Abstract

Introduction

There are certain debates you can count on every NBA season. One of them is a week-long debate in late January surrounding which players should and shouldn't have been named All-Stars. Both fans and sports analysts can't agree on who the top-12 players in each conference are, so how are we even supposed to figure out who should be on the 75th Anniversary Team?

The NBA 75th Anniversary Team was selected by a blue-ribbon panel of current and former NBA players, coaches, general managers, and team and league executives, WNBA legends, sportswriters, and broadcasters. Voters were asked to select the 75 Greatest Players in NBA History without regard to position. Panelists did not rank their selections. Current and former players were not allowed to vote for themselves. Note, Dave DeBusschere is the only player not included in this project dataset.

In this project, we attempt to settle this debate statistically; by comparing the statistics of players listed on the 75th Anniversary to those who were "snubbed'.

Those labeled as "snubbed" mean that players did not make the anniversary list; however, public sentiment suggests that they should've. A list of snubbed player names is generated from public opinion, in which suggested names from NBA reporters and analysts are used.

Data

This project statistically compares player statistics and requires data collected from the library, "nbastatR," an interface for professional basketball data in R. Data sources *include, but are not limited to:* NBA Stats API, Basketball Insiders, Basketball-Reference, HoopsHype, and RealGM. Additionally, this project use functions such as "players_careers" and "get_nba_players_ids" to collect career regular-season statics for all players in this project. The statics retrieved from "nbastatR" are from players listed on the NBA's ″75th Anniversary**"** list and those who are labeled as "snubbed" from the anniversary list.

Statistical Methods

Since patterns for player's statistics vary greatly depending on their position, and both listed players and snubbed players could possess a variety of positions, in order to perform further analysis we need to utilize the idea of multivariate analysis and look at conditional distributions based on different levels of positions. Thus, we would generate a column called "position" that stores the main position of each player in the whole dataset, and then split the whole dataset into subsets for next steps.

Now, in order to explore the crucial factors contributing to the establishment of 75th anniversary list, we would use 2 pipelines to analyze each of the sub-dataset. First of all, we would run multiple linear regression to find out significant factors contributing to the decision of listed/snubbed; then, we would perform permutation test of all distinct features against the decision of listed/snubbed, accompanied by other forms of test such as T-test of difference in means, F-test and chi-square test, although the rather small-sized dataset could impact some of those tests, which we could analyze and compare.

Finally, concluding from results of previous pipelines, we would have a general idea on the possible factors that contribute to the 75th anniversary list.

Expected Outcomes/Conclusions.

In conclusion, our expected outcome is that the top 75 NBA players' qualification is highly related to some, but not all, factors in the dataset we collected through nbastatR. In other words, we assume selected players to have better performance in the regular seasons than other non-selected players. We plan to look at conditional distributions of both the selected and non-selected players on the player's position in our statistical method. By such a method, we can explore the decision criterions for distinct positions of players and conclude the important factors more integrally. However, there are also some limitations of our project. First of all, since many players have multiple main positions and each of them has to be in only one subset as the total number of observations is rather small, there could be some issues on the conditional distributions. Also, the NBA rules could be different based on different periods. Overall, the project has utilized lots of reliable methods, and results are expected to express a general idea on the importance of different factors in the establishment of the 75th anniversary list.

Plus, while we currently have focused on the statistical comparisons between snubbed players and listed players, one of our future alternative strategies would be to develop an alternative list of 75 players, which requires us to replace some of the currently listed players

with snubbed players. Then, to explore possibilities of such replacements, the Bayesian model could be one available method for such analysis.