Target Variable

The target variable in this project is the fantasy football points for the 2023 NFL season. The goal is to predict the total fantasy points a player will accumulate based on their historical performance data from previous seasons. Fantasy football points are calculated from key player statistics such as yards from scrimmage, touchdowns, receptions, passing yards, and other vital metrics. These statistics, which directly influence the overall fantasy points a player can achieve, form the core outcome of interest in the predictive model.

Predictors

The predictors for the project are performance statistics from the 2019-2022 seasons, as these contribute significantly to fantasy football scoring. Yards from scrimmage, which combine both rushing and receiving yards, are a critical predictor, as are the total touchdowns a player scores, whether through rushing, receiving, or passing. Receptions, particularly important for points-per-reception (PPR) leagues, offer another key metric in determining player value. Rushing attempts, passing yards, and passing touchdowns provide further insights into a player's offensive contributions, especially for quarterbacks. Other predictors include targets, fumbles, and interceptions, all of which can influence a player's fantasy score. These metrics are chosen because of their direct and significant impact on fantasy football points, and studying historical trends in these statistics helps forecast future player performance.

Exploration of the Dataset

The dataset comprises player statistics from the 2019 to 2022 seasons, focusing on the top 200 players each year. This selection aligns with the structure of a typical 12-person fantasy football league, where each team has 16 players. The data from each year is stored in separate CSV files but also compiled into one "master" Excel workbook, allowing for more streamlined data analysis across all years. Each row in the dataset represents a player's season, and the dataset will ultimately contain around 800 rows, with 200 players for each of the four years.

The variables in the dataset include player name, position, team, games played, and various performance metrics. The key variables are player names, which is categorical, as are the position and team that the player represents. Games played, rushing attempts, and targets are recorded as integers, while all performance-related statistics such as yards from scrimmage, receptions, touchdowns, passing yards, passing touchdowns, fumbles, and interceptions are continuous variables. Each row in the dataset captures the performance of one player in one season, and the dataset will feature around 15 to 20 columns that account for these metrics.

The dataset serves as the foundation for building machine learning models that will predict fantasy points for the 2023 season. By leveraging historical data from 2019 to 2022 and focusing on the most relevant predictors for fantasy football scoring, the project aims to provide accurate forecasts of player performance for fantasy football enthusiasts. The selection of 200 players per year ensures the analysis focuses on the most impactful athletes in a fantasy league, and the data structure supports detailed trends analysis across multiple seasons.