Clustering NBA Players Offensively

Nehal Chigurupati

November/December 2023

1 Objective

The objective of this project is to cluster together NBA players based on their offensive game, as determined by their performance on the following types of plays:

- Isolation
- Cut
- Off screen
- Pick and roll ball handler
- Pick and roll roller
- Post up
- Spot up

The clusters could be used by a team in a number of ways. If a front office is looking for a hidden gem in free agency to replace a player, they could look at everyone within the cluster associated with the departing player. Strategy-oriented coaches could use the model to craft lineups and plays, and development coaches could use clusters to find appropriate film in the off-season.

2 Methodology

For each play type, the following statistics over the 2023-24 season were collected on a per player basis (averaging over games):

- number of possessions
- frequency
- average points per possession
- total points

- field goals made

- field goals attempted

- field goal percent

- effective field goal percent

- free throw frequency

- turnover frequency

- shooting foul frequency

- and one frequency

- score frequency

For each play type, the corresponding player data was scaled and normalized. Three different types of models were tested: KMeans, spectral clustering with affinity matrix calculated via RBF, and spectral clustering with affinity matrix calculated via nearest neighbors. For each model type, seven different models were trained, one for each play type.

For KMeans, the number of clusters for each play type model was computed via the maximum of silhouette scores over a range from 2 to 20 clusters. The same technique was used for spectral clustering.

Next, all the models were trained. The average of all Davies-Bouldin scores for the KMeans models was computed, and compared to the same averages for the Spectral RBF and nearest neighbor models. The average score was lowest on the KMeans models, leading to it being chosen.

3 Results

The KMeans models were trained on the following numbers of clusters:

Iso: 3 clusters
Cut: 3 clusters
Off screen: 3 clusters
PnR Handler: 3 clusters
PnR Roller: 2 clusters
Post up: 3 clusters
Spot up: 2 clusters

The average Davies-Bouldin score over all KMeans models was 1.39, indicating a moderately strong clustering. To visualize the clusters (a difficult task, because the feature space is greater than three dimensions), two different techniques were used.

First, PCA on two components was run on the data to project into 2-space, then displayed on a scatterplot with points colored in accordance with their label. The clusters were well visible. **Second**, strip plots were produced to show how the clusters vary over a specific stat.

3.1 Isolation

3.1.1 Defining each cluster

There are three isolation clusters. The seven statistics in which the clustering was most clearly found were:

- number of isolation possessions per game (POSS)

- frequency of isolation plays every game (FREQ%)

- points per isolation possession (PPP)

- average PPG from isolation shots (PTS)

- FGM in isolation per game (FGM)

- FGA in isolation per game (FGA)

- FG% in isolation (FG%)

- EFG% in isolation (EFG%)

The range of values in these statistics for each isolation cluster are listed below.

Cluster	POSS	FREQ%	PPP	PTS	FGM	FGA	FG%	EFG%
0	0-2.5	2.5-12.5	0-.85	0-1.5	0-.6	0-2	0-42	0-42
1	2.5-7	11-22.5	.75-1.25	1.5-7.5	.55-2.4	1.9-5.2	35-60	35-58
2	0-2.5	2.5-12.5	.75-1.75	.75-3	.2-1.2	0-2	39-85	38-105

Table 1: Isolation Cluster Values

Cluster 0 players attempt isolation plays relatively infrequently, and aren't very productive in such plays. Cluster 1 players attempt a lot of isolation plays, and are moderately efficient at generating points off such possessions. Cluster 2 players don't attempt many isolation plays, but are incredibly talented scorers one-on-one.

A quick gut check with the lists of players belonging to each cluster confirms the above analysis. Cluster 2 includes skilled, tall shooting bigs like Nikola Jokic and Chet Holmgren, who can score reliably over any defender. Cluster 1 includes power and small forwards in the 6'6-6'9 range like Jayson Tatum and LeBron, who back down defenders in isolation frequently and can score moderately efficiently on such plays. Finally, cluster 0 includes shifty point guards who can generate space in isolation, but typically take tough shots that result in a basket less frequently than the post moves of players like Jokic.

Cluster 0 players include:

- Cade Cunningham

- Darius Garland

- Dennis Schroder

- Chris Paul

- Collin Sexton

- Jamal Murray

- Josh Giddey
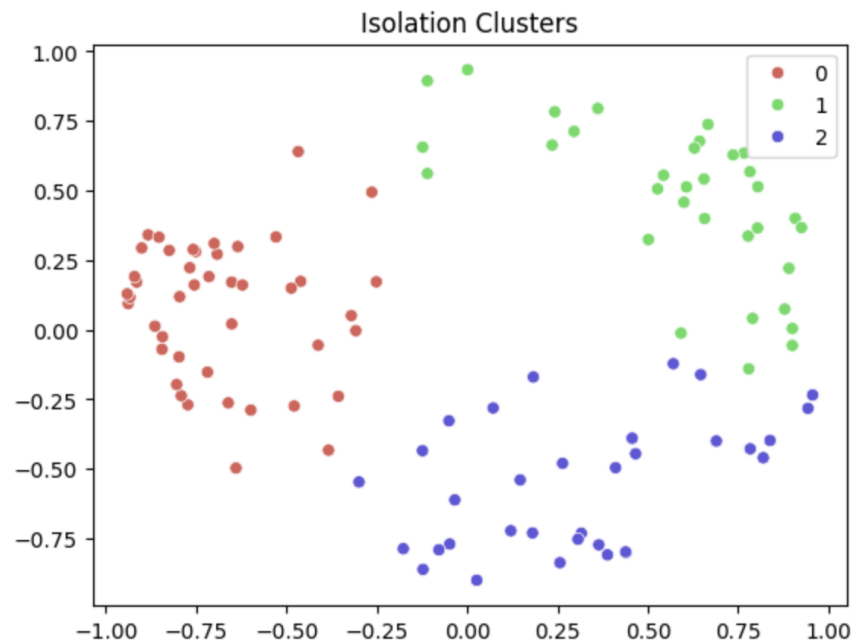
Cluster 1 players include:

- Luka Doncic

- LeBron James

- Jimmy Butler

- Anthony Edwards

- Kevin Durant

- Brandon Ingram

4

- DeMar DeRozan
- Jayson Tatum

Cluster 2 players include:

- Jaren Jackson Jr.
- Nikola Jokic
- Bam Adebayo
- Anthony Davis
- Victor Wembanyama
- Chet Holmgren
- Evan Mobley

3.1.2 Clusters under PCA

The axes are the two PCA components. It seems like the clustering algorithm is working moderately well for isolation plays.

3.1.3 Variance of clusters with average number of isolation possessions per game
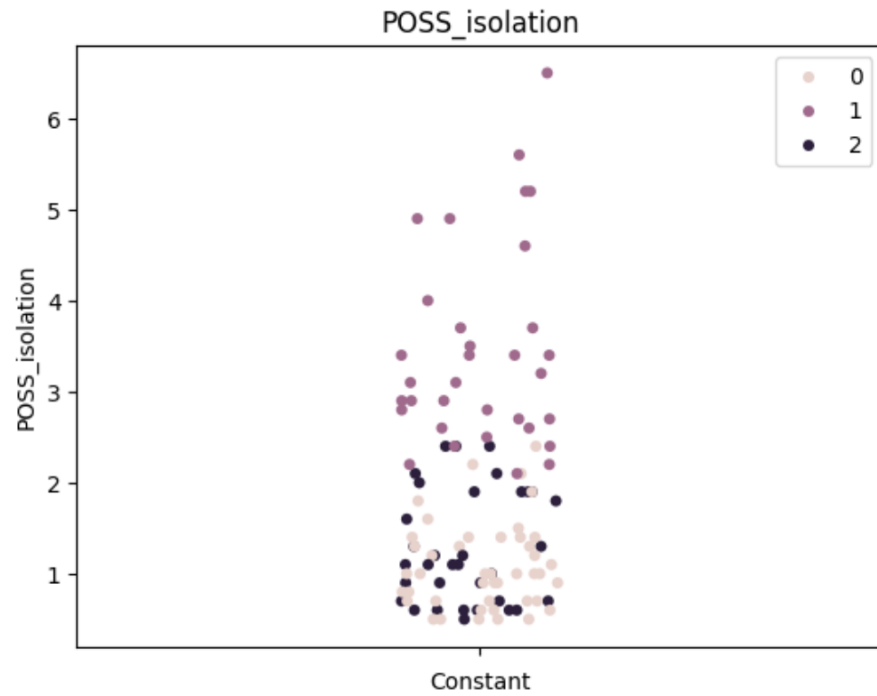

The strip plot doesn't show a very strong clustering (meaning that the model is clustering more heavily over other variables). Cluster 0 and 2 players average between 0 and 2.5 isolation possessions per game, while cluster 1 players average between 2.5 and 7 isolation possessions per game.

3.1.4 Variance of clusters with the average frequency of isolation plays per game
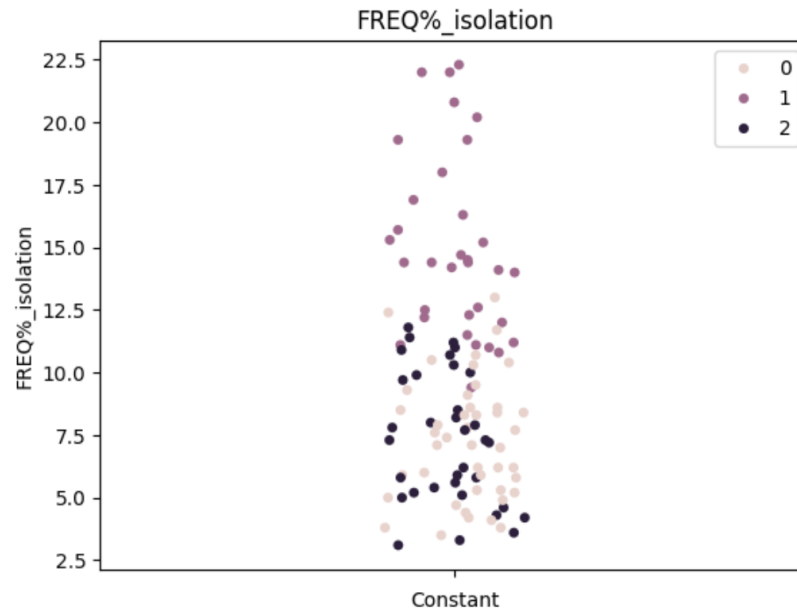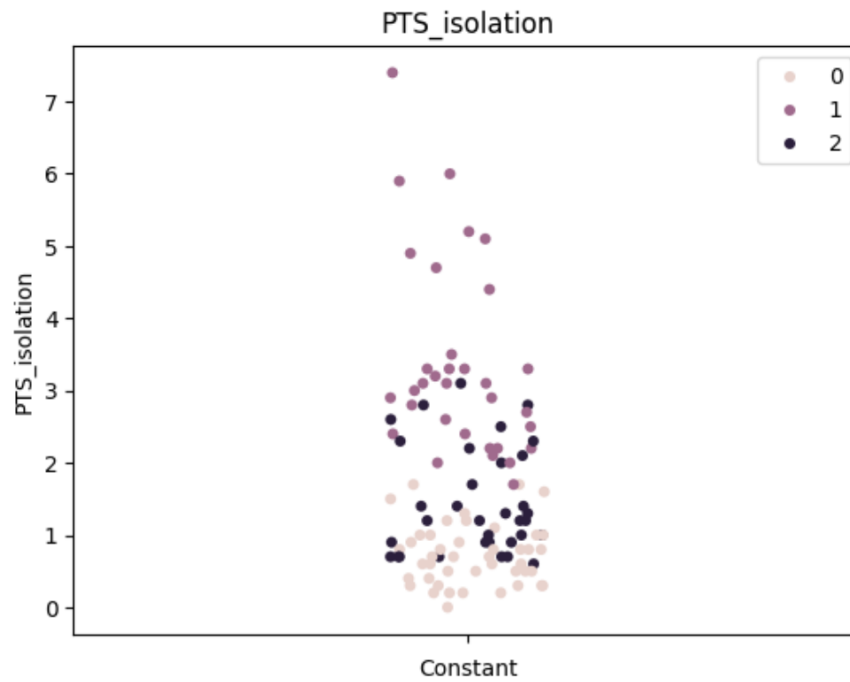

FREQ%_isolation

There isn't great clustering here either, but the trend is clear, and similar to that of the last statistic: cluster 1 players attempt a lot of isolation plays (in this case, relative to other types of plays), while a smaller proportion of cluster 0 and 2 players' offensive games come from isolation.

### 3.1.5 Variance of clusters with points per possession on isolation plays
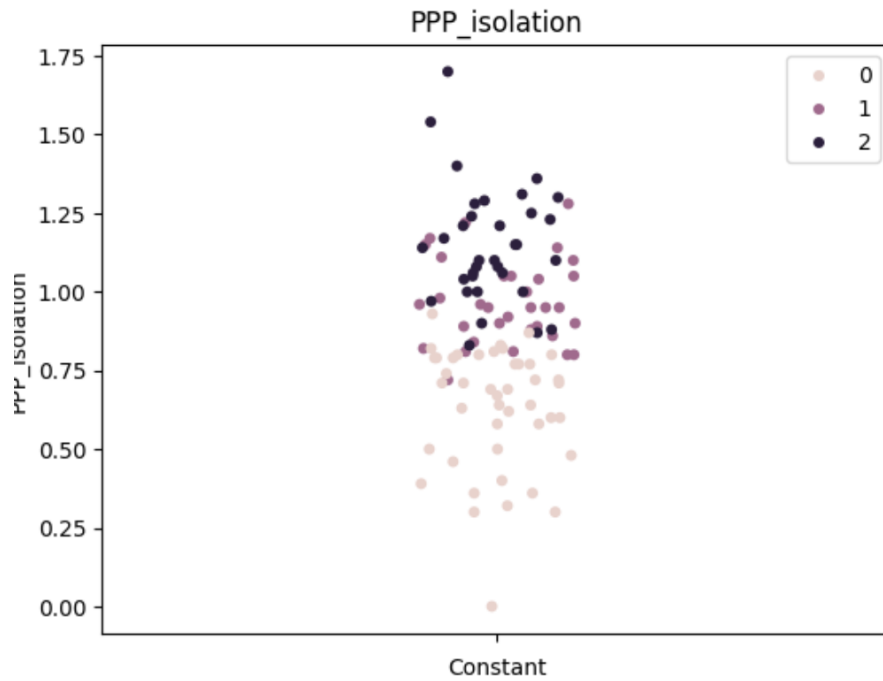
Cluster 0 players are the least productive on isolation plays, scoring between .25 and .75 points per possession. Cluster 1 players are middle-of-the-road, scoring between .75 and 1.25 points per possession. Cluster 2 players, on the other hand, are excellent isolation players who can reliably generate points.

3.1.6 Variance of clusters with average PPG from isolation attempts
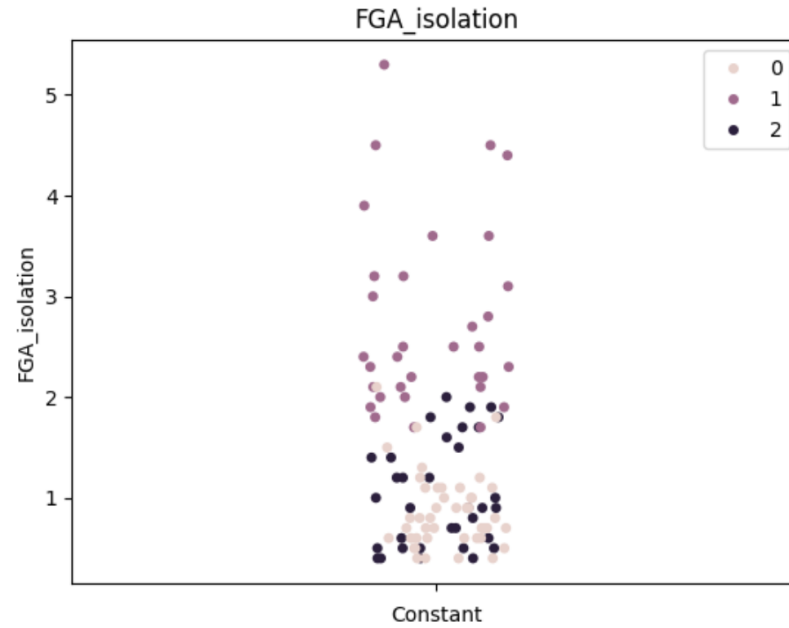

PTS_isolation

Cluster 0 players generate, on average, between 0 and 1.5 points per game from isolation attempts. Cluster 1 players generate between 2 and 7 points per game from isolation. Finally, cluster 2 players generate between .7 and 3 points per game from isolation.

### 3.1.7 Variance of clusters with field goals made in isolation
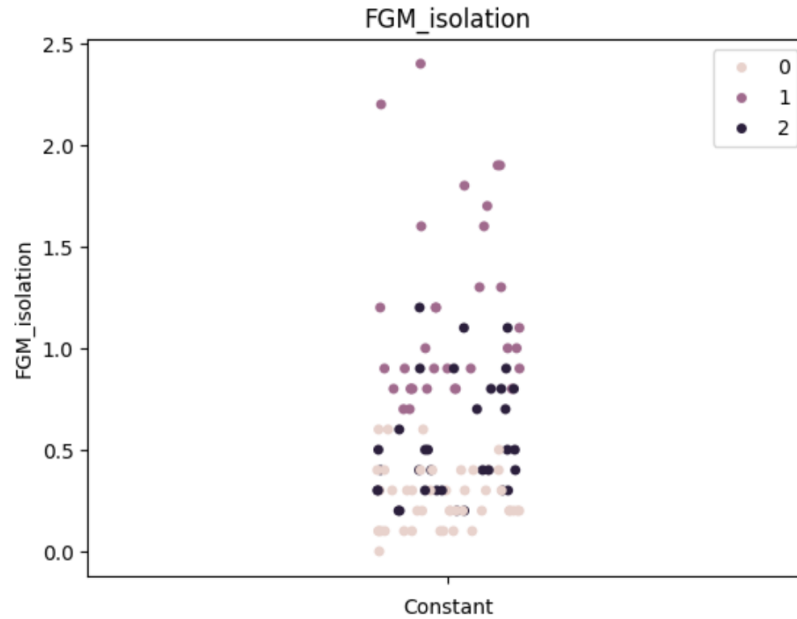
Cluster 0 players average between 0 and .6 field goals made per game in isolation. Cluster 1 players average between .6 and 2.4 isolation field goals per game. Finally, cluster 2 players average between .25 and 1.25 field goals made in isolation per game.

3.1.8 Variance of clusters with field goal attempts in isolation
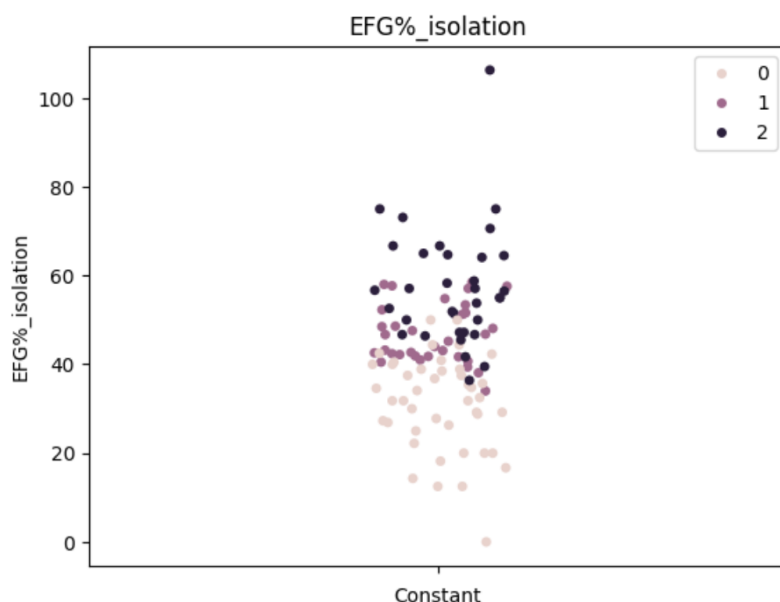

FGA_isolation

Cluster 0 players attempt between 0 and 2 isolation shot attempts per game. Cluster 2 players attempt a similar range of shot attempts per game. Cluster 1 players attempt a lot of isolation shots per game, ranging from 2 to 5.5. scorers, yet do not attempt as many isolation shots as cluster 1 players.

3.1.9 Variance of clusters with field goal percentage in isolation
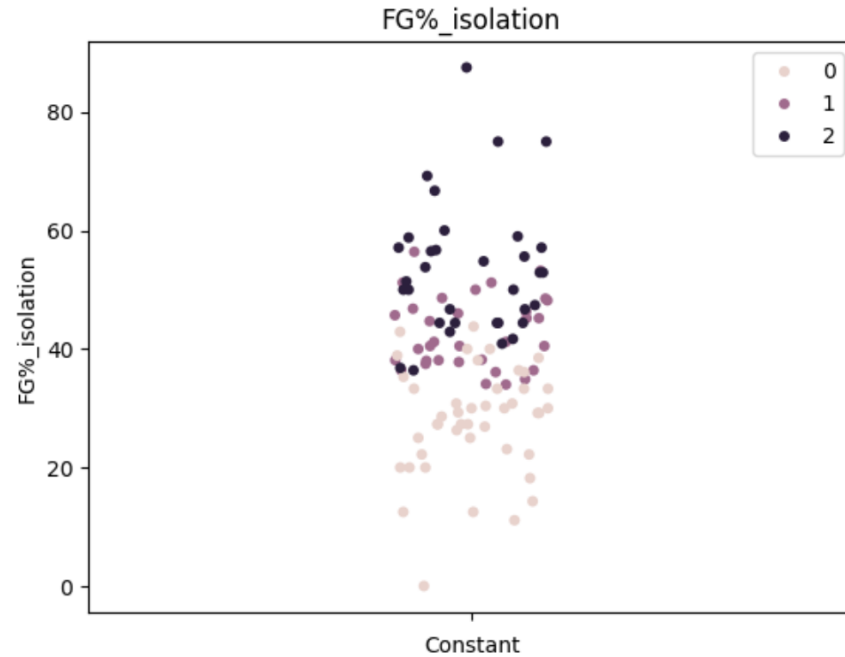
FG%_isolation

Cluster 0 players make between 0 and 41% of their isolation shot attempts. Cluster 1 players make between 38 and 60% of their isolation shot attempts. Cluster 2 players shoot at an extremely efficient rate, making between 40 and 80 percent of their isolation field goal attempts.

3.1.10 Variance of clusters with effective field goal percentage in isolation

The ranges here echo those of the field goal percentage statistic. Cluster 0 players don't shoot efficiently in isolation, at effective field goal percentages between 0 and 43%. Cluster 1 players score at an effective field goal percentage between 39 and 60%. Finally, cluster 2 players are extremely efficient, with an effective field goal percentage between 39 and 100%.

3.2 Cuts

3.2.1 Players in each cluster

Cluster 0 included, among others, the following players:

- Giannis Antetokounmpo

- Obi Toppin

- Jonas Valanciunas

- Goga Bitadze

- Kevin Durant

- Kristaps Porzingis

- De'Andre Hunter

- Keldon Johnson

Cluster 1 included:

- Jusuf Nurkic
- Kevon Looney
- Brook Lopez
- Clint Capela
- Nikola Vucevic
- Marvin Bagley III
- Bam Adebayo
- Alperen Sengun

Cluster 2 included:

- Evan Mobley
- Jarrett Allen
- Domantas Sabonis
- Ivica Zubac
- Anthony Davis
- Nikola Jokic
- Rudy Gobert
- Walker Kessler
- Jalen Duren

At first glance, the offensive identities of these clusters are not clear, so further analysis of the strip plots is required.

3.2.2 Clusters under PCA
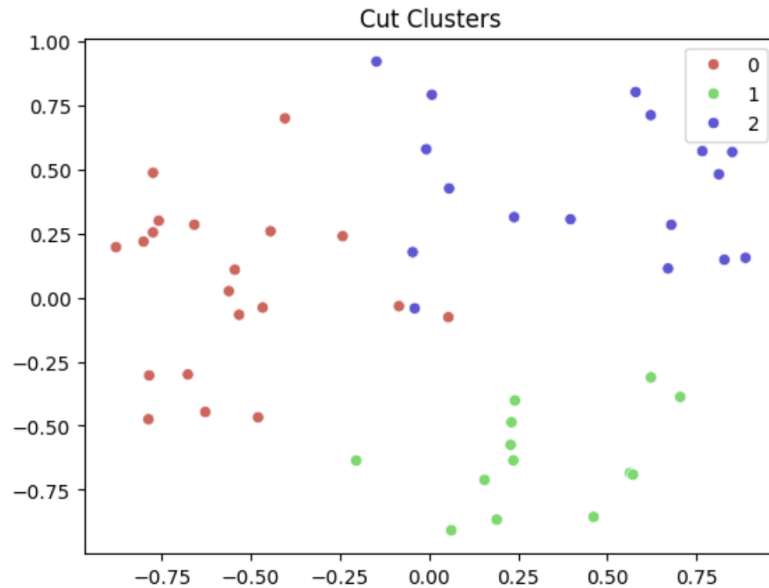
The clustering doesn't appear to be working well on cuts. Let's see how the clusters vary over individual statistics.

3.3 Off screen

3.4 Pick and roll ball handler

3.5 Pick and roll roller

3.6 Post up

3.7 Spot up

4 Further work

Right now, I'm working on a few different extensions of this project:

- Graphics to holistically visualize a player's strengths/weaknesses as determined by their cluster

- A similar analysis on 2022-23 data to determine areas of improvement

- A similar analysis, but for player defense against the same types of plays