**Craig Dacosta**  
**Data Science Programming CMP-262**  
**Final Project Summary**  
**English Premier League Player Statistics: 2023-2024 Season**

As a soccer fan, particularly of the English Premier League (EPL), I chose to base my project on player statistics for the 2023-2024 season. The dataset included 22 columns and over 500 rows. Initially, I wasn’t which questions to explore, but a recent discussion with a friend about players performance expectations for the season led me to investigate several questions:

1. **Who were the top ten players last season based on goals scored?**  
   The bar chart showed that Erling Haaland from Manchester City was the leading player, scored over 17 goals during the season. Manchester City won the league that year and history shows that the number one player usually comes from the number one team.
2. **Did the top ten players meet or exceed their predicted/expected goals based on the 2022-2023 season?**  
   The analysis showed that 70% of the top ten players exceeded their expected goals. This is a valuable statistic for fantasy league players when selecting players for their teams.
3. **Did younger, more physical but less experienced players score more goals than older, less physical but more experienced players?**  
   Using data from all 583 players in the dataset, I categorized players into two age groups: 17-25 (younger) and 26-39 (older). The results revealed that older players (26-39) scored more goals on average than their younger counterparts.

**Potential Target Audience:**  
The insights from this analysis could benefit various groups, including:

* **Sports commentators and analysts**: To enhance discussions and provide insights during matches.
* **Sports betting shops**: To refine odds and predictions.
* **Fantasy league players**: To make informed decisions about team selection.
* **Fans**: To better their understanding and engagement with the games.

Things that I could continue to analyze are, which players played full 90 minutes games to analyze their fitness, percentage on shots on target, midfielder with the most assist, fouls committed etc. All these can be beneficial information for Fans, coaching staff and recruits.

For this project, I used Jupyter Notebook for exploratory data analysis, data cleaning, and overall analysis. Key Python packages included pandas for data manipulation, matplotlib for visualization, and numpy for numerical computations.