Avinash, Justin, Olusegun, Jaryt MSA 6440
3/5/2021

Project Proposal: Re-Categorization of NBA Positions

Proposal

Have you even been a fan of something for a long time and wondered, "How is [this thing I like] now related to what it was [X] years ago?" For us, this thing we like is the National Basketball Association (NBA). If you are a sports basketball, you know there are five players on the court per team. These five players have position names that often indicate their function on the court.

For example, the position "center" was known as "the big guy." Usually, the two tallest players on the court, who, on offense, would stay close to basket. If passed the ball; he'd have been close to the hoop to back-down the defense, and attempt a short-shot or layup for easy points.

Another position-function is called, "point guard." This is the playmaker, the quarterback, if you will. On offense, he would control the tempo of the game and facilitate scoring opportunities for his team. Point guards are often valued more for their assists than for their scoring. Though score, specifically, also highly valued for jump shot percentages.

Our contention is that these positions meant something concrete in terms of describing their on-court function/output as expressed by their respective statistics. To reify, using positions we discussed above: centers would have few 3-point shot attempts and point guards would have many assists, both, as compared to other positions. We believe, what was meaningful in the categorization of players, e.g. "point guard" or "center," has been lost.

Research Question

New dynamics, techniques, strategies, coaching, etc. have transformed the game of basketball from individuals performing five discrete position-functions—Center, Power Forward, Point Guard, Strong Forward, and Shooting Guard—to players wearing multiple hats, performing significantly different from players of the past in their same position. How could we take players' statistics and recategorize them into more meaningful function-titles?

Data Mining Techniques

We begin with web scraping the data to obtain the dataset. Next, we find which columns we may not need in our data, such as a player's age or what team they belong to. After, we plan to use variable selection as our main methodology for recategorizing the position of a player and finding the optimal number of positions for the number of clusters. We hope to have more of an optimal value compared to the number of positions currently in our dataset. Finally, we will use a combination of dimensionality reduction techniques and cluster analysis to help create visuals representations of these new categories.

Description

https://www.basketball-reference.com/leagues/NBA_2021_per_game.html

Format

This data frame contains the following columns along with an updated number of rows:

Rk — Rank, alphabetically ordered by surname.

Pos — Position: `C`, Center; `PF`, Power Forward; `PG`, Point Guard; `SF`, Strong Forward; `SG`, Shooting Guard.

Age — Player's age on February 1 of the season.

Tm — Team: 30 categorical variable of 3-letter team name, e.g. `CLE`, Cleveland Cavaliers; `DET`, Detroit Pistons; `MIA`, Miami Heat.

G — Games, number of games played by player.

GS — Games Started.

MP — Minutes Played Per Game.

FG — Field Goals Per Game, total shots made per game.

FGA — Field Goal Attempted Per Game.

FG% — Field Goal Percentage.

3P — 3-Point Field Goals Per Game.

3PA — 3-Point Field Goal Attempts Per Game.

3P% — 3-Point Field Goal Percentage.

2P — 2-Point Field Goals Per Game.

2PA — 2-Point Field Goal Attempts Per Game.

2P% — 2-Point Field Goal Percentage.

eFG% — Effective Field Goal Percentage. (This statistic adjusts for the fact that a 3-point field goal is worth one more point than a 2-point field goal.)

FT — Free Throws Per Game

FTA — Free Throw Attempts Per Game

FT% — Free Throw Percentage

ORB — Offensive Rebounds Per Game

DRB — Defensive Rebounds Per Game

TRB — Total Rebounds Per Game

AST — Assists Per Game

STL — Steals Per Game

BLK — Blocks Per Game

TOV — Turnovers Per Game

PF — Personal Fouls Per Game

PTS — Points Per Game