

# Premier League Prediction Analysis

...

September, 2022

# Project Overview

Sports statistics and analytics has been increasingly popular across all sports as analytical tools have become more available and powerful.

Higher volumes and precision of data has become more readily available but the dynamics of team sports has remained the same.

Our goal with this project is to use historical scoring data from the Premier League in English Football (or Soccer) to create predictive models that can influence sports betting, and team decision making.

# Applicability and Metrics Used

## Useful Team Statistics:

- Total Wins/Losses
- Home Wins/Losses/Draws
- Away Wins/Losses/Draws
- Goal Differential (gd)
- Goals Scored/ Goals For (gf)
- Goals Allowed/ Goals Against (ga)
- Expected Goals (xg)
- Expected Goals Against (xga)
- Player Salary
- Yearly Expenditure
- Yellow Cards
- Red Cards
- Fouls
- Possession Time
- Clean Sheets
- Corners
- Clearance off line

## Other Questions to Explore:

- How much does a red card affect the probability of winning a soccer?
- Are teams more successful in home or away games?
- Is there really a best league?
- How would “all time great” teams fare in today’s leagues?
- Can you predict the Ballon D’Or Winner?
- Does wealth = success for a team?
- Can you predict the world cup winner?
- Do “better” players really give a better chance of winning the world cup?
- Do sponsors indicate anything about how well a team performs (i.e. Nike vs Adidas vs Puma)
- What statistic drives viewers per game/ ticket sales for popular teams?
- Does age relate to points in any way?

# **Project objective:**

Our specific objective after digging through available data is to:

1. Create a predictive model for Premier League team scoring and...
2. Create a random forest model that pits each team against each other and predicts outcomes for the 2022 season.

# Data Sources

The majority of our data is pulled from the website [Sports Reference](#). More specifically the soccer portion of this website know as [FB REF](#). This website has vast amount of sports data ready to be downloaded via CSV files. The data can be filtered by player, league, individual clubs, seasons, and includes many different comparison tools as well.

The screenshot displays the FBRef website, a comprehensive source for football statistics. The header includes navigation links for various sports (Baseball, Football, Basketball, Hockey, Soccer) and a search bar. The main content area is divided into several sections:

- Football Stats and History:** A section for searching over 193,000 players, with a search bar and a "Search" button. Below this, there are links to "Trending Player Pages" and "Player Comparison".
- Football Squads:** A section for viewing club statistics, with a "View a Club:" dropdown menu and a "Go!" button.
- What's Happening:** A section for the latest news and updates, including "BIG updates coming to FBRef.com soon", "Site News", "Recent blog posts", and "Join FBRef's Linker Program".
- FBRef.com Coverage Totals:** A section showing the total number of countries, competitions, squads, players, and match reports covered by the site.
- Sports Reference Blog:** A section for the latest blog posts, including "Goal Logs Added to FBRef", "Capology: Worst Data Explainer", and "NWSL and U.S. Draft Data Now on FBRef".
- Upcoming Dates:** A section for upcoming dates, including "October 2: 2022 NWSL regular season ends", "October 9: 2022 M.L.S. regular season ends", "October 16: El Clásico: Real Madrid vs FC Barcelona", "October 28-30: 2022 NWSL Championship weekend", and "November 1-2: Final matchday of UEFA Champions League group".