The NBA awards the Most Valuable Player (MVP) at the end of every regular season. The award normally goes to the player that contributed the most to their team’s success. The winner of this award usually plays for a team that is ranked among the best in the NBA, and their statistical contribution is often a large part of that reason. A group of league associates votes for the regular season award, splitting their votes for every player between the top five. Whichever player receives the most first-place votes wins the award. Before a season begins, different analysts from around the league and sports stations make predictions based on a player’s past performance to determine who will be the MVP for the upcoming season. These predictions are then altered as the season progress, and players either play up to the preseason hype or are outplayed by someone that was not talked about. The goal of my capstone project is to look at past MVP winners, assign weighted values to different statistical categories based on their correlation to the odds of winning the award, and with that in place, try to predict who will be the MVP for the current season.

This data set contains the Assists Per game, Minutes Per game, Points per game, and other key stats used in tracking NBA players’ individual impact on the team. The data included in the model can be used to predict the next NBA MVP or to predict who will lead the league in a certain statistic. In recent years the NBA has started to rely heavily on data analytics, using historical data to create new offensive and defensive strategies. One of the most notable analytic shifts in the NBA was seen with the Houston Rockets in the last 6 years. The team found that based on historical data, the mid-range shot was the least efficient shot in basketball and started to discourage their players from taking that shot. The team eventually built its entire offense based on that methodology, recruiting players and coaching staff that fit the metric. The data can also be used to analyze the types of shots favored over the years.

The dataset contains various stats and metrics used to measure players’ contributions over an NBA season. The data was acquired from Kaggle, and it is the NBA data set for 1982 – 2022 NBA player statistics with MVP votes. In order to solve this problem, I plan to sift through the data and compare the results to past NBA MVP winners. Once I’m sure that the data can replicate historical results, I can then use it to predict future results. I plan to use key metrics such as offensive rating, win share, and past MVP votes to predict future winners. By looking closely at what stats have historically contributed to MVP votes, I expect to be able to then predict future award winners based on career stats.

This MVP prediction can be used by anyone interested in the NBA; the dataset, once completed, can be viewed by fans of both data science and basketball, as well as fans of solely basketball. The goal of this set is to return results that are easy to read and understand so that a friend group with no programming knowledge can run the code and use the predictions for their own fantasy leagues and in-season bets. The data will also be able to provide odds of a player winning the award based on earlier assigned weighted statistical values; the odds column will be helpful for persons interested in making bets and people simply interested in tracking their favorite players’ position in the race. Once the model has been trained, it should be able to take on updated data from the NBA season and make accurate predictions using the new data as well as the old data.