

Analyzing the Relationship Between NBA Draft Picks and Player Success

By: Yash Dave, Khanh Nguyen, Jeffery Jones

Background:

Since its inception in 1947, the NBA draft has been an important tool for NBA teams to rebuild or retool their roster by adding promising young high school and college players. In the modern-day draft format, there are 60 picks, and the pick order is determined by the team's record in the previous season; the worse a team's record is, the higher they pick in the draft. While many prominent NBA stars have been drafted early, like LeBron James who was drafted first overall in 2003, there is the potential for a "diamond in the rough" to be found in later picks, like Nikola Jokic who was drafted 41st overall in 2014. Teams employ numerous strategies throughout the season to try to maximize the benefit they can gain from their draft picks, from trading players for additional draft picks or vice versa, to intentionally losing games in order to ensure a higher draft pick. Whether a team is struggling and looking to rebuild, or a team is in the championship hunt, draft picks are valuable assets that teams are always looking to utilize effectively to meet their goals.

Objectives/Motivation:

Since NBA draft picks are important pieces for both contending and struggling teams, we wanted to dive deeper into player performance and success as it relates to the draft pick number that a player is selected. We want to analyze the probability of a player becoming an All-NBA player based on what pick they were drafted and the number of NBA All-Star selections for players drafted 1-60. Additionally, we want to analyze and understand what pick range produced the most players winning the various end of year awards (Most Valuable Player, Defensive Player of the Year, and Sixth Man of the Year). Finally, we wanted to see how a player's draft position affects a team's success in the playoffs. Do winning teams tend to be made of earlier picks? In conducting these analyses, we hope to provide NBA teams with useful statistics to aid in their strategy when deciding how to best utilize their draft picks.

Defining Success:

When defining player success, we decided to focus on the metrics of whether a player had earned All-NBA status, been selected as an NBA All-Star, or won one of the following end-of-season awards: Most Valuable Player (MVP), Defensive Player of the Year (DPOY), Sixth Man of the Year (6MOTY). We decided to focus on these measures instead of individual player statistics since those measures can vary drastically depending on a player's position. The measures we selected are largely independent of the player's position, therefore we decided to use them when evaluating player success.



DRAFT

Data Sources

Data Set	Description	Important Variables	Time Periods	Format	Source
NBA Draft Data (1990 - 2021)	This data set is our primary data set and provides information about every player who has been drafted by an NBA team from 1990 to 2021. It includes information about each player such as what pick they were drafted, what team they were drafted to, what college they were drafted from, as well as each player's career statistics.	Player Name, Pick Number, Pick Year	1990-2021	CSV	NBA Draft Data 1990-2021
Historic MVP, DPOY, 6MOTY Data	These are three separate data files all collected from the same data source. They each contain data about the player who won MVP, DPOY, or 6MOTY for each season since the awards were established. They all have a similar format, with each data set containing information about the player that won the award in a given season such as their name, team, and individual statistics for the season that they won the award.	Player Name, Season	1955-2023	Excel File	MVP Data DPOY Data 6MOTY Data
Historic All-NBA Team Data	At the end of each season, the top players in the league are voted to either the first, second, or third team All-NBA team. This data set contains information about each season and the players that were elected to the first second and third team in each season.	Season, Team, Player Name	1946-2023	Excel File	Historic All-NBA Team Data
NBA All-Star Data (1951-2021)	At the middle of each season, the top players in the NBA are selected as NBA All-Stars. This data set contains information about each player that was selected as an NBA All-Star in a particular season, and what team they played for.	Player Name, Team, Season	1951-2021	CSV	NBA All-Star Data
NBA Team Playoff Data (2000-2021)	At the end of each season, 16 total NBA teams compete in the playoffs. The Eastern and Western conference each have eight teams, playing a best of seven series. This data contains counting stats, wins, and losses for these teams over the past twenty years.	Team, Wins, Losses, Plus-Minus, Win Percentage, Season	2000-2021	CSV	NBA Team Playoff Data
NBA Player Data (1991-2021)	This data file contains NBA player information starting in 1991, and ending in 2021. It keeps track of each player's stats for each season including their team, and various trackable stats such as points per game, minutes per game, and games played.	Team, Player, Year, Minutes Per Game, Games Played	1991-2021	CSV	NBA Player Data

Data Manipulation Methods - Player draft success

The dataset is refined to only include essential columns: Pick ('Rk', 'Pk'), Team ('Tm'), Player name ('Player'), and draft year ('DraftYear').

Datasets detailing MVP awards, Defensive Player of the Year awards, Sixth Man of the Year awards, All NBA teams, and All-Star appearances are introduced.

For MVP, DPOY, and Sixth Man datasets, the data is aggregated by player, counting the number of seasons each player received an award.

The All NBA dataset undergoes a thorough cleaning: columns are renamed for clarity, columns corresponding to team positions are unified, and the occurrences of players in different NBA teams are aggregated.

The All-Star dataset is aggregated to count the number of All-Star appearances per player.

The final dataset was modified by introducing a new column 'Pick_Range'. This was achieved using the `pd.cut()` function to segregate players based on their draft pick number into different bins (ranges), such as '1-10', '11-20', etc.

For several accolades (like being in any All NBA team, having an MVP award, etc.), binary flags were introduced to the dataset. These flags (columns) have a value of '1' if the player has achieved the accolade and '0' otherwise.

Secondary datasets are progressively merged with the primary draft data using outer joins. This ensures representation of players across datasets, even if they didn't secure specific awards.

After integrating secondary datasets with the primary draft data, missing or unmatched records yield NaN values. These NaN values are filled with zeros, implying that the respective players didn't achieve certain accolades.

After the merging process, the data is filtered to represent players drafted between 1990 and 2021, keeping the analysis focused on this timeframe.

The structure of the All NBA dataset introduced complexity due to multiple positions and players. This was addressed by chaining different lambda functions.

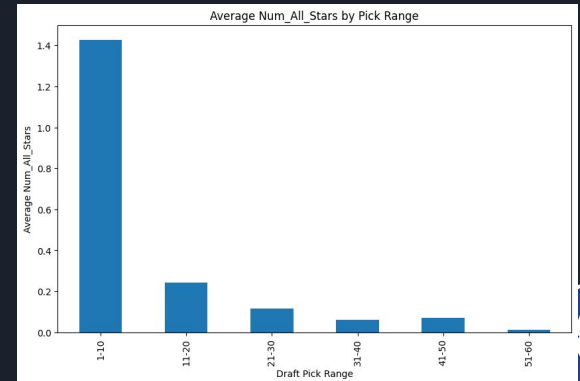
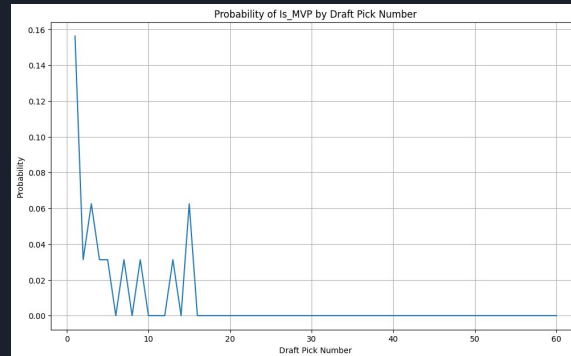
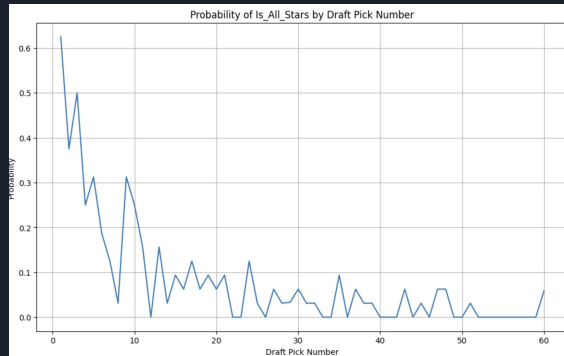
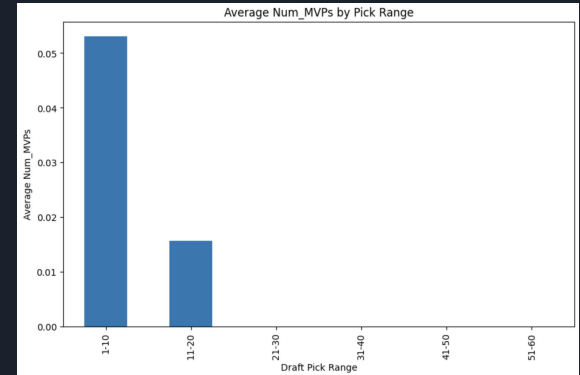


Visualization & analysis - Player draft success

A function `plot_metric_by_pick_range` was defined to create bar plots that show the average of a given metric (like the number of MVP awards) by draft pick range. The code then iterated over various metrics ('Num_MVPs', 'Num_DPOYs', 'Num_6MOTYs', etc.) to visualize each metric's average across different draft pick ranges.

Another function, `plot_probability_by_pick_range`, was defined to plot the probability of a player receiving a particular accolade based on their exact draft pick number, rather than by a range. The code then iterated over the binary achievement metrics to visualize each metric's probability across draft pick numbers.

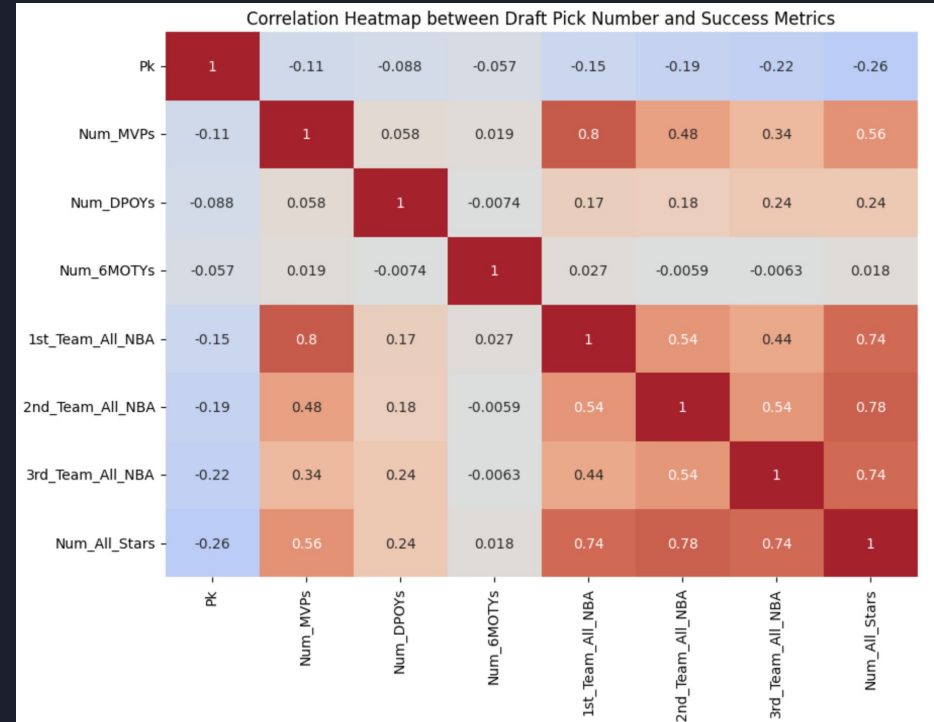
The bar plots and probability plots reveal that players picked earlier in the draft (like '1-10') generally have higher chances of achieving accolades like MVPs, All-Star appearances, etc.



Visualization & analysis - Player draft success

A subset of the dataset was taken to focus on the draft pick number and the various success metrics. A correlation matrix was computed for these columns, and this matrix was visualized using a heatmap to discern relationships between the draft pick number and the metrics of success.

The heatmap shows a negative correlation between the draft pick number and all the success metrics, indicating that a lower draft pick number (earlier in the draft) is usually associated with greater professional success. It also shows that many of the achievements are positively correlated with each other, meaning that if you're a top NBA talent you're likely to win awards in multiple categories.



Data Manipulation Methods - Team draft success

To effectively combine datasets that use both abbreviations and full team names, we generated a dictionary (nba_abbr) that maps abbreviations to their corresponding team names.

We focused on specific columns of the NBA Playoff data: 'team', 'wins', 'losses', 'win_percentage', 'plus_minus', and 'season'. Filtering was applied to retain data after the 2005 season. Team names were cleaned to ensure consistency with the NBA Team Abbreviation Data.

The dataset contains information about individual players' performances was cleaned to include only relevant columns, and players from after the year 2001. Non-standard characters in player names were filtered using replacement and string manipulations. Teams represented by abbreviations were replaced with their full names from the nba_abbr dictionary. Players were filtered based on the minimum played games and average game duration, ensuring that only significant contributors were included.

The playoff data and the player season data were combined based on the team and season. This new data frame was then merged with the draft data based on the 'Player' column. Before this merge, the draft data was stripped of unnecessary characters and was decoded to ensure compatibility.

For players who weren't drafted but played in the NBA, their draft pick position was filled with a value of 61, essentially tagging them as undrafted.

We iterate over all seasons and teams to compute the average draft pick for each team in each season. This was added as a 'Pick_Mean' column to the team_playoff dataframe.

Based on the number of wins a team had in the playoffs, a new column 'Exit' was created to indicate the stage at which a team exited the playoffs.

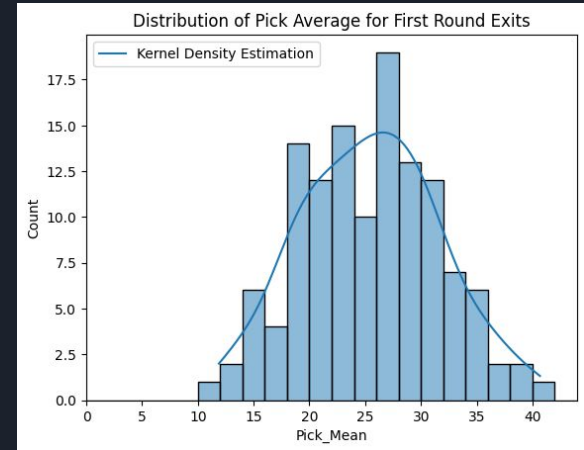
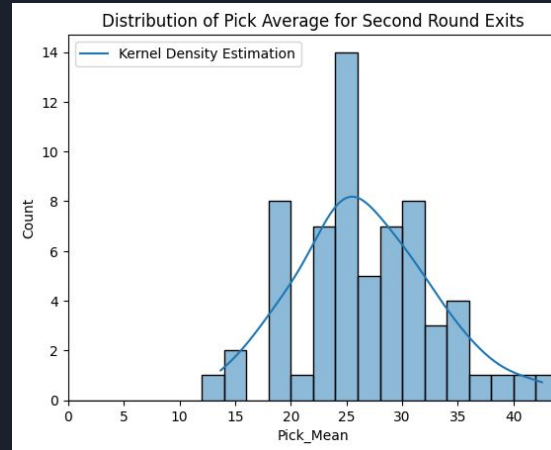
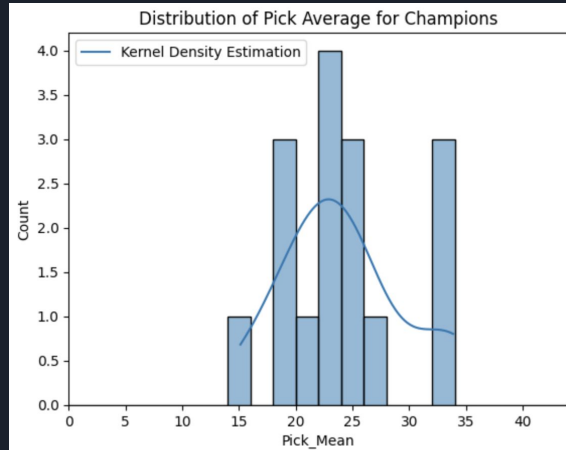
This categorization allowed for an exploration of whether a team's average draft position had any correlation with their playoff success.



Visualization & analysis - Team draft success

For each exit stage (e.g., 'First Round', 'Conference Finals'), a histogram is created to provide insights into the distribution of teams' average draft picks based on their playoff exit points. From these visualizations, one could discern that successful teams reaching the playoffs generally have their pick mean distribution centered around 20 - 25 picks on average.

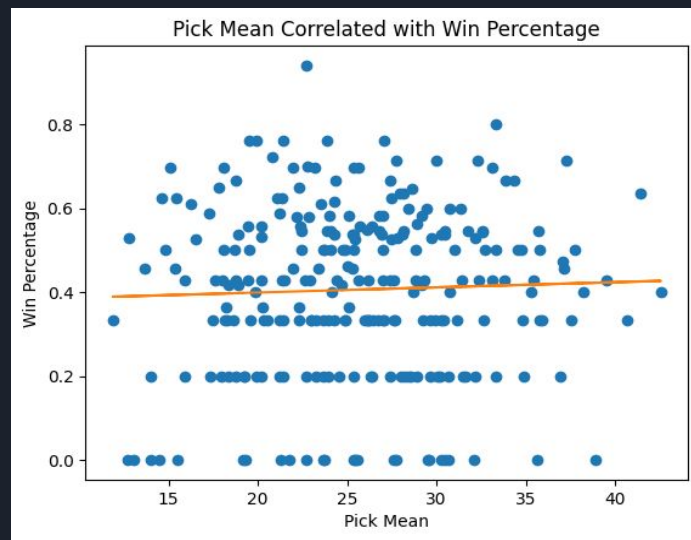
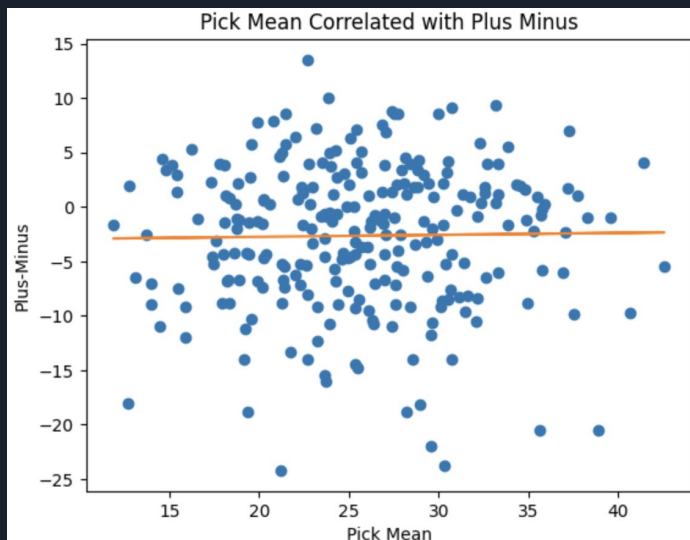
From the visualization, we have observed that teams that make it to the playoffs, regardless of their exit stage, generally have an average draft pick distribution centered around the 20 - 25 range. This could suggest that teams do not solely rely on top draft picks to achieve success in the league. Further research could be done to look at the importance of team play as opposed to looking for high drafted players for talent purposes.



Visualization & analysis - Team draft success

Additionally, two scatter plots were created in order to measure the Pick Mean's impact on other performance metrics such as Plus Minus and Win Rate. A line of best fit was added to depict the general trend for both of these plots.

The scatter plots illustrated a subtle relationship between the average draft pick and performance metrics like Plus Minus and Win Rate. The trend lines suggest only a marginal impact, reinforcing the idea that a team's average draft position doesn't significantly influence these performance indicators.



Statement of Work:

The team collaboration went well. We were able to communicate effectively via Slack and regular meetings. All group members helped in all parts of the project, and finished tasks in a timely manner. While all members of the group were willing to aid in all parts of the project, reflecting back we all agree that having more clearly defined roles and responsibilities would be beneficial in future collaborations.

Yash Dave:

Worked on data cleaning of primary and secondary datasets for individual player analysis. Helped in making the final slides for the report..

Khanh Nguyen:

Worked on visualizations and analysis for individual player analysis. Helped in making making the final slides for the report.

Jeffery Jones:

Worked on data cleaning, visualizations, and analysis for team draft analysis. Helped in making the slides for the final report.

References

Data Sources:

- “20 Years of NBA Draft Data.” Kaggle, 22 Feb. 2022, www.kaggle.com/datasets/benwieland/nba-draft-data.
- “NBA & ABA All-League Teams.” Basketball, www.basketball-reference.com/awards/all_league.html. Accessed 2 Oct. 2023.
- “NBA & ABA Defensive Player of the Year (Hakeem Olajuwon Trophy) Award Winners.” Basketball, www.basketball-reference.com/awards/dpoy.html. Accessed 2 Oct. 2023.
- “NBA & ABA Sixth Man of the Year (John Havlicek Trophy) Award Winners.” Basketball, www.basketball-reference.com/awards/smoy.html. Accessed 2 Oct. 2023.
- “NBA All Star Games 1951-2021.” Kaggle, 2 Nov. 2021, www.kaggle.com/datasets/darinhawley/nba-all-star-games-19512021?resource=download.
- “NBA MVP & ABA Most Valuable Player Award Winners.” Basketball, www.basketball-reference.com/awards/mvp.html. Accessed 2 Oct. 2023.
- “NBA Team Stats.” Kaggle, 7 Feb. 2022, www.kaggle.com/datasets/mharvnek/nba-team-stats-00-to-18?select=nba_team_stats_playoffs_00_to_21.csv.
- Vinco, Vivo. “1991-2021 NBA STATS.” Kaggle, 15 Apr. 2022, www.kaggle.com/datasets/vivovinco/19912021-nba-stats/data.

Background Information and Image:

- Chris Tomasson. “Nuggets Star Nikola Jokic Went from ‘being Drafted during a Taco Bell Commercial’ to the NBA ‘Witnessing Greatness.’” *Denver Gazette*, 11 June 2023, denvergazette.com/sports/denver-nuggets/nuggets-star-nikola-jokic-went-from-being-drafted-during-a-taco-bell-commercial-to-the/article_e2c7cada-07db-11ee-8f11-f753453bbc8b.html.
- Gerrard, Brayden. “What Are the Odds: A Statistical Analysis of Tanking in the NBA.” *Medium, Towards Data Science*, 18 Apr. 2021, towardsdatascience.com/what-are-the-odds-a-statistical-analysis-of-tanking-in-the-nba-2c5fe228cd67.
- Press, The Associated. “Today in Sports - LeBron James Is Drafted #1 Overall by the Cleveland Cavaliers.” *AP News*, AP News, 25 June 2023, apnews.com/sports/nba-draft-cleveland-cavaliers-034f7753e43d609e8769aee860a37103#:~:text=2003%20%E2%80%94%20NBA%20Draft%3A%20St.,first%20pick%20by%20Cleveland%20Cavaliers.
- Staff, NBA.com. “All-Time NBA Draft History.” *NBA.Com*, NBA.com, 23 June 2023, www.nba.com/news/history-draft.
- Image: <https://www.sportsmediawatch.com/2018/06/nba-draft-ratings-drop-espn/>