
# Who is the next NBA GOAT?

Mitch Larocque
Max Van Sickle

# Introduction:

- Both avid basketball fans
- With the advent of the NBA's 75th anniversary they released a list of the top 76 players of all time
- A very nostalgic season has had us reminiscing about the past but also looking towards the future
- In the Modern NBA, most discussions are about the best right now (LeBron, KD, Curry), not young talent

**Question: Which current NBA players under the age of 25 will end up among the legends?**

# Data:

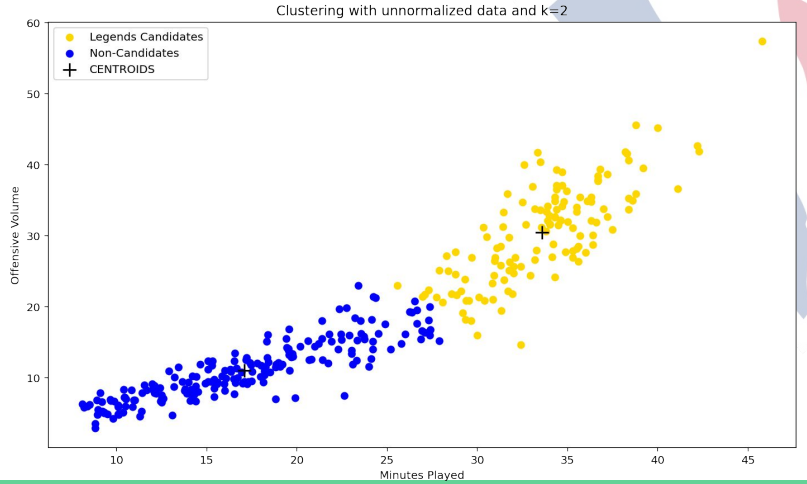
- We used two datasets both from basketball-reference.com
- The first dataset contained statistics about the 76 players on the NBA 75th anniversary team
- The second dataset was statistics for current players in the NBA
- Averaged out last three seasons for all players, kept players <=25 Years old, played >8 Min per game, >5 games per season.
- Only kept comparable data

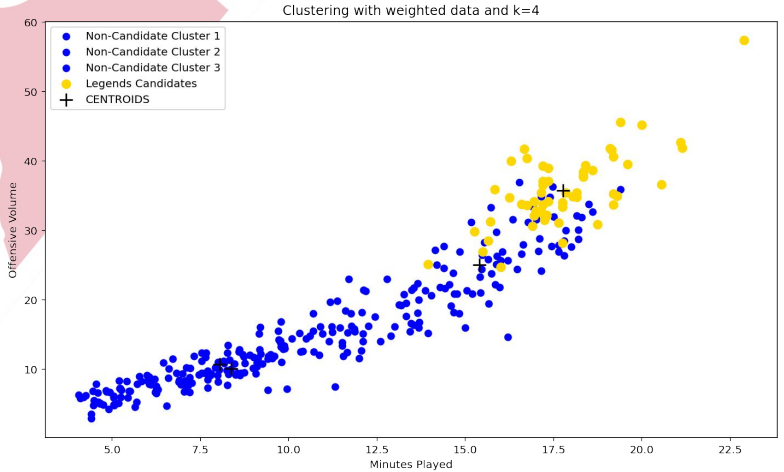
| | Player | From | To | G | MP | PTS | TRB | AST | STL | BLK | FG% | 3P% | FT% | WS | WS/48 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | Kareem Abdul-Jabbar | 1970.0 | 1989.0 | 1560.0 | 36.8 | 24.6 | 11.2 | 3.6 | 0.9 | 2.6 | 0.559 | 0.056 | 0.721 | 273.4 | 0.228 |
| **12** | Wilt Chamberlain | 1960.0 | 1973.0 | 1045.0 | 45.8 | 30.1 | 22.9 | 4.4 | NaN | NaN | 0.540 | NaN | 0.511 | 247.3 | 0.248 |
| **32** | LeBron James | 2004.0 | 2022.0 | 1316.0 | 38.2 | 27.0 | 7.4 | 7.4 | 1.6 | 0.8 | 0.504 | 0.345 | 0.734 | 242.6 | 0.232 |
| **40** | Karl Malone | 1986.0 | 2004.0 | 1476.0 | 37.2 | 25.0 | 10.1 | 3.6 | 1.4 | 0.8 | 0.516 | 0.274 | 0.742 | 234.6 | 0.205 |
| **35** | Michael Jordan | 1985.0 | 2003.0 | 1072.0 | 38.3 | 30.1 | 6.2 | 5.3 | 2.3 | 0.8 | 0.497 | 0.327 | 0.835 | 214.0 | 0.250 |

**Top 5 highest win shares of the 75th anniversary team**

# Our Model

- Preliminary Outlier detection to get exploratory results of best young player candidates based on highest WS
- 1: UV Decomposition to fill in missing values for 75th Team Players
  - Blocks, Steals, 3P% not recorded until 1974, 1979
- 2: K-Means clustering with un-normalized data
- 3: Normalized data to include better weights for certain stats (WS/48, scoring stats)
- 4: Finally, run k-means clustering to produce satisfactory clusters
  - Goal was to maintain a high number of legends with a low number of candidate young players



# Results:

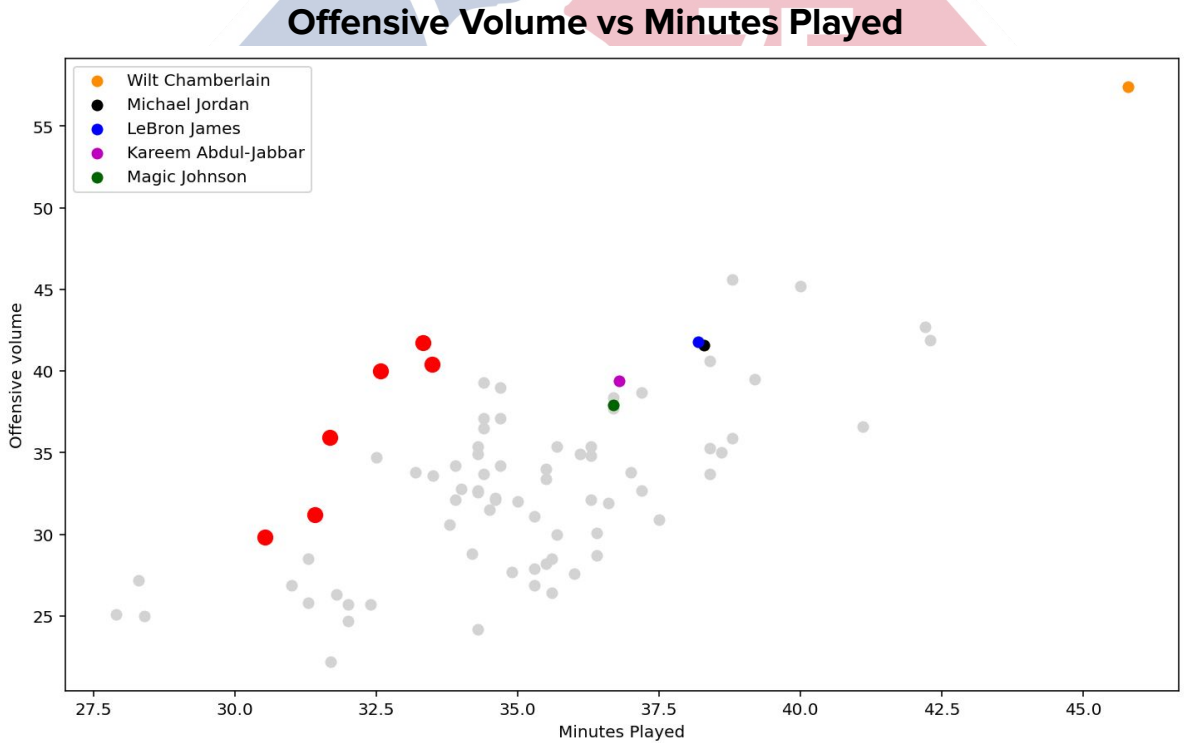
**The top six current young NBA players who are most likely to be among the legends are:**

| | Player | PER |
|---|---|---|
| 0 | Nikola Joki | 31.3 |
| 1 | Zion Williamson | 27.1 |
| 2 | Luka Don | 25.3 |
| 3 | Karl-Anthony Towns | 23.1 |
| 4 | Kristaps Porzi | 21.3 |
| 5 | John Collins | 20.6 |

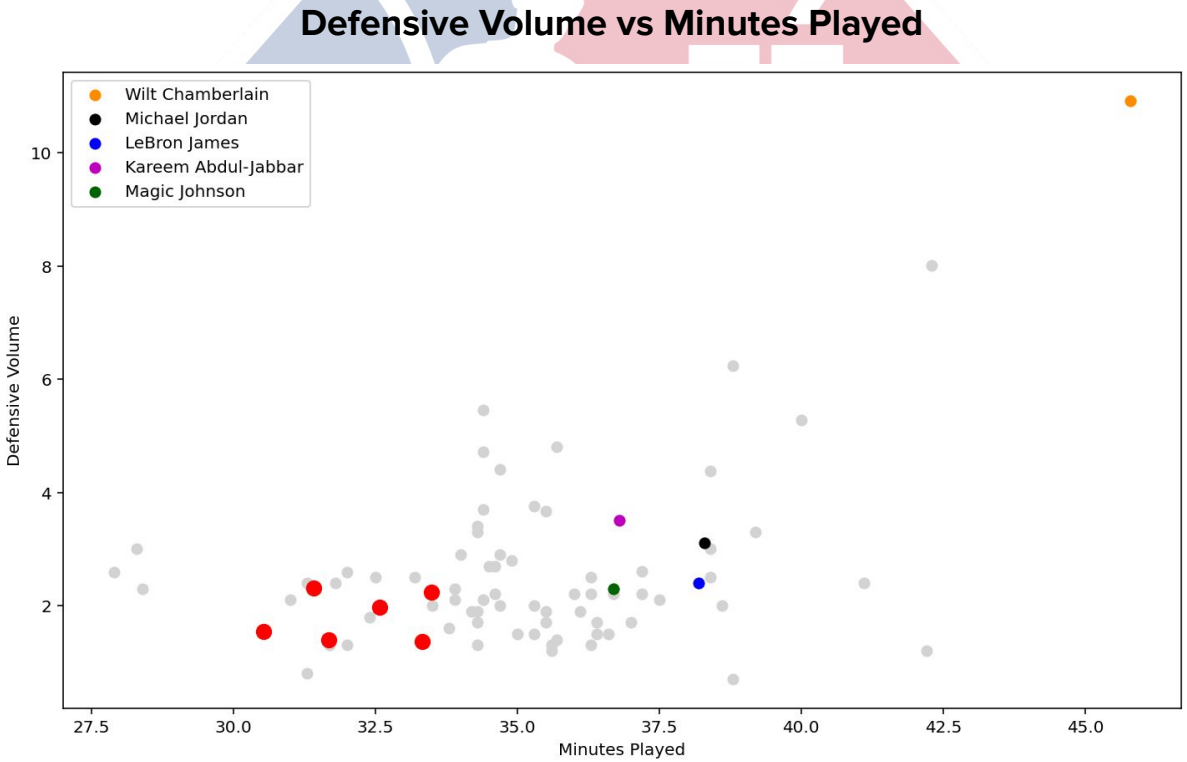
**Here we ordered the top 6 players by the most accepted modern advanced statistic for ‘best player’ which is PER (player efficiency rating). The answer to our original question then is:**

**Nikola Jokic is the best young player in the NBA and is most likely to be among the legends.**

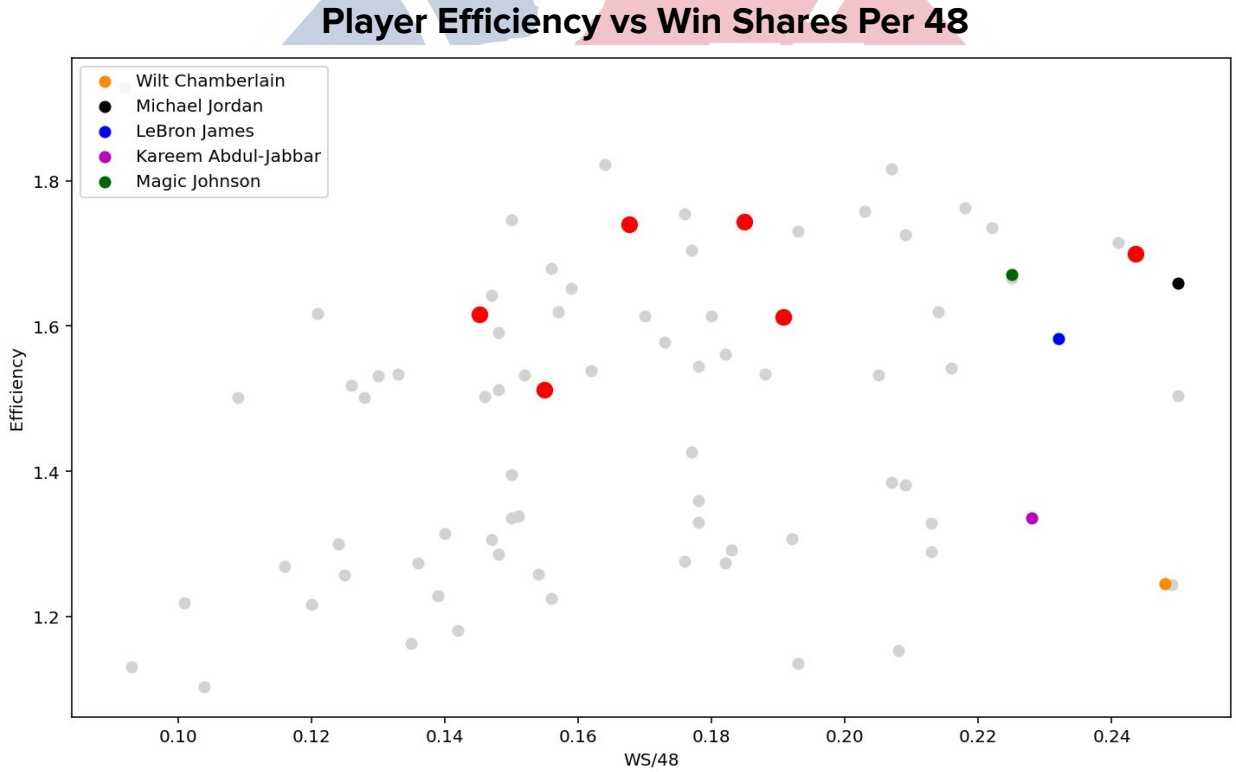
# Results best young players compared to legends:



# Results best young players compared to legends:



# Results best young players compared to legends:



# Conclusion and Modifications

- Some of our methods used our personal discretion to determine values:
  - Weighting each statistic and its importance
  - Choosing K-values for clustering
  - Ranking final 6 young players by PER
- Could have been determined with linear models to remove any bias

- Using other methods like neural nets instead of K-means Clustering
  - NNs could decompose given statistics for each player into abstract concepts to place players into categories more effectively

# Thank You!

Questions?