# Analysis of Player Offensive Efficiency using AI


**Presented By:**

Group 7

John Baldi
Shweta Madhale
Babafemi Sorinolu
Su Zhang

# Introduction

Leveraging AI techniques in the sports world yields insightful and remarkable reports, and organizations invest millions into research and development each year


source: AIWS

# Problem Statement

- NBA player offensive efficiency has been tracked using several metrics in the past
    - Field Goal % (FG%)
    - Effective Field Goal % (eFG%)
    - True Shooting % (TS%)
- Can we use AI to create an improved metric that adjusts for the difficulty of the shots being taken?
- Goal: be able to use our metric to provide recommendations to teams on who the most efficient players are

NBA

# Techniques to tackle the problem

- K-means clustering
- Reward-Penalty scoring
- Rank Biased Overlap (RBO)

- Other approaches tried
    - Naive-bayes
    - SVM
    - Linear regression
    - K-nearest neighbor
    - Random Forest

# Previous Work

Estimating an NBA player's impact on his team's chances of winning.
 - Sameer K.Deshpande, Shane T.Jensen

- Overcome limitation of traditional measurement for players' efficiency by including game context such as game clock and shot clock.
- Win probability framework using Bayesian linear regression model to estimate an individual player's impact on the court.
- Use several posterior summaries to derive rank-ordering of players within their team and across the league.

| | Lineup | Impact Score | Minutes |
|---|---|---|---|
| 1 | Stephen Curry, Klay Thompson, Andre Iguodala David Lee, Andrew Bogut | 2.98 | 780.25 |
| 2 | Chris Paul, J.J. Redick, Matt Barnes Blake Griffin, DeAndre Jordan | 2.88 | 88.57 |
| 3 | Stephen Curry, Klay Thompson, Andre Iguodala David Lee, Jermaine O'Neal | 2.82 | 31.75 |
| 4 | George Hill, Lance Stephenson, Paul George David West, Roy Hibbert | 2.58 | 1369.38 |
| 5 | Mario Chalmers, Ray Allen, LeBron James Chris Bosh, Chris Andersen | 2.57 | 34.28 |
| 6 | Patrick Beverley, James Harden, Chandler Parsons Terrence Jones, Dwight Howard | 2.51 | 589.97 |
| 7 | Mario Chalmers, Dwyane Wade, LeBron James Chris Bosh, Chris Andersen | 2.46 | 26.2 |
| 8 | C.J. Watson, Lance Stephenson, Paul George David West, Roy Hibbert | 2.42 | 118.27 |
| 9 | John Wall, Bradley Beal, Trevor Ariza Nene Hilario, Marcin Gortat | 2.38 | 384.03 |
| 10 | Patrick Beverley, James Harden, Chandler Parsons Donatas Motiejunas, Dwight Howard | 2.38 | 65.58 |

# Previous Work

**Efficiency and productivity evaluation of basketball players' performance.**

-José Vitor Senatore, Gilbert Fellingham and Leonardo Lames

- To overcome limitations of the Ei (Efficiency Index) and GRS (Game-related-statistic), new index BEi (Basketball Efficiency Index) and BPi (Basketball Productivity Index) is introduced.
- Main purpose of these indexes is to summarize the players’ performance in terms of both efficiency and productivity.
- The new indexes also overcome the bias derived from both the players’ playing time and game pace.

$$\text{BEi} = \text{PTSeq} + \text{OREBeq} + \text{ASTeq} - \text{TOVeq} + \text{DREBeq} + \text{STLeq} + \text{BLKeq} - \text{PFeq}$$

$$\text{BPi} = \text{Participation Factor} * \text{BEi}$$

# Previous Work

Measuring players' importance in basketball using the generalized Shapley value.
- Rodolfo Metulini, Giorgio Gnecco

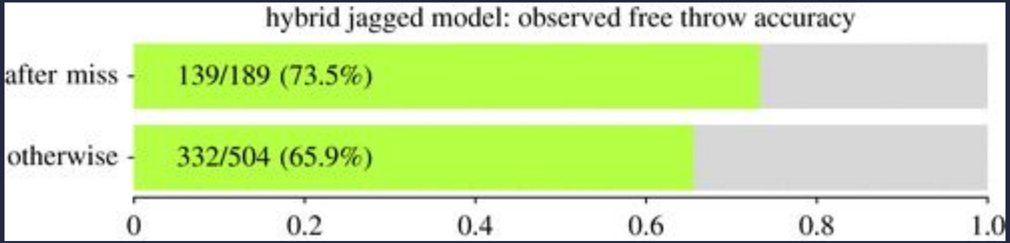
- Evaluate each players' importance and the players' marginal contribution to the utility of an ordered subset of players.
- Through a generalized version of Shapley value, to determine the probability a certain lineup to win the game.
- Using greedy algorithm, by re-computing the generalized Shapley value conditions to find the best lineup.

| Player | (1st step) | (2nd step) | (3rd step) | (4th step) |
|---|---|---|---|---|
| **Donovan Mitchell** | **0.0551** | – | – | – |
| **Mike Conley** | 0.0535 | **0.0821** | – | – |
| **Joe Ingles** | 0.0383 | 0.0547 | **0.0941** | – |
| **Royce O'Neale** | 0.0512 | 0.0707 | 0.0924 | **0.1076** |
| Bojan Bogdanović | 0.0485 | 0.0664 | 0.0922 | n.a. |
| Jordan Clarxson | 0.0337 | 0.0231 | n.a. | n.a. |
| George Niang | 0.0398 | n.a. | n.a. | n.a. |
| Miye Oni | 0.0000 | n.a. | n.a. | n.a. |
| Trent Forrest | n.a. | n.a. | n.a. | n.a. |
| Derrick Favors | n.a. | n.a. | n.a. | n.a. |

# Previous Work

Predictive Bayesian selection of multistep Markov chains, applied to the detection of the hot hand and other statistical dependencies in free throws. - Joshua C. Chang

- Debate of hot hand effect addressed multivariate methods from recent analyses.
- Using multistep Markovian models to predict the probability of missing free throws.
- Using single player (Lebron James) to exam hot hand phenomenon after missing free throws.



# More Previous Work

- Analysis of NBA Players and Shot Prediction Using Random Forest and XGboost Models.
  - Sahar A. EL Rahman
- Sports Analytics Algorithms for performance prediction.
  -Konstantinos Apostolou, Christos Tjortjis
- Basketball Win Percentage Prediction using Ensemble-based Machine Learning. -Dhruv Sikka, Rajeswari D
- Machine Learning Techniques for Analyzing Athletic Performance in Sports using GWO-CNN model.
  - G.Radhakrishnan, T.Parasuraman, D.Harigaran

# Other Relevant Techniques

1. K-means cluster analysis is used to divide game results into different groups.

2. Descriptive statistics were used to calculate for each team performance indicator relative to match outcome.

3. Binary logistic regression was used to build a linear probability model to predict match win.

4. The Gray Wolf Optimization (GWO) based Machine Learning (CNN) Technique is used for the purpose of performance plan prediction in order to bring about an improvement in the degree of quality achieved.

# Dataset Overview

<table>
<tr><td>Key</td><td>Attribute</td></tr>
<tr><td>Column labels</td><td><table>
<tr><td>Game ID</td><td>Final Margin</td><td>Period</td><td>Shot Distance</td><td>Points Type</td></tr>
<tr><td>Shot Clock</td><td>Dribbles</td><td>Touch Time</td><td>Closest Defender</td><td>FGM</td></tr>
<tr><td>Closets Defender Player ID</td><td>Closest Defender Distance</td><td>Player Name</td><td>Player ID</td><td>Location</td></tr>
</table></td></tr>
<tr><td>Data Samples</td><td>128,069</td></tr>
<tr><td>Period</td><td>Oct 2014 - Mar 2015</td></tr>
<tr><td>Source</td><td>Kaggle</td></tr>
</table>

# Implementation Details

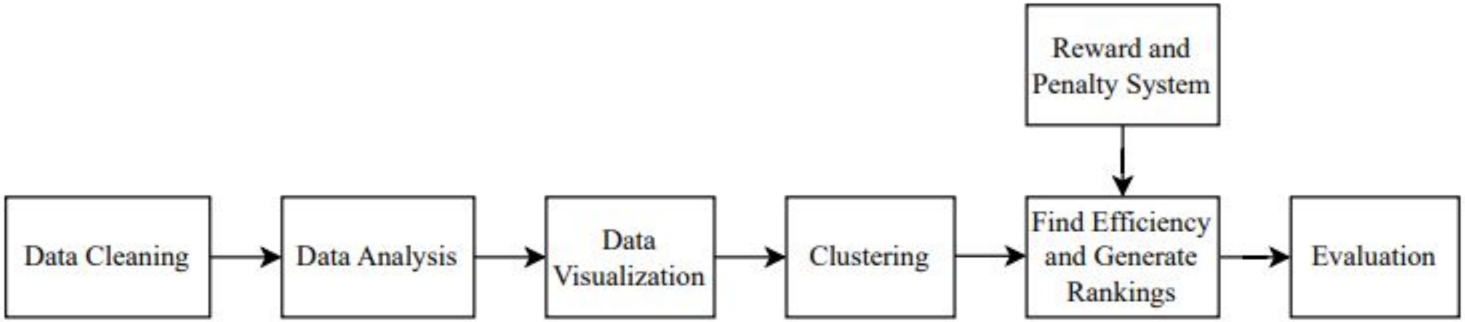



**Fig. Flow of Implementation**

## Data Cleaning

- Remove irrelevant features
- Fixed formatting
- Dropped empty data

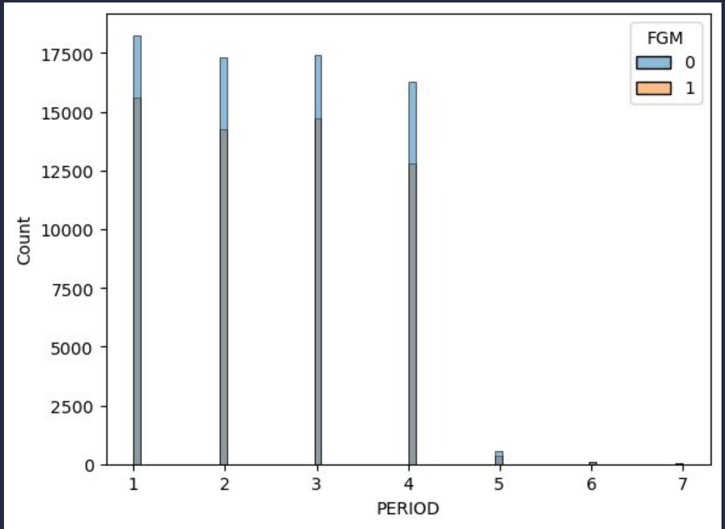
## Data Analysis

- Found defender's blocking efficiency
- Create ranking for defenders
- Aggregate features for each player

## Data Visualization

- Plot histograms and count plots
- Heatmap for correlation matrix

# Visualized Data



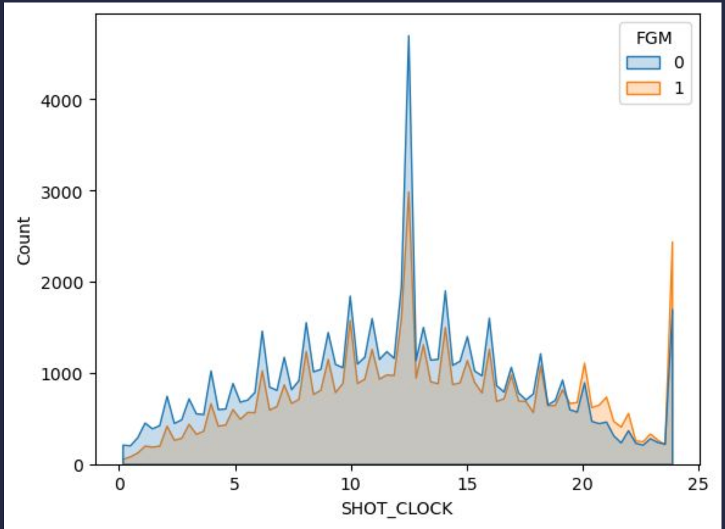




Fig. Histogram of period and shot clock against attempted shots

# Visualized Data

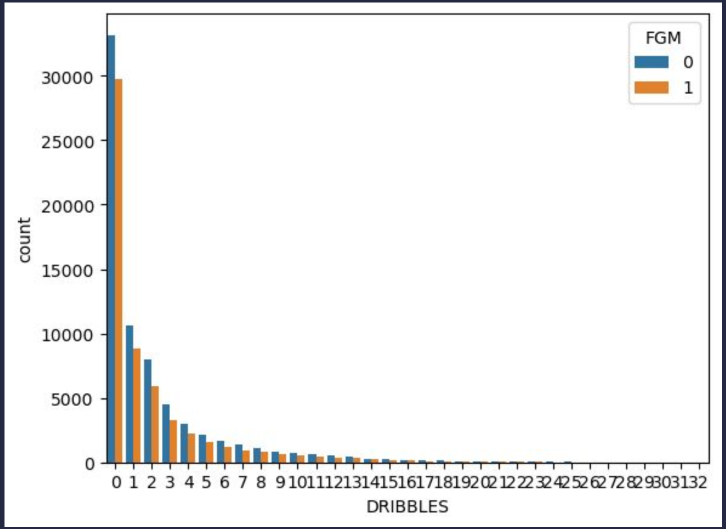



Fig. Count plot of dribbles against attempted shots

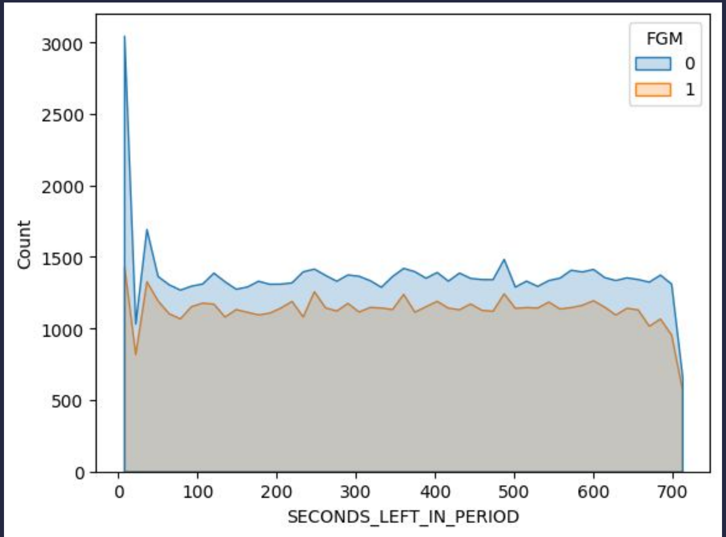
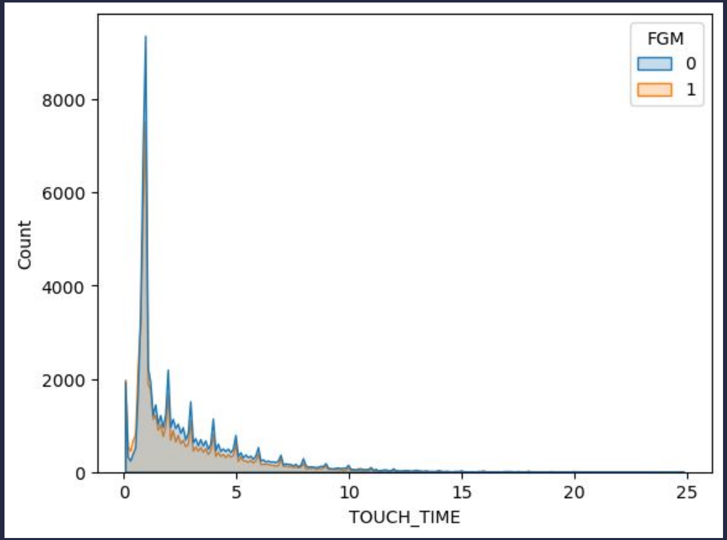



Fig. Histogram of seconds left in period against attempted shots

# Visualized Data



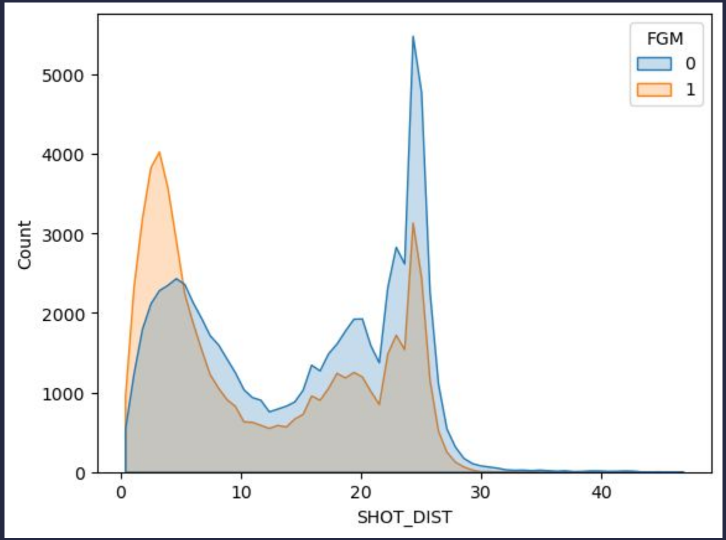




Fig. Histogram of touch time and shot distance  against attempted shots

# Visualized Data

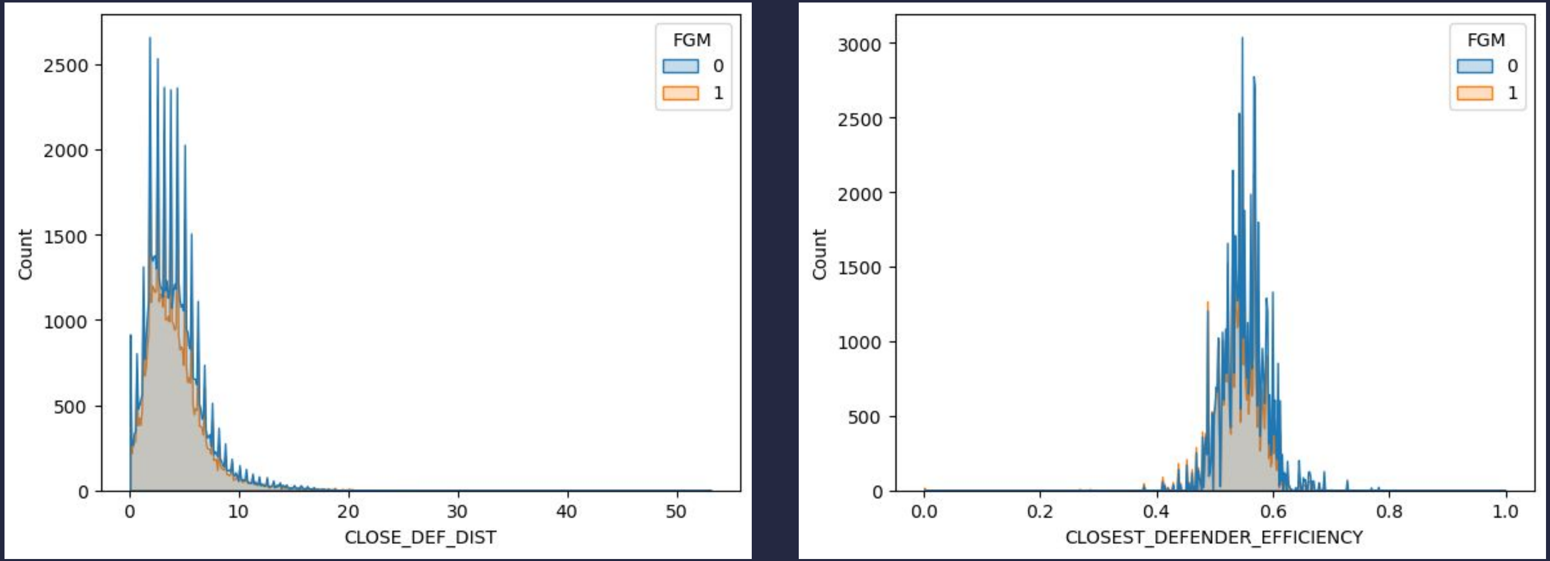



Fig. Histogram of closest defender distance and efficiency  against attempted shots

# Correlation Matrix

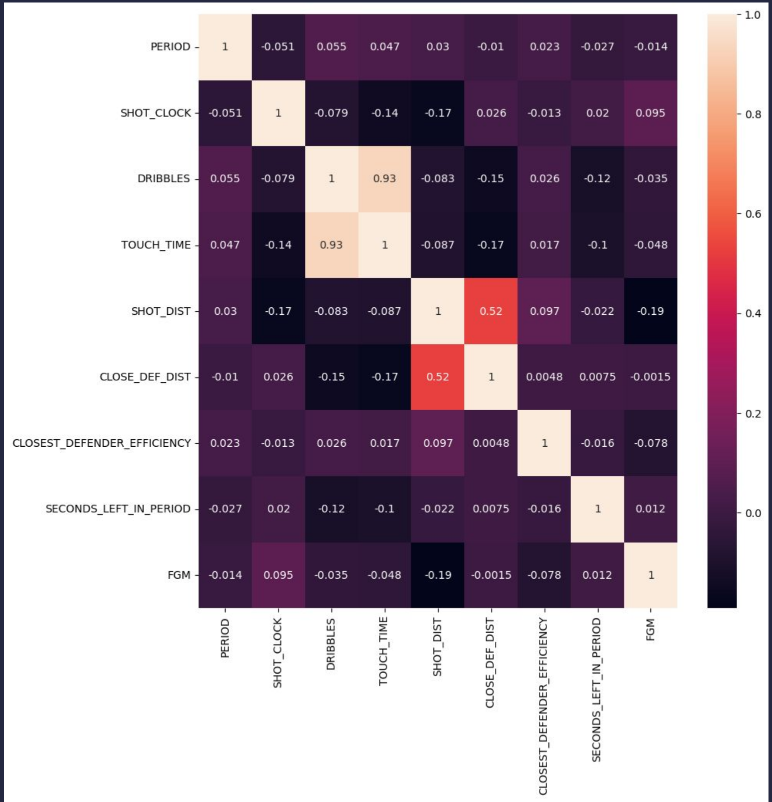
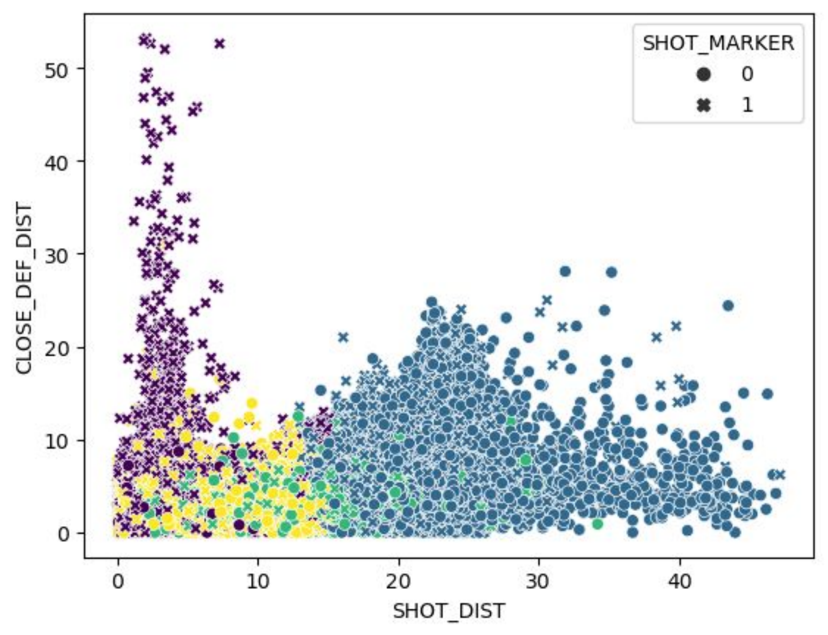



Fig. Heatmap for correlation between the features

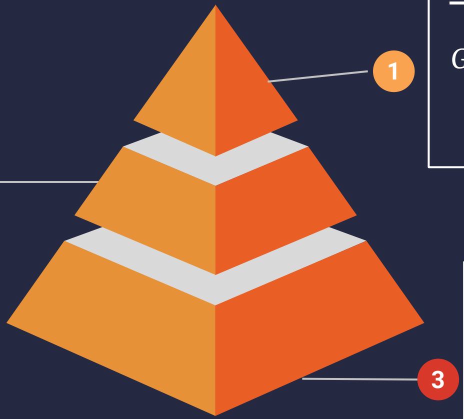
# Clustering




- Method : K-Means
- Total clusters : 4
- Clustering context : Successful shots per attempted shots
- Clusters based on **shot difficulty**

| Cluster | Cluster Context |
|---|---|
| 0 (Blue) | 0.3738 |
| 1 (Green) | 0.4119 |
| 2 (Purple) | 0.6007 |
| 3 (Yellow) | 0.4843 |

# Estimating Player's Offensive Efficiency




## Find threshold of clusters

*Get average value of each feature and find* ***average threshold efficiency***

## Reward and Penalize

*Using threshold for each feature, reward and penalize each player based on cluster efficiency vs average efficiency*

## Find efficiency and rank players

*With rewards and penalties set, get adjusted efficiency, rank players based on this*

# Comparing Metrics

## Our Ranking basis

- Adjusted Efficiency based on:
  - ❑ Period
  - ❑ Shot Clock
  - ❑ Dribbles
  - ❑ Touch time
  - ❑ Shot distance
  - ❑ Closest Defender Distance
  - ❑ Closest Defender Efficiency
  - ❑ Seconds left in period

## ESPN Ranking basis

- True Shooting % based on:
  - ❑ Points adjusted 3-point efficiency
  - ❑ 2-point efficiency
  - ❑ Points Adjusted Free throw efficiency

## Rank Biased Overlap (RBO)

- Similarity measure between rankings

$$RBO(S,T,p) = (1-p)\sum_{d=1}^{\infty}(p^{d-1}).A_d$$

**Overlap % = 58.084%**

# Evaluating Results

- Overlap of 58.084% is a positive outcome and isn't meant to be 100%
    - Our proposed metric should have similarity and difference with current metrics
    - Incomplete data likely resulted in disagreements in rankings
- Our ranks accomplished our initial goal
    - Includes the best shooters who consistently take more difficult shots
    - Does not include Centers who shoot a high percentage but take easier shots

| Our Ranks: player_name | Our Ranks: Adjusted_Efficiency | Rank | ESPN Ranks: player_name | ESPN Ranks: TS% |
|---|---|---|---|---|
| kyle korver | 0.597 | 1 | Kyle Korver, ATL | 0.699 |
| jj redick | 0.532 | 2 | Tyson Chandler, DAL | 0.697 |
| al horford | 0.531 | 3 | Brandan Wright, DAL/BOS/PHX | 0.66 |
| chris paul | 0.523 | 4 | Luke Babbitt, NO | 0.639 |
| dirk nowtizski | 0.518 | 5 | Stephen Curry, GS | 0.638 |
| anthony davis | 0.508 | 6 | DeAndre Jordan, LAC | 0.638 |
| stephen curry | 0.505 | 7 | Kevin Durant, OKC | 0.633 |
| courtney lee | 0.505 | 8 | Meyers Leonard, POR | 0.631 |
| david west | 0.500 | 9 | Rudy Gobert, UTAH | 0.627 |
| wesley matthews | 0.491 | 10 | Jonas Valanciunas, TOR | 0.623 |
| chris bosh | 0.490 | 11 | JJ Redick, LAC | 0.622 |
| serge ibaka | 0.490 | 12 | Chris Andersen, MIA | 0.622 |
| klay thompson | 0.489 | 13 | Hassan Whiteside, MIA | 0.619 |
| nene hilario | 0.488 | 14 | Aron Baynes, SA | 0.618 |
| khris middleton | 0.486 | 15 | James Johnson, TOR | 0.617 |

# Conclusion and Future Analysis

- Using our metric for offensive efficiency, teams can build more successful rosters and gameplans
- Future Analysis for better results
  - Get more complete and current dataset
  - Gather additional data points about the context of a shot
    - Player movement
    - Shot angle
    - Additional defense data

| Our Best 2014-2015 Lineup |
| --- |
| PGs - Chris Paul, Stephen Curry |
| SGs - Kyle Korver, JJ Redick |
| SFs - Klay Thompson, Khris Middleton |
| PFs - Dirk Nowitzki, David West |
| Cs - Al Horford, Anthony Davis |

# References


1. Senatore, J., Fellingham, G., & Lamas, L. “Efficiency and productivity evaluation of basketball players' performance.”, Motriz: Revista de Educação Física, 2022, doi: 10.1590/s1980-657420220004922
2. M. S. Oughali, M. Bahloul and, S. A. El Rahman, "Analysis of NBA Players and Shot Prediction Using Random Forest and XGBoost Models," 2019 International Conference on Computer and Information Sciences (ICCIS), Sakaka, Saudi Arabia, 2019, pp. 1-5, doi: 10.1109/ICCISci.2019.8716412.
3. K. Apostolou and C. Tjortjis, "Sports Analytics algorithms for performance prediction," 2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA), Patras, Greece, 2019, pp. 1-4, doi: 10.1109/IISA.2019.8900754.
4. Chang, Joshua C.. “Predictive Bayesian selection of multistep Markov chains, applied to the detection of the hot hand and other statistical dependencies in free throws.” Royal Society Open Science 6 (2017): n. Pag.
5. Deshpande, Sameer K. and Jensen, Shane T.. "Estimating an NBA player’s impact on his team’s chances of winning" Journal of Quantitative Analysis in Sports, vol. 12, no. 2, 2016, pp. 51-72.
6. Kaggle dataset - https://www.kaggle.com/datasets/dansbecker/nba-shot-logs

# Thank you!