MIST 5635 Group Project Deliverable 1

By: Ryan Cullen, Alec Zerona, Warren Paintsil, and Will Hensley

Link to Kaggle dataset: Fifa World Cup 2022: Complete Dataset

Data Description

The dataset we chose from Kaggle contains comprehensive information about all matches from the FIFA World Cup 2022. The dataset provides a snapshot of each game's dynamics with various features, and we plan on adding more features to the dataset for easier interpretation (i.e. a binary feature - if goals scored > goals against then win = 1, else loss = 0 for each match). It will include not only outcomes for each team but also inferred tactical strategies implemented, game flow, and factors contributing to match results.

Key Features for Analyzing Team Success in FIFA World Cup 2022

The dataset's features cover various aspects of team performance, including offensive output (goals, attempts, assists), defensive effectiveness (goal preventions, forced turnovers), ball control and movement (possession, passes completed), set pieces and special situations (corners, penalties), tactical elements (line breaks, defensive pressures), and discipline (cards).

The key features we plan on using include:

1. **Number of goals (team1/team2)**
   - Direct indicator of offensive success and match outcomes

2. **Possession (team1/team2)**
   - Reflects a team's ability to control the game

3. **Total attempts (team1/team2)**

   o Measures offensive aggression and pressure

4. **On target attempts (team1/team2)**

   o Indicates the quality and accuracy of offensive plays

5. **Assists (team1/team2)**

   o Reflects teamwork and the ability to create scoring opportunities

6. **Passes completed (team1/team2)**

   o Indicates team cohesion and tactical effectiveness

7. **Crosses completed (team1/team2)**

   o Measures effectiveness in creating scoring chances from wide areas

8. **Goal preventions (team1/team2)**

   o Reflects defensive effectiveness

9. **Forced turnovers (team1/team2)**

   o Indicates defensive pressure and ability to regain possession

10. **Yellow cards and Red cards (team1/team2)**

   o Reflects discipline and potential impact on team performance

11. **Offsides (team1/team2)**

   o Can indicate offensive strategy and timing

12. **Corners (team1/team2)**

   o Potential for set-piece scoring opportunities

13. **Penalties scored (team1/team2)**

   o Direct impact on score and potential game-changing moments

14. **Completed line breaks (team1/team2)**

   o Indicates ability to penetrate opponent's defensive structure

15. Defensive pressures applied (team1/team2)

- o Reflects defensive strategy and intensity

By understanding these features we can gain an understanding on what contributes to a team's success in the tournament. We want to consider creating composite metrics or conducting detailed descriptions to reveal more complex relationships between the features and the team success.

Problem Statement

The FIFA World Cup 2022 dataset seems complex with 88 columns; after analyzing all features, we have decided to investigate the key factors that influence team performance and match outcomes throughout the FIFA 2022 tournament. Our analysis will focus on answering this question: **what are the most significant statistical indicators of a team's success in the tournament?** Some questions that may help answer this main question include the three following questions:

1. How do possession and passing statistics correlate with match outcomes?
2. Is there a relationship between disciplinary actions (fouls, cards) and a team's progression in the tournament?
3. Can we identify playing styles or tactical approaches? If yes, which were most effective?

We will use a variety of statistical methods we have learned in class. We are still discussing which models we would prefer which will be discussed and showcased in the next deliverables.