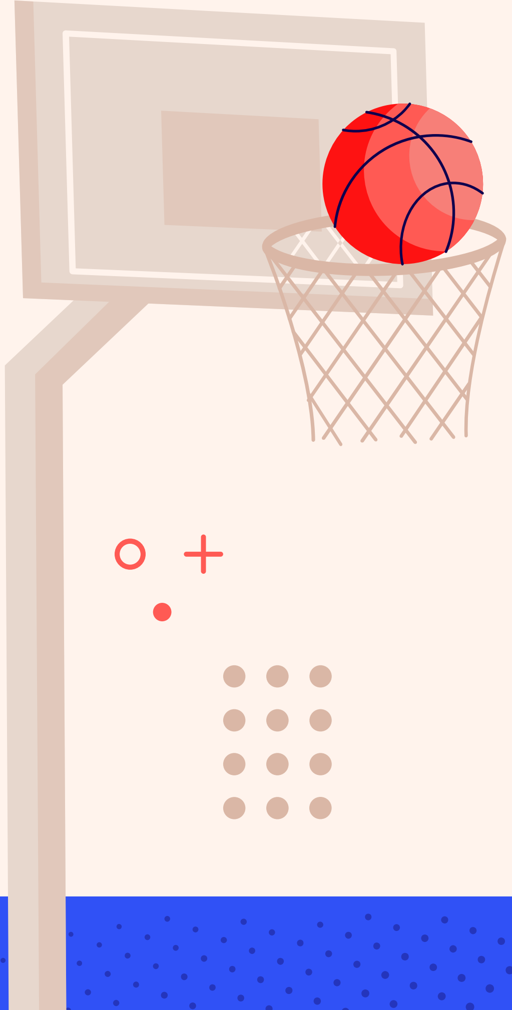


NBA Shot Selection

Ibrahim, Mason, Kabir, Kapil

TABLE OF CONTENTS

01

Research Question + Motivation

How has the distribution and efficiency of NBA shot selection evolved across eras, and what do player clustering patterns reveal about the most optimal team compositions in each era?

- Player tendencies vary widely — some specialize in restricted area finishes, others thrive on midrange pull-ups or catch-and-shoot 3s.
- Team strategies often emphasize certain zones (e.g., Moreyball: 3s and layups), creating tension or synergy between a player's habits and a team’s system.
- Efficiency tradeoffs: A player shooting from their comfort zone may be more efficient even if it deviates from team norms — but alignment can lead to more seamless offensive execution.
- Clustering analysis helps us identify groups of players based on shot profiles (e.g., paint-heavy slashers vs. perimeter snipers), and assess which groups yield higher FG% in specific eras.
- Era-based trends reveal how the value of shot locations has shifted — e.g., a midrange-heavy style in the 2000s vs. modern 3PT-heavy spacing.

02

Data and Methods

Data Sources

HoopR

Package data from the NBA Stats API
Past 30 years (1996-2024)

Basketball Reference

1996 scoring efficiency (team-level) not available in hoopR

Data frame was manually created for this year

Used in scoring efficiency regression

Data Manipulation

Preprocessing

- Calculate Standardized Field Goal Attempt Frequencies
- Merge team and player data
- Player bias scores were calculated by comparing individual shot frequencies to their team's average shot frequencies

Clustering

- Player Shot Location (Restricted_Area_FGA, Corner_3_FGA)
- Team Shot Location (Similar to player, but for teams)
- Created for 1996-2024

Regression

- ESPN Team Stats (hoopR) - Scoring Efficiency
- Clustering data was aggregated to determine count of player types per team (ex: num_paint_players)
- Scoring efficiency regressed count of players based on cluster

Key Variables

Shot Frequency Variables

Restricted_Area_Freq_Player = Restricted_Area_FGA / Total_FGA

Paint_Freq_Player = In_The_Paint_Non_RA_FGA / Total_FGA

Mid_Range_Freq_Player = Mid_Range_FGA / Total_FGA

Corner_3_Freq_Player = (Left_Corner_3_FGA + Right_Corner_3_FGA) / Total_FGA

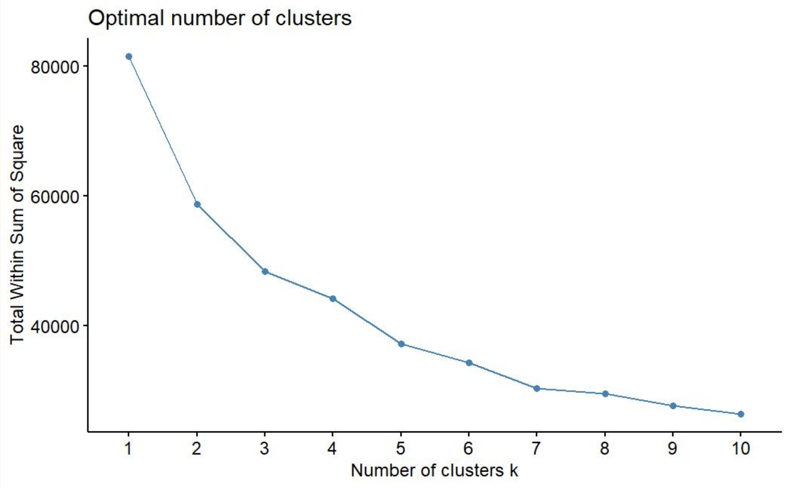
Above_Break_3_Freq_Player = Above_the_Break_3_FGA / Total_FGA

Backcourt_Freq_Player = Backcourt_FGA / Total_FGA

Methodology

Clustering Players by Shot Selection Preferences

- Players grouped into clusters using K-Means based on their shot frequency profiles
- Used elbows to determine the optimal number of clusters
- Clusters represented distinct player types



Methodology

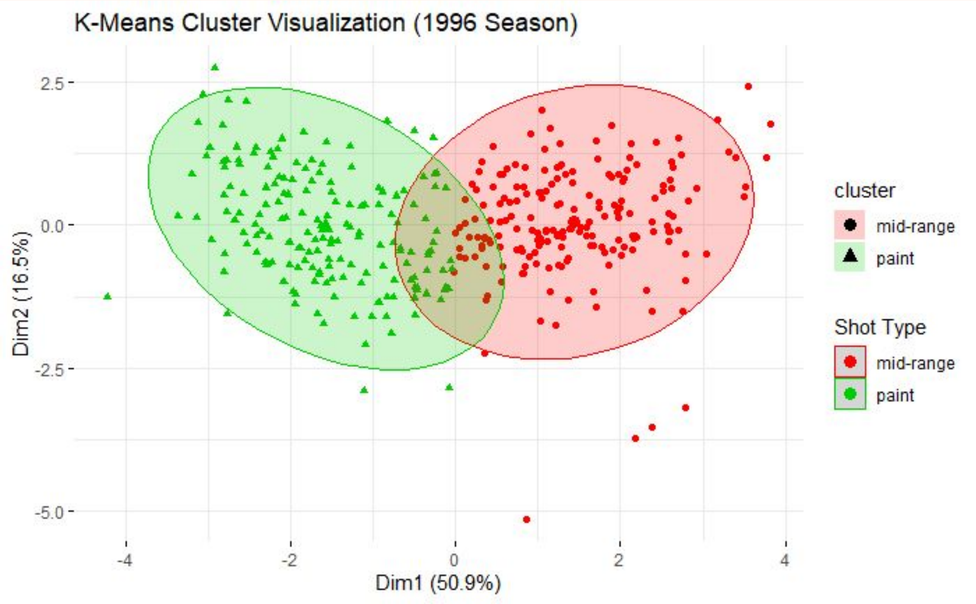
Multi-Season Analysis

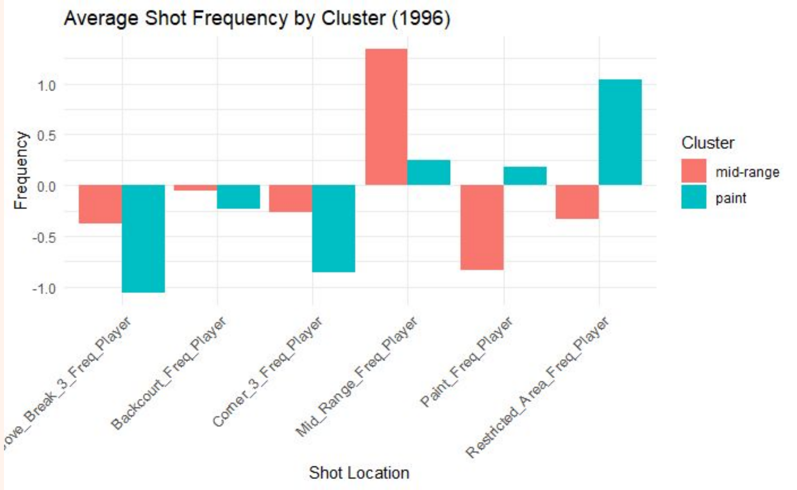
- Comparisons across different eras in the NBA
- Statistical significance and directionality of coefficients were interpreted from regressions

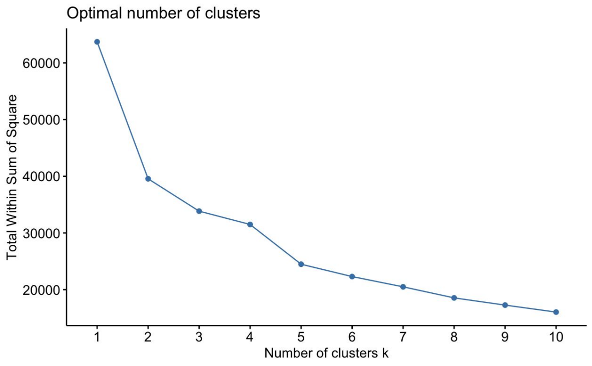
03

Results

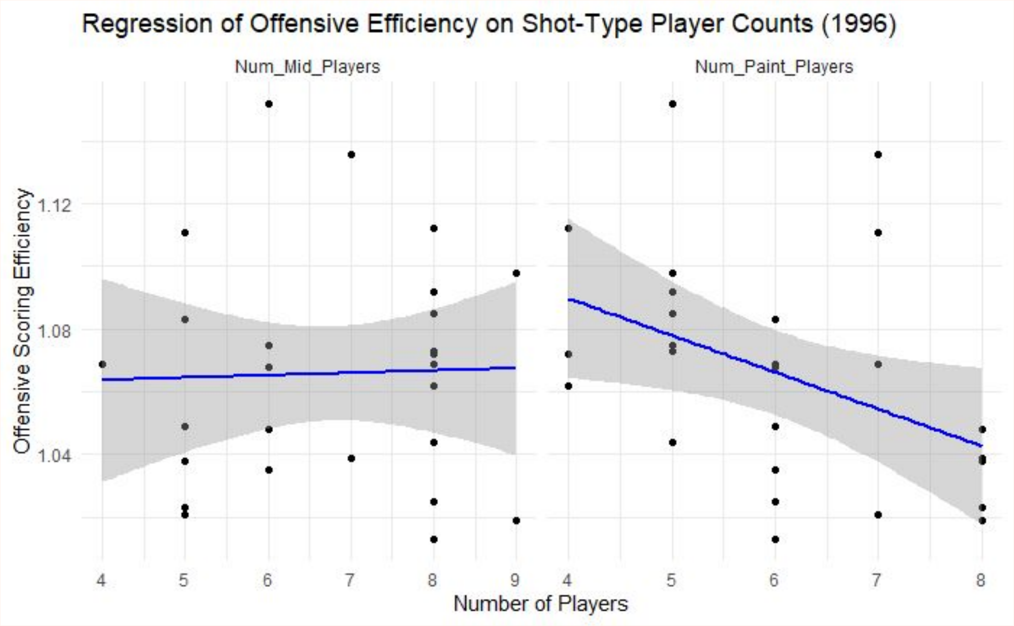
1996


K-Means Cluster Visualization (1996 Season)
Dim1 (50.9%)
Dim2 (16.5%)
2.5
0.0
-2.5
-5.0
-4
-2
0
2
4
cluster
mid-range
paint
Shot Type
mid-range
paint


Average Shot Frequency by Cluster (1996)
Frequency
1.0
0.5
0.0
-0.5
-1.0
ove_Break_3_Freq_Player
Backcourt_Freq_Player
Corner_3_Freq_Player
Mid_Range_Freq_Player
Paint_Freq_Player
Restricted_Area_Freq_Player
Shot Location
Cluster
mid-range
paint

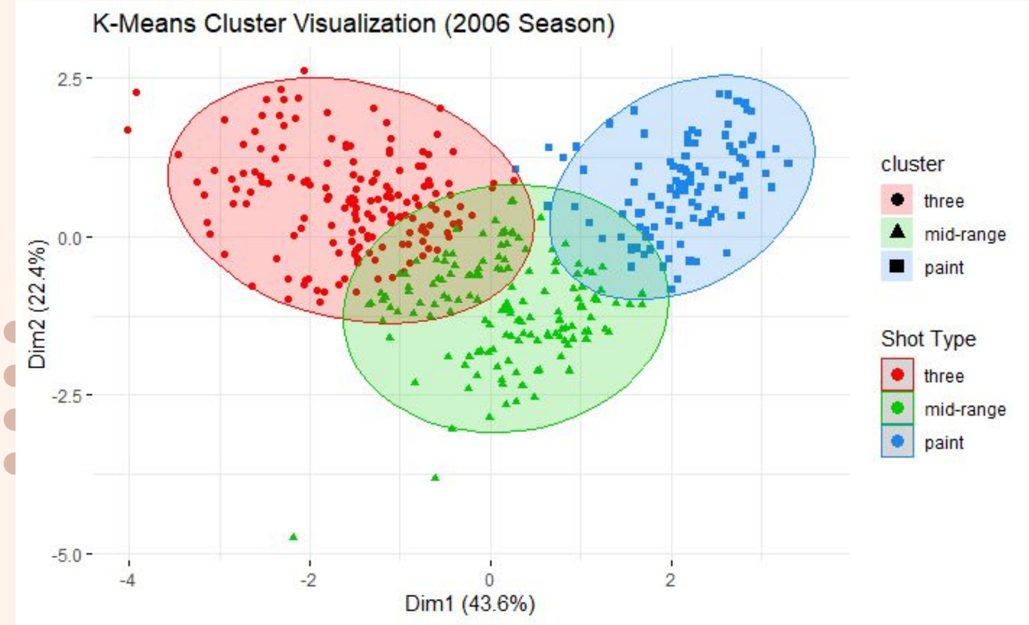

Optimal number of clusters
Total Within Sum of Square
60000
50000
40000
30000
20000
1
2
3
4
5
6
7
8
9
10
Number of clusters k

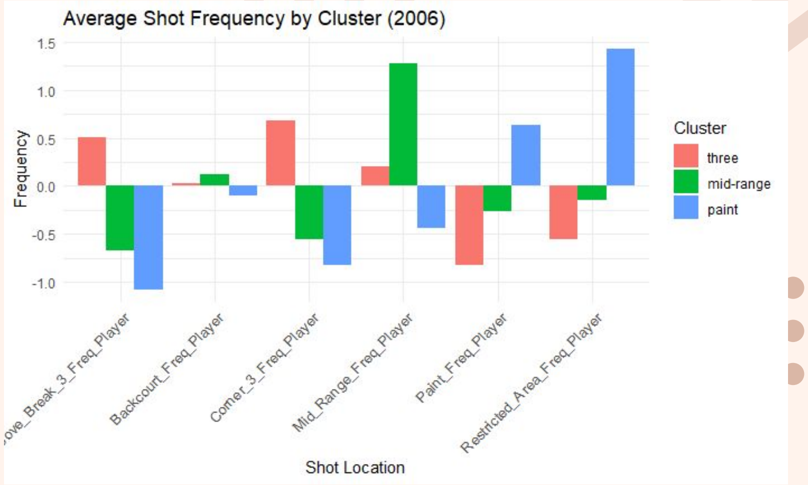
Scoring Efficiency - Regression (1996)

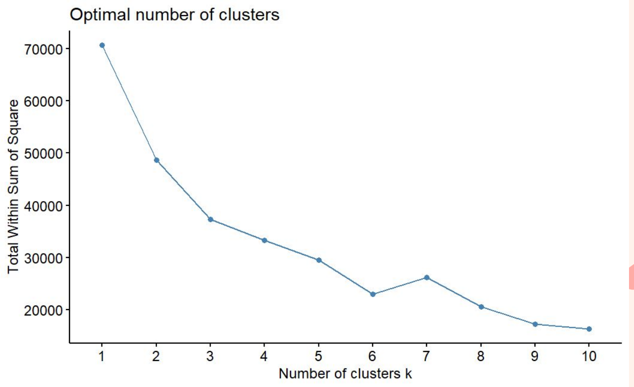



```
Residuals:
      Min        1Q    Median        3Q       Max
-0.046796 -0.026016 -0.001939  0.010251  0.085467

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        1.190144   0.060322  19.730  6.5e-16 ***
Num_Paint_Players -0.014604   0.005808  -2.515   0.0194 *
Num_Mid_Players   -0.005341   0.005151  -1.037   0.3106
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.0336 on 23 degrees of freedom
  (3 observations deleted due to missingness)
Multiple R-squared:  0.2164,    Adjusted R-squared:  0.1482
F-statistic: 3.175 on 2 and 23 DF,  p-value: 0.06059
```

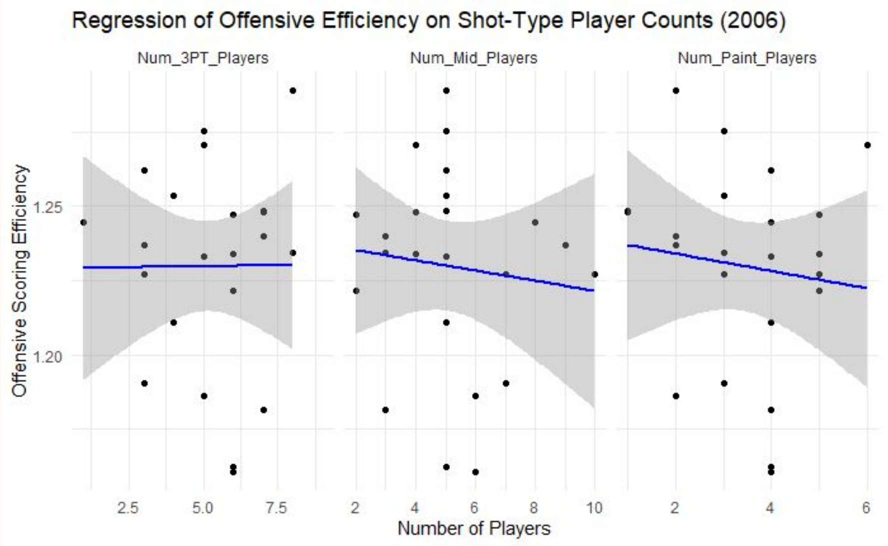
2006


K-Means Cluster Visualization (2006 Season)
2.5
0.0
-2.5
-5.0
Dim2 (22.4%)
-4
-2
0
2
Dim1 (43.6%)
cluster
three
mid-range
paint
Shot Type
three
mid-range
paint


Average Shot Frequency by Cluster (2006)
1.5
1.0
0.5
0.0
-0.5
-1.0
Frequency
ove_Break_3_Freq_Player
Backcourt_Freq_Player
Corner_3_Freq_Player
Mid_Range_Freq_Player
Paint_Freq_Player
Restricted_Area_Freq_Player
Shot Location
Cluster
three
mid-range
paint

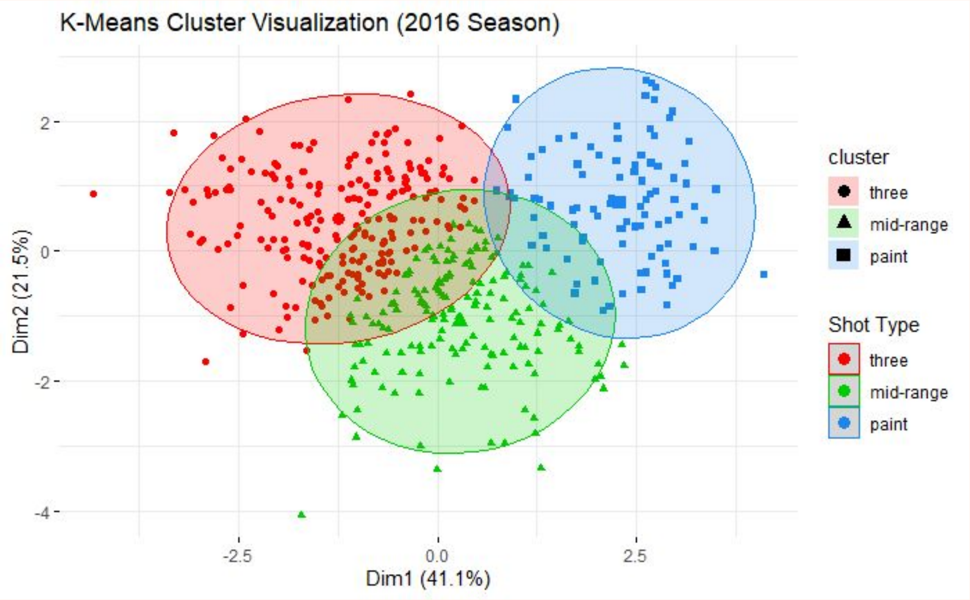

Optimal number of clusters
70000
60000
50000
40000
30000
20000
Total Within Sum of Square
1
2
3
4
5
6
7
8
9
10
Number of clusters k

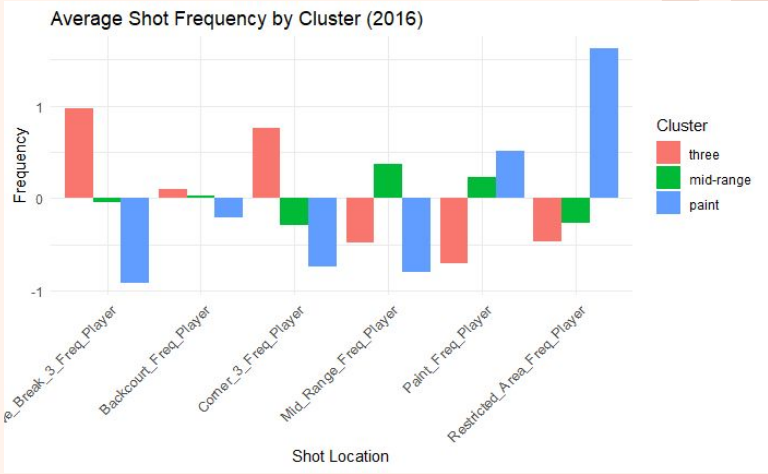
Scoring Efficiency Regression (2006)

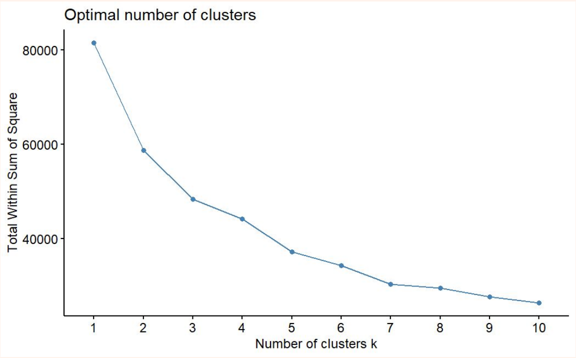



```
Residuals:
     Min       1Q   Median       3Q      Max
-0.05770 -0.01838  0.00840  0.01325  0.06737

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        1.331344   0.084036  15.843  2.1e-12 ***
Num_Paint_Players -0.007806   0.007100  -1.099    0.285
Num_Mid_Players   -0.007466   0.006467  -1.155    0.263
Num_3PT_Players   -0.007108   0.007148  -0.994    0.333
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03533 on 19 degrees of freedom
  (7 observations deleted due to missingness)
Multiple R-squared:  0.07821,	Adjusted R-squared:  -0.06734
F-statistic: 0.5373 on 3 and 19 DF,  p-value: 0.6624
```

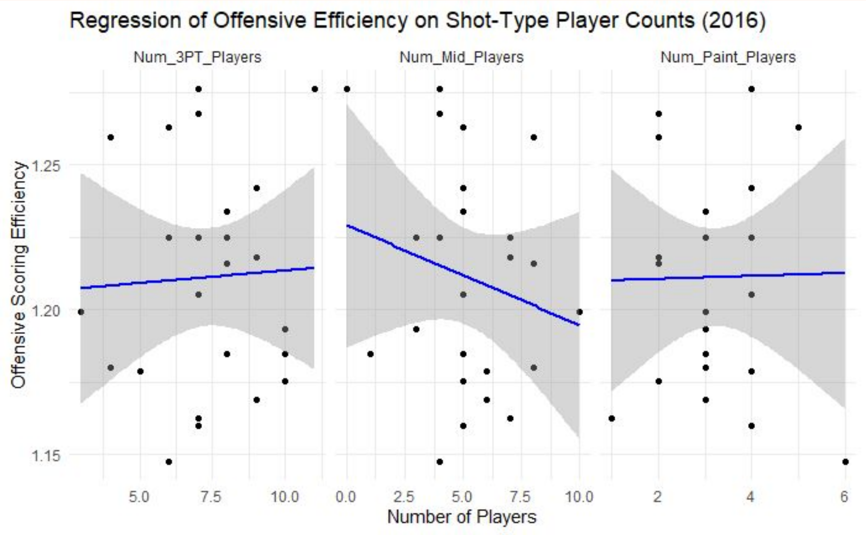
2016


K-Means Cluster Visualization (2016 Season)
Dim2 (21.5%)
Dim1 (41.1%)
2
0
-2
-4
-2.5
0.0
2.5
cluster
three
mid-range
paint
Shot Type
three
mid-range
paint


Average Shot Frequency by Cluster (2016)
Frequency
1
0
-1
e_Break_3_Freq_Player
Backcourt_Freq_Player
Corner_3_Freq_Player
Mid_Range_Freq_Player
Paint_Freq_Player
Restricted_Area_Freq_Player
Shot Location
Cluster
three
mid-range
paint


Optimal number of clusters
Total Within Sum of Square
80000
60000
40000
1
2
3
4
5
6
7
8
9
10
Number of clusters k

Scoring Efficiency Regression (2016)




```
Residuals:
      Min        1Q    Median        3Q       Max
-0.063994 -0.029137 -0.000105  0.028429  0.058647

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        1.307544   0.094956  13.770 1.15e-11 ***
Num_Paint_Players -0.005664   0.009230  -0.614    0.546
Num_Mid_Players   -0.007606   0.006138  -1.239    0.230
Num_3PT_Players   -0.005274   0.006468  -0.815    0.424
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.04016 on 20 degrees of freedom
  (6 observations deleted due to missingness)
Multiple R-squared:  0.07367,	Adjusted R-squared:  -0.06528
F-statistic: 0.5302 on 3 and 20 DF,  p-value: 0.6667
```

04

Interpretation + Insights

Interpretation

1996

- Paint + Mid-Range Players dominated
- Inverse relationship between Paint players and Offensive Scoring Efficiency
  - P-Value: 0.0194 (statistically significant)

2006

- Mid-Range players on the rise
- Increase in paint shots outside of restricted area
- No clear relationship with offensive scoring efficiency

2016

- Dominated by 3PT and Paint Players
- Not statistically significant, but downward trend for mid-range players
- Could possibly go back to two clusters in the future

Insights

While a clear, direct relationship between the number of players in each cluster and offensive efficiency was not found there are still valuable insights from the analysis:

- Team’s cannot have gaps in shooting
- 3 point shots/shooters are increasingly valuable
- Teams with inside players need shooters to bridge the gap

Estimated Best Team Composition

Estimating Best Team By Minimizing Loss Function Of Shooter And Paint Player Proportions:

2023-24: Est. Ideal Team Composition is: 33% Shooters, 22% Inside Players, 44% Versatile

Based by minimizing: $L(P_{shooters}, P_{inside}) = \Sigma\, E_i(P_i - P_{shooter})^2 + E_i(P_{ii} - P_{inside})^2$, Where E_i is the team's offensive efficiency and P is the proportion of shooters/inside players.

- Displays How Teams Need Outside Shooting Ability, Even With Strong Inside Players
- Close To Average Proportions, But Slightly More Shooters And Versatile Players

The End