cho vị trí này
- Hình dạng của biểu đồ giúp người chơi dễ dàng nhận ra phong cách và điểm mạnh đặc trưng của từng thủ môn, chẳng hạn như thiên về phản xạ nhanh hay có khả năng chơi chân tốt.
- Một thủ môn có hình dạng biểu đồ mở rộng mạnh về các chỉ số như Phản xạ và Đổ người thường phù hợp với phong cách chơi phòng ngự phản công nhanh, vì họ xử lý tốt trong các tình huống đối mặt.
- Nếu biểu đồ tập trung vào Ra vào và Phát bóng, thủ môn này thích hợp với chiến thuật pressing hoặc kiểm soát bóng từ phía sau.
- Một thủ môn có chỉ số đồng đều trong tất cả các khía cạnh cho thấy họ là người đáng tin cậy, phù hợp với hầu hết các lối chơi.

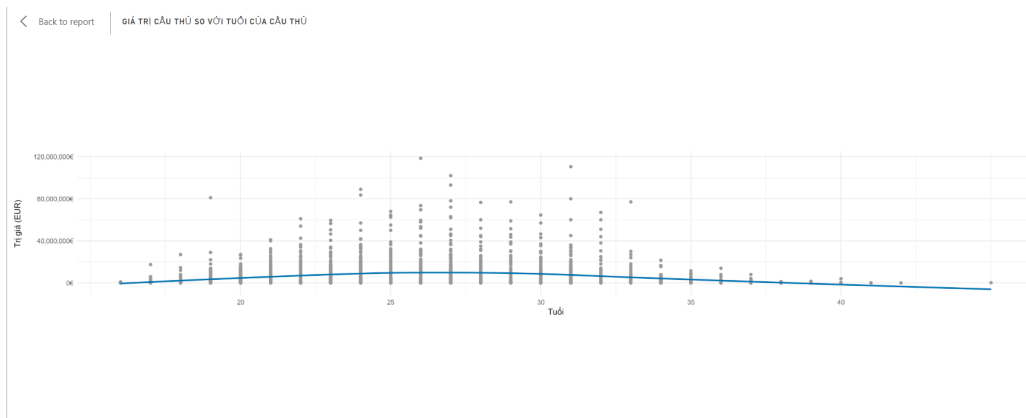
### Nhận xét:

- **Chỉ số nổi bật về các kỹ năng thủ môn:** Một số cầu thủ như A. Abang và A. Abdi có chỉ số GK Reflexes cao, thể hiện khả năng phản xạ nhanh.
- **Mối quan hệ giữa các kỹ năng:** Một số cầu thủ có sự cân bằng giữa các kỹ năng, trong khi những người khác có sự chênh lệch lớn giữa các chỉ số.

### Kết luận:

- **Sự khác biệt và tương quan giữa các kỹ năng thủ môn:** Biểu đồ radar này cung cấp cái nhìn tổng quan về sự khác biệt và tương quan giữa các kỹ năng thủ môn của các cầu thủ khác nhau. Nó cho phép đánh giá và so sánh một cách toàn diện các khía cạnh kỹ thuật khác nhau của thủ môn.

## 15. Biểu đồ giá trị cầu thủ so với độ tuổi của cầu thủ



Hình 15: Biểu đồ giá trị cầu thủ so với độ tuổi của cầu thủ

### Lý do chọn biểu đồ:

- **Hiển thị mối quan hệ giữa hai biến số:** Giúp dễ dàng nhận thấy mối quan hệ giữa "Tuổi" và "Trị giá", phân tích cách một biến số có thể ảnh hưởng đến biến số khác.
- **Xác định xu hướng:** Giúp xác định xu hướng tổng quát của dữ liệu, như liệu có mối quan hệ tuyến tính hay phi tuyến tính giữa các biến số.
- **Phát hiện ngoại lệ:** Dễ dàng phát hiện các điểm dữ liệu ngoại lệ hoặc bất thường, nhận ra những trường hợp đặc biệt trong dữ liệu.
- **Phân loại dữ liệu:** Các điểm dữ liệu có thể được mã hóa màu sắc hoặc hình dạng khác nhau để thể hiện thêm các chiều thông tin, giúp so sánh và phân loại dữ liệu theo nhiều nhóm khác nhau.
- **Dễ hiểu và trực quan:** Trình bày dữ liệu rõ ràng và trực quan, giúp người xem dễ dàng nắm bắt thông tin và rút ra các kết luận hữu ích.



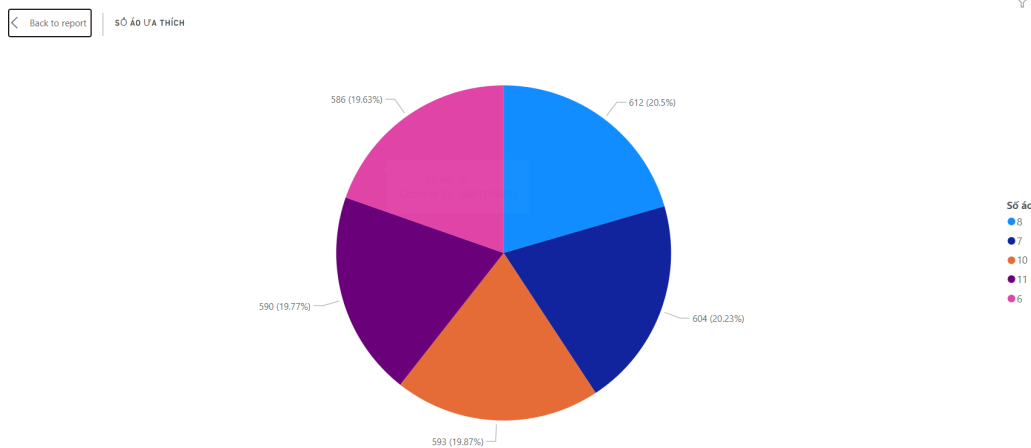
### Nhận xét:

- **Xu hướng tổng quan:** Biểu đồ thể hiện mối quan hệ giữa giá trị cầu thủ (Trị giá) và độ tuổi (Tuổi) của họ. Các cầu thủ trẻ hơn (từ 16 đến khoảng 26 tuổi) có xu hướng có giá trị cao hơn so với các cầu thủ lớn tuổi.
- **Giá trị cao nhất:** Các cầu thủ có giá trị cao nhất thường ở độ tuổi từ 20 đến 25, với một số cầu thủ đạt giá trị tối đa khoảng 120 triệu Euro.
- **Biến động giá trị theo tuổi:** Có sự biến động lớn về giá trị cầu thủ ở các độ tuổi trẻ (16-20 tuổi), cho thấy sự cạnh tranh và tiềm năng phát triển cao.

### Kết luận:

- **Giá trị cầu thủ và tuổi tác:** Các cầu thủ trẻ hơn có xu hướng có giá trị thị trường cao hơn, điều này có thể phản ánh tiềm năng phát triển và khả năng đầu tư vào tương lai. Ngược lại, giá trị của các cầu thủ lớn tuổi giảm dần, cho thấy rằng tuổi tác có ảnh hưởng tiêu cực đến giá trị thị trường của họ. Thông tin này hữu ích cho việc quản lý đội bóng và chiến lược chuyển nhượng cầu thủ.

## 16. Biểu đồ tròn (Pie Chart): Tỷ lệ giữa các màu áo



Hình 16: Biểu đồ tròn: Tỷ lệ giữa các màu áo

### Lý do chọn biểu đồ:

- **So sánh tỷ lệ phần trăm:** Hiệu quả trong việc hiển thị tỷ lệ phần trăm của từng phần tử trong tổng thể.

- **Trực quan và dễ hiểu:** Cung cấp cái nhìn tổng quan nhanh chóng và dễ hiểu.
- **Hiển thị sự phân chia dữ liệu:** Giúp hiển thị rõ ràng sự phân chia của các phần tử.
- **Dễ dàng phân biệt:** Mỗi phần có thể được mã hóa màu sắc khác nhau, giúp dễ dàng phân biệt.

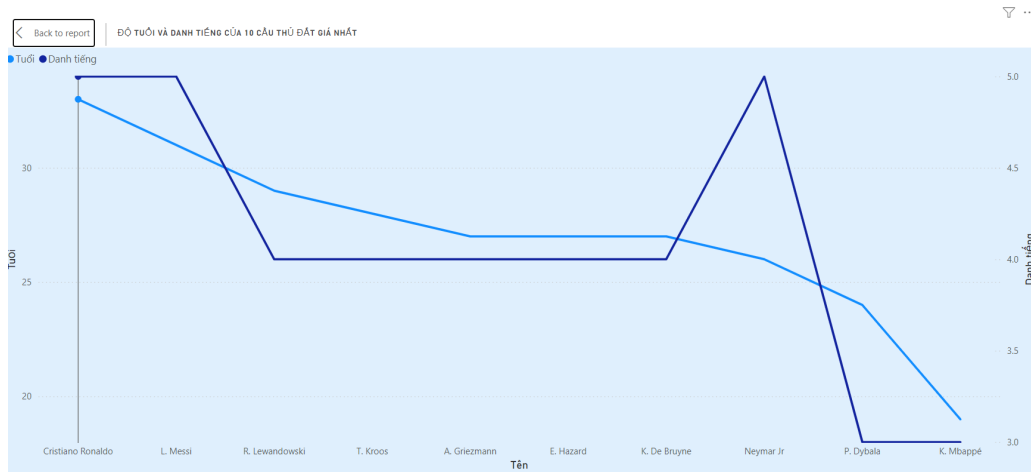
**Nhận xét:**

- **Số áo yêu thích phổ biến:**
  - Số áo 8 chiếm tỷ lệ cao nhất với 20.5
  - Số áo 7 cũng có sự ưa chuộng cao, chiếm 20.23
  - Các số áo 10, 11 và 6 đều có tỷ lệ ưa thích tương đối đều nhau, khoảng 19.63% - 19.87%.
- **Phân bố đồng đều:** Tỷ lệ ưa thích các số áo khá đồng đều, không có sự chênh lệch quá lớn giữa các số áo.

**Kết luận:**

- **Sở thích số áo:** Số áo 8 và 7 là những số áo được ưa chuộng nhất. Các số áo khác như 10, 11 và 6 cũng có sự ưa thích tương đương. Biểu đồ này hữu ích trong việc hiểu rõ hơn về xu hướng và sở thích của cộng đồng bóng đá.

## 17. Biểu đồ đường (Line Chart): Độ tuổi và danh tiếng của 10 cầu thủ đắt giá nhất



Hình 17: Biểu đồ đường: Độ tuổi và danh tiếng của 10 cầu thủ đắt giá nhất

### Lý do chọn biểu đồ:

- **Hiện thị xu hướng rõ ràng theo thời gian:** Hiệu quả trong việc hiển thị sự biến đổi của dữ liệu theo thời gian.
- **So sánh giá trị dữ liệu:** Dễ dàng so sánh các giá trị dữ liệu tại các thời điểm khác nhau.
- **Phát hiện biến động:** Giúp phát hiện các biến động hoặc dao động trong dữ liệu.
- **Phân tích nhiều chuỗi dữ liệu:** Cho phép hiển thị nhiều chuỗi dữ liệu trên cùng một biểu đồ.
- **Dễ đọc và trực quan:** Hiển thị các thông tin một cách rõ ràng và dễ hiểu.

### Nhận xét:

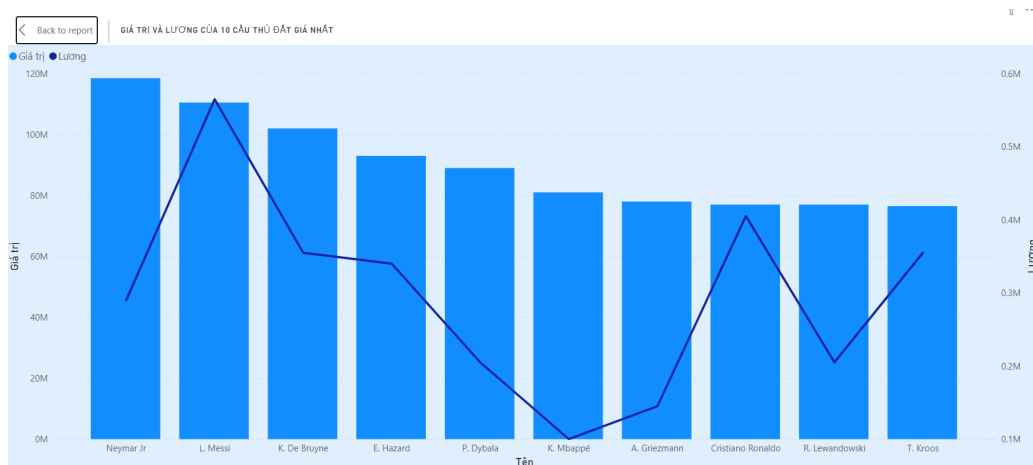
- **Sự nổi bật về tuổi và danh tiếng:**
  - Cristiano Ronaldo và Lionel Messi có độ tuổi lớn hơn nhưng vẫn duy trì được danh tiếng rất cao.
  - Cầu thủ trẻ nhất trong danh sách là Kylian Mbappé, người có danh tiếng rất cao ở tuổi trẻ.
- **Mối liên hệ giữa tuổi và danh tiếng:**

- Danh tiếng không giảm đi nhiều theo tuổi, đặc biệt là những cầu thủ hàng đầu như Ronaldo và Messi.
- Danh tiếng của các cầu thủ trẻ như Mbappé và Dybala rất cao, cho thấy tiềm năng phát triển lớn trong tương lai.

### Kết luận:

- **Tuổi và danh tiếng:** Danh tiếng của các cầu thủ không chỉ phụ thuộc vào tuổi tác mà còn phụ thuộc vào tài năng và thành tích của họ. Các cầu thủ như Cristiano Ronaldo và Lionel Messi vẫn duy trì được danh tiếng cao dù đã lớn tuổi. Các cầu thủ trẻ như Kylian Mbappé có tiềm năng phát triển danh tiếng lớn.

## 18. Biểu đồ cột và đường (Line and Clustered Column Chart): Mối quan hệ giữa giá trị và lương của 10 cầu thủ đắt giá nhất



Hình 18: Biểu đồ cột và đường: Mối quan hệ giữa giá trị và lương của 10 cầu thủ đắt giá nhất

### Lý do chọn biểu đồ:

- **Hiển thị nhiều loại dữ liệu:** Biểu đồ kết hợp hiển thị giá trị và mức lương của cầu thủ.
- **So sánh trực quan:** Giúp so sánh trực quan các giá trị dữ liệu.
- **Nhận diện xu hướng và mẫu:** Giúp nhận diện xu hướng tổng quát của dữ liệu.
- **Hiển thị thông tin chi tiết:** Kết hợp cả hai loại biểu đồ giúp hiển thị thông tin chi tiết và toàn diện hơn.

- **Phân tích hiệu quả:** Giúp hiểu rõ hơn mối quan hệ giữa hai biến số.

**Nhận xét:**

- **Giá trị cao nhất:**
  - Neymar Jr có giá trị cao nhất, xấp xỉ 180 triệu Euro.
  - Lionel Messi và Kevin De Bruyne cũng có giá trị cao, lần lượt xấp xỉ 150 triệu Euro và 120 triệu Euro.
- **Lương cao nhất:**
  - Lionel Messi có lương cao nhất, gần 70 triệu Euro.
  - Cristiano Ronaldo và Neymar Jr cũng có lương cao, lần lượt khoảng 60 triệu Euro và 55 triệu Euro.
- **Sự khác biệt giữa giá trị và lương:**
  - Neymar Jr và K. Mbappé có giá trị cao nhưng lương không phải là cao nhất, thể hiện tiềm năng phát triển lớn của họ.
  - Lionel Messi và Cristiano Ronaldo có lương rất cao, phản ánh sự nổi tiếng và thành tích của họ.

## Kết luận chung

Bộ dữ liệu này mang đến cái nhìn toàn diện về đặc điểm và khả năng của các cầu thủ trên toàn thế giới. Thông qua việc phân tích, một số kết luận quan trọng đã được rút ra:

### 1. Sự đa dạng trong kỹ năng và giá trị cầu thủ

- Các cầu thủ có sự khác biệt rõ rệt về kỹ năng, chiều cao, cân nặng, và giá trị thị trường, phản ánh chiến lược và yêu cầu đa dạng trong bóng đá hiện đại.

### 2. Mối quan hệ giữa các biến số

- Chiều cao, sức mạnh, và độ thăng bằng:** Có mối liên hệ chặt chẽ, thể hiện vai trò quan trọng của thể chất trong hiệu suất thi đấu.
- Tuổi tác và giá trị thị trường:** Tuổi tác có ảnh hưởng lớn đến giá trị thị trường, với các cầu thủ trẻ thường được đánh giá cao hơn nhờ tiềm năng phát triển.

### 3. Phân bố giá trị cầu thủ theo quốc gia và câu lạc bộ

- Một số quốc gia và câu lạc bộ nổi bật với các cầu thủ đắt giá, thể hiện sự chênh lệch về tài năng và đầu tư trong bóng đá toàn cầu.

### 4. Xu hướng trong các chỉ số kỹ thuật

- Các cầu thủ tấn công và phòng ngự có những đặc điểm kỹ thuật riêng biệt, giúp định hình vai trò chiến thuật của họ trong đội hình.

### 5. Tầm quan trọng và ứng dụng thực tiễn

- Phân tích này giúp cung cấp dữ liệu hỗ trợ cho việc ra quyết định trong quản lý đội bóng, định giá cầu thủ, và lập kế hoạch chiến lược.
- Kết quả cũng mở ra cơ hội nghiên cứu sâu hơn về xu hướng bóng đá toàn cầu, từ việc đào tạo cầu thủ đến xây dựng đội hình.

## 4 Đánh giá

Họ và Tên	Tích cực tham gia đóng góp ý kiến	Khả năng hoàn thành công việc	Tư duy sáng tạo trong công việc	Chất lượng đạt được	Số lần tham gia cuộc họp
Trần Anh Quân	9	10	10	10	6
Đặng Anh Tuấn	10	9	9	10	6
Lê Thanh Tiến	10	10	9	10	6
Trịnh Hoàng Phú	9	10	9	10	6
Nguyễn Hoàng Linh	10	10	9	9	6
Trần Phước Vinh	9	10	10	9	6

Bảng 1: Bảng đánh giá các thành viên trong nhóm

## **5 Tài liệu tham khảo**