



Fifa eda stats

DỮ LIỆU BÓNG ĐÁ

# BÁO CÁO CUỐI KỲ

Trình bày bởi nhóm 13



FIFA19



## Fifa\_eda\_stats 2019

# Dự án

Tổng quan kế hoạch  
của bài thuyết trình hôm nay

- Giới thiệu về dự án
- Khái quát về dữ liệu
- Mục tiêu dự án
- Các kỹ thuật xử lý dữ liệu và trực quan
- Tiền xử lý dữ liệu
- Lựa chọn biểu đồ trực quan và đánh giá
- Tổng kết
- Các hạn chế và vấn đề phát sinh

# Giới thiệu về Nhóm 13

## Nhóm thực hiện

Nhóm 13 rất hào hứng khi nhận được project lần này, đây là một project khá thú vị, nó đúng về lĩnh vực mà nhóm yêu thích nói riêng và đa số mọi người nói chung, nhóm làm bài báo cáo này với tâm thế những người ở giai đoạn đầu trong lĩnh vực phân tích, tâm huyết là thứ mà nhóm dành cho bài báo cáo này.



# Thành viên nhóm:

Nguyễn Hoàng Linh  
21110333

Trịnh Hoàng Phú  
21110370

Trần Anh Quân  
21110394

Lê Thanh Tiến  
21110409

Đặng Anh Tuấn  
21110434

Trần Phước Vinh  
21110450

# Giới thiệu về Dự án

## 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 ở khắp các châu lục, là ngành công nghiệp trị giá hàng tỷ đô la.

## 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.
- Đưa ra các chiến thuật phù hợp.
- Việc đánh giá, phân tích dữ liệu đã trở thành yếu tố quan trọng.



# Khái quát về dữ liệu

**Bộ dữ liệu FIFA 2019** chứa thông tin chi tiết  
về hơn 18.000 cầu thủ bóng đá trên toàn thế  
giới.

# Các thuộc tính của bộ dữ liệu

01

## Thông tin cầu thủ

Name,  
Age,  
Nationality,  
Body Type

02

Thông tin CLB và hợp đồng  
Club,  
Joined,  
Loaned From,  
Contract Valid Until

03

## Lương và giá trị

Value,  
Wage,  
Release Clause

04

## Kỹ năng và phong độ

Overall, Potential, Preferred Foot,  
International Reputation,  
Weak Foot, Skill Moves

05

Các chỉ số về kỹ năng  
Crossing, Finishing, HeadingAccuracy,  
ShortPassing, Volleys, Dribbling,  
Curve, FKAccuracy, LongPassing,  
BallControl, Positioning,  
Vision, Penalties

06

Thể chất và phong độ trên  
sân  
Acceleration, SprintSpeed,  
Agility, Reactions, Balance,  
ShotPower, Jumping, Stamina,  
Strength, LongShots,  
Aggression, Interceptions,  
Marking, StandingTackle,  
SlidingTackle, Composure

07

## Chỉ số về thủ môn

GKDiving,  
GKHandling,  
GKKicking, GKPositioning,  
GKReflexes

# Mục tiêu dự án

01

Mục tiêu tổng quát

02

Mục tiêu chi tiết

# 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 nhằm hiểu rõ hơn về cấu trúc và ý nghĩa của bộ dữ liệu.

# Mục tiêu chi tiết

01

Tổng quan

02

Thống kê

03

Câu lạc bộ

04

So sánh câu lạc  
bộ

05

Chỉ số cầu thủ

06

Giá trị cầu thủ

# Các kỹ thuật xử lý dữ liệu và trực quan

01

Các kỹ thuật xử lý dữ liệu

02

Các kỹ thuật trực quan

# Các kỹ thuật xử lý dữ liệu

01

## Kiểm tra cấu trúc dữ liệu:

- Đảm bảo rằng các cột quan trọng như Age, Overall, Value, Wage,... có định dạng phù hợp.
- Loại bỏ các giá trị bị thiếu hoặc không hợp lệ.
- Kiểm tra các giá trị bất thường, ví dụ: Một số cầu thủ trong bộ dữ liệu có nhiều trường có giá trị N/A.

01

## Chuẩn hóa định dạng dữ liệu:

- Thay hoặc loại bỏ các kí hiệu (M,K) và kí hiệu tiền tệ (€) cho một số trường và chuyển đổi sang kiểu số cho các biến định lượng.

# Các kỹ thuật trực quan

01

## Tách cột Work Rate:

- Các giá trị trong cột Work Rate có dạng Tấn công/Phòng thủ chẵng hạn : “Medium/Medium”.
- Thực hiện tách thành 2 cột là:
- Attack Work Rate (Chỉ số tấn công ):

Attack Work Rate =

```
IF(  
    CONTAINSSTRING('fifa_eda_stats'[Work Rate], "/"),  
    LEFT('fifa_eda_stats'[Work Rate], FIND("/", 'fifa_eda_stats'[Work Rate]) - 1),  
    BLANK()  
)
```

# Các kỹ thuật trực quan

01

## Tách cột Work Rate

- Defense Work Rate (Chỉ số phòng thủ):

Defense Work Rate =

```
IF(  
    CONTAINSSTRING('fifa_eda_stats'[Work Rate], "/"),  
    MID(  
        'fifa_eda_stats'[Work Rate],  
        FIND("/", 'fifa_eda_stats'[Work Rate]) + 1,  
        LEN('fifa_eda_stats'[Work Rate])  
    ),  
    BLANK()  
)
```

# Các kỹ thuật trực quan

02

## Áp dụng cho biểu đồ Radar

- 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:

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

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

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

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

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

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

# Các kỹ thuật trực quan

02

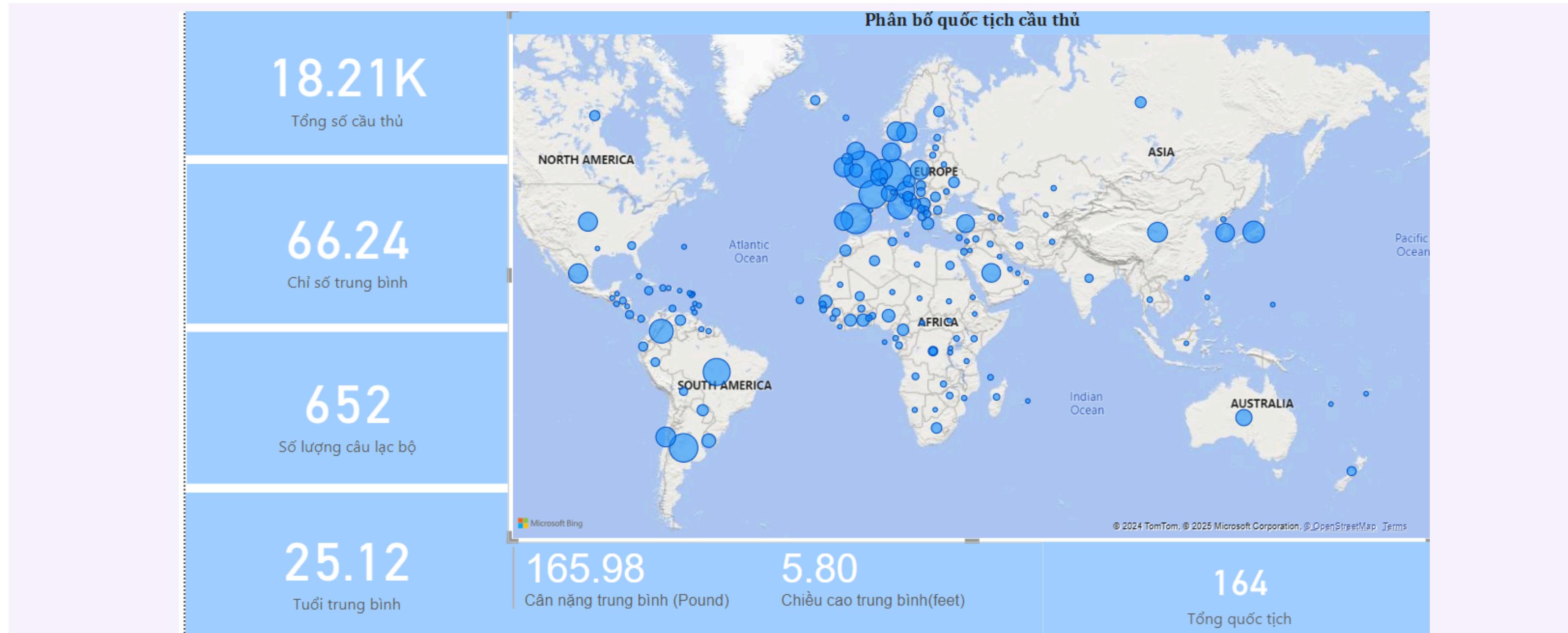
## Áp dụng cho biểu đồ Radar

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:

$$\text{GK\_Speed} = (\text{[Acceleration]} + \text{[SprintSpeed]})/2$$

Lựa chọn  
biểu đồ trực  
quan và  
đánh giá

# Mục tiêu tổng quan



# Mục tiêu thống kê

Biểu đồ thanh ngang



Biểu đồ phân tán



Biểu đồ khoảng tin cậy



Biểu đồ tròn

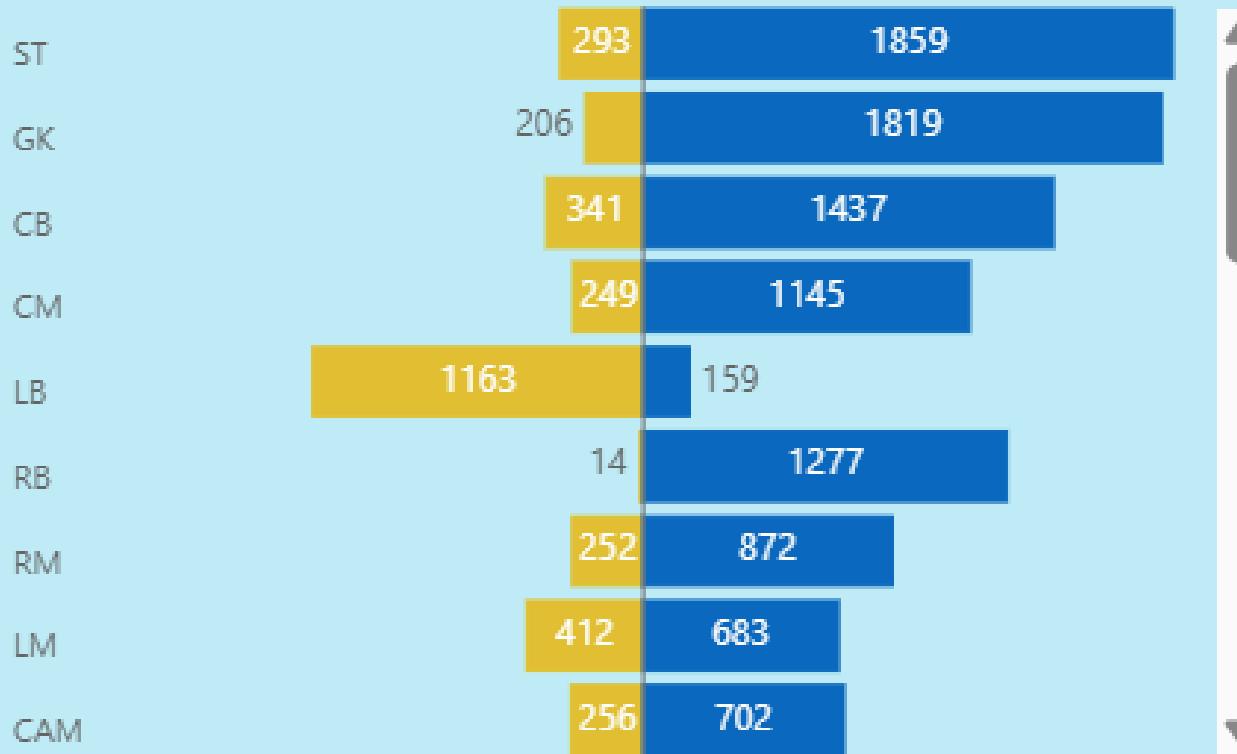


Biểu đồ vùng xếp  
chồng

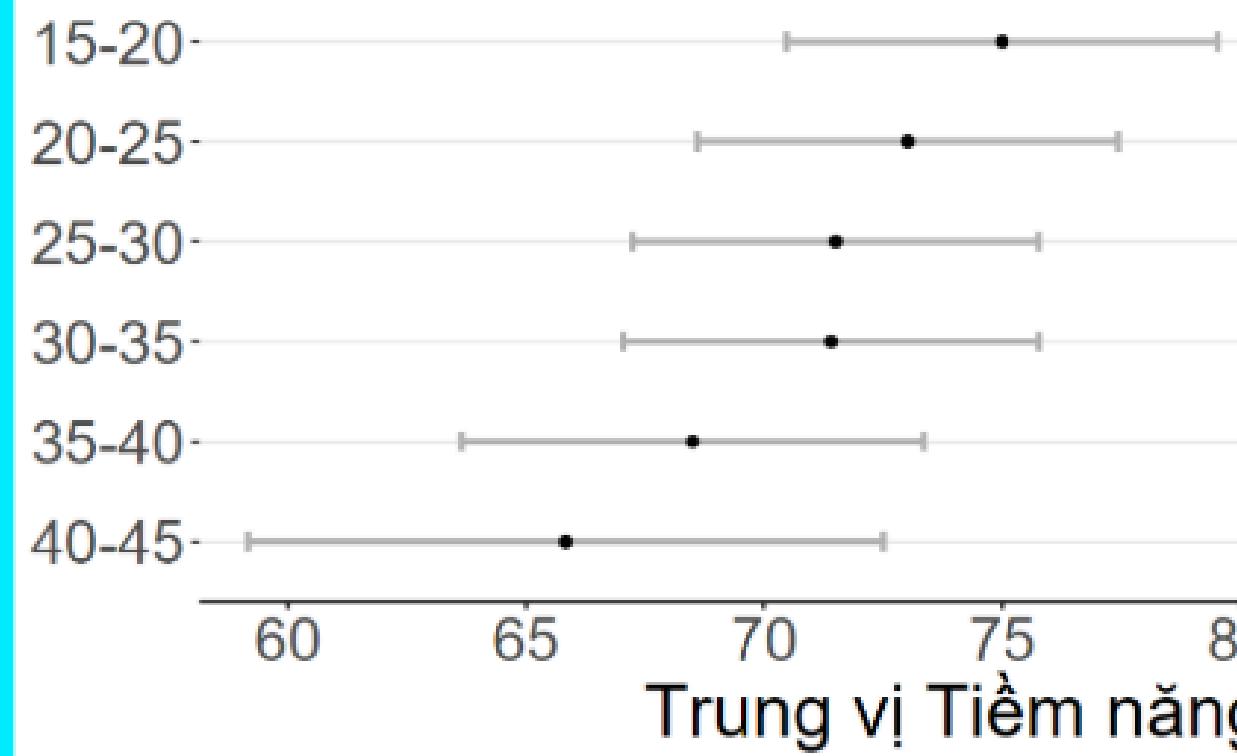


### Chân thuận của cầu thủ theo từng vị trí

Left ● Right

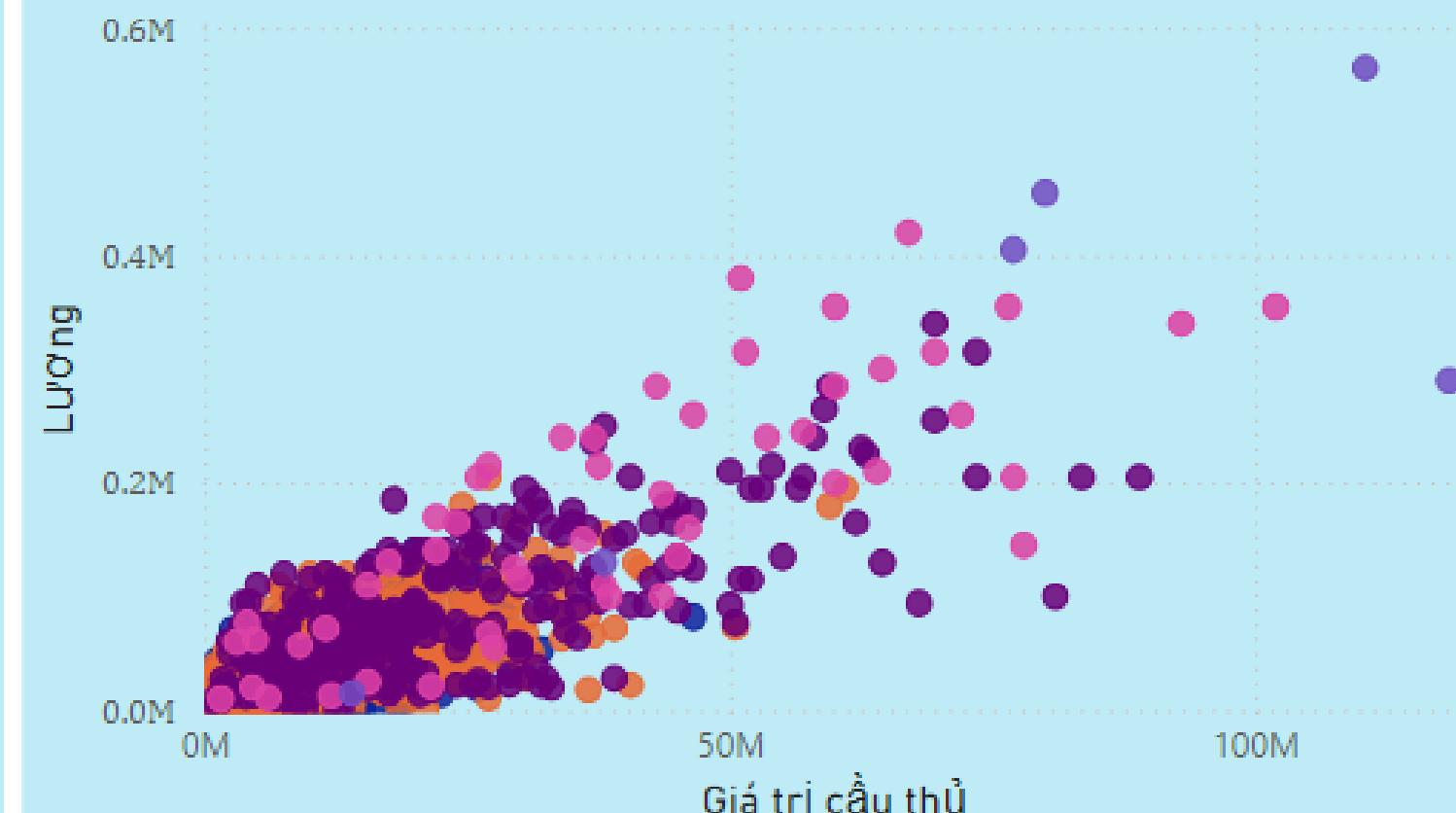


### Khoảng tin cậy giữa độ tuổi và tiềm năng

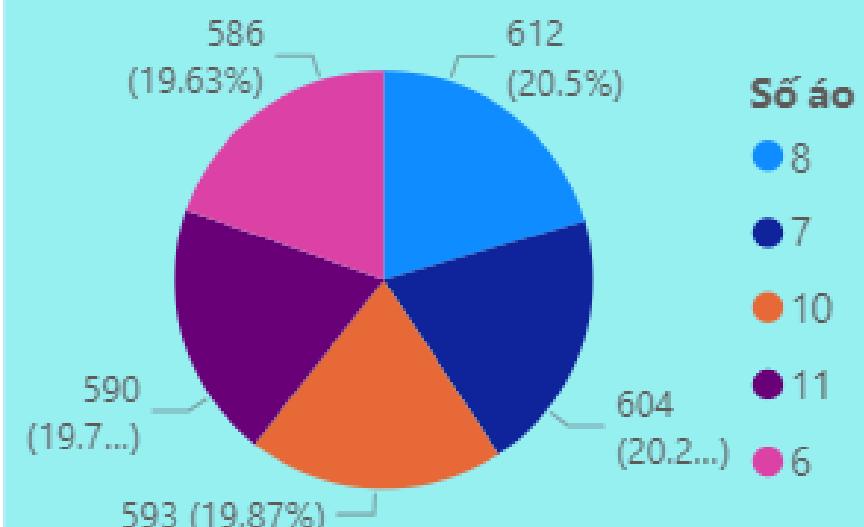


### Lương và giá trị cầu thủ theo danh tiếng

Danh tiếng ● (Blank) ● 1 ● 2 ● 3 ● 4 ● 5

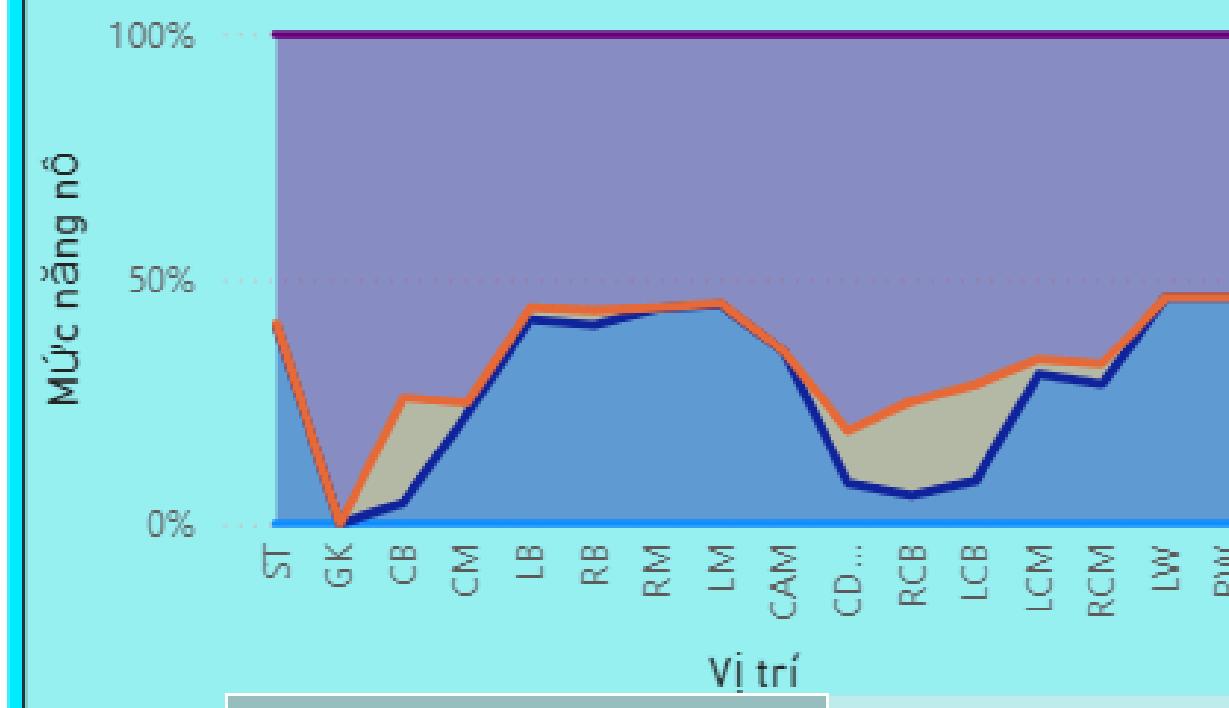


### Số áo ưa thích



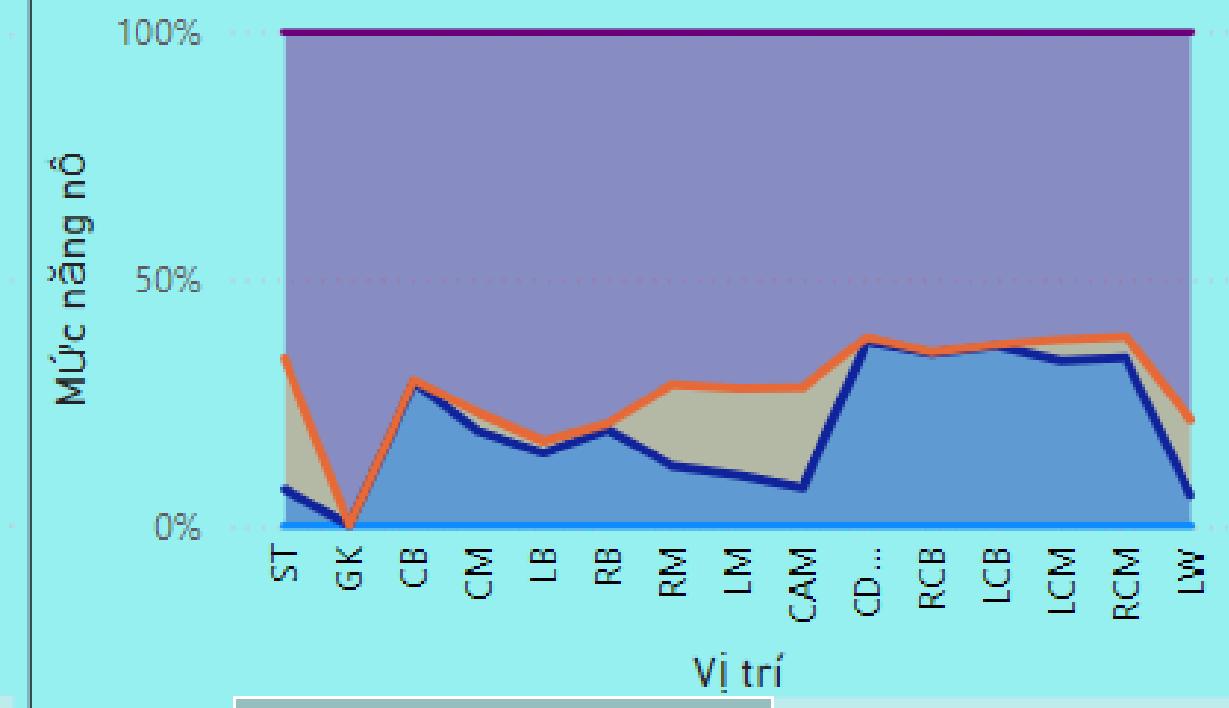
### Độ năng nổ tấn công theo vị trí

Độ năng nổ tấn công theo vị trí ● (Blank) ● High ● Low ● Medium



### Độ năng nổ phòng ngự theo vị trí

Độ năng nổ phòng ngự theo vị trí ● (Blank) ● High ● Low ● Medium



# Mục tiêu câu lạc bộ

Biểu đồ thanh cụm



Biểu đồ Treemap



Biểu đồ vùng

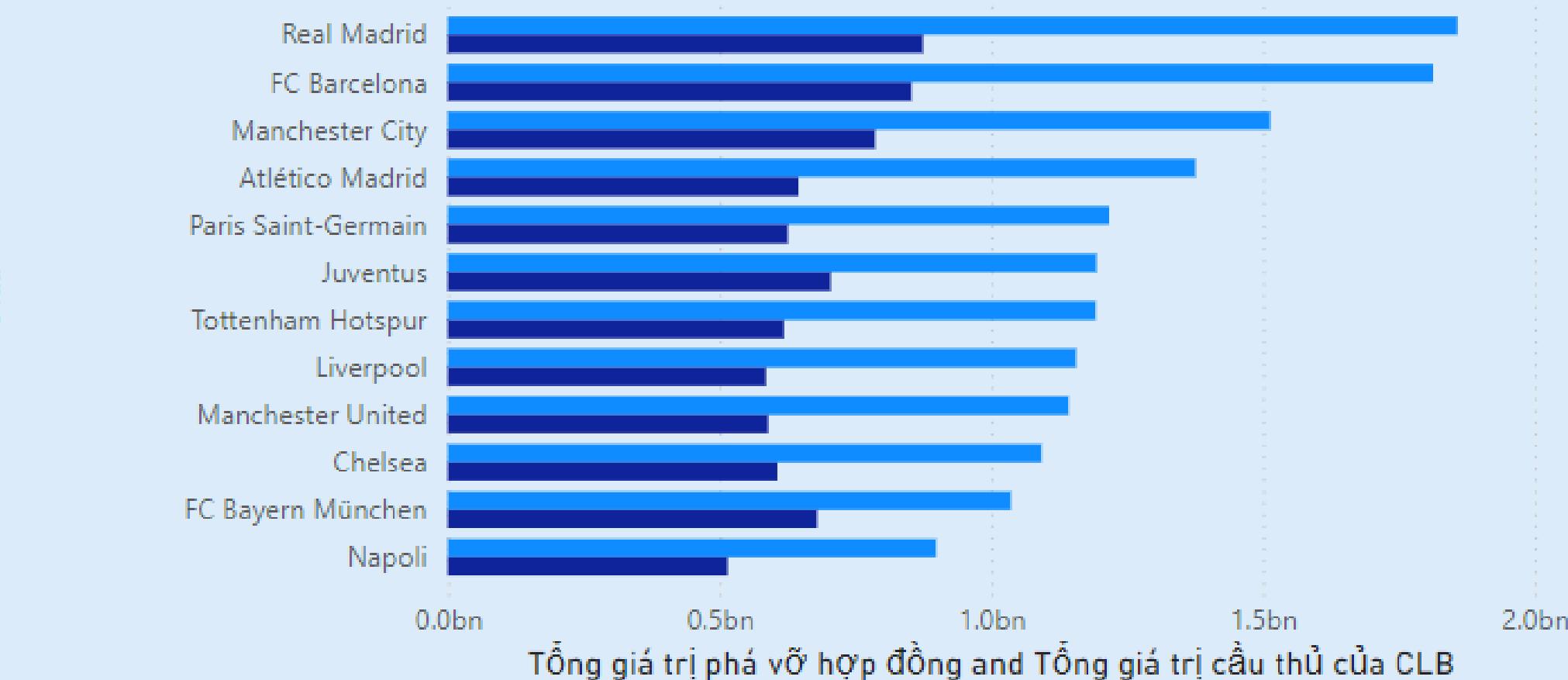


CLB	15-20	20-25	25-30	30-35	35-40	40-45	Tổng
Zaglebie Sosnowiec	2	7	8	7	1		<b>25</b>
Zaglebie Lubin	2	8	11	5	1		<b>27</b>
Yokohama F. Marinos	5	9	11	4		1	<b>30</b>
Yeovil Town	4	16	2	4	1	1	<b>28</b>
Yeni Malatyaspor	2	8	14	6			<b>30</b>
Wycombe Wanderers	1	9	4	9	2		<b>25</b>
Wolverhampton Wanderers	9	14	7	3			<b>33</b>
Wolfsberger AC	5	8	8	5			<b>26</b>



## Tổng giá trị phá vỡ hợp đồng và Tổng giá trị cầu thủ của từng CLB

● Tổng giá trị phá vỡ hợp đồng ● Tổng giá trị cầu thủ của CLB



## Quốc tịch cầu thủ trong các CLB



# Mục tiêu so sánh câu lạc bộ

Biểu đồ cột và đường



Biểu đồ Violin



Biểu đồ đường



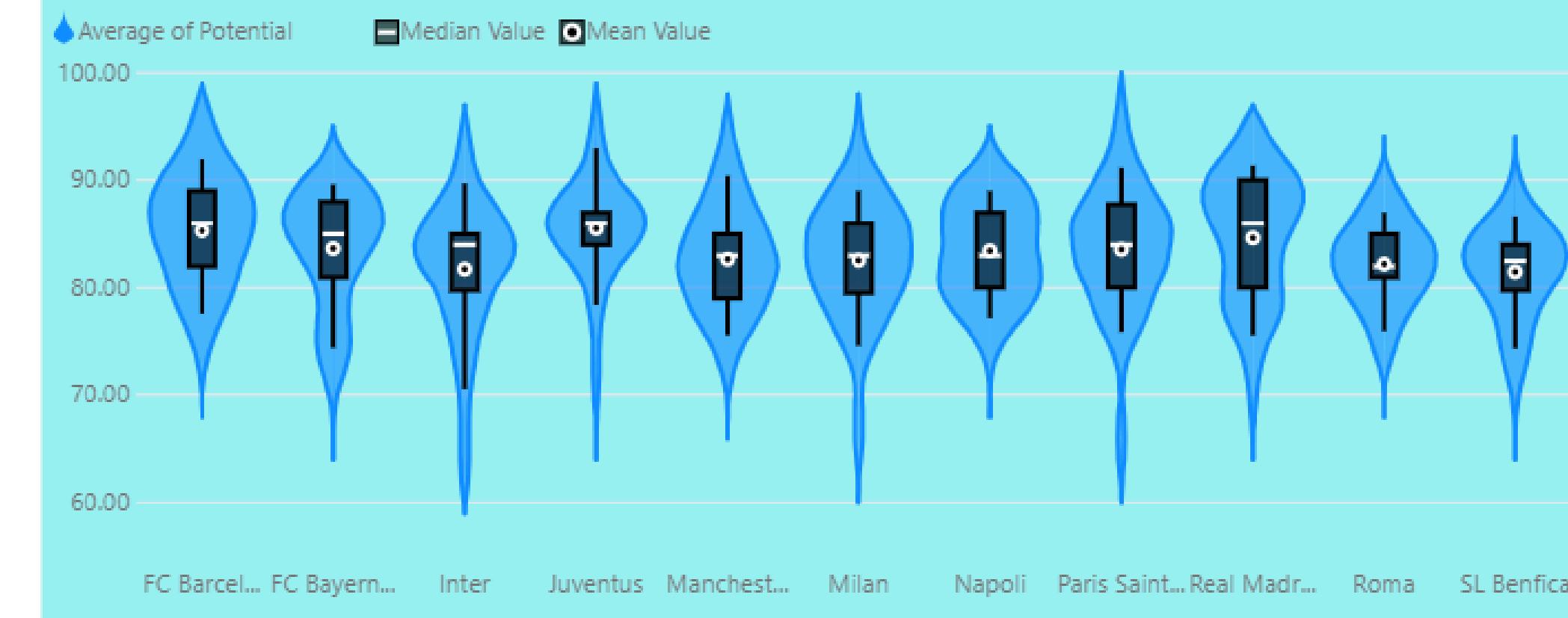
Biểu đồ khu vực



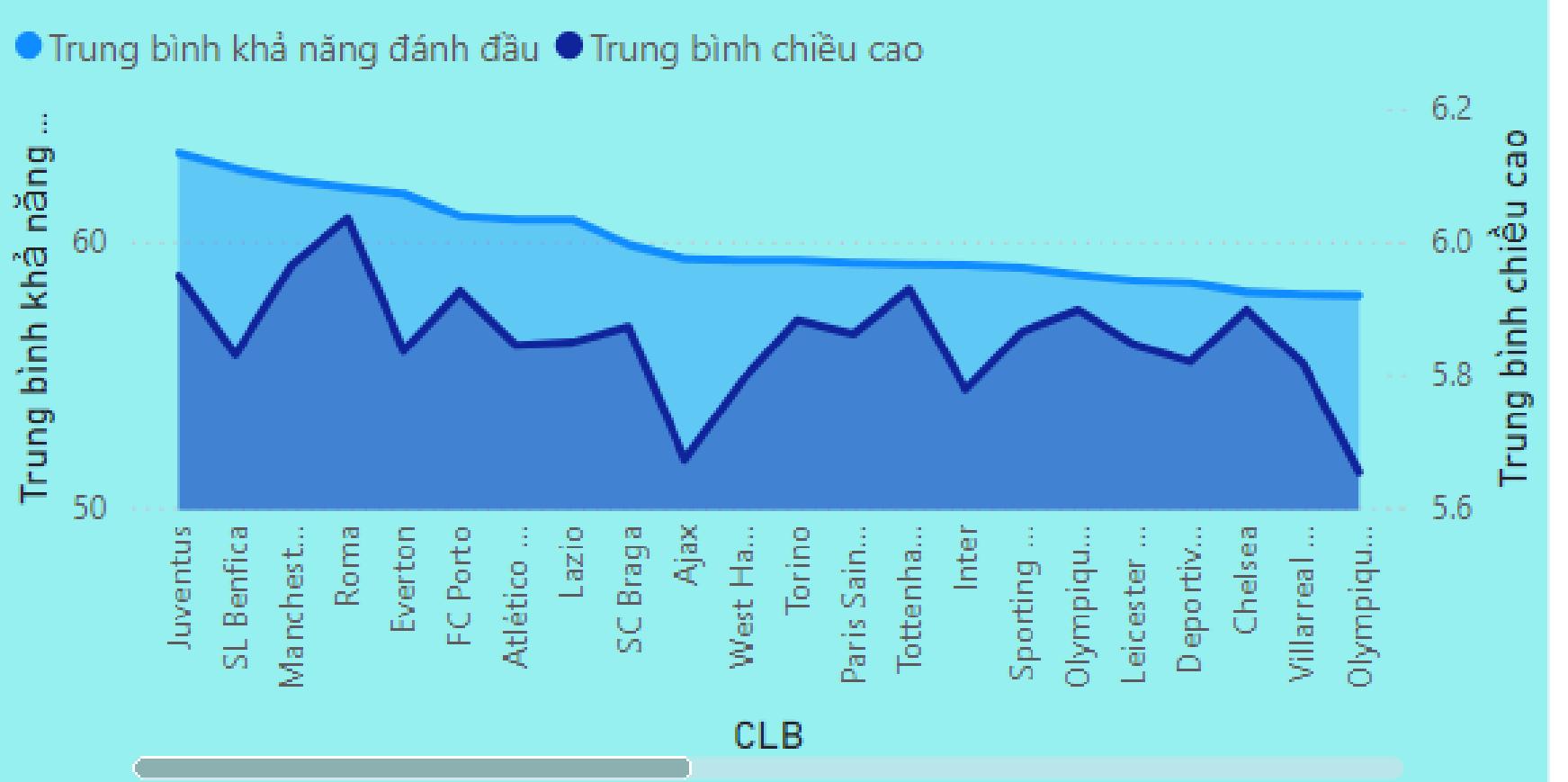
## Tổng giá trị và Quỹ lương của CLB



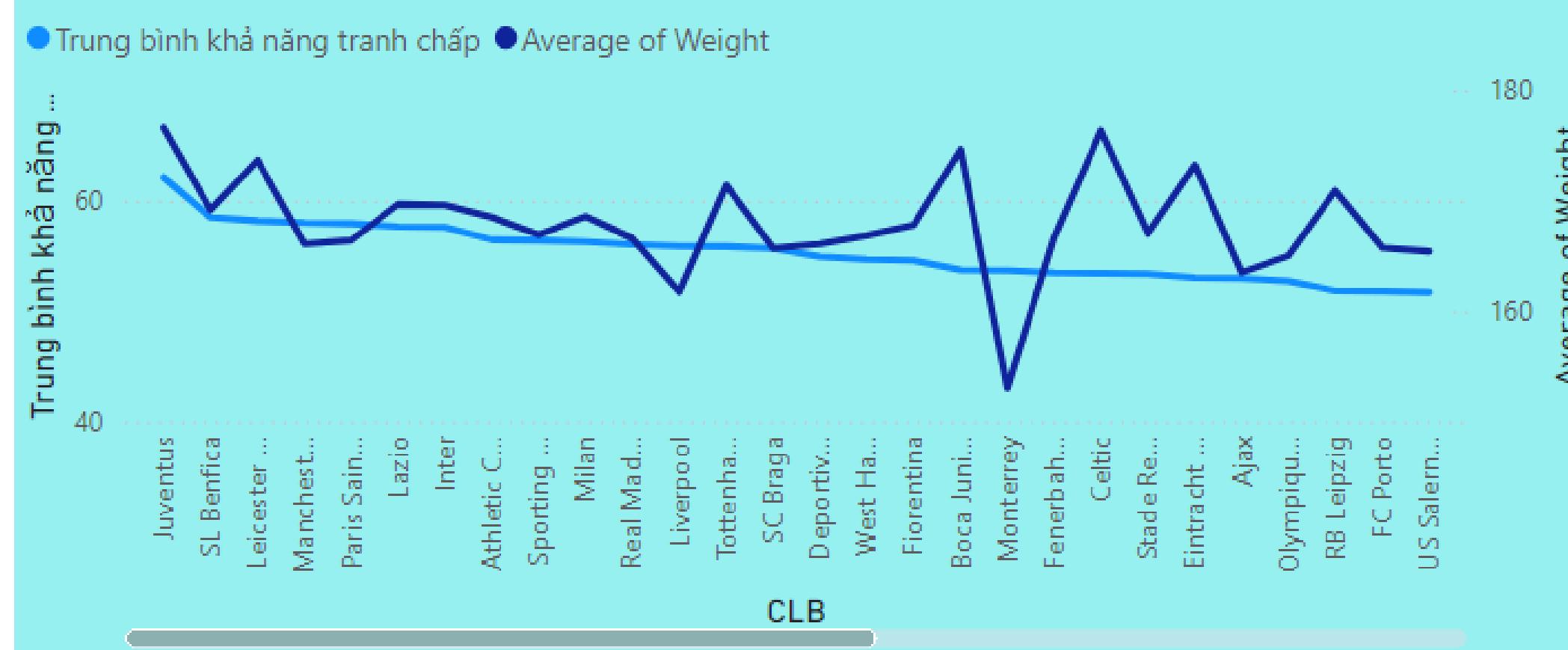
## Tiềm năng trung bình của top 10 CLB tốt nhất



## Khả năng đánh đầu và chiều cao trung bình của các CLB



## Khả năng tranh chấp và Cân nặng trung bình của các CLB



# Mục tiêu chỉ số cầu thủ

Biểu đồ thanh cụm



Biểu đồ Radar  
cầu thủ



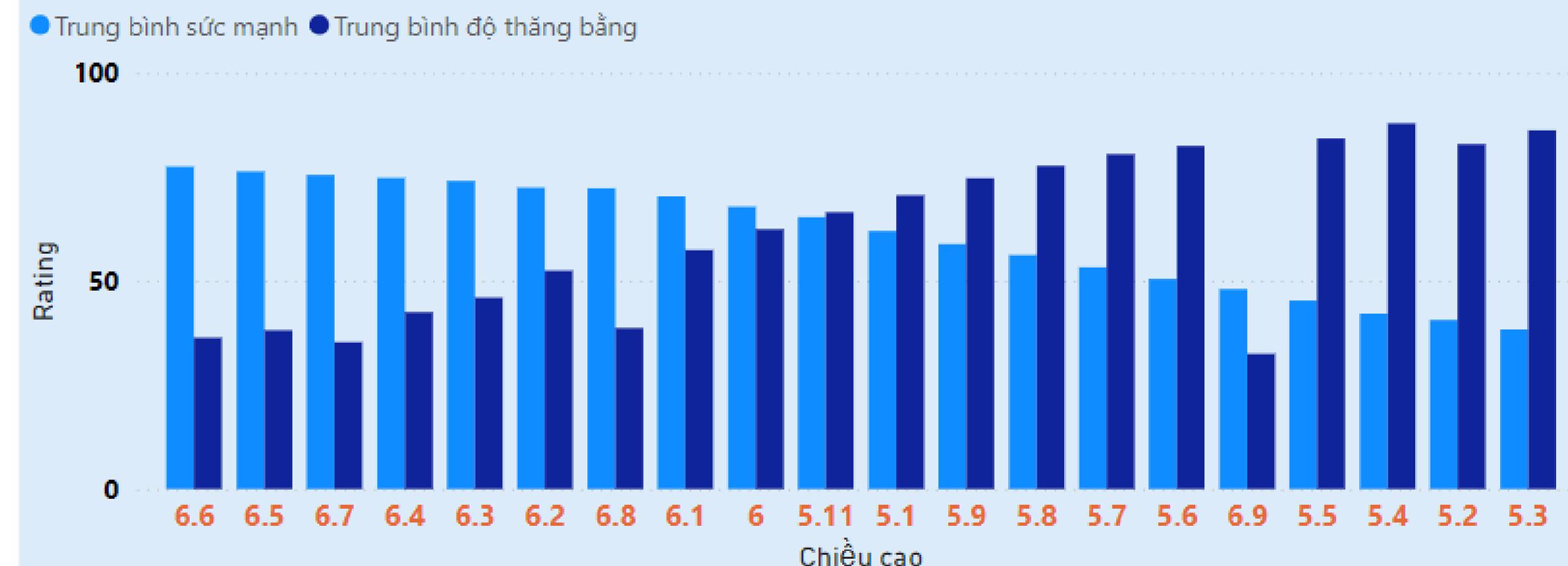
Biểu đồ Radar thủ  
môn



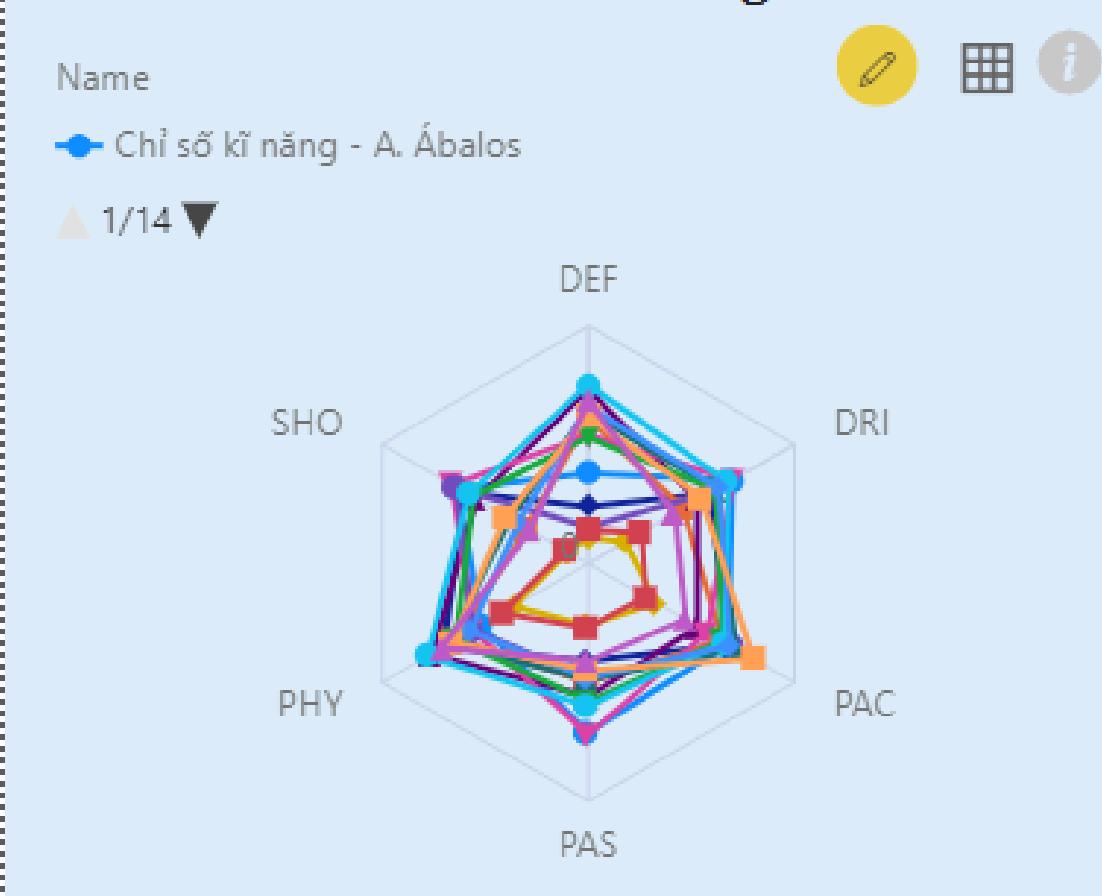
## Cầu thủ

Tên	CLB	Vị trí	OVR
Cristiano Ronaldo	Juventus	ST	94
L. Messi	FC Barcelona	RF	94
Neymar Jr	Paris Saint-Germain	LW	92
E. Hazard	Chelsea	LF	91
K. De Bruyne	Manchester City	RCM	91
L. Modric	Real Madrid	RCM	91
Sergio Ramos	Real Madrid	RCB	91
D. Godín	Atlético Madrid	CB	90
David Silva	Manchester City	LCM	90
R. Lewandowski	FC Bayern München	ST	90
T. Kroos	Real Madrid	LCM	90
A. Cisneros	Atletico Madrid	CAM	90
<b>Total</b>			

## Sức mạnh và độ thăng bằng của cầu thủ theo chiều cao



## Radar kĩ năng



## Thủ môn

Tên	CLB	OVR
De Gea	Manchester United	91
J. Oblak	Atlético Madrid	90
M. Neuer	FC Bayern München	89
M. ter Stegen	FC Barcelona	89
T. Courtois	Real Madrid	89
G. Buffon	Paris Saint-Germain	88
H. Lloris	Tottenham Hotspur	88
S. Handanovic	Inter	88
K. Navas	Real Madrid	87
Ederson	Manchester City	86
Alisson	Liverpool	85
W. Szczesny	Juventus	85

## Radar kĩ năng thủ môn



# Giá trị cầu thủ

Biểu đồ giá trị theo độ tuổi  
của cầu thủ



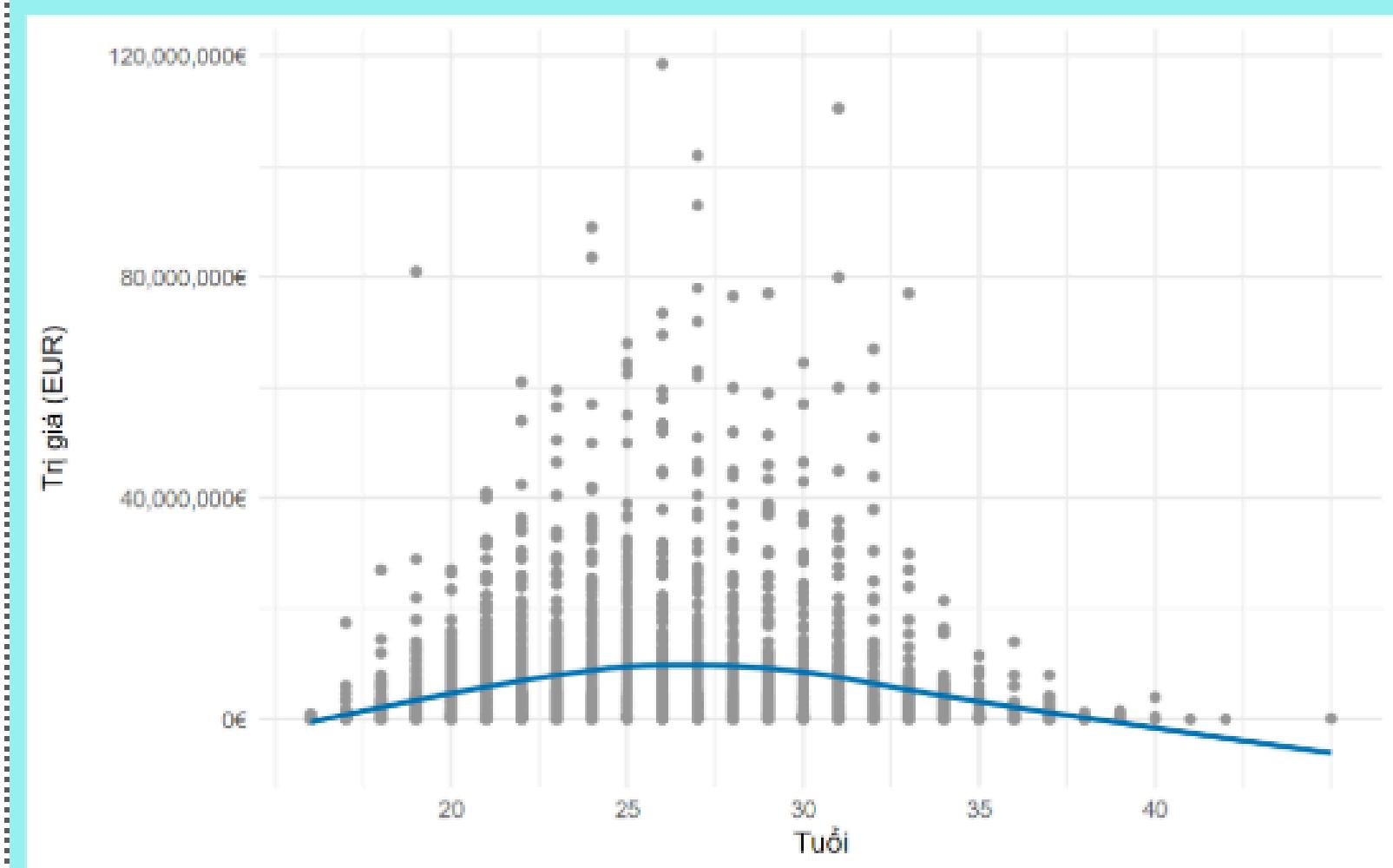
Biểu đồ giá trị và lương



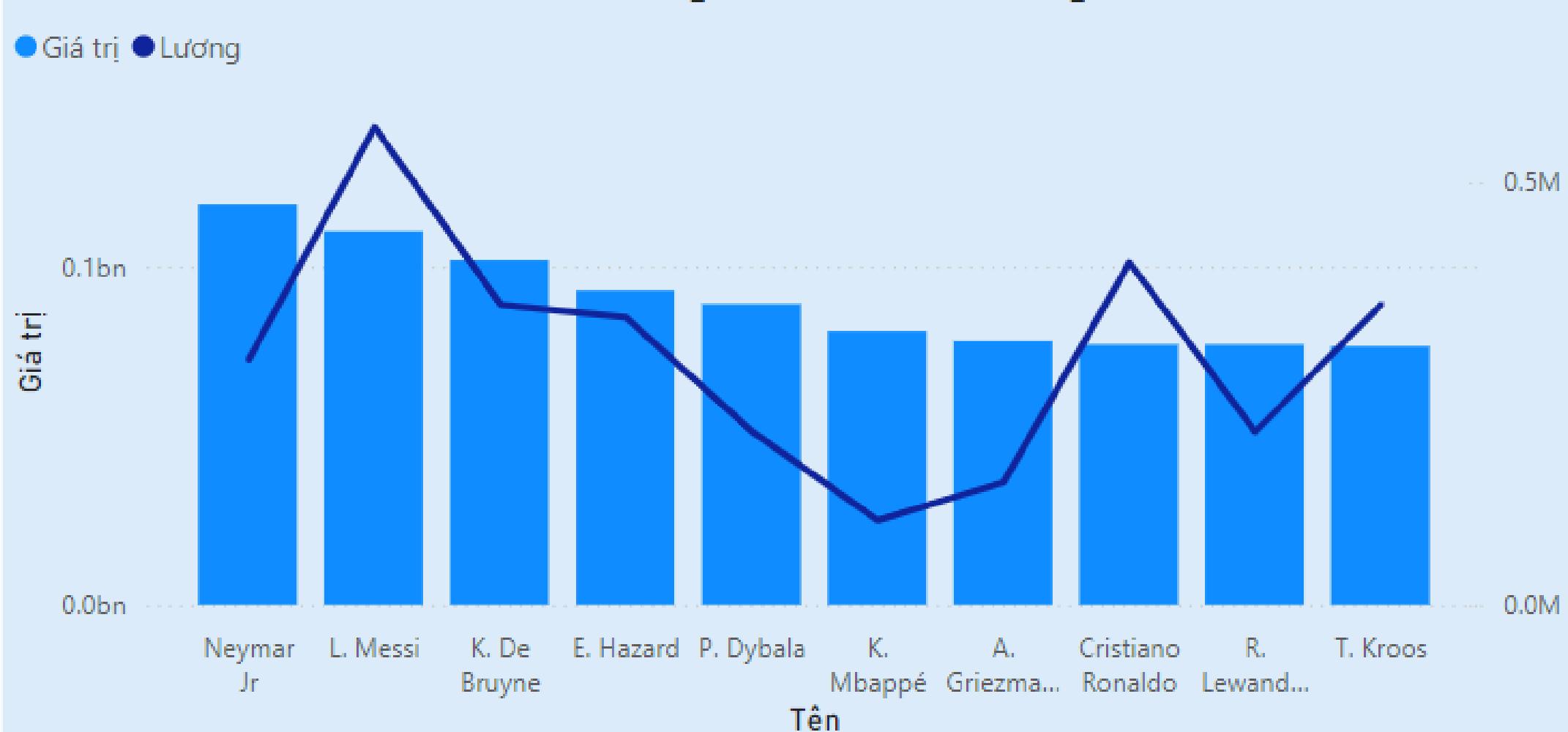
Biểu đồ độ tuổi và danh tiếng



## Giá trị cầu thủ so với tuổi của cầu thủ



## Giá trị và Lương của 10 cầu thủ đắt giá nhất



## Độ tuổi và Danh tiếng của 10 cầu thủ đắt giá nhất



Tên	Giá trị	Lương
L. Messi	110500000	565000
Cristiano Ronaldo	77000000	405000
K. De Bruyne	102000000	355000
T. Kroos	76500000	355000
E. Hazard	93000000	340000
Neymar Jr	118500000	290000
P. Dybala	89000000	205000
R. Lewandowski	77000000	205000
A. Griezmann	78000000	145000
K. Mbappé	81000000	100000

# Tổng kết

Sự đa dạng trong kỹ năng và giá trị cầu thủ

Mối quan hệ giữa các biến số

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

Xu hướng trong các chỉ số kỹ thuật

# Các hạn chế và vấn đề phát sinh



# Khi thực hiện trực quan hóa bộ dữ liệu FIFA 2019, một số hạn chế và khó khăn phổ biến gặp phải bao gồm:

- 01 Kích thước và độ phức tạp của dữ liệu
- 02 Chất lượng dữ liệu
- 03 Sự chồng chéo và mối quan hệ phức tạp
- 04 Lựa chọn biểu đồ phù hợp
- 05 Khán giả mục tiêu
- 06 Giới hạn công cụ trực quan hóa

# Kích thước và độ phức tạp dữ liệu

Dữ liệu lớn: Với hơn 18.000 câu thủ và nhiều cột thông tin, việc xử lý và trực quan hóa toàn bộ dữ liệu có thể dẫn đến quá tải, làm chậm hiệu suất Power BI hoặc các công cụ khác. Dữ liệu đa chiều: Các chỉ số kỹ năng, tài chính, và thể chất tạo thành nhiều tầng thông tin, gây khó khăn trong việc chọn biểu đồ phù hợp để trình bày hiệu quả.

# Chất lượng dữ liệu

Dữ liệu bị thiếu hoặc không nhất quán: Một số cột như Loaned From hoặc Release Clause có thể bị thiếu hoặc không đồng nhất, làm giảm tính chính xác của biểu đồ. Định dạng không chuẩn hóa: Các cột như Value (ví dụ: €110.5M) và Wage (ví dụ: €200K) yêu cầu xử lý và chuyển đổi trước khi sử dụng, gây mất thời gian.

# Sự chồng chéo và mối quan hệ phức tạp

Tương quan nhiều biến: Các chỉ số như Overall, Potential, và Value có mối quan hệ mật thiết, việc trình bày sao cho dễ hiểu mà không bị chồng chéo là một thách thức. Mối quan hệ phi tuyến tính: Một số quan hệ (ví dụ: tuổi và tiềm năng) không tuân theo quy luật tuyến tính, khó trực quan bằng các biểu đồ cơ bản.

# Lựa chọn biểu đồ phù hợp

Hạn chế biểu đồ tương tác: Power BI có thể gặp khó khăn trong việc xử lý các biểu đồ tương tác phức tạp khi dữ liệu quá lớn. Trực quan hóa các chỉ số kỹ năng: Với nhiều chỉ số như Dribbling, Finishing, SprintSpeed, Radar Chart có thể bị rối nếu hiển thị quá nhiều giá trị cùng lúc.

# Khán giả mục tiêu

Đa số người đọc bộ dữ liệu này thường là các hlv hoặc người chơi, nếu đưa cho người không biết gì về bóng đá có thể gây khó hiểu

# Giới hạn công cụ trực quan hóa

Hạn chế của Power BI: Không hỗ trợ tốt cho một số biểu đồ phức tạp như Violin Plot hoặc Scatter Matrix. Hiệu năng giảm khi xử lý các tập dữ liệu lớn với nhiều phép tính. Khả năng tuỳ chỉnh: Một số yêu cầu trực quan phức tạp đòi hỏi khả năng lập trình hoặc tích hợp công cụ bổ sung, vượt ngoài khả năng mặc định của Power BI.

Xin cảm ơn!