

BÀI TẬP VỀ BIỂU ĐỒ

PHẦN 1: GIỚI THIỆU VỀ TẬP DỮ LIỆU

1.1 Tập dữ liệu:

Tên tập dữ liệu: **Top Football Leagues Scorers**.

Đây là tập dữ liệu tất cả các giải đấu bóng đá (2016-2019) và được crawl từ trang web hàng đầu là [Infogol](#) bằng Selenium.

Trong tập dữ liệu có bảng xếp hạng và thống kê từ một số giải đấu hàng đầu từ khắp nơi trên thế giới, bao gồm Ngoại hạng Anh, Giải vô địch Anh, La Liga Tây Ban Nha, Serie A của Ý, Bundesliga của Đức, Ligue 1 của Pháp, MLS Mỹ và Serie A của Brazil,...

Nguồn: [Top Football Leagues Scorers \(kaggle.com\)](#)

Dữ liệu nhóm phân tích:  Top Football Leagues Scorers(2016-2019)

1.2 Thành phần trong tập dữ liệu:

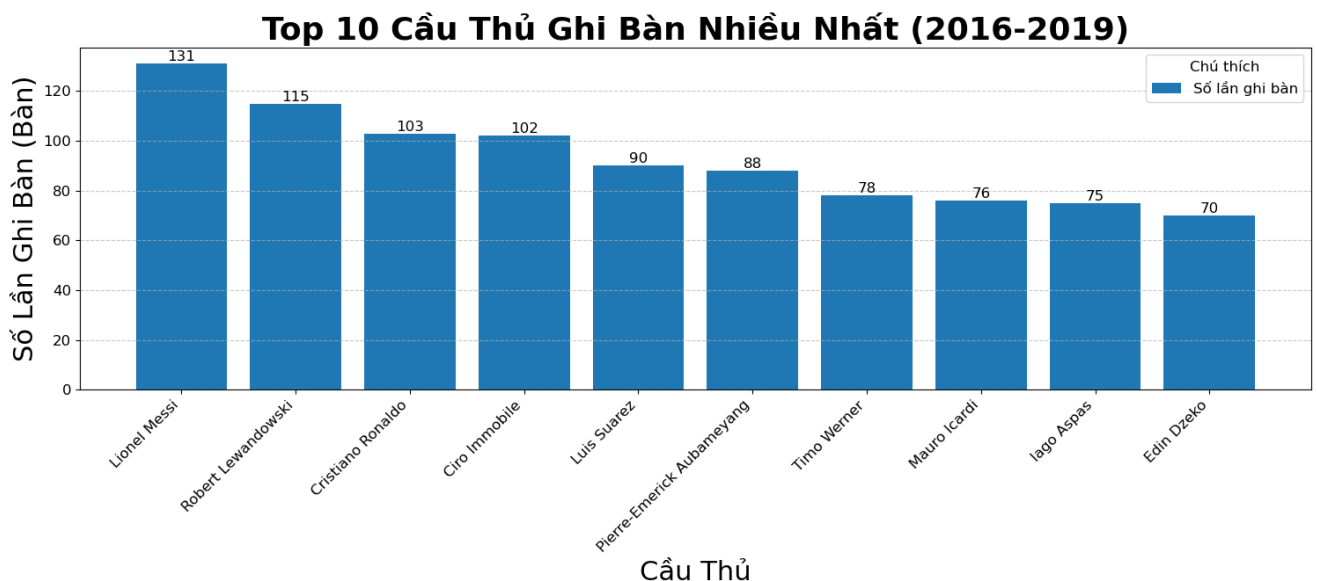
1. Country: Tên quốc gia tổ chức giải đấu.
2. League: Tên giải đấu.
3. Club: Tên câu lạc bộ.
4. Player names: Tên cầu thủ.
5. Matches Played: Số trận đã thi đấu.
6. Substitution: Tổng số lần thay cầu thủ.
7. Mins: Số phút thi đấu.
8. Goals: Số bàn thắng đã dành được.
9. xG: Số bàn thắng kỳ vọng ghi được trong mùa giải.
10. xG Per Avg Match: Số bàn thắng kỳ vọng trung bình mỗi trận đấu.
11. Shorts: Số lần sút bóng.
12. OnTarget: Số lần sút bóng trúng khung thành.

13. Shots Per Avg Match: Số lần sút bóng trung bình mỗi trận đấu.
14. On Target Per Avg Match: Số lần sút bóng trúng đích trung bình mỗi trận đấu.
15. Year: Năm thi đấu.

PHẦN 2: VẼ BIỂU ĐỒ

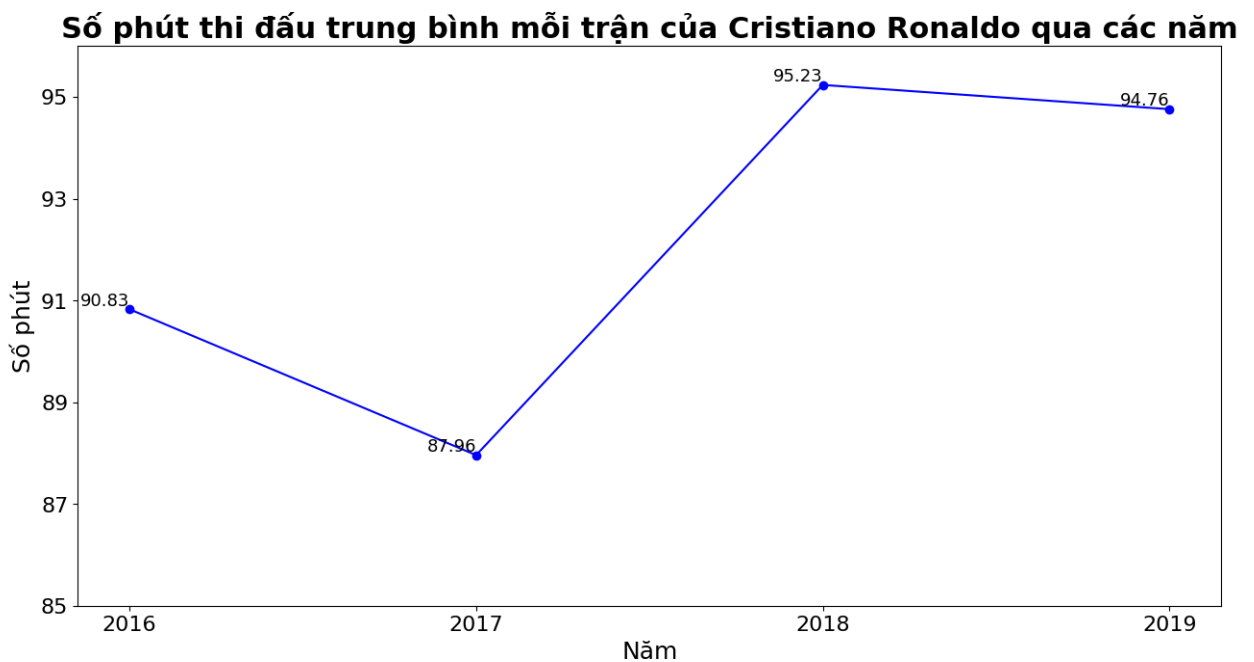
2.1 Biểu đồ cột (Bar Chart)

Mô tả: Biểu đồ cột (Bar Chart) thể hiện số lần ghi bàn nhiều nhất của top 10 cầu thủ trong khoảng thời gian từ 2016 đến 2019. Mỗi cột đại diện cho một cầu thủ, trục x hiển thị tên cầu thủ và trục y thể hiện tổng số lần ghi bàn của họ trong các giải đấu. Các thanh cột được sắp xếp theo thứ tự giảm dần, từ cầu thủ ghi bàn nhiều nhất đến cầu thủ thứ 10. Số bàn thắng cụ thể của mỗi cầu thủ được hiển thị trên đỉnh của từng thanh cột để cung cấp thông tin chi tiết, biểu đồ còn thêm các đường lưới và chú thích để giúp đơn giản hóa việc đọc và hiểu biểu đồ.



2.2 Biểu đồ đường (Line Chart)

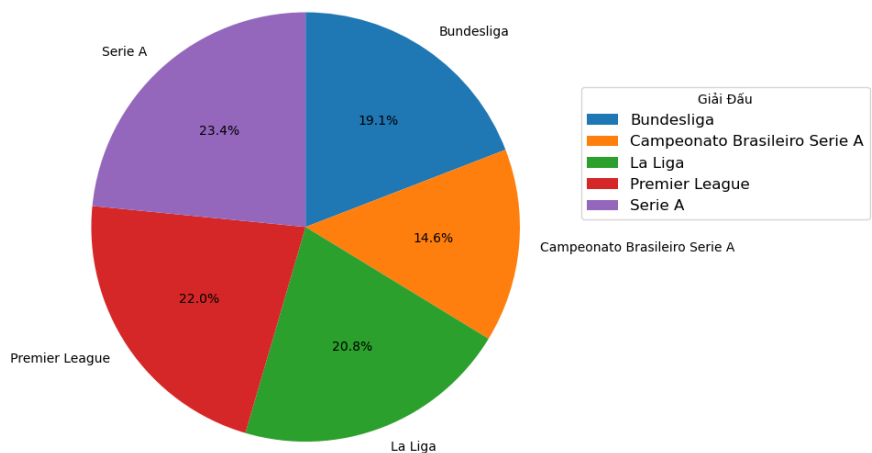
Mô tả: Biểu đồ thể hiện sự biến động của số phút thi đấu trung bình mỗi trận của Cristiano Ronaldo qua các năm trong khoảng thời gian từ 2016 đến 2019. Trục x của biểu đồ là năm, trong khi trục y là số phút trung bình mỗi trận. Mỗi điểm trên đường biểu diễn một năm cụ thể, và đường nối giữa các điểm tạo nên một hình thể linh hoạt giúp nhận diện xu hướng tổng thể của dữ liệu. Các hình tròn đánh dấu trên đường biểu diễn giá trị cụ thể của số phút trung bình mỗi trận tại từng năm. Thông tin chi tiết được hiển thị bên trên từng điểm, giúp người đọc có cái nhìn tổng quan và nhận biết sự biến động của thời lượng thi đấu của Cristiano Ronaldo qua các năm.



2.3 Biểu đồ tròn (Pie Chart)

Mô tả: Biểu đồ tròn (Pie Chart) thể hiện phân bổ ghi bàn theo giải đấu trong năm 2016. Mỗi phần trăm trên biểu đồ đại diện cho tỷ lệ ghi bàn của mỗi giải đấu so với tổng số ghi bàn trong năm đó. Các phần trên biểu đồ được chia thành các phần tương ứng với từng giải đấu, với mỗi nhánh đại diện cho một giải. Kích thước của mỗi phần phản ánh tỷ lệ ghi bàn của giải đấu đó trong tổng số ghi bàn của năm 2016. Các nhãn được đặt bên ngoài mỗi phần để chỉ rõ tên của từng giải đấu và phần trăm tương ứng.

Phân bổ ghi bàn theo giải đấu (2016)



2.4 Biểu đồ điểm (Scatter Plot)

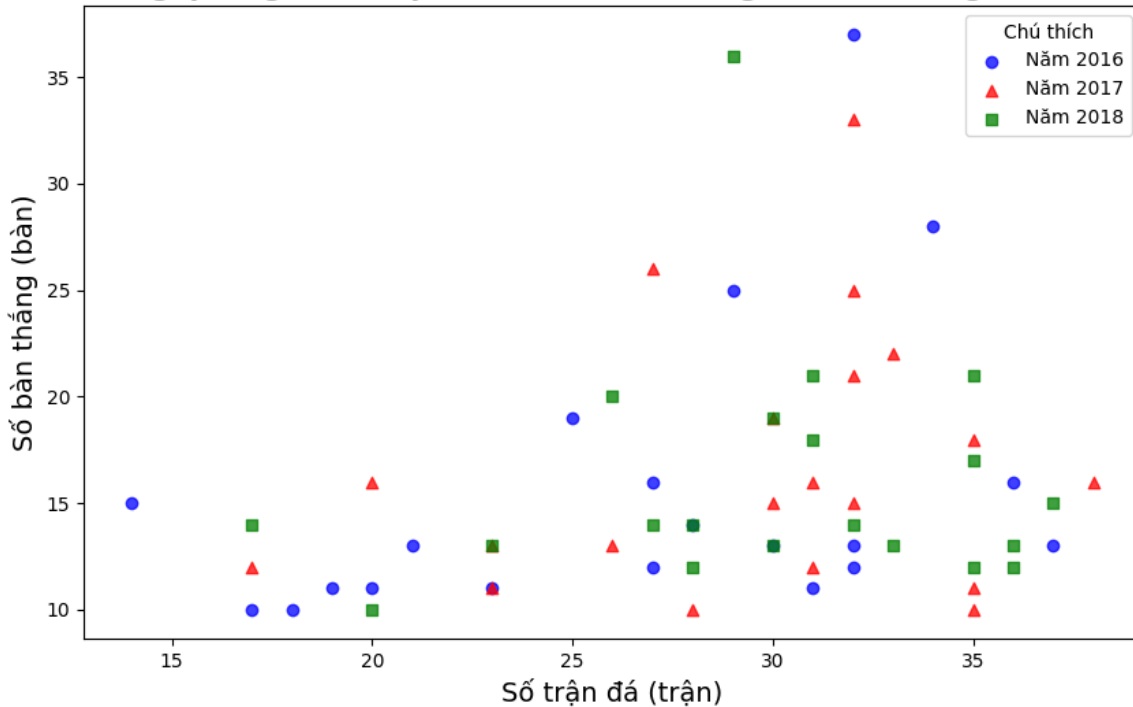
Mô tả: Biểu đồ điểm (Scatter Plot) thể hiện mối quan hệ giữa Số trận đấu và Số bàn thắng của đội bóng Tây Ban Nha (TBN) trong giải La Liga qua các năm 2016, 2017 và 2018. Dưới đây là mô tả chi tiết của biểu đồ:

1. **Trục X (hoành độ):** Biểu đồ sử dụng trục X để biểu diễn số trận đấu (Matches_Played). Đây là trục ngang của biểu đồ, thể hiện thông tin

về số lượng trận đấu mà đội bóng Tây Ban Nha đã tham gia trong mỗi năm.

2. **Trục Y (tung độ):** Trục Y thể hiện số bàn thắng (Goals). Đây là trục đứng của biểu đồ, biểu thị số bàn thắng mà đội bóng Tây Ban Nha đã ghi được trong mỗi năm.
3. **Điểm Dữ Liệu (Scatter Points):** Mỗi điểm trên biểu đồ tương ứng với một năm và đại diện cho một cặp giá trị (số trận đấu, số bàn thắng) của đội bóng TBN trong giải La Liga.
4. **Màu Sắc và Ký Hiệu Đặc Trưng:** Các điểm dữ liệu được phân biệt bằng màu sắc và ký hiệu đặc trưng. Năm 2016 được biểu diễn bằng điểm màu xanh (hình tròn màu xanh dương), năm 2017 được biểu diễn bằng điểm màu đỏ (tam giác màu đỏ), và năm 2018 được biểu diễn bằng điểm màu lục lam (hình vuông màu lục lam).
5. **Tiêu Đề:** "Mối tương quan giữa số trận đấu và số bàn thắng của TBN (LaLiga) (2016-2018)".
6. **Chú Thích (Legend):** Một chú thích được thêm vào để giải thích ý nghĩa của mỗi nhóm dữ liệu. Trong trường hợp này, nó chứa các nhãn "2016", "2017", và "2018" tương ứng với màu sắc và ký hiệu của từng năm.

Mối tương quan giữa số trận đấu và số bàn thắng của TBN (Laliga) (2016-2018)

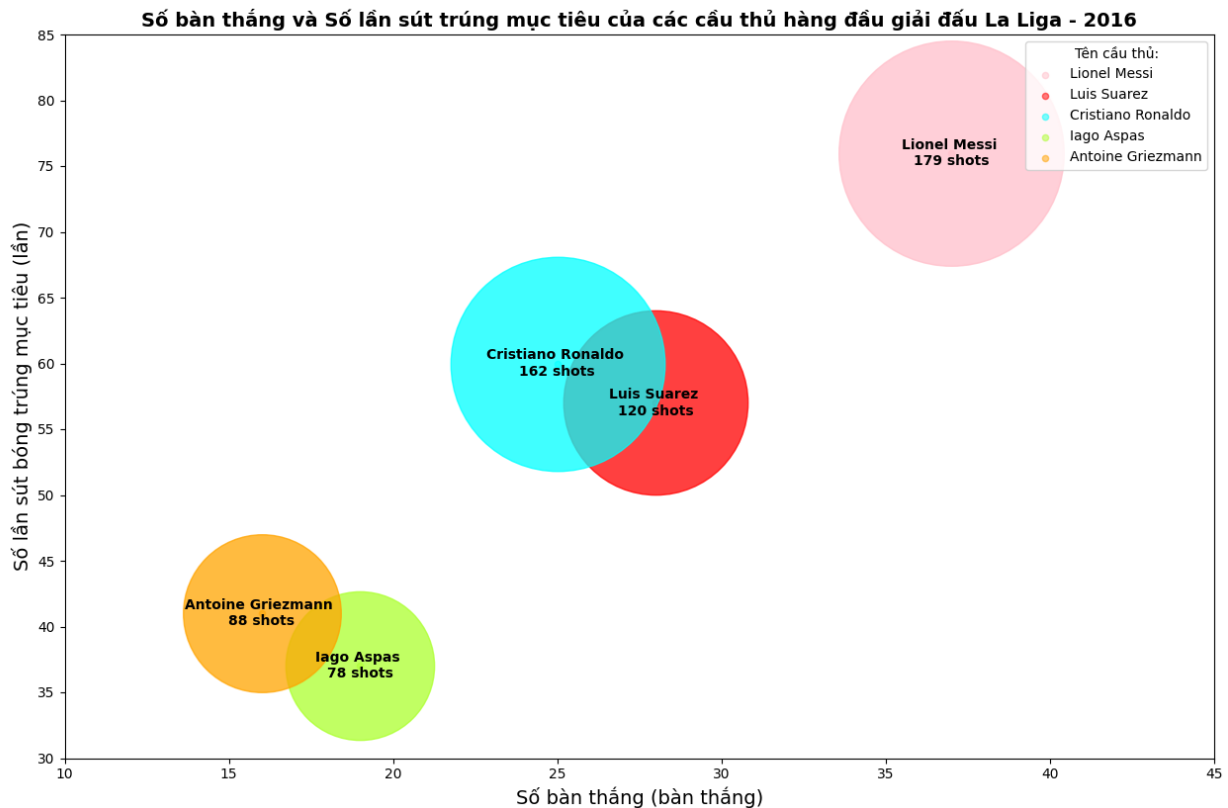


2.5 Biểu đồ bong bóng (Bubble Chart)

Mô tả: Biểu đồ bong bóng (Bubble Chart) trình bày thông tin về Số bàn thắng và Số lần sút trúng mục tiêu của năm cầu thủ hàng đầu trong giải đấu La Liga năm 2016. Dưới đây là mô tả chi tiết của biểu đồ:

1. **Trục X (hoành độ):** Biểu đồ sử dụng trục X để biểu diễn số bàn thắng của từng cầu thủ hàng đầu.
2. **Trục Y (tung độ):** Trục Y biểu diễn số lần sút bóng trúng mục tiêu của từng cầu thủ.
3. **Kích Thước Điểm (s):** Kích thước của điểm trên biểu đồ phản ánh số lần sút của từng cầu thủ. Điều này làm cho điểm lớn hơn nếu cầu thủ có số lần sút nhiều hơn.
4. **Màu Sắc:** Mỗi điểm trên biểu đồ có một màu sắc khác nhau để phân biệt giữa các cầu thủ.
5. **Chú Thích (Annotations):** Mỗi điểm có một chú thích hiển thị tên cầu thủ, số lần sút của họ và một biểu tượng dạng hình tròn.

6. **Chú Thích Legend:** Biểu đồ có một chú thích ghi tên của các cầu thủ tương ứng với màu sắc của điểm tương ứng.
7. **Tiêu Đề:** "Số bàn thắng và Số lần sút trúng mục tiêu của các cầu thủ hàng đầu giải đấu La Liga - 2016".
8. **Trục Ticks (Chia Cột):** Các giá trị trên trục X và Y được chia thành các cột để giúp người đọc dễ dàng đọc giá trị của các điểm.



2.6 Biểu đồ phân phối tần suất (Histogram)

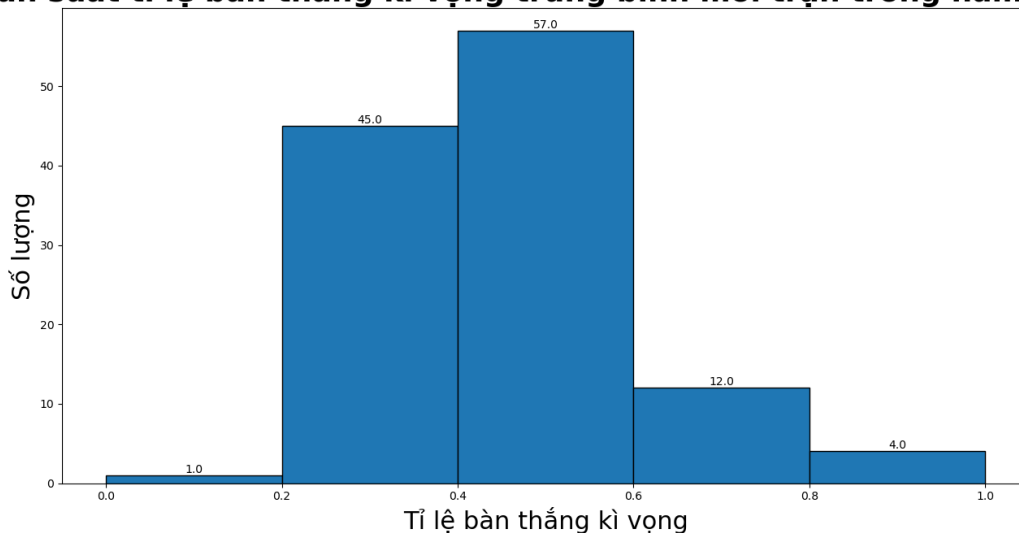
Mô tả:

Biểu đồ tần suất (histogram) mô tả phân phối của dữ liệu tỉ lệ bàn thắng trung bình mỗi trận trong năm 2018. Dữ liệu này được chia thành các khoảng (bins) từ 0 đến 1 với bước nhảy là 0.2, và trục x biểu thị tỷ lệ bàn thắng kỳ vọng (expected goals per average match), trong khi trục y thể hiện số lượng quan sát trong từng khoảng.

Mỗi cột trên biểu đồ đại diện cho một khoảng giá trị, và chiều cao của cột thể hiện số lượng các quan sát trong khoảng đó. Chú thích số lượng được thêm vào mỗi cột để cung cấp thông tin chi tiết hơn về phân phối của dữ liệu.

Biểu đồ giúp hiểu rõ hơn về phổ biến của các giá trị xG Per Avg Match trong năm 2018 và có thể giúp bạn nhận biết các xu hướng hoặc đặc điểm đặc biệt trong dữ liệu này.

Tần suất tỉ lệ bàn thắng kỳ vọng trung bình mỗi trận trong năm 2018

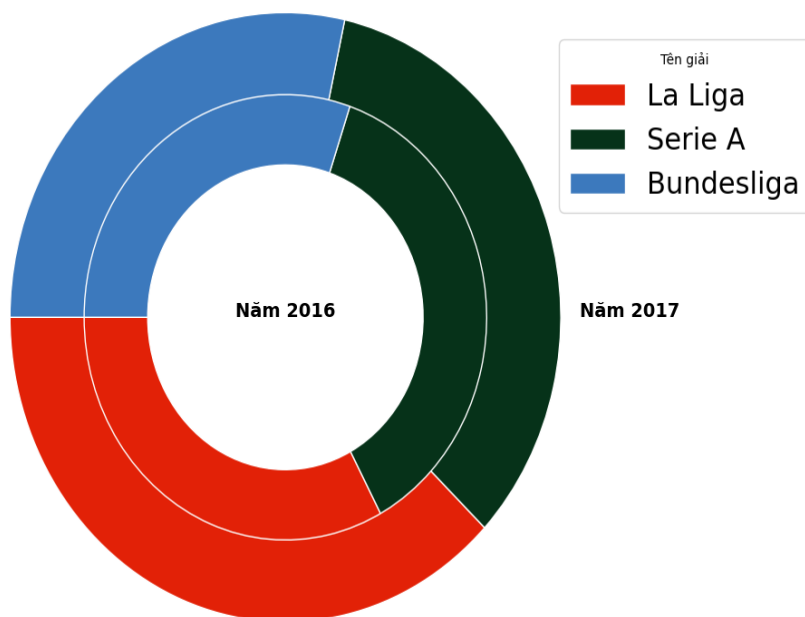


2.7 Biểu đồ doughnut (Doughnut Chart)

Mô tả:

Biểu đồ Doughnut mô tả phân phối của tổng số bàn thắng (Goals) trong một năm (2017 và 2016) giữa các giải đấu La Liga, Serie A, và Bundesliga. Mỗi phần trong biểu đồ đại diện cho một giải đấu, và kích thước của phần đó thể hiện tỷ lệ phần trăm của tổng số bàn thắng mà giải đấu đó đóng góp trong tổng số bàn thắng của ba giải đấu được so sánh với nhau.

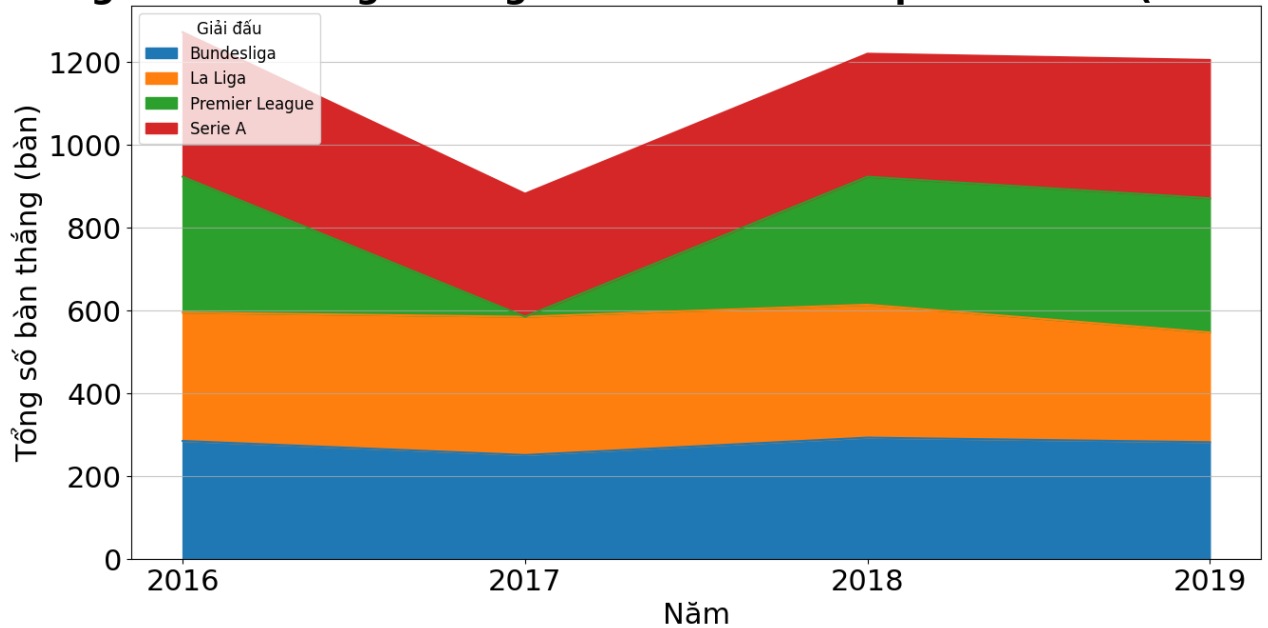
Đóng góp trong tổng số bàn thắng của ba giải đấu được so sánh với nhau trong năm 2016 & 2017



2.8 Biểu đồ vùng (Area Chart)

Mô tả: Biểu đồ vùng (Area Chart) thể hiện tổng số bàn thắng của 4 giải đấu (Bundesliga, La Liga, Premier League, Serie A) qua các năm từ 2016 đến 2019 trong tập dữ liệu Top Football Leagues Scorers. Mỗi màu sắc khác nhau trên biểu đồ đại diện cho một giải đấu và chiều cao của mỗi vùng thể hiện tổng số bàn thắng ghi được của giải đấu đó trong năm tương ứng. Biểu đồ này giúp nhìn nhận sự đóng góp bàn thắng của giải đấu đó qua các năm.

Tổng số bàn thắng ở các giải đấu khác nhau qua các năm (2016-2019)



2.9 Biểu đồ mật độ (Density Plot)

Mô tả: Biểu đồ mật độ (Density Plot) thể hiện số bàn thắng ghi được của 5 giải thi đấu (La Liga, Serie A, Bundesliga, Premier League, Campeonato Brasileiro Serie A) qua các năm từ 2016 đến 2019 trong tập dữ liệu Top Football Leagues Scorers. Trục x của biểu đồ biểu diễn các năm, trong khi trục y thể hiện tổng số bàn thắng ghi được của 5 giải đấu qua các năm. Mỗi điểm trên đồ thị mật độ đại diện cho một năm và các vùng mật độ có màu sắc lên xuống thể hiện mức độ tăng giảm của số lượng bàn thắng ghi được. biểu đồ mật độ giúp quan sát biến động của sự phân bố bàn thắng ghi được của 5 giải đấu qua các năm một cách trực quan.

