**Project Proposal**

**2023 – 2024 Soccer Player Statistics and Their FC25 Ratings**

# Introduction

EA Sports FC is a soccer video game franchise who’s ratings of players hold a large weight amongst fans and football communities around the world. Statistics are a key indicator in a soccer players performance, and have a great deal to do with their ratings in EAFC. In this project, I plan to use FBref[[1]](#footnote-1) statistics and performance metrics to see if there is a correlation between the players 2023-24 statistics and the players FC25 ratings that I obtain from Kaggle[[2]](#footnote-2), as well as find out which clubs and leagues have the highest average FC25 rating.

# Data

The first data source I will use in this project is FBref1, a website that tracks in-depth statistics about every player that plays in the top 5 leagues in the world (Premier League, La Liga, Serie A, Bundesliga, Ligue 1). The dataset I will be using contains 2852 rows of match statistics of each player for the 2023-2024 season. The actual dataset contains around 40 columns, but I only plan on using 10 of the most relevant features, seen in the data dictionary. I plan to scrape this data from FBref1 using the Python package called Selenium.

The second data source I plan to use is from Kaggle2, which is a downloadable CSV file that contains the EA Sports FC25 overall ratings and physical/player attributes for every player in the world. This was scraped from the EA Sports Website by Davis Nyagami. The CSV file contains 16,159 rows of these ratings and statistics for each player, based on the 2023-2024 season. The file has 58 columns, but again I will only be using 10 of the most relevant features.

I plan on merging the two datasets using horizontal integration on the primary key of Player Name. I will **intersect** the datasets using an inner join, so that only players that are in both datasets (play in the top 5 leagues) will be included in the final dataset. As a result, I will be left with the 2852 players from the FBref1 dataset, with their FC25 ratings and attributes added on. The data will need to be cleaned and transformed in many ways. A few examples include, the renaming of columns, the creation of a Goal Contributions column (Goals + Assists), and a summation of the statistics for “duplicate” players (players who were transferred to a new team midseason, and therefore appear twice).

A discrepancy between the two datasets that I anticipate is if a player changed teams between last season and this season. The EAFC data will have the up to date team but I will be using the FBref1 team data because this is the team that they played for during the 2023-24 season and therefore have achieved these statistics with. A limitation of this analysis is that I am only examining offensive statistics (Goals, assists, and expected goals) which leaves defenders and goalkeepers as possible outliers when it comes to their FC25 ratings compared to their goal contributions.

*Table 1 Data Dictionary*

|  |  |  |  |
| --- | --- | --- | --- |
| Column | Type | Source | Description |
| Player | Text | Both | The name of the player |
| Nation | Text | Both | The nation the player represents in international competitions |
| Position | Text | Both | The position(s) the player plays |
| Club | Text | Both | The club the player plays for |
| League | Text | Both | The league the club is in |
| Age | Numeric | Both | The age of the player |
| Matches Played | Numeric | FBref | The number of matches the player played in the 2023-2024 season |
| Goals | Numeric | FBref | The number of goals the player scored in the 2023-2024 season |
| Assists | Numeric | FBref | The number of assists the player got in the 2023-2024 season |
| xG | Numeric | FBref | The number of goals the player was expected to score in the 2023-2024 season |
| Overall | Numeric | Kaggle | The overall rating of the player in the FC25 game |
| Height | Numeric | Kaggle | The height of the player in inches |
| Weight | Numeric | Kaggle | The weight of the player in pounds |
| Pace | Numeric | Kaggle | The pace rating of the player in the FC25 game |
| Shooting | Numeric | Kaggle | The shooting rating of the player in the FC25 game |
| Passing | Numeric | Kaggle | The passing rating of the player in the FC25 game |
| Dribbling | Numeric | Kaggle | The dribbling rating of the player in the FC25 game |
| Defending | Numeric | Kaggle | The defending rating of the player in the FC25 game |
| Physicality | Numeric | Kaggle | The physicality rating of the player in the FC25 game |
| Play Style | Text | Kaggle | The play style of the player |

# Proposed Analysis

This project aims to analyze the effects of players statistics from the 2023-24 season on the players EA FC25 ratings. I also would like to explore some broader ideas about the clubs and leagues, as well as the relationship of different physical and player attributes with their number of goals and assists achieved last season. My potential research questions include:

* Is there a strong correlation between a players goal and assist tally and their FC25 overall rating? – My hypothesis is yes, but I would like to see how strong the correlation is.

* Which club has the highest average FC25 player rating?
* Which league has the highest average FC25 player rating?
* Is there a strong correlation between a players age and their FC25 overall rating?
* Which player/physical attributes are most important for a player with a high amount of goal contributions?

1. 1 <https://fbref.com/en/comps/Big5/2023-2024/stats/players/2023-2024-Big-5-European-Leagues-Stats> [↑](#footnote-ref-1)
2. <https://www.kaggle.com/datasets/nyagami/ea-sports-fc-25-database-ratings-and-stats?select=male_players.csv> [↑](#footnote-ref-2)