Derrick Chun, Kelsey Keate, Pavan Kumar, Cole Ouyang

DS 110 - Final Project Paper

Professor Gold

May 1st, 2024

Introduction

The Most Valuable Player (MVP) award is an annual award given by the National Basketball Association to the best performing basketball player of the regular season. This trophy has been awarded to one player each year since the 1955-1956 season. A collection of factors go into deciding which basketball player will receive this nomination. A panel of 100 sports writers and broadcasters are selected by the NBA to vote on who wins the MVP award. These panelists take into account a number of factors including individual statistics like Points Per Game (PPG), Assists Per Game (APG), Rebounds Per Game (RPG), Blocks Per Game (BPG), and Steals Per Game (SPG). On top of these, they also look into the basketball player's team success throughout the season and in their conference. The problem that we noticed for deciding the MVP is that this process could become convoluted with many subjective opinions from possibly biased voters. With that, the goal of our project is to simplify this process by creating a model that can accurately predict which players could be the top five MVP candidates for the 2023-2024 season solely based on the statistics from the 2019-2020, 2020-2021, 2021-2022, and 2022-2023 seasons.

Previous Work

When looking at the statistics from the previous few seasons that include 2019-2020, 2020-2021, 2021-2022, and 2022-2023 the MVPs tend to have a common trend of retaining higher than average statistics when looking at PPG, RPG, APG, and more. For this project our central question is if we can accurately predict which players could be the top 5 MVP candidates for the 2023-2024 season strictly based off of past season statistics and its correlation to MVP voting results. The MVP award takes into account various factors "such as statistical performance, team success, leadership, and overall impact on the game." (Rose, 2023). While leadership and overall impact on the game are important, these are both subjective metrics that cannot be quantitatively measured. The team success statistic is a given because there have only been two times where the MVP was given to a player on a losing team. With that, statistical performance is the only metric that is measurable and differentiable between NBA players. When comparing each of the MVP stats of PPG, RPG, APG, BPG, and SPG to the average stats per player that season, the MVP had above average stats in most, if not all, the categories (ESPN, 2024). These results show that statistics do have a play in the determination of the MVP. It has also been noted that the higher the statistics are for defensive rebounds and shooting efficiency, the more likely the team is to win and the more likely the player on that team will win the MVP award (Cabarkapa D, Deane MA, Fry AC, Jones GT, Cabarkapa DV, Philipp NM, 2022). There is a positive correlation between higher statistics and the chance of getting MVP for NBA basketball players on better teams. Although Thunder lost 5 games in the first round of playoffs back in the 2016-2017 season, the playoffs don't matter in Russell Westbrook's case. He "averaged a triple-double on the season" and "led the league" (Cavanaugh, 2024). This proves that although the MVP does take team performance into account it is primarily an individual stats driven award. Nikola Jokic is expected to be the MVP for the 2023-2024 season. He averaged "26.4 points, 12.4 rebounds, and 9.0 assists per game during the regular season" (Conway, 2024). These numbers are significantly above average which is rare for one player to have all three of these statistics to be above average meaning that his stats are important when taking MVP into account. Overall, the better or the higher the statistics are for a player, the more likely they are to receive the MVP award.

Methodology - Data Preparation

In order to prepare the data to help figure out the machine learning, statistical analysis, and visualizations, the 2019, 2020, 2021, and 2022 season player statistics were downloaded from kaggle as .csv files. These four seasons worth of player data were gathered and in these data sets five main variables were observed: Points Per Game (PPG), Assists Per Game (APG), Blocks Per Game (BPG), Steals Per Game (SPG), and Rebounds Per Game (RPG). The datasets originally presented these statistics as totals throughout the season, not per game statistics. The totals were then divided by the amount of games played to get accurate per game statistics. The Number of Games were also observed for machine learning and exploratory statistics. All the other variables that were observed in the data sets were dropped because they played no significance in determining MVP in this project. With all of this, the four seasons were then merged into one cohesive dataset.

A new dataset was then created to include NBA MVP voting results from the NBA website. This was made by putting the MVP voting data into excel and to look at the list of the top players from each of the four seasons mentioned previously. The players who received MVP votes were ranked based on the amount of total First Place, Second Place, and Third Place votes they received. From there, the new column was used as labels in both the decision tree classifier and KNN classifier in the code. This then expanded the number of positive labels from 5 to 25 as the top 5 ranked players from each season were added. The four seasons were then merged into one dataset. When merged, the same data parameters will be kept while the parameters that differ from each other in each dataset will be excluded.

In order to determine the 2023-2024 NBA MVP, a fifth dataset was added which was the 2023-2024 season. The 2023-2024 season was downloaded to look at this season’s player’s statistics. This dataset will then be used to help predict this season’s MVP voting results by using machine learning trained with the previous four seasons of historic data. Some of the statistics for the five seasons gave us the totals of the player's Points, Assists, Blocks, Steals, and Rebounds so in order to get the statistics for Per Game, new columns were created based on the raw data that looks at the Per Game statistics. To prevent any NaN errors, the default data was filled with None for the code to process even though the data is missing for some variables. The datasets were then merged and customized to create categorical variables.

Methodology - Statistics

For our first statistical analysis, a T-Test was used to compare the PPG (Points Per Game) values between players ranked 5 or less (most likely to be MVP) and players ranked 6 or higher (not likely to be MVP). With that, the null hypothesis is that there is no significant difference in the number of PPG between two groups of players based on rank while the alternative hypothesis states that there is a significant difference.

For our second and third statistical analysis', two ANOVA tests were performed to analyze the difference between the means of Points Per Game (PPG), Assists Per Game (APG), Rebounds Per Game (RPG), Steals Per Game (SPG), Blocks Per Game (BPG), and the Number of Games (G). The first ANOVA test was done for all players and the second ANOVA test was done for MVPs. With that for both ANOVA tests, the null hypothesis states that all these means are equal to each other while the alternative hypothesis states that at least one mean is not equal.

These tests were used to see if there is a statistically significant difference between certain player statistics and their place in MVP voting for the year. We also utilized the various statistics to identify the correlation of how each player's statistics relate to its possibility of winning the MVP.

Methodology - Machine Learning

In order to begin the machine learning methods, we started with creating a decision tree classifier. This is used in order to make a prediction on the 2023-2024 dataset of the player's statistics by testing certain thresholds players have to reach in certain stats to win MVP. The parameters can vary by changing the thresholds. For example, on one branch of the tree it may be checking whether or not a player had more than 20 points per game, the parameters could change to 22 points per game which then allows us to see how that affects the accuracy. The target features of the decision tree are PPG, APG, RPG, BPG, and SPG while the labels for the decision tree are the MVP voting results. The previous four seasons worth of data were then trained and tested by using the current season's player statistics. This decision tree classifier is predicted to return the players and their corresponding stats who were predicted to be at the top of MVP voting.

The next machine learning method that was used was the K Nearest Neighbors classifier. This was used because it is best at creating a boundary to classify and predict data that is nearest to that boundary line. It was used to predict for each K value of 1-10 to show players that correspond with each K values condition. This classifier takes the same target values as before (Points Per Game, Assists Per Game, Rebounds Per Game, Blocks Per Game, and Steals Per Game) into account. Since there is only one MVP named per season, using this as the primary label would result in an artificially high accuracy when using the test data. In the original model, only an extremely small number of data points would be True. To help combat this issue, we used data from NBA MVP voting sheets that breaks down each player that received a vote. With that, an empty list was created to store the results and a loop was made to iterate from 1 to 10 to try different values of the parameter for the KNN classifier. The X_new_season was made in order to make predictions using the trained classifier on the new data and from there indices where the predicted values of less than six were selected and stored. This allows for the classifier to retrieve player names that correspond to the predictions from the dataframe and store them.

Results - Statistics

For the T-Test, we set our significance value to 0.05. After running the T-Test, the result was a p-value of approximately 1.419 x 10^-36 which is well under 0.05. Since our null hypothesis is that there is no significant difference in the number of PPG between two groups of players based on rank and our alternative hypothesis states that there is a significant difference, the result of the test is that we can reject the null hypothesis and conclude that there is a significant difference in the number of PPG between two groups of players based on rank.

For the first ANOVA test, we compared all the averages of all the players' statistics together. With that, we set our significance value to 0.05 and found that there were six calculated p-values since there were six averages to be calculated. The p-value for each stat is as follows:

- PPG = 0.878170
- APG = 0.309543
- RPG = 0.040865
- SPG = 0.350394
- BPG = 0.107298
- G = 0.380720

Only one of these p-values was below the significance value of 0.05 which was the Rebounds Per Game (RPG). Even though only RPG had a significant difference, the null hypothesis was rejected because at least one pair of means is not equal to the rest.

For the second ANOVA test, we compared all the averages of the MVP's player statistics together. Then, we set our significance value to 0.05 once again. After running the second ANOVA test, there were another six calculated p-values since there were six means to be calculated. The p-value for each stat is as follows:

- PPG = 0.008273
- APG = 0.004331
- RPG = 0.007545
- SPG = 0.006274
- BPG = 0.005029
- G = 0.007005

All of these p-values are well below the significance value of 0.05. Since our null hypothesis is that all the mean values are equal to each other and the alternative hypothesis states that at least one pair is not equal, the results of the test concludes that all group means are significantly different to each other meaning that we can reject the null hypothesis.

Results - Machine Learning

For the decision tree classifier, it resulted in nine players to be predicted to be classified as the Top Players that were ranked 1-5 for the current 2023-2024 season. The nine players predicted were:

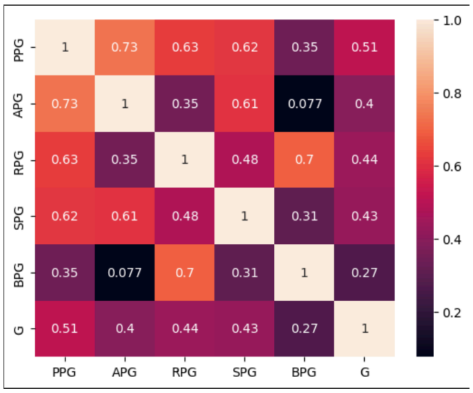
- Giannis Antetokounmpo with Predicted Rank of 3.0
- Anthony Davis with Predicted Rank of 1.0
- Kevin Durant with Predicted Rank of 5.0
- Joel Emiid with Predicted Rank of 3.0
- Tyler Herro with Predicted Rank of 3.0
- LeBron James with Predicted Rank of 5.0
- Damian Lillard with Predicted Rank of 4.0
- Anfernee Simons with Predicted Rank of 4.0
- Trae Young with Predicted Rank of 3.0

From these results, the predicted rank allows us to see that the better the predicted rank for a player, the more the model believes that they fit the criteria for MVP. For example, a player with a rank of 1.0 is more fit for MVP than a player with a rank of 5.0. We can compare these results to the ongoing MVP vote and find out that one player in the ongoing vote matched with our results: Giannis Antetokounmpo. The performance metrics changed if the parameter values were changed meaning that different accuracies were calculated.

For the K Nearest Neighbors classifier, it works based on the idea that similar players are going to be closer together. In this code's KNN, it results in a list containing the names of the 2023-2024 players predicted at k-values ranging from 1-10 within a rank of less than 6 to get the top 5 ranked players. The players that were listed in multiple k values were the most noticeable players (Joel Embiid and Giannis Antetokounmpo) due to the fact that their stats were a better fit to the model so they are more likely to appear even when the k value is changed. The lower k value a player has, the better fitting they are on the training data for the MVP status. Every k-value up until k = 8 had players listed under them; however, the k-values of 8, 9, and 10 had no players listed under since we only trained the model with 25 MVP candidates. Once the model expanded to 8-10 closest neighbors, there were 0 players who fit the criteria. As you get higher k values, the accuracy goes down denoting that no players were not a good fit on the training data.

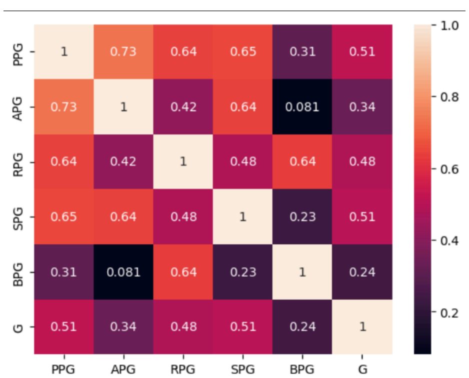
Visualizations:

Heat Map for 2019-2020 Season:

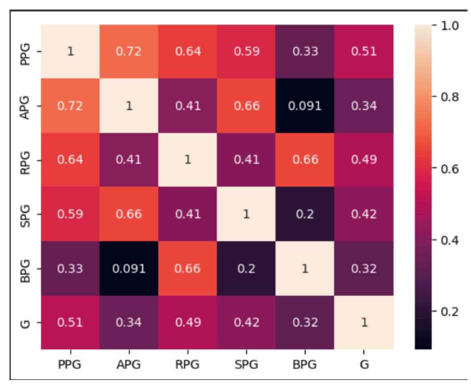


Heat Map for 2020-2021 Season:

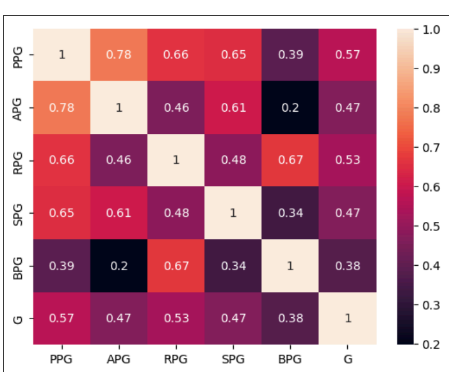
Heat Map for 2021-2022 Season:



Heat Map for 2022-2023 Season:

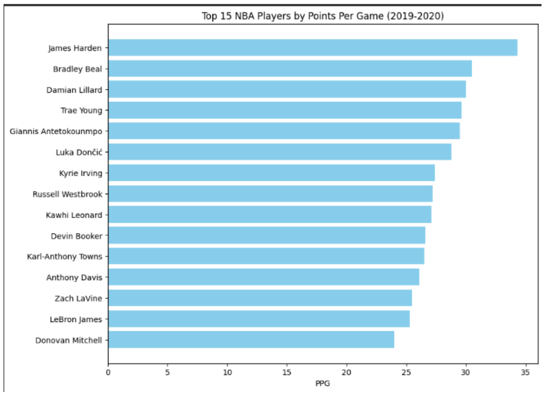


Heat Map for 2023-2024 Season:

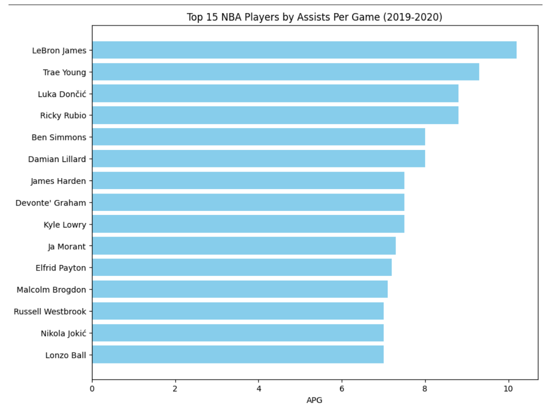


These heatmaps from 2019-2020, 2020-2021, 2021-2022, 2022-2023, and 2023-2024 tell the correlations between each X feature. All the heatmaps conclude that the Points Per Game and Assists Per Game have the biggest effect in determining MVP.

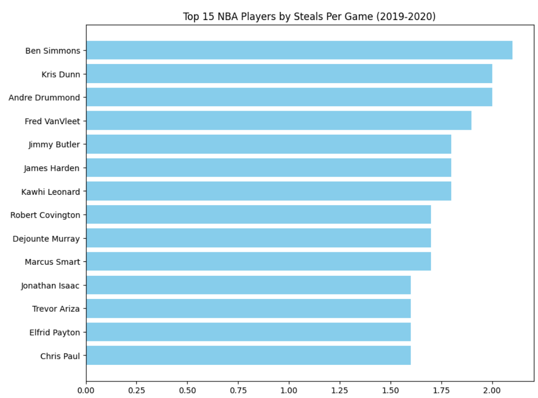
Top 15 NBA Players by PPG 2019-2020:



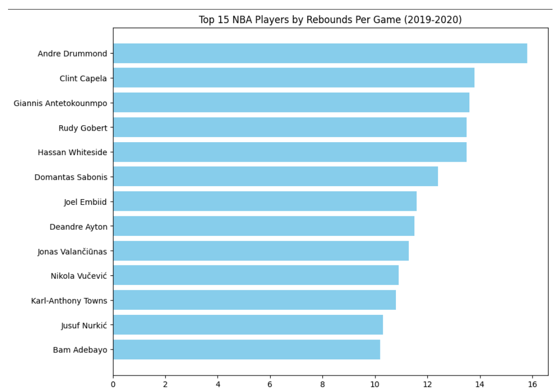
Top 15 NBA Players by APG 2019-2020:



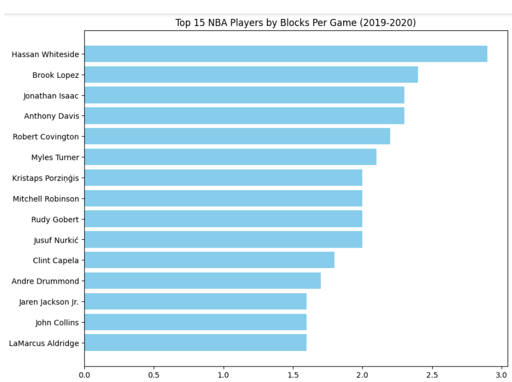
Top 15 NBA Players by SPG 2019-2020:



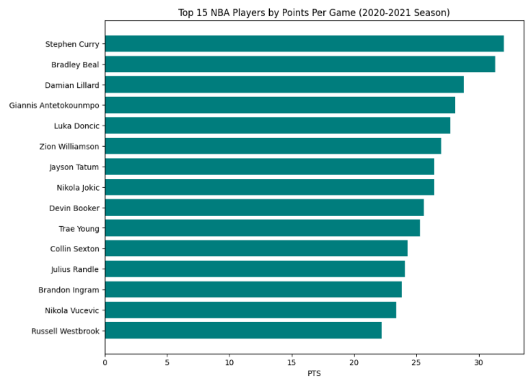
Top 15 NBA Players by RPG 2019-2020:



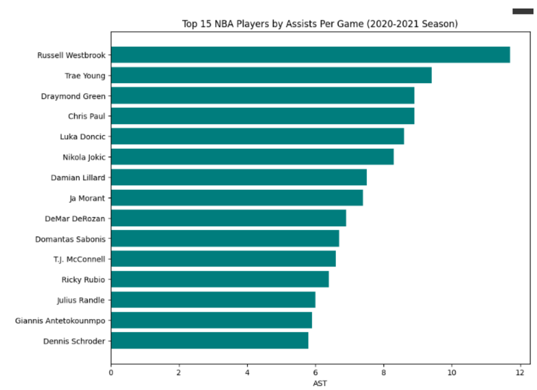
Top 15 NBA Players by BPG 2019-2020:



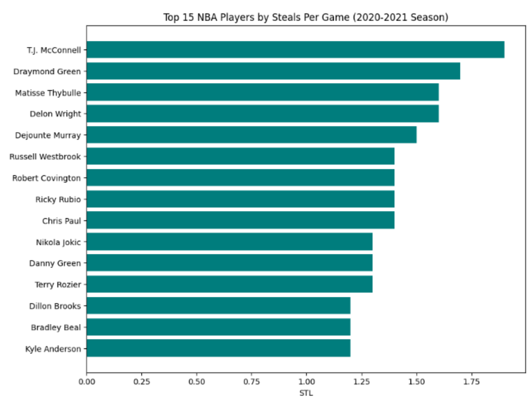
Top 15 NBA Players by PPG 2020-2021:



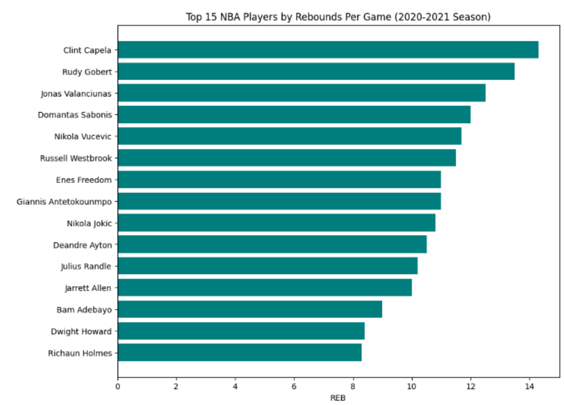
Top 15 NBA Players by APG 2020-2021:



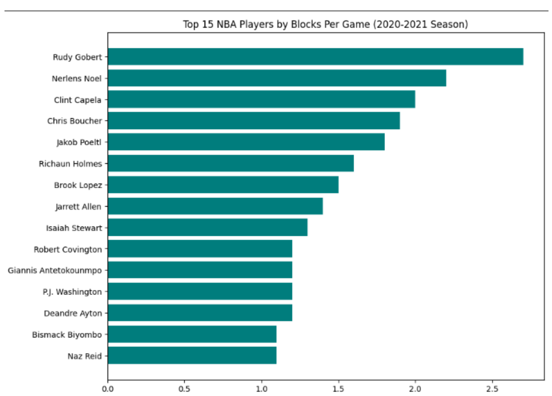
Top 15 NBA Players by SPG 2020-2021:



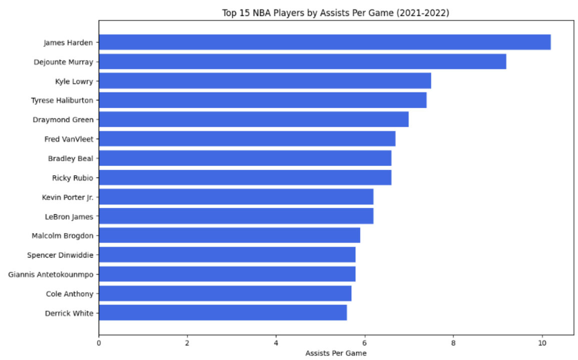
Top 15 NBA Players by RPG 2020-2021:



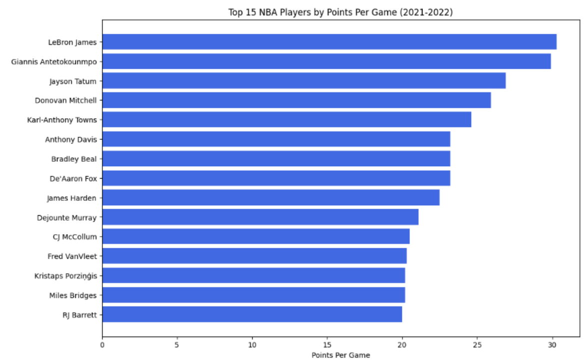
Top 15 Players by BPG 2020-2021:



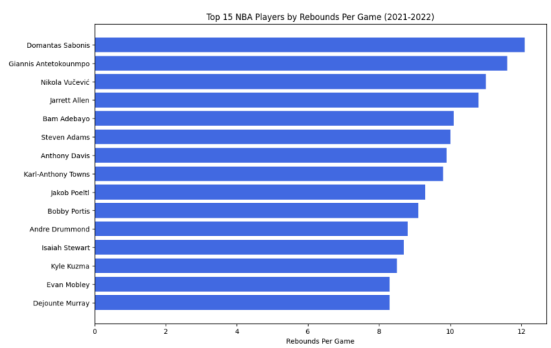
Top 15 NBA Players by APG 2021-2022:



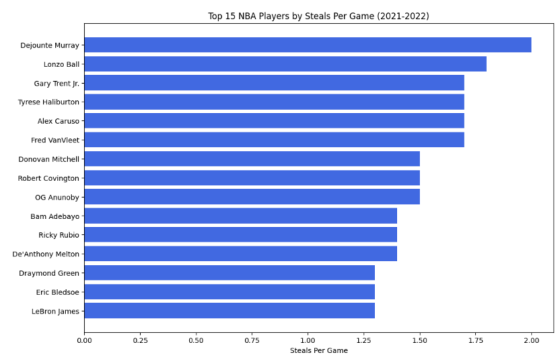
Top 15 NBA Players by PPG 2021-2022:



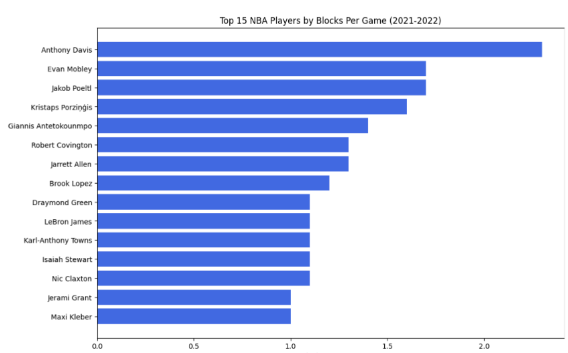
Top 15 NBA Players by RPG 2021-2022:



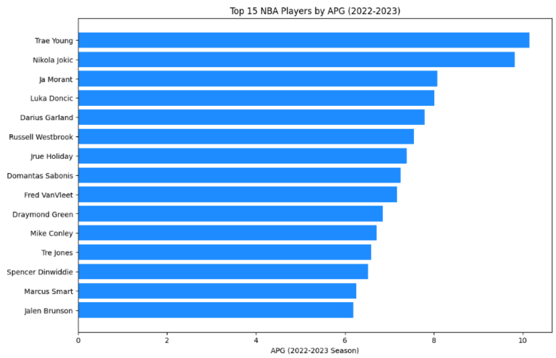
Top 15 NBA Players by SPG 2021-2022:



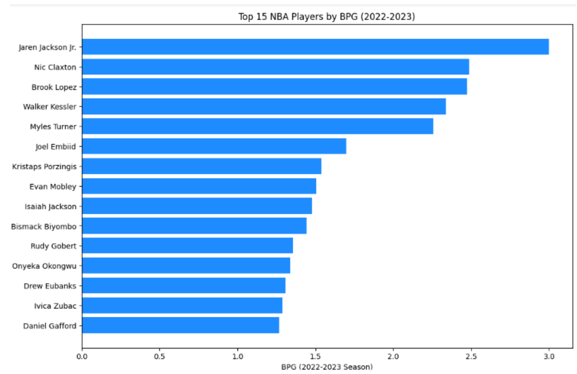
Top 15 NBA Players by BPG 2021-2022:



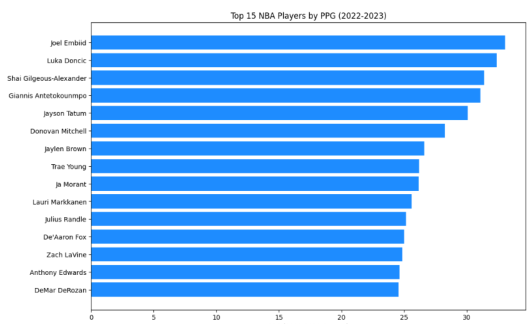
Top 15 NBA Players by APG 2022-2023:



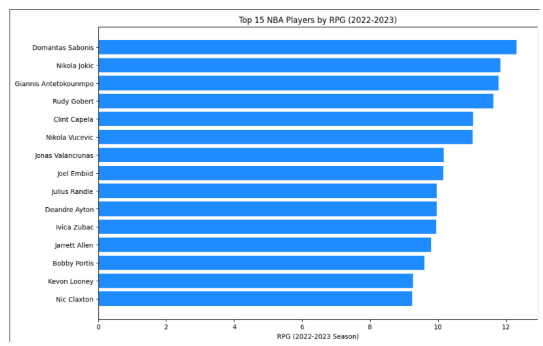
Top 15 NBA Players by BPG 2022-2023:



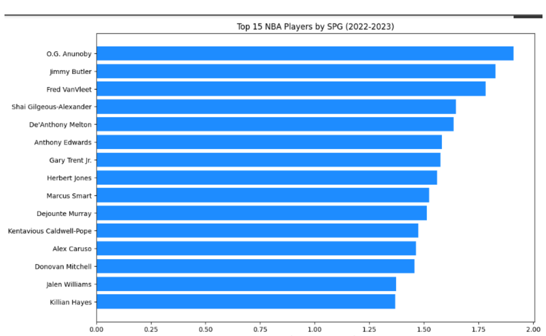
Top 15 NBA Players by PPG 2022-2023:



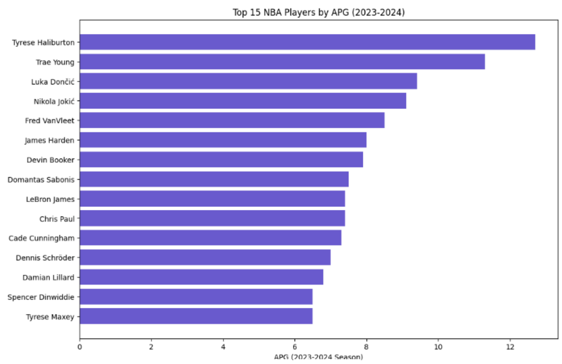
Top 15 NBA Players by RPG 2022-2023:



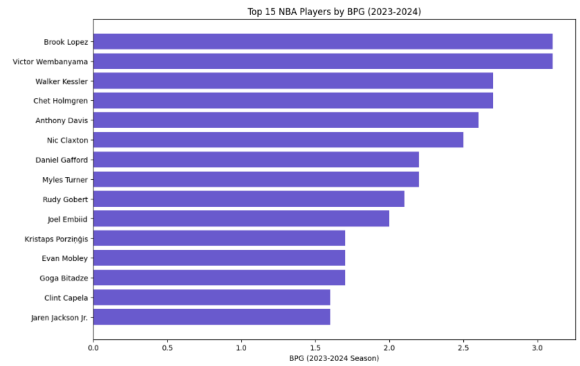
Top 15 NBA Players by SPG 2022-2023:



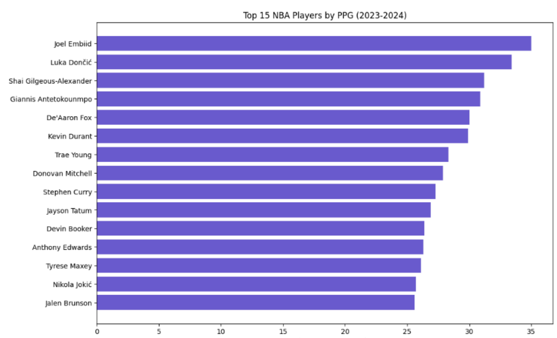
Top 15 NBA Players by APG 2023-2024:



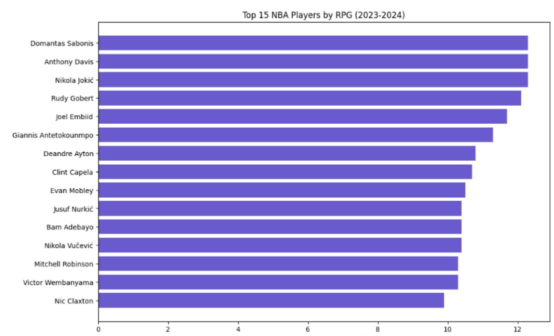
Top 15 NBA Players by BPG 2023-2024:



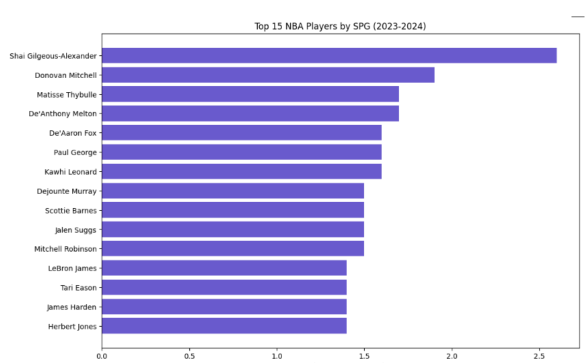
Top 15 NBA Players by PPG 2023-2024:



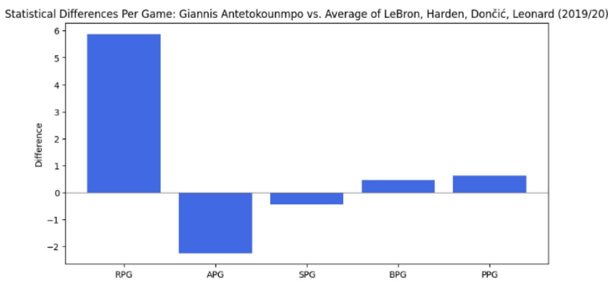
Top 15 NBA Players by RBG 2023-2024:



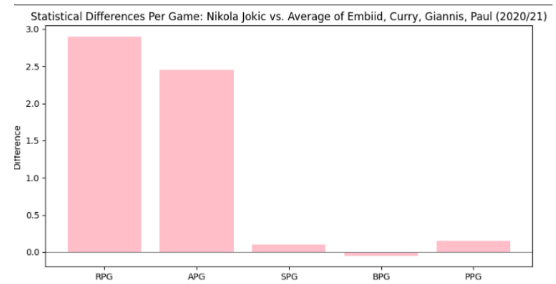
Top 15 NBA Players by SPG 2023-2024:



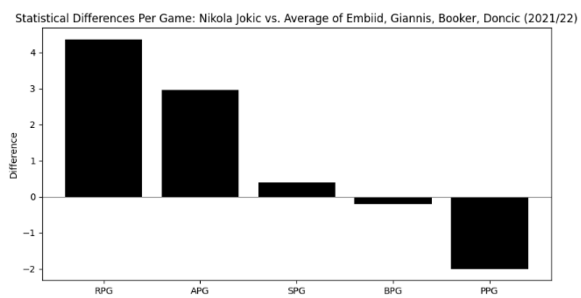
Bar Graph from 2019-2020:



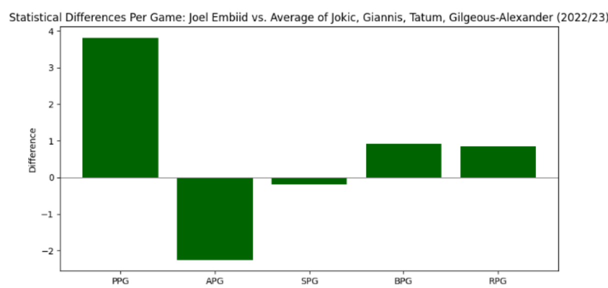
Bar Graph from 2020-2021:



Bar Graph from 2021-2022:

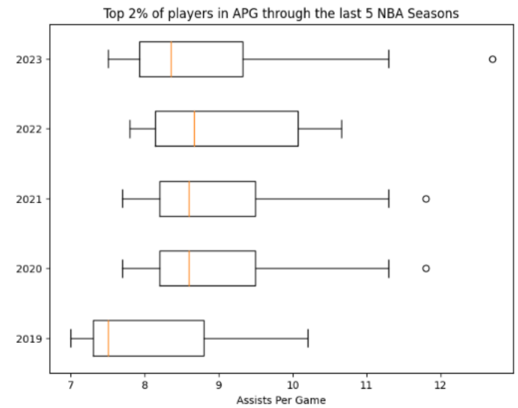


Bar Graph from 2022-2023:

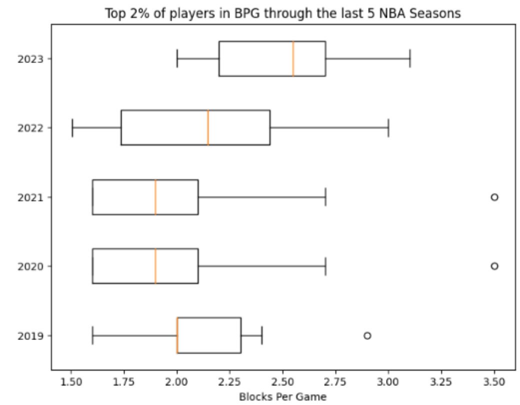


These bar graphs represent the visual comparison of MVPs of each season from 19-20, 20-21, 21-22, and 22-23 with the Next Top 4 Players in MVP Voting. The bar graphs emphasize the fact that the MVP from each season have significantly different statistics when compared to the next top four players from that season. This is important to identify as we have mentioned before that statistics has a huge say in who is voted MVP.

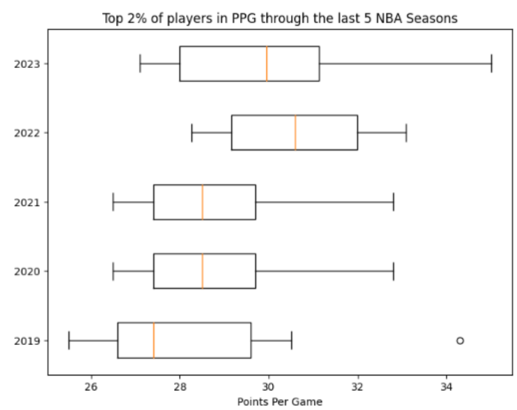
Top 2% of Players in APG:



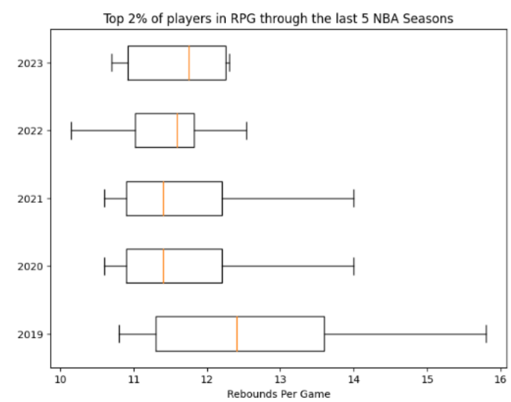
Top 2% of Players in BPG:



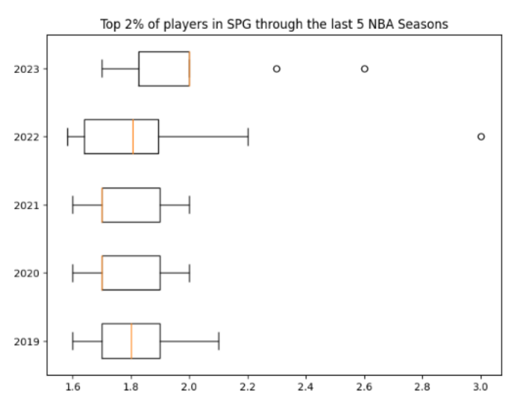
Top 2% of Players in PPG:



Top 2% of Players in RPG:

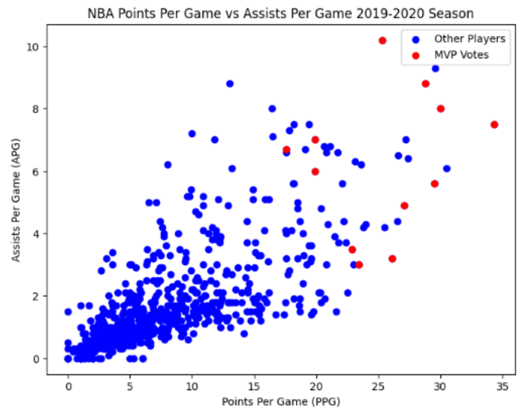


Top 2% of Players in SPG:

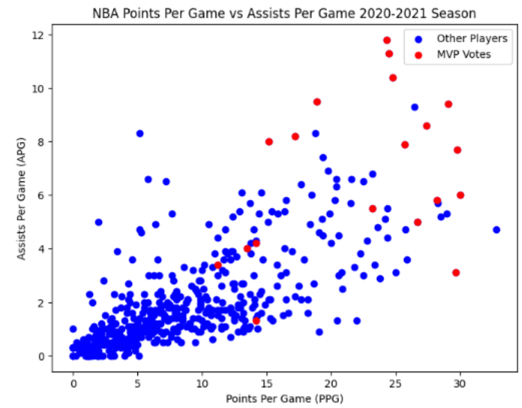


These box and whisker plots show the top 2% of players in each statistical category that was observed in the code. The different shapes between each box and whisker plot shows how the data between each season is distributed within the top 2% of players. On top of this, some of them have outliers meaning that some of the top players in that season had above average stats compared to the other players.

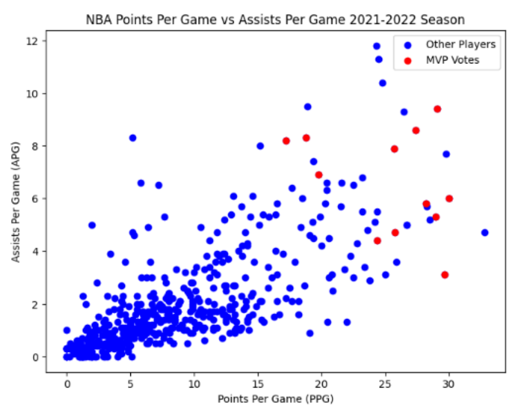
PPG vs APG 2019-2020:



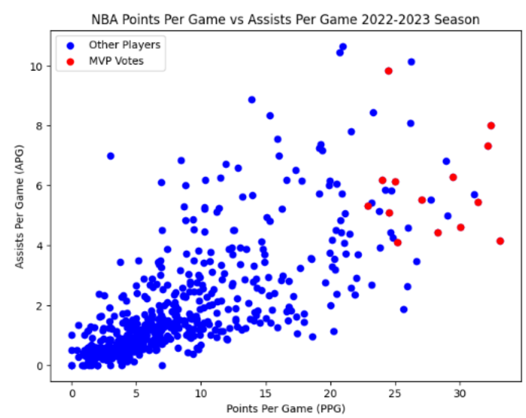
PPG vs APG 2020-2021:



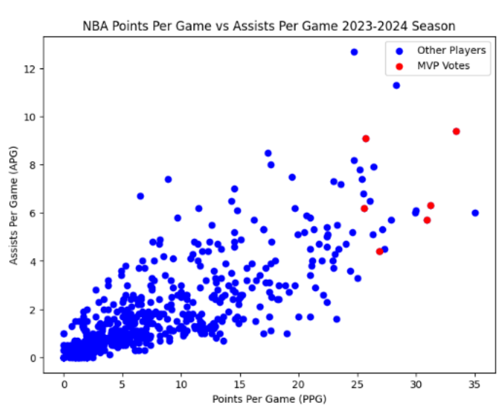
PPG vs APG 2021-2022:



PPG vs APG 2022-2023:

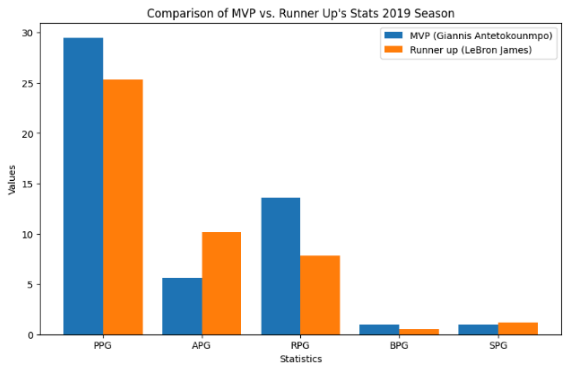


PPG vs APG 2023-2024:

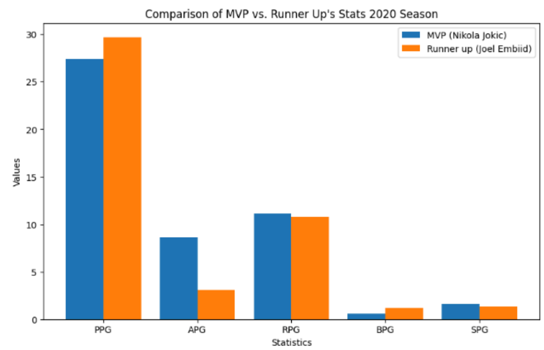


These scatterplots from the 2019-2020, 2020-2021, 2021-2022, 2022-2023, and 2023-2024 seasons show only two specific statistics: Points Per Game and Assists Per Game. These graphs make it noticeable that for each season, the NBA MVP had higher Points Per Game and Assists Per Game stats when compared to all the other NBA players.

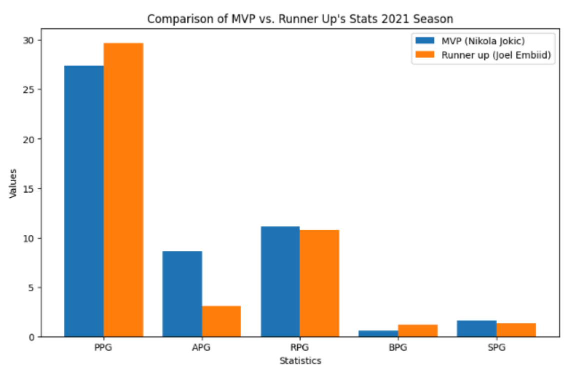
MVP vs Runner Up’s in 2019 Season:



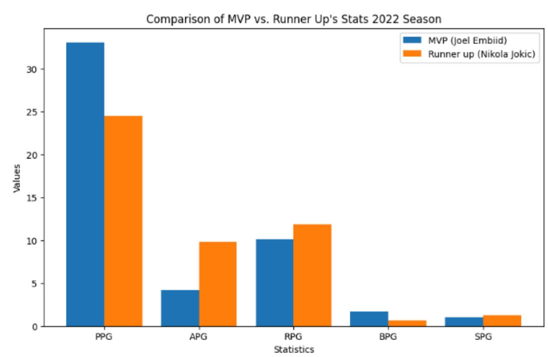
MVP vs Runner Up’s in 2020 Season:



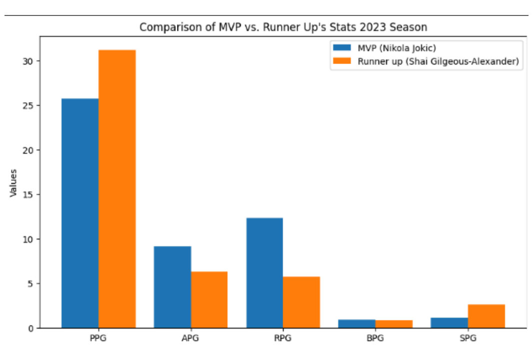
MVP vs Runner Up’s in 2021 Season:



MVP vs Runner Up’s in 2022 Season:



Projected MVP vs Runner Up’s in 2023 Season:



Conclusion

In both the K Nearest Neighbors and Decision Tree classifiers, the machine learning models were accurate considering the limitations of the data that went into the code. In the decision tree, Tyler Herro and Anfernee Simons were outliers while in K Nearest Neighbors, Fred Vanvleet was an outlier. Both classifiers took four years of previous NBA MVP data into account rather than considering the entirety of past NBA data and the players were classified solely based on correlated rank. Additionally, individual statistics are not the only criteria considered when selecting the NBA's MVP. Other metrics such as team success, player popularity, and leadership are included. The MVP is also voted on by a panel consisting of 100 people and subjective opinions may influence the process of determining the recipient of the award.

Although there were limitations, both the K Nearest Neighbors and Decision Tree classifiers predicted precise results in identifying the top MVP candidates. The machine learning models predicted players who had above average individual statistics compared to the rest of the league. The NBA players in the data frame who were predicted to have above average statistics are more likely to be ranked Top 5 in the code. By comparing the results of the predicted machine learning models and the 2023-2024 MVP voting, there were corresponding players. Therefore, both machine learning models that were performed in the code were accurate at identifying the top MVP candidates for the 2023-2024 NBA season.

Credits

Derrick Chun:

Regarding cleaning and organizing data, Derrick was supposed to organize the raw datasets (raw), clean the raw CSV datasets, and merge the four seasons (years) of datasets into one by incorporating the particular season after the name of each player. **As a result,** Derrick organized all datasets from 19-20 season player stats, 20-21 season player stats, 21-22 season player stats, 22-23 season player stats, and 23-24 season player stats along with researching all the MVP voting stats for four seasons: 19-20, 20-21, 21-22, 22-23 and typed every single cell on a separate excel sheet and converted into a CSV file. Derrick merged these files into one total MVP voting results CSV file and a total player stats CSV file. Which was heavily used for all parts of the methodology including data preparation, two statistics, and the two method learnings.

Derrick volunteered and was responsible for coding the two machine learning methods: Decision Tree Classifier and the K nearest-neighbors to predict the MVP candidates for 23-24 season players (Ranked under 5) utilizing the merged total player stats CSV file, total MVP voting results in CSV file and 23-24 player stats CSV file. **As a result,** Derrick did individually code the two machine learning methods the Decision Tree Classifier and the K nearest-neighbors. Derrick did one stretch, to code beyond requirements, with the Decision Tree Classifier to visualize the most important x features that were judged heavily when predicting MVP candidates for the 23-24 season. It is placed at the end of the Decision Tree Classifier Code.

Derrick also took the initiative and took responsibility for coding the T-test and the ANOVA for the two exploratory statistics. **As a result,** Derrick coded both methods for the exploratory statistics as well as setting the null hypothesis and alternative for each method, and wrote the result if the null hypothesis was rejected or not for each method.

Derrick was assigned to work on one visualization which was the seaborn heatmap. **As a result,** Derrick coded a total of five seaborn heat maps for each season (19-20, 20-21, 21-22, 22-23, 23-24) from player stats from each season with corresponding x features to visually show the correlation between each x feature in total player stats.

In addition, Derrick worked on making the presentation along with other members. Derrick added content details and explanations for each slide; as well as the conclusion slide. Derrick added visualizations of the code method and results on the slides for better understanding from the audience during the presentation.

Cole Ouyang:

Cole was originally assigned to complete the K Nearest Neighbors model for the machine learning portion of the coding section, 2 visualizations, and assist in cleaning all the data sets used. Cole did create the original K Nearest Neighbors, but it was eventually scrapped due to overfitting. Instead, Cole took on another visualization to complete. Cole completed 5 boxplot visualizations that compared player statistics from 2019-2023, 5 scatter plots that compared players who had received MVP votes to those who had not from 2019-2023, and 5 comparative bar charts that compared the statistics of the MVP to the MVP runner up from 2019-2023.

In regards to cleaning data, Cole helped to create the 5 cohesive data sets from 2019-2023 that were used in many visualizations. Cole unified the datasets and altered them to make sure all of the statistics were “per game” statistics. Cole also volunteered to help with the presentation of the project and presented half of the slides while also helping to work on the creation of the slides.

Kelsey Keate:

Kelsey was originally assigned to make the presentation, present the presentation, and write the paper for this project.

In order to make the presentation, Kelsey found the presentation layout on Slides Carnival and wrote the basic layout of the presentation. In order to write the basic layout of the presentation, Kelsey had to interpret the code, with the help of her partners, and write about the central question, the methods, the results, and the fun fact. Some of the information she wrote was based off of old information so some of the presentation had to be redone. To present the presentation, Kelsey volunteered to present the information to the DS 110 class. Beforehand, Kelsey practiced with one of her partners in order to make sure the presentation was under four minutes and all the information that was said would be relevant to mention.

In order to write the paper, Kelsey did research on what exactly goes into determining the NBA’s MVP. She wrote the introduction based on the information that she found along with the central question. She also found many sources to help back up the central question which helped with the “Previous Work” section. To do the methodology section of the paper, Kelsey had to go back and see how the data was prepped for the coding. She also had to go back and understand how the code of the statistics and machine learning was done to write in detail. To do the results section, Kelsey had to run the code and understand how and why the code gave the results that it did for both the statistics and machine learning parts. To include the visualizations from the code, Kelsey screenshotted all of them and put them into the paper. She wrote brief descriptions about the visualizations that needed some explaining. With all of this, Kelsey was finally able to write the conclusion based off of what the group learned as a whole and based off of the results of the code.

Pavan Kumar:

Pavan was originally assigned to work on some of the descriptive visualization for the provided data.

To do this, Pavan aggregated data from online centered around the statistics of all NBA players from each NBA season starting from 2019-2020. Once this was done, he aided the rest of the group in tailoring each dataset enough for the visualizations to be done efficiently. Many of the datasets had unnecessary information that made it difficult for some variables to be isolated, so those were cleaned up.

Once the datasets were cleaned up, he worked on visualizing (in bar graph format) the top 15 NBA players in each of the PPG, APG, RPG, BPG, and SPG statistics. This was done for the season 19/20 through the current 23/24 season. After that was done Pavan worked on the visualization (in comparison bar graph format) of the stats of the MVP for each season compared to the average of the next 4 (in MVP voting standings) players. This was done for each season from season 19/20 to the 22/23 season.

Works Cited

Cabarkapa D, Deane MA, Fry AC, Jones GT, Cabarkapa DV, Philipp NM, et al. (2022) Game statistics that discriminate winning and losing at the NBA level of basketball competition. PLoS ONE 17(8): e0273427. https://doi.org/10.1371/journal.pone.0273427

Cavanaugh, S. (2024). Who is the lowest seeded player to win MVP in NBA? stats explored. https://www.sportskeeda.com/basketball/who-lowest-seeded-player-win-nba-mvp-stats-explored

Conway, T. (2024). Nikola Jokić voted NBA MVP over SGA, Luka Dončić in player poll: "he's unstoppable." https://bleacherreport.com/articles/10117967-nikola-jokic-voted-nba-mvp-over-sga-luka-doncic-in-player-poll-hes-unstoppable

ESPN. "NBA Awards: Most Valuable Player." ESPN. Accessed March 25, 2024. https://www.espn.com/nba/history/awards/_/id/33.

Rose, J. (2023). What does MVP mean in basketball? https://www.metroleague.org/what-does-mvp-mean-in-basketball/

McGregor, G. (2024). *SN NBA Awards 2024: Voting results for player of the year, All-International Teams, best young player & more*. The Sporting News. https://www.sportingnews.com/us/nba/news/sn-nba-awards-2024-voting-results-best-player/122e61e3599c2b7923491d87#MVP

Works Cited(URL) for Dataset from Kaggle.com:

1. Dataset: 2019-2020 NBA players' individual Stats:
   a. URL: https://www.kaggle.com/datasets/nicklauskim/nba-per-game-stats-201920/data
2. Dataset: 2020-2021 NBA players' individual Stats:
   a. URL: https://www.kaggle.com/datasets/umutalpaydn/nba-20202021-season-player-stats
3. Dataset: 2021-2022 NBA players' individual Stats:
   a. URL: https://www.kaggle.com/datasets/vivovinco/nba-player-stats
4. Dataset: 2022-2023 NBA players' individual Stats:
   a. URL: https://www.kaggle.com/datasets/amirhosseinmirzaie/nba-players-stats2023-season/data
5. Dataset: 2023-2024 NBA players' individual Stats:
   a. URL: https://www.kaggle.com/datasets/bryanchungweather/nba-player-stats-dataset-for-the-2023-2024

Works Cited(URL) for Dataset from pr.nba.com:

6. Dataset: 2019-2020 NBA MVP Voting Results:
    a. URL: https://ak-static.cms.nba.com/wp-content/uploads/sites/46/2020/09/2019-20-Kia-NBA-Most-Valuable-Player-Voting-Results.pdf
7. Dataset: 2020-2021 NBA MVP Voting Results:
    a. URL: https://ak-static.cms.nba.com/wp-content/uploads/sites/46/2021/06/2020-21-NBA-MVP-Voting-Results-Announcement.pdf
8. Dataset: 2021-2022 NBA MVP Voting Results:
    a. URL: https://ak-static.cms.nba.com/wp-content/uploads/sites/46/2022/05/2021-22-Kia-NBA-Most-Valuable-Player-Voting-Results.pdf
9. Dataset: 2022-2023 NBA MVP Voting Results:
    a. URL: https://ak-static.cms.nba.com/wp-content/uploads/sites/46/2023/05/2022-23-Kia-NBA-Most-Valuable-Player-Voting-Results.pdf