Final Report: Top 15 Playoff Scorers per 48 Minutes in the 2024 NBA Playoffs

1. Team members and project name

Project: "Top 15 Playoff Scorers per 48 Minutes in the 2024 NBA Playoff"

Shanglin Jiang(sjiang83@usc.edu)

USCid: 1967702034

Yufan Yao(yufanyao@usc.edu)

USCid: 1450125536

2. Description

This project examines scoring efficiency in the 2024 NBA Playoffs by using points per 48 minutes as a standardized performance measure. During the playoffs, players are assigned very different roles and receive uneven amounts of playing time, which makes raw statistics such as total points or points per game difficult to compare across players. As a result, scoring output is adjusted to a fixed 48-minute scale so that players with different minute loads can be evaluated on a more comparable basis. The analysis focuses on identifying players who were able to score efficiently in limited minutes and contrasts them with players whose scoring totals were primarily driven by extended playing time. This approach provides a clearer picture of how scoring efficiency varies by role during high-stakes playoff games.

3. Data

Where the Data Comes From

All the data came from Basketball-Reference, a public website commonly used to access NBA statistics. The data were collected from the page that provides per-game player statistics for the 2024 NBA Playoffs. Instead of using a pre-made dataset, the data were collected

directly using Python. Specifically, we used pandas.read_html() to extract HTML tables from the webpage and pandas to clean and process the data into a structured format. No external APIs were used in this data collection process.

3.1 Number of Data Samples

There were about 180 to 200 players in the first set of data who had played in at least one playoff game. Players who didn't play or only played a little were taken out so that the efficiency would be more accurate. There were still more than 150 players left after filtering. We chose the top 15 players based on points per 48 minutes for more in-depth study.

4. Cleaning, looking at, and showing data in a graph

4.1 Cleaning Up the Data

There were several things that had to be done to get the data ready for analysis. The Basketball-Reference tables were cleaned up by removing repeated header rows and records that were missing important information, such as minutes played or points scored. Text variables were turned into numbers so that calculations could be done. Players who had zero or almost zero minutes were left out so that efficiency values wouldn't go up. Then, they figured out other things, like the total number of points scored, the total number of minutes played, and the points per 48 minutes, which was the main way to measure efficiency.

Total Minutes = Games Played × Minutes per Game

Total Points = Games Played times Points Per Game Points per 48 Minutes = (Total Points ÷ Total Minutes) × 48

4.2 Examining the Information

The study focused on figuring out how well players scored. We ranked players by points per 48 minutes and then compared that to more common stats like total points, minutes played, and points per game. This helped us see the difference between scoring based on efficiency and scoring based on volume. We used descriptive statistics to see how efficiency was spread out among all the players in the playoffs. This project is based on exploratory and descriptive analysis; no formal hypothesis testing or regression modeling was performed.

4.3 Making Data Look Good

We used Matplotlib to make graphs that help us understand the data. The 15 players with the most points per 48 minutes are shown in a horizontal bar chart. A scatter plot that shows minutes played against points per 48 minutes shows that some players are very good at scoring but don't get a lot of time on the field. A comparison bar chart shows how total points can depend on playing time instead of scoring rate by putting total points next to efficiency values.

4.4 Putting Data on the Internet

The data/ directory has both unprocessed and processed datasets. The raw data that was scraped is stored in data/raw/, while the cleaned datasets that are used for analysis and visualization are stored in data/processed/. Because of this structure, you can do the project again without having to scrape the source website again. The requirements.txt file has a list of all the libraries that the project needs to work.

4.5 Hypothesis and Results

This analysis looks at whether scoring efficiency changes based on playing time, whether some players seem to be underused, and whether high scoring totals are mostly due to long minutes played. The results show that scoring and efficiency don't always go hand in hand.

Some players have a lot of points per 48 minutes even though they don't play much, while others who score a lot drop in ranking when efficiency is taken into account. This means that old ways of scoring don't show how well a player did in the playoffs.

5. Changes to the First Proposal

There were some changes along the way, but the project mostly went as planned.

The minimum playing time threshold was raised after small sample sizes showed very high or very low values. To make the results more stable, data from several seasons was added. We were able to fix problems with accessing data by changing the data sources and updating t he code that scrapes the data.

Adding more visualizations also made it easier to compare scores based on volume and efficie ncy.

6. In the future

It may be possible to reduce contextual bias in the future by examining game scenarios such a s garbage time or by utilizing competitive minutes to assess an individual's efficiency. Adding usage rate or shooting efficiency and comparing how well a player does are other way s to extend the game.