DSCI 599 PROJECT REPORT

on

NBA Play-by-Play Analysis and Clustering

By Dhandeep Suglani, Bhavya Ramani, Saurabh Yadgire

Abstract:
This project aims to understand the dynamics of basketball performance at the player and team levels using play-by-play data from NBA games. The analysis seeks to uncover the underlying factors contributing to successful basketball strategies by examining the interactions between players on the court. Focusing on the Los Angeles Lakers, this project aims to help the organization identify potential trade targets based on the efficiency metric, calculating plus/minus scores along with star players from the organization to assess the impact of critical role players and lastly, clusters players from around the league to find players with complementing playing styles.

Introduction:
Finding good role players can help you decide between a good and great team. Thus, we need to analyze specific metrics that allow for finding good role players. However, noticing players that complement the team's star players is just as important. The traditional statistics may not showcase the player's ability, but defining custom metrics such as plus/minus and efficiency is essential. Then, clustering allows for discovering similar players who are likely to perform well alongside star players, leading to better team performance. Furthermore, these metrics aid in making decisions regarding player contracts, enabling teams to get talent at a lower cost. This project will specifically cater to the Los Angeles Lakers and their stars, LeBron James and Anthony Davis.

Dataset Details:
Contains NBA play-by-play for regular seasons, including the 2019 and 2020 seasons.

Methodology:

1. Efficiency Analysis:

Assists per game(APG) is a valuable metric for looking at only the raw assist total of a player. However, APG does not showcase the entire story; other factors, such as turnovers, bad passes, and violations, are also indicators of bad playmaking. These "negative plays" can hinder a player's playmaking ability and overall effectiveness to the team. Roleplayers, especially, need to be effective in the short time they are given; relying solely on APG is not enough. Players who commit a high number of assists will showcase higher turnovers, but some players exhibit high assists and low turnovers. Players like Lebron and Luka Doncic may showcase higher APG but have higher turnovers. On the other hand, players like Tyus Jones and Draymond Green(Figure 1) reflect their efficiency in higher efficiency totals.

As mentioned, defining a new metric incorporating negative plays is essential in assessing a players overall passing and playmaking ability. The formula is simple, "playmaking efficiency" is calculated: (1 / play length) x (1 + Assists: Turnovers ratio). The metric now rewards players

who make efficient plays while maintaining low bad possessions. This goes beyond just the raw APG numbers. Using these additional details allows teams to target truly efficient players and make winning plays in playmaking without slowing down the offense. Now, one of the metrics for finding players for the Laker's playmaking cause is completed.
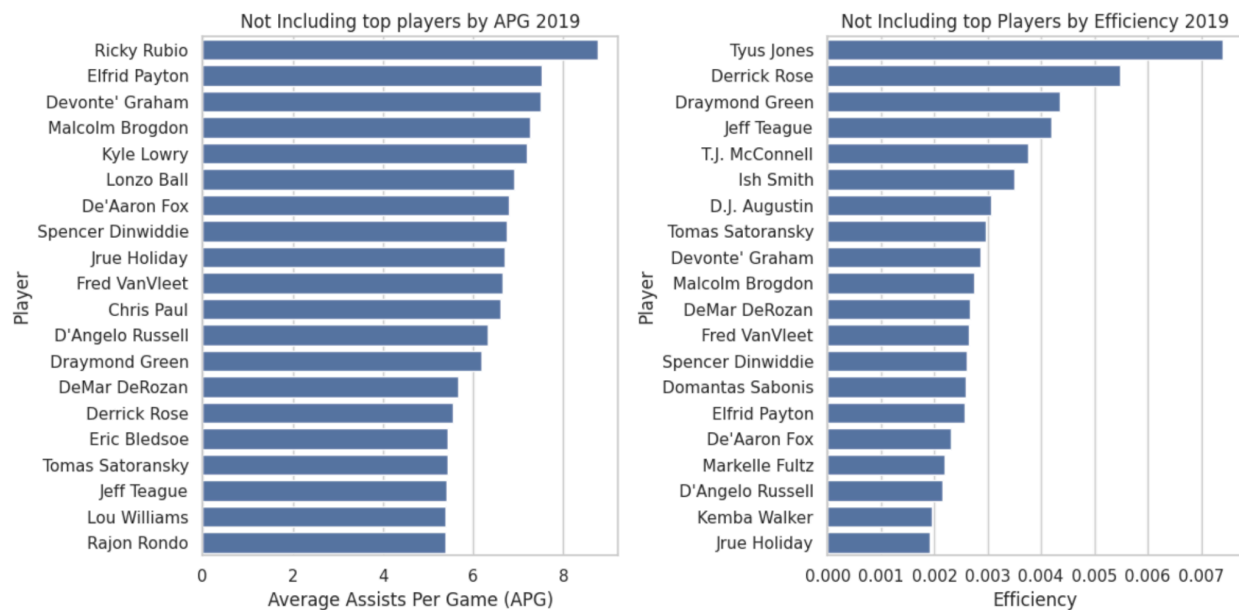

Figure 1

2. Plus/Minus Analysis

For analyzing the plus-minus scores for role players of Los Angeles Lakers our project adopts a step by step methodology to find out all game data relevant to Lakers from all NBA plays during the regular season. The next step then is to capture and find players involved in plays, calculate plus-minus scores to find the impact of selected players on game's outcome and how they assist the star players on the team.

In order to analyze the plus-minus scores, the first step adopted was to exclude all the invalid plays after which only 0 (no score), 1 (free throw), 2 (regular field goal), or 3 (three-pointer) scores are allowed. This ensured that all the calculations in the steps followed later on are based on correct plays, maintaining accuracy. Project also has functions which out all the games where the Los Angeles Lakers are playing out of all regular season games.

The next step after getting the necessary data consisting of all Lakers games, we classified each play based upon the scoring team which is either the Lakers or the opponents. We achieved this by manipulating the data entries to check which team's score increased after the play. Moreover, we make a finite list of all players on the court for both teams. By tracking each play we increase the accuracy in computing the aggregate plus-minus scores, especially in scenarios where players are substituted or there are any changes in the starting line-up.

An important step in the project was to calculate the total plus-minus scores for all role players in all the games they played in. We did this with respect to them playing with or without stars. The plus-minus value shows the difference in points when a certain players is on the court. For our project, LeBron James is selected as the superstar for 2018-2019 NBA season and LeBron James and Anthony Davis are selected as superstars for 2019-2020 season. We compared the aggregate plus-minus scores of all role players when they played with the star players versus when they played with star players off the court.

Our functions categorize player involvement into 1) on-court 2) off-court w.r.t. to star players. After that, we track the points scored by each team during these scenarios and update the plus-minus scores accordingly. We implement this using a simple check to see if the stars are on the court and modify the scoring impact based on whether or not stars are playing on the court. The final plus-minus for each player is then calculated by taking an average of the impact they create overall games they played.

Hence by analyzing the plus-minus scores, we were able to understand how role players can impact the game positively and how potential substitutions can be made to aid star players. This can have a large impact on team strategy and can allow the head coach to make better decisions regarding potential trade targets and also throws light on the performance dynamics of role players in different game situations.

	Player Name	Total Games Played	Plus/Minus w Stars	Plus/Minus wo Stars
0	Lonzo Ball	47	68.47	97.47
1	Michael Beasley	26	1.55	164.40
2	Reggie Bullock	19	32.31	133.64
3	Kentavious Caldwell-Pope	82	58.02	107.93
4	Alex Caruso	25	12.18	153.76
5	Josh Hart	67	62.64	103.31
6	Brandon Ingram	52	77.91	88.04
7	Kyle Kuzma	70	124.35	41.60
8	Rajon Rondo	46	52.18	113.76

Figure 2
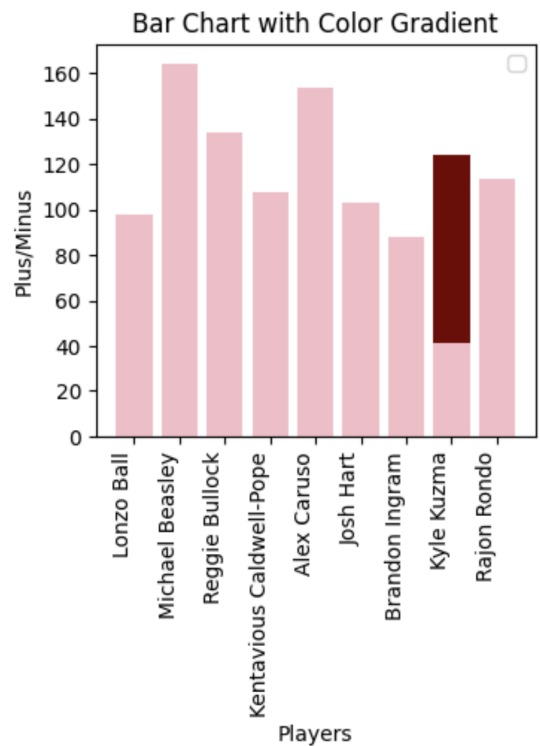
Bar Chart with Color Gradient

Figure 3

3. Clustering Analysis:

The goal is to optimize player combinations for the Los Angeles Lakers alongside LeBron James and Anthony Davis while trying to minimize total spending. Using plus-minus and efficiency scores, we compare Lakers' players with the top 10 players league-wide for the respective season. Thus by implementing the k-means clustering algorithm, we group Lakers' players with those from the rest of the league who showcase similar statistics. These clusters identify potential trade options for the Lakers, and show which players to target for free agency,

Following are the clustering results upon clustering:
2018-19 season:
- Cluster 1: Rajon Rondo
- Cluster 2: Lonzo Ball
- Cluster 3: Kentavious Caldwell-Pope, Brandon Ingram, Miles Bridges, Kelly Oubre Jr.

- Cluster 4: Thabo Sefolosha, Markelle Fultz, JaVale McGee, Torrey Craig, Terrance Ferguson, Josh Jackson, Jason Smith, Justin Jackson

2019-20 season:
- Cluster 1: Reggie Bullock
- Cluster 2: Josh Hart, Kyle Kuzma, Kentavious Caldwell-Pope, Alex Caruso, Johnathan Motley, DeAndre' Bembry, Malcolm Miller, DaQuan Jeffries, Jonathan Isaac, Nicolas Batum, Dewan Hernandez
- Cluster 3: Brandon Ingram, Rajon Rondo, Lonzo Ball, Tim Frazier
- Cluster 4: Adam Mokoka, Ryan Broekhoff

Conclusion & Future Plans:

In this project we explored creation of custom key metrics such as efficiency and also performed plus-minus analysis to find players whose play style is suited to star players on the Lakers. We used k-means clustering to discover such players without spending extra money for contract extensions. Potentially, we can further improve these metrics and exploring different clustering methods to fine tune the roster selection process. In addition to that, the insights obtained from this project can be used to drive future decisions regarding player contracts which can lead to signing role players for cheap. This project showcases the ability of clustering and other data analysis in the NBA.

Link to code:

[Efficiency](#)

[Plus Minus](#)

[Clustering](#)