Exploring Football Dynamics: A Comparative Study of Playing Styles, Home Advantage Effects, and Shooting Positions in the Top-5 European Leagues

DEEPAK CHAURASIYA, RAN XU, SAHIL KADADEKAR, New York University, USA

This research presents a comprehensive review of football analytics, synthesizing insights from pivotal studies that have reshaped the field with a primary focus on performing data analytics. It begins by emphasizing the holistic integration of player tracking data and tactical metrics, showcasing the dynamic potential of advanced analytics in valuing individual player actions. The application of AI in football analytics is exemplified by enriching our understanding of player behavior profiles and game patterns using game statistics. An exploration of the analytics movement in player-side sports emphasizes its transformative impact on club strategy and management. A comprehensive review on sports big data outlines shifts from simple statistics to model-based evaluation, explicit to implicit sports features, and social network to knowledge graph analysis. This research provides profound insights into player performance, team strategies, and the broader implications for talent identification and management strategies through big data analytics in football.

Additional Key Words and Phrases: Sports, Analytics, Big Data, Pyspark, Soccer, Visualization

ACM Reference Format:
Deepak Chaurasiya, Ran Xu, Sahil Kadadekar. 2023. Exploring Football Dynamics: A Comparative Study of Playing Styles, Home Advantage Effects, and Shooting Positions in the Top-5 European Leagues. 1, 1 (December 2023), 13 pages.

1 INTRODUCTION

The landscape of football analytics has undergone a profound metamorphosis in recent years, spurred by the assimilation of multifaceted data sources and the application of sophisticated analytical methodologies [7]. This comprehensive literature review aims to embark upon an exhaustive exploration of seminal studies that have wielded considerable influence, contributing significantly to the ever-evolving panorama of football analytics.

In tandem with the burgeoning influence of data-driven methodologies, recent advancements underscore the integration of cutting-edge technologies such as machine learning and predictive modeling in football analytics. These technological interventions have not only revolutionized the evaluation of player performance but have also provided novel insights into team dynamics, enabling a granular understanding of on-field strategies and the intricate interplay between various contributing factors.

Furthermore, the increased accessibility and utilization of sports big data have facilitated a shift from traditional statistical evaluations to more nuanced, model-based approaches, offering a more comprehensive understanding of player dynamics, team interactions, and strategic decision-making processes [2]. This transition has not only augmented the depth of insights derived from analytics but has also paved the way for the exploration of implicit sports features, knowledge graph analysis, and intricate social network dynamics within the sporting ecosystem.

In this study, our primary objective is to develop methodologies tailored for the analysis of sports data within a distributed framework. The evolving landscape of soccer data, characterized by its increasing richness and granularity, necessitates scalable approaches capable of efficiently processing substantial volumes of information. As the intricacies of soccer-related datasets continue to expand, there is a growing demand for methods that not only ingest extensive

Author's address: Deepak Chaurasiya, Ran Xu, Sahil Kadadekar, New York University, P.O. Box 1212, Brookyln, New York, USA.

data sets but also possess the capacity to derive meaningful and actionable analytics. This research aims to address these challenges by proposing innovative techniques that enable comprehensive analysis of soccer data within the context of a distributed computing framework.

2 LITERATURE REVIEW

Goes et al. [7] advocate a holistic paradigm shift in football analytics, urging the integration of player tracking data and tactical metrics. Their research emphasizes the symbiotic potential of merging diverse data sources for deeper insights into player behaviors and team strategies, with advanced analytics, particularly clustering methods, serving as potent tools for nuanced football analysis. Simultaneously, Decroos et al.'s [3] milestone study introduces SPADL, a specialized language for event stream data in soccer, and VAEP, a groundbreaking framework assigning values to player actions. VAEP's holistic valuation, nuanced consideration of game context, and predictive capabilities offer a dynamic perspective on player contributions, differentiating actions enhancing scoring chances with positive values from those diminishing chances with negative values.

In 2020, García-Aliaga et al. [6] demonstrate AI's application in football analytics, categorizing players based on technical-tactical behavior using game statistics. This AI-driven approach significantly contributes to understanding behavior profiles and game patterns, providing pragmatic insights for optimizing player placement in club youth teams. The findings suggest a promising avenue for leveraging AI in talent identification and player development, ultimately enhancing player progression and performance.

Foster et al. [5] highlight the transformative impact of the analytics movement in player-side sports, emphasizing its influence on club strategy, management dynamics, and the increased demand for analytics talent from diverse industries. The broader adoption of analytics is expected to drive continuous innovations in player squad assembly, game strategy, and health and fitness, extending its impact beyond professional sports to broader business and societal contexts. In a comprehensive review, Bai et al. [2] delineate the shift in sports big data from simple statistic evaluation to model-based approaches, explicit to implicit sports features, and from social network to knowledge graph analysis. Despite progress, challenges such as predicting athlete performance in knowledge graphs and establishing unified sports big data platforms persist, making the review a valuable guide for navigating obstacles and maximizing the benefits of sports big data in analytics.

In 2022, Fonti et al. [4] examine the rising trend of scholars leveraging sports settings for management research. Their review offers a balanced exploration of using sports data in this context, providing reflections and a roadmap for future studies. Simultaneously, Herold et al.'s [8] investigation into positional tracking and data-driven video intervention in football highlights a pivotal moment in technology application on the field. Despite marginal improvements, the study prompts questions about video feedback transferability and underscores the need for collaborative efforts between coaches and researchers in integrating data science seamlessly into football practice.

Collectively, these studies weave a rich tapestry of football analytics, showcasing the integration of diverse data sources, advanced analytical techniques, and the transformative potential of AI. The literature not only underscores the evolving landscape of football analytics but also provides profound insights into player performance, team strategies, and the broader implications for talent identification and management strategies.

3 EXPERIMENT SETTING

This section describes the dataset employed in our analysis and notes the PySpark configurations used during the execution of our experiment within a distributed computing environment.

3.1 Dataset

We utilize the data provided by Statsbomb to analyze the performance of Soccer teams, individual players, and details of an instance in the game. StatsBomb is a prominent sports analytics company that provides comprehensive and granular Soccer data. The dataset is broadly divided into 5 large classes, namely, Competitions, Matches, Lineups, Events, and 360-degree. Each of these classes is interrelated and their raw form is present in multiple JSON files. The size description of the data is shown in Table 1.

- **Competitions:** The competition JSON contains attributes of various competitions, multiple seasons of each competition, and their primary keys. For instance, one competition block would describe the *"2018/19"* season of the *"English Premier League"* competition
- **Matches:** The matches JSON lists out attributes of all the matches belonging to a particular season. It details the home and away teams participating in the match, their managers, the final score, and other features of the match. For instance, one match block would describe *"Match 36: Arsenal v/s Liverpool"* of the *"2018/19"* season of the *"English Premier League"* competition
- **Lineups:** The lineups JSON describes the players participating in a particular match. It contains the names of the players, their team, their position on the field, their nationality, whether they started the match or were substituted later, and primary keys. There is one JSON file corresponding to one match, identified by match_id. For example, one lineup block would describe *Mesut Özil* (an Arsenal player) and his aforementioned attributes for the *"Match 36: Arsenal v/s Liverpool"* match of the *"2018/19"* season of the *"English Premier League"* competition
- **Events:** The events JSON is a timestamped data containing all the events that occurred in a particular match. The file lists the type of event, time of occurrence, players involved, and other event-specific attributes. Similar to the lineups JSON, there is one event JSON of a given match, identified by match_id. There are 33 distinct events in the data, their distribution is shown in Figure 1. For instance, one event block would describe a *"Goal"* event, the time of the goal, the scoring team, the scoring player, the body part used to score the goal, and many other attributes of the *"Match 36: Arsenal v/s Liverpool"* match of the *"2018/19"* season of the *"English Premier League"* competition
- **360-degree:** 360 JSON contains the view of the whole pitch (visible to the camera) at the time of any event in the match. It lists the visible area of the pitch and the position of each player in that area. This data is very granular and is not available for all the matches. We would only be using this data while trying to create a video playback of the whole game.

All of Statsbomb's data[1] is proprietary but they have made a sample available to the public for research. The public version of the data can be found here.

Class	Entity Count	Size
Competitions	18	33KB
Matches	67	6MB
Lineups	3199	67.2MB
Events	3199	9.67GB
360-degree	211	1.45GB

Table 1. Dataset size description

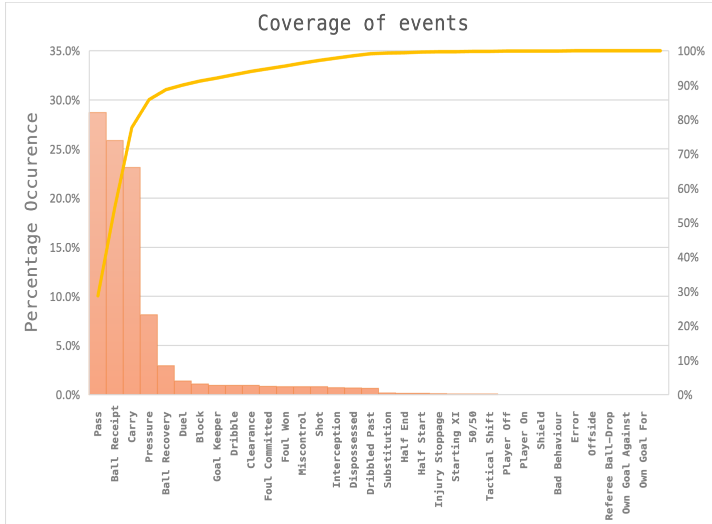


Fig. 1. Distribution of events

3.2 Environment

The analysis is being carried out on Pyspark version 3.1.2 using Scala version 2.12.14 and Python 3.8.13. With no dynamic allocation, spark is configured with 2 instances, 8530MB executor memory, 8 executor cores, and 30720MB driver memory. The code to replicate the analysis shown in this paper can be found here.

4 DATA FLOW PIPELINE

This section describes the essential steps involved in this project, as shown in Figure 2. For mid-review, we have completed the data ingestion and data cleaning part of the pipeline and are starting to work on the data analysis.

4.1 Data Ingestion

In the initial step of our data analysis process, we focus on uploading the datasets to HDFS, enabling subsequent data cleaning and analysis using PySpark. The process begins with the acquisition of a ZIP file from the original website, which is subsequently uploaded to the local Dataproc environment. Following this, the ZIP file is unpacked locally, allowing access to the individual datasets contained within. These datasets are then transferred to HDFS. Then we used the multiline option in Pyspark employed for reading the JSON files within each of the designated folders: matches, lineup, and events.

4.2 Data Cleaning

In the subsequent phase of our data processing workflow, our focus turns to data cleaning. Given that the original datasets are in JSON format without a predefined schema, our objective is to transform them into a structured format suitable for further analysis. PySpark is instrumental in achieving this goal.

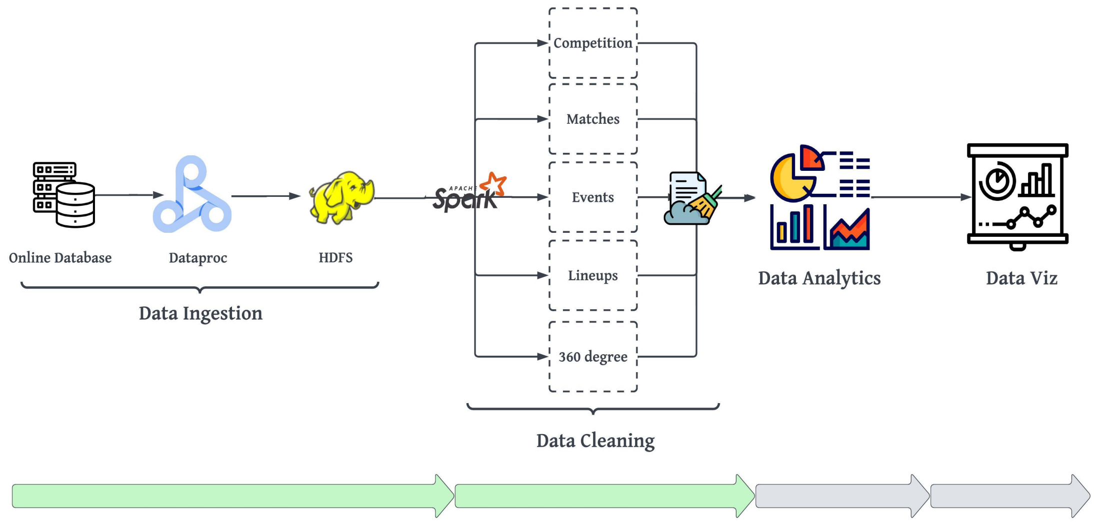


Fig. 2. Data Flow Pipeline

Utilizing PySpark's Select operation, we strategically choose the attributes deemed essential for our analysis. This step helps streamline the datasets by retaining only the relevant information, paving the way for a more concise and focused exploration.

Once the selection process is done, we convert the data into a standard parquet format. We end up with three parquet files—one for competitions, one for lineups, one for matches, and one for 360-degree data. When it comes to event data, we merge all the 3199 JSONs into 33 parquets, with each parquet containing event data for a specific type across all matches. These files are designed to be easily joined, making it simple to access the specific data needed for our analysis.

This format enhances interoperability and facilitates downstream analysis using various tools and platforms. As a result, we obtain distinct parquet files, each serving a specific analytical purpose.

4.3 Data Analysis and Visualization

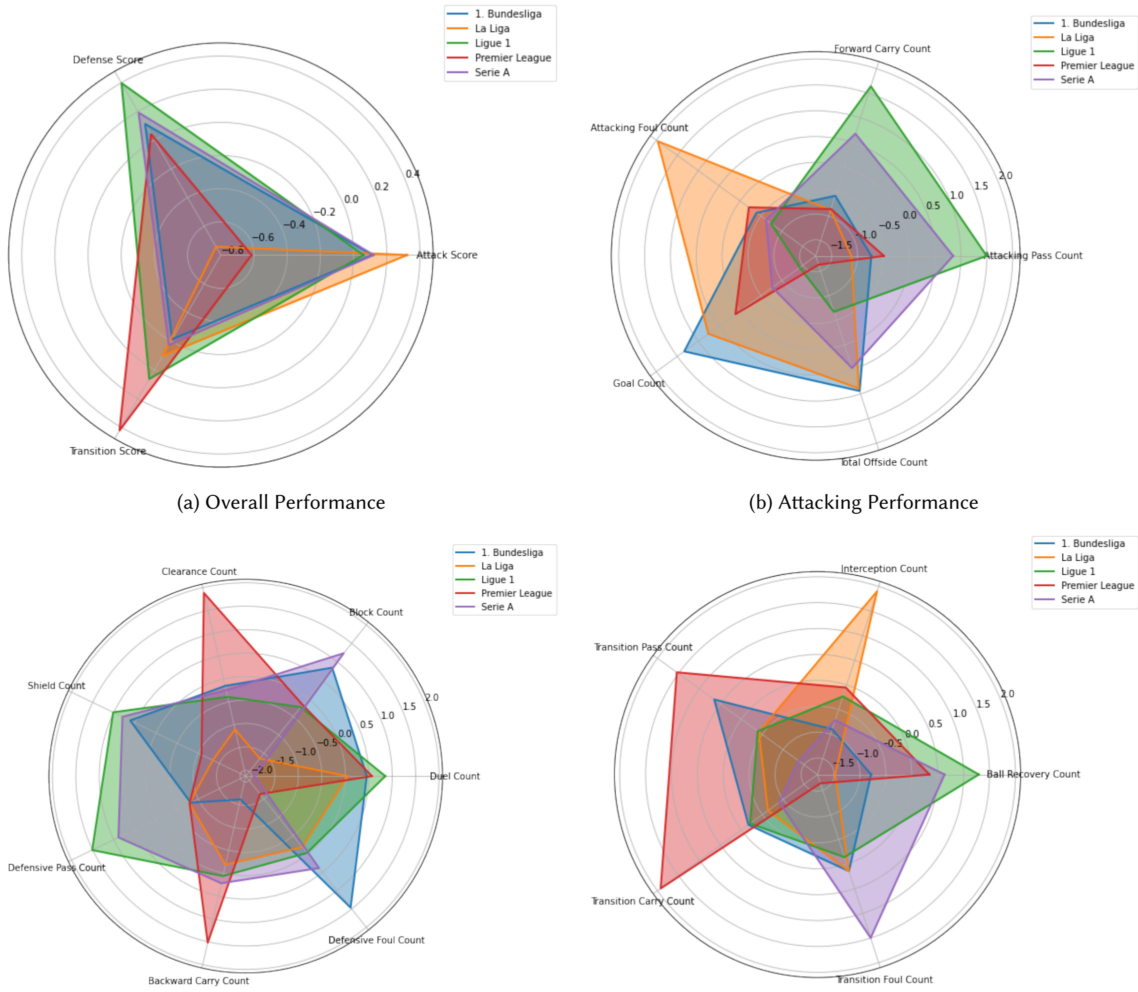
There are mainly four questions that we focus to analyze about the dataset.

- **Comparative study of playing styles across leagues**
- **Is Home Advantage a myth?**
- **Comprehensive Analysis of Barcelona's Performance over Time**
- **Shot Location of Teams**

5 ANALYSIS AND FINDINGS

5.1 Comparative Study of Playing Styles Across Leagues

In this comprehensive analysis, we delve into the distinct playing styles and strategic nuances exhibited by football teams across five premier European leagues: Bundesliga, La Liga, Ligue 1, Premier League, and Serie A for the 2015/16 season. By scrutinizing key performance metrics encompassing attack, defense, and transition phases, we unravel



(a) Overall Performance

(b) Attacking Performance

(c) Defensive Performance

(d) Attacking Transition Performance

Fig. 3. Comparision of Top-5 Leagues in Football (Europe) over Attacking, Defensive and Transition metrics

the unique identities that define each league. La Liga's emphasis on individual skill and longer passes in the attack, the Premier League's aggressive and dynamic attacking approach, Ligue 1's robust defense, Serie A's focus on ball retention, and the Bundesliga's well-rounded performance emerge as defining characteristics. As we explore each league's strengths and playing philosophies, a nuanced understanding of their footballing dynamics unfolds, showcasing the diverse and captivating nature of the beautiful game across these European football powerhouses. Fig. 3a summarises this comparison of Europe's top 5 leagues.

5.1.1 Attack Analysis. La Liga distinguishes itself with a unique attacking style, placing a strong emphasis on individual skill as evidenced by the league's highest average dribble count. This highlights a strategic commitment to creating goal-scoring opportunities through the flair and artistry of individual players, defining La Liga's attacking philosophy.

As shown in Fig. 3b, La Liga's proclivity for longer passes in the attacking phase underscores a strategic focus on creating opportunities from a distance, utilizing the vision and passing ability of players to unlock defenses with precision over longer distances.

In contrast, the Premier League showcases an aggressive attacking approach, consistently leading in the average shot count. This statistic reflects the league's dynamic and goal-oriented playing style, where teams actively seek opportunities to shoot and score. Moreover, both Premier League and Serie A teams display disciplined attacking movements, reflected in lower average total offside counts. This disciplined approach signifies a strategic and organized attacking style, where teams prioritize well-timed and coordinated offensive maneuvers. The emphasis on disciplined attacking structures minimizes errors in the final third and underscores a shared commitment to strategic play in both leagues.

5.1.2 Defence Analysis. Fig.3c shows the indepth analysis of how teams in these leagues defend. Ligue 1 distinguishes itself with a robust defensive style, leading in the average duel count and emphasizing physicality and defensive resilience. This underscores the league's commitment to one-on-one battles and the importance of winning defensive duels.

The Premier League showcases a proactive defensive approach, particularly in blocking shots, highlighting a resolute mindset to deny opponents goal-scoring opportunities. Premier League teams also lead in average clearance count, reflecting a pragmatic strategy to efficiently clear defensive zones, emphasizing the league's focus on maintaining a solid defensive foundation.

Serie A's defensive emphasis is evident in the highest average shield count, showcasing a focus on ball retention in defensive scenarios and contributing to a composed and possession-oriented defensive style. La Liga, with a preference for longer defensive passes, potentially aims to switch play and relieve pressure from the defensive third, emphasizing precision and vision in distribution. Serie A's controlled defensive passes indicate a possession-oriented style, while Premier League's lead in backward carry count signifies a tendency to play out from the back with controlled backward ball carries. The Premier League's dominance in backward average carry distance suggests an ability to initiate attacks from deep defensive positions, showcasing strategic play from their own half. Serie A's higher average defensive foul count may reflect a physical defensive approach or tactical fouls to disrupt opponent attacks, contributing to the league's defensive solidity. Ligue 1's higher average defensive foul share suggests a more aggressive defensive style, emphasizing assertiveness and intensity in defensive plays.

5.1.3 Transition Analysis. According to Fig. 3d, Premier League teams exhibit a dynamic transition game, covering the most distance with their passes during transitional phases, indicating a strategic willingness to switch play quickly. Additionally, both Premier League and Serie A teams focus on quick and precise passing, as reflected in their higher average transition pass counts, showcasing an emphasis on maintaining ball control and accuracy in transitions. The Premier League takes the lead in average transition carry count, highlighting a proactive approach to ball progression during transitional phases. Teams from the Premier League further showcase dominance in average transition carry distance, emphasizing their ability to cover significant ground while carrying the ball, indicative of effective and expansive transitions from defense to attack. Premier League's superiority in average transition attacking carry distance underscores their proficiency in ball progression from deep defensive positions to advanced areas during transitional play. However, the higher average transition foul count and share in the Premier League suggest a more intense style of play during transition phases, characterized by a greater number of fouls committed compared to other leagues.

5.1.4 Overall Performance. Bundesliga teams demonstrate a well-rounded performance across attack, defense, and transition phases. La Liga prioritizes a strong attacking style, contributing to high overall scores, mainly driven by impressive attack metrics. Ligue 1 teams focus on a solid defense with a balanced attack, resulting in strong overall scores, particularly in defense and transition. The Premier League stands out for its aggressive attacking and strong

Fig. 4. Comparision of Home Advantage of Teams belonging to different Leagues

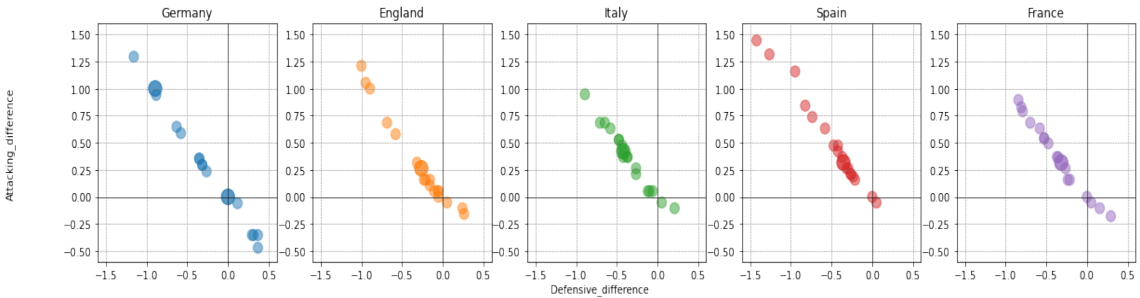
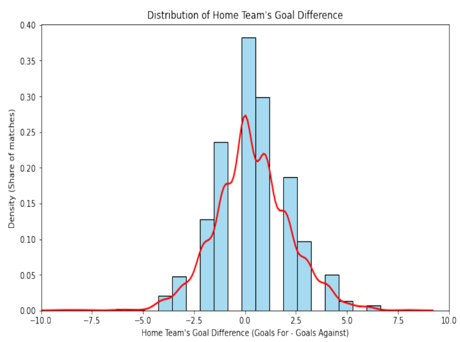
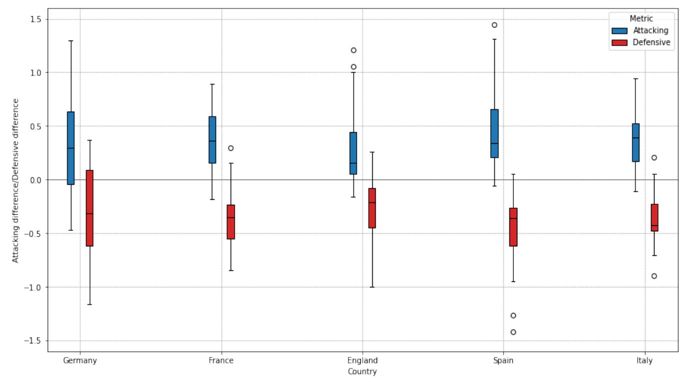


Fig. 5. Variation of Home Advantage in Teams belonging to different Leagues



(a) Distribution of Home team's Goal Difference



(b) Variation in prevalence of Home Advantage in various Leagues

Fig. 6

defensive approach, yielding the highest overall scores across all phases. Serie A teams exhibit a balanced performance, reflected in strong overall scores indicative of their well-rounded playing styles.

In conclusion, these findings can guide players in aligning playing styles with football leagues for informed career decisions. These insights benefit individuals aiming to optimize talents within the diverse football landscape, while also aiding sports recommendations to attract newcomers, enhancing overall engagement and expanding viewership. The research contributes to the broader football ecosystem by increasing accessibility and appeal.

5.2 Is Home Advantage a Myth?

In our analysis, we focused on a specific season, namely 2015/2016, to delve into the dynamics of home and away team performance. During this season, the mean score for the home team stood at 1.6, while the mean score for the away team was 1.26, indicating a noticeable difference favoring the home team.

Furthermore, we examined the home advantage by calculating the score difference between the home and away teams for each match and visualizing it on a plot(Fig.6a). The plotted data revealed a distinct skewness to positive, emphasizing a discernible advantage for the home teams throughout the season. And also the percentage of the home team winning is 44.4%, while the percentage of the away team winning is 28.6%.

Considering these findings, it is evident that a home advantage does exist in the context of the 2015/2016 season.

Expanding our analysis, we examined the home advantage at both the team and country levels. To do this, we calculated the mean score for each team when playing as the home team and as the away team. The home advantage was then derived by subtracting these values.

The resulting plot revealed a prevalent trend: a majority of teams exhibited a home advantage, as indicated by positive home advantage scores. This observation suggests that teams generally perform better when playing on their home turf.

As shown in Fig.4, each country displayed a diverse range of home advantage scores, with some teams showing negative home advantage. Of particular interest is the observation that Germany stands out with the most negative home advantage among the countries analyzed.

Expanding our analytical scope, we investigated the performance dynamics of teams by examining their scores when acting as the home team and away team. We use their score as attacking and their opponent's score in that match as defensive. So we calculate the attacking and defensive separately when each team is as home team and away team. And then we calculate the difference as attacking difference and defensive difference.

Fig. 5 based on these differences provided a comprehensive visualization of team dynamics among countries. This approach allowed us to gain insights into how teams perform both offensively and defensively, shedding light on their strengths and weaknesses in different match scenarios. We can see from the plot that most of the teams are more attacking and less defensive when playing at home. We can see the scatter plot intuitively that England has less difference when they are away or at home. Germany varies from each team, and Italy has a larger difference than Spain and France. And Fig. 6b gives us a more accurate way of our result, which allies our intuition in the scatter plot that Germany has a broad play style and England has the smallest influence by the home advantage.

5.3 Barcelona's Performance Over the Years

The dataset under scrutiny offers a comprehensive overview of FC Barcelona's league performance across multiple seasons, honing in on crucial metrics related to attack, defense, and transition. By analyzing key indicators such as shot counts, goal tallies, defensive clearances, and transition passes, this examination aims to unravel the intricate details of Barcelona's playing style. The data not only serves as a quantitative record of the team's on-field exploits but also provides valuable insights into the strategic evolution of the club over time. This exploration delves into the nuances of Barcelona's footballing philosophy, offering enthusiasts and analysts a deeper understanding of the team's strengths, patterns, and tactical choices throughout various periods.

5.3.1 Attacking Style. Barcelona consistently maintained a high shot count, peaking at 9.05 in 2012/2013, showcasing their proactive approach in creating goal-scoring opportunities. Their clinical finishing is highlighted by an average goal count of 1.18 per game in the same season, reflecting efficiency in converting chances.

The team's commitment to intricate passing play peaked at 186.42 in 2012/2013, emphasizing their possession-oriented attacking strategy and focus on strategic ball circulation. The consistent average dribble count underscores Barcelona's reliance on individual skill in the attacking phase, with a peak of 19.10 dribbles per game in 2013/2014 indicating an increased emphasis on one-on-one play.

5.3.2 Defensive Style. Barcelona's defensive resilience is consistently strong, with an average shield count consistently above 1.27 and a peak of 1.80 in 2012/2013, reflecting a robust defensive structure. The team effectively disrupts opponent plays with a steady average interception count, reaching a peak of 12.76 in 2012/2013, indicating sharp defensive anticipation.

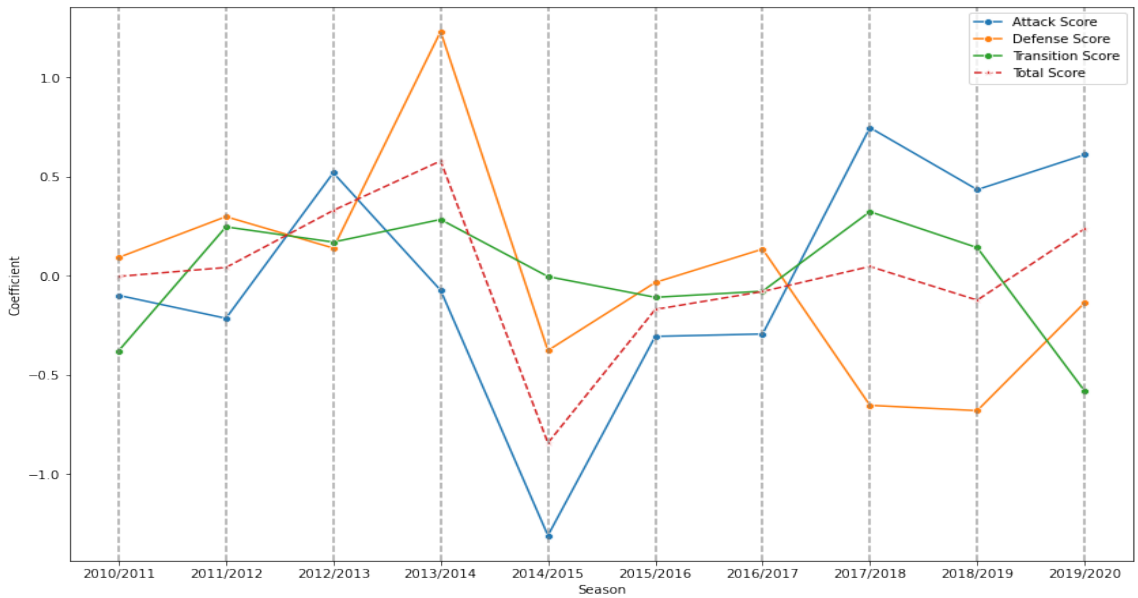


Fig. 7. Change in Barcelona's playing style over years

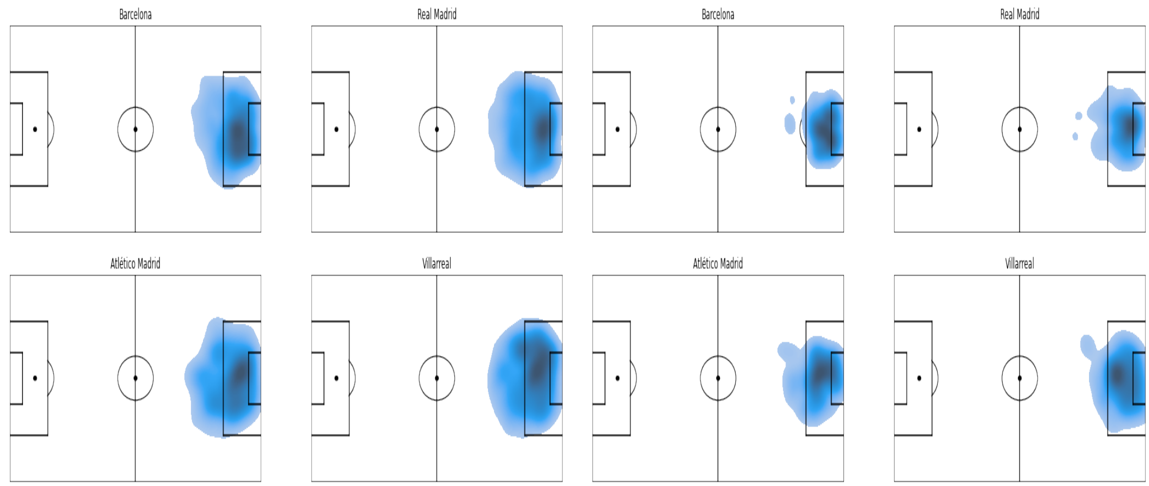
While the average clearance count remains stable, Barcelona's ability to handle defensive situations is evident, with a peak of 26.55 clearances per game in 2013/2014 suggesting a phase where defensive resilience was put to the test. The team maintains defensive discipline, as reflected in consistently low average defensive foul counts, with the lowest value of 2.67 in 2011/2012 indicating a season marked by defensive solidity without resorting to fouls.

5.3.3 Transition Scores. Barcelona excels in quick ball recovery, consistently maintaining counts above 46.90, and a peak of 49.56 in 2013/2014 underscores their proficiency in swift possession regain. The forward carry count reflects their efficient progression from defense to attack, reaching a peak of 161.91 in 2012/2013, indicating a season of particularly effective ball advancement. The commitment to preventing counterattacks is evident in Barcelona's consistently low average transition foul count, with the lowest value of 1.02 in 2013/2014 indicating a season where transitions were managed with minimal fouls. The team's transition strategy involves quick and precise passing, with the highest transition pass count of 22.17 in 2013/2014 pointing to a season marked by effective ball circulation during transitions.

In conclusion, Barcelona's tactical performance, as evidenced by numerical metrics, reflects a harmonious balance between attacking prowess, defensive solidity, and efficient transitions. *The major shift in these scores represent a change in philosophy of the team playing style. Fig. 7 shows drastic change in the year 2013/14, 2014/15 and 2017/18. Interestingly, these snapshots coincide with managerial (coach) changes at the club.* One can infer that these metrics sufficiently showcase the playing style of the club.

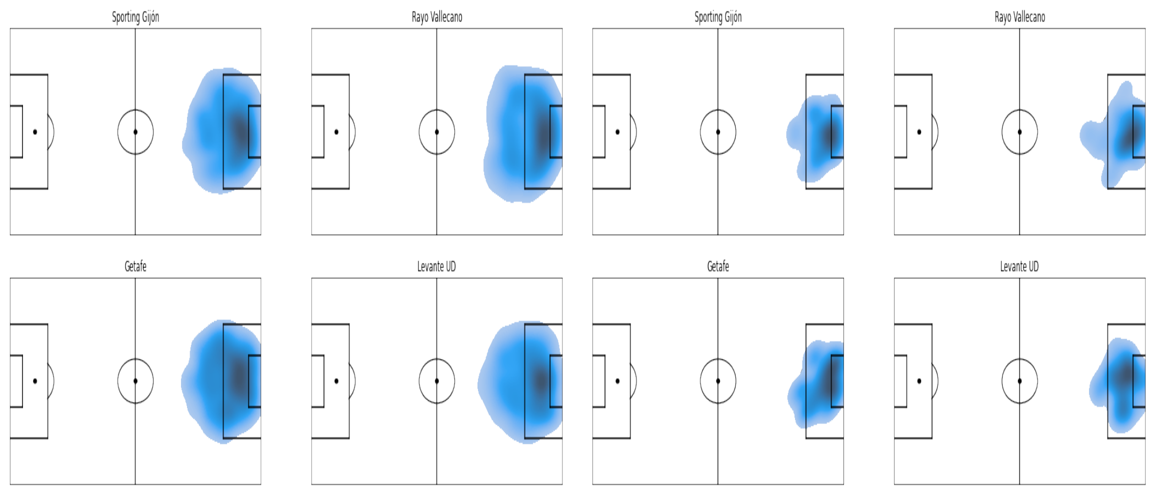
5.4 Shot Location of teams

5.4.1 Shots on Goal. Shots on goal represent the attempts a team makes towards the opponent's goal, indicating their attacking intent.



(a) Shot Heatmap of Top-4 teams

(b) Goal Heatmap of Top-4 teams



(c) Shot Heatmap of Bottom-4 teams

(d) Goal Heatmap of Bottom-4 teams

Fig. 8. Heatmap of Shot and Goals (Shots which resulted in Goals)

- **Barcelona, Atletico Madrid, Villarreal, and Real Madrid (Top-4 Teams):** The concentration of attacks near the center of the field aligns with their strategic approach, as reflected in their consistently high shot counts. The emphasis on concentrated attacks may contribute to a higher frequency of attempts on goal, putting additional pressure on opposing defenses.
- **Sporting Gijon, Rayo Vallecano, Getafe, and Levante UD (Bottom-4 Teams):** In contrast, the less successful teams tend to have lower shots on goal counts. This could indicate a less aggressive or less focused offensive strategy. It's possible that these teams face challenges in breaking through strong defensive setups, leading to shots being comparatively less concentrated at any particular location of the pitch.

5.4.2 Successful Shots on Goal. Successful shots on goal represent the subset of shots that result in goals. This metric provides a more refined measure of a team's effectiveness in converting goal-scoring opportunities into actual goals.

- **Barcelona, Atletico Madrid, Villarreal, and Real Madrid (Top-5 Teams):** The Top team although have highly concentrated density near the center of the Penalty box, presence of successful shots from far away shows the higher quality of the players playing for these teams.
- **Sporting Gijon, Rayo Vallecano, Getafe, and Levante UD (Bottom-5 Teams):** Albeit having a similar number of goals from far away, the heatmap shows that these teams are unsuccessful in converting those opportunities into goals. These heatmaps not only show the difference in player's strengths in different teams but also the difference in player quality of top teams v/s bottom teams.

5.5 Conclusion and Future Work

In conclusion, our comparative analysis of playing styles across football leagues raises questions about the myth of home advantage, delves into the nuanced performance trends of Barcelona over time, and scrutinizes shot locations of teams. Moving forward, improving visualization techniques and incorporating real-time analysis can enhance our understanding of evolving playing dynamics. Access to more public data is crucial, as it would significantly contribute to a more comprehensive and insightful exploration of football performance across leagues.

REFERENCES

[1] [n. d.]. StatsBomb Soccer Data. https://statsbomb.com/what-we-do/soccer-data/. Accessed: November 20, 2023.

[2] Zhongbo Bai and Xiaomei Bai. 2021. Sports Big Data: Management, Analysis, Applications, and Challenges. *Complexity* 2021 (2021), 6676297.

[3] Tom Decroos, Lotte Bransen, Jan Van Haaren, and Jesse Davis. 2019. Actions Speak Louder than Goals: Valuing Player Actions in Soccer. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 1851–1861.

[4] F. Fonti, J.-M. Ross, and P. Aversa. 2023. Using Sports Data to Advance Management Research: A Review and a Guide for Future Studies. *Journal of Management* 49, 1 (2023), 325–362.

[5] G. Foster, N. O'Reilly, and Z. Naidu. 2021. Playing-Side Analytics in Team Sports: Multiple Directions, Opportunities, and Challenges. *Frontiers in Sports and Active Living* 3 (2021).

[6] A. García-Aliaga, M. Marquina, J. Coterón, A. Rodríguez-González, and S. Luengo-Sánchez. 2021. In-game behaviour analysis of football players using machine learning techniques based on player statistics. *International Journal of Sports Science & Coaching* 16, 1 (2021), 148–157.

[7] F.R. Goes, L.A. Meerhoff, M.J.O. Bueno, D.M. Rodrigues, F.A. Moura, M.S. Brink, M.T. Elferink-Gemser, A.J. Knobbe, S.A. Cunha, R.S. Torres, and K.A.P.M. Lemmink. 2021. Unlocking the potential of big data to support tactical performance analysis in professional soccer: A systematic review. *European Journal of Sport Science* 21, 4 (2021), 481–496.

[8] M. Herold, M. Kempe, L. Ruf, L. Guevara, and T. Meyer. 2022. Shortcomings of applying data science to improve professional football performance: Takeaways from a pilot intervention study. *Frontiers in Sports and Active Living* 4 (2022).