

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA TOÁN-TIN HỌC



TRỰC QUAN HÓA DỮ LIỆU

**Đề tài: ỨNG DỤNG Trực Quan Hóa Dữ Liệu trong Phân tích
và So sánh Chỉ số Cầu thủ Bóng đá**

Giáo viên hướng dẫn: Tô Đức Khánh

Nhóm thực hiện: Nhóm 14

Lớp: 21TTH_KDL

Thành viên nhóm:

1. Trần Hoàng Anh - 20280004
2. Lê Thị Mỹ Tiên- 20280096
3. Hồ Quốc Tuấn - 20110435
4. Phạm Đức Mạnh - 20110341
5. Nguyễn Đức Mạnh– 20280065
6. Trương Quốc Khang - 20280051

Mục Lục

1. Tổng quan	3
1.1 Giới Thiệu	3
2. Kiến thức về bộ dữ liệu	4
2.1 BỘ DỮ LIỆU SPORTS DATA ANALYSIS	4
2.1.1 Giới thiệu về bộ dữ liệu	4
2.1.2 Các bước kiểm tra và xử lý dữ liệu	5
2.1.3 Hướng sử dụng bộ dữ liệu trong báo cáo	8
3. Phương Pháp Đề Xuất Trực Quan Hóa Dữ Liệu	11
3.1 Bản Đề Xuất Trực Quan Hóa Dữ Liệu	11
3.1.1. Bản đề xuất trực quan dữ liệu	11
4. NỘI CÁC BIỂU ĐỒ THÀNH DASHBOARD	25
4.2. Dashboard	25
.....	29
.....	29
Triển khai	30
5.1 Giới thiệu Streamlit	30
5.2. Những thứ còn thiếu sót	31
Kết luận và hướng phát triển	31

Tổng quan

1.1 Giới Thiệu

Trong thời đại dữ liệu số hóa phát triển vượt bậc, dữ liệu đã trở thành chìa khóa quan trọng đối với sự thành công của bất kỳ cá nhân hay tổ chức nào trong nhiều lĩnh vực, từ kinh doanh, y tế, giáo dục đến giải trí [1][2]. Vì vậy, việc sử dụng các kỹ thuật trực quan hóa dữ liệu không chỉ là một kỹ năng hữu ích mà còn là yếu tố cần thiết đối với sinh viên ngày nay. Môn học Trực quan hóa dữ liệu không chỉ giúp sinh viên nắm vững các khái niệm cơ bản mà còn khuyến khích áp dụng chúng vào các bài toán thực tiễn, qua đó nâng cao năng lực phân tích và truyền tải thông tin.

Đề tài cuối kỳ của môn học tập trung vào phân tích và trực quan hóa dữ liệu từ các tập dữ liệu khác nhau, bao gồm sports, Olympic 2024, và clothing chain. Sau quá trình khảo sát, chúng tôi quyết định chọn bộ dữ liệu **sports data analysis** vì tính thực tiễn của nó cùng với cấu trúc dữ liệu rõ ràng, dễ tiếp cận cho người dùng mới. Bộ dữ liệu này bao gồm thông tin chi tiết về 18,207 cầu thủ với 57 biến, như ID, tên, tuổi, quốc tịch, câu lạc bộ, giá trị chuyển nhượng, lương, và các chỉ số kỹ thuật. Chi tiết về bộ dữ liệu sẽ được trình bày tại phần \ref. Bộ dữ liệu cung cấp một cái nhìn tổng quan về hiệu suất và tiềm năng của các cầu thủ, đồng thời mở ra nhiều hướng khai thác khác nhau.

Mục tiêu chính của bài báo cáo là xây dựng các biểu đồ và dashboard trực quan, thể hiện những khía cạnh nổi bật và các mối quan hệ quan trọng trong dữ liệu. Ngoài ra, chúng tôi đã triển khai thành công bài báo cáo trên local server thông qua thư viện Streamlit, giúp nội dung của báo cáo trở nên gần gũi và ứng dụng hơn trong thực tế.

Tổng kết, những đóng góp chính trong báo cáo của chúng tôi bao gồm 3 nội dung chính:

1. Phân tích bộ dữ liệu, xác định các đặc điểm nổi bật và xây dựng các biểu đồ phù hợp với đặc tính của dữ liệu.
2. Kết hợp các biểu đồ thành các dashboard có tính liên kết, qua đó tạo ra một thiết kế hoàn chỉnh, phù hợp với tập dữ liệu.
3. Triển khai thành công các biểu đồ và dashboard trên local server thông qua thư viện Streamlit, mang lại tính thực tiễn cao cho nội dung môn học.

Kiến thức về bộ dữ liệu

2.1 BỘ DỮ LIỆU SPORTS DATA ANALYSIS

2.1.1 Giới thiệu về bộ dữ liệu

Bộ dữ liệu Sports Data Analysis chứa thông tin chi tiết về 18,207 cầu thủ bóng đá, được tổng hợp trong file dữ liệu `fifa_eda_stats.csv`. Dữ liệu bao gồm 57 cột (biến số) khác nhau, cung cấp một bức tranh toàn diện về đặc điểm cá nhân, hiệu suất, và các chỉ số chuyên môn của từng cầu thủ.

Dưới đây là một số trường giá trị tiêu biểu từ bộ dữ liệu:

Thông tin cơ bản về cầu thủ:

- ID: Mã định danh duy nhất của mỗi cầu thủ.
- Name: Tên cầu thủ.
- Age: Tuổi của cầu thủ.
- Nationality: Quốc tịch của cầu thủ.
- Club: Câu lạc bộ hiện tại của cầu thủ.
- Position: Vị trí thi đấu sở trường (ví dụ: RF - Right Forward, ST - Striker).

Chỉ số kỹ năng tổng quát:

- Overall: Điểm đánh giá tổng thể (tối đa là 100), đại diện cho kỹ năng hiện tại.
- Potential: Điểm đánh giá tiềm năng (tối đa là 100), dự đoán hiệu suất tương lai.

Giá trị thị trường:

- Value: Giá trị thị trường ước tính của cầu thủ (ví dụ: €110.5M).
- Wage: Mức lương hàng tuần của cầu thủ (ví dụ: €565K).
- Release Clause: Chi phí giải phóng hợp đồng.

Thông tin cá nhân:

- Preferred Foot: Chân thuận (Left hoặc Right).
- Height: Chiều cao (ví dụ: 5'7).
- Weight: Cân nặng (ví dụ: 159 lbs).

Chỉ số chuyên môn:

- Dribbling: Kỹ năng rê bóng.
- Crossing: Khả năng tạt bóng.

- Finishing: Khả năng dứt điểm.
- Acceleration và SprintSpeed: Tốc độ khởi đầu và chạy nước rút.
- Stamina: Thể lực.

Chỉ số phòng ngự và thủ môn:

- StandingTackle và SlidingTackle: Khả năng cản phá bóng.
- GKReflexes và GKDiving: Phản xạ và khả năng bay người của thủ môn.

Ví dụ minh họa:

- Cầu thủ **Lionel Messi** (ID: 158023) đến từ Argentina, đang thi đấu cho câu lạc bộ **FC Barcelona** với điểm đánh giá tổng thể Overall = 94 và tiềm năng Potential = 94. Anh có giá trị thị trường là **€110.5M** và mức lương hàng tuần là **€565K**. Messi thuận chân trái, chơi ở vị trí **Right Forward (RF)** và nổi bật với kỹ năng rê bóng Dribbling = 97 cùng chỉ số tốc độ cao Acceleration = 91.
- Cầu thủ **Cristiano Ronaldo** (ID: 20801) đến từ Bồ Đào Nha, hiện đang thi đấu cho **Juventus**. Với cùng điểm tổng thể Overall = 94, anh có giá trị thị trường **€77M** và mức lương **€405K**. Ronaldo thuận chân phải, chơi ở vị trí **Striker (ST)**, sở hữu thể hình vượt trội với chiều cao **6'2** và cân nặng **183 lbs**, cùng khả năng dứt điểm xuất sắc Finishing = 94.

2.1.2 Các bước kiểm tra và xử lý dữ liệu

1. Kiểm tra missing values

Bước đầu tiên trong xử lý dữ liệu là kiểm tra các giá trị bị thiếu (missing values) trong bộ dữ liệu. Chúng tôi sử dụng lệnh **isnull()** trong Python để xác định số lượng giá trị bị thiếu ở từng cột.

Kết quả kiểm tra: Dưới đây là danh sách các cột có giá trị bị thiếu và số lượng giá trị NULL tương ứng:

Cột dữ liệu	Số lượng giá trị NULL
Club	241
Preferred Foot	48
International Reputation	48
Weak Foot	48
Skill Moves	48

Work Rate	48
Body Type	48
Position	60
Jersey Number	60
Joined	1,553
Loaned From	16,943
Contract Valid Until	289
Height	48
Weight	48
Release Clause	1,564

Ngoài ra, một số cột kỹ năng (ví dụ: **Crossing, Finishing, Dribbling, BallControl, Acceleration**) cũng có đúng **48 giá trị bị thiếu**, cùng các chỉ số liên quan đến thủ môn (**GKDividing, GKHandling, GKKicking, GKPositioning, GKReflexes**).

2. Xử lý missing values

Dựa trên kết quả kiểm tra, chúng tôi áp dụng các phương pháp xử lý như sau:

1. Đối với các cột có 48 giá trị NULL:

- Các cột như **Preferred Foot, Body Type, Height, Weight**, và các chỉ số kỹ năng có cùng số lượng giá trị bị thiếu (**48**) được giả định là bị trùng lặp dữ liệu.
- Chúng tôi quyết định xóa toàn bộ các hàng chứa **48 giá trị NULL** để làm sạch dữ liệu.

2. Đối với các cột có số lượng NULL lớn (**Loaned From, Joined, Release Clause**):

- **Loaned From:** Cột này chứa tới **16,943 giá trị NULL**, chiếm phần lớn bộ dữ liệu. Vì đây là thông tin không quan trọng trong phân tích, chúng tôi quyết định xóa toàn bộ cột này.

3. Phân tích cơ bản với hàm describe:

Sử dụng hàm describe trong Python, chúng tôi đã tính toán các thống kê cơ bản bao gồm số lượng (count), trung bình (mean), độ lệch chuẩn (std), giá trị nhỏ nhất (min), các phần tư (25%, 50%, 75%) và giá trị lớn nhất (max) của các cột dữ liệu số trong bộ dữ liệu

	ID	Age	Overall	Potential
count	18207.000000	18207.000000	18207.000000	18207.000000
mean	214298.338606	25.122206	66.238699	71.307299
std	29965.244204	4.669943	6.908930	6.136496
min	16.000000	16.000000	46.000000	48.000000
25%	200315.500000	21.000000	62.000000	67.000000
50%	221759.000000	25.000000	66.000000	71.000000
75%	236529.500000	28.000000	71.000000	75.000000
max	246620.000000	45.000000	94.000000	95.000000

	International Reputation	Weak Foot	Skill Moves	Jersey Number
count	18159.000000	18159.000000	18159.000000	18147.000000
mean	1.113222	2.947299	2.361308	19.546096
std	0.394031	0.660456	0.756164	15.947765
min	1.000000	1.000000	1.000000	1.000000
25%	1.000000	3.000000	2.000000	8.000000
50%	1.000000	3.000000	2.000000	17.000000
75%	1.000000	3.000000	3.000000	26.000000
max	5.000000	5.000000	5.000000	99.000000

	Crossing	Finishing ...	Penalties	Composure
count	18159.000000	18159.000000	...	18159.000000
mean	49.734181	45.550911	...	48.548598
std	18.364524	19.525820	...	15.704053
min	5.000000	2.000000	...	5.000000
25%	38.000000	30.000000	...	39.000000
50%	54.000000	49.000000	...	49.000000
75%	64.000000	62.000000	...	60.000000
max	93.000000	95.000000	...	92.000000

	Marking	StandingTackle	SlidingTackle	GKDividing
count	18159.000000	18159.000000	18159.000000	18159.000000
mean	47.281623	47.697836	45.661435	16.616223
std	19.904397	21.664004	21.289135	17.695349
min	3.000000	2.000000	3.000000	1.000000
25%	30.000000	27.000000	24.000000	8.000000
50%	53.000000	55.000000	52.000000	11.000000
75%	64.000000	66.000000	64.000000	14.000000
max	94.000000	93.000000	91.000000	90.000000

	GKHandling	GKKicking	GKPositioning	GKReflexes
count	18159.000000	18159.000000	18159.000000	18159.000000
mean	16.391596	16.232061	16.388898	16.710887
std	16.906900	16.502864	17.034669	17.955119
min	1.000000	1.000000	1.000000	1.000000
25%	8.000000	8.000000	8.000000	8.000000
50%	11.000000	11.000000	11.000000	11.000000
75%	14.000000	14.000000	14.000000	14.000000
max	92.000000	91.000000	90.000000	94.000000

2.1.3 Hướng sử dụng bộ dữ liệu trong báo cáo

1. Phân tích thông tin cầu thủ

- Hiện thị thông tin cơ bản của cầu thủ như tên, tuổi, quốc gia, vị trí thi đấu, và câu lạc bộ.
- Cung cấp bảng tóm tắt các chỉ số chuyên môn (Overall, Potential) và kỹ năng cá nhân (Finishing, Dribbling, Passing, etc.).

2. So sánh các cầu thủ

- So sánh trực quan các chỉ số:
Người dùng có thể chọn từ 2 đến nhiều cầu thủ để so sánh các chỉ số nổi bật (ví dụ: tốc độ, sút bóng, rê bóng) thông qua các biểu đồ dạng thanh hoặc radar.
 - Ví dụ: So sánh giữa Lionel Messi và Cristiano Ronaldo về kỹ năng rê bóng, dứt điểm, và chuyền bóng.
- So sánh theo vị trí thi đấu:
Cho phép so sánh cầu thủ cùng vị trí (ví dụ: tiền đạo với tiền đạo) để đánh giá các yếu tố chuyên môn hóa theo vai trò trên sân.

3. Phân tích theo nhóm cầu thủ

- Xếp hạng cầu thủ theo tiêu chí:
Hiển thị danh sách các cầu thủ hàng đầu dựa trên các tiêu chí như chỉ số tổng quát (Overall), tiềm năng (Potential), hoặc kỹ năng cụ thể (ví dụ: tốc độ, rê bóng, phản xạ).
- Phân cụm cầu thủ:
Phân chia cầu thủ thành các nhóm theo phong cách chơi hoặc chỉ số (ví dụ: cầu thủ tấn công nhanh, tiền vệ chuyên tốt, thủ môn phản xạ cao).

4. Phân tích chỉ số theo thời gian

- Hiển thị biểu đồ phát triển của các cầu thủ trẻ qua các năm, dựa trên chỉ số Potential và Overall.
 - Ví dụ: Dự đoán thời điểm đỉnh cao phong độ của một cầu thủ dựa trên dữ liệu hiện có.

5. Tìm kiếm cầu thủ tiềm năng

- Bộ lọc cầu thủ:
Dashboard sẽ cung cấp các bộ lọc cho phép tìm kiếm cầu thủ dựa trên độ tuổi, quốc gia, vị trí thi đấu, hoặc chỉ số cụ thể.
 - Ví dụ: Lọc ra các cầu thủ dưới 21 tuổi có tiềm năng trên 80 và chỉ số tốc độ cao hơn 85.
- Khám phá tài năng trẻ:
Xác định các cầu thủ trẻ có chỉ số Potential cao nhưng chỉ số Overall hiện tại chưa nổi bật, từ đó gợi ý các mục tiêu tuyển dụng tiềm năng.

6. Trực quan hóa và tương tác

- Biểu đồ radar: Hiển thị sự phân bố kỹ năng của cầu thủ, giúp người dùng dễ dàng nhận diện điểm mạnh và yếu.
- Biểu đồ thanh: So sánh cụ thể các chỉ số giữa cầu thủ.
- Bản đồ nhiệt: Hiển thị các kỹ năng nổi bật nhất của cầu thủ ở từng vị trí thi đấu.
- Giao diện tương tác: Cho phép người dùng nhấp chọn cầu thủ để xem thông tin chi tiết và tùy chỉnh so sánh theo nhu cầu.

7. Định hướng ứng dụng thực tế

- Hỗ trợ các câu lạc bộ và nhà quản lý: Dashboard có thể trở thành công cụ tham khảo để đánh giá và tuyển dụng cầu thủ.
- Ứng dụng trong trò chơi: Dashboard được thiết kế mô phỏng theo các tính năng trong game, giúp người dùng vừa phân tích dữ liệu vừa trải nghiệm cảm giác như đang "quản lý đội bóng" trong môi trường ảo.

Phương Pháp Đề Xuất Trực Quan Hóa Dữ Liệu

3.1 Bản Đề Xuất Trực Quan Hóa Dữ Liệu

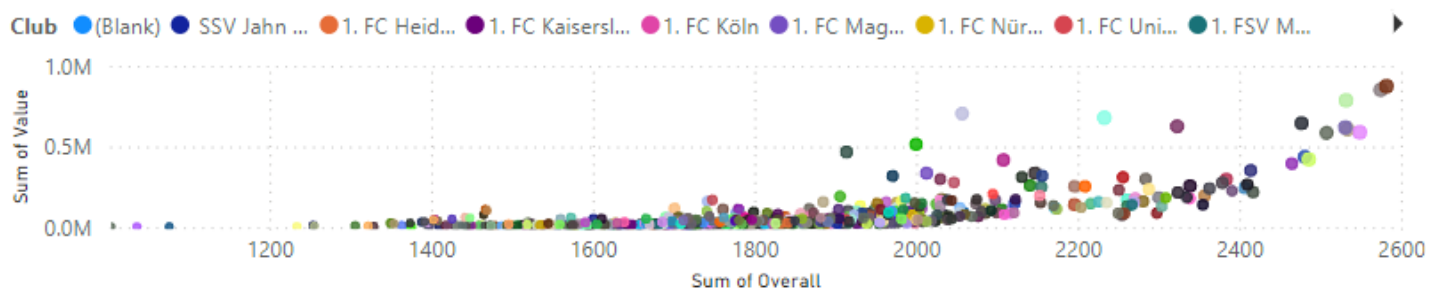
3.1.1. Bản đề xuất trực quan dữ liệu

Dựa trên bộ dữ liệu FIFA, chúng tôi đề xuất một hệ thống trực quan hóa toàn diện bao gồm 8 loại biểu đồ khác nhau:

1. Bubble Chart: Phân tích mối quan hệ giữa Overall, Wage, Value và International Reputation
2. Shape Map: Phân bố địa lý của cầu thủ trên toàn cầu
3. Radar Chart và Bar Chart: Phân tích chi tiết các chỉ số của cầu thủ
4. Tunnel Chart: Phân tích phân bố dạng cơ thể
5. Donut Chart: Tỷ lệ cầu thủ theo vị trí thi đấu
6. Pie Chart: Tỷ lệ cầu thủ theo chân thuận
7. Histogram with Points Chart: Mối quan hệ giữa Value và Release Clause
8. Line and Stacked Column Chart: So sánh tổng giá trị và điều khoản giải phóng
9. Histogram Chart: Phân tích giá trị theo độ tuổi

Bubble Chart:

Sum of Overall, Sum of Value and Sum of International Reputation by Club



Biểu đồ Bubble Chart được sử dụng để phân tích mối quan hệ giữa các yếu tố quan trọng trong việc đánh giá cầu thủ, bao gồm: chỉ số tổng thể (Overall), mức lương (Wage), giá trị thị trường (Value) và uy tín quốc tế (International Reputation). Việc trực quan hóa các yếu tố này trong một biểu đồ duy nhất cho phép dễ dàng nhận diện các nhóm cầu thủ, phát hiện mối quan hệ giữa các chỉ số, cũng như tìm kiếm những bất thường hoặc ngoại lệ trong dữ liệu.

2. Các mục tiêu trực quan mong muốn đạt được

Mục tiêu đầu tiên của việc sử dụng biểu đồ Bubble Chart là để **phân tích mối quan hệ giữa chỉ số tổng thể (Overall) và mức lương (Wage)** của cầu thủ. Việc này giúp xác định xem cầu thủ có chỉ số cao có xu hướng nhận mức lương cao hay không và ngược lại, từ đó tìm ra những bất thường trong định giá của các cầu lạc bộ.

Mục tiêu tiếp theo là phân tích **tác động của uy tín quốc tế (International Reputation)** đối với mức lương và giá trị chuyển nhượng của cầu thủ. Cầu thủ có uy tín quốc tế cao có thể có mức lương hoặc giá trị thị trường vượt trội so với những cầu thủ có chỉ số tương đương nhưng ít được biết đến.

Một mục tiêu quan trọng khác là **phân nhóm cầu thủ** dựa trên các đặc điểm như chỉ số, lương, giá trị và uy tín quốc tế. Việc này giúp phân loại cầu thủ thành các nhóm như “ngôi sao” (chỉ số cao, lương và giá trị cao) và “cầu thủ tiềm năng” (chỉ số thấp nhưng giá trị và uy tín quốc tế cao). Điều này hỗ trợ câu lạc bộ trong việc ra quyết định về chiến lược chuyển nhượng và tuyển dụng.

Cuối cùng, biểu đồ Bubble Chart cũng hỗ trợ việc **phát hiện các outliers** (các giá trị bất thường), chẳng hạn như cầu thủ có mức lương hoặc giá trị không tương xứng với chỉ số tổng thể và uy tín quốc tế. Những outliers này có thể là dấu hiệu của các vấn đề trong chiến lược quản lý nhân sự hoặc định giá cầu thủ.

3. Tại sao chart phù hợp với mục tiêu đã đề ra

Bubble Chart là lựa chọn phù hợp vì nó cho phép trực quan hóa mối quan hệ giữa nhiều biến đồng thời trong một biểu đồ duy nhất. Các cầu thủ được thể hiện dưới dạng các vòng tròn (bubbles), trong đó:

- Trục x thể hiện chỉ số tổng thể (Overall).
- Trục y thể hiện mức lương (Wage).
- Kích thước của các vòng tròn biểu thị giá trị thị trường (Value).
- Màu sắc của các vòng tròn biểu thị uy tín quốc tế (International Reputation).

Biểu đồ này cho phép nhận diện dễ dàng các nhóm cầu thủ khác nhau, quan sát sự phân bố của các cầu thủ theo các tiêu chí quan trọng, đồng thời phát hiện những mối quan hệ hay bất thường giữa các chỉ số.

4. Nhận xét và kết luận

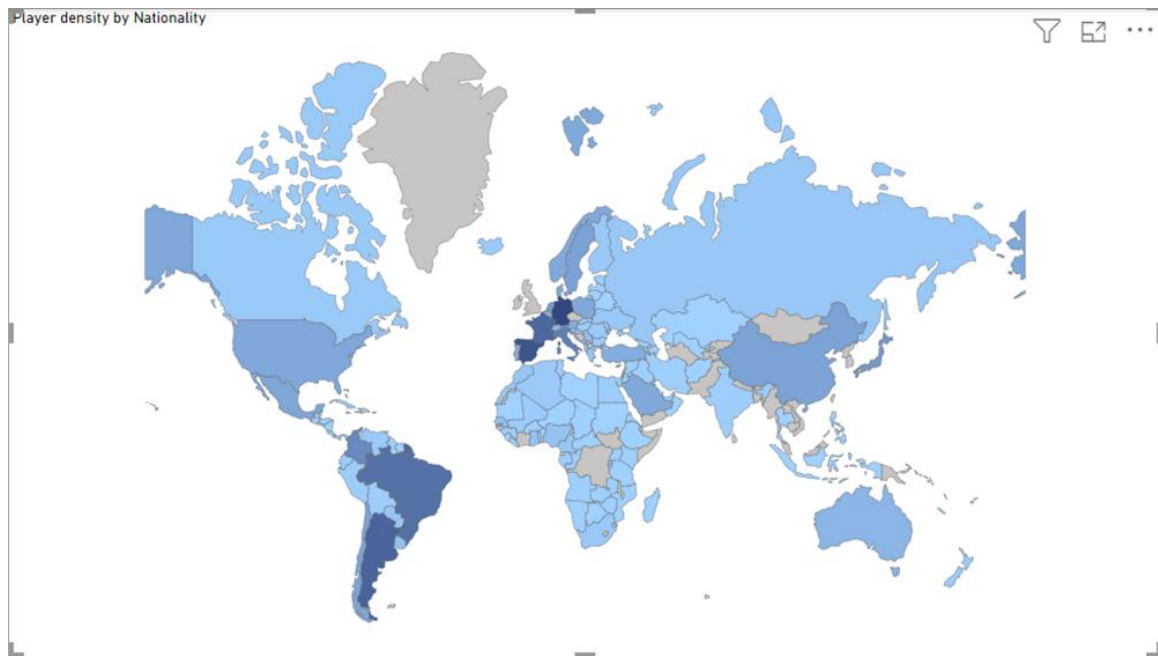
Sau khi phân tích biểu đồ Bubble Chart, ta có thể rút ra một số nhận xét quan trọng:

- Các nhóm cầu thủ có mức uy tín quốc tế thấp chiếm đa số trong dữ liệu. Hầu hết các nhóm cầu thủ theo câu lạc bộ có tổng điểm chỉ số Overall từ 1400 đến 2400, và giá trị thị trường dao động trong khoảng 0.2M đến 0.4M.
- Mối quan hệ giữa chỉ số tổng thể (Overall) và giá trị thị trường (Value) là **tương quan tích cực**: khi chỉ số tổng thể cao, giá trị thị trường của cầu thủ cũng cao. Tuy nhiên, nhóm cầu thủ có giá trị thị trường cao nhưng chỉ số tổng thể ở mức trung bình chiếm tỷ lệ nhỏ, chứng tỏ rằng giá trị thị trường không hoàn toàn phụ thuộc vào chỉ số tổng thể.

- Một số **outliers** vẫn tồn tại, như các cầu thủ có mức lương cao hoặc giá trị thị trường lớn nhưng chỉ số tổng thể thấp, cho thấy các yếu tố ngoài năng lực như danh tiếng hoặc chiến lược chuyển nhượng cũng có ảnh hưởng đáng kể.

Từ kết quả này, có thể kết luận rằng khi tuyển một cầu thủ, câu lạc bộ không chỉ cần xem xét chỉ số tổng thể mà còn cần cân nhắc thêm các yếu tố như uy tín quốc tế, giá trị thương mại, và chiến lược phát triển đội bóng. Việc phát hiện các outliers và phân loại cầu thủ theo các nhóm đặc điểm sẽ giúp câu lạc bộ xây dựng chiến lược nhân sự và chuyển nhượng hiệu quả hơn.

Shape chart



Biểu đồ Shape Chart sử dụng bản đồ thế giới với màu sắc thay đổi từ nhạt đến đậm để biểu thị số lượng cầu thủ bóng đá tại từng quốc gia. Màu nhạt đại diện cho ít cầu thủ và màu đậm cho nhiều cầu thủ, giúp dễ dàng nhận diện sự phân bố cầu thủ. Biểu đồ Shape Chart là công cụ lý tưởng để thể hiện sự phân bố cầu thủ theo quốc gia, giúp người xem nhanh chóng nhận ra các quốc gia có số lượng cầu thủ cao hoặc thấp.

Mục tiêu trực quan:

- Nhận diện sự khác biệt số lượng cầu thủ giữa các quốc gia.
- Xác định quốc gia có số cầu thủ cao nhất và thấp nhất.
- Phân tích lý do sự khác biệt dựa trên các yếu tố như văn hóa thể thao, cơ sở hạ tầng, và tiếp xúc với thể thao.

Nhận xét và kết luận:

- Các quốc gia có nhiều cầu thủ nhất là ở Châu Âu và Nam Mỹ, trong khi Châu Á và Châu Phi có ít cầu thủ hơn.
- Một số quốc gia như Greenland không có cầu thủ.
- Sự khác biệt có thể do văn hóa thể thao, cơ sở hạ tầng, và chính sách phát triển thể thao.
- Việc phát triển thể thao ở những quốc gia ít cầu thủ cần sự đầu tư vào cơ sở hạ tầng và đào tạo.

Biểu đồ Radar Chart: Đánh giá toàn diện các chỉ số cầu thủ



Biểu đồ Radar Chart sử dụng hình đa giác để thể hiện các chỉ số của cầu thủ, với mỗi đỉnh của đa giác đại diện cho một chỉ số. Các chỉ số có giá trị cao sẽ đẩy các đỉnh ra xa trung tâm, giúp so sánh các chỉ số của cầu thủ một cách trực quan. Người dùng có thể lựa chọn cầu thủ cụ thể qua Slicer hoặc Filter, nếu không chọn, biểu đồ sẽ hiển thị chỉ số trung bình của tất cả cầu thủ.

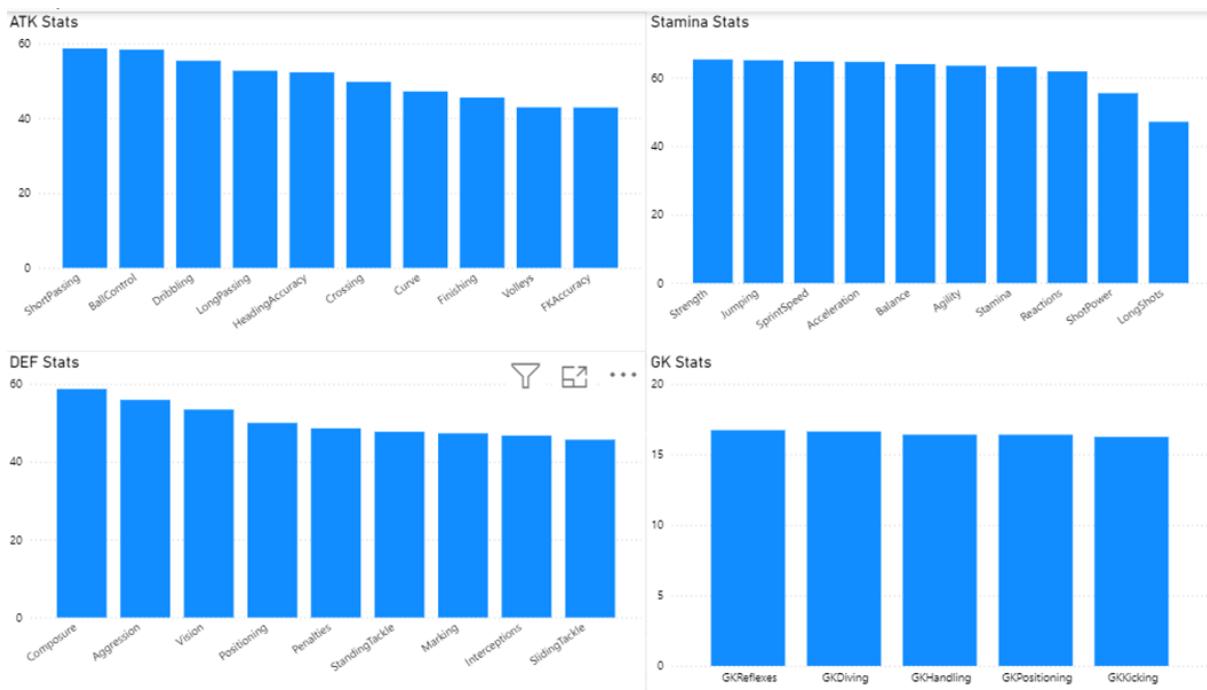
Mục tiêu trực quan

- Đánh giá sự toàn diện của cầu thủ ở các chỉ số khác nhau như tấn công, phòng thủ, sức mạnh, và thủ môn.
- So sánh các cầu thủ để nhận diện sự khác biệt về kỹ năng.
- Xác định điểm mạnh và yếu của các cầu thủ để hỗ trợ quá trình sắp xếp vào các vị trí phù hợp trong đội hình.

Nhận xét và kết luận

- **Chỉ số tấn công:** Cầu thủ có chỉ số Short Passing (Kỹ năng chuyền ngắn) cao nhất, các chỉ số còn lại tương đương nhau.
- **Chỉ số phòng thủ:** Chỉ số Composure (Sự bình tĩnh khi xử lý tình huống) cao nhất, các chỉ số còn lại gần như tương đương.
- **Chỉ số sức mạnh:** Chỉ số Strength (Sức mạnh cơ bắp) cao nhất, trong khi Long Shots (Kỹ năng sút xa) thấp nhất.
- **Chỉ số thủ môn:** Các chỉ số của thủ môn gần như đồng đều.
- Cầu thủ có ATK Stats vượt trội thích hợp cho vị trí tiền đạo, trong khi DEF Stats cao sẽ phù hợp với vị trí hậu vệ.

Biểu đồ Bar Chart



Biểu đồ Bar Chart được sử dụng để thể hiện chi tiết các giá trị của các chỉ số cầu thủ, những chỉ số đã được thể hiện trong biểu đồ Radar Chart. Mỗi thanh trong biểu đồ sẽ tương ứng với một chỉ số và có độ dài phản ánh giá trị của chỉ số đó.

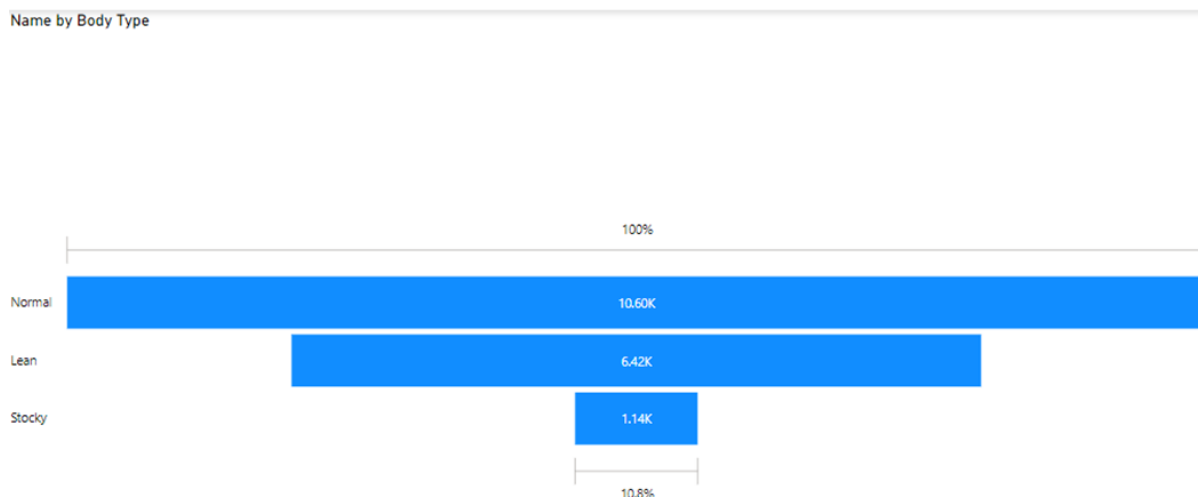
Mục tiêu trực quan

- **Xem chi tiết giá trị các chỉ số cầu thủ,** đặc biệt là các chỉ số được thể hiện trong Radar Chart.
- **Khắc phục điểm yếu** của Radar Chart khi không thể hiện rõ ràng giá trị chi tiết của từng chỉ số.

Nhận xét và kết luận

Bar Chart giúp làm rõ chi tiết các chỉ số của cầu thủ, hỗ trợ việc so sánh các giá trị chính xác giữa các cầu thủ. Biểu đồ này khắc phục được nhược điểm của Radar Chart khi không thể hiện đầy đủ các giá trị của các chỉ số cụ thể, từ đó giúp người dùng dễ dàng phân tích và đánh giá cầu thủ một cách chi tiết hơn.

Biểu đồ Tunnel Chart



Biểu đồ Tunnel Chart được sử dụng để thể hiện số lượng cầu thủ bóng đá được phân loại theo dạng cơ thể. Ba nhóm chính trong biểu đồ bao gồm: Normal (bình thường), Lean (thon gọn) và Stocky (vạm vỡ). Biểu đồ này giúp phân tích sự phân bố dạng cơ thể trong bóng đá và nhận diện các xu hướng cơ thể của cầu thủ ngoài ra biểu đồ cũng giúp nhận diện rõ ràng tỷ lệ phân bố các dạng cơ thể trong bóng đá, từ đó đưa ra các chiến lược đào tạo và lựa chọn vị trí thi đấu phù hợp.

Mục tiêu trực quan

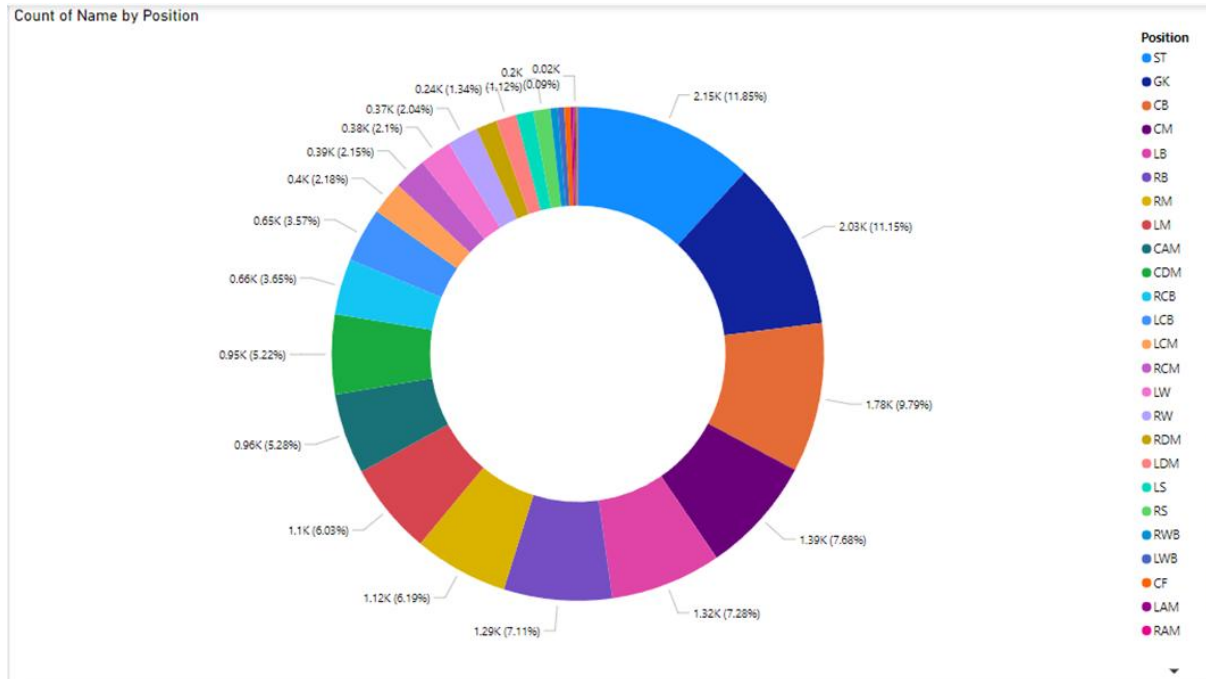
- **Phân tích sự phân bố dạng cơ thể** trong bóng đá.
- **Xác định các dạng cơ thể phổ biến hoặc hiếm** trong thể thao, hỗ trợ việc đào tạo và tìm kiếm vị trí thi đấu phù hợp với từng dạng cơ thể.

Nhận xét và kết luận

- **Nhóm Normal:** Chiếm số lượng lớn nhất (100%), phù hợp với đa dạng vị trí và phong cách thi đấu.
- **Nhóm Lean:** Chiếm khoảng 60.6% so với nhóm Normal, đóng vai trò quan trọng ở những vị trí đòi hỏi tốc độ hoặc sức mạnh vượt trội.

- **Nhóm Stocky:** Chiếm tỷ lệ nhỏ nhất (10.8% so với nhóm Normal), phản ánh rằng vóc dáng to lớn không phổ biến trong bóng đá hiện đại, nơi sự nhanh nhẹn và kỹ thuật được đánh giá cao hơn.

Biểu đồ Donut Chart



Biểu đồ Donut Chart được sử dụng để thể hiện tỷ lệ số cầu thủ phân bố giữa các vị trí thi đấu trong bóng đá. Biểu đồ này giúp so sánh sự phân bố số lượng cầu thủ ở từng vị trí, từ đó đánh giá được xu hướng sử dụng cầu thủ trong các vị trí khác nhau.

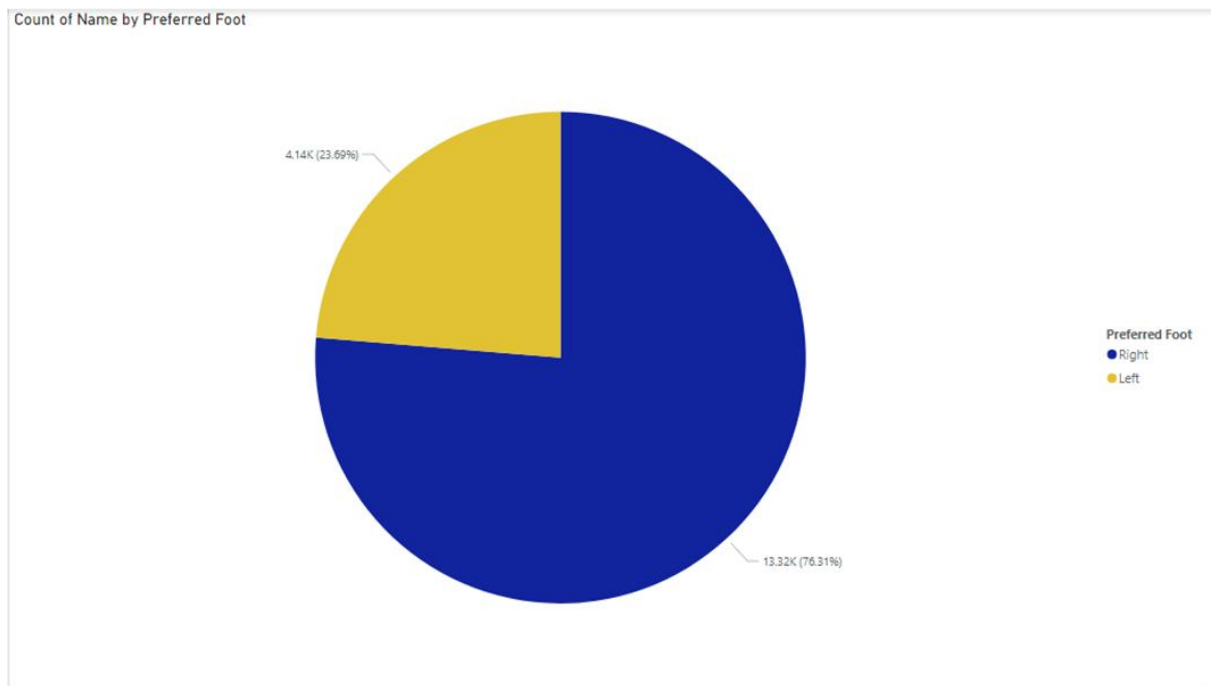
Mục tiêu trực quan

- **Xác định vị trí thi đấu phổ biến nhất và ít phổ biến nhất.**
- **Đánh giá xu hướng sử dụng cầu thủ ở các vị trí thi đấu khác nhau.**

Nhận xét và kết luận

- **Vị trí ST (Striker):** Chiếm tỷ lệ lớn nhất, cho thấy vị trí này là một trong những vị trí thi đấu quan trọng và phổ biến nhất trong bóng đá.
- **Vị trí RF (Right Forward):** Chiếm tỷ lệ nhỏ nhất, phản ánh rằng đây là một vị trí ít phổ biến trong đội hình, có thể chỉ xuất hiện trong một số chiến thuật cụ thể.

Biểu đồ Pie Chart



Biểu đồ Pie Chart được sử dụng để thể hiện tỷ lệ cầu thủ thuận chân phải so với cầu thủ thuận chân trái trong bóng đá. Biểu đồ này giúp trực quan hóa sự phân bố của các cầu thủ theo tính chất thuận tay/chân của họ, từ đó đưa ra những đánh giá về xu hướng và tác động đến chiến thuật thi đấu.

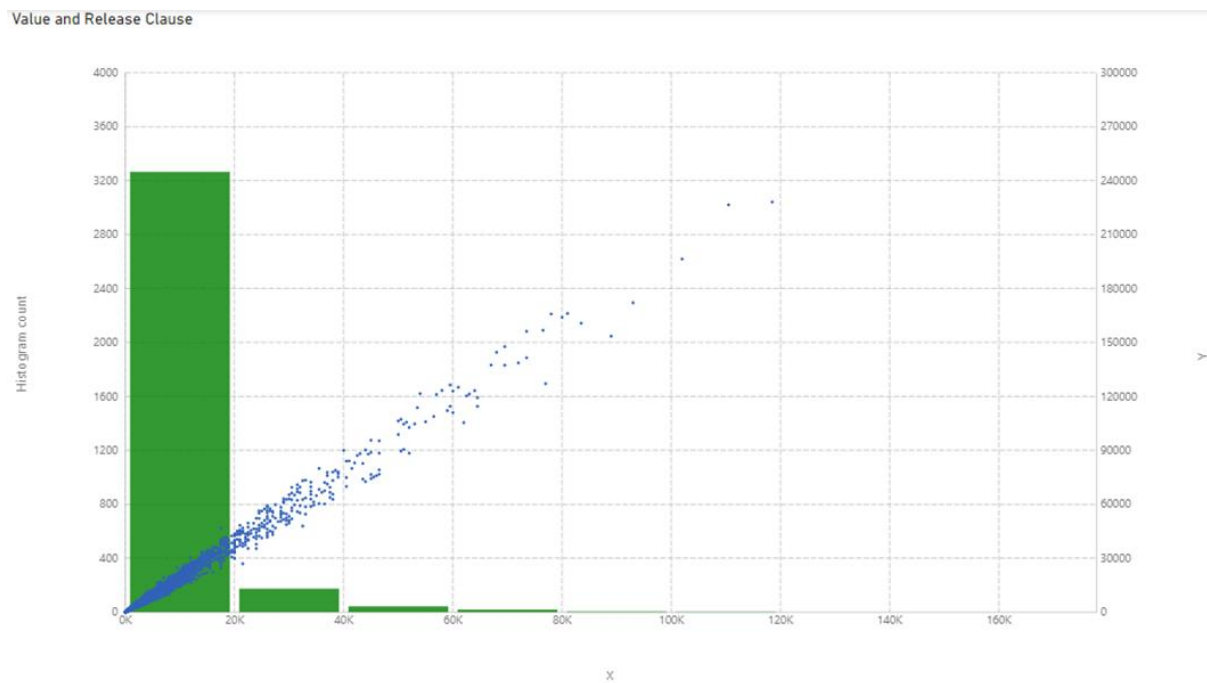
Mục tiêu trực quan

- **Xác định xu hướng chung** về tỉ lệ cầu thủ thuận chân phải và chân trái trong bóng đá.
- **Đánh giá sự chênh lệch** giữa các nhóm và tác động của nó đến các chiến thuật thi đấu.

Nhận xét và kết luận

- **Đại đa số cầu thủ thuận chân phải:** Điều này phản ánh sự phổ biến tự nhiên, vì phần lớn dân số thế giới cũng thuận tay/chân phải.
- **Nhóm cầu thủ thuận chân trái** chiếm tỷ lệ ít hơn: Tuy nhiên, cầu thủ thuận chân trái có thể tạo ra những chiến thuật mới lạ, bổ sung sự đa dạng và bất ngờ cho đội hình.

Biểu đồ Histogram with Points Chart



Biểu đồ Histogram with Points Chart kết hợp hai loại biểu đồ: scatter plot để thể hiện mối quan hệ giữa giá trị X (Value) và giá trị Y (Release Clause), và histogram để thể hiện tần suất phân bố của các giá trị X (Value) tại các khoảng giá trị khác nhau.

Mục tiêu trực quan

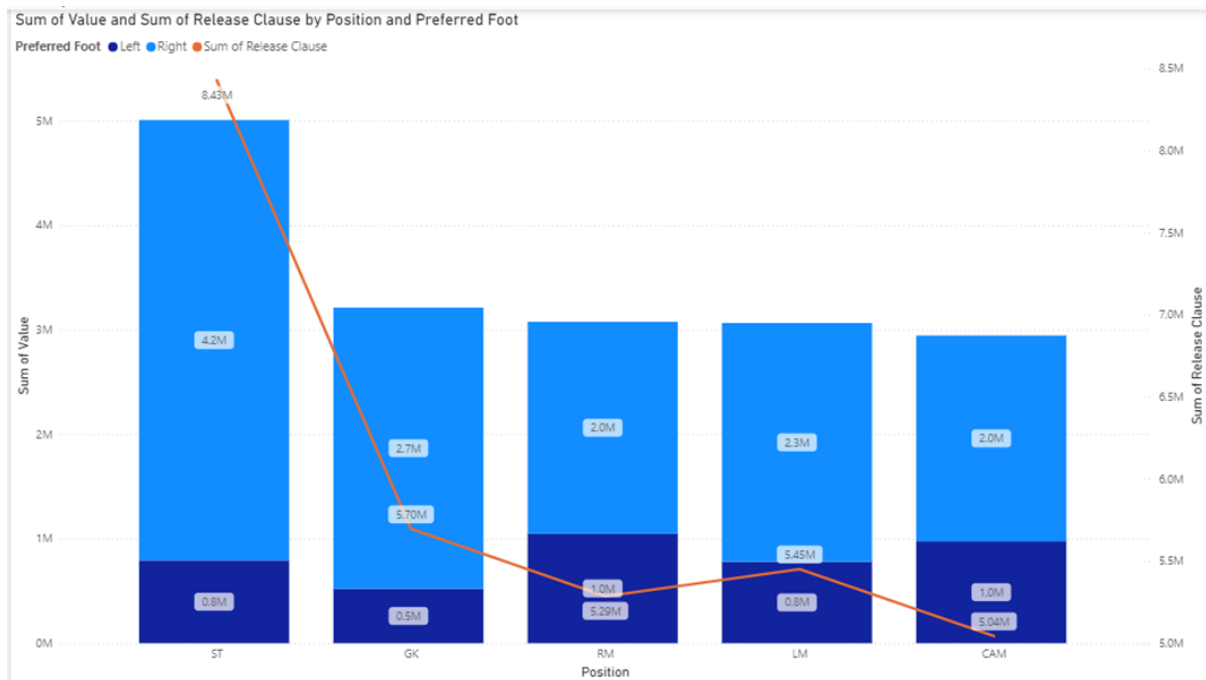
- **Phân tích mối tương quan** giữa giá trị cầu thủ (Value) và điều khoản giải phóng hợp đồng (Release Clause).
- **Quan sát sự phân bố** của các giá trị X (Value), hiểu rõ về tần suất xuất hiện của các giá trị này trong các khoảng giá trị khác nhau.

Nhận xét và kết luận

- **Mối quan hệ giữa X (Value) và Y (Release Clause):** Có thể thấy một xu hướng tăng, khi giá trị cầu thủ (Value) cao hơn thì điều khoản giải phóng hợp đồng (Release Clause) cũng có xu hướng tăng theo.
- **Phân bố không đồng đều:** Các giá trị X (Value) không phân bố đều, một số giá trị xuất hiện với tần suất cao, trong khi một số giá trị lại ít phổ biến.
- **Điểm bất thường:** Một số điểm có Release Clause rất cao nhưng giá trị cầu thủ (X) không tăng tương ứng, chỉ ra những bất thường trong cách định giá hoặc chính sách của các câu lạc bộ.
- **Ứng dụng thực tế:** Biểu đồ giúp đánh giá mối quan hệ giữa giá trị cầu thủ và điều khoản giải phóng hợp đồng, từ đó hỗ trợ các quyết định chiến lược trong việc đánh giá và tuyển dụng cầu thủ.

Line and stacked column chart

Biểu đồ Line and Stacked Column Chart so sánh tổng giá trị cầu thủ (Sum of Value) và tổng điều khoản giải phóng (Sum of Release Clause) theo từng vị trí cầu thủ (Position) và chân thuận (Preferred Foot). Trục X thể hiện các vị trí cầu thủ, trục Y trái thể hiện tổng giá trị cầu thủ, trục Y phải thể hiện tổng điều khoản giải phóng.



Mục tiêu trực quan

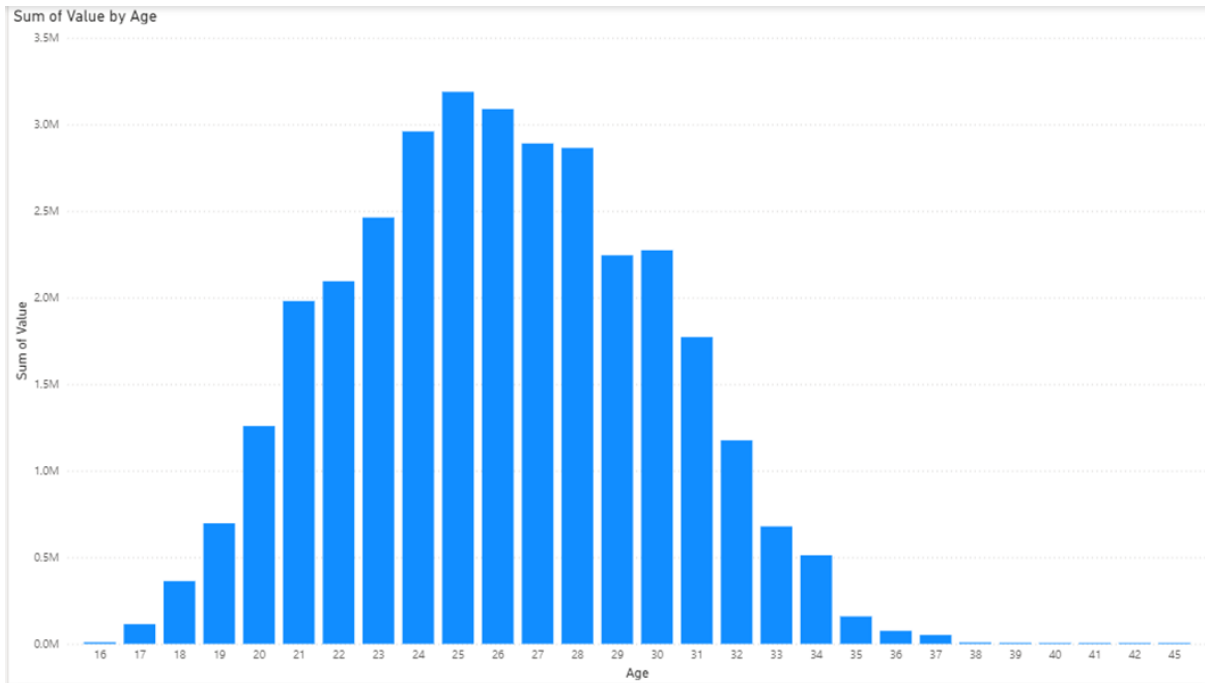
- So sánh tổng giá trị và tổng điều khoản giải phóng của các cầu thủ theo vị trí thi đấu, giúp đánh giá sự khác biệt về tổng giá trị và tổng điều khoản giải phóng ở từng vị trí.
- Ảnh hưởng của chân thuận (Preferred Foot): Xem xét sự tác động của việc chân thuận đến tổng giá trị và điều khoản giải phóng của cầu thủ ở các vị trí khác nhau.

Nhận xét và kết luận

- Vị trí ST: Đây là vị trí có tổng giá trị cầu thủ cao nhất, với chân phải chiếm phần lớn so với chân trái. Tổng điều khoản giải phóng cũng cao nhất cho vị trí này.
- Các vị trí khác: Tổng giá trị của các vị trí còn lại khá đồng đều, nhưng chân phải vẫn chiếm ưu thế. Vị trí CAM có tổng điều khoản giải phóng thấp nhất trong số các vị trí.
- Ứng dụng thực tế: Biểu đồ giúp phân tích giá trị và điều khoản giải phóng của cầu thủ dựa trên vị trí thi đấu và chân thuận, từ đó đưa ra chiến lược hợp lý cho đội bóng.

Biểu đồ Histogram Chart

Biểu đồ Histogram Chart thể hiện tổng giá trị cầu thủ theo độ tuổi, trong đó trục X biểu thị độ tuổi (Age) và trục Y biểu thị tổng giá trị (Sum of Value) của các cầu thủ.



Mục tiêu trực quan

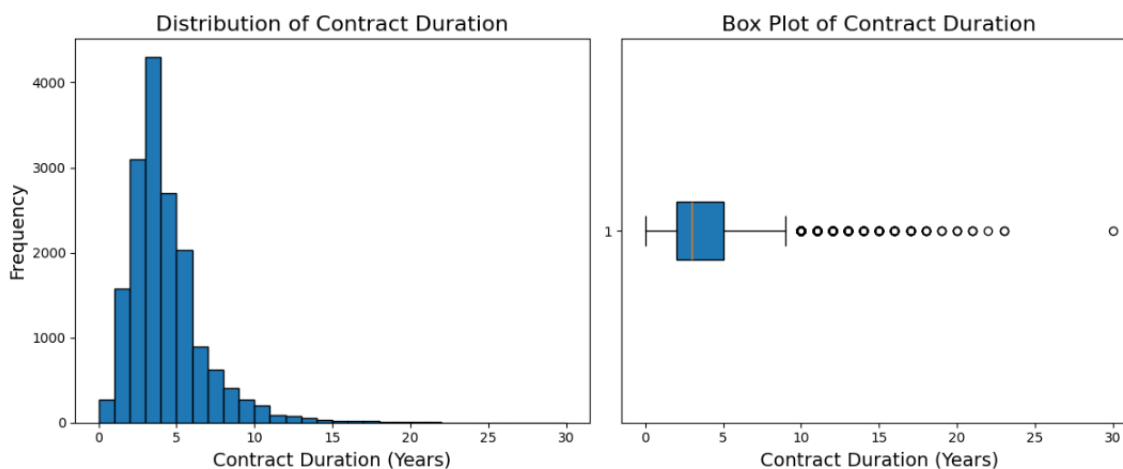
- **Xác định xu hướng giá trị theo độ tuổi:** Phân tích mối quan hệ giữa độ tuổi và tổng giá trị, từ đó tìm ra nhóm tuổi có giá trị cao nhất.
- **Hỗ trợ ra quyết định:** Xác định độ tuổi tiềm năng để đầu tư hoặc phát triển cầu thủ.

Nhận xét và kết luận

- **Phân bố chuẩn:** Biểu đồ có dạng phân bố chuẩn với đỉnh ở độ tuổi 24-25, tổng giá trị cầu thủ tăng nhanh từ độ tuổi 16 và giảm từ độ tuổi 26 trở đi.
- **Nhóm tuổi vàng:** Nhóm tuổi từ 22 đến 28 chiếm phần lớn tổng giá trị, chứng tỏ đây là độ tuổi vàng để đầu tư và phát triển cầu thủ.
- **Ứng dụng thực tế:** Biểu đồ giúp tối ưu hóa chiến lược mua bán cầu thủ, xác định xu hướng phát triển để đưa ra quyết định hợp lý trong việc đầu tư vào các cầu thủ trẻ hoặc kinh nghiệm.

Biểu đồ Histogram và Boxplot

- Histogram: Để phân tích phân phối thời gian hợp đồng Contract_Duration của các cầu thủ, giúp hiểu rõ về thời gian hợp đồng phổ biến.
- Box Plot: Để phát hiện các giá trị ngoại lệ (hợp đồng quá ngắn hoặc quá dài) và đánh giá sự biến động của thời hạn hợp đồng.



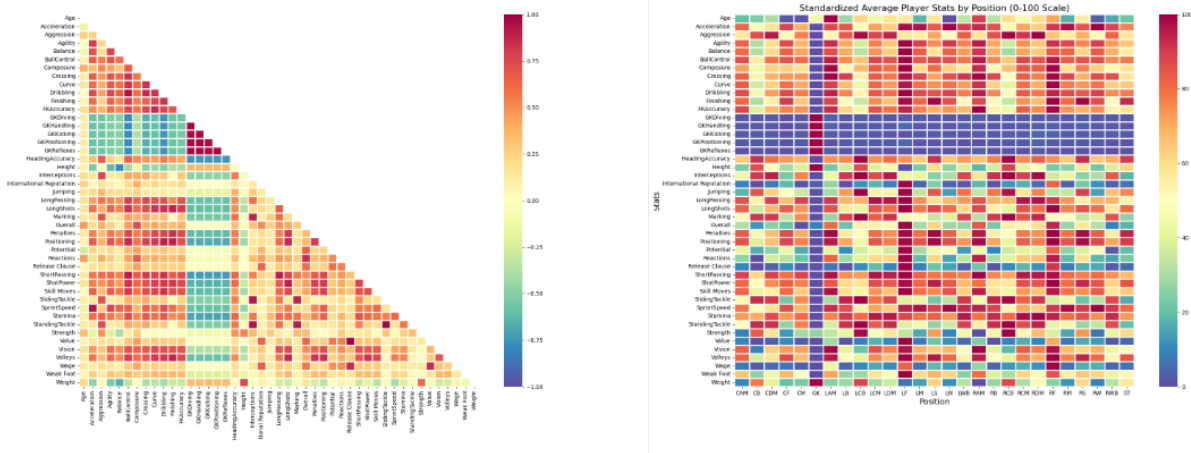
Mục tiêu trực quan

- **Histogram:** Xác định khoảng thời gian hợp đồng phổ biến nhất, ví dụ: từ 3-5 năm.
- **Box Plot:** Nhận biết các hợp đồng bất thường (hợp đồng quá dài hoặc quá ngắn) và độ phân tán của dữ liệu.

Nhận xét và kết luận.

- **Histogram:**
 - Nền đỉnh cao ở khoảng 3-5 năm, phần lớn cầu thủ có hợp đồng trong khoảng thời gian này.
 - Đuôi dài về bên phải cho thấy một số ít cầu thủ có hợp đồng dài hơn.
- **Box Plot:**
 - Giá trị ngoại lệ phía bên phải thể hiện một số cầu thủ có hợp đồng dài hơn đáng kể so với phần lớn.

Biểu đồ heatmap về sự tương quan giữa các chỉ số



Mô tả biểu đồ Heatmap

- **Heatmap 1:** Thể hiện ma trận tương quan giữa các biến số trong DataFrame. Biểu đồ này giúp nhận diện mối quan hệ tuyến tính giữa các chỉ số cầu thủ.
- **Heatmap 2:** Thể hiện giá trị trung bình đã được chuẩn hóa (trên thang điểm 0-100) của các chỉ số cầu thủ theo từng vị trí thi đấu. Biểu đồ này giúp so sánh và đánh giá các chỉ số giữa các vị trí khác nhau.

Thông tin mong muốn từ biểu đồ

- **Heatmap 1:** Hiển thị mối tương quan giữa các chỉ số cầu thủ và giúp huấn luyện viên hiểu mối quan hệ giữa các yếu tố quan trọng.
- **Heatmap 2:** So sánh các chỉ số cầu thủ theo từng vị trí thi đấu, giúp xác định vị trí nào có các chỉ số vượt trội.

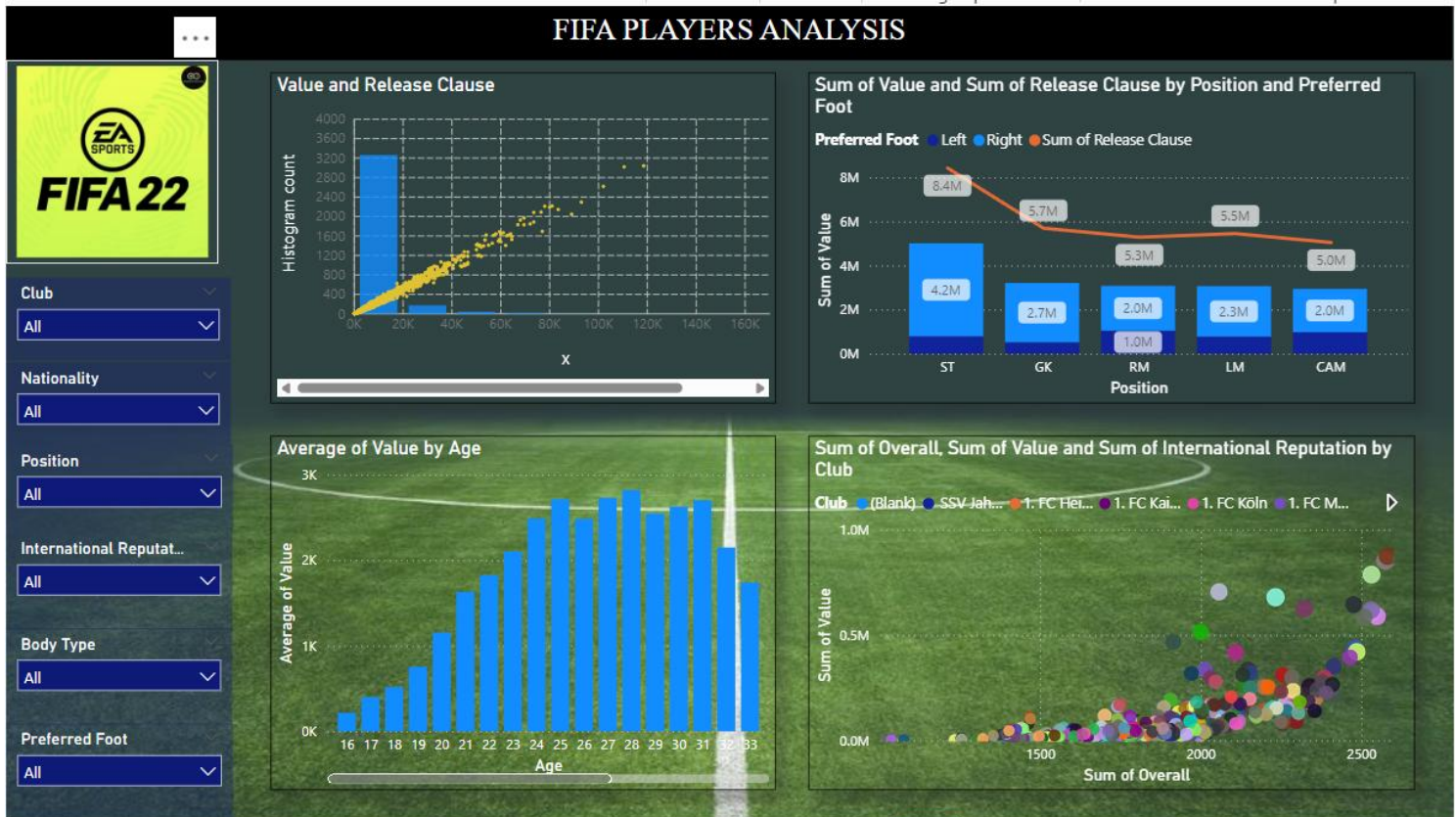
Nhận xét và kết luận

- **Tương quan dương mạnh:**
 - Các cặp biến số có giá trị tương quan gần +1, cho thấy chúng có xu hướng tăng hoặc giảm cùng nhau. Ví dụ: **Overall** và **Potential** thường có tương quan dương mạnh, tức là cầu thủ có chỉ số tiềm năng cao thường có chỉ số tổng thể cao.
- **Tương quan âm mạnh:**

- Các cặp biến số có giá trị tương quan gần -1, cho thấy chúng có xu hướng ngược chiều nhau. Ví dụ: **Age** và **Potential** có thể có tương quan âm, vì cầu thủ trẻ thường có tiềm năng phát triển tốt hơn so với những cầu thủ lớn tuổi.
- **Tương quan yếu:**
 - Các cặp biến số có giá trị tương quan gần 0 cho thấy không có mối quan hệ rõ ràng, như giữa **Height** và **Weight**, những yếu tố này không có sự ảnh hưởng rõ rệt đến nhau.

NỘI CÁC BIỂU ĐỒ THÀNH DASHBOARD

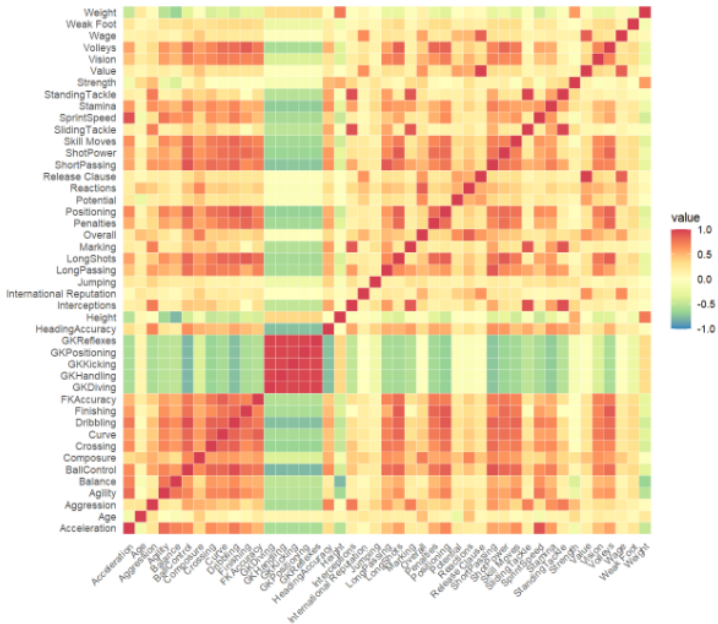
4.2. Dashboard



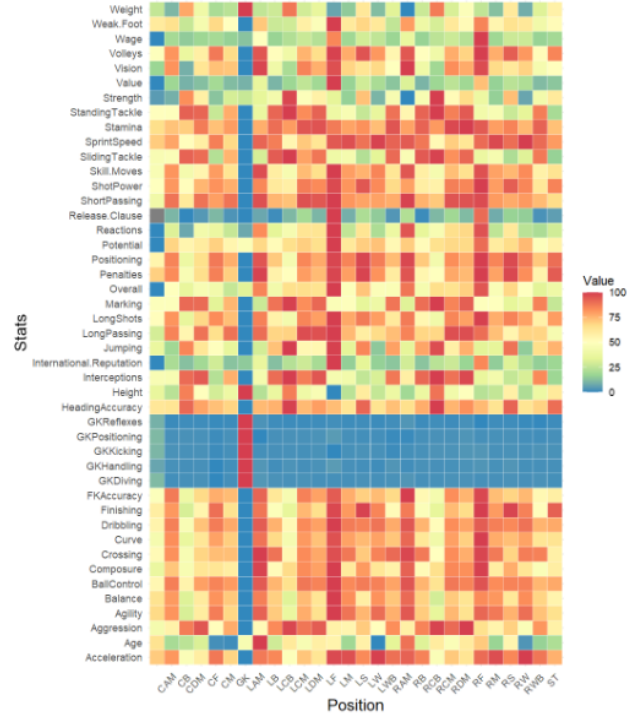
- Cầu thủ có điểm Overall cao nhất
- Cầu thủ có lương cao nhất
- Cầu thủ đáng giá nhất
- Cầu thủ tiềm năng cao nhất
- Cầu thủ lớn tuổi nhất
- ...

FIFA PLAYERS ANALYSIS

Correlation heatmap of stats vs stats



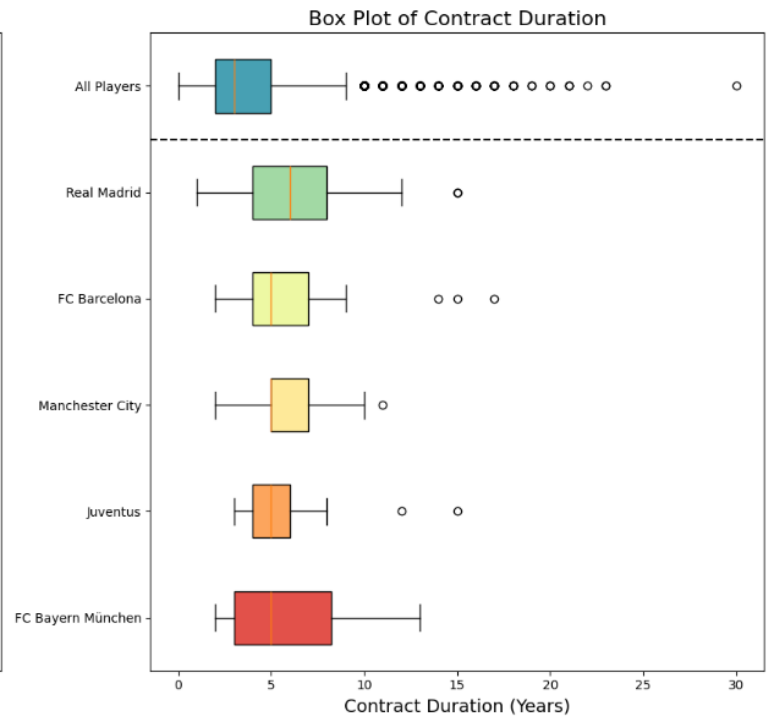
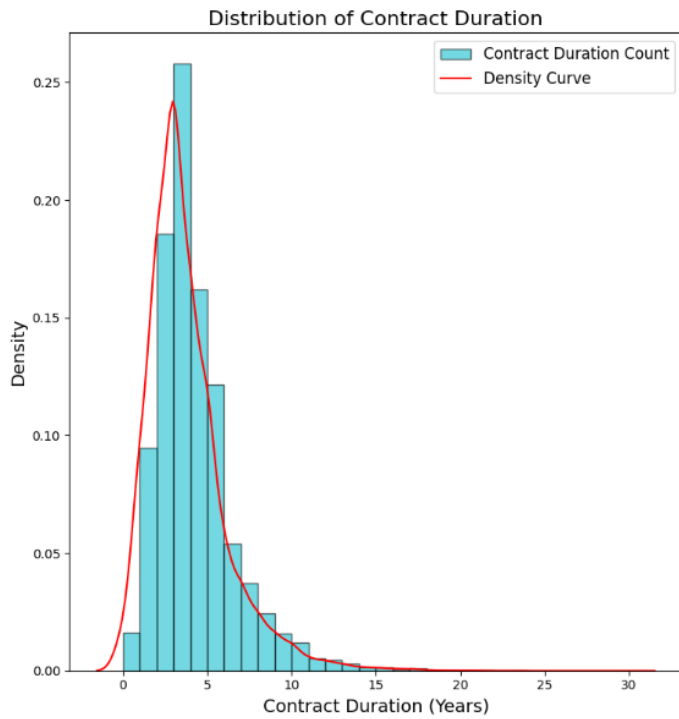
Correlation heatmap of stats vs positions



FIFA PLAYERS ANALYSIS



FIFA PLAYERS ANALYSIS



FIFA PLAYERS ANALYSIS



Triển khai

Để triển khai mô hình ta sử dụng Streamlit [5]

5.1 Giới thiệu Streamlit

Streamlit là một framework mã nguồn mở dành cho Python, cho phép các nhà phát triển dễ dàng tạo ra các ứng dụng web tương tác để trình bày các mô hình học máy và dữ liệu. Streamlit nổi bật nhờ tính đơn giản, dễ sử dụng và khả năng tích hợp mạnh mẽ với các công cụ dữ liệu phổ biến.



Dưới đây là một số điểm chính về Streamlit:

Cách tính năng chính của Streamlit

- **Dễ Dàng Sử Dụng:** Streamlit không yêu cầu kiến thức về HTML, CSS hay JavaScript. Bạn chỉ cần viết mã Python, và Streamlit sẽ tự động tạo ra giao diện web cho bạn.
- **Tương Tác Theo Thời Gian Thực:** Streamlit cho phép bạn tạo các giao diện người dùng tương tác theo thời gian thực mà không cần phải làm mới lại trang web. Bạn có thể dễ dàng tạo các widget tương tác như sliders, text inputs, buttons, v.v.
- **Tích Hợp Tốt với Các Thư Viện Dữ Liệu:** Streamlit tích hợp dễ dàng với các thư viện phân tích và hiển thị dữ liệu như Pandas, NumPy, Matplotlib, Plotly, Altair, và nhiều thư viện khác.
- **Cộng Đồng và Tài Liệu Hỗ Trợ:** Streamlit có cộng đồng người dùng rộng lớn và tài liệu phong phú, giúp bạn dễ dàng tìm kiếm hỗ trợ và học hỏi.

Lợi ích khi sử dụng Streamlit

- **Nhanh Chóng Phát Triển Ứng Dụng:** Với Streamlit, bạn có thể biến các script Python thành các ứng dụng web đẹp mắt và chuyên nghiệp chỉ trong vài phút.

- **Chia Sẻ Dễ Dàng:** Các ứng dụng Streamlit có thể dễ dàng được chia sẻ với người khác thông qua việc triển khai trên các dịch vụ đám mây hoặc máy chủ nội bộ.
- **Tùy Chỉnh Cao:** Bạn có thể dễ dàng tùy chỉnh giao diện và chức năng của ứng dụng để phù hợp với nhu cầu cụ thể của dự án.

5.2. Những thứ còn thiếu sót

Nhóm em sẽ còn cải thiện thêm về phần streamlit nữa

Kết luận và hướng phát triển

Kết luận

Bài làm không chỉ sử dụng các biểu đồ trực quan để phân tích dữ liệu cầu thủ bóng đá mà còn kết hợp chúng thành những dashboard tổng hợp, giúp trình bày thông tin một cách trực quan và dễ hiểu. Việc ghép nối các biểu đồ thành dashboard đã giúp tạo ra một cái nhìn tổng thể về các yếu tố quan trọng như giá trị cầu thủ, điều khoản giải phóng hợp đồng, các chỉ số về thể lực, vị trí thi đấu và độ tuổi. Dashboard này không chỉ hỗ trợ huấn luyện viên và các nhà quản lý bóng đá trong việc phân tích và đánh giá cầu thủ mà còn làm nổi bật những mối quan hệ phức tạp giữa các chỉ số, từ đó giúp đưa ra các quyết định chiến lược chính xác hơn.

Tổng kết:

Bài làm đã thành công trong việc ứng dụng các biểu đồ trực quan kết hợp với các dashboard phân tích dữ liệu cầu thủ bóng đá. Các dashboard này không chỉ giúp dễ dàng nhận diện và so sánh các chỉ số quan trọng giữa các cầu thủ mà còn cho phép người dùng có cái nhìn toàn diện về sự phân bố và mối tương quan giữa các chỉ số. Các biểu đồ như Histogram, Heatmap, và Radar Chart khi được kết hợp vào các dashboard đã giúp nâng cao khả năng phân tích và hỗ trợ ra quyết định, mang lại giá trị thực tiễn trong công tác huấn luyện và quản lý đội bóng. Hướng phát triển: Cải thiện giao diện người dùng (UI): Tạo ra các dashboard với giao diện người dùng thân thiện hơn, cho phép người dùng dễ dàng tương tác và khám phá các dữ liệu một cách trực quan, thậm chí có thể tùy chỉnh các biểu đồ theo yêu cầu.

Áp dụng phân tích dự báo: Kết hợp các mô hình học máy và phân tích dự báo vào các dashboard để giúp dự đoán sự thay đổi của các chỉ số cầu thủ, từ đó hỗ trợ các chiến lược phát triển đội bóng trong tương lai.

Triển khai hệ thống local server: Việc triển khai hệ thống server nội bộ (local server) có thể giúp lưu trữ và quản lý dữ liệu một cách hiệu quả hơn, giảm độ trễ khi truy xuất thông tin từ các dashboard.