Project proposal

Team Name: Elizabeth Berenguer, Vincent Huang, Mina Kim, Ashir Raza

2020-10-09

Section 1. Introduction

We are investigating various statistics for every player who has played in the NBA within the 1996-2019 seasons, ranging from their home country and college of origin, draft year and round, and various stats such as career average points, rebounds, and assists per game.

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 peak 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 end up creating an average salary column, that will definitely go up but that might not be a reliable pattern due to the natural inflation of the salary cap.

Section 2. Data description

This dataset was found on Kaggle (https://www.kaggle.com/justinas/nba-players-data) and was originally collected using the NBA Stats API. The person who created this dataset filled in missing rows of data manually using data from the Basketball Reference. Each observation in the dataset represents a player and their corresponding qualities/ draft stats/ game stats. The variables in the dataset include different aspects pertaining to the player- whether it be information about how/ when they were drafted, physical characteristics (height, weight), and game stats (career average number of games, rebounds).

Section 3. Glimpse of data

```
library(tidyverse)
nba <- read_csv("data/all_seasons.csv")
```

```
glimpse(nba)
```

```
## Rows: 11,145
## Columns: 22
## $ X1                <dbl> 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14,...
## $ player_name       <chr> "Dennis Rodman", "Dwayne Schintzius", "Earl Curet...
```

```
## $ team_abbreviation <chr> "CHI", "LAC", "TOR", "DAL", "MIA", "HOU", "LAL", ...
## $ age               <dbl> 36, 28, 39, 24, 34, 38, 25, 28, 29, 28, 27, 27, 2...
## $ player_height     <dbl> 198.12, 215.90, 205.74, 203.20, 205.74, 200.66, 1...
## $ player_weight     <dbl> 99.79024, 117.93392, 95.25432, 100.69742, 108.862...
## $ college           <chr> "Southeastern Oklahoma State", "Florida", "Detroi...
## $ country           <chr> "USA", "USA", "USA", "USA", "USA", "USA", "USA", ...
## $ draft_year        <chr> "1986", "1990", "1979", "1995", "1985", "1981", "...
## $ draft_round       <chr> "2", "1", "3", "1", "1", "2", "1", "1", "Undrafte...
## $ draft_number      <chr> "27", "24", "58", "9", "10", "29", "10", "27", "U...
## $ gp                <dbl> 55, 15, 9, 64, 27, 52, 80, 77, 71, 82, 9, 1, 13, ...
## $ pts               <dbl> 5.7, 2.3, 0.8, 3.7, 2.4, 8.2, 17.2, 14.9, 5.7, 6....
## $ reb               <dbl> 16.1, 1.5, 1.0, 2.3, 2.4, 2.7, 4.1, 8.0, 1.6, 1.5...
## $ ast               <dbl> 3.1, 0.3, 0.4, 0.6, 0.2, 1.0, 3.4, 1.6, 1.3, 3.0,...
## $ net_rating        <dbl> 16.1, 12.3, -2.1, -8.7, -11.2, 4.1, 4.1, 3.3, -0....
## $ oreb_pct          <dbl> 0.186, 0.078, 0.105, 0.060, 0.109, 0.034, 0.035, ...
## $ dreb_pct          <dbl> 0.323, 0.151, 0.102, 0.149, 0.179, 0.126, 0.091, ...
## $ usg_pct           <dbl> 0.100, 0.175, 0.103, 0.167, 0.127, 0.220, 0.209, ...
## $ ts_pct            <dbl> 0.479, 0.430, 0.376, 0.399, 0.611, 0.541, 0.559, ...
## $ ast_pct           <dbl> 0.113, 0.048, 0.148, 0.077, 0.040, 0.102, 0.149, ...
## $ season            <chr> "1996-97", "1996-97", "1996-97", "1996-97", "1996...
```