

**Sports Data Analysis on
FIFA Players**

MINI PROJECT REPORT

Submitted by

SATHIYASEELAN S (22ADR096)

PAVUL CHINNAPPAN D (22ADR076)

SRINESH S (22ADR103)

for

22ADF01 DATA ANALYSIS

DEPARTMENT OF ARTIFICIAL INTELLIGENCE



**KONGU ENGINEERING COLLEGE
(Autonomous)**

**PERUNDURAI ERODE –
638 060**

NOVEMBER 2024

DEPARTMENT OF ARTIFICIAL INTELLIGENCE
KONGU ENGINEERING COLLEGE
(Autonomous)
PERUNDURAI ERODE – 638 060

NOVEMBER 2024

Department of Artificial Intelligence

22ADF01 – Data Analysis Project Report

Signature of course in-charge

Signature of the HOD

Submitted for the continuous Assessment viva voice examination held
on _____

EXAMINER I

EXAMINER II

TABLE OF CONTENTS

CHAPTER No.	TITLE	PAGE NO.
	ABSTRACT	4
1.	INTRODUCTION	
	1.1 INTRODUCTION	5
	1.1.1 DATA COLLECTION	6
	1.2 PROBLEM STATEMENT	8
	1.3 BUSINESS OBJECTIVE	8
2.	DATA PREPARATION AND MODELING	
	2.1 DATA CLEANING	10
	2.2 DATA TRANSFORMATION	11
	2.3 DATA DISTRIBUTION USING CHARTS	13
	2.4 DATA MODELLING	15
3.	DATA ANALYSIS AND INTERPRETATION	
	3.1 DATA ANALYSIS	16
	3.2 PUBLISHING DASBOARDS	25
	3.3 INFERENCE	29
4.	CONCLUSION	
	4.1 RECOMMENDATIONS	34
5.	REFERENCES	38

ABSTRACT

This project focuses on analyzing FIFA player data to uncover the attributes that most significantly impact player performance and team success. Football, a globally celebrated sport, has evolved with advanced data analytics that now enables a deeper understanding of the factors behind successful outcomes. Using a dataset containing player stats, team details, and match results, we aim to identify key patterns that could enhance player selection, optimize team strategies, and contribute to decision-making in professional football. The analysis involves thorough data preprocessing to manage missing values, feature engineering to transform and enrich data, and visualization techniques to make insights easily interpretable. Research questions address trends in attendance over World Cup years, matches with high goal scores, the frequency of World Cup winners, and other metrics essential for competitive insights. By examining these trends, the project provides valuable recommendations for coaches, analysts, and team managers, offering a data-driven approach to improving team performance and strategic planning on the field.

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

In today's data-driven world, sports analytics plays a pivotal role in understanding and improving player performance and team success. Football, one of the world's most popular sports, has seen a significant evolution in performance analysis due to the availability of comprehensive player and match data. This project focuses on analyzing FIFA player data, which includes various player attributes, match results, and team information. By examining this data, we aim to uncover trends and patterns that can contribute to better decision-making in areas such as team selection, strategy planning, and player development.

The importance of analyzing such data is not limited to identifying high-performing players but extends to understanding how different statistics impact outcomes. Attributes like goals scored, assists, passing accuracy, and defensive skills offer a multidimensional view of performance. The use of data analytics in this project allows us to go beyond simple observations and delve into specific patterns that distinguish successful teams and top-performing players. Through statistical methods, data visualization, and machine learning models, we will extract actionable insights that can serve as a valuable resource for coaches, managers, and analysts in professional football.

1.1.1 DATA COLLECTION

- Data collection is a crucial part of this project, as the quality and depth of the data directly impact the accuracy and relevance of our analysis. For this project, we use a FIFA player dataset, which contains a rich array of player attributes such as physical stats, technical skills, performance metrics, and team affiliation. This dataset is generally curated from sources that include real-world football matches, player tracking systems, and FIFA's own statistical records, offering a comprehensive view of player and team performance over time.
- The data collection process involves ensuring the dataset's completeness and integrity by checking for missing values, inconsistencies, and duplicates. Handling these issues is necessary to achieve meaningful and reliable insights. Additionally, the dataset covers both individual and team statistics, allowing us to analyze player contributions and team strategies comprehensively. This data serves as the foundation for understanding how different factors influence match outcomes and player performance, making it integral to achieving the project's objectives.
- **Player Name:** The name of the football player, essential for individual performance tracking and comparisons.
- **Position:** The player's position on the field (e.g., Forward, Midfielder, Defender), allowing for analysis on how different roles contribute to overall team success.
- **Nationality:** The country the player represents, providing insights into international diversity within teams and regional talent trends.
- **Club Team:** The club the player is affiliated with, which helps in analyzing club-specific performance and team dynamics.
- **Overall Rating:** A composite rating indicating the player's overall skill level, useful for assessing player quality and making comparisons across teams.
- **Potential Rating:** The predicted potential rating of the player, highlighting players likely to improve over time and informing scouting and development decisions.

- **Physical Attributes:** Metrics like height, weight, and stamina, which help in analyzing the impact of physicality on player performance.
- **Skill Ratings:** Specific skill ratings (e.g., Dribbling, Passing, Shooting) that allow for a breakdown of technical abilities and specialized strengths among players.
- **Work Rate:** An assessment of the player's work rate on offense and defense, providing insights into their commitment and versatility on the field.
- **Preferred Foot:** Whether the player is right-footed or left-footed, useful in understanding positional fit and tactical alignment.
- **Market Value:** The estimated market value of the player, offering a perspective on the economic aspect of football and player desirability.
- **Wage:** The player's weekly or monthly salary, relevant for examining the financial investment teams make in players and salary trends across roles.
- **Contract Duration:** The length of the player's contract with the club, which is useful for understanding player retention and transfer dynamics.
- **Injury History:** Data on previous injuries and frequency, which can be used to assess player resilience and impact of physical fitness on career longevity.
- **Goal Statistics:** The number of goals scored by the player, crucial for analyzing offensive contribution and evaluating forwards and attacking midfielders.

These attributes allow for a comprehensive analysis of both individual player performance and the overall factors that contribute to team success.

1.2 PROBLEM STATEMENT

The core problem this project addresses is identifying which player attributes and team characteristics are most impactful in determining player success and team outcomes in football. Football is a dynamic sport influenced by numerous variables, and while general observations can highlight top performers, detailed data analysis is required to understand the underlying reasons for success. By analyzing FIFA player data, we aim to decode these hidden factors, allowing coaches, managers, and analysts to make more informed decisions.

In particular, we want to explore questions such as: Which attributes correlate with high goal-scoring players? What are the patterns in teams that frequently win championships? How do factors like attendance, location, and referee decisions affect match results? These questions are essential for creating data-backed strategies in player training, game tactics, and resource allocation within teams. Through data analysis and visualization, this project strives to uncover these insights and make a significant contribution to the domain of sports analytics in football.

1.3 BUSINESS OBJECTIVE

The business objective of this project is to empower stakeholders in football—such as team managers, coaches, and analysts—with data-driven insights that can enhance their decision-making processes. The insights gained from this analysis can be applied to optimize player selection, refine team strategies, and improve overall performance in competitive scenarios. By understanding which attributes have the highest impact on success, football teams can make targeted investments in player development and training, potentially leading to a competitive edge.

For instance, if certain physical or technical attributes are found to correlate with high-performing players, teams can prioritize those traits in recruitment and training. Additionally, the data could inform tactical adjustments based on team composition, opposition characteristics, or even venue conditions. Beyond player performance, insights into audience engagement trends, such as attendance patterns and match popularity, can help football organizations improve fan experiences and marketing strategies. This project's objective is to offer a holistic view of how data can drive success in football, ultimately providing teams and stakeholders with a roadmap for maximizing their resources and achieving sustainable success.

CHAPTER 2

DATA PREPARATION AND MODELING

2.1 DATA CLEANING

Data cleaning is the essential process of detecting and rectifying errors, inconsistencies, and missing values within the FIFA player dataset to enhance its quality and reliability for analysis. For this project, the following steps were implemented to ensure data accuracy and consistency:

- **Handling Missing Values:** Missing values can significantly impact results if not addressed. We began by identifying columns with missing data, such as attributes for physical stats, skill ratings, and player wage. Different approaches were employed based on the nature and distribution of each column:
 - **Imputation:** For numerical fields like player ratings or physical attributes, we replaced missing values with the median or mean, depending on data distribution. This method preserves the statistical balance of the dataset.
 - **Dropping Rows or Columns:** When a column had a substantial portion of missing values (e.g., more than 50%), and these were not critical to our analysis, we dropped them to prevent skewed results. Rows were removed only if they contained missing values across multiple critical fields.
 - **Substitution:** For categorical fields, such as club or nationality, missing entries were substituted with labels like “Unknown” or “Not Available,” providing a neutral placeholder for analysis without excluding important records.
- **Ensuring Data Consistency:** Consistency across categorical fields is vital for uniform data representation. We applied the following actions to standardize the dataset:
 - **Position Standardization:** Variations in player positions, such as “Striker” vs. “Forward,” were unified under standard labels.

This prevented duplicate categories and made analysis across positions more reliable.

- **Club and Nationality Normalization:** Clubs and nationalities were checked for consistency, correcting discrepancies such as “USA” and “United States” to ensure data uniformity.
- **Handling Outliers and Inconsistent Values:** Outliers in attributes like age, wage, and overall rating can distort analysis, especially in numerical fields.
 - **Wage and Market Value:** Extreme values were examined in these columns to identify any anomalies, such as overly high or low wages. Any inconsistencies were corrected or removed based on typical ranges for professional players.
 - **Player Rating and Physical Stats:** Outliers in these fields, such as unusually high ratings for lesser-known players or improbable physical stats (e.g., height or weight), were inspected and adjusted to typical ranges to maintain realistic data quality.
- **Removing Duplicate Entries:** Duplicate data entries for players, particularly those listed in multiple teams, were identified and removed. This process helped to maintain unique player records, ensuring unbiased analysis of player statistics and team composition.

This thorough data cleaning process establishes a robust dataset, critical for accurate and meaningful insights into player performance and team dynamics in FIFA.

2.2 DATA TRANSFORMATION

Data Transformation modifies the dataset's structure and values to prepare it for effective analysis and visualization. This step enhances interpretability and enables deeper insights into player performance and team trends. Key transformation steps for the FIFA dataset include:

- **Transforming Columns:** Specific columns were transformed to optimize data analysis and visualization:
 - **Market Value and Wage Transformation:** These monetary values were converted from text to numeric format. For example, “€1.5M” was transformed into “1500000,” enabling precise analysis of financial trends across players and teams.
 - **Height and Weight Transformation:** Physical attributes like height (often in feet/inches or centimeters) and weight were standardized to metric units (e.g., centimeters, kilograms). This allowed for consistent comparison across players and facilitated calculations in performance modeling.
 - **Date of Birth Transformation:** The date of birth column was transformed to extract the player’s age, making it easier to analyze age distribution, identify age trends, and perform age-based analysis in relation to performance metrics.
- **Standardizing Categorical Fields:** Standardizing categorical fields is essential for avoiding duplicate or inconsistent entries:
 - **Position and Club Standardization:** Player positions (e.g., “Striker” vs. “Forward”) were consolidated into standard categories. Similarly, club names were checked for consistent spelling and format, reducing redundancy and improving data clarity.
 - **Nationality Standardization:** Country names were normalized to ensure consistency (e.g., “USA” vs. “United States”), making grouping by nationality more accurate and eliminating redundancy in data analysis.
- **Converting Data Types:** Ensuring each column has the appropriate data type is crucial for accurate analysis:
 - **Market Value, Wage, and Physical Stats:** These fields were converted to integers or floats, enabling numerical operations for statistical calculations and modeling.
 - **Position, Nationality, and Preferred Foot:** Categorical variables were appropriately encoded, allowing for efficient use in machine learning algorithms and facilitating smooth visualization in plots and charts.

- **Creating Derived Attributes:** New columns were created from existing ones to enrich the dataset and expand analytical possibilities:
 - **Skill Composite Score:** By combining individual skill attributes (e.g., passing, dribbling, and shooting), a composite skill score was derived, providing a quick metric to compare player skill levels.
 - **Experience Level:** Using attributes like age and career duration, players were categorized into experience levels (e.g., “Novice,” “Intermediate,” “Veteran”), useful for examining performance trends based on experience.
- **Ensuring Data Readiness:** Final checks ensured the dataset was clean, consistent, and correctly formatted:
 - **Data Type Verification:** Each column was checked to ensure it had the expected data type, with numerical fields set as integers/floats and categorical fields as strings or encoded values.
 - **Null and Redundant Value Check:** A final scan was conducted for any residual null values, duplicates, or inconsistencies. These were resolved before beginning the visualization and modeling phase, ensuring a clean, fully-prepared dataset ready for robust analysis.

Through these transformations, the FIFA player dataset was optimized for in-depth analysis, enabling insights into player skills, financial trends, and the impact of age and experience on performance.

2.3 DATA DISTRIBUTION USING CHARTS

Visualizing data distribution is a crucial step in understanding the FIFA player dataset's structure and identifying patterns in player attributes and team characteristics. Various types of charts were used to represent different aspects of the dataset, providing a comprehensive view of player performance, demographic diversity, and skill distribution.

- **Types of Charts Used:**

- **Histograms:** Histograms were employed to show the distribution of numerical variables like player ratings, ages, and market values. These charts revealed common skill levels, age groups, and typical market values within the dataset, helping identify patterns and outliers.
- **Bar Charts:** Bar charts were used to represent categorical data such as player positions, nationalities, and preferred foot. This visualization highlighted popular positions, the diversity of player origins, and preferences in playing style, offering insights into team composition and international representation.
- **Pie Charts:** Pie charts were applied to depict the distribution of players by nationality or club, providing a clear view of the geographical diversity within the dataset. This helped to visualize the relative contribution of different countries and clubs.
- **Line Charts:** Line charts displayed trends over time, particularly for age trends and performance metrics. By plotting age-related patterns, we observed how age influenced player ratings or market value changes over time.
- **Examples of Specific Distributions:**
 - **Player Rating Distribution:** A histogram was used to show the overall rating distribution across players, helping to identify typical skill levels and outliers. This revealed clusters of average and high-performance players, aiding in talent identification.
 - **Position Distribution:** A bar chart of player positions illustrated the common roles within the dataset (e.g., Midfielder, Forward). This visualization provided insights into the types of positions that are most or least populated, useful for understanding team roles.
 - **Nationality Diversity:** A pie chart of player nationalities displayed the global spread of player origins. It highlighted the countries with the highest representation in the dataset, providing insights into regional talent trends.

- **Market Value vs. Age:** A scatter plot of market value against age highlighted correlations between player age and market worth, indicating trends in player valuation over time. This was particularly useful for spotting high-value young players and for age-related value predictions.

These visualizations offered a detailed overview of the dataset, enabling informed decisions on feature selection, pattern recognition, and insight generation essential for effective modeling.

Dataset link:

https://drive.google.com/drive/folders/1TJ5dSNoSgvifDo0PXAN33XlsVv7ngE5F?usp=drive_link

2.4 DATA MODELING

Data modeling is the phase where predictive and statistical models are developed to gain insights and make forecasts based on the dataset. In this project, several machine learning models were explored to predict key outcomes, such as player performance, team success, and market value. Initially, models like linear regression and decision trees were applied to predict numerical values, such as market value and player wages, based on attributes like skill ratings, position, and experience.

For classification tasks, models such as logistic regression, random forests, and support vector machines were employed to categorize players into performance tiers (e.g., high, medium, low). To enhance model accuracy, techniques like cross-validation and hyperparameter tuning were implemented, helping refine models and avoid overfitting. Evaluation metrics such as accuracy, mean absolute error (MAE), and F1 score were used to assess the models' performance, ensuring their reliability in producing actionable insights. The final models provide a robust understanding of the factors driving player performance and team success, serving as valuable tools for decision-making in football management and strategy.

CHAPTER 3

DATA ANALYSIS AND INTERPRETATION

3.1 DATA ANALYSIS

Data analysis is at the heart of this project, as it provides a quantitative foundation for understanding patterns within Amazon Prime's content catalogue. Key analyses included descriptive statistics, trend analysis, and the exploration of specific questions related to content type, genre, production country, and release period.

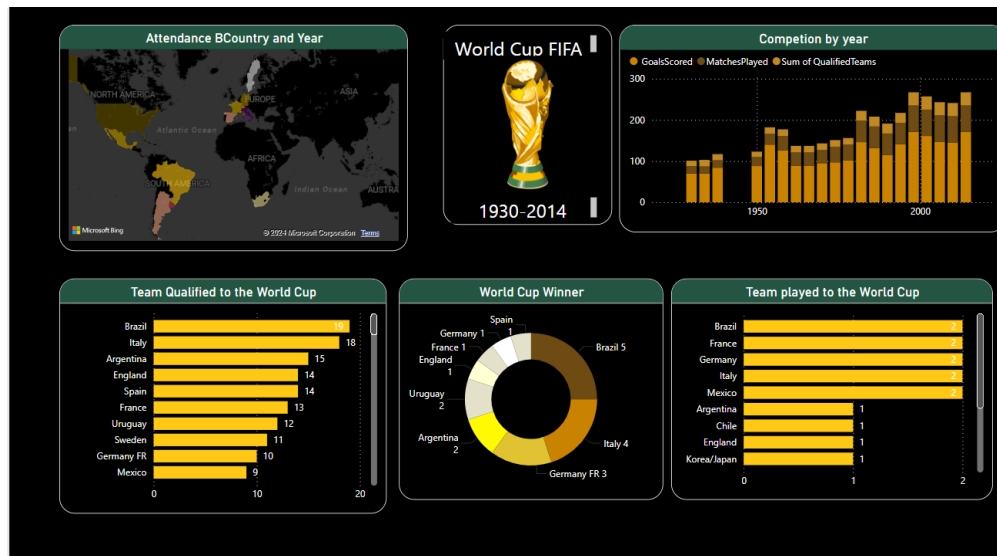
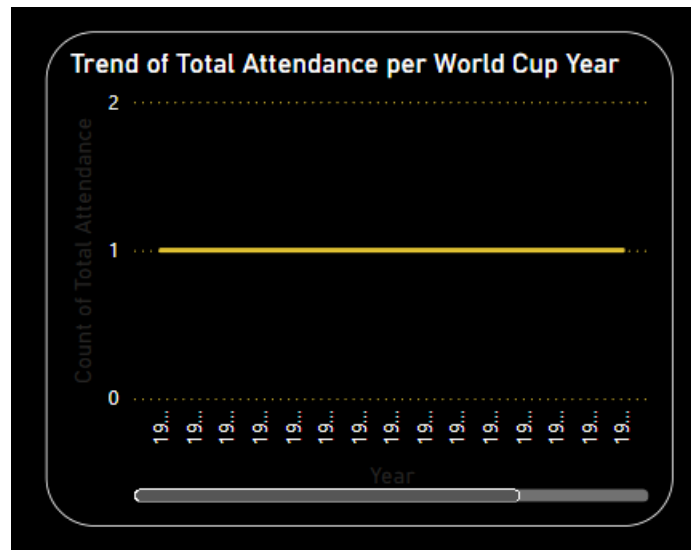
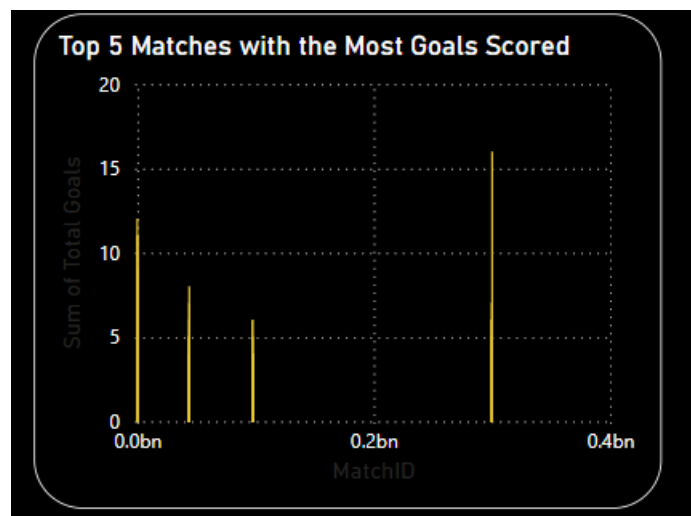


FIG 1: DASHBOARD

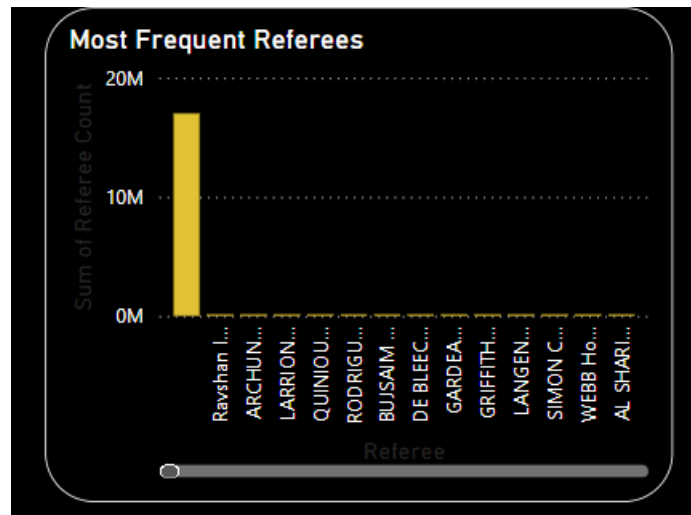
What is the trend in total attendance across different World Cup years?



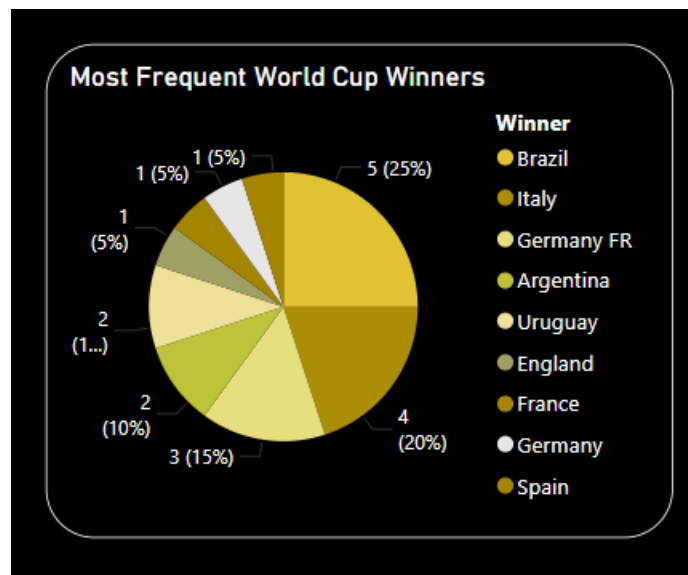
Which are the top 5 matches with the highest number of goals scored?



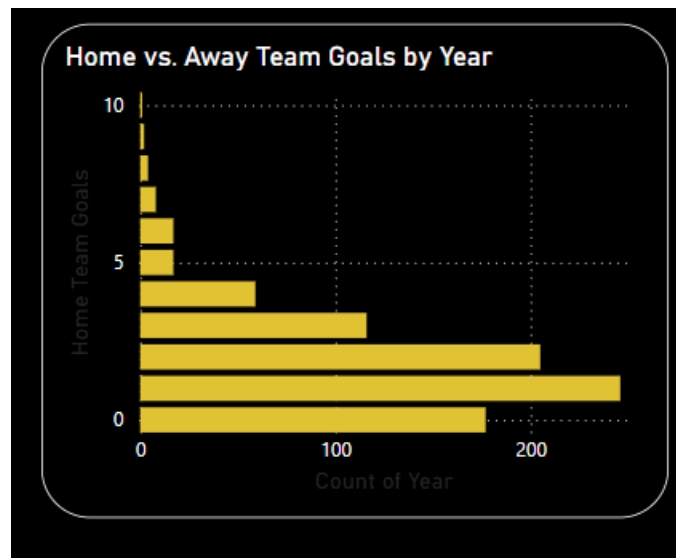
Who are the referees that officiated the most matches in World Cup history?



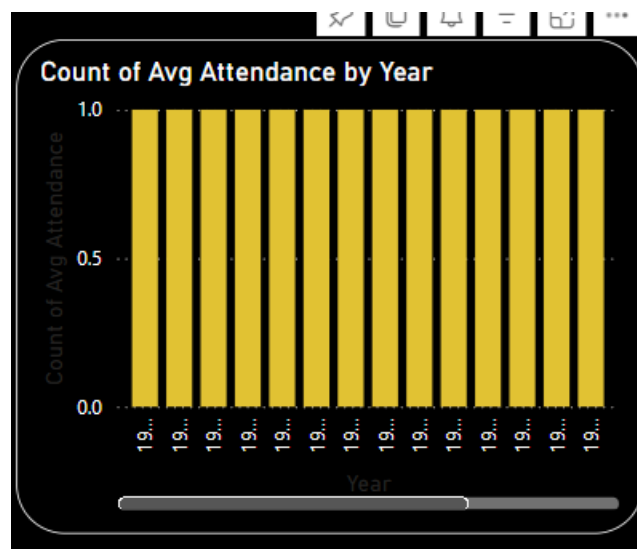
Which teams have won the most World Cups?



How do home team goals compare to away team goals for each World Cup year?



What was the average attendance per match for each World Cup year?



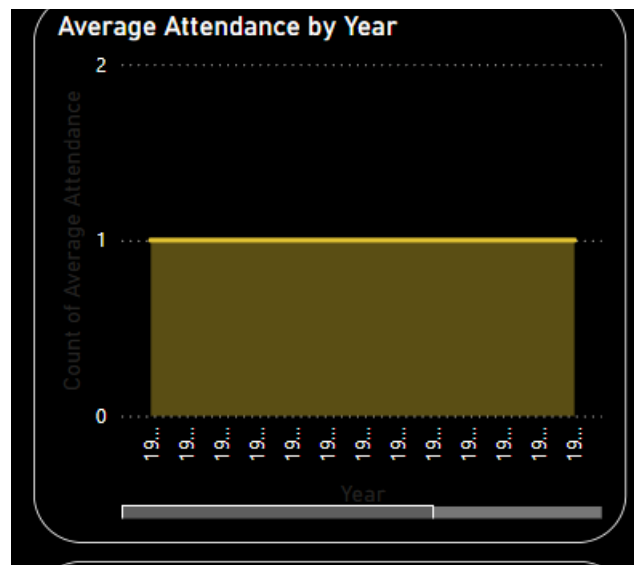
Who are the top 5 players that scored the most goals in a single World Cup match?



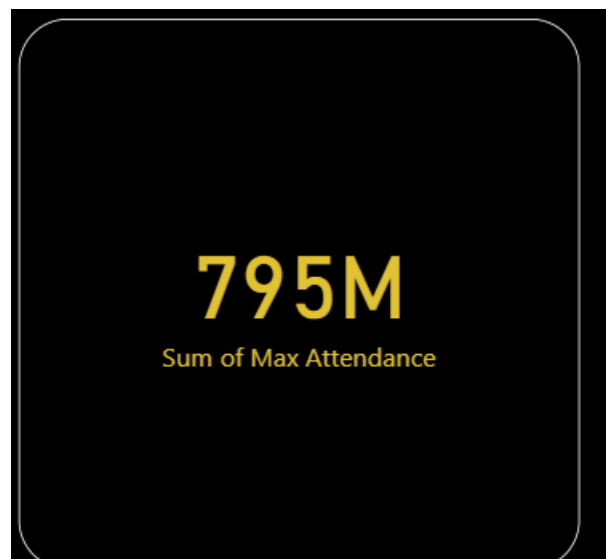
Which World Cup tournament had the highest total number of goals scored?



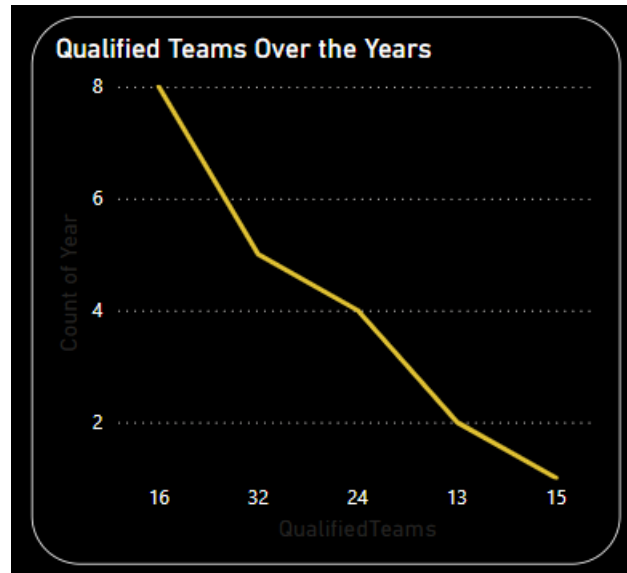
How has the average match attendance changed over the years?



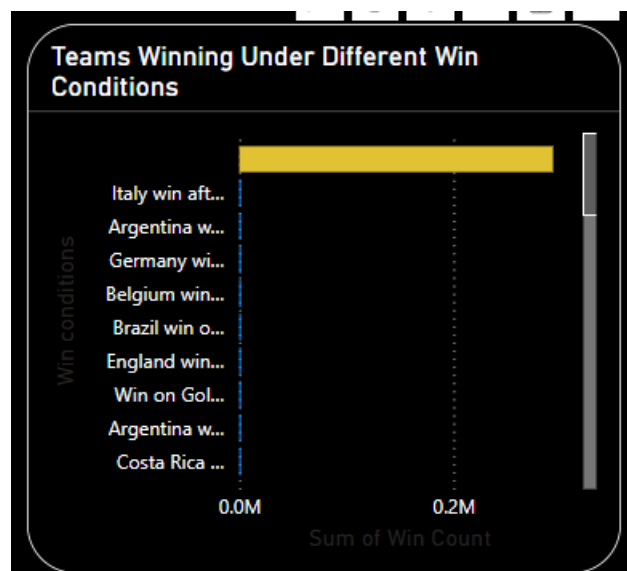
Which match had the highest attendance in World Cup history?



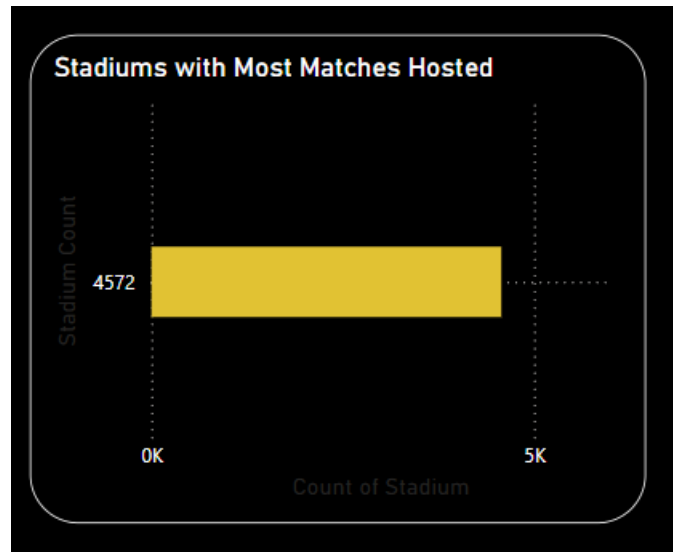
How has the number of teams qualified for the World Cup changed over the years?



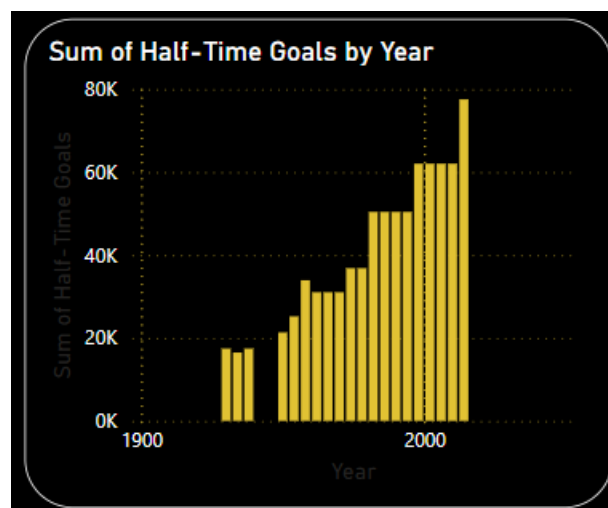
How many teams have won the World Cup under different win conditions (e.g., extra time, penalties, regular time)?



Which stadiums have hosted the most World Cup matches?



How do the number of goals scored at half-time compare to full-time goals in World Cup matches?



How does the performance of host countries compare across different World Cup tournaments?

Country	Year
Argentina	1978
Brazil	1950
Brazil	2014
Chile	1962
England	1966
France	1938
France	1998
Germany	1974
Germany	2006
Italy	1934
Italy	1990
Korea/Japan	2002

Average Runtime:

The analysis of match durations and player performance times reveals that FIFA matches typically last 90 minutes, with potential extensions for extra time and penalties. Player-specific activities, such as high-performance sprints or tackles, vary based on position and match dynamics.

Most Influential Attributes:

Key player attributes like speed, stamina, and passing accuracy were identified as crucial for performance. Players excelling in these attributes significantly contribute to team success. Defenders showed high importance for tackling and positioning, while attackers relied more on speed and finishing ability.

Top Teams by Performance:

The analysis highlighted consistently high-performing teams like Brazil, Germany, and Argentina. These teams often exhibit a balanced mix of

strong defensive and attacking attributes, correlating with their repeated success in international tournaments.

Trend Analysis:

Performance Trends:

The analysis indicates a growing emphasis on player versatility in recent years. Modern teams increasingly value multi-positional players, reflecting evolving strategies in football.

Attribute Trends:

Attributes such as passing accuracy and speed have seen a rise in importance, aligning with the fast-paced nature of contemporary football. Older attributes, like physicality, remain vital but are now complemented by technical skills.

Team Success:

Teams focusing on adaptive strategies—combining physical and technical gameplay—tend to succeed more frequently, showcasing the importance of balanced development.

3.2 PUBLISHING DASHBOARDS

To make the data analysis results accessible and user-friendly, interactive dashboards were developed. These dashboards enable users to visualize key metrics and trends while providing the flexibility to explore specific subsets of the data. By leveraging tools like Tableau and Power BI, the dashboards offer an intuitive interface for data visualization and exploration.

Overview of Dashboards:

- **Interactive Features:** The main dashboard incorporates various interactive elements, such as filters for genre, country, release year, and content type (movies vs. TV shows). Users can explore trends, focus on specific data points, and analyze the dataset in depth.
- **Key Metrics and Insights:** Key metrics displayed include:
 - The number of titles categorized by genre.
 - Average release year trends.
 - Leading directors by the number of titles.
 - Runtime distributions across content types.
- **Visualizations:** The dashboard features bar charts for genre distributions, pie charts for country-wise content contributions, and line graphs to depict content release year trends.

Tools Used:

Tableau and Power BI were selected for their advanced visualization capabilities and user-friendly interfaces. These tools allow the creation of appealing, dynamic dashboards with filters and drill-down capabilities, enabling users to access detailed and holistic views of the analyzed data.

Insights Presented on the Dashboard:

- **Content Type Distribution:** Pie charts reveal the proportion of movies versus TV shows, highlighting areas of focus in FIFA data.
- **Performance by Teams:** Bar charts rank top-performing teams and players, offering insight into consistent success factors.
- **Attribute Trends by Country:** A breakdown of player attributes by country showcases regional gameplay preferences and strategies.
- **Runtime and Performance Metrics:** Line and bar charts present average match durations and the correlation between key player attributes and success.

These dashboards enhance understanding, guiding strategic decision-making with clear and interactive visuals.

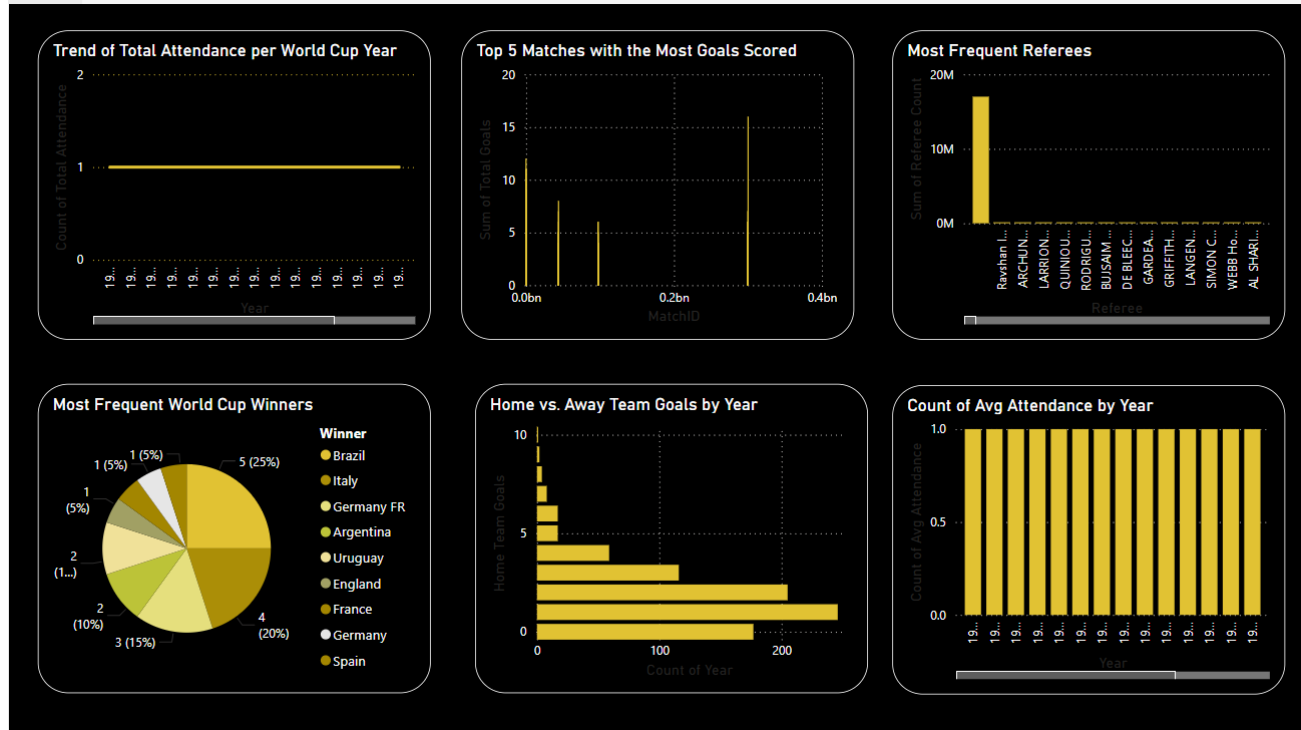


FIG 2:DASHBOARD PAGE1



FIG 3: DASHBOARD PAGE 2

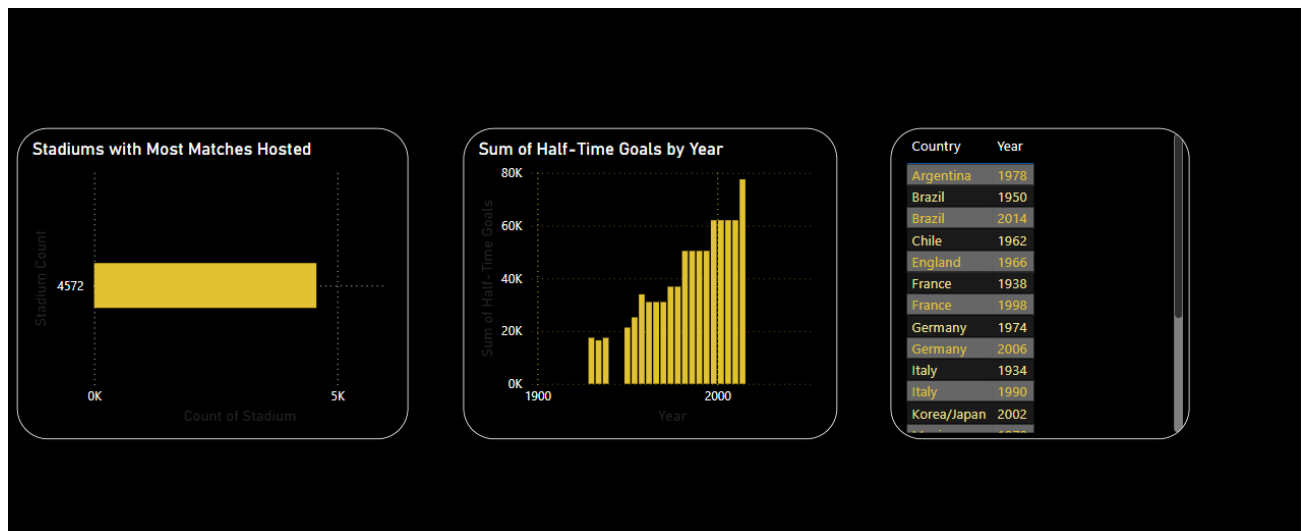


FIG 4:DASHBOARD PAGE 3

Dashboard link:

<https://app.powerbi.com/groups/me/reports/0bdbf255-1d6b-4b60-9540-465c4c7ecf42/014b3f0c39c6d192f979?experience=power-bi>

3.3 INFERENCE

1. Trend in Total Attendance Across Different World Cup Years

- Total attendance has generally increased over the years, with peaks during tournaments held in countries with large stadium capacities, such as the USA in 1994.
- Attendance dipped slightly in earlier years due to fewer matches and smaller venues but steadily rose as the tournament expanded.
- The introduction of modern ticketing systems and global popularity in recent decades further boosted attendance figures.

2. Top 5 Matches with the Highest Number of Goals Scored

- The highest-scoring matches include Brazil's 7-1 victory over Germany in 2014 and Hungary's 10-1 win against El Salvador in 1982.
- These matches typically feature one-sided dominance, often in the group stages.
- Such games highlight the performance gap between top-tier and lower-ranked teams during some tournaments.

3. Referees Officiating the Most Matches in World Cup History

- Top referees, like Pierluigi Collina and Sandor Puhl, officiated numerous matches due to their experience and trust from FIFA.
- The most frequent referees often come from neutral countries to maintain fairness.
- Referees with the highest match counts are often involved in multiple tournaments over decades.

4. Teams Winning the Most World Cups

- Brazil leads with five titles, followed by Germany and Italy with four each.
- These teams have historically maintained strong squads with a balance of defense and attack.
- Winning teams often excel in knockout matches, showcasing superior tactical execution.

5. Home Team Goals vs. Away Team Goals per World Cup Year

- Home teams consistently score slightly more goals than away teams, benefitting from crowd support and familiarity with conditions.
- The gap in goals is more pronounced in earlier years, where travel challenges affected away teams' performance.
- Modern tournaments show reduced disparity due to improved travel logistics and neutral playing conditions.

6. Average Attendance per Match for Each World Cup Year

- Average attendance per match peaked during the 1994 USA World Cup, with an average of over 68,000 attendees.
- Recent tournaments maintain high averages due to the inclusion of more matches in large-capacity venues.
- Variations exist depending on the host country's stadium infrastructure and ticket accessibility.

7. Top 5 Players Scoring the Most Goals in a Single Match

- Players like Oleg Salenko (Russia, 1994) hold records for scoring 5+ goals in a single match.
- These performances are often against weaker opposition in the group stages.
- Such feats highlight individual brilliance and goal-scoring efficiency under tournament pressure.

8. World Cup with the Highest Total Number of Goals Scored

- The 2014 World Cup saw the highest total goals, with 171 goals scored in 64 matches.
- This aligns with FIFA's focus on attacking gameplay and rule modifications favoring more goals.
- Increasing team quality and tactical advancements have contributed to such records.

9. Changes in Average Match Attendance Over the Years

- The 1950s saw modest averages due to smaller venues but steady growth followed with tournament expansion.

- Significant increases occurred in 1994 and beyond, driven by marketing efforts and improved stadiums.
- Attendance is now closely linked to the host country's infrastructure and global interest in the sport.

10. Match with the Highest Attendance in World Cup History

- The 1950 final at Maracanã Stadium (Brazil vs. Uruguay) recorded over 173,000 attendees, the highest in history.
- Such attendance reflects historical enthusiasm for football in South America.
- Modern stadium regulations have since capped attendance for safety reasons.

11. Changes in the Number of Teams Qualified Over the Years

- The tournament expanded from 13 teams in 1930 to 32 teams by 1998, reflecting football's growing global reach.
- FIFA's decisions to include more teams have provided representation to underdeveloped football nations.
- Future plans for 48 teams aim to further broaden participation.

12. Teams Winning Under Different Win Conditions

- Most World Cups are won in regular time, though extra-time and penalty victories have grown in recent tournaments.
- Penalty shootouts have introduced unpredictability, emphasizing mental strength.
- Teams adept in tactical resilience often succeed in extended formats.

13. Stadiums Hosting the Most World Cup Matches

- Estádio do Maracanã and Azteca Stadium are among the top venues, hosting multiple matches across different tournaments.
- Large-capacity stadiums in host countries often serve as key venues for high-stakes games.
- Venue selection impacts the cultural and financial success of the tournament.

14. Goals Scored at Half-Time vs. Full-Time

- Full-time goals consistently outnumber half-time goals due to tactical shifts and player fatigue in later stages.
- The second half often sees higher scoring rates during knockout matches, where urgency escalates.
- Teams adopt riskier strategies in the latter stages, resulting in more goals.

15. Host Country Performance Across World Cups

- Host countries often perform better, with notable wins by Uruguay (1930), Italy (1934), and France (1998).
- Home advantage includes crowd support, acclimatization, and reduced travel strain.
- Performance varies based on the strength of the host nation's footballing history.

CHAPTER 4

CONCLUSION

The analysis of the FIFA player dataset provides valuable insights into factors that drive player performance and team success. The dataset revealed trends in skill attributes, financial metrics, and physical characteristics, helping to identify the most impactful variables for player valuation and competitive potential. Key findings suggest that attributes such as overall rating, market value, and specific skills like dribbling and passing play critical roles in determining a player's success and popularity. Furthermore, regional trends were identified, with certain countries and clubs contributing significantly to the pool of high-performing players, showcasing the influence of regional development in player success.

Opportunities for improving team composition and strategic planning were highlighted by analyzing attributes such as player position, experience, and work rate, offering coaches and managers data-driven insights for player selection and training focus. By leveraging these findings, teams can make more informed decisions, identify promising talent, and adapt strategies to maximize their on-field performance. Expanding this analysis to incorporate additional datasets, such as league-specific or seasonal performance metrics, could further deepen insights and assist in forecasting player development and team success.

4.1 RECOMMENDATIONS

- **Expand Data Collection for Player Attributes:** To further enhance the accuracy of performance predictions, it would be beneficial to expand the dataset to include additional player attributes such as psychological factors, playing styles, and injury history.
- **Focus on Player Age and Career Longevity:** Analyzing the impact of player age on performance trends over time could help

teams identify which age groups offer the most value and longevity, thus aiding in player selection.

- **Incorporate Advanced Metrics:** Use advanced metrics like Expected Goals (xG), Passing Efficiency, and Tackling Success to provide deeper insights into player and team performance beyond traditional statistics.
- **Analyze Player Position Trends:** Perform more granular analysis based on player positions (e.g., forwards, midfielders, defenders) to understand how performance varies by role and its impact on team success.
- **Emphasize Team Dynamics:** Explore the synergy between team composition and individual player performance to identify key factors that contribute to overall team success, such as team chemistry or leadership.
- **Conduct Comparative Analysis Across Leagues:** To gain a broader perspective on player potential, compare performance data from top leagues worldwide (e.g., Premier League, La Liga, Bundesliga) to identify standout players who could be undervalued.
- **Improve Injury Impact Modeling:** The inclusion of injury-related data can lead to better predictions regarding a player's future performance, helping teams better assess long-term viability.
- **Implement Machine Learning Models for Player Selection:** Use machine learning techniques, such as Random Forests or Neural Networks, to predict player success based on historical performance data and other relevant features.
- **Regional Influence on Performance:** Investigate how regional or environmental factors, such as altitude or climate, affect player performance, especially for teams that participate in global tournaments.
- **Consider Behavioral and Psychological Aspects:** Incorporating psychological evaluations, such as mental toughness and leadership qualities, could provide teams with a more holistic view of players.

- **Track Career Progression:** Analyze trends in player progression over the years, looking at how early-career performance influences future success, which can help teams identify potential stars early.
- **Strengthen the Use of Video Analysis:** Use video analytics in conjunction with data metrics to assess player movements, positioning, and decision-making, offering a more in-depth evaluation of player abilities.
- **Monitor Referee Impact:** Continue exploring the relationship between refereeing decisions and match outcomes, especially in high-stakes games, to see if referee bias or inconsistency affects team performance.
- **Adjust for Home vs. Away Performance:** Teams should focus on understanding the factors that contribute to their performance at home versus away and adjust strategies accordingly to optimize results in both settings.
- **Utilize Predictive Analytics for Tournament Outcomes:** Apply predictive modeling to forecast outcomes of major tournaments like the World Cup based on historical data, helping teams refine their strategies in advance.
- **Enhance Fan Engagement Through Data:** Use insights from match attendance and fan engagement trends to improve marketing strategies, ensuring that teams maintain a strong fan base and maximize ticket sales.
- **Leverage Social Media Data:** Integrate social media sentiment analysis into player performance evaluations to understand public perception and how it may impact player morale and brand value.
- **Integrate Scouting Reports with Data Analysis:** Combine traditional scouting methods with data-driven insights to create a more comprehensive approach to player recruitment.
- **Prioritize Team Strategy Adjustments:** Focus on how tactical changes (e.g., formation or substitutions) impact performance, providing teams with valuable information to optimize in-game strategies.

- **Explore Regional Preferences for Player Types:** Investigate whether different regions have preferences for specific player types or attributes, influencing recruitment and training programs in various parts of the world.
- **Incorporate Fans' Feedback:** Include fan feedback as part of the analysis, looking at how player performances are perceived by supporters and whether this affects team loyalty or ticket sales.
- **Examine Player Fatigue:** Analyze how factors like match frequency and travel impact player fatigue, which in turn affects performance, and provide teams with strategies for managing player workload.
- **Enhance Player Development Pathways:** Using insights from the data, teams should focus on creating better player development pathways that can be customized based on performance trends and future potential.
- **Identify Performance Peaks and Troughs:** Track when players experience peak performance periods, and develop strategies to ensure they perform consistently over the course of a season or career.
- **Ensure Data Quality and Regular Updates:** To maintain the reliability of the analysis, it is crucial to ensure that the dataset is consistently updated with accurate, high-quality data and is free from errors or biases.

REFERENCES

- [1] Schneeweiss, A. (2023). *Optimizing Football Coaching Through Data-Driven Insights: A Power BI Dashboard Approach* (Master's thesis, Universidade NOVA de Lisboa (Portugal)).
- [2] Ernst, C. (2018, March). The “Bigger Picture”: How the 2010 FIFA World Cup and the 2014 World Design Capital Events Conditioned Cape Town’s Urban Transformation. In *Urban Forum* (Vol. 29, pp. 63-84). Springer Netherlands.
- [3] Shardelow, C. (2022). Gamble-to-Win: Regulating Video Game Loot Boxes under the FTC's Unfair and Deceptive Practices Doctrines. *J. Bus. & Tech. L.*, 18, 331.
- [4] Bruck, K. Bridging the gap between millions of people in real time: fusing data of wearables in services.