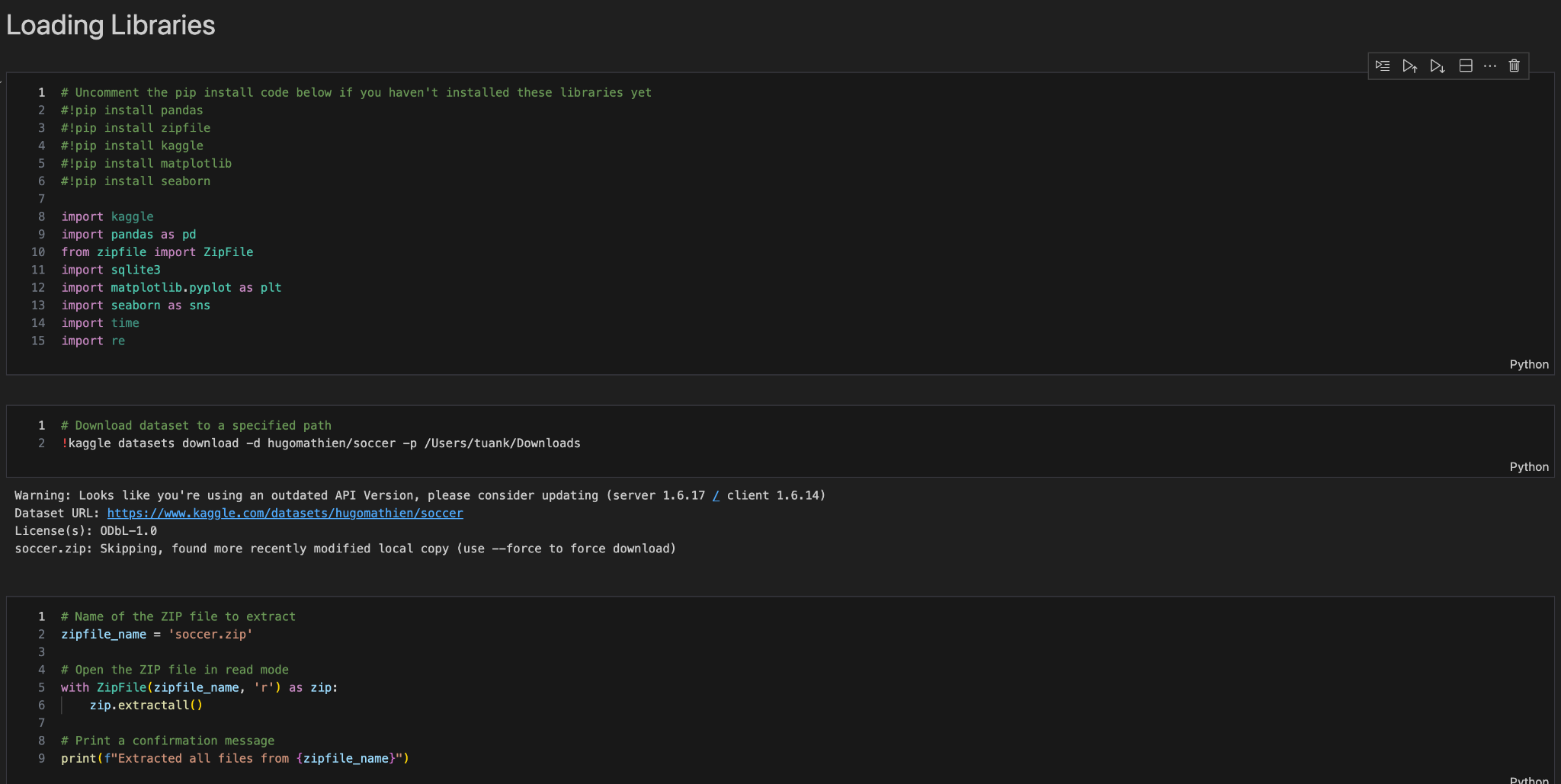
**Introduction**

The performance of soccer players and teams is a multifaceted subject, influenced by various factors ranging from tactical approaches to individual player skills. This report aims to delve into the core metrics of player and team performance across major European soccer leagues over a span of five seasons. The research question guiding this analysis is: "How do player and team performance core metrics vary across the major European soccer leagues between 5 seasons, and what factors contribute to these performance differences?"

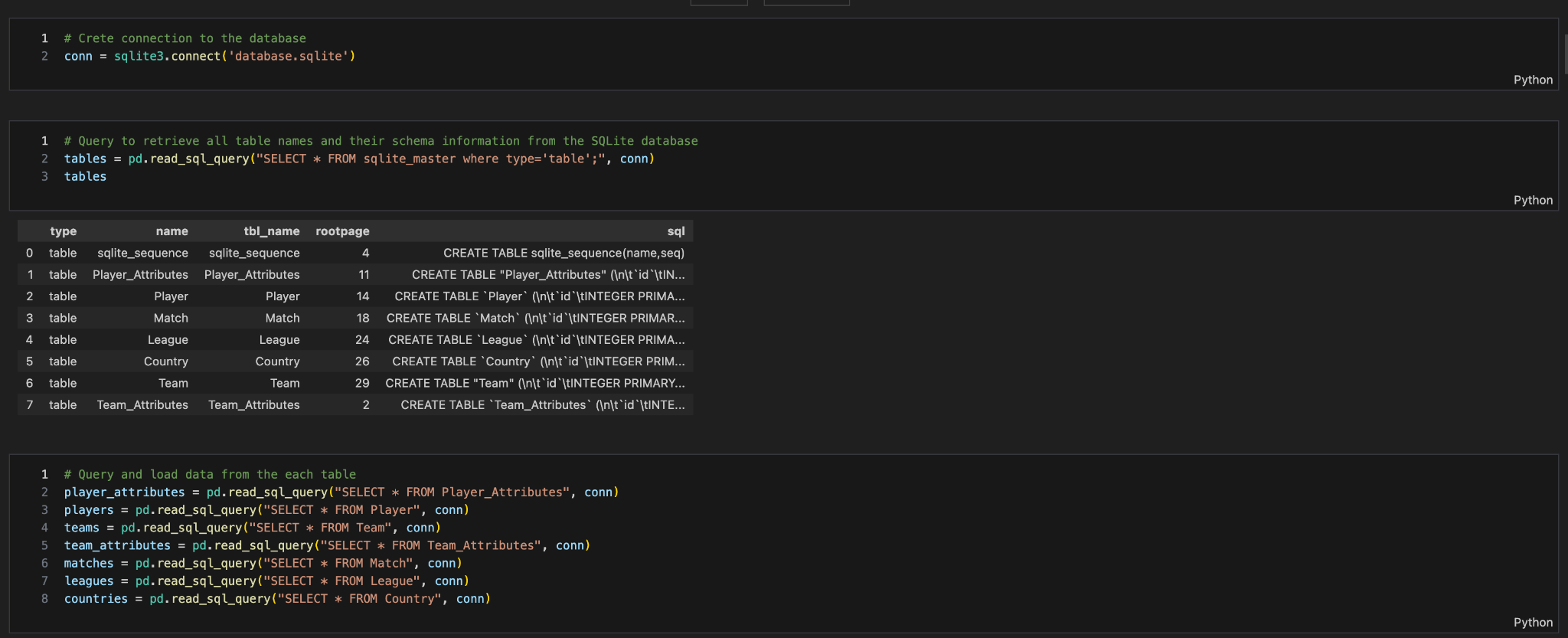
We will explore various performance metrics to address this question, including goals scored, wins, build-up play speed, and player ratings. By comparing these metrics across the top 5 leagues, including the English Premier League, La Liga, Serie A, Bundesliga, and Ligue 1, we aim to identify patterns and trends that highlight the distinctive characteristics of each league. Understanding these variations can provide valuable insights for coaches, analysts, and fans and contribute to a deeper appreciation of the dynamics within European soccer.

**Loading libraries and collecting data:**

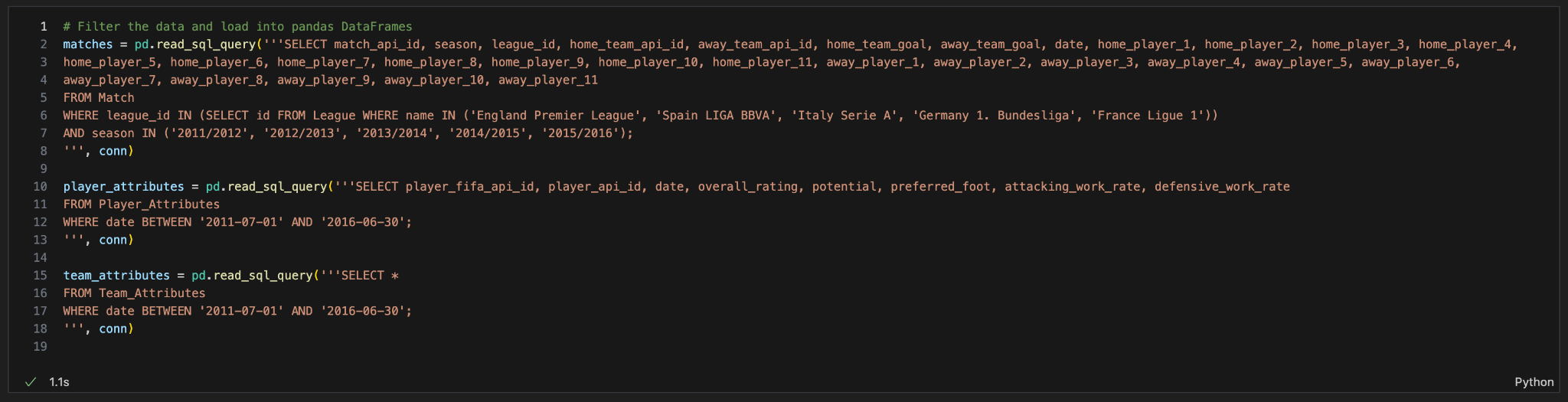


This project's data collection involved utilizing two primary datasets to analyze European soccer performance. The main dataset, "European Soccer Database," was sourced from Kaggle and uploaded by Hugo Mathien. It encompasses data from 11 European countries, covering league championships and seasons from 2008 to 2016. To focus the analysis, the dataset was narrowed down to include only the top leagues and the most recent seasons. Additionally, a secondary dataset, "Football Data from Transfermarkt," provided by David Cariboo on Kaggle, was intended to offer supplementary information on club and league relationships. However, due to a mismatch in seasons between the datasets, the integration of this secondary dataset was ultimately unsuccessful.

**Import data into pandas DataFrame:**

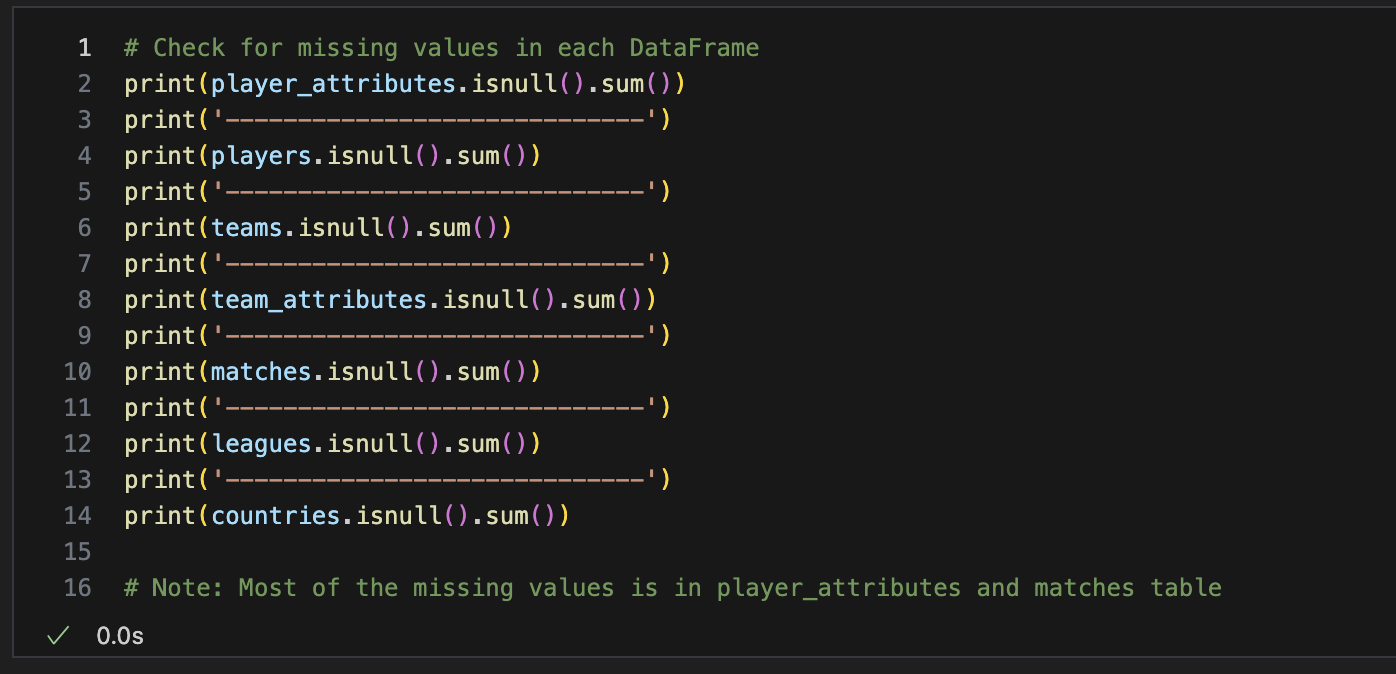
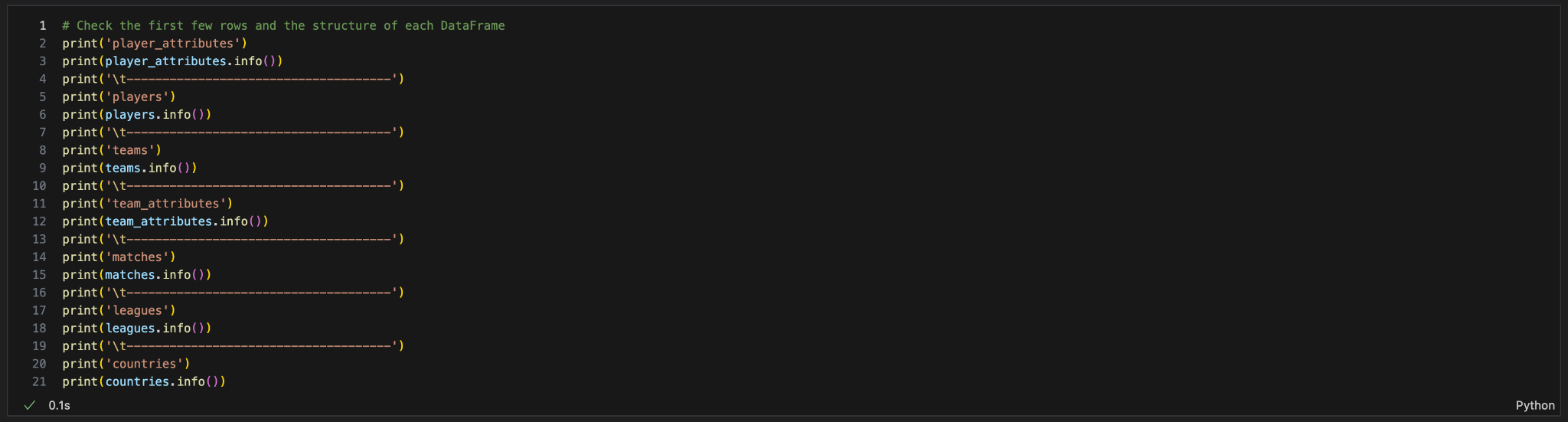


# After this step, I look at every table to decide what should be on the User’s Story/requirements to write them out and ask ChatGPT to create one that will use all the information that I wanted with interest from a User Story that I found on Baraa’s Youtube video (Baraa, 2024). One challenge I found during this process was that the data was comprehensive, specifically including 11 European Countries and 8 seasons, so I had to filter out to avoid trying to use every bit of information and ended up nowhere.



My main approach was to analyze only the 5 biggest leagues of all time based on the Global Football Ranking and the 5 latest seasons.

# **Data Exploration**

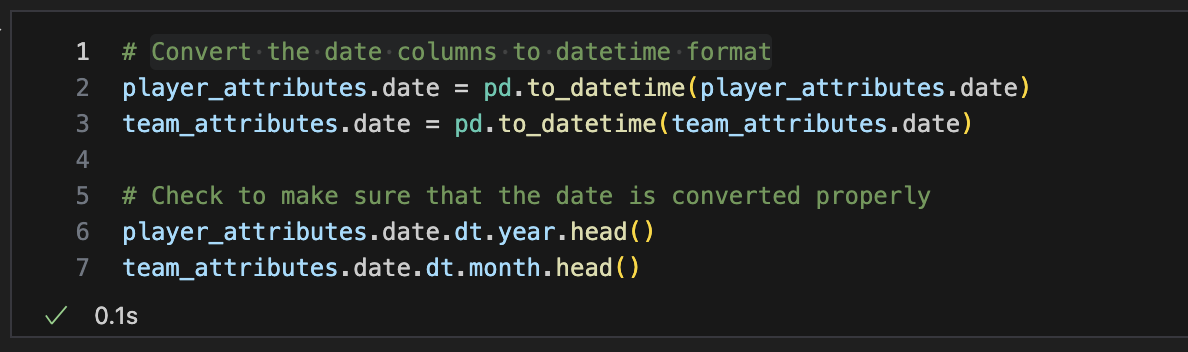




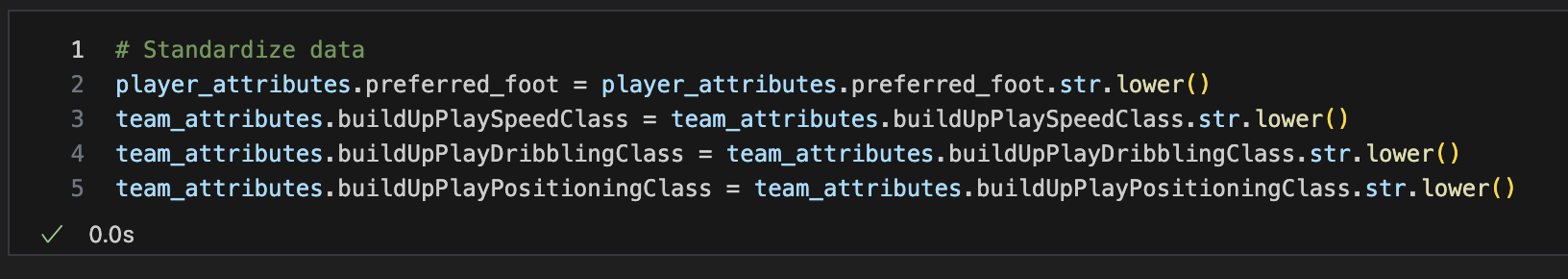
Due to the lack of a clear connection between the player and team in the first dataset, I was unable to create a proper join between them. To address this, I decided to incorporate another dataset. I chose one from Transfermarkt, uploaded by David Cariboo on Kaggle, to fill in the gaps. However, a significant issue arose: the data from Transfermarkt was from a different season than the initial dataset, resulting in mismatched data.

**Data Validation**

2.1 Convert the date columns to datetime format:

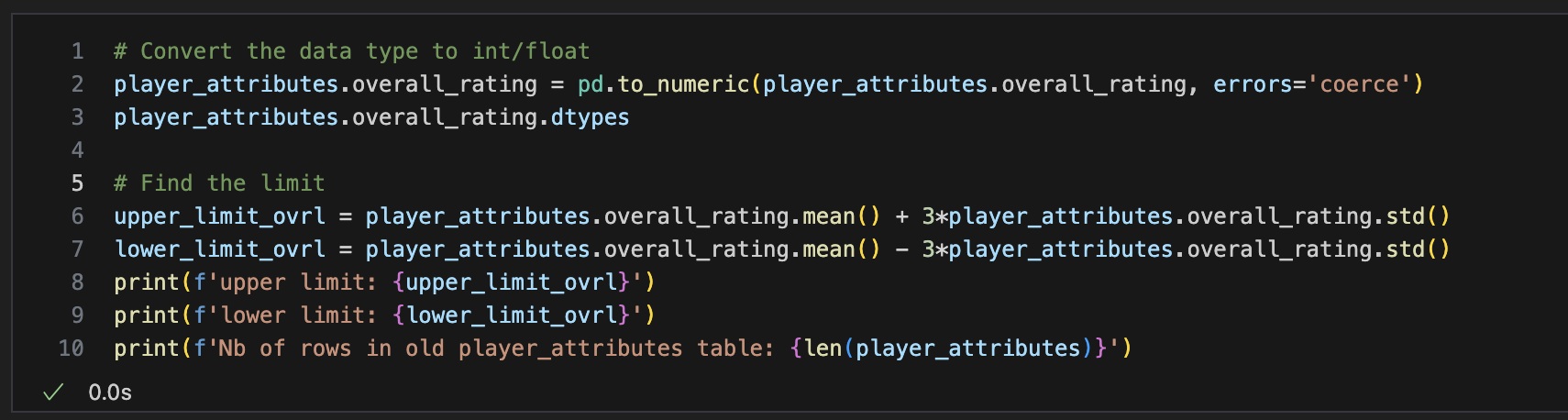


2.2 Standardize data



**Data Manipulation**

3.1 Reduce outliers by removing ratings that are outside 3 standard deviations:



In this process, I also used a box plot and bar charts to visualize the number of outliers before and after filtering.

3.2 Filtering, Merging, and Enriching Match Data for Analysis



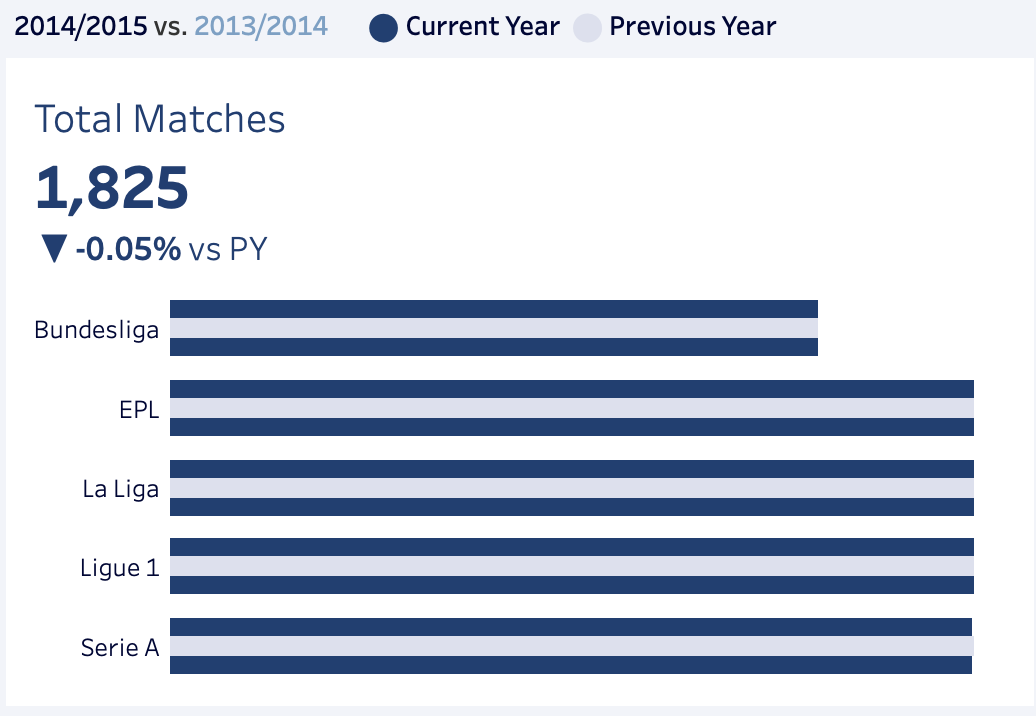
****

**Data Visualization & Analysis**



**Figure 1:** Radio button for the user to choose the season

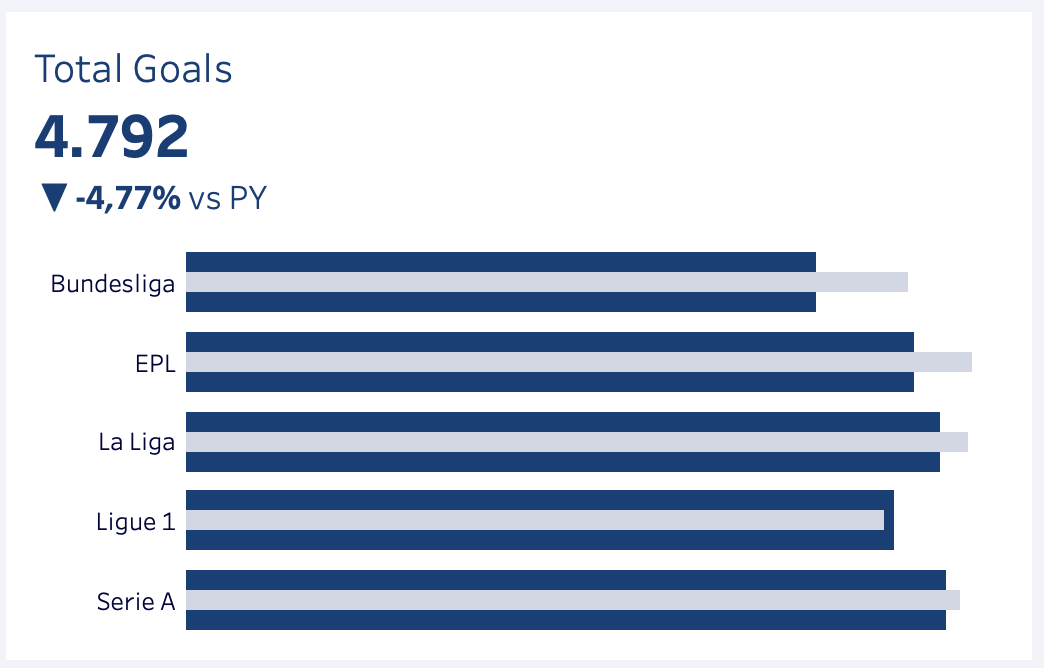
To enhance user interaction, I created a parameter that allows users to choose a specific or all seasons using a radio button. This parameter enables users to view a comparison between the selected season and the previous season. Some charts would display two graphs for each metric: one in navy representing the current season and one in grey representing the previous season. This interactive feature allows users to compare match and goal statistics easily fluctuations across different seasons, providing a clearer understanding of trends and fluctuations over time.



**Figure 2:** BAN (Big Ass Numbers) and the total number of matches in different leagues

One of the main factors affecting goal rates and relevant statistics in football is the number of matches held between leagues. Based on the chart, the number of matches throughout the five seasons remains around 1825, with a minor dip to 1804 in 2011. This consistency in match numbers suggests that goal statistics should not fluctuate significantly due to match count alone. The bar chart shows two bars, one in navy representing the current season and one in grey for the previous season, as the legend indicates. However, the data is cut off between the 2011/2012 and 2015/2016 seasons, so there is no second bar for the 2010/2011 season.

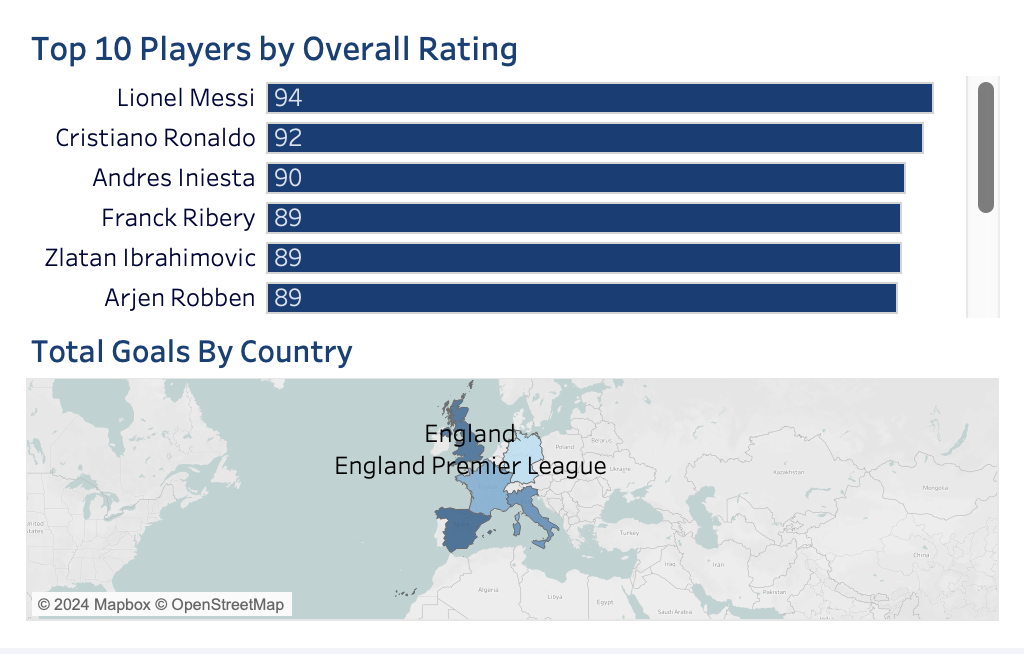
This analysis disproves my sixth hypothesis, which proposed that there would be an increase in matches over this period due to more teams joining the big leagues. The data indicates that the number of matches has remained stable, thus nullifying the expected increase. The absence of comparative data for the earliest season further emphasizes the stability of match numbers in the subsequent seasons, supporting the conclusion that fluctuations in goal rates are likely influenced by factors other than the number of matches.



**Figure 3:** BAN and the total number of goals in different leagues

With regard to the number of goals, the data fluctuates significantly more than the number of matches. The chart, similar to the previous Figure 2 with bar charts comparing the current and previous seasons, includes BANs for clarity. Notably, there is a downtrend in the number of goals following a peak of 5032 goals across all five leagues during the 2013/2014 season. Of the five leagues, La Liga from Spain has proved its offensive attacking style by having the highest number of goals scored per season, and so does all seasons, aligning with my third hypothesis.

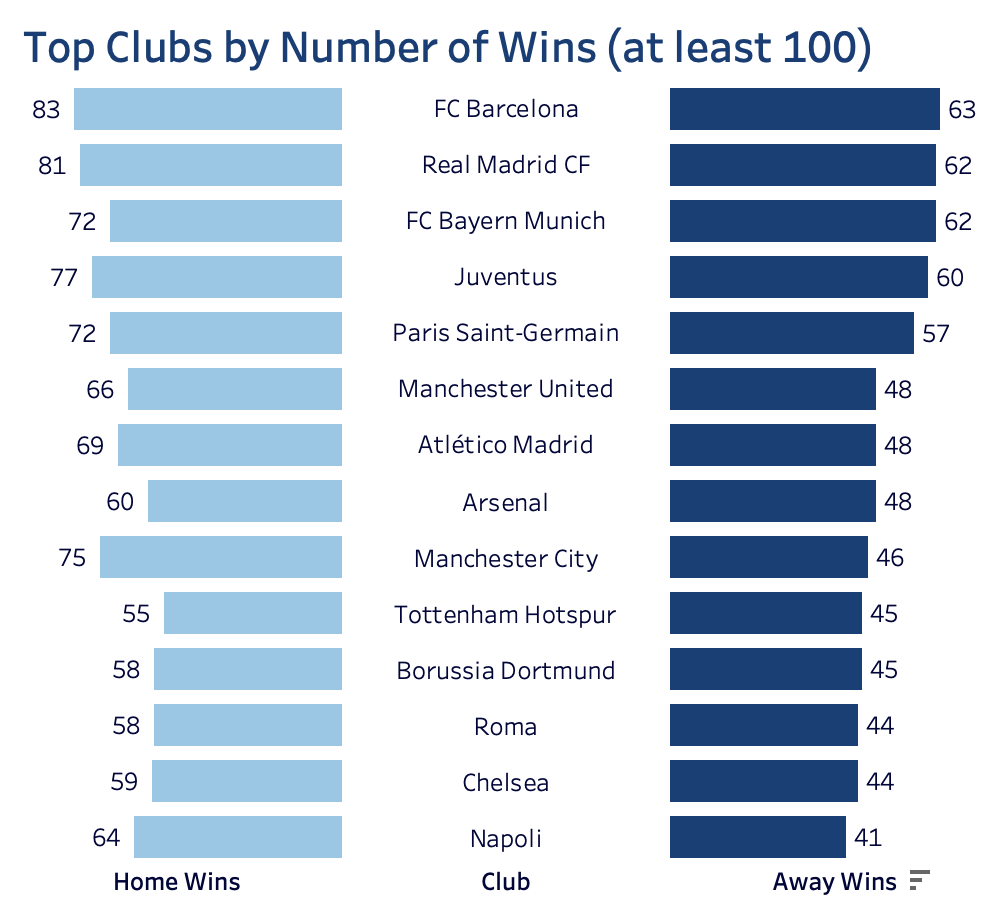
This observation disproves my fifth hypothesis, which suggested a significant upward trend in the number of goals from 2011/2012 to 2015/2016, indicating improving competitiveness and quality over time. Instead, the data shows that after reaching a high point, the number of goals declined, suggesting that factors other than an increase in competitiveness and quality, such as changes in team strategies, player performance, or defensive improvements, might influence goal-scoring trends. Therefore, while match numbers remained stable, the variability in goal statistics highlights the complexity of football dynamics beyond mere match counts.



**Figure 4: Map of countries and the top 10 players by Fifa Overall Rating**

The map indicates that Spain's La Liga had the highest number of goals, shaded the darkest among the five countries. This is further supported by the presence of high-rating players in La Liga. Out of the top 10 players with the highest FIFA overall ratings, nine have spent most of their careers in La Liga, according to their profiles on Transfermarkt (Transfermarkt, np). This finding aligns with my second hypothesis, which posited that La Liga would have the most players with the highest ratings due to its competitive nature and emphasis on youth development.

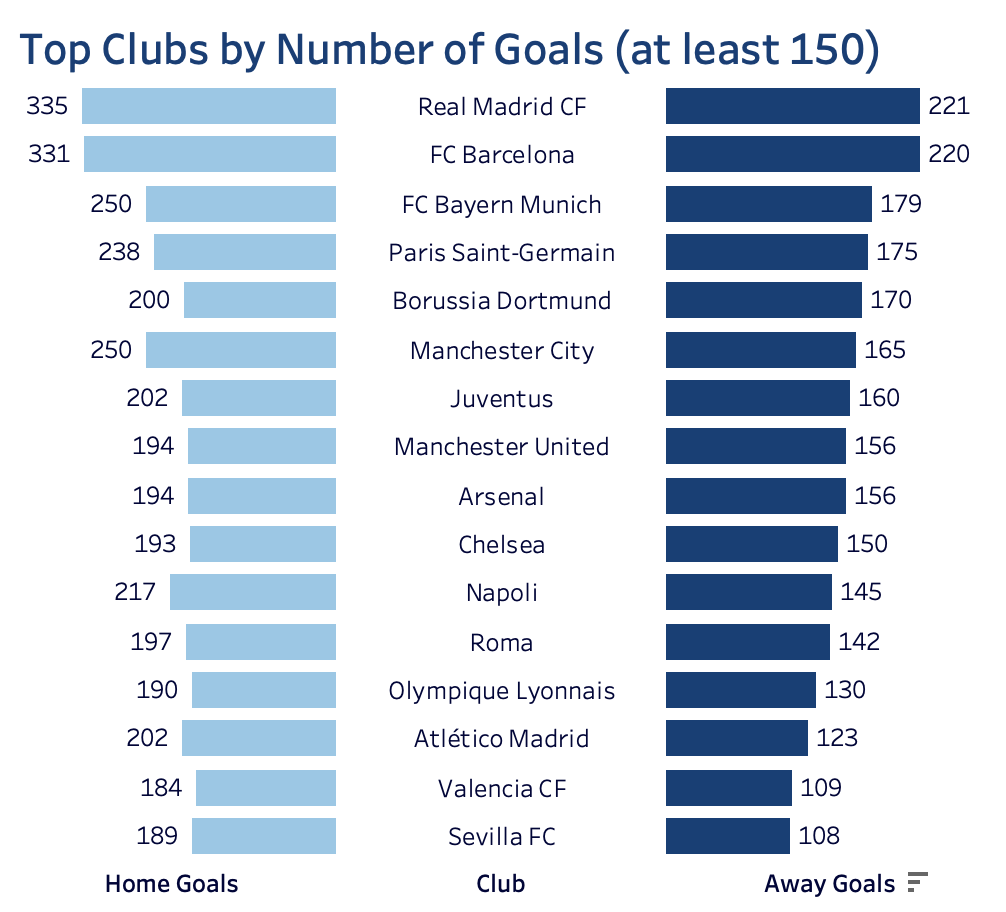
The high concentration of top players in La Liga can be attributed to the league's robust scouting and training programs, which nurture young talent and develop their skills to the highest standards. Clubs such as FC Barcelona and Real Madrid have renowned youth academies, like La Masia and La Fábrica (Football academies in Spain, np), respectively, producing numerous world-class players. La Liga's competitive environment also continually pushes players to improve and maintain high-performance levels. This combination of excellent youth development systems and a challenging competitive landscape makes La Liga a magnet for top talent, leading to its dominance in goal statistics and player ratings.



**Figure 5: Top clubs by number of wins**

For this analysis, I assume that clubs that play most of their matches in a particular country represent that country's league. Based on this implication, La Liga includes Barcelona, Real Madrid, Atlético Madrid, and Roma (AS Roma). These clubs account for most teams in this figure, demonstrating their dominance in terms of wins. This selection reflects the strength and competitiveness of La Liga, as these clubs consistently perform at high levels, contributing significantly to the league's overall success.

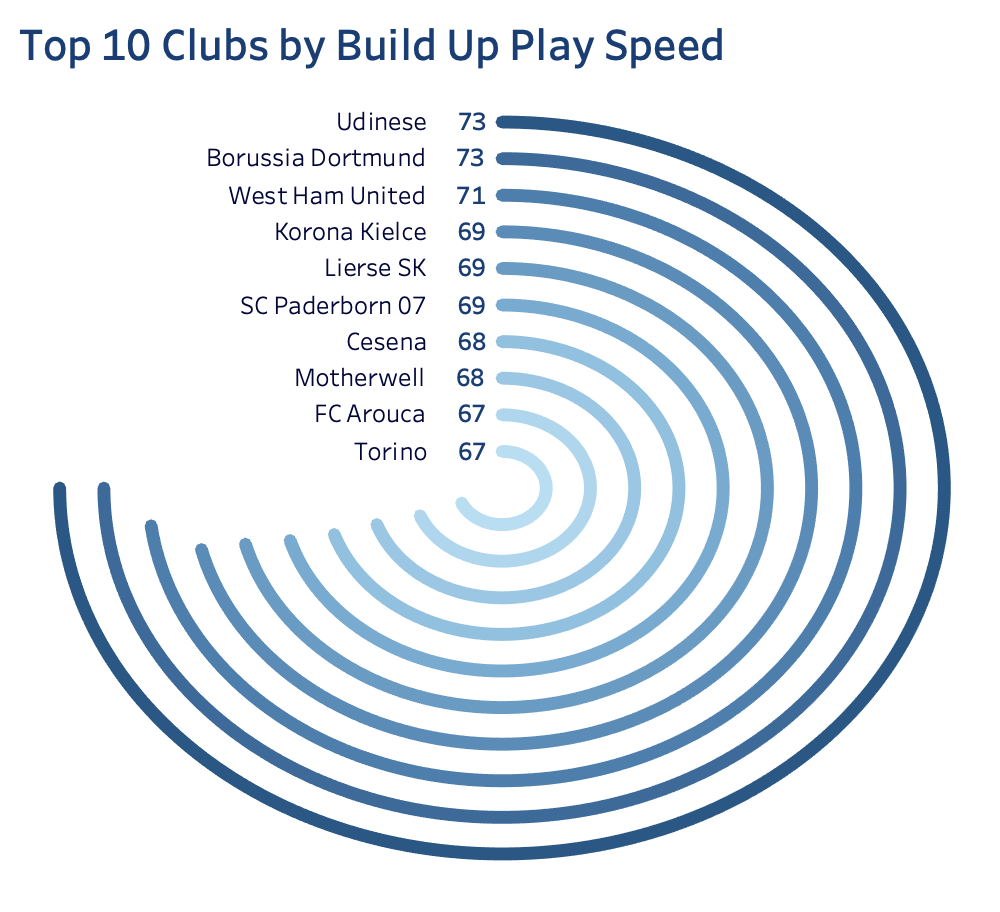
The concentration of top-performing clubs in La Liga underscores the league's competitive nature and ability to attract and develop top talent. Barcelona and Real Madrid have a long history of domestic and international success. At the same time, Atlético Madrid has established itself as a formidable competitor. However, the inclusion of AS Roma indicates a potential misclassification in the dataset, suggesting the need for data validation. Overall, the dominance of these clubs highlights their consistent performance and the high-stakes competitive environment of La Liga.



**Figure 6: Top clubs by number of goals scored**

In this figure, most of the goals scored by clubs away and at home are primarily contributed by La Liga clubs. Similar to Figure 5, the leading clubs in terms of goals scored are Real Madrid and Barcelona. Real Madrid and Barcelona scored 550+ goals each, while the third-highest goal scorer only achieved 429 goals. This shows a significant gap in offensive performance between these top two clubs and the rest, underscoring their dominant attacking capabilities.

The large goal differential highlights the exceptional quality of players and the tactical prowess of Real Madrid and Barcelona. Their ability to consistently outscore other teams by such a margin speaks to their superior offensive strategies, player talent, and overall team synergy. This dominance not only strengthens their positions within La Liga but also reinforces their reputations as two of the most formidable clubs in global football.

****

**Figure 7: Top clubs by build-up play speed**

Although La Liga clubs excel in the number of goals scored and the number of wins, they seem less competitive in build-up play speed. This indicates that these clubs might prioritize other strategies and factors that impact their overall performance more efficiently. The data suggests that build-up play speed might not be a crucial factor for success in La Liga, as their dominance in other areas, such as scoring and winning matches, demonstrates their ability to achieve high performance through different tactical approaches. Clubs like Real Madrid and Barcelona may focus more on possession-based play, quick counter-attacks, or individual brilliance to create scoring opportunities rather than relying on rapid build-up play.

This analysis shows that while build-up play speed can be an effective strategy for some teams, it is not the only path to success. Despite their slower build-up play, La Liga clubs' success in other metrics highlights the diversity of playing styles that can lead to victories. Factors such as player quality, tactical flexibility, and efficient use of resources play a significant role in achieving high performance. This reinforces the idea that football has no one-size-fits-all strategy, and teams can achieve success through various means tailored to their strengths and competitive environments.

**Interpret**

Based on the analysis of the graphs, the following conclusions can be drawn:

* The analysis highlights that La Liga clubs, particularly Barcelona and Real Madrid, consistently lead in both the number of goals scored and total wins across seasons. These clubs are significantly more prolific in their offensive output compared to other European teams. Therefore, teams in other leagues should anticipate higher competition from La Liga clubs and consider their superior scoring capabilities when formulating strategies and preparing for matches.
* The substantial goal differential between Real Madrid, Barcelona, and other clubs underscores their exceptional offensive skills and tactical expertise. With Real Madrid and Barcelona each surpassing 550 goals, while the next highest club scores only 429, this suggests that these top teams have been particularly effective in their attacking strategies. Clubs aiming to compete with these giants should analyze their approaches to scoring and consider adopting similar tactics to enhance their own offensive performance.
* Build-up play speed varies among leagues, with La Liga clubs showing slower build-up play compared to their counterparts. Despite this, their success in scoring and winning matches indicates that they may prioritize different strategies such as counter-attacking or possession-based play. Teams should consider these alternative approaches and how they can leverage different styles of play to achieve success in their own leagues.
* The stability in the number of matches played across seasons suggests that factors other than match volume are influencing goal-scoring trends. This reinforces the idea that changes in team strategies, player performance, or defensive improvements may have a more significant impact on goal statistics than the number of matches alone. Clubs and analysts should focus on these aspects when evaluating performance and competitiveness.

In conclusion, the variation in goal-scoring performance across different seasons highlights the dynamic nature of football. Teams should remain adaptable and responsive to these fluctuations when planning their strategies and preparing for future competitions. Understanding these trends and their underlying causes will be crucial for maintaining a competitive edge in the evolving landscape of European football.

**Recommendation**

Based on the analysis, several recommendations can be made to enhance performance and strategic decision-making in football:

* **Invest in Offensive Strategies for Top Leagues:** The data indicates that La Liga clubs excel in terms of goals scored and overall wins. To maintain or enhance competitiveness, other leagues and clubs should focus on strengthening their offensive strategies. This may involve investing in high-quality attacking players, improving tactical approaches, and prioritizing training that enhances goal-scoring capabilities.
* **Adapt to Tactical Flexibility:** While build-up play speed is less emphasized by La Liga clubs, other leagues might benefit from adopting or refining this strategy. Teams should explore diverse tactical approaches and adapt their play styles based on the strengths and weaknesses of their opponents. Further research into the effectiveness of different playing styles could provide insights into which tactics could be beneficial for various clubs and leagues.
* **Focus on Youth Development and Scouting:** The dominance of La Liga clubs in player ratings highlights the importance of robust youth development programs and scouting systems. Other leagues should consider investing in their youth academies and scouting networks to develop top talent. This could involve enhancing training facilities, recruiting skilled coaches, and fostering partnerships with local talent development programs.
* **Monitor and Adjust to Market Trends:** The fluctuations in goal statistics and match performance underscore the need for clubs to stay informed about trends and changes in the football market. Teams should continuously monitor performance metrics and adjust their strategies accordingly. This includes keeping an eye on changes in player performance, team strategies, and league dynamics.
* **Enhance Data Integration and Analysis:** To improve future analyses, it is essential to integrate data from various sources more effectively and ensure consistency across datasets. Clubs and analysts should focus on improving data accuracy and completeness to gain more actionable insights. Enhanced data collection methods and analytical tools can provide a clearer picture of performance trends and competitive dynamics.

By implementing these recommendations, clubs and leagues can better align their strategies with performance metrics, improve competitiveness, and adapt to the evolving landscape of football. Continued research and data refinement will also contribute to more effective decision-making and strategic planning in the sport.

**Conclusion**

In conclusion, the analysis of soccer performance data reveals critical insights into trends and dynamics across major European leagues. The data underscores the influence of league-specific factors, such as offensive strategies, player ratings, and tactical approaches, on team performance and competitive outcomes. Notably, La Liga stands out with its superior goal-scoring records and high player ratings, reflecting the league's emphasis on attacking play and robust talent development. In contrast, other leagues may benefit from diversifying their tactical approaches and investing in youth development to enhance competitiveness.

The findings offer valuable information for clubs, coaches, and analysts seeking to understand performance trends and make strategic decisions. By examining factors like goal statistics, match counts, and build-up play speeds, stakeholders can better understand what drives success in different leagues. Future research that further explores these factors, along with improved data integration and analysis, will provide more nuanced insights into football dynamics.

The research question, "How do player and team performance core metrics vary across the major European soccer leagues between 5 seasons, and what factors contribute to these performance differences?" is addressed by analyzing key performance indicators and trends. The insights gained from this analysis can inform strategic decisions related to team management, player acquisitions, and tactical planning, ultimately contributing to enhanced performance and competitive positioning within the football industry.

**Work Cited**

Baraa (2024) Tableau Complete Project End-to-End | Like I Do in My Real Projects, YouTube. Available at: <https://www.youtube.com/watch?v=dahrmqT5GD4&t=300s> (Accessed: 05 August 2024).

Category: Football academies in Spain (2022) Wikipedia. Available at: <https://en.wikipedia.org/wiki/Category:Football_academies_in_Spain> (Accessed: 05 August 2024).

Football League rankings (no date) Global Football Rankings. Available at: <https://www.globalfootballrankings.com/> (Accessed: 05 August 2024).

Lionel Messi (no date) Transfermarkt. Available at: <https://www.transfermarkt.com/lionel-messi/profil/spieler/28003> (Accessed: 05 August 2024).