JAYPEE INSTITUTE OF INFORMATION TECHNOLOGY, NOIDA

B.TECH 5th SEMESTER

Report for Open Source Lab

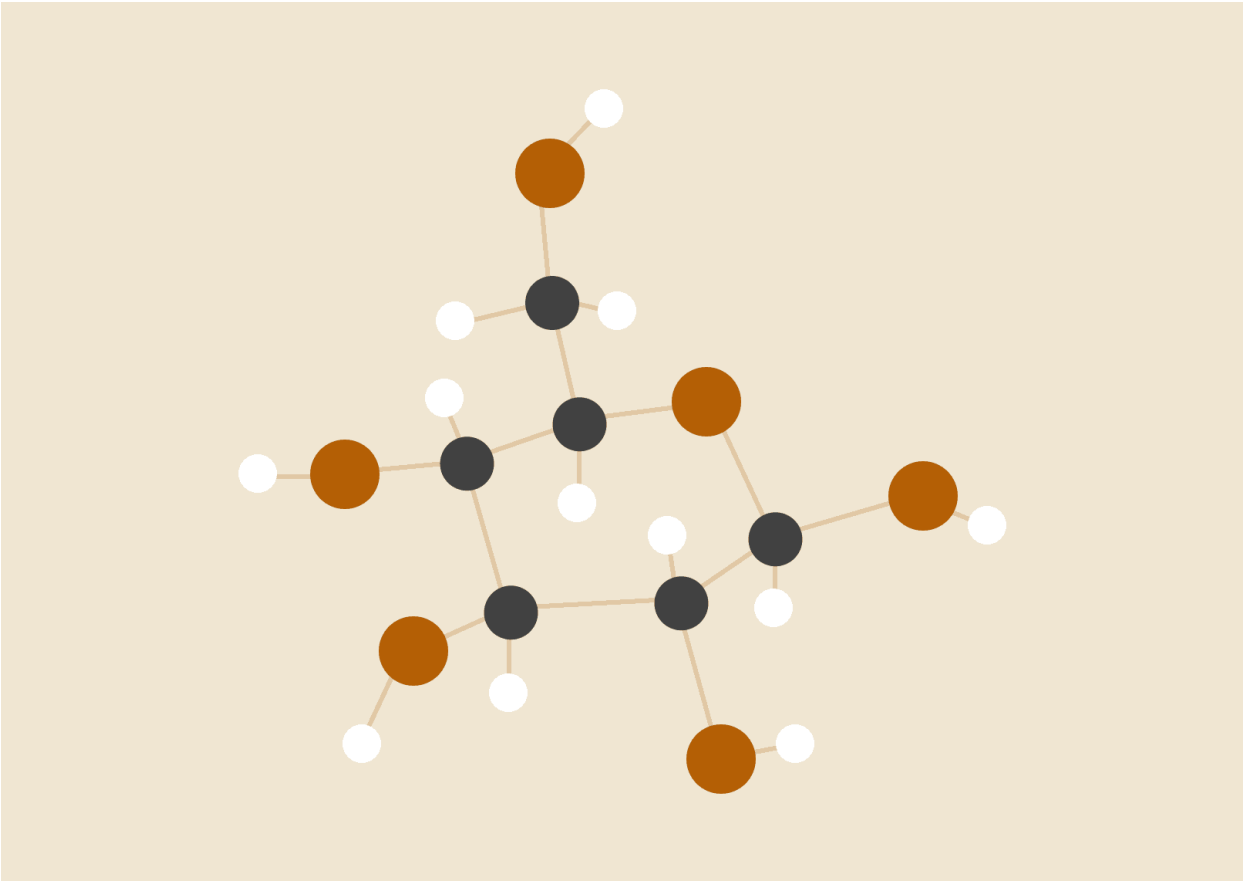
Project Based Learning

" Football Match Analysis"

Batch -B6

VAIBHAV VINAYAK	-	22103160
PRACHI NIGAM	-	22103175
DEV AGARWAL	-	22103177
MAYANK SINGH SIKARWAR	-	22103156

Football Match Analysis



VAIBHAV VINAYAK - 22103160

PRACHI NIGAM - 22103175

DEV AGARWAL - 22103177

MAYANK SINGH SIKARWAR - 22103156

21.11.2024

INTRODUCTION

Football analytics has become an essential tool for understanding and improving player performance, match strategies, and overall team outcomes. This project leverages data from the StatsBomb API to create a comprehensive analysis tool using Streamlit. It allows users to explore match data, visualize player and team metrics, and gain insights through interactive visualizations.

OBJECTIVES

To provide a user-friendly platform for analyzing football matches and player statistics. To visualize key aspects of matches, such as heatmaps, pass networks, and team formations. To explore predictive modeling techniques for analyzing football data.

FEATURES AND IMPLEMENTATION

By leveraging advanced technologies and data science techniques, we can extract valuable information from match data to inform strategic decisions, improve training regimens, and enhance the overall fan experience.

Key Features:

1. Data Collection and Integration:

    - Real-time tracking data: Collect player tracking data using statbombs api which stores data (position, speed, acceleration, etc.) collected using wearable devices or video analysis tools .
    - Event data: Capture events like goals, shots, tackles, passes, and fouls using this collected data by effectively preprocessing it.

2. Performance Metrics and Analysis:

    - Player performance metrics:
        - Individual player statistics (goals, assists, shots, passes, tackles, etc.)
        - Player workload.
        - Player movement patterns and heat maps
        - Player passing networks and efficiency
    - Team performance metrics:
        - Possession statistics
        - Shot conversion rates
        - Defensive pressure and intensity
3. Advanced Analytics:

    - Machine learning models: Use machine learning algorithms to predict match outcomes, identify key performance indicators, and detect patterns in player and team behavior.
    - Statistical analysis: Employ statistical techniques (e.g., hypothesis testing, correlation analysis) to quantify the significance of performance metrics.
    - Data visualization: Create interactive visualizations (e.g., charts, graphs, heat maps) to communicate insights effectively.

4. Tactical Analysis:
    - Formation analysis: Analyze team formations and tactical setups.
    - Set-piece analysis: Study set-piece strategies (corners, free kicks, penalties) to identify strengths and weaknesses.
    - Opponent analysis: Analyze opponent's strengths, weaknesses, and preferred tactics

to develop effective counter-strategies.

Implementation Considerations:

- Data Infrastructure: Use the statbombs api to collect match data.
- Data Quality and Cleaning: Ensure data accuracy and consistency through rigorous data cleaning and validation processes.
- Data Analysis Tools: Utilize advanced data analysis tools to extract meaningful insights.
- Visualization Tools: Employ powerful visualization tools(eg. Seaborn, Matplotlib, MPL-Soccer) to create engaging and informative visualizations.
- Domain Expertise: Collaborate with football experts (coaches, analysts, scouts) to interpret data and generate actionable insights.

Match Selection and Summary

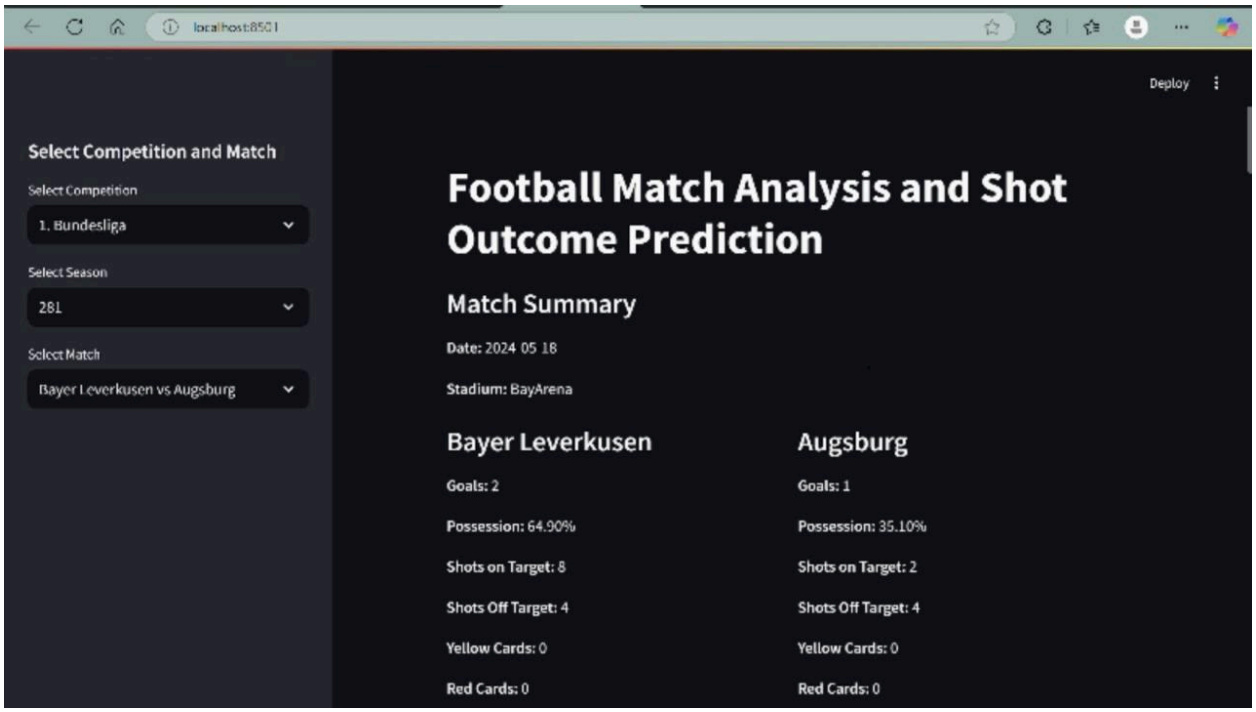
Users can select competitions, seasons, and specific matches through an intuitive sidebar interface. The app displays a detailed summary of matches, including:

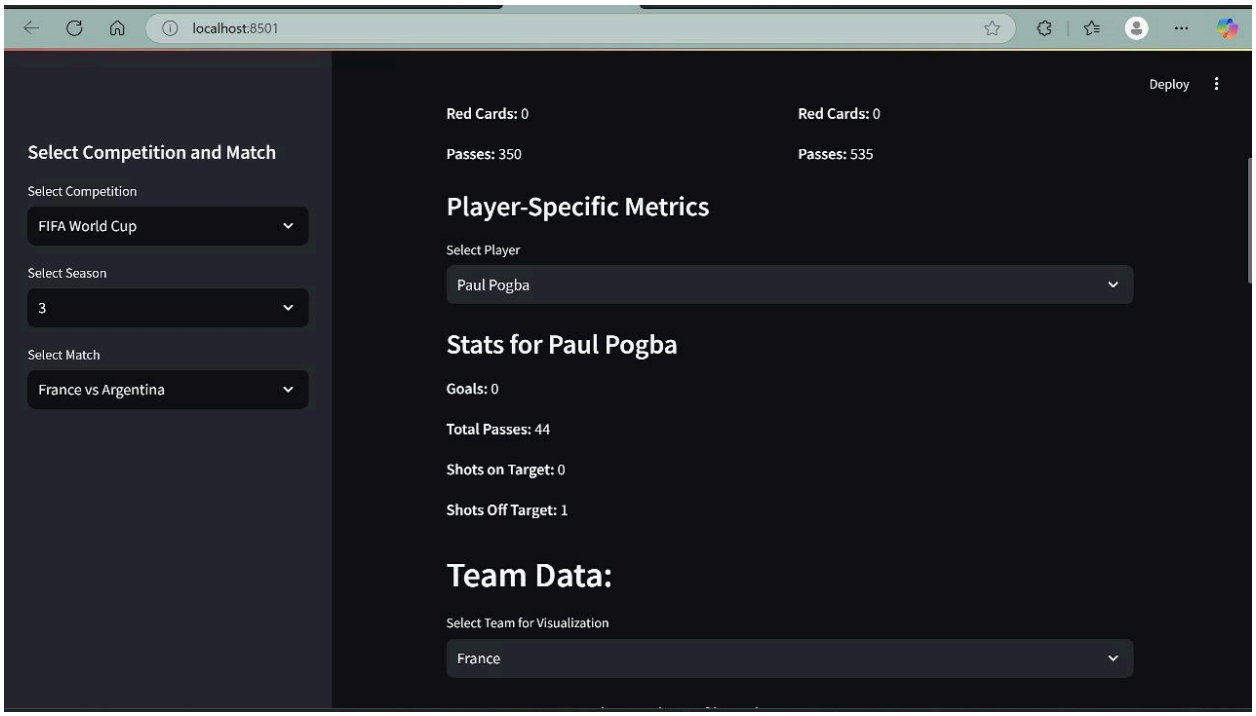
Goals scored by each team.

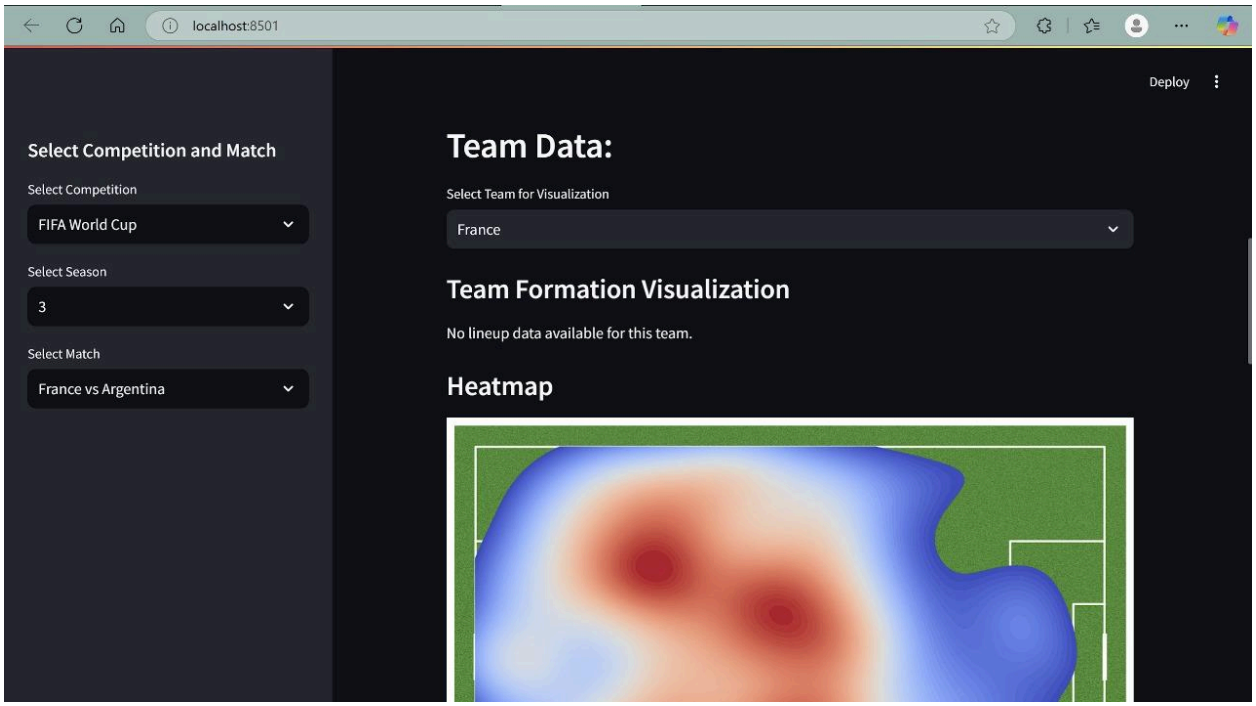
Possession percentages based on the number of passes.

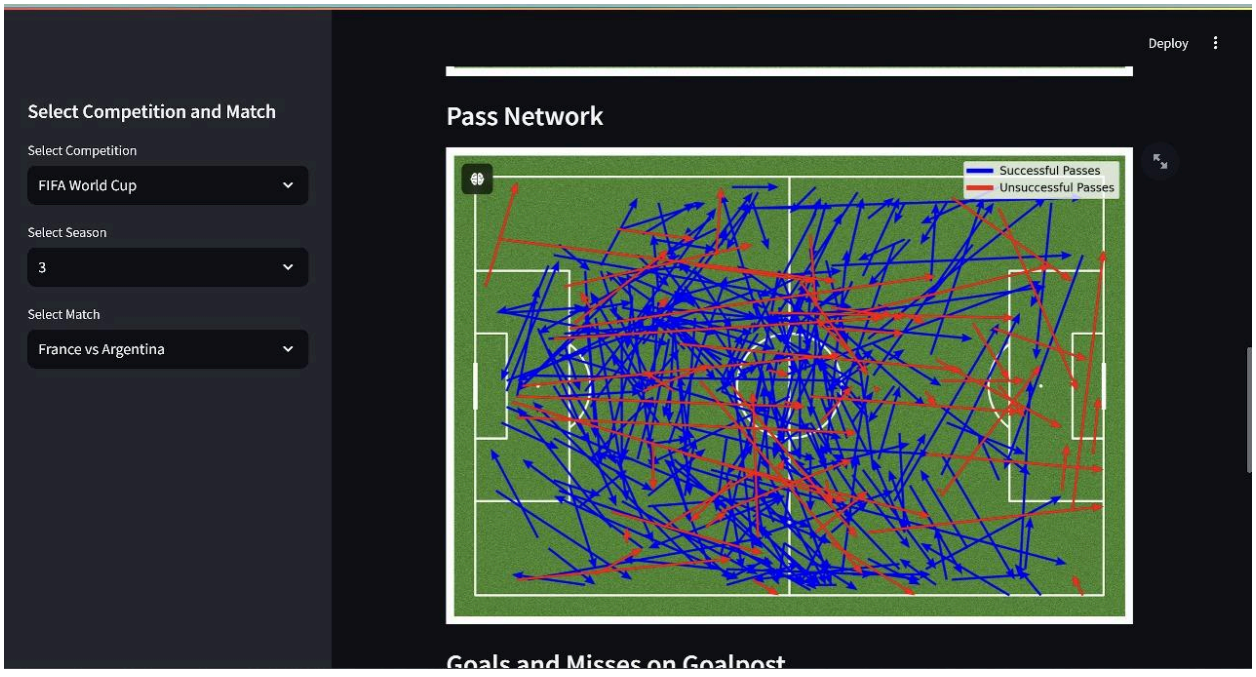
Shots on/off target, yellow/red cards, and total passes.

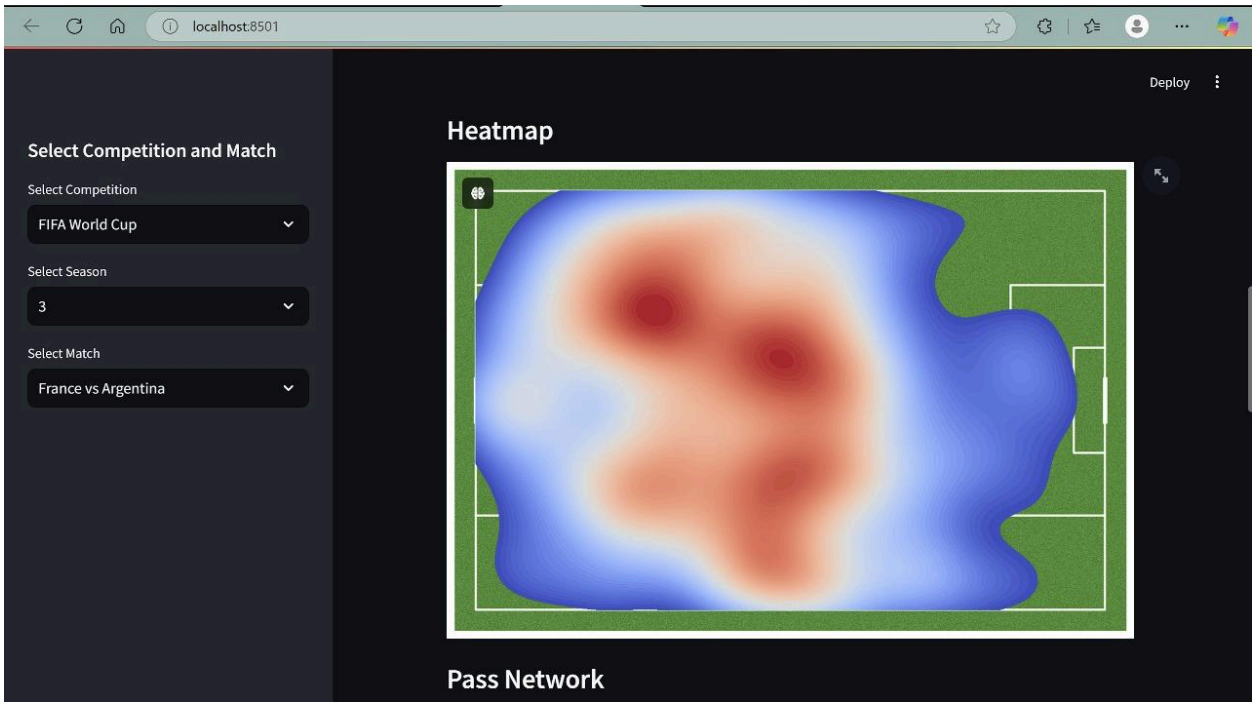
Streamlit Dashboard


localhost:8501
Deploy
Select Competition and Match
Select Competition
1. Bundesliga
Select Season
281
Select Match
Bayer Leverkusen vs Augsburg
Football Match Analysis and Shot Outcome Prediction
Match Summary
Date: 2024 05 18
Stadium: BayArena
Bayer Leverkusen
Goals: 2
Possession: 64.90%
Shots on Target: 8
Shots Off Target: 4
Yellow Cards: 0
Red Cards: 0
Augsburg
Goals: 1
Possession: 35.10%
Shots on Target: 2
Shots Off Target: 4
Yellow Cards: 0
Red Cards: 0


localhost:8501
Deploy
Red Cards: 0
Red Cards: 0
Passes: 350
Passes: 535
Select Competition and Match
Select Competition
FIFA World Cup
Select Season
3
Select Match
France vs Argentina
Player-Specific Metrics
Select Player
Paul Pogba
Stats for Paul Pogba
Goals: 0
Total Passes: 44
Shots on Target: 0
Shots Off Target: 1
Team Data:
Select Team for Visualization
France


localhost:8501
Deploy
Select Competition and Match
Select Competition
FIFA World Cup
Select Season
3
Select Match
France vs Argentina
Team Data:
Select Team for Visualization
France
Team Formation Visualization
No lineup data available for this team.
Heatmap


Deploy
Select Competition and Match
Select Competition
FIFA World Cup
Select Season
3
Select Match
France vs Argentina
Pass Network
Successful Passes
Unsuccessful Passes
Goals and Misses on Goalpost


localhost:8501
Deploy
Select Competition and Match
Select Competition
FIFA World Cup
Select Season
3
Select Match
France vs Argentina
Heatmap
Pass Network


localhost:8501
Deploy
Select Competition and Match
Select Competition
FIFA World Cup
Select Season
3
Select Match
France vs Argentina
Goals and Misses on Goalpost
1.0
0.8
0.6
0.4
0.2
0.0
Goal (1) / Miss (0)

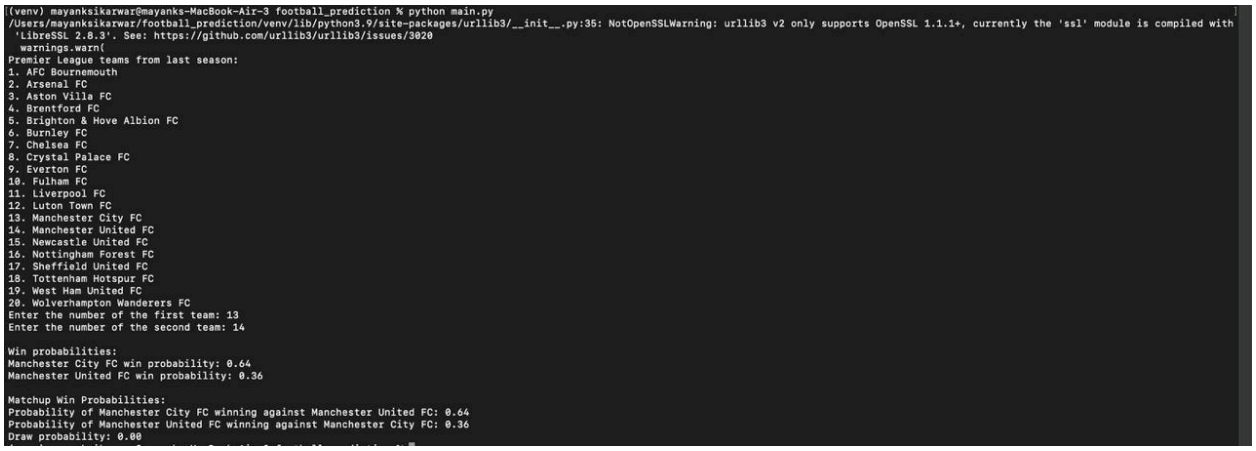
Player-Specific Metrics:

- Users can select individual players to view performance metrics such as:
- Goals scored.
- Total passes and assists.
- Shots on and off target.

Data Visualizations:

- Team Formation Visualization: Displays the formation of players on the pitch based on lineup data.
- Heatmaps: Provides a density map of team activity across the pitch, highlighting key areas of action.
- Pass Networks: Visualizes successful and unsuccessful passes using arrows on a pitch diagram.
- Goals & Misses Visualization: Plots shot locations on the goalpost to show areas of success and failure.

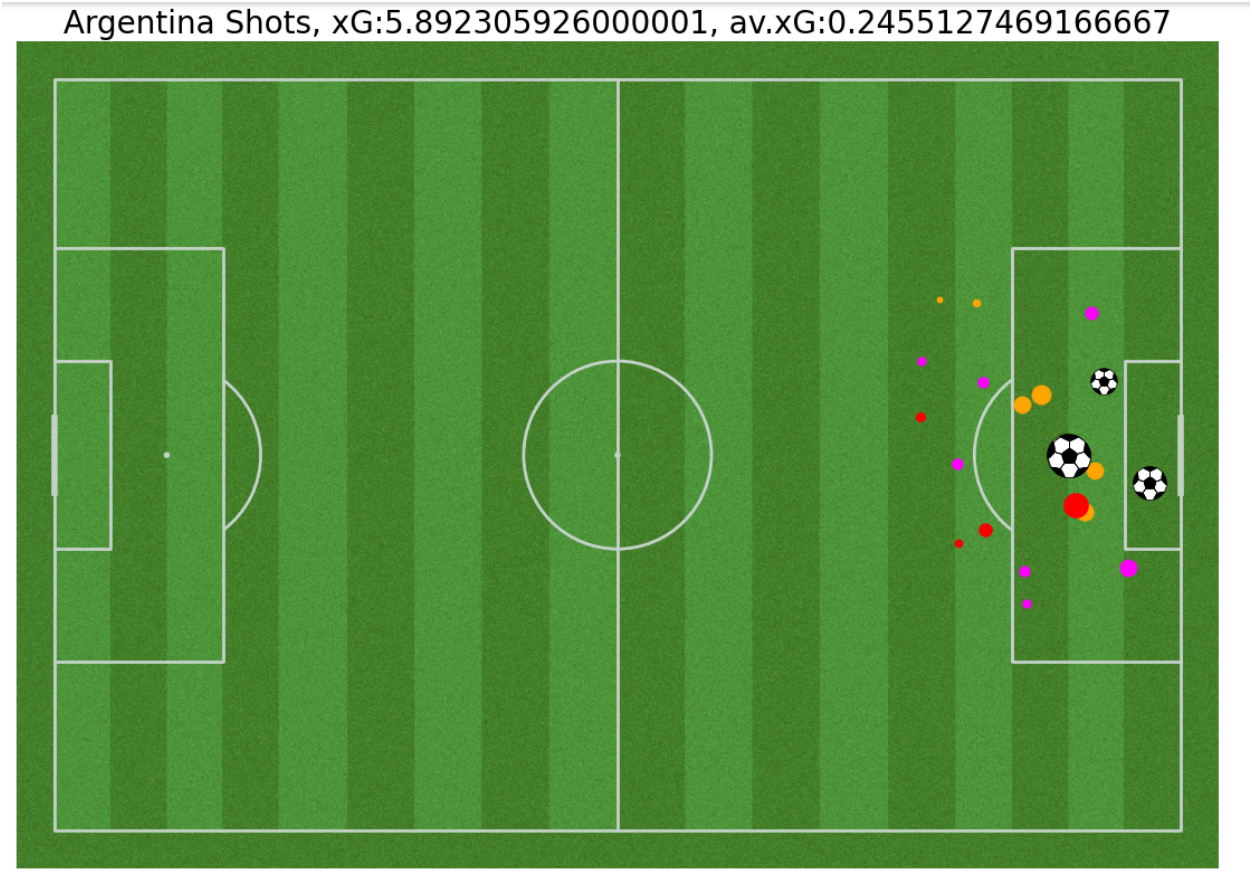
Match winner prediction model:

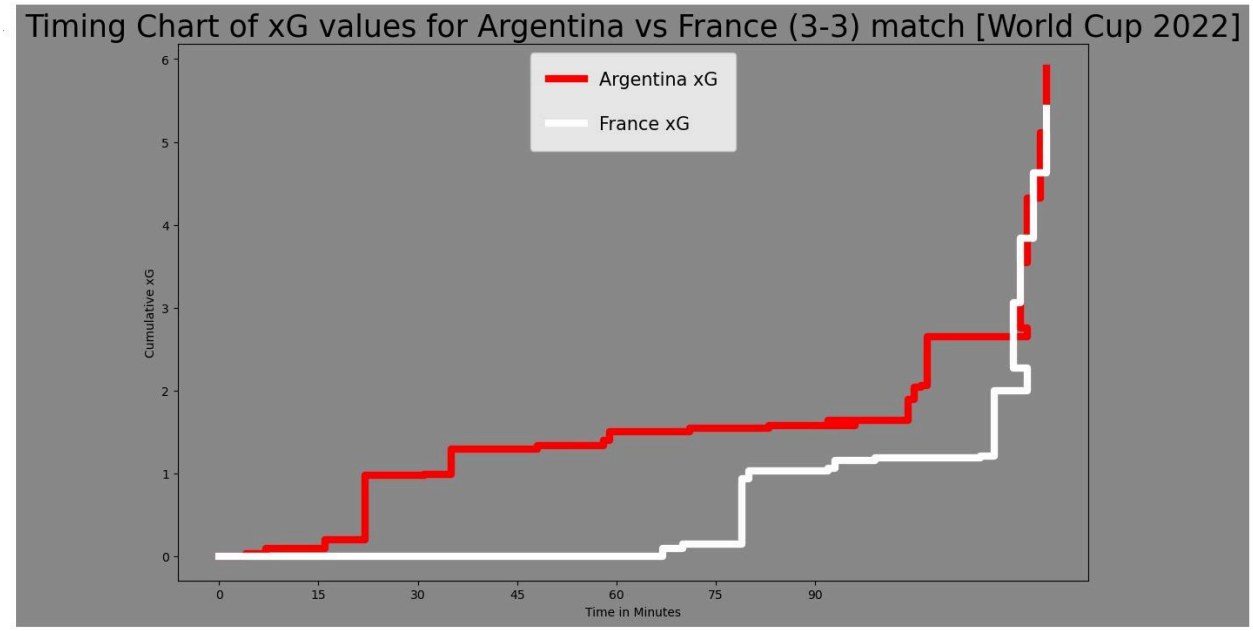


```
(venv) mayanksikarwar@mayanks-MacBook-Air-3 football_prediction % python main.py
/Users/mayanksikarwar/football_prediction/venv/lib/python3.9/site-packages/urllib3/__init__.py:35: NotOpenSSLWarning: urllib3 v2 only supports OpenSSL 1.1.1+, currently the 'ssl' module is compiled with
 'LibreSSL 2.8.3'. See: https://github.com/urllib3/urllib3/issues/3020
  warnings.warn(
Premier League teams from last season:
1. AFC Bournemouth
2. Arsenal FC
3. Aston Villa FC
4. Brentford FC
5. Brighton & Hove Albion FC
6. Burnley FC
7. Chelsea FC
8. Crystal Palace FC
9. Everton FC
10. Fulham FC
11. Liverpool FC
12. Luton Town FC
13. Manchester City FC
14. Manchester United FC
15. Newcastle United FC
16. Nottingham Forest FC
17. Sheffield United FC
18. Tottenham Hotspur FC
19. West Ham United FC
20. Wolverhampton Wanderers FC
Enter the number of the first team: 1
Enter the number of the second team: 19

Win probabilities:
AFC Bournemouth win probability: 0.36
West Ham United FC win probability: 0.27

Matchup Win Probabilities:
Probability of AFC Bournemouth winning against West Ham United FC: 0.36
Probability of West Ham United FC winning against AFC Bournemouth: 0.27
Draw probability: 0.36
```

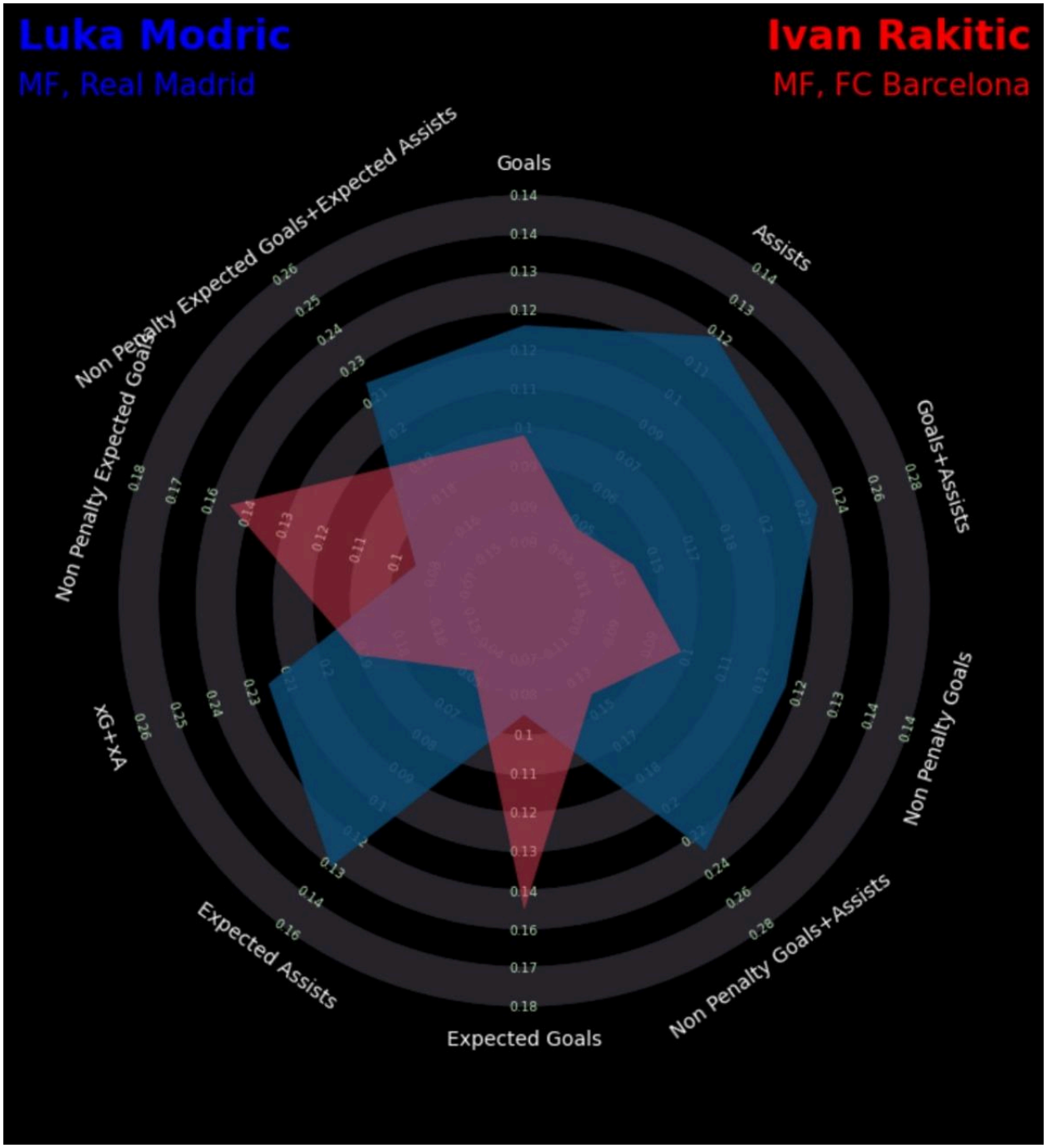
```
(venv) mayanksikarwar@mayanks-MacBook-Air-3 football_prediction % python main.py
/Users/mayanksikarwar/football_prediction/venv/lib/python3.9/site-packages/urllib3/__init__.py:35: NotOpenSSLWarning: urllib3 v2 only supports OpenSSL 1.1.1+, currently the 'ssl' module is compiled with
 'LibreSSL 2.8.3'. See: https://github.com/urllib3/urllib3/issues/3020
  warnings.warn(
Premier League teams from last season:
1. AFC Bournemouth
2. Arsenal FC
3. Aston Villa FC
4. Brentford FC
5. Brighton & Hove Albion FC
6. Burnley FC
7. Chelsea FC
8. Crystal Palace FC
9. Everton FC
10. Fulham FC
11. Liverpool FC
12. Luton Town FC
13. Manchester City FC
14. Manchester United FC
15. Newcastle United FC
16. Nottingham Forest FC
17. Sheffield United FC
18. Tottenham Hotspur FC
19. West Ham United FC
20. Wolverhampton Wanderers FC
Enter the number of the first team: 11
Enter the number of the second team: 18

Win probabilities:
Liverpool FC win probability: 0.82
Tottenham Hotspur FC win probability: 0.45

Matchup Win Probabilities:
Probability of Liverpool FC winning against Tottenham Hotspur FC: 0.82
Probability of Tottenham Hotspur FC winning against Liverpool FC: 0.45
Draw probability: -0.27
```

Other Visualizations:





Player comparison Radar Chart:



Tools and Technologies

StatsBomb API: Source of football match data.

Streamlit: Framework for creating the interactive web application.

mplsoccer: Library for football-specific visualizations.

Machine Learning Models:

Feature engineering and predictions were performed using Random Forest Classifier with feature scaling via StandardScaler.

Challenges

Incomplete or missing data for certain matches limited the scope of some visualizations. The shot outcome prediction model requires further development to provide meaningful results. Handling position data for lineup visualizations needed careful preprocessing due to variations in player roles.

Conclusion and Future Work

This project successfully demonstrates the potential of integrating data analysis, visualization, and machine learning in football analytics. While the core features like match summaries and visualizations work robustly, the predictive aspects can be further enhanced by:

Refining the machine learning model with more data and advanced techniques.

Incorporating additional features, such as player fatigue or defensive pressure, for better shot outcome predictions.

Expanding the platform to include comparisons between matches, teams, or players across multiple competitions.

Appendix

Source Code: Provided in the attached files (footballapp.py and notebooks for visualizations).

Tools Used: Python, Streamlit, mplsoccer, scikit-learn, pandas, matplotlib.