International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Visual Analysis of FIFA World Cup Data

Renu Yadav[1], Swarnima Arya[2], Abhimanyu Tiwari[3], Mr. Rajesh Sharma[4]

[1,2,3]School of Computer Science & Engineering, Galgotias University, Greater Noida, India
[4]Assistant Professor (SCSE), Galgotias University, Greater Noida, India

Abstract
The FIFA World Cup has become an international event during the last thirty years, and the 2022 final in Qatar is expected to draw over three billion spectators. This study examines the variables affecting national football success by using data from 68 World Cup-playing nations between 1994 and 2022. The influence of national sports prowess, as seen by Olympic accomplishments, and football heritage, such as hosting history and World Cup victories, are among the important factors that determine success, according to a generalized linear mixed model (GLMM). It's interesting to note that the study finds a negative relationship between performance and the Human Development Index (HDI), although past football involvement has no bearing. Additionally important are interaction effects, which include population size, economic status, and football tradition.

Keywords: FIFA World Cup, data analysis, football performance, macro factors, econometrics.

INTRODUCTION

Considered the most popular sport globally, football reaches its peak in the FIFA World Cup, an international tournament that brings countries together through competition and enthusiasm. Ever since it started in 1930, the World Cup has experienced a significant increase, with almost 10,000 athletes from more than 200 countries taking part. This impressive growth showcases the importance of football in globalizing, as advanced and emerging countries vie for attention globally.

Over the last thirty years, football has evolved into a lucrative industry worth billions of dollars, closely connected to the worldwide economy and advancements in society. The success of national teams mirrors the overall progress and reputation of a nation. Participating successfully in the World Cup represents not just athletic skill but also helps boost the economy by generating income from broadcasting rights, sponsorships, and merchandise sales.

The significance of analyzing success factors in international competitions is highlighted by the prominence of football giants like Brazil, Germany, and Argentina, as well as Europe's "Big Five Leagues" (La Liga, Serie A, Ligue 1, Bundesliga, and the Premier League). Comprehending these factors provides understanding of how global football resources are distributed and the ways to achieve competitive success.

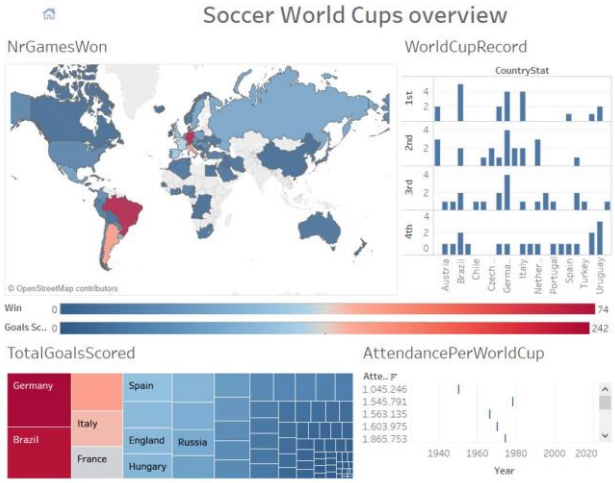

Fig. 1. Overview dashboard of all the World Cups.

This research examines how national teams have performed in the last eight World Cups (1994-2022), looking at factors such as socio-economic development, football history, and national athletic prowess. The research uses various data sources, such as FIFA and the International Olympic Committee (IOC), along with econometric analysis, to uncover important factors that impact success. The results are intended to assist policymakers, teams, and stakeholders in creating strategies to improve football performance and promote global development.

BACKGROUND AND RELATED WORK

Various techniques have been suggested for displaying sports data, such as soccer, to provide understanding of performance and match dynamics. Nevertheless, a lot of these methods depend on intricate processes that are challenging for laypeople to grasp, or utilize basic graphics that lack substance. A major problem is the lack of effective integration of interactions among visualizations, like connecting various graphics or coordinating views, which could improve how they are used. Users without experience in data visualization might find these tools less user-friendly and accessible in the absence of interactive features.

The majority of soccer visualizations either concentrate on an individual player's actions in a game or on displaying combined statistics, frequently overlooking the chance to analyze data from various angles. Soccer games produce large quantities of data, which consist of numerous events recorded by cameras, wearable sensors, and GPS technology. Nevertheless, only a small portion of the data is thoroughly examined, leaving a significant amount of its possibilities unexplored. This is partially due to the fact that current techniques often depend on easily available data, overlooking the potential insights that can be gained from examining data from various perspectives.

Additionally, various individuals invested in soccer, including teams, supporters, talent scouts, and experts, have unique requirements and hopes. The variety of preferences within different groups of players makes it challenging to develop a one-size-fits-all platform, as each group prioritizes distinct aspects of the game. Platforms such as Soccer STATS and Who Scored mainly serve particular user groups, mainly offering fundamental data or interactive graphics centered around a sole viewpoint. The difficulty is in providing extensive, complex visual representations that can cater to different user demographics.

Current dashboards commonly display fixed charts and tables showing key metrics such as passes, possession, and shots on goal, with no interactive features. Adding more imaginative, engaging functions like filtering matches or players and exploring timelines could greatly improve analysis and provide a more thorough and adaptable comprehension of the topic.

DATASETS

The data used in this study were collected from various datasets containing information from all FIFA World Cups. Initially, the data included basic details such as teams, final results, referees, and stadiums. Over time, the scope of data collection expanded, incorporating additional statistics such as goals, assists, corner kicks, and fouls. By recent editions of the World Cup, the dataset had grown to include many more match details. These included pass success percentages, the number of fouls, tackles, total distance covered by teams, and more. Modern datasets now provide comprehensive insights, capturing event details like passes, player involvement, ball location on the pitch, areas covered by players, and shot locations, including the position from which players attempt to score.

For this project, we used several datasets to design an interactive and user-friendly visualization for non-experts. These datasets included:

- **World Cup matches**: This dataset contains the results of all soccer matches from all World Cups, along with halftime scores, referees, and stadiums.
- **World Cups**: This dataset provides statistics for countries that reached the semi-finals, including total matches, goals scored, and attendance figures for each World Cup.
- **World Cup players**: A dataset containing detailed information about individual player statistics, such as appearances, goals scored, assists, and other performance metrics across all World Cups.

The focus of this analysis was on the total number of games won, goals scored (broken down by first half, second half, and total), and the number of times each country finished in the top four in each World Cup. Additionally, we highlighted improvements in data collection across World Cup editions, using metrics like distance covered, pass accuracy, ball possession, and shot details for each game, offering a comprehensive view of team performance across tournaments.

FORMULATION OF PROBLEM

In this part, we outline the activities that can be tackled using the chosen datasets, ranging from simple analyses to more intricate inquiries. To begin with, we offer a comprehensive summary of the data, beginning with every team and World Cup event. This broad viewpoint is later narrowed to concentrate on the best-performing nations, leading to a detailed examination of specific national teams. This method involves starting with a broad view and only delving into specific information when needed, ensuring a clear grasp of overall patterns before focusing on specifics.

The dataset covers every World Cup game from 1930 till now, allowing for an initial analysis of different national teams' performances. By examining key indicators like wins, goals, and attendance, we can pinpoint the top-performing teams in the history of the World Cup. This initial examination provides a distinct view of the teams that have consistently excelled in the tournament.

As we delve deeper into the analysis, the recent World Cup datasets, particularly the player-level statistics, enable us to compare teams in various aspects. The aim is to evaluate team performance through analyzing aspects like scoring patterns, possession levels, and defensive capabilities. This improved method aims to discover the root causes influencing the success or failure of each team, aiding

in the comprehension of their overall strategies Continuing to narrow down, we concentrate on the role of each player in the team's achievements. Studying data about goal scorers, such as where goals are scored and which players are involved, provides important information about team strategies and tactical choices. This data at the player level provides a more in-depth insight into the unique strengths and weaknesses of each team, improving the overall assessment of team performance and offering strategic insights into their playing styles in the World Cup.

VISUALISATION DESIGN

The visualization aspect of this research was constructed utilizing Tableau, a platform selected for its user-friendly interface and robust interactive capabilities. The visual design is intended to enhance comprehension of FIFA World Cup data by capitalizing on human perceptual strengths to effectively identify patterns and trends.

A. Techniques Employed

The visualizations were crafted to accommodate a diverse audience, including those without expertise in data visualization. To promote clarity and functionality, the design process commenced with the identification of essential tasks and target demographics, establishing objectives such as facilitating comparative analysis and pinpointing the most significant factors influencing match results. The selected visual formats encompass heat maps, tree maps, stacked bar charts, horizontal bar charts, side-by-side bar charts, and scatter plots, all of which are straightforward yet powerful in conveying data insights.

Color is a pivotal element in the visualization framework, with neutral shades (e.g., light gray) denoting average outcomes and vibrant colors (e.g., blue and red) representing contrasting data points. This strategy allows users to swiftly recognize critical patterns and trends, such as the most successful teams throughout World Cup history.

Three interactive dashboards were developed to examine the data from various angles: an overview dashboard, a goals dashboard, and a statistics dashboard. A primary menu links these dashboards, facilitating smooth navigation. Each dashboard features a home button, enabling users to easily return to the main screen, thereby enhancing the overall user experience.

B. Interaction and Linked Design

Interactive components have been incorporated to improve the exploration and extraction of data. Features such as hovering, zooming, and selecting data points enable users to concentrate on particular areas of interest. These interactions facilitate multiple queries per visualization, allow for detailed examination of specific data points, and enable filtering based on attributes chosen by the user. For example, users can filter data by country to assess performance trends over time or to examine detailed statistics of players.

The interactive dashboards are interconnected, meaning that selections made in one view are simultaneously reflected in all other views where the data is present. For instance, when a user selects a country, they can observe its geographic performance distribution, match statistics, and related visualizations all at once.

C. Application to the Datasets

Utilizing the World Cup matches dataset, the dashboards offer insights into metrics such as total games won, goals scored, and attendance figures. The World Cup players dataset enhances this analysis by allowing users to evaluate individual player performance, including goals scored, assists, and contributions to matches. The visualizations also integrate data from significant events in World Cup

history, allowing users to interactively explore both team and player dynamics. This design enables users to efficiently extract insights, rendering the exploration of FIFA World Cup data both accessible and informative.

APPLICATIONS OF VISUALISATION IN WORLD CUP DATA EXPLORATION

This section illustrates practical applications of visualizations and analyses using FIFA World Cup datasets. The focus is on overall performance across all tournaments, the accomplishments of historically top-performing teams, and insights into goal-scoring trends.

A. Comprehensive Overview of All World Cups

The dataset encompassing all World Cup matches from 1930 to the latest tournaments allows for an extensive examination of global football trends. Through the overview dashboard, users can assess various metrics, including the total matches played, goals scored, and attendance figures across all tournament editions. For instance, visual representations reveal a significant increase in participation and attendance over the years, figure.6 indicative of football's growing international appeal. The data also highlights the consistent success of specific regions, particularly European and South American nations, which frequently advance to the later stages of the competition.

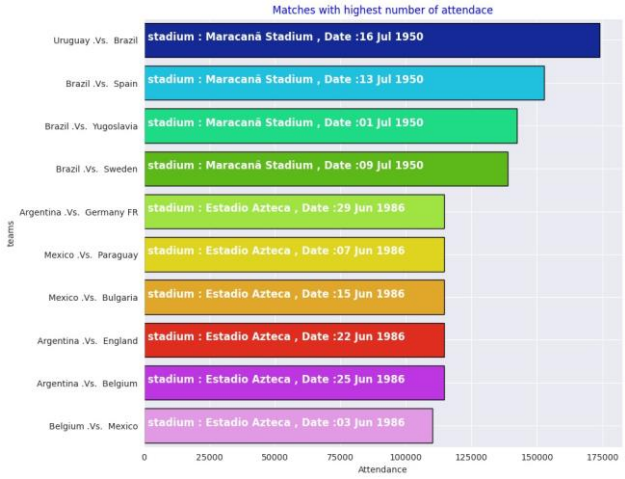

Fig. 6. *Highest attendance of team over the years..*

B. Historical Performance of Top Teams

The evaluation of historically top-performing teams is supported by visualizations that rank nations based on their victories, goals scored, and appearances in the top four positions. Notably, countries such as Brazil, Germany, and Italy consistently emerge as leaders in these performance metrics, having secured multiple World Cup titles and reached the finals or semi-finals in various tournaments. These findings are depicted through heat maps and bar charts, facilitating a straightforward comparison of performance across different decades. Furthermore, trends illustrating the emergence of new contenders alongside the enduring dominance of traditional football powerhouses are emphasized figure 7.

SUMMARY OF LAST FOUR IN ALL SOCCER WORLD CUP TOURNAMENTS

Year	Venue nation(s)	Venue of final	Winner	Runner-up	Third	Fourth
1930	Uruguay	Montevideo	**Uruguay**	Argentina	United States*	Yugoslavia*
1934	Italy	Rome	**Italy**	Czechoslovakia	Germany	Austria
1938	France	Paris	**Italy**	Hungary	Brazil	Sweden
1950	Brazil	Rio de Janeiro	**Uruguay**	Brazil	Sweden**	Spain**
1954	Switzerland	Bern	**West Germany**	Hungary	Austria	Uruguay
1958	Sweden	Solna^	**Brazil**	Sweden	France	West Germany
1962	Chile	Santiago	**Brazil**	Czechoslovakia	Chile	Yugoslavia
1966	England	London	**England**	West Germany	Portugal	USSR
1970	Mexico	Mexico City	**Brazil**	Italy	West Germany	Uruguay
1974	West Germany	Munich	**West Germany**	Netherlands	Poland	Brazil
1978	Argentina	Buenos Aires	**Argentina**	Netherlands	Brazil	Italy
1982	Spain	Madrid	**Italy**	West Germany	Poland	France
1986	Mexico	Mexico City	**Argentina**	West Germany	France	Belgium
1990	Italy	Rome	**West Germany**	Argentina	Italy	England
1994	United States	Pasadena^^	**Brazil**	Italy	Sweden	Bulgaria
1998	France	Paris	**France**	Brazil	Croatia	Netherlands
2002	Japan/South Korea	Yokohama	**Brazil**	Germany	Turkey	South Korea
2006	Germany	Berlin	**Italy**	France	Germany	Portugal
2010	South Africa	Johannesburg	**Spain**	Netherlands	Germany	Uruguay
2014	Brazil	Rio de Janeiro	**Germany**	Argentina	Netherlands	Brazil
2018	Russia	Moscow	**France**	Croatia	Belgium	England

* Position as per FIFA records but debatable.
** Position as per points in final round of 4 teams.

Solna^ is adjacent to Stockholm.
Pasadena^^ is adjacent to Los Angeles.

Note: East Germany appeared in the tournament once, in 1974.

FIG.7. Total number of games won by each country across all World Cups till 2018.

C. Analysis of Goal-Scoring Trends

Examining goal-scoring patterns offers valuable insights into team and player strategies. The goal dashboard enables users to investigate data regarding the number of goals scored by each team, the locations of shots taken, and the timing of goals. For instance, the analysis uncovers trends such as the prevalence of high-scoring matches during group stages compared to the more defensive strategies employed in knockout rounds. Additionally, individual player contributions are scrutinized, showcasing their impact on overall team performance figure 8.

Fig. 8 Total number of goals scored by each country across all World Cups

CONCLUSION AND FUTURE WORK

This study introduced an interactive dashboard aimed at analyzing FIFA World Cup data, specifically tailored to be accessible for users with minimal experience in data visualization. By incorporating a variety of visual representations and interactive elements, the dashboard allows users to investigate soccer data from diverse angles, such as team performance, individual player contributions, and goal-scoring patterns throughout all World Cups. Through various use cases, we illustrated how the dashboard supports data-driven insights into the history of the World Cup and the strategies employed by different teams. Although the current version offers significant insights, a comprehensive assessment of its usability and intuitiveness is essential. Future research could evaluate the dashboard's effectiveness

for both novice and expert users, potentially utilizing advanced methodologies like eye-tracking technology to further refine its design. Moreover, the analytical scope could be broadened to incorporate predictive modelling, allowing users to anticipate outcomes based on historical data and team performance metrics. Another promising avenue for future investigation is to enhance the focus on individual player analysis. While the existing datasets provide useful insights into player statistics, the addition of more detailed data could facilitate event-level predictions, such as goals, assists, or defensive actions. Furthermore, expanding the analysis to juxtapose national team performances with club league data could yield a more comprehensive understanding of the dynamics within global football. A key strength of the interactive dashboard is its capacity to spatially visualize data, revealing patterns associated with formations, possession, and attacking strategies. By improving these spatial visualizations, users can achieve a more profound understanding of tactical adjustments and their effects on match outcomes. This research establishes a robust foundation for future advancements in soccer data visualization and analysis, with potential applications extending beyond the World Cup to other areas of the sport.

REFERENCES

1. World cups statistics, 2018. https://gitlab.com/djh_or/ 2018-world-cup-stats/blob/master/world cup 2018 stats.csv Accessed; August, 2020.
2. G. L. Andrienko, N. V. Andrienko, G. Budziak, T. von Landesberger, and H. Weber. Exploring pressure in football. In Proceedings of the 2018 International Conference on Advanced Visual Interfaces, AVI, pages 54:1–54:3, 2018.
3. T. Blascheck, M. Burch, M. Raschke, and D. Weiskopf. Challenges and perspectives in big eye-movement data visual analytics. In Proceedings of the 1st International Symposium on Big Data Visual Analytics, pages 17–24, 2015.
4. T. N. Company. World football report - Nielsen, 2018. https://www. nielsen.com/us/en/insights/report/2018/world-football-report/ Accessed:August, 2020.
5. FIFA. FIFA World Cup™ comparative statistics 1982-2014, 2014. https://resources.fifa.com/image/upload/ comparative-statistics-1982-2010-519730.pdf?cloudid=o3dx8kjlopctz35v460o Accessed: August, 2020.
6. C. G. Healey and J. T. Enns. Attention and visual memory in visualization and computer graphics. IEEE Transactions on Visualization and Computer Graphics, 18(7):1170–1188, 2012.
7. D. A. Keim. Solving problems with visual analytics: Challenges and applications. In Proceedings of Machine Learning and Knowledge Discovery in Databases - European Conference, pages 5–6, 2012.
8. J. L. S. Malqui, N. M. L. Romero, R. Garcia, H. Alemdar, and J. L. D. Comba. How do soccer teams coordinate consecutive passes? A visual analytics system for analysing the complexity of passing sequences using soccer flow motifs. Computers and Graphics, 84:122–133, 2019.
9. R. R. Memmert. Match analysis, big data and tactics: Current trends in elite soccer. German Journal of Sports Medicine, 3, 2018.
10. C. Perin, R. Vuillemot, C. D. Stolper, J. T. Stasko, J. Wood, and S. Carpendale. State of the art of sports data visualization. Computer Graphics Forum, 37(3):663–686, 2018.
11. J. C. Roberts. Guest editor's introduction: special issue on coordinated and multiple views in exploratory visualization. Information Visualization, 2(4):199–200, 2003.

12. M. Ryoo, N. Kim, and K. Park. Visual analysis of soccer players and a team. Multimedia Tools and Applications, 77(12):15603–15623, 2018.
13. B. Shneiderman. The eyes have it: a task by data type taxonomy for information visualizations. In Proceedings 1996 IEEE Symposium on Visual Languages, pages 336–343, 1996.
14. SoccerSTATS. SoccerSTATS, 2019. https://www.soccerstats.com Accessed: August, 2020.
15. C. Solberg, Harry Arne Gratton. Broadcasting the World Cup, pages 47–62. Palgrave Macmillan UK, London, 2014.
16. R. Theagarajan, F. Pala, X. Zhang, and B. Bhanu. Soccer: Who has the ball? Generating visual analytics and player statistics. In 2018 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR, pages 1749–1757. IEEE Computer Society, 2018.
17. E. R. Tufte. The Visual Display of Quantitative Information. Graphics Press, 1992.
18. W. M. P. van der Aalst. Process Mining - Data Science in Action, Second Edition. Springer, 2016.
19. C. Ware. Information Visualization: Perception for Design. Morgan Kaufmann, 2004.
20. C. Ware. Visual Thinking: for Design. Morgan Kaufmann Series in Interactive Technologies, Paperback, 2008.
21. R. M. G. Wesley, M. Eldridge, and P. Terlecki. An analytic data engine for visualization in Tableau. In T. K. Sellis, R. J. Miller, A. Kementsietsidis, and Y. Velegrakis, editors, Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD, pages 1185–1194. ACM, 2011.
22. WhoScored. Whoscored?, 2019. https://www.whoscored.com Accessed: