**Introduction and Motivation:**

I’ve been a big fan of the NBA since I was a kid. I remember going through the newspaper each day and reading the box scores. Looking for who has the best stats and how they impacted the game. As I’ve gotten older, I’ve learned about advanced basketball statistics that give new insight into the game. Which led me to think, what goes into the recipe that creates a great player? I want to know which stats correlate the most with being a valuable basketball player, and other factors that may impact that.

**Methods:**

I went to a few websites to find the data that I used. I got per game stats and advanced stats from Basketball Reference’s website and got that by copying and pasting csv data into an excel sheet, which I then moved into a Data Frame in python. I also tried to get data from Cleaning the Glass, another basketball website with advanced stats, but unfortunately, the names of some players were not listed the same which was causing some big issues when I tried merging all this data into one dataset. So, I stuck with the stats I got off Basketball Reference’s website instead. I also narrowed down the list of players to only players that averaged at least 15 minutes per game, so there would be less outliers based on small sample sizes.

**Results:**

Which of the five main statistics (Points, rebounds, assists, steals, blocks) are correlated the most with a players’ value?

Points vs vorp Points vs WS

A graph of blue dots

Description automatically generatedA blue dots on a white background

Description automatically generated

Points had the strongest correlation between one of the five main stats (points, rebounds, assists, steals, blocks) and Vorp, which stands for value over replacement player. It also had the highest correlation between one of the 5 main stats and win shares, which is an advanced stat that estimates how many wins a player contributed to over the season. The correlation coefficient was .67 and .62, respectively, indicating a pretty strong positive linear relationship.

Next, I wanted to see what type of shot would be best for a player to shoot to contribute to their value: 2-point shots, 3-point shots, or free throws.

FTA to vorp

A blue dots on a white background

Description automatically generated

3PTA vs vorp

A diagram of a number of dots

Description automatically generated

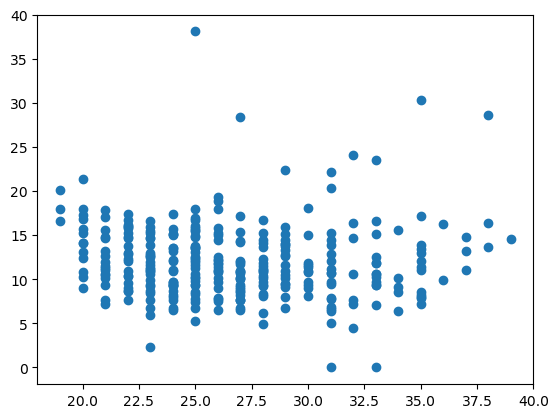
2PTA vs vorp

A diagram of a number of dots

Description automatically generated

As we can tell from these graphs, shooting free throws have a stronger correlation to a player’s value above replacement player compared to 3-point shooting and 2-point shooting. I thought it was interesting that 3-point shooting was the lowest, considering the trend in the NBA of more and more 3 pointers being shot every season. A player may look at 3-point shooter as their best offensive role to play, but perhaps driving in more and drawing fouls would allow them to be more valuable.

Age vs TOV%



For this question, I was curious whether age has an impact on a player’s ability to take care of the ball. Younger players are known for making more mistakes, as they are still learning how to play in the NBA. However, according to this, there is no correlation between age and the estimate of number of turnovers committed every 100 plays.

TOV vs VORP TOV% vs VORP

A diagram of a number of dots

Description automatically generated A diagram of a number of dots

Description automatically generated

Next, I wanted to know if there was a negative correlation between the number of turnovers in a game, and the players value above a replacement player. However, I discovered that instead of a negative linear relationship, there is a weak positive linear relationship between turnovers per game and VORP. This may be because the best players tend to have the ball more, and so they would turnover the ball more. Though testing it with turnover percentage showed a lack of relationship between the two variables.

Conclusion:

This data was helpful for answering some of my questions about what stats are correlated with the overall value of a player. Scoring points was the most correlated, as was shooting free throws for the type of shot. Turnovers did not correlate to a lack of value as much as I thought, while age doesn’t seem to make much of a difference on how much a player turns it over. I felt fulfilled by this and feel that there is a lot more I could do to explore this data and find interesting correlations.