# Title

The Player Performance Lab

By: Daniel Tshiani

# Description of the problem to be analyzed

It is difficult to scout soccer players in the US due to the size of the US. This app facilitates scouting for recruiters, coaches, and fans when it comes to U.S. soccer.

# Literature review

**He, M., Cashucho, R., & Knobbe, A. (2015).**  
Football player’s performance and market value. In Machine Learning and Data Mining for Sports Analytics Workshop at ECML/PKDD 2015. Retrieved from <https://dtai.cs.kuleuven.be/events/MLSA15/papers/mlsa15_submission_8.pdf>  
This study developed a model to estimate the market value of La Liga players, focusing on forwards due to positional biases found in data from Transfermarkt and WhoScored. By narrowing the scope, the authors aimed to improve predictive accuracy for offensive players. This paper help me brainstorm potential predictors to include in my model to make salary predicitions.

**Lee, S., & Harris, J. (2012).**  
Managing excellence in USA Major League Soccer: An analysis of the relationship between player performance and salary. Soccer & Society, 13(4), 533–550. https://doi.org/10.1080/13606719.2012.674389  
This paper examined the relationship between performance metrics and salary among MLS players. It identified salary distribution disparities, notably highlighting the “star player effect,” where a few high-profile players earn disproportionately more than others. It introduced me to the idea of income inequality in the MLS and made me realize OLS regression will be a difficult approach for a linear model.

**Li, C., Kampakis, S., & Treleaven, P. (2022).**  
Machine learning modeling to evaluate the value of football players (arXiv:2207.11361). arXiv. <https://arxiv.org/pdf/2207.11361>  
This paper employed machine learning algorithms to explore how both on-field and off-field factors (such as fan popularity) influence player salaries. It illustrates the complexity of modeling player value and underscores the importance of including commercial appeal in addition to performance metrics. This paper open my eyes to the idea that key performance indicators are not limited to metric on the field.

# Proposed data sources

American Soccer Analysis API released a package called “Itscalledsoccer” in April of 2025. This package frequently updates U.S. soccer data. I will use this package as my primary source of data. Within this package, data ranges from men and women professional soccer data all the way to boys and girls academy data. For each league, there are data frames on individual player stats, team stats, salaries, gameflow, managers, referees, and more.

If needed, I will join data from API-Football and I will also scrape the web for additional data. Other papers have used transfer market data and data from the who scored website, so I will keep those in mind if more data is needed.

# Use Cases & Actor Descriptions

* actors can be
  + soccer scouts/recruiters
  + coaches and analyst
  + soccer fans
* What questions will the app answer?
  + Performance-related questions like how do 2 players compare in terms of goals, assists, and key passes?
  + Who are the most effective players given a position, league, age group, etc…
  + Can we predict the current or future market value of young players?
* Dynamic
  + Users will be able to filter league, age, position, stats, etc. They will be able to interact with different variables, and modify statistical models (for example to select features for market value predictions)

Use case summary: A soccer scout is looking for potential signings and wants to evaluate a player’s performance metrics, compare them to similar players, and predict their market value. Using the app, they can filter by position, league, and nationality, visualize player performance, and use a regression model to estimate transfer value. The app helps them make data-driven recruitment decisions.

# Ethics

All the data used in this app comes from publicly available sources, and I made sure to follow any terms and conditions that apply. There were no restrictions violated, and I didn’t collect or use any private or sensitive information.

# General Layout

The app will feature a clean, user-friendly interface designed to guide users through the process of exploring, comparing, and analyzing soccer player performance data. A sidebar or top navigation bar will allow users to filter data by key attributes such as league, season, position, nationality, and age. These filters will dynamically update the main content area, allowing for an interactive and personalized user experience.

The main panel will be divided into several sections or tabs. The Overview tab will display summary statistics, top-performing players, and quick insights based on selected filters. In the Player Comparison tab, users will be able to select two or more players and view side-by-side comparisons using radar charts, bar graphs, and stat tables. The Explore Players tab will offer a searchable and sortable table of player stats, allowing deeper dives into individual performance. Finally, the Market Value Prediction tab will let users choose which performance metrics to include in a predictive model that estimates a player’s current or future market value. A final tab will provide details about the app’s data sources, ethical considerations, and project background.

# Planned options for:

User Selection and Manipulation:

Users will be able to interactively filter the dataset based on several variables, including player position, league, nationality, age range, and season. The app will also allow users to sort and search player stats within an interactive table and select specific players for comparison. In addition, users will be able to customize which performance metrics to include in the analysis, such as goals, assists, key passes, minutes played, and more.

User Exploratory Numerical/Graphical Analysis

the app will support exploratory data analysis through both summary statistics and interactive visualizations. Users will be able to view leaderboards, distribution plots (e.g., histograms of goals scored by position), and scatterplots comparing selected metrics. Comparative visuals will allow users to evaluate individual player performance relative to peers, league averages, or specific filters they've selected.

User Statistical Models and Test

The app will include a simple linear regression model that predicts a player's market value based on selected performance indicators. Users will be able to choose which variables to include in the model and visualize the model output, including predicted values and model fit diagnostics. If time permits, additional statistical tools—such as hypothesis testing or correlation matrices—may also be added to support deeper analysis of player performance trends.

# Responsibilities

Daniel will be responsible for the entirety of the project. The aim is to submit it by May 6th.