From Kick-off to Glory: Visualizing the Evolution of FIFA World Cups

Bhargavi Dwivedi, Utkarsh Pratap Singh Jadon

1 INTRODUCTION

Our project titled 'From kick-off to Glory: Visulizing the Evoluation of FIFA World Cups' explores the transformation of national team strategies, performance trends, and the overall progression of FIFA World Cup tournaments since their beginning in 1930. The driving force behind our study is to gain insights into the expansion of the tournament, the dynamics of participating teams, and the global impact on the World Cup. We will be analyzing the FIFA World Cup dataset from Kaggle for our research [1].

Our goal is to depict the historical evolution of the World Cup, focusing on aspects like the variety of teams participating, the total number of matches played, and the overall goal counts. We aim to investigate whether scoring goals is the sole factor in securing World Cup victories. Our project includes data visualizations of various aspects such as the frequency of podium finishes per country, goal statistics, audience sizes, and home vs. away team performance, to comprehensively illustrate the progression of the World Cup over the years and address our main objective [8]. We plan to use a range of visualization methods such as interactive bar graphs, line charts, box plots, radar charts, bubble charts, and violin plots, based on our initial project assessment. Additionally, we developed a network function to examine team interactions from 1930 to 2014, helping us further explore the factors leading to World Cup success. Our analysis will cover several perspectives, including Timeline Analysis, Performance Metrics, and Team and Player Network Analysis

2 RELATED WORK

In charting the path for our project, several pivotal studies have been instrumental. The research paper "PassVizor: Toward Better Understanding of the Dynamics of Soccer Passes" [7] sheds light on soccer pass dynamics. It underscores the significance of passing interactions in forging scoring opportunities and introduces PassVizor, a visual analytics tool for in-depth analysis of pass dynamics. This tool helps identify evolving pass strategies through a unique glyph-based design, offering insights into various match phases, including player identities, spatial factors, and team formations. This research significantly contributes to understanding a team's shifting pass strategies within a game.

The comprehensive survey "A survey of competitive sports data visualization and visual analysis" [2] offers a broad view of the field, introducing a taxonomy for sports data visualization. It categorizes data into spatiotemporal and statistical groups and outlines key tasks like feature presentation, comparison, and prediction. The survey also examines different visualization techniques and their relationship with these tasks, providing an extensive resource for applying visualization methods in competitive sports data.

- *Bhargavi Dwivedi is with Ohio State University. E-mail: dwivedi.30@osu.edu*
- *Utkarsh Pratap Singh Jadon is with Ohio State University. E-mail: jadon.1@osu.edu*

Another pivotal work, "Feature-Driven Visual Analytics of Soccer Data" [4], introduces an advanced system for analyzing detailed, position-based soccer data. Recognizing soccer's widespread appeal and research significance, this method incorporates Visual Analytics to provide multi-player and event-based views. It selects and visualizes critical features for specific analytical objectives, aiding soccer analysts in pinpointing key events. The system's modular architecture is both flexible and expandable, enhancing soccer data analysis and setting a benchmark for sports Visual Analytics.

Furthermore, the paper "Forecasting the FIFA World Cup – Combining result- and goal-based team ability parameters" [6] offers a unique perspective on predicting FIFA World Cup outcomes. It compares result-based Elo ratings and goal-based ODM ratings in forecasting, using Monte Carlo simulations to determine teams' probabilities of advancing in the tournament. This study stands out for its comprehensive approach, encompassing evaluation on previous World Cups and extending its application to domestic league football matches.

These studies collectively underpin our project's methodology, offering a wealth of knowledge and innovative approaches that guide our exploration into the evolution of FIFA World Cups.

3 METHODS

In the methodology adopted for this study, three distinct datasets from Kaggle were utilized: WorldCupPlayers.csv, WorldCup-Matches.csv, and WorldCups.csv. Each dataset plays a crucial role in supporting the research. WorldCupPlayers.csv, encompassing every player's World Cup journey, provides detailed player-specific data. WorldCupMatches.csv presents a comprehensive match-by-match account, offering a chronological narrative of the games. Meanwhile, WorldCups.csv encapsulates the overall results of the tournaments. The data cleaning process was meticulously carried out, involving the rectification of inconsistencies in player names, addressing missing values, and standardizing team names to ensure uniformity. This rigorous data preparation process was crucial to ensure the reliability and accuracy of the subsequent analysis.
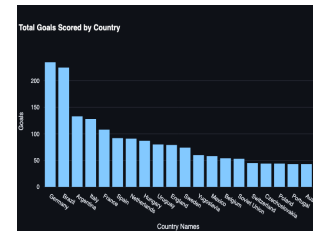

Fig. 1: Highest Goal Scorers

Figure 1 shows the highest goal scorers in FIFA World Cup History since its inception in 1930 till 2014. Germany is the highest goal scorer, followed by Brazil, Argentina, and Italy.

Figure 2a shows the increase in number of goals scored every world cup, depicting the growing competitiveness of the tournament. Figure
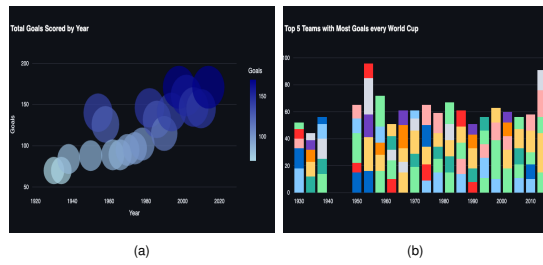
Fig. 2: World Cup Trends

2b shows the top 5 teams with most goals in every World Cup. We can see that the teams like Germany, Brazil, Argentina, Italy, have been a constant participant in these world cups uniformly throughout time.

For the presentation and analysis of data, Streamlit was employed to design an interactive dashboard, while NetworkX [3] was used to analyze team interactions. The dashboard was structured to highlight three primary areas of analysis: goal analysis, attendance and participation analysis, and cup analysis. In the goal analysis section, a bar chart depicting total goals scored by each country (Figure1), a bubble chart (Figure2a) illustrating total goals scored per year from 1930 to 2014, and a box plot (Figure 2b) showing the top 5 teams in FIFA World Cup from 1930 to 2014 were included. For the attendance and participation analysis, three distinct visualizations were created: a line chart showing average attendance per match over the years (Figure 3a), and a network analysis of team interactions from 1930 to 2014 (Figure 4). A significant facet of the World Cup's evolution is its expansion in terms of participating teams. A radar chart (Figure 3b) was devised to visualize this growth, highlighting the increasing number of teams qualifying for the tournament over the years. This visualization clearly depicts the progressive inclusivity of the World Cup, reflecting the tournament's growth from its inception to the present day. Lastly, the cup analysis segment featured a bar chart of matches with the highest attendance, another bar chart detailing the number of podium finishes by country, and a violin plot displaying home and away goals per year [5], each offering interactive features for detailed exploration.

Fig. 4: (a) - (f) Team Network Analysis for years 1930, 1954, 1966, 1986, 2006, 2014

towards a more interconnected and globalized sporting event, mirroring the World Cup's development into the expansive, inclusive competition it is today.

This comprehensive approach, combining data from multiple sources and utilizing a variety of visualization techniques, facilitated a multi-faceted analysis of the World Cup, revealing insights into team performances, audience engagement, and historical trends in the tournament.
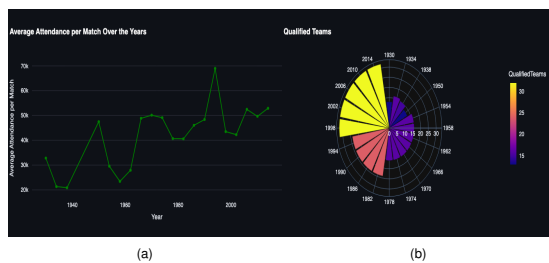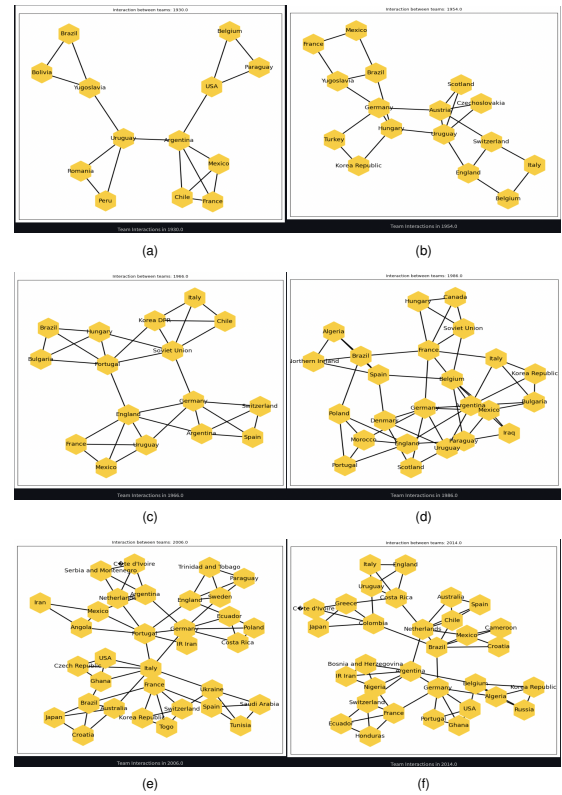

Fig. 3: Growth in Participation

Table 1 depicts the evolution of how different teams have interacted with each other over different editions of world cups, and plausible reasoning for the change.

Over the decades, the FIFA World Cup has witnessed a remarkable increase in team interactions, as evidenced by the network charts from various years (as seen in the network charts for 1930, 1954, 1966, 1986, 2006, and 2014). This evolution is characterized by the growing number of participating teams and the complexity of match interactions. Initially, the World Cup started with a modest number of teams and matches, indicative of the tournament's nascent stage. The comparison of network charts over time demonstrates a trajectory

Table 1: Evolution of Team Interactions in FIFA World Cups

Year	# of Teams (Nodes)	# of Matches (edges)	Remarks
1930	13	16	Initial tournament, less connectivity
1954	16	26	Increased teams, more interactions
1966	16	32	Stable number of teams, increased matches
1986	24	52	Expansion of tournament, significant growth in interactions
2006	32	64	Modern era, high connectivity
2014	32	64	Consistent with modern format, maximum interactions

4 RESULTS

In the initial segment of the analysis, a bar chart (Figure 1) was constructed to visualize the total goals scored by countries. It was observed that Germany has amassed the highest goal count, closely followed by Brazil, Argentina, and Italy. This aligns with our project's goal to decipher the connection between a nation's goal-scoring record and their World Cup victories. Moreover, the visualization detailing the top goal-scoring teams in each World Cup (Figure 2b) demonstrated the consistent presence of these countries among the top scorers, bolstering the theory that goal-scoring prowess plays a crucial role in World Cup triumphs.
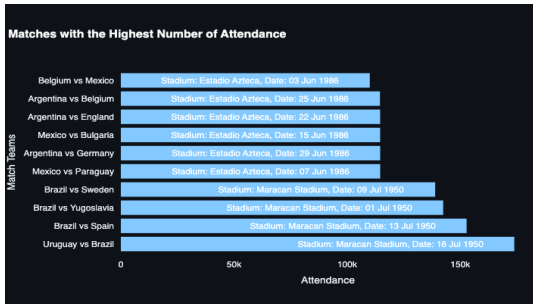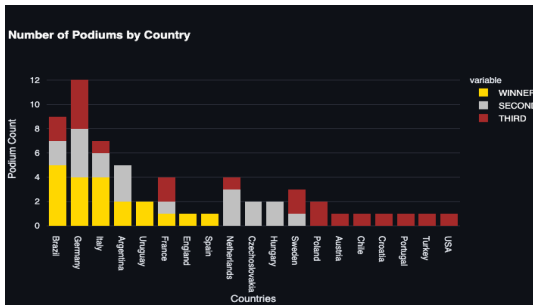

Fig. 5: Highest Single Match Attendance

Fig. 6: Podium Counts

The cup analysis segment offered a bar chart of podium finishes by country (Figure 6), which, when scrutinizes only the World Cup winners, revealed Brazil surpasses Germany in terms of titles despite a lower goal total. Italy also emerged with more victories than

Argentina, despite scoring fewer goals. This evidence suggests that while goal-scoring is important, it is not the singular factor in securing World Cup success. The analysis of the matches with the highest attendance (Figure 7) suggested that Brazil's frequent presence in such matches might correlate with their impressive performance, possibly due to the impact of audience support.
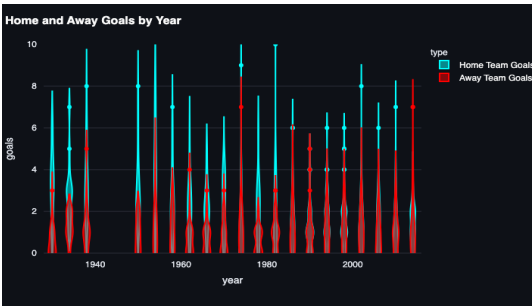

Fig. 7: Home vs Away Analysis

The study's final results, presented through a violin plot (Figure 6), indicated a clear home-field advantage, signifying the importance of stadium familiarity and fan backing. This comprehensive analysis of goal trends highlighted the significant role of both home advantage and audience engagement, reinforcing the notion that World Cup success is a multifaceted achievement influenced by both on-field prowess and supportive external factors.

The extensive data analysis and visualization provided a nuanced perspective on the dynamics that define FIFA World Cup tournaments, illustrating that success in football's most esteemed competition extends beyond the traditional metric of goal-scoring.

5 CONCLUSIONS

An exhaustive analysis and visualization of data were conducted, culminating in the creation of an interactive dashboard that facilitates an in-depth examination of World Cup trends. This tool proved to be invaluable in distilling the complex dynamics of tournament victories beyond mere goal tallies. It was revealed that the quantity of goals scored by a team is not the preeminent indicator of their World Cup title count, challenging traditional metrics of soccer success.

The role of fan support was also highlighted as a critical component in enhancing team performance. It was demonstrated that the psychological advantages provided by home supporters play a significant role in a team's triumph. These insights underscore the conclusion that success in World Cup competitions is influenced by a combination of factors, including but not limited to, the support teams receive from their fans, suggesting a nuanced view of what contributes to the making of a world champion in the realm of football.
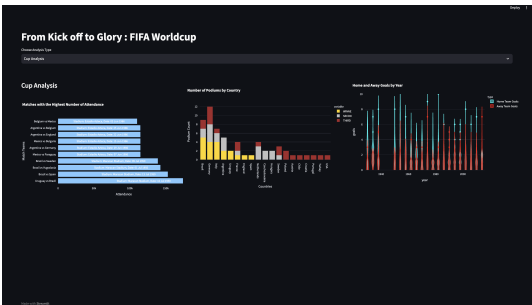

Fig. 8: Proposed Visualization

Table 2: Task Mapping

Overview Task	Solution	Why?
How does the evolving team count in the FIFA World Cup reflect its expansion?	Bar charts illustrating the growth in the number of participating teams, highlighting the tournament's expansion.	Visually communicate the FIFA World Cup's expansion and growth scale.
How does the evolution of the FIFA World Cup mirror global soccer trends and its international influence?	Box plots show match growth, highlighting tournament expansion	Summarizes key statistics, revealing annual match trends.
How has the FIFA World Cup's competitiveness evolved, as shown by goal scoring trends?	Line graphs illustrating the growth in the number of goals scored, providing insights into the tournament's increasing competitiveness	Line graphs compare tournaments, spotlighting key shifts in goal scoring and competitiveness.
Examine the impact of home-field advantage on international sports competition dynamics.	Pie chart illustrating and comparing the performances of home and away teams	It visually clarifies the impact of home-field advantage on international sports dynamics by showing team performance.

User Interface Design

- Filtering Options: Allow users to filter visualizations based on specific World Cup editions, teams, and 1performance metrics

- Comparative Analysis Tools: Include tools for users to compare the evolution of multiple teams and tournaments side by side

- Interactive Widgets: Implement widgets for dynamic filtering and user interaction, enhancing the overall user experience

Table 1 maps the overview tasks, with its solution and its requirement. Figure 1 shows two of the proposed visualizations.

LIST OF FIGURES

LIST OF TABLES

REFERENCES

[1] A. Becklas. Fifa world cup dataset: https://www.kaggle.com/datasets/abecklas/fifa-world-cup.

[2] M. Du and X. Yuan. A survey of competitive sports data visualization and visual analysis. *Journal of Visualization*, 24:47–67, 2021. doi: 10.1007/s12650-020-00687-2

[3] A. Hagberg, P. Swart, and D. S Chult. Networkx.

[4] H. Janetzko, D. Sacha, M. Stein, T. Schreck, D. Keim, and O. Deussen. Feature-driven visual analytics of soccer data. 09 2014. doi: 10.1109/VAST .2014.7042477

[5] Jaspersoft. What is a radar chart?: https://www.jaspersoft.com/articles/what-is-a-radar-chart: :text=a

[6] P. Robberechts and J. Davis. Forecasting the FIFA world cup – combining result- and goal-based team ability parameters. In U. Brefeld, J. Davis, J. Van Haaren, and A. Zimmermann, eds., *Machine Learning and Data Mining for Sports Analytics. MLSA 2018*, vol. 11330 of *Lecture Notes in Computer Science*, p. Page Range. Springer, 2019. doi: 10.1007/978-3-030 -17274-9_2

[7] X. Xie, J. Wang, H. Liang, D. Deng, S. Cheng, H. Zhang, W. Chen, and Y. Wu. Passvizor: Toward better understanding of the dynamics of soccer passes. *CoRR*, abs/2009.02464, 2020.

[8] A. Yadav. Fifa world cup match data analysis: https://github.com/imajeetyadav/fifa-world-cup-match-data-analysis/blob/master/fifa