Vaulting into Victory: Optimizing for US Gymnastics Medal Count at the Paris 2024 Olympics

Benjamin Thorpe

12-16-2023

```
library(readr)
library(dplyr)
```

```
nba_data <- read_csv("NBA_Play_Types_16_23.csv")

# Sort the data
nba_data <- nba_data |>
  arrange(PLAYER_ID, PLAY_TYPE, SEASON) |>
  filter(PLAY_TYPE %in% c("Isolation", "Spotup", "PnR Ball-Handler"))

# Creating the next season's PPP and PPP_PCTL columns
nba_data <- nba_data |>
  group_by(PLAYER_ID, PLAY_TYPE) |>
  mutate(PPP_next = lead(PPP),
         PPP_PCTL_next = lead(PPP_PCTL))

# Handling the 2022-23 season by replacing with NA
nba_data <- nba_data |>
  mutate(PPP_next = ifelse(SEASON == "2022-23", NA, PPP_next),
         PPP_PCTL_next = ifelse(SEASON == "2022-23", NA, PPP_PCTL_next)) |>
  filter(POSS > 19,
         !is.na(PPP_next))
```

Steps: 1. Decide on case study idea - What teams best develop spotup shooters, effective cutters, and efficient PnR ball-handlers 2. Write introduction on why this is important + background info about the topic 3. Describe data source in detail (where I found it, website it's from, companies involved, what key variables mean, etc.) 4. Do 3-5 visualizations explaining some things about the data - use past project as inspiration 5. Write goals of the study + any expectations I have 6. Come up with model structure I will use to accomplish the goals - multilevel linear model where level one is player's attributes (num possessions, age, PPP, frequency) and level two is the team they are on that year, response will be PPP_PCTL_next to account for league-wide efficiency changes and will examine team-level coefficients 7. Write methodology section using the above 8. Run the model and visualize results nicely in a table 9. Write a couple results paragraphs 10. Write conclusion, limitations and future study, and overall summary of the case

Background

In the contemporary landscape of the National Basketball Association (NBA), the ability of teams to develop players' offensive skills has become a focal point of team strategy and success. This emphasis stems from the evolving nature of the game, where player's offensive versatility and efficiency have become paramount to earning playing time. As the pace of play has increased in every season but one going back to 2012-2013 (Scaletta), having players who can maintain strong efficiency numbers while the volume of shots taken continues to increase year over year is essential for teams looking to be successful. Additionally, as the use of analytics has become widely accepted in the league over the past 15 years, the number of three point shot attempts taken every year has went up in each season since 2010-2011 (Wal). This reliance on shooting has meant a transition to more guard-heavy and "small-ball" lineups and a decline in teams building around a dominant big man, which is demonstrated by guards "using" almost 50%

of offensive possesions in the NBA as of 2018 (Thinking). Here, this usage rate is defined as "the proportion of possessions used by a player by either shooting, winning free throws, or committing a turnover" (Thinking).

With these trends in mind, I wanted to determine if any teams stand out, either positively or negatively, in developing their players' offensive game. I will be focusing on the development of more guard-related attributes in this study, however as the league continue to see an influx of taller athletes with guard-like skills, it should be noted that these abilities can be improved upon by all NBA players. In particular, I will be examining the play types of isolation, pick and roll ball handler, and spot-up shooter to try and identify which teams are the best at developing these specific skills, as each is an important part of offense in the NBA today. Isolation plays, which rely on a player's ability to score one-on-one, demand high skill levels in ball-handling, shooting, and decision-making. Similarly, proficiency in the pick and roll as a ball handler necessitates a blend of vision, timing, and scoring ability, making it a cornerstone of modern offensive systems. Meanwhile, the spot-up shooter role, critical for spacing the floor and capitalizing on defensive lapses, hinges on precise shooting and quick decision-making. The teams that excel in cultivating these skills not only enhance their offensive firepower but also gain a tactical edge. This ability to develop offensive prowess in players aligns with the league's shift towards a more dynamic, fast-paced, and perimeter-oriented style of play, as discuessed above. Consequently, identifying which NBA teams are best at developing these specific skills provides insights into their potential for long-term success and adaptability in an ever-changing basketball environment. This analysis not only reflects on the teams' coaching and training methodologies but also on their talent scouting and player utilization strategies, making it a multifaceted and critical aspect of understanding team dynamics and future prospects in the league.

Data Description

For my analysis, I am using the "NBA_Play_Types_16_23" dataset available on Dominic Samangy's github page (DomSamangy). Dominic is currently an analyst for the New Orleans Pelicans and has over time created a great set of sports analytics resources on a google sheet called "Guide to Sports Analytics", and I found this set of data, which he scraped and uploaded into a csv, from the "Data Resources" tab.

The data itself comes from NBA.com, which has an immense database of all different kinds of basketball statistics from the NBA. The play type data I will be utilizing here is via Synergy Sports (though publicly available on NBA.com), and at a high level contains data describing how well players perform when they are involved in different play types over the course of a season, and comprises data from the 2015-2016 to 2022-2023 seasons. A row will contain a player's identifying information, the season which the data is from, a certain play type, their points per possession (PPP, explained in more detail below) and its associated percentile in that season, and multiple other metrics that I will not be using in my analysis. There are 11 different play type categories, however I will only be focusing on three as mentioned in the introduction to specifically look at a few key areas of offense. How efficient a player is at a given play type is defined here by their PPP, which is the average points per possession of that player, in that play type, in that year. So if on all possessions in a season where a player takes a catch and shoot shot their team scores an average of 1.2 points, this would be their PPP for their "spotup" row.

As far as data cleaning, I mostly kept the dataset as is besides filtering for my three play types of interest. The only additional change I made was excluding rows where a player had fewer than 20 possessions of a certain play type in a given season, as below this mark there were many outliers that may have skewed my findings (though this was chosen arbitrarily). My goal here is to determine which teams have the best offensive development program, and including data on extremely small sample sizes from players would not be beneficial as my response variable is a percentile, and thus does not take frequency into account since I want to maintain a focus on efficiency. This resulted in a final dataframe with 5,193 observations which I used to complete my analysis.

Citations

Scaletta - https://www.lineups.com/articles/why-nba-game-pace-is-at-historic-high/ Wal - https://medium.com/@gwal325/how-the-nba-has-changed-in-the-past-20-years-and-insights-to-win-23f8e9f17643 Thinking Machines Data Science - https://stories.thinkingmachin.es/nba-in-30-years/