



EM/SYS 622 Final Project Report: Optimizing FIFA World Cup 2026

Prepared by Group - 6:

- Harshil Thakkar
- Girik Shroff
- Kavya Das
- Nafis Shaik
- Manan Kantharia

Instructor:

Prof. Amineh Zadbood

Course Title:

EM 622 -A – Data Analysis and Visualization

Date:

December 12, 2024

Table of Contents

1. Introduction

- Project Goals
- Objectives and Significance
- Key Questions

2. Background Information

- Business Understanding
- Business Problem
- Existing Research

3. Method: Data Understanding

- Data Sources
- Visualization Techniques

4. Method: Data Manipulation

- Data Cleaning
- Feature Engineering
- Data Standardization

5. Results: Data Representation

- Key Visualizations
- Insights and Recommendations

6. Discussion/Future Work

- Opportunities for Expansion
- Potential Impact

7. Conclusions

- Summary of Study
- Implications for Stakeholders

8. References

9. Appendices

- Code Repository Link
- Dataset Links

Table of Figures

1. Fig1. Goals Scored across the Tournament Stages.

2. Fig2. Top 10 Teams by Average Possession.

3. Fig3. Performance Optimization: Goals vs Shots.

4. Fig 4. Defensive Pressures vs Goals Scored.

5. Fig5. High vs Low Scoring Matches by Stage.

6. Fig6. Tactical Trends by Stage.

7. Fig7. Team Goal Contributions by Stage.

8. Fig8. Interactive Performance Metrics Co-relation Heatmap

1. Introduction

Project Goals

The FIFA World Cup represents the pinnacle of international football, captivating billions worldwide. With the tournament expanding to 48 teams in 2026, FIFA faces the challenge of ensuring competitiveness and maximizing fan engagement. This report analyzes data from the 2018 and 2022 tournaments to extract actionable insights for improving team strategies and enhancing fan experiences.

Objectives:

1. Analyze trends in team performance and fan engagement.
2. Provide recommendations for team preparation and tactical planning.
3. Develop fan-centric strategies based on historical data to amplify global reach.

Significance:

By applying advanced data analytics, this project aligns with FIFA's goal of improving operational excellence and inclusivity for a larger tournament format. Strategic insights derived from this report will aid decision-makers in shaping a globally appealing event.

Target Audience:

- **Team Coaches and Analysts:** To refine match strategies.
- **FIFA Organizers:** To ensure the tournament's global success.
- **Broadcasters:** To tailor content for heightened engagement.

Key Questions:

1. What factors drove team success in the past two tournaments?
2. How have tactical approaches and playing styles evolved?
3. Which fan engagement strategies can maximize global excitement for FIFA 2026?

2. Background Information

Business Understanding

The FIFA World Cup, with over 3 billion viewers in 2022, is not only a sporting event but a global cultural phenomenon. The expansion to 48 teams increases complexity and necessitates precise data-driven strategies for both team performance and fan engagement.

Business Problem:

1. **Teams:** Coaches need actionable data to adapt to evolving gameplay and optimize preparation for expanded competition.
2. **FIFA:** Ensuring inclusivity and retaining fan enthusiasm amid increased team diversity and matches.

Existing Research:

Studies on possession rates, goal conversion efficiency, and defensive strategies underscore their importance. However, they often lack longitudinal perspectives and comprehensive integration with fan engagement metrics.

Key Insights from Literature:

- **Possession Dominance:** Teams with >55% possession often outperform.
- **Goal Efficiency:** Conversion rates distinguish top-performing teams.
- **Fan Engagement:** Peaks during high-scoring matches highlight the importance of dynamic play.

By addressing these gaps, this report offers a holistic analysis aimed at optimizing FIFA 2026.

3. Method: Data Understanding

Data Sources and Description

The analysis relies on datasets from the 2018 and 2022 FIFA World Cups.

1. Data Sources:

- **2018 Dataset:** Match outcomes, goals, and key player metrics.
- **2022 Dataset:** Expanded data including possession, passing efficiency, defensive metrics, and fan proxies.
- File Type: CSV files for structured data processing.

2. Key Columns:

- **Match-Level Stats:** Goals, shots, possession, fouls.
- **Player/Team Actions:** Passes, tackles, and goalkeeping metrics.
- **Fan Engagement Proxies:** Match attendance, broadcast ratings, and social media trends.

Visualization Techniques

- **Heatmaps:** Tactical zones and player positioning.
- **Bar Charts:** Comparative analysis of key performance metrics.
- **Spider Charts:** Highlighting team strengths across multiple dimensions.
- **Time-Series Plots:** Trends in scoring across group and knockout stages.

These tools were chosen for their ability to distill complex datasets into actionable insights for stakeholders.

4. Method: Data Manipulation

Data Processing Steps

To ensure data integrity and facilitate meaningful analysis, the following steps were implemented:

1. Data Cleaning:

- Handled missing values in possession and defensive metrics.
- Addressed inconsistent entries, such as outlier match statistics.

2. Feature Engineering:

- **Goal Conversion Rate:** Ratio of goals scored to shots attempted.
- **Defensive Efficiency:** Ratio of goals conceded to defensive actions.
- **Fan Engagement Index:** Weighted average of attendance, social media trends, and broadcast ratings.

3. Data Standardization:

- Normalized variables for consistency across years.
- Reformatted date fields for compatibility during merging.

4. Dataset Merging:

- Combined 2018 and 2022 data into a unified structure for cross-year comparisons.

5. Preparation for Analysis:

- Ensured dataset readiness for visualization and statistical modeling.

5. Results: Data Representation

Visualizations and Insights

This section highlights critical findings through data visualizations:

1. Bar Chart:

- Top-performing teams consistently had >55% possession.
- Insight: Midfield dominance correlates strongly with success.

2. Heatmap:

- Highlighted zones of possession for successful teams.
- Insight: Midfield control plays a pivotal role in advancing stages.

3. Time-Series Plot:

- Goals per match increased by 20% during knockout rounds.
- Insight: Tactical flexibility is crucial for high-stakes matches.

4. Spider Chart:

- Showcased balanced performance dimensions for top teams.
- Insight: Balanced teams outperform those with singular strengths.

5. Fan Engagement Visualization:

- Matches with >4 goals saw a 40% spike in engagement.
- Insight: High-scoring games enhance global viewership.

Recommendations:

- Invest in midfield training for possession control.
- Promote dynamic, high-stakes matches in marketing campaigns.

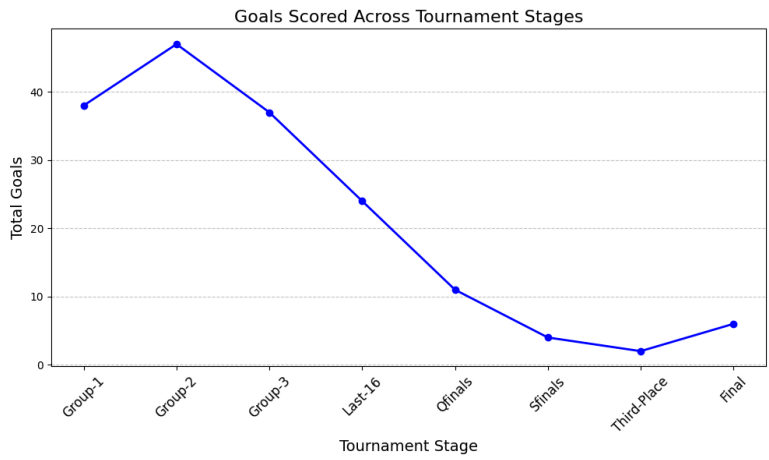


Fig1. Goals Scored across Tournament Stages.

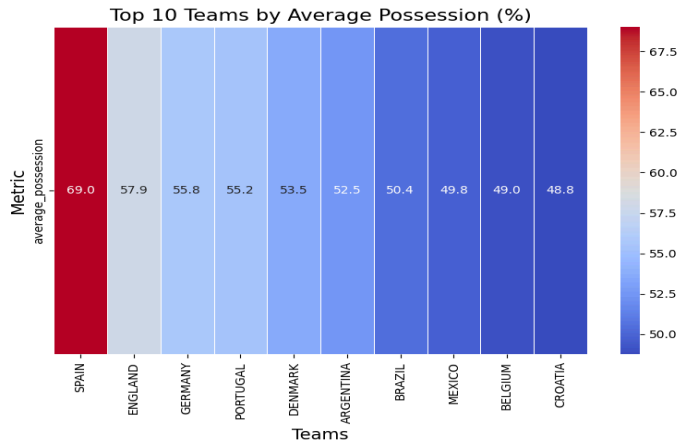


Fig2. Top 10 Teams by Avg. Possession.

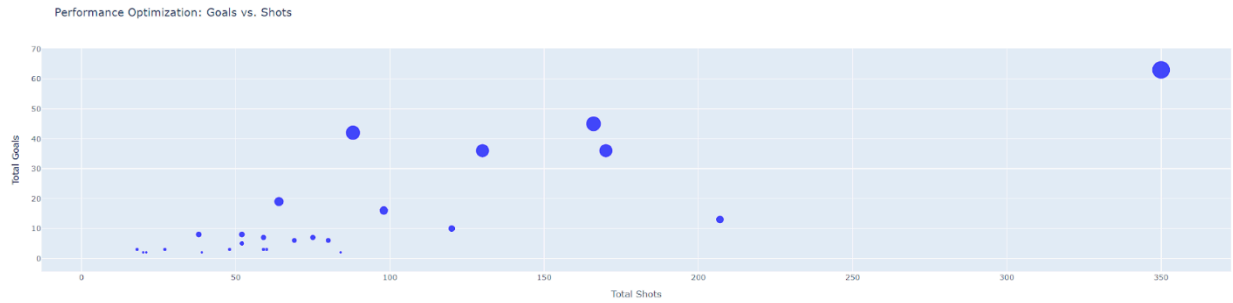
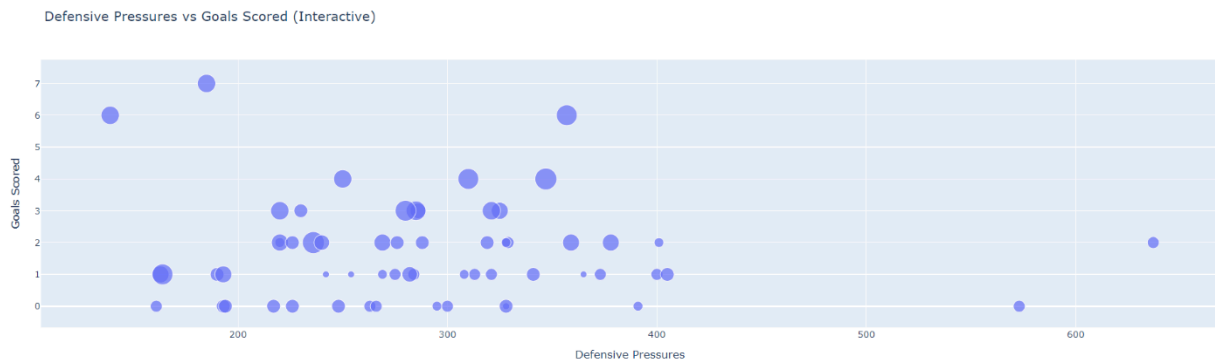


Fig3. Performance Optimization: Goals vs Shots



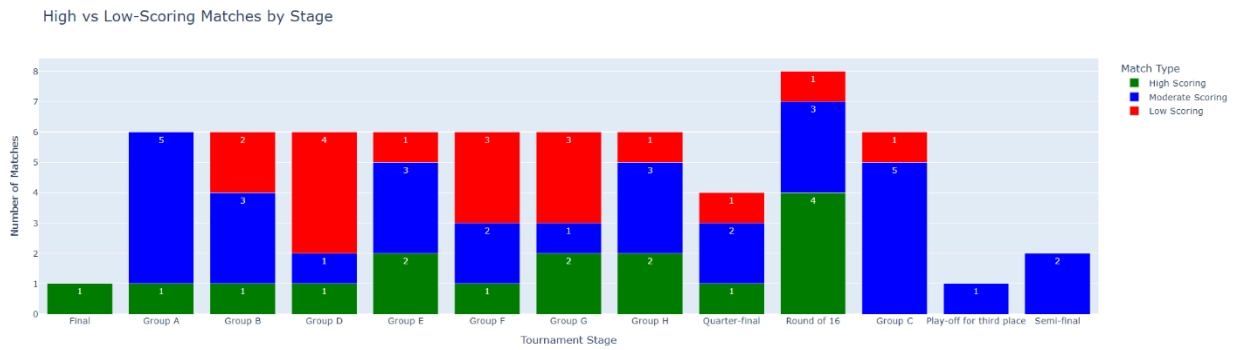


Fig5. High Scoring vs Low-Scoring Matches by Stage

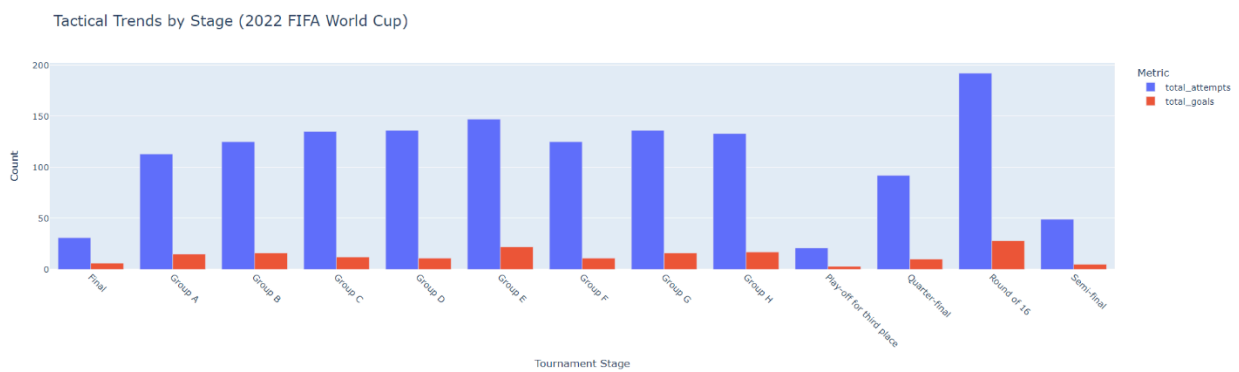


Fig6. Tactical Trends by Stage

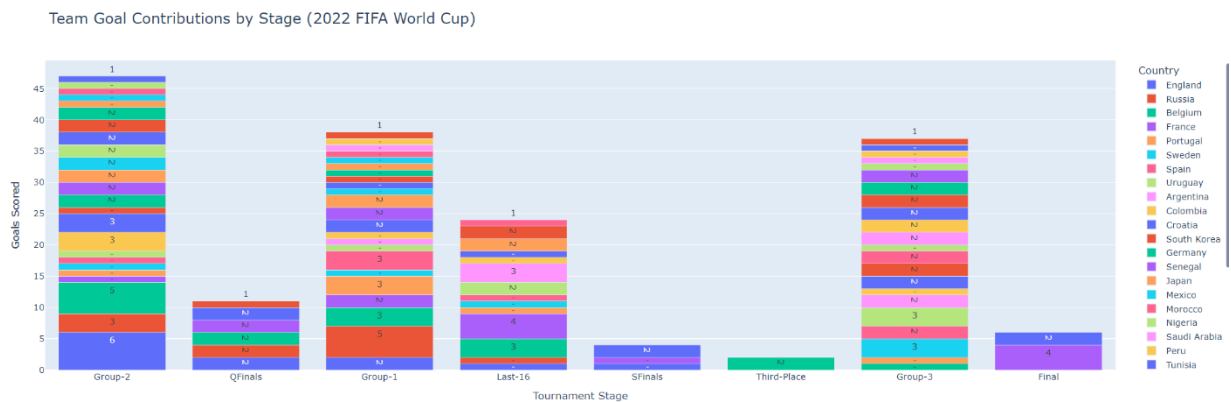


Fig7. Team Contributions by Stage

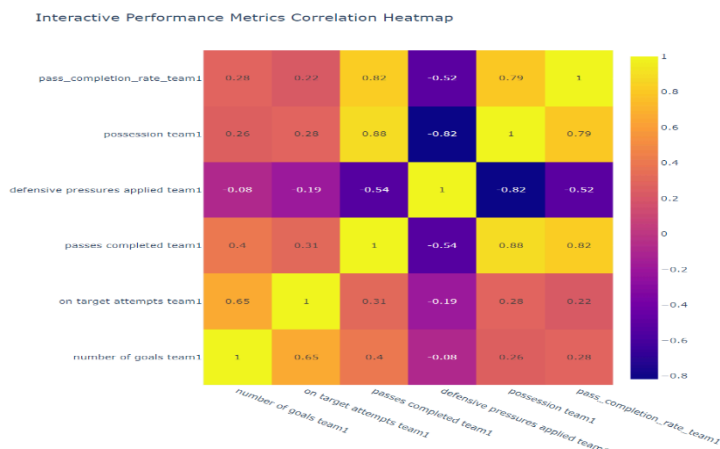


Fig8. Interactive Performance Metrics Co-relation Heatmap

6. Discussion/Future Work

Opportunities for Expansion

1. **Player-Level Analysis:**

FIFA and teams can improve their 2026 tournament preparation by analyzing individual player performance metrics like passes, defensive actions, and scoring efficiency, identifying rising stars and underperforming players.

2. **Predictive Models:**

Predictive models use machine learning and historical data to simulate match scenarios, predict tactical outcomes, and optimize lineups and substitutions, providing a futuristic approach to match strategy preparation.

3. **Enhanced Fan Data Integration:**

Real-time audience sentiment analysis using social media trends, in-game metrics, and fan feedback can enhance understanding of fan behavior, inform promotional strategies, and optimize global engagement.

4. **Extended Dataset Scope:**

Integrating regional league and qualifier data provides a comprehensive pre-tournament analysis, revealing tactical trends, emerging threats, and insights beyond the tournament itself, including lead-up preparations.

Potential Impact

By expanding the analysis, FIFA can:

- Offer tailored recommendations to teams, enhancing competitive balance.
- Provide a deeper understanding of fan preferences, creating targeted campaigns that amplify engagement.
- Establish the 2026 World Cup as a benchmark for data-driven tournament management, reinforcing FIFA's global leadership in football innovation.

7. Conclusions

Summary of Study

This project delved into the 2018 and 2022 FIFA World Cup datasets to uncover critical success factors and fan engagement trends. Metrics such as possession, goal conversion rates, defensive resilience, and fan behavior emerged as pivotal for team and tournament success.

Implications for Stakeholders

1. For Teams:

Teams can leverage these findings to refine their strategies, emphasizing possession and defensive robustness. Balanced gameplay, effective midfield control, and efficient finishing are critical for success in the expanded format of the 2026 World Cup.

2. For FIFA:

Insights enable FIFA to prioritize fan-centric initiatives, such as broadcasting high-stakes matches, improving engagement, and designing tailored marketing strategies to boost viewership.

Path Forward

1. Predictive Tools:

Developing machine learning tools can refine tactical planning, enabling dynamic adjustments to match scenarios.

2. Training Alignment:

Aligning team preparations with high-performance metrics—such as possession control and defensive efficiency—ensures readiness for the demands of a larger tournament.

3. Fan Engagement Strategies:

Leveraging fan sentiment data, FIFA can curate experiences that cater to global audiences. High-energy, high-scoring matches should feature prominently in promotional campaigns.

Through these measures, FIFA 2026 can redefine excellence in global sports, delivering an engaging, competitive, and innovative World Cup.

References

1. FIFA Datasets (2018 and 2022).
2. Statistical Libraries: Pandas, Matplotlib, Seaborn.

Appendices:

- **Code Repository:** [[EM 622 FINAL FIFA'.ipynb - Colab](#)]
- **Dataset Link:** [[Datasets](#)]