

# NBA MVP 2021-22 Prediction - Which individual player performance statistics influence the Most Valuable Player (MVP) award prediction in the NBA\*

Linear Regression model analysis

Shoubhik Bhattacharya

27 April 2022

## Abstract

This paper presents an NBA player performance statistics analysis to determine which NBA statistics are important in determining the MVP. The analysis consists of linear regression models and visualizations conducted by the statistical programming language R. The results build upon existing research on predicting MVP players from previous seasons by adding newer variables with the newer data from the ongoing 2021-22 NBA season.

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Data</b>	<b>2</b>
2.1	Variable selection . . . . .	2
<b>3</b>	<b>Methodology and Visualizations</b>	<b>4</b>
<b>4</b>	<b>Model</b>	<b>13</b>
<b>5</b>	<b>Results</b>	<b>14</b>
<b>6</b>	<b>Discussion</b>	<b>14</b>
6.1	Weaknesses and next steps . . . . .	15
	<b>References</b>	<b>16</b>

---

\*Code and data are available at: [https://github.com/ShoubhikBhattacharya/STA304\\_Final](https://github.com/ShoubhikBhattacharya/STA304_Final)

# 1 Introduction

The National Basketball Association (NBA) is a North American professional men's basketball league comprising of 30 teams from the United States and Canada. Every year, the NBA awards the best performing player in the regular season with the Most Valuable Player (MVP) Award. Although this decision is made through a vote from a panel of sportswriters and broadcasters, player performance statistics are a key component that is utilized to formulate this decision. Every year players with the highest levels of performance, reflected by points, assists, rebounds and other statistics are awarded the MVP award. Since top NBA players such as Nikola Jokic and Joel Embiid reflect similar levels of performance in terms of points, assists and rebounds, the decision every year is extremely close! While there are qualitative factors to making this decision, advanced statistics such as PER (Player Efficiency Rating), WS (Win Share) and Box Plus-Minus (BPM) are usually the differentiating factors between MVP candidates. While our analysis will look into these new advanced statistics, we will also consider adding variables such as player scores in the NBA 2k22 video game to see if there might be an effect.

Since the NBA has a wide range of statistics to measure player performance and how they contribute to their respective teams success, this paper aims to determine which such statistics are key indicators of an MVP season in the NBA. In this paper, we will look at data for the 2021-22 NBA season from Basketball-Reference.com (2022) for the ongoing regular season to analyze performances throughout the season and make a prediction. We will also use previous years data from Basketball-Reference.com (2022) to make comparison with previous year MVP's and MVP candidates.

This analysis builds on existing literature by creating a regression model with additional newer statistics that may provide greater predictability. We discuss these variables and provide a prediction using our model for our MVP for 2021-22 in our data section. Using these variables we build a linear regression model(s) to help make an accurate MVP prediction. We discuss key insights on the variables we picked and excluded from our model in the data section. Lastly, we discuss the limitations and weaknesses of our data, model and analysis.

## 2 Data

The data for this project as mentioned earlier was obtained from [www.basketballreference.com](http://www.basketballreference.com) Basketball-Reference.com (2022). The data on this website is free and easily downloadable in a .csv format, through which we were able to conduct our analysis.

For conducting this analysis we used the R programming language R Core Team (2020), and libraries such as Wickham et al. (2019), Xie (2022) and Wickham (2016).

We obtained two separate datasets - The first consisting of regular NBA player statistics such as points, rebounds, assists, turnovers, effective field goal %, and other relevant figures used to measure box-score player analysis (analysis done in game). The second dataset consist of advanced player performance metrics such as PER (player efficiency rating), True Shooting %, Win shares (WS) , Value over Player Replacement (VORP), Box Plus-Minus (BPM) and Usage rate. Both datasets consist of player ages and minutes played (MP) along with individual player names.

### 2.1 Variable selection

To create our model, we need to pick variables that we think best measure a players impact on the court. We need variables that help us pick a player that can help his team win while on the court rather than boost his own individual statistics in game regardless of the team's outcome. To do so, we look at a bunch of regular and advanced statistics and pick our variables for the model. It will be beneficial to define each of the variables we are considering for our model.

### 2.1.1 Variable Definitions

First, it will be beneficial to define each of the variables. Points (PTS), assists (AST) and rebounds (REB) are the baseline basketball statistics that are used to define player performances with PTS measuring the sum of how many 2-pointer and 3-pointer baskets an individual scores. AST represents every pass that leads to a point scored, attributed to the passer. REB measures the collection of a missed shot attempt by the opposition offensive team, attributed to the player collecting the basket. Steals (STL) and blocks (BLK) are defensive statistics that tell us how many times a player can cause their opposition to turn the ball over (lose possession of the ball). (Basketball-Reference.com (2022))

Field goal percentage (FG.) is a percent that tells us the number of baskets made relative to shots attempted by a player. This is measured for 2 and 3 pointers. Similarly, Free throw percentage (FT.) measures the number of free throws made relative to the free throws attempted by a player. Other shooting statistics such as effective field goal % (eFG.) and True Shooting % (TS) helps measure how efficient a player's shooting is for 2-pointers, 3-pointers and free throws. (Basketball-Reference.com (2022))

While these statistics are a great baseline tool to observe individual player performance, a large variety of advanced statistics are used in the NBA to measure how efficient and impactful an individual is when on the court.

Advanced statistics are the most interesting measure as they combine regular statistics as they convert raw statistics such as PTS, AST and REB into measures that can be interpreted directly for efficiency and impact on the court. We picked a series of the most important advanced statistics for our model which we define below - Player Efficiency Rating (PER) is a vital statistic that "sums up all a player's positive accomplishments, subtracts the negative accomplishments, and returns a per-minute rating of a player's performance" (Hollinger (2007)). Usage % (USG.) is an estimate of the percentage of team plays carried out by a player when he was on the floor. Higher usage may indicate that more plays may be carried out by the individual player, implying he is of greater importance to the team. Win Shares (WS) are an estimate that calculate the number of wins contributed by a player. This estimate ranges from 0 (lowest) to 1 (highest) and is calculated from an equation accounting for several offense and defense basketball statistics. Box Plus Minus (BPM) is a "box score estimate of the points per 100 possessions that a player contributed above a league-average player, translated to an average team." Simply put, this statistic measures the basketball player's contribution offensively and defensively while on the court. Lastly, Value Over Replacement Player (VORP) is the exact same measure as BPM, while including minutes played by the basketball player. (Basketball-Reference.com (2022))

### 2.1.2 Final Variable choice

Looking at the above definitions, it is an obvious decision to include all the advanced statistics to predict the MVP as they provide a great indication of players impact on the court. Moreover, including shooting statistics such as eFG. and TS. will help us predict a player's efficiency in scoring while on the court. We will not include statistics such as PTS, AST, REB, STL and BLK in our analysis as the advanced statistics already account for them while measuring player impact. Novel additions in our paper will be the addition of statistics such as player age (Age) and Minutes Played per game (MP) and Games Played (G). I believe adding these statistics could influence the prediction of our MVP for the following reasons - Basketball players (just like any other sport) experience periods in their careers known as "player primes." These prime years are when the player performs at their peak level and are more likely to win an MVP award during this time. As players experience primes mid-career (at around 25-30 years old), including player ages in our model may help contribute to our MVP prediction. Moreover, MP helps indicate the importance of a player on a team as higher minutes played indicates that the player may have a higher impact for the team and vice versa. Lastly, GP is a crucial factor that is considered in MVP voting as being healthy throughout the season indicates greater value to a team. Due to this, they will be a key indicator in our MVP prediction model.

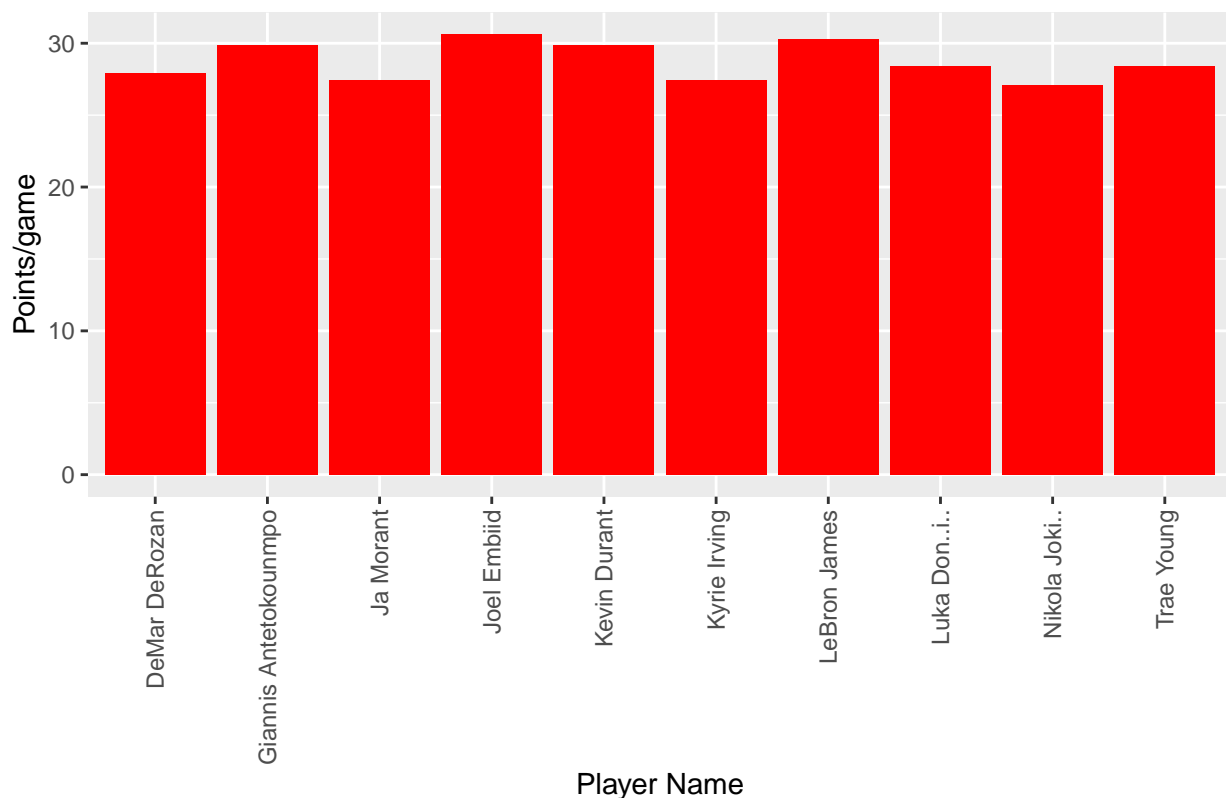
PREDICTION <- 2 + -0.007(Age) + -0.002(G) + -0.043(MP) + 1.19(eFG.) + -3.396(TS.) + 0.026(WS) + 1.310(WS.48) + 0.010(BPM) + 0.010(PER) + 0.030(VORP) + 0.009(USG.)

### 3 Methodology and Visualizations

After choosing our variables, it makes sense to observe the best performing player in the 2021-22 NBA season for each of the metrics (variables). For the variables chosen, we will choose the best performing players for each variable and create our own shortlist for the MVP Race! We will obtain a list of previous top 10 MVP candidates from the past 5 years and use it to create our model. Using this data, we will apply the statistics of each shortlisted player in the model and see which player obtains the highest score. That would provide us with the MVP Prediction!

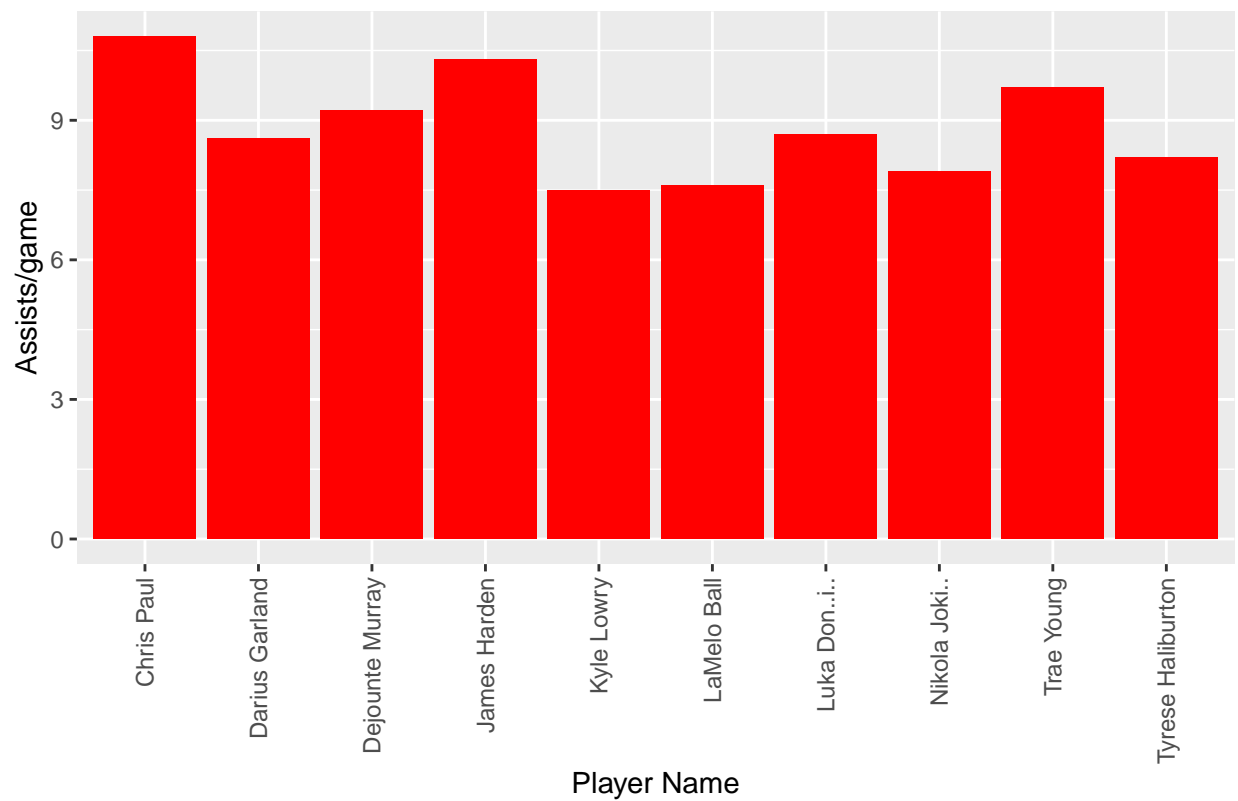
To make the shortlist, we start off by creating visualisations of our variables. We take a look at PTS, AST and REB first. Although, we will not be including these directly into our model, they are important components in calculating the advanced statistics. We look at points per game as it tells us how well top scoring players perform each game rather than all games combined. This helps us account for per game performance rather than overall scoring, not giving an advantage to players who played more games.

Figure 1: Player points per game (Top 10)



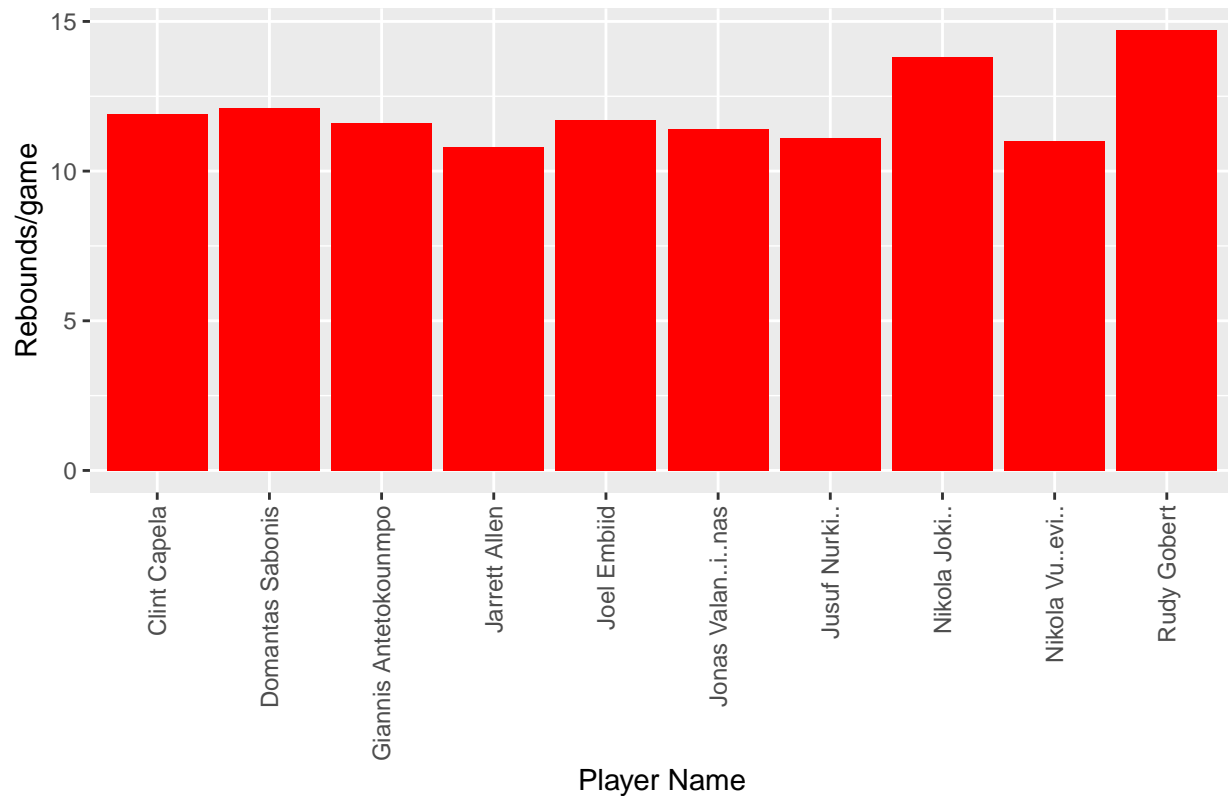
From Figure 1??, we see that Giannis Antetounmpo, Joel Embiid, Kevin Durant, LeBron James and Trae Young appear to be the highest scorers in the league. This is no surprise as they are regarded as the best players in the NBA, in the 2021-22 season.

Figure 2: Player Assists per game (Top 10)



Moving onto assists in Figure 2???, we see different names come up. Chris Paul, James Harden and Trae Young and Dejounte Murray are to top assisters in the league.

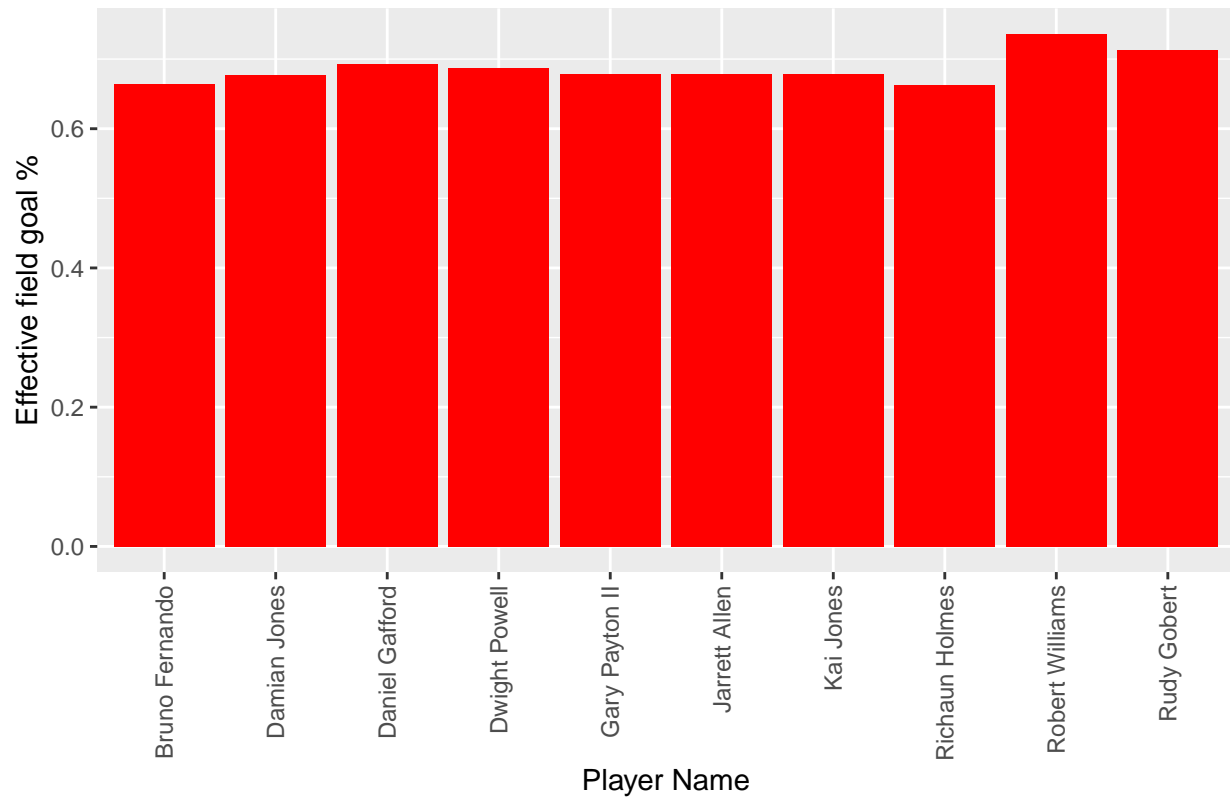
Figure 3: Player Rebounds per game (Top 10)



Finally, Nikola Jokic and Rudy Gobert are the top rebounders in the league as per Figure 3???. Now, It will be interesting to see if these names appear when we measure their positions with respect to advanced statistics such as BPM, WS, VORP and PER.

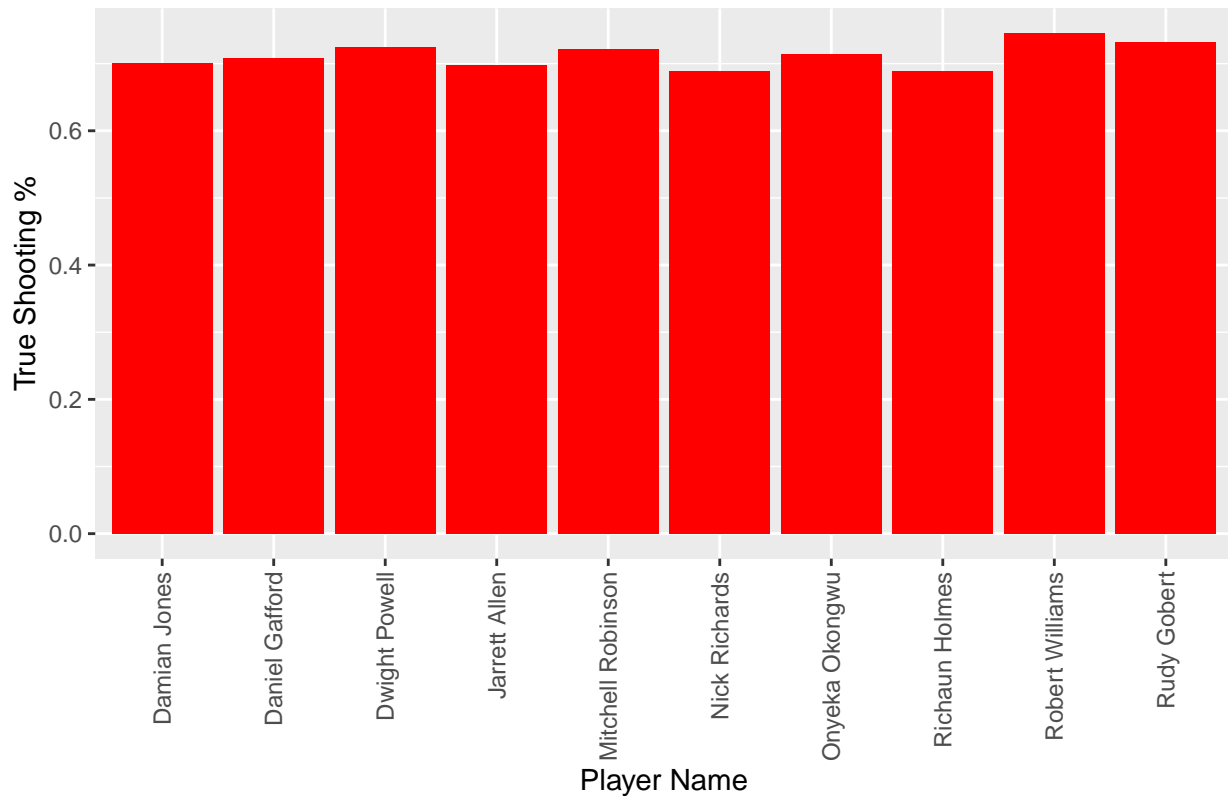
Before that, shooting statistics will give us a key insight on the efficiency of players on the court.

Figure 4: Effective field goal % (Top 10)



Looking at Figure 4??, we see a few new names. Robert Williams tops the eFG.% followed by Rudy Gobert and Daniel Gafford.

Figure 5: True Shooting % (Top 10)

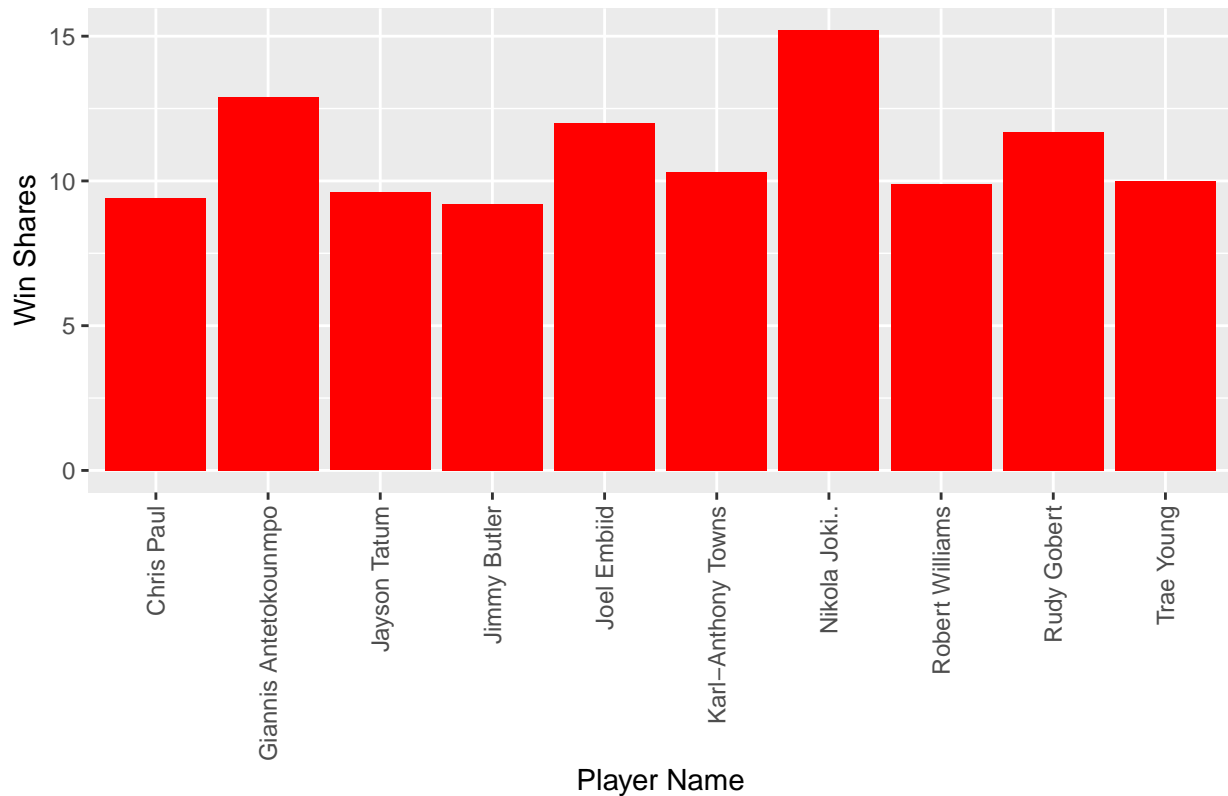


From Figure 5??, we see similiar names such as Robert Williams and Rudy Gobert appear, topping the list. These players are centers and usually score when they are nearer to the basket. This may be a key reason why we see such a high field goal percentage for this position.

Now, moving on to the more interesting advanced statistics!!!

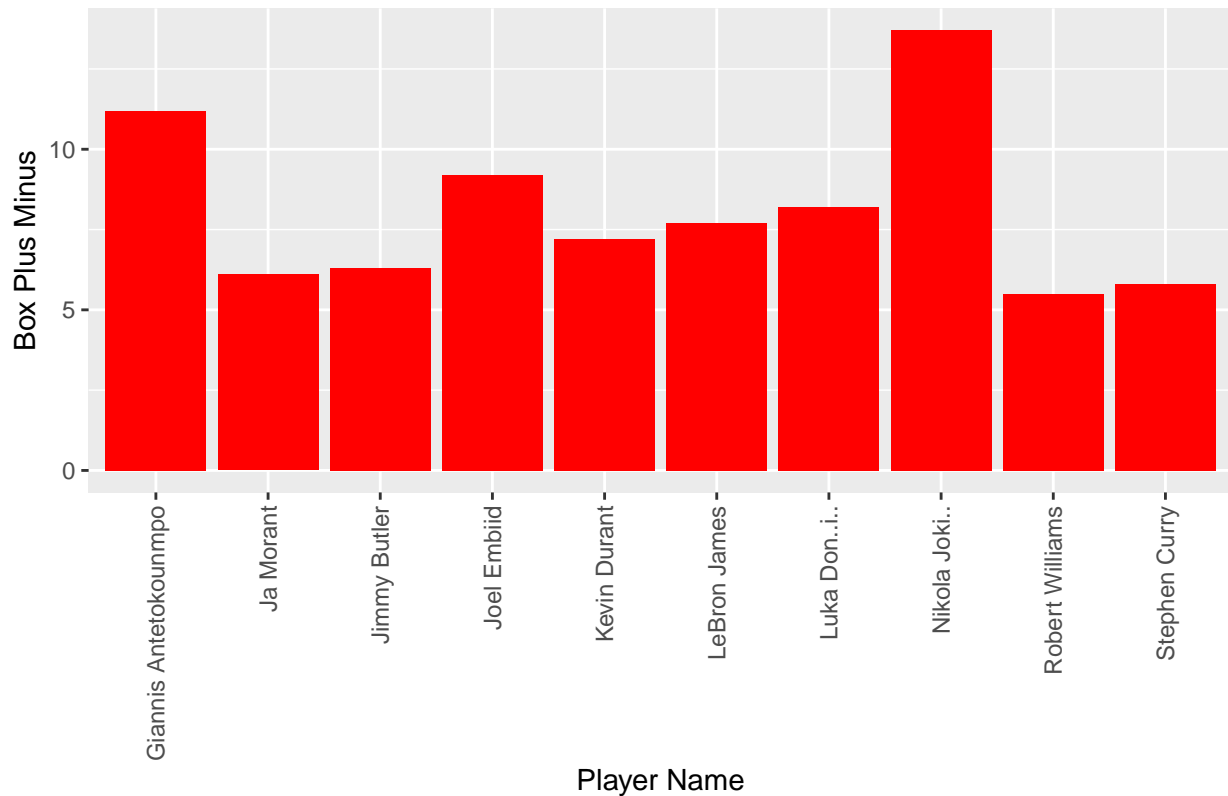


Figure 6: Player Win Shares (Top 10)



From Figure 6??, we see that Nikola Jokic, Giannis Antetokounmpo, Rudy Gobert and Joel Embiid contribute significantly to their teams wins than other league players in the 2021-22 NBA season. Let's see if these players have a large impact using our other advanced statistics.

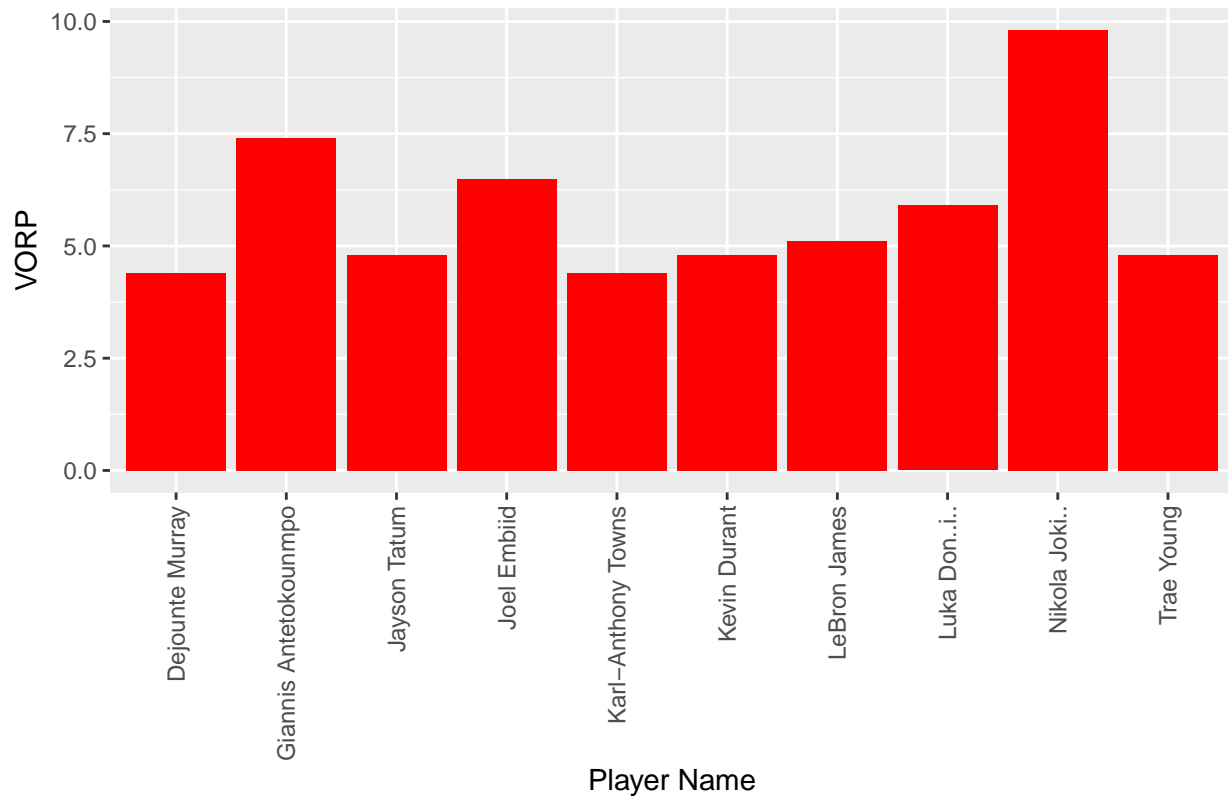
Figure 7: Box Plus Minus (Top 10)



Looking at BPM in Figure 7??, we see similar names! Nikola Jokic and Giannis Antetokounmpo lead this chart by a landslide! They are followed by Joel Embiid, Luka Doncic, LeBron James and Kevin Durant. Not surprisingly, we expect these big name players to contribute heavily to their teams when on the court.

Does VORP, which adjust BPM for minutes played, show a different trend?

Figure 8: Value Over Replacement Player (Top 10)



In Figure 8??, while we do observe the similar trend as expected there are a few key insights. Nikola Jokic and Giannis Antetokounmpo don't lead anymore in VORP by a landslide like they did for BPM. The differences between other players such as Joel Embiid, Luka Doncic, LeBron James and Kevin Durant compared to Giannis and Jokic are much smaller! Finally, we do observe a few new names in the list with Trae Young, Karl Anthony Towns and Jayson Tatum being added.

Figure 9: Player Usage Rate (Top 10)

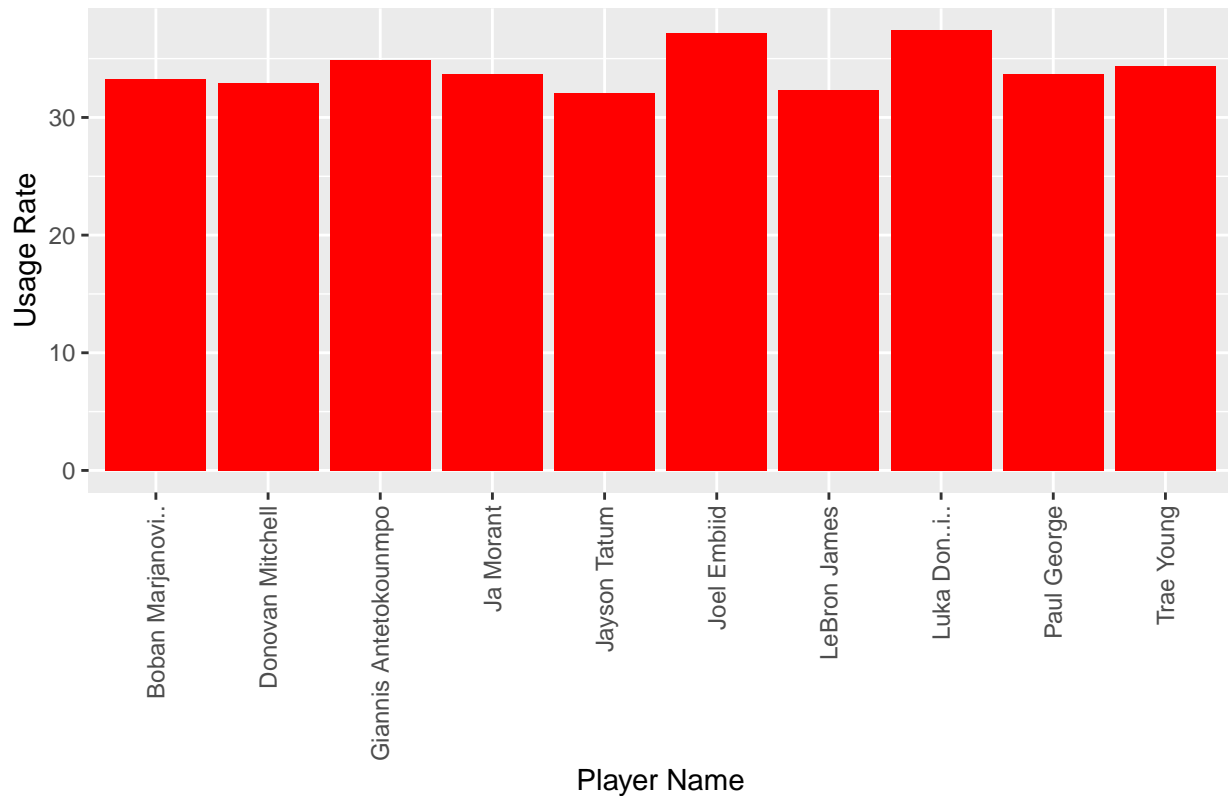
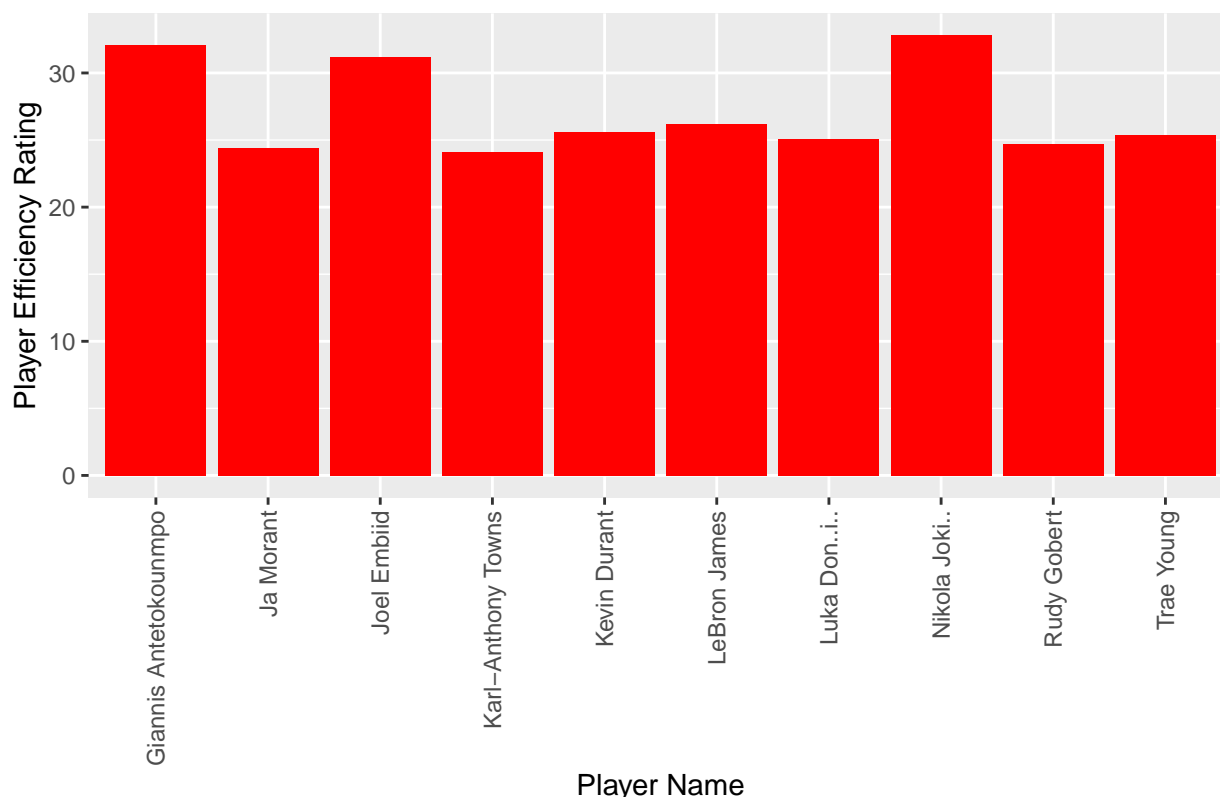


Figure 9?? shows us the player usage rates. Joel Embiid, Luka Doncic and Giannis Antetokounmpo have the highest usage rates followed by the likes of Trae Young, Paul George and Ja Morant.

Lastly, and most importantly, we look at PER. As explained previously, PER measures a players per-minute productivity while on the court. This statistic, is essentially, what we look for in MVP candidates - measuring how well a player plays and is a true measure of their performance on the court. This is differentiated from WS and BPM, as it doesnt look at their contributions to their team winning but rather how impressive an individual players performance truly is in terms of scoring.

Figure 10: Player Efficiency Rating (Top 10)



Looking at Figure 10??, we see the same names appear - Nikola Jokic, Giannis Antetokounmpo and Joel Embiid. Similarly, we see the same set of followers with these top 3 candidates - Luka Doncic, LeBron James, Trae Young, Rudy Gobert and Kevin Durant.

Based on this exploratory data analysis, we shortlist 10 candidates for our MVP candidate in the 2021-22 season in no particular order whatsoever - Nikola Jokic Giannis Antetokounmpo Joel Embiid Luka Doncic Kevin Durant Trae Young LeBron James Rudy Gobert Chris Paul and, Robert Williams

## 4 Model

As explained previously, this model will compare how MVP winners of previous years statistics compare to the current MVP candidates based on our advanced statistics analysis. We will use the previous years MVP's as our independent variable to compare against various regular and advanced statistic values and henceforth, attempt to create a prediction of who the 2021-22 NBA MVP may be.

To do so, we compile the statistics of 10 MVP candidates from 2016 to 2021 for each of the years and create a linear model for the variables we chose earlier. Using this, we will obtain estimates that will be employed to calculate the Final MVP score for each of our chosen candidates for the 2021-22 NBA season.

Our model unfortunately does not find any significance for any of the variables at the 5% or 10 % level of significance. We discuss this in the next section, but will use this model for our prediction. We will proceed with our prediction using these estimates. Using the equation given below, we produce our MVP Prediction results for the 2021-22 season -

$$\text{PREDICTION} < -2 + -0.005(\text{Age}) + 0.004(\text{G}) + 1.21(\text{eFG.}) + -3.236(\text{TS.}) + \#0.0392(\text{WS}) + 0.120(\text{BPM}) + 0.012(\text{PER}) + -0.139(\text{VORP}) + 0.005*(\text{USG.})$$

## 5 Results

Figure 11: MVP Candidates and their final MVP Scores

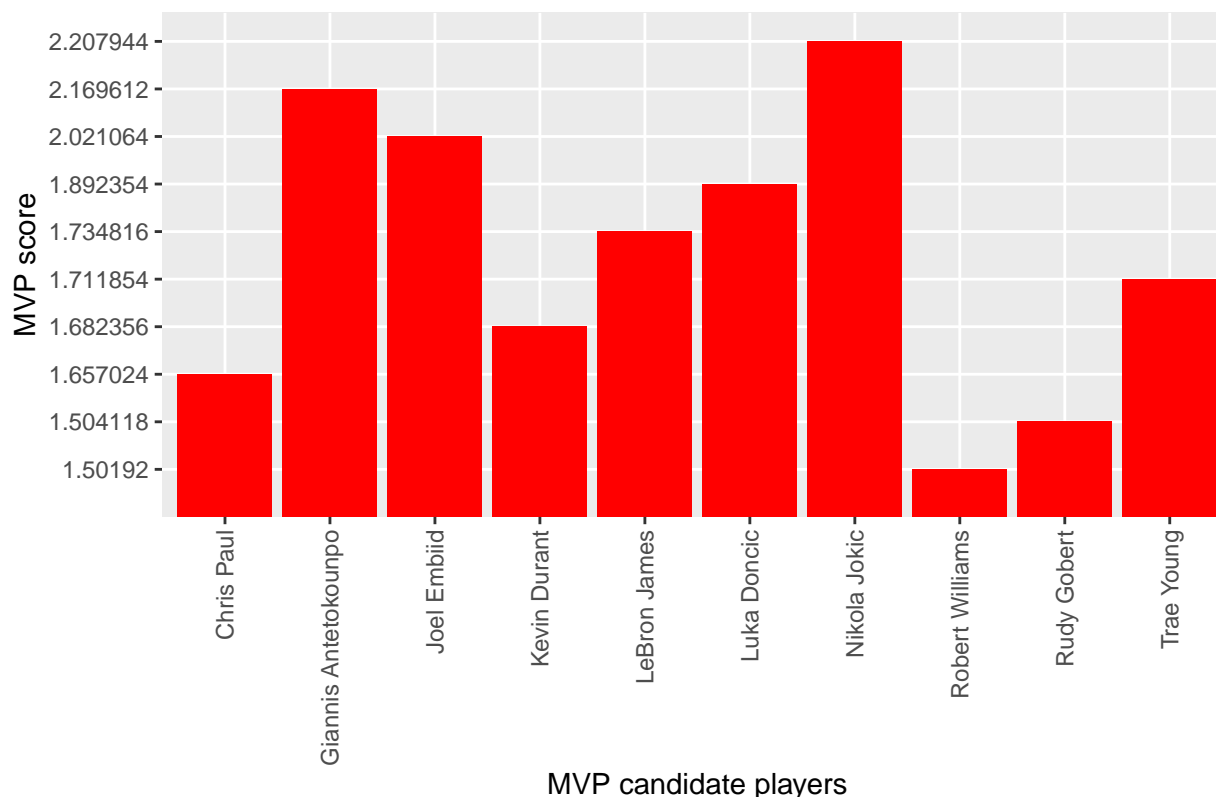


Figure 11?? shows us the top 10 MVP Candidates with their final scores from our prediction model. It can be seen that Nikola Jokic wins the MVP according to our model! He is followed by Giannis Antetokounmpo and Joel Embiid respectively. This is not a surprising result as these three candidates have been projected to be the finalists with Nikola Jokic being the favourite. (NBA.com (2022)) It is important to note however, that this model is not an accurate depiction of how MVP's are announced. The announcement of the MVP is a long, rigorous process involving several sports broadcasters' opinions. This means that while statistical variables may play a role, they do not solely decide the MVP, regardless of how great the statistics may be! Qualitative factors involving broadcaster judgment is essential in declaring the MVP.

## 6 Discussion

There are surprising outcomes observed for variables chosen for our linear model. Advanced statistic values such as BPM, PER and WS all demonstrate higher than expected p-values, implying it may not be a crucial factor for determining the MVP. This is surprising as we would expect the Most Valuable player on the court to be the one having the greatest impact when on the court and contributing significantly to one's own team's wins. Upon further research, I was able to identify the importance of qualitative assessments by media journalists and sportswriters over quantitative statistics in making the MVP decision. Since the 1980-81 season, voting powers were distributed to the media rather than having a player vote. Voters were allowed to rank five candidates of their choice where the first place obtains 10 points, second place gets seven, third place gets five, fourth place gets three and finally, fifth place gets one vote. The player obtaining the most cumulative points for that season from all eligible voters would obtain the MVP award. Voter Bias has proved to be a major issue in deciding the MVP each year as factors such as lobbying, voter fatigue and recency bias have contributed to poor decision making in regards to awarding the MVP trophy. (Corvo

(2021))

Lobbying occurs when the media vouches for a certain player over the other due to excessive and exceptional campaigning. A prime example of this was in the 1989-90 season when Magic Johnson beat Charles Barkley to win the MVP award despite having lower first place votes, resulting in a rather controversial decision. Lobbying recently has become more normal as NBA teams promote their players through social media platforms such as Twitter, Instagram and Facebook in order to win the MVP. Album covers comprising of Utah Jazz's superstar Rudy Gobert for the 2019 season, or the Anthony Davis "Unibrow" MVP toys sold by the New Orleans Pelicans in 2015 are campaigning efforts in order to lobby stars to win the MVP. Lobbying is indeed very powerful as teams consisting of MVP winners from past years such as the Milwaukee Bucks (Giannis Antetokounmpo in 2018-19) and Oklahoma City Thunder (Russell Westbrook in 2016-17) beat their counterparts with lesser convincing statistics and performances, but with greater marketing campaigns! (Corvo (2021))

Recency bias occurs in the NBA MVP award decision from the fact that one tends to favour performances of players towards the end of the season, closer to the announcement date of the MVP. Final period points per game and wins contributed are extremely significant in deciding the MVP as per HarvardSports (2015). In our model, an indication of this could be Devin Booker of the Phoenix Suns. Booker finds himself in the official MVP shortlist as per Basketball-Reference.com (2022), but fails to make the top 10 in any of our chosen variables that we have examined (both regular and advanced metrics). His performances towards the end of this 2021-22 NBA season has gained a lot of attention, and hence, won him a spot in the MVP race.

The most crucial factor that continues to affect NBA MVP decisions every year is voter fatigue. Voter fatigue stems from the fact that people hate to vote the same player as MVP every year to avoid losing the thrilling and exciting aspect of the MVP announcement. The biggest example of voter fatigue can be associated with the 2011-12 NBA season, where the MVP was named Derrick Rose rather than LeBron James. LeBron James beat Derrick Rose in almost every respect - points, shooting percentages, impact, winning and still managed to fall shy of Rose. Voters wanted change and were impressed by Derrick Rose's passion and determination of reviving the Chicago Bulls franchise that had struggled for over a decade after Michael Jordan's departure. Moreover, LeBron was made to look like the 'bad guy' after he left the Cleveland Cavaliers to join the Miami Heat after he had promised fans he was loyal to Cleveland. Emotional factors such as this one affect the game and people love to see new faces in the MVP circuit. (Brainwash (2019)) Even though our model predicts the MVP winner to be Nikola Jokic for the second time in a row, it will not be surprising to see the likes of Joel Embiid win the award as he has never received the award. His performances over the year are extremely impressive, but the statistics show that Jokic has had a far more significant impact in terms of winning the MVP. It will be exciting to see who wins!!!

## 6.1 Weaknesses and next steps

A major limitation of our model is the insignificance of the variables chosen. Although we discussed potential reasons for this in the earlier paragraph, there may be other ways the model could have been created. For example, our model accounts for shooting estimates such as True shooting percentage and effective field goal percentage. These values are higher for players who play the Center (C) position where they attempt field goals from nearer to the basket compared to Point Guards (PG) or Shooting Guards (SG), who take more shots and attempt more difficult three-pointers. Improving variables by accounting for factors such as player position, role in the team (role-player or starter) may provide newer, more interesting results. Moreover, our data compares estimates of MVP's from the 2016-17 season to the 2021-22 season in the NBA. This may lead to bias as we consider the MVP choosing criteria across these seasons to be the same. Creating a model that considers changes in the criteria may form a more comprehensive model for analysis.

## References

- Basketball-Reference.com. 2022. “Basketball Statistics and History.” Basketball-Reference.com.
- Brainwash. 2019. “What Is Voters’ Fatigue in the NBA and Is It Affecting More Than Just the MVP?”
- Corvo, Michael. 2021. “How Voting Is Done for the NBA MVP and Its Evolution.” [clutchpoints.com/how-voting-is-done-for-the-nba-mvp-and-its-evolution/](https://clutchpoints.com/how-voting-is-done-for-the-nba-mvp-and-its-evolution/).
- HarvardSports. 2015. “NBA MVP VOTING: HOW PLAYING ON THE WEST COAST AND LATE-SEASON SURGES AFFECT THE RACE.”
- Hollinger, John. 2007. “What Is PER?”
- NBA.com. 2022. “Kia NBA MVP Tracker: Latest Performances from Leading Contenders.” <https://www.nba.com/news/kia-nba-mvp-tracker-2022>.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. “Welcome to the Tidyverse.” *Journal of Open Source Software* 4 (43): 1686.
- Xie, Yihui. 2022. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*.