Investigating NBA Player Statistics

Team Name: Team member 1, Team member 2, Team member 3, Team member 4

2020-10-28

Your written report goes here! Before you submit, make sure your code chunks are turned off with `echo = FALSE` and there are no warnings or messages with `warning = FALSE` and `message = FALSE`

Our dataset consists of various statistics for every player who has played in the NBA within the 1996-2020 seasons, ranging from their home country and college of origin, draft year and round, and various stats such as average points, rebounds, and assists per game. This dataset was found on Kaggle (https://www.kaggle.com/justinas/nba-players-data) and was originally collected using the NBA Stats API by Justinas Cirtautus, who created this dataset by filling in missing rows of data manually using data from the Basketball Reference website.

Each observation/case in the dataset represents a player and their corresponding qualities/ draft stats/game stats for a specific season. Some variables such as draft number remain constant while other such as ppg change depending on the season. A player who played for ten seasons within the timespan analyzed will have ten observations. The variables in the dataset include different aspects pertaining to the player- whether it be information about how/ when they were drafted, what country and college they are from, physical characteristics (height, weight), and game stats (number of games played, rebounds).

Basketball is a constantly evolving game, and how NBA players played twenty years ago vs today vary tremendously. From the rise of the three point shooter to the fall of the big man, it's amazing how a game's rules can be so fluid over time. Our group wishes to use this dataset to try and find more about how the highest flight of professional basketball has evolved over the years. Check the following article for more information: https://www.espn.com/nba/story/_/id/29113310/seven-ways-nba-changed-michael-jordan-bulls.

Research Question: How have the quality traits NBA teams desire in a draftee changed over time?

We will be looking at season to determine the time, the draft round and number to determine the relative importance of each player, and a way to combine a player's height, weight, and ingame statistical information in order to present an ideal position or skillset desired by teams for that draft year. We might additionally scrape data from other sources about a player's salary to determine more about their career's overall value.

Our general hypothesis is that over the past twenty or so seasons, shorter and lighter players have been prioritized in drafts. Scoring, assisting, and true shooting averages will also likely have an upwards trend, while rebounding will probably have a downwards trend. Teams will have shifted from desiring a tall and heavy rebounder and shot-blocker towards a smaller and quicker shooter with more offensive capabilities.

If we take a closer look at our data set by year, we can see that although there are some outliers, consistent data on players across the league started being kept around the year 1981. For this reason, we will pick 1981 as our start year for our analysis and filter out any recorded data from before 1981. We will also cap off our analysis before the 2019 season, as that season was ongoing at the time of upload, so metrics like gp, pts, reb, etc. would not be complete.

This data set is biased towards the offensive side of the ball, with there being only defensive statistics in this data set, so contributions given by players who devoted much of their time to the defensive side of the ball might be neglected.

To start, we will create a new dataset with just newly drafted players. This will remove all returning players to allow us to simply