
FootyViz: Top 5 Leagues

CSC_51052_EP - Data Visualization - 21.12.2024

Enzo Pinchon
Institut Polytechnique de Paris
enzo.pinchon@ip-paris.fr

Mathieu Antonopoulos
Institut Polytechnique de Paris
antonopoulos@ip-paris.fr

1 Introduction

"Football is not just a game; it's an art form. You have to be creative, think outside the box, and always be one step ahead of your opponent." – Ronaldinho. These words from one of the sport's greatest icons capture the essence of football's universal appeal and its underestimated complexity. Football, often referred to as "the beautiful game," is the most popular sport in the world, captivating billions of fans across the globe. As the most followed sport globally, football generates an immense amount of data each year. Every minute of each game is analyzed, and every step of every player is tracked. This vast amount of data is crucial not only for professional teams but also for fans, analysts, and the growing betting industry. Given the wide accessibility and richness of football data, we chose to explore this in our project, focusing on visualizing key statistics and performance trends from the top five European leagues over three seasons: 2021-2022, 2022-2023, and 2023-2024. For context, football games are organized into seasons, typically lasting from August to May. During each season, teams within the same league face each other twice (home and away matches). Leagues usually consist of around 20 teams, and each league is specific to a country. Each team is composed of 20 to 25 players. While many visualization tools already exist for football, we aimed to differentiate our project by offering a user-friendly interface, emphasizing design and animation. Most existing websites follow a standard format, with few offering advanced graphic visualizations and more raw data. Our project is structured around different pages, each providing a unique perspective on the data: Players, Leagues, Teams, Matches, and more. This document outlines the methods and techniques used to build the platform, from data collection and preprocessing to visualization design and implementation.

2 Data Acquisition and Preparation

2.1 Data Sources

For this project, we leveraged multiple data sources to ensure comprehensive coverage of football statistics. Our primary source was FootyStats, a platform known for its extensive football data offerings, including downloadable CSV datasets containing information on teams, players, matches, and leagues.

To achieve a holistic dataset, we aggregated multiple CSV files covering players, matches, and clubs across the top five European leagues for three consecutive seasons: 2021-2022, 2022-2023, and 2023-2024. Each season contributed three datasets per league (players, matches, and clubs), resulting in a total of 45 datasets.

In addition, we enriched our datasets using web scraping techniques. A Python script utilizing WikiData API and SerpApi was used to collect football club logos and country flags. For players' photos, we accessed a public JSON file provided by SoccerWiki. GeoJSON files containing geographical data were also used to facilitate the creation of interactive map visualizations. We used the publicly available GeoJSON files from Eurostat.

2.2 Data Characteristics

The different types of datasets used in this project are detailed below. Given the extensive number of features, only the most relevant ones are highlighted as examples:

- **Teams Dataset:** CSV file containing 293 team-relative features, including:
 - teamName, commonName, season, country
 - Performance metrics: matchesPlayed, wins, draws, losses
 - Home/Away metrics: matchesPlayedHome, winsAway,
- **Players Dataset:** CSV file containing 277 player-relative features, including:
 - Personal details: fullName, age, birthday, position, salaryUSD
 - Club details: currentClub, shirtNumber

- **Matches Dataset:** CSV file containing 66 features for each match, including:
 - Match details: homeTeam, awayTeam, scores, dates, stadium, referee
- **Logos / Flags Dataset:** CSV files containing URLs to football club logos and country flags.
- **Players Pictures Dataset:** JSON file including photo URLs and player names.

2.3 Data Cleaning and Preprocessing

To ensure data consistency and usability, several preprocessing steps were applied. Missing values in the datasets were handled by either replacing them with null values or setting them to zero, depending on the context. Additionally, player names were standardized to resolve variations across datasets, such as differences caused by nicknames or abbreviations. These steps were crucial for maintaining data integrity and ensuring integration across multiple sources.

3 Visualization

3.1 Map View

The Map View 1 serves as the home page, offering an intuitive overview of the geographical distribution of football teams across cities in the top five leagues. Each city hosting at least one team is represented by a distinct green circle. Upon hovering over a city, a menu appears on the left side of the screen, showcasing the team logos and names for all teams located in the given city. Each logo is clickable, redirecting the user to the Team View page[3.2]. This feature allows users to geographically locate teams while grouping them by their respective leagues (each country corresponds to a league).

A toggle button further enhances the Map View by switching from the default circle markers to team logos, displayed as pins on the map.

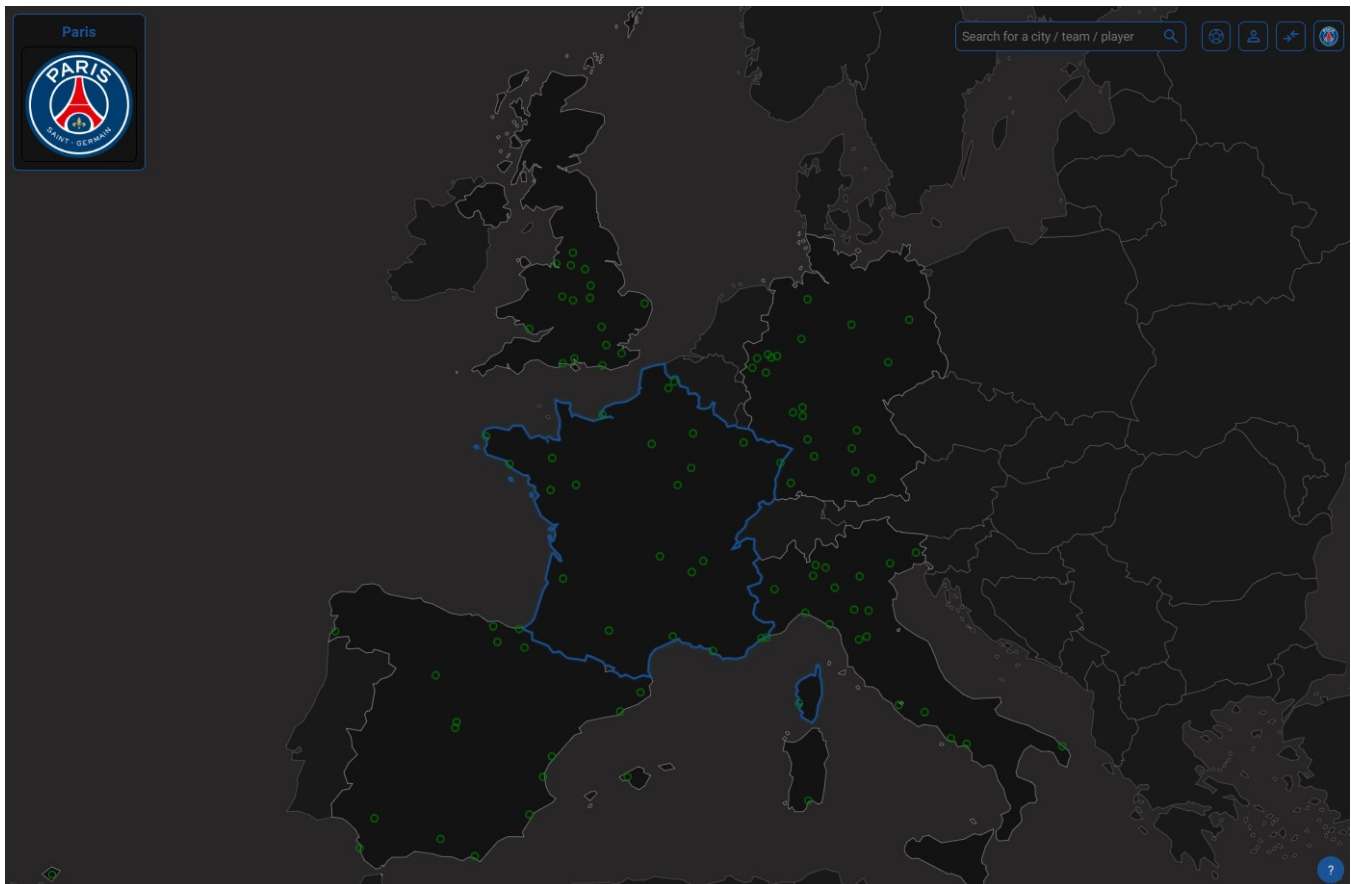


Figure 1: Capture of the Map View page when clicking on Paris

3.1.1 Design Choices and Implementation

The Map View is implemented using D3.js, with GeoJSON files providing geographical data for country and city boundaries.

When team logos are displayed, cities with multiple teams are indicated by green dots to visually distinguish them and avoid overlapping logos. We used a dark theme, which highlights the markers and creates contrast with the background. When hovering over a country, its borders glow with animation aligning with the project's color scheme of black and light blue.

An integrated search bar is included to assist users in quickly locating specific cities by typing team names, city names or even player names. As users type, the search bar is responsive, providing feedback with animation. Upon pressing enter, an animation is triggered, guiding the user to the relevant city on the map. If a correct city is typed, the search box turns green, confirming the match. If the city is not found, the search box turns red, indicating an invalid search term. This feature offers an intuitive and dynamic way to navigate the map.

The menu at the top right corner includes buttons to navigate directly to all the main pages of the project as well as the search bar and the toggle button.

3.2 Team View

The Team View provides a clear and interactive summary of a selected team, combining detailed statistics and visualizations. It aims to provide both a quick overview and detailed insights about performances.



Figure 2: Capture of the Team View page

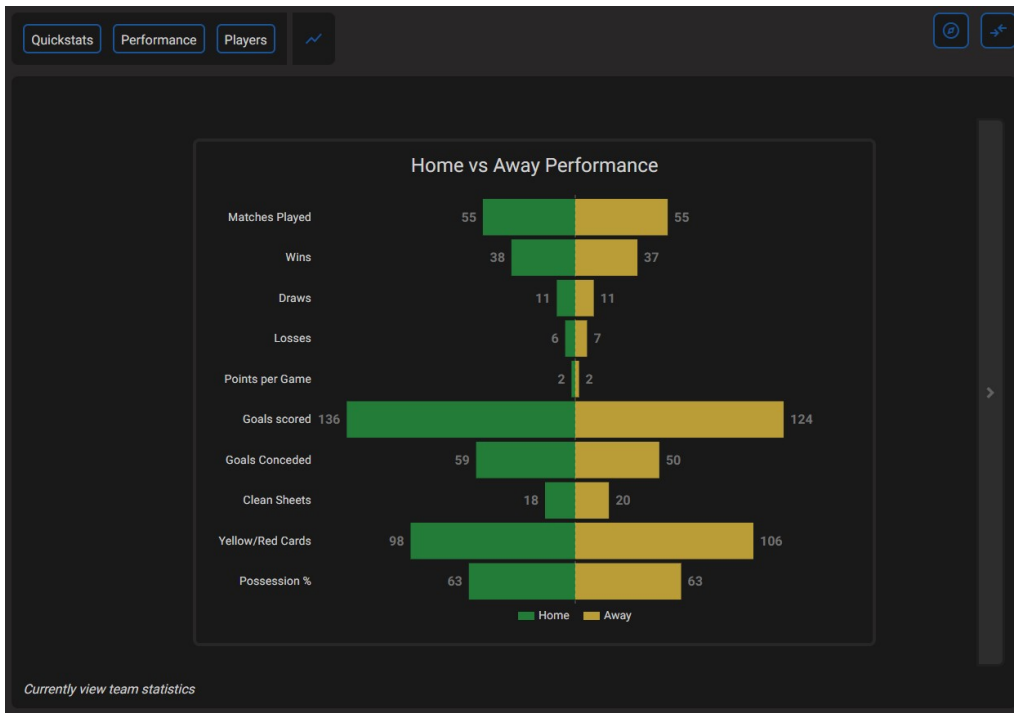


Figure 3: Capture of the bidirectional chart comparing home and away performance statistics for PSG

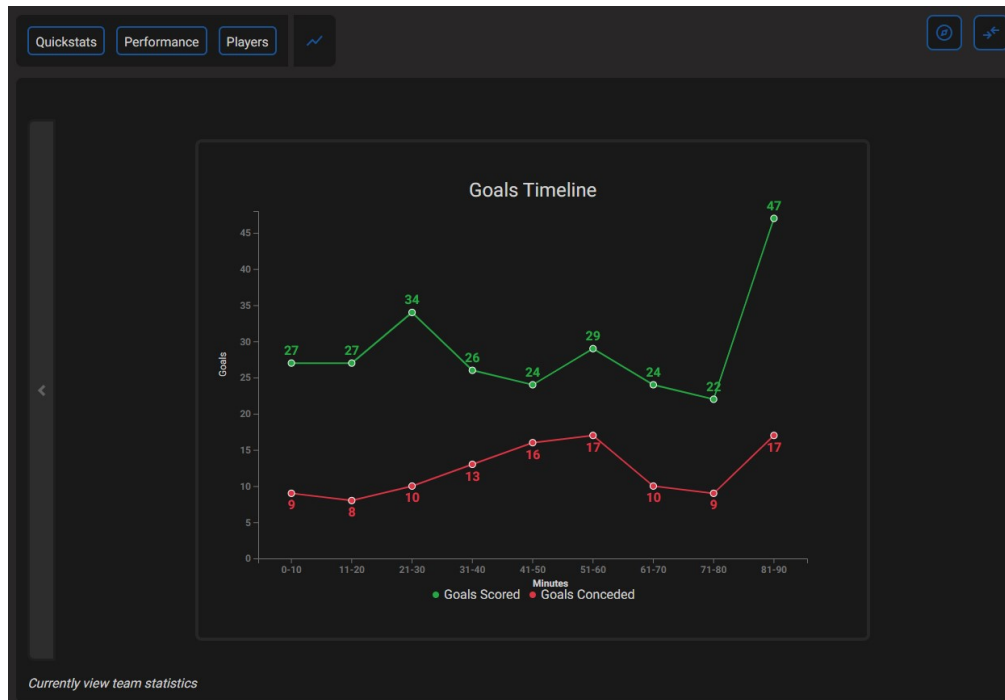


Figure 4: Capture of the goals timeline chart for PSG

3.2.1 Design Choices and Implementation

Left Panel: The left panel displays the following key information about the team 2:

- **Team Logo** and basic details about the team.
- **Team Name**, which also serves as a search bar to select other teams.
- **League:** Team's current league.
- **League Rank:** The team's position in the league standings.
- **Ratio:** Shows the number of wins, losses, and draws.

Then, the Match Records section shows the team's league matches for the current season in reverse chronological order. Each match includes the date, the opponent and the final score. Then a final column displays the result, whether the team won, lost, or drew the game. Each state being written in a specific color (e.g., Green for Win, Red for Loss) to visually emphasize the team's form. Each row in the Match Records table is clickable, redirecting users to a detailed Match View page [3.5] for more information about any specific game.

Right Panel: The right panel focuses on visualizing the team's performance statistics. Upon loading, the Quick Stats section is displayed, featuring three Donut Charts for a quick overview 2:

- **Shots:** Displays the total number of shots, divided into on-target and off-target shots.
- **Goals:** Shows the number of goals scored and conceded by the team.
- **Results:** Shows the number of wins, losses, and draws.

The Donut Charts provide a clear and intuitive visualization of the team's performance distribution, such as the proportion of on-target vs. off-target shots or the team's win vs. loss ratio.

Next, clicking the Performance button in the menu will load a Bi-Directional Bar Chart 3, comparing the team's home and away performance across various metrics (e.g. wins, draws, goals). It allows a clear comparison of the team's performance playing at home or away. Estimating the home advantage of each team.

Finally, by clicking the right arrow on the Performance menu, users can view the Goals Timeline Chart 4. This chart tracks goals scored and conceded throughout a match, divided into 10-minute intervals. This visualization helps to analyze the flow of goals throughout a match and deduct team's dynamics (e.g. whether a given team usually scores or concedes goals at the beginning of a match).

For the technical implementation, all charts were created using D3.js to leverage its flexibility and customization capabilities. We opted to load all the data upfront when the page loads. While this approach results in a slight initial loading time, it ensures that users can seamlessly switch between teams, seasons, or charts without experiencing any delays afterward. To enhance the user experience during the initial load, we added a loading screen.

Additional Features:

- **Players Button:** Redirects to the Player View [3.4] page, where the players list of the selected team is displayed.
- **League Stats Toggle:** Allows users to switch between the team's performance stats and the league's average stats, enabling to put in perspective team stats compared to the average team performances within the league.
- **Season Selector:** User can choose from three available seasons. Changing the season triggers an event listener that updates all visualizations (charts, match records, rankings, etc.) to reflect the selected season. By default, the selector is set to "All Seasons," which aggregates data from the three available seasons. An event listeners updates all the stats every time a new season is selected.
- **Team Comparison:** Redirects to the Team Comparison View [3.3], where the current team is filled.
- **Back to Map:** Redirects to the Map View [3.1] home page.

3.3 Team Comparison View

The Team Comparison View 5 provides an interactive interface for users to compare two teams based on their performance over a selected season or across all seasons. It also displays confrontations between the two teams if playing in the same league.

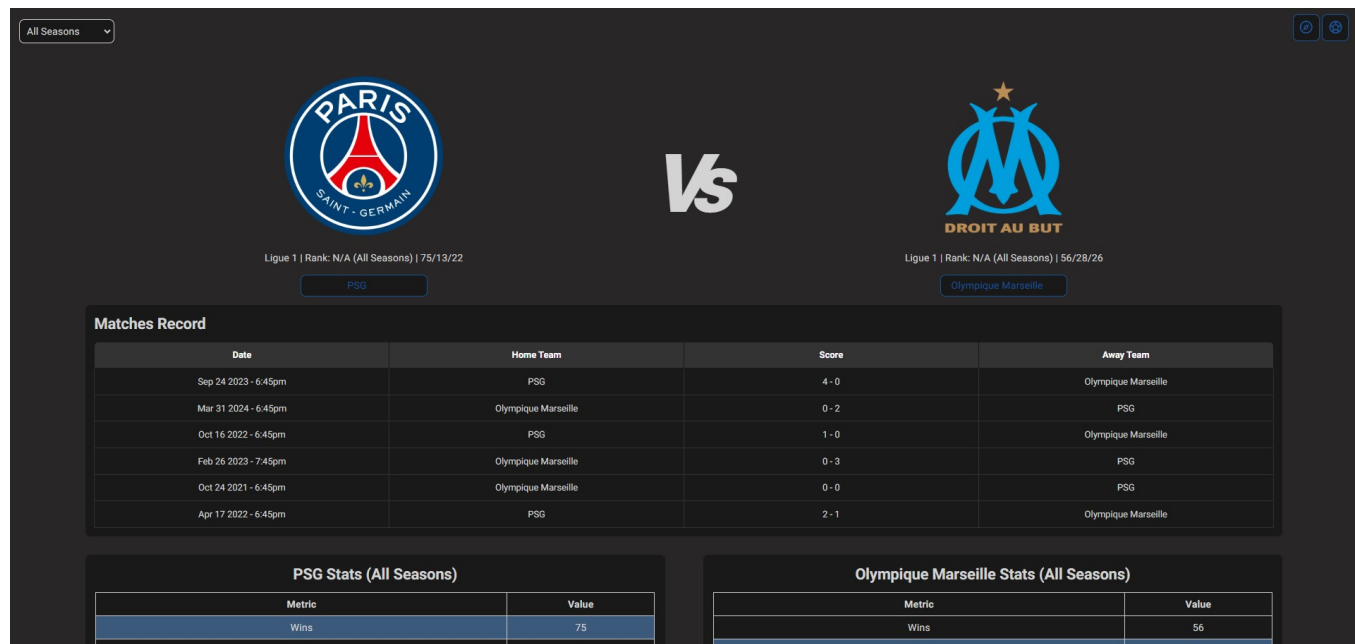


Figure 5: Capture of the Team Comparison page when comparing PSG and OM

3.3.1 Design Choices and Implementation

Team Selection: User can select two teams to compare. Each team section includes an input field where users can search for and select a team from a dynamic list of available teams. Then the logo and basic information of the selected teams are displayed (e.g., team name, league). Logos are clickable and redirect to the Team View [3.2] page.

Matches Record Once the teams are selected, a table shows the Match Record between the two teams. This table contains information about each match (e.g., Date, Score). If the teams are playing in different leagues, the matches record table is hidden.

Team Stats: This section presents a side-by-side comparison of the two teams' statistics in table format, evaluating various metrics such as Wins, Goals Scored, Cards, and more. For each metric, the team with the highest value is highlighted in blue, making it easy to identify which team leads in each category.

3.4 Player View

The player view was created to visualize players datasets (from 2021 to 2024), covering all the 5 leagues and containing 227 columns of data for each player. The challenge was to process and present this data in a way that is both clear and interactive.

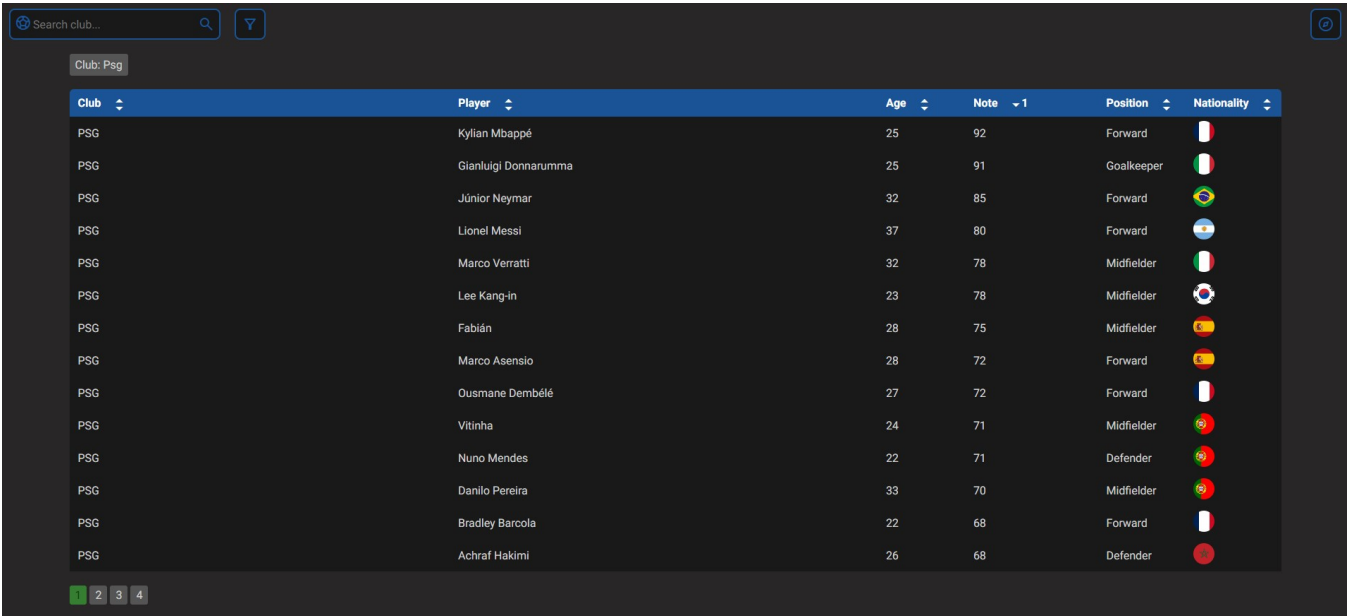


Figure 6: Capture of the Player View list.

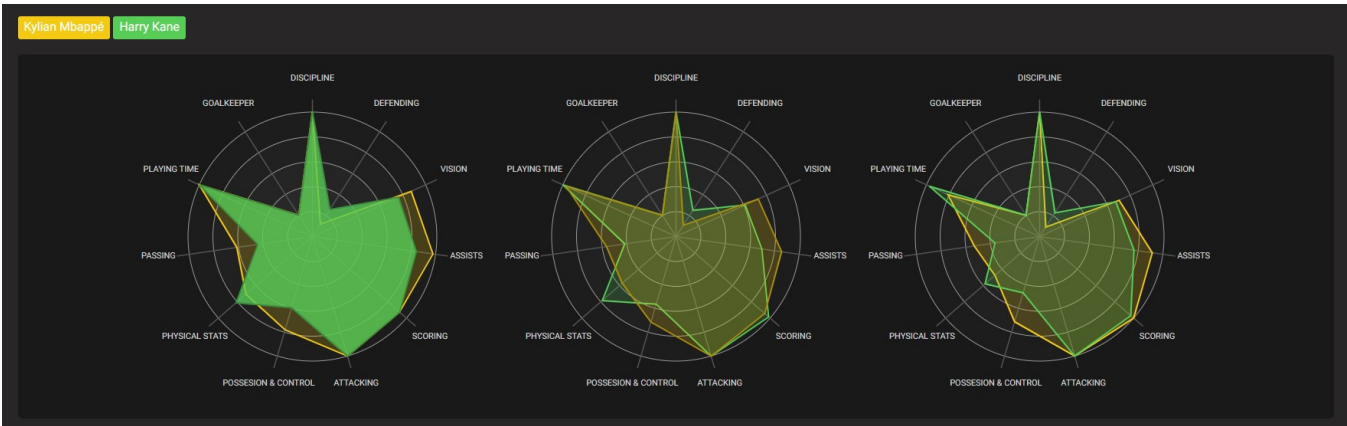


Figure 7: Capture of the Player View radar charts.

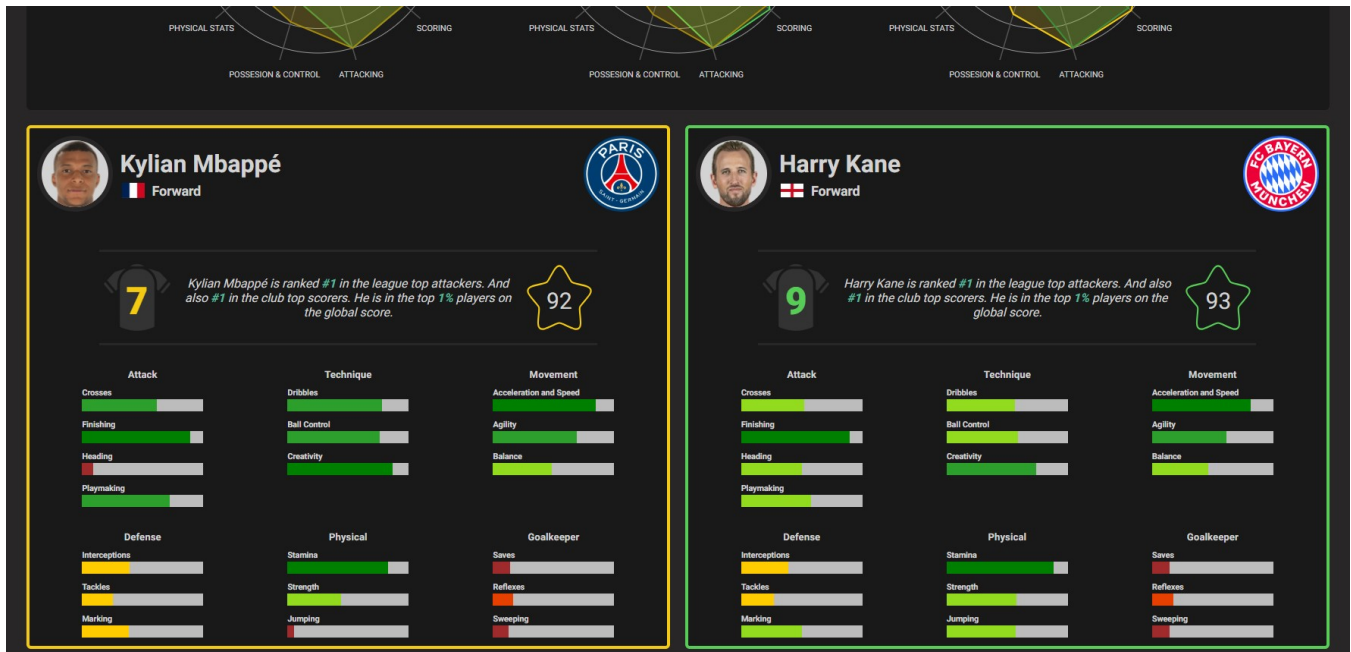


Figure 8: Capture of the Player View, player cards.

3.4.1 Design Choices and Implementation

To handle the dataset's complexity, all columns were aggregated manually to ensure precision and relevance. Some attributes were dynamically reversed (e.g., fouls committed) or treated as bonuses (e.g., number of penalties saved by the goalkeeper). This allowed the creation of two complementary data structures for analysis.

The first structure focused on season-specific performance, visualized through radar charts 7. These charts displayed eleven key indicators, each calculated as a weighted combination of normalized attributes, such as average goals per match or aerial duel success percentiles. This representation provided an interactive way to explore player performance across the three seasons.

The second structure averaged player statistics across all seasons 8, extracting 19 indicators grouped into six categories. This structure ensures more stable metrics and comprehensive evaluations. It was used to calculate a global score for each player, consisting of two components:

1. **Primary Score:** This score is a weighted sum of key indicators (shown on the player cards). The focus indicators depend on the player's role; for example, forwards are evaluated heavily on finishing, while goalkeepers are rated on saves. This score forms the majority of the global ranking.
2. **Secondary Score:** This score incorporates additional statistics from the player card and is adjusted based on the player's average playing time over the three seasons. Players with less play time are penalized to prevent inflated scores due to exceptional performance in a few matches. For example, players who only participated to one or two games within a season might have unrealistically high scores if unadjusted.

The secondary score is then combined with the primary score to produce a final balanced ranking. This approach ensures fairness and accounts for variability in the dataset, creating a harmonious evaluation for all players.

The data visualization was designed to highlight these insights effectively. Radar charts dynamically illustrate aggregated statistics for each season, providing a visual summary of strengths and weaknesses. A table view allows users to explore detailed statistics with filters for position, nationality, and club. Player cards further summarized performance, showing both key and secondary indicators alongside comparisons with the broader player distribution 8.

This combination of manual aggregation, score design, and clear visualization provides a robust and user-friendly way to explore complex player data while maintaining analytical depth and fairness.

3.5 Match View

The Match View section 9 is designed to provide detailed information and statistics about a specific match. The implementation comprises two main parts: a list of matches displayed in a table in reverse chronological order (from the most recent to the oldest), and detailed match statistics presented dynamically upon selecting a game from the table.



Figure 9: Enter Caption

3.5.1 Design Choices and Implementation

The detailed view is designed to present key match information intuitively 9. It prominently features team logos, names, and match score, as well as contextual details such as half-time scores, referee information, stadium details or game week. Key statistics, including total shots, shots on target, expected goals (XG), and disciplinary actions (yellow/red cards), are organized in a grid layout for easy comparison between the teams. For the possession percentage, a visual bar chart illustrates the distribution.

Data are preloaded upon page initialization, with a loading screen displayed to inform the user. Consistent with the Team View [3.2] design, the decision to preload all data ensures seamless transitions between matches and statistical views, maintaining a fast and responsive interface for the user.

4 League View

The League View gives a global view of a given league over a given season. Displaying information about the best clubs, best players as well as the season table.

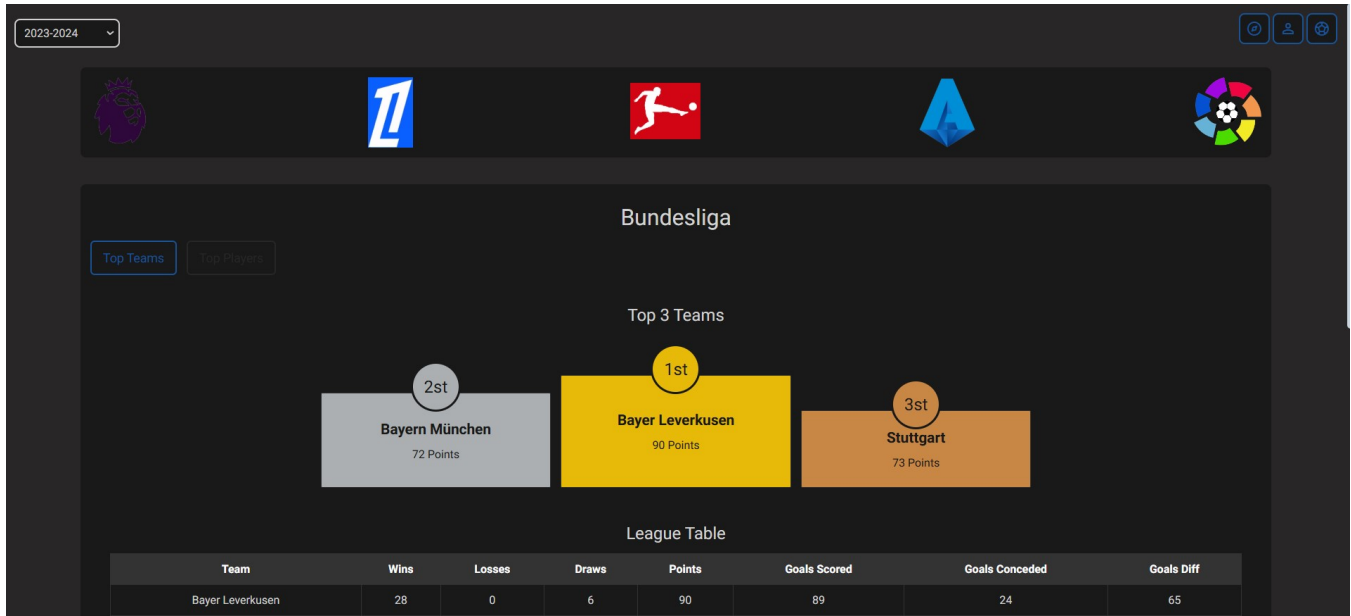


Figure 10: Capture of the League View in the teams tab.

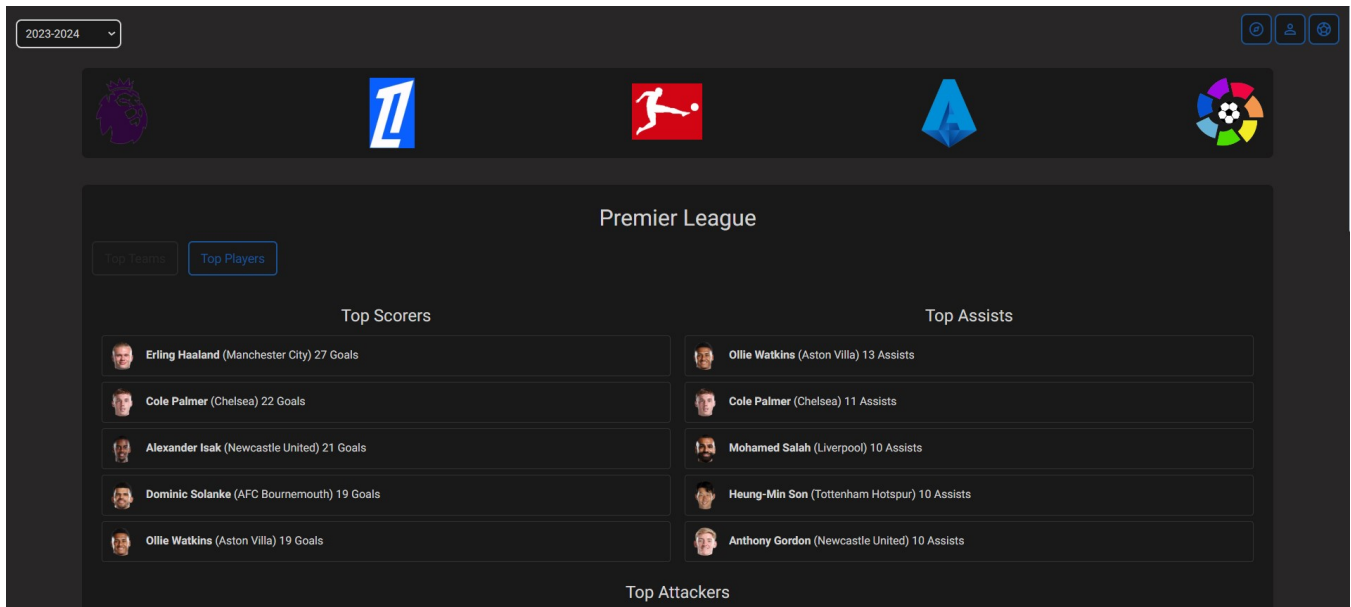


Figure 11: Capture of the League View in the players tab.

4.1 Design Choices and Implementation

A league selector is displayed 11 with logos for the top five leagues. Users can switch between leagues by clicking on their logos. The page also includes tab navigation with two main options: "Top Teams" and "Top Players." The "Top Teams" section highlights the top 3 teams in the league and provides a detailed league table with statistics such as wins, losses, points, and goal difference. The "Top Players" section presents a list of the highest-performing players in the league for each position (e.g., Midfielder, Defender) and top scorers.

5 Tutorial

Given the relative complexity of the project and the numerous pages, we included a tutorial to make sure that even users unfamiliar with football can easily explore the different sections. Once launched, by clicking on the question mark button located at the bottom right corner of the home page, the tutorial guides the user through each section of the project explaining every functionalities and features. This user-friendly guide aims to enhance the overall experience and help users fully understand how to interact with and navigate through the platform.

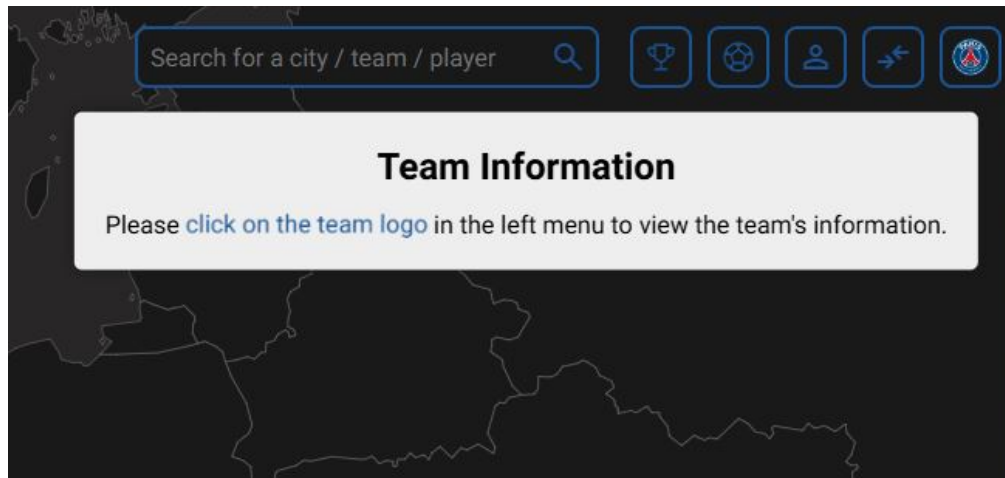


Figure 12: Capture of the tutorial on the Map View.

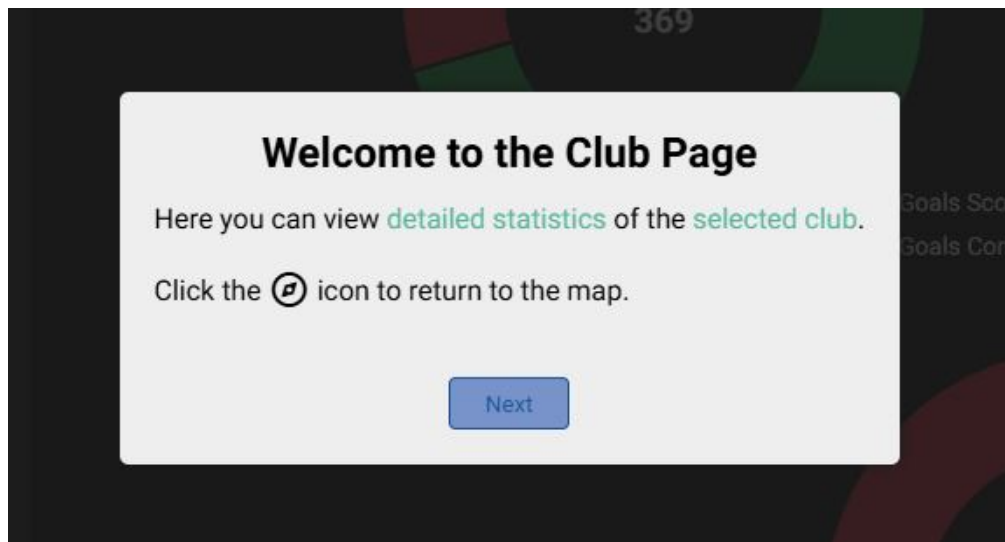


Figure 13: Capture of the tutorial on the Team View page.

5.1 Design Choices and Implementation

- **Minimizing Tutorial Interruptions:** Safeguards were implemented to avoid unintended navigation away from the tutorial. During explanatory sessions, user navigation is temporarily restricted to ensure focus. If the user attempts an action that might lead to leaving the current page or ending the tutorial prematurely, a warning popup is triggered. This mechanism reinforces continuity and minimizes frustration while maintaining a smooth tutorial experience.
- **Contextual and Interactive Design:** The tutorial operates directly within the website interface, offering contextual assistance tailored to each page. Through strategically placed pop-ups and subtle animations, the tutorial highlights relevant sections and provides clear instructions on their functionality. After each explanation, users are granted the freedom to explore the page at their own pace before proceeding, balancing guidance with autonomy.
- **Concise and Accessible Communication:** The tutorial adopts a minimalist approach to text, delivering information through concise and clear language. This ensures that users are not overwhelmed with excessive details while still gaining an adequate understanding of the platform's features. The tone is designed to emulate a personal guide, fostering an accessible and user-friendly learning environment.
- **Seamless Integration and Accessibility:** The tutorial is initiated through a clearly visible question mark button located at the bottom right of the home page. Its integration within the platform ensures that it complements, rather than disrupts, the user experience. The design carefully balances its informative role with maintaining the overall aesthetic and usability of the website.

This combination of elements ensures the tutorial guides users effectively, helping both football enthusiasts and newcomers navigate and engage with the platform confidently.

6 Conclusion and Future Work

This project provided a comprehensive platform to visualize and analyze football data from the top five European leagues, leveraging data processing, visualization techniques, and an intuitive user interface. We aimed to create an engaging and user-friendly tool for football enthusiasts.

Despite our satisfaction with the current achievements, the project faced some limitations. Certain columns in the datasets contained incomplete or missing values, such as statistics on successful tackles or long passes which would have been useful to improve player rating. Addressing these gaps would require further data preprocessing or enrichment, which is a potential area for improvement. Additionally, the datasets were vast and complex, necessitating a significant time investment for understanding and performing meaningful aggregations. For example, the player-level analysis alone required approximately 10 hours of effort to achieve the final results presented in this work.

Another area for future development might be the integration of machine learning techniques. For instance, predictive models could analyze player performance or team outcomes, offering match simulations. While these features were outside the scope of the current project, we considered adding this functionality. However, the limitations regarding data quality and availability made it challenging within the given deadline.