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 Players in each cluster

Cluster 0 seems to include shifty guards. A few players in this cluster are:

- Cade Cunningham
- Darius Garland
- Dennis Schroder
- Chris Paul
- Collin Sexton
- Jamal Murray
- Josh Giddey

Cluster 1 includes a lot of forwards and guards who can both drive and shoot moderately efficiently:

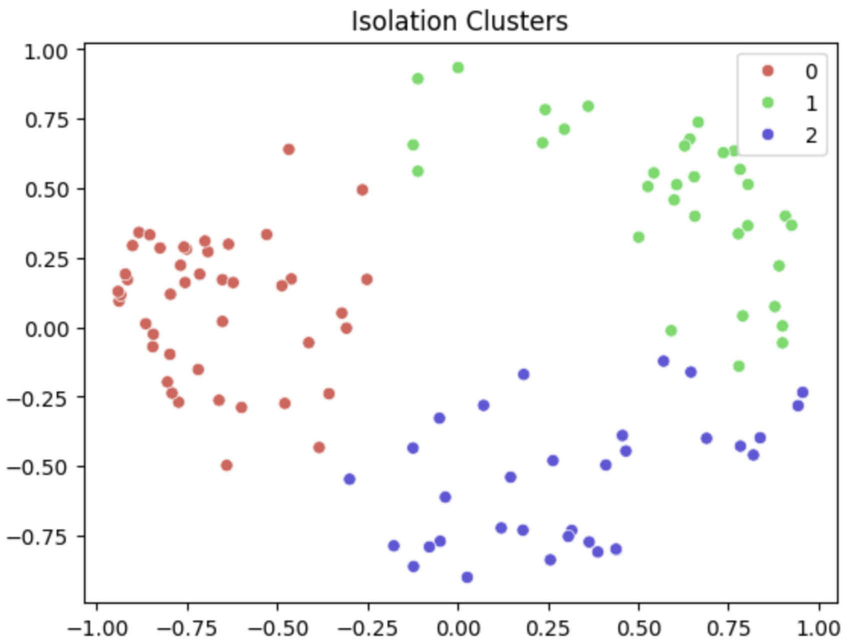
- Luka Doncic
- LeBron James

- Jimmy Butler
- Anthony Edwards
- Kevin Durant
- Brandon Ingram
- DeMar DeRozan
- Jayson Tatum

Cluster 2 looks like it includes bigs, as well as guards who are extraordinarily efficient at shooting in iso. The cluster includes:

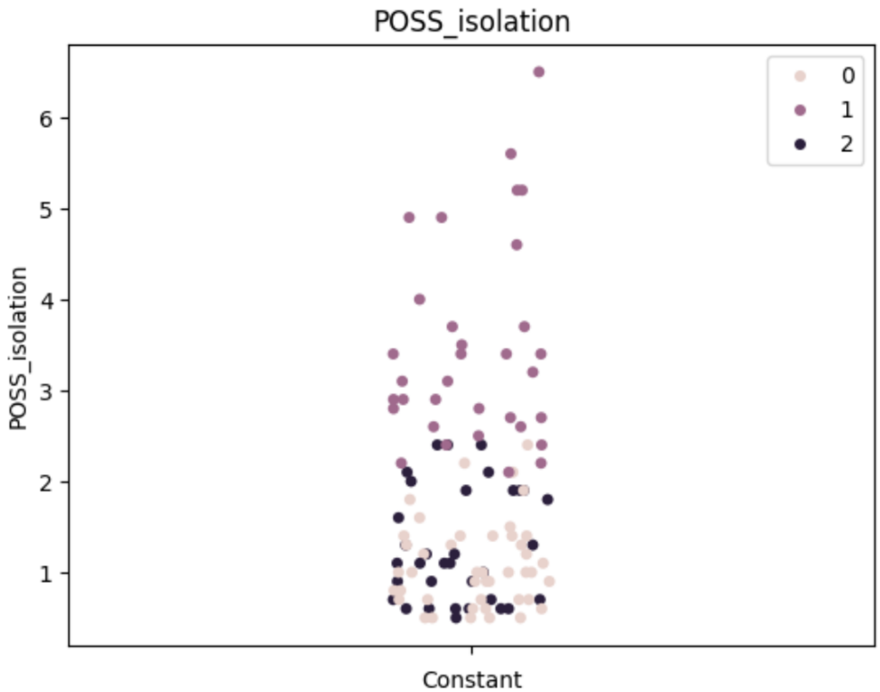
- Jaren Jackson Jr.
- Nikola Jokic
- Bam Adebayo
- Anthony Davis
- Victor Wembanyama
- Chet Holmgren
- Evan Mobley
- Steph Curry
- Tyrese Haliburton

3.1.2 Clusters under PCA



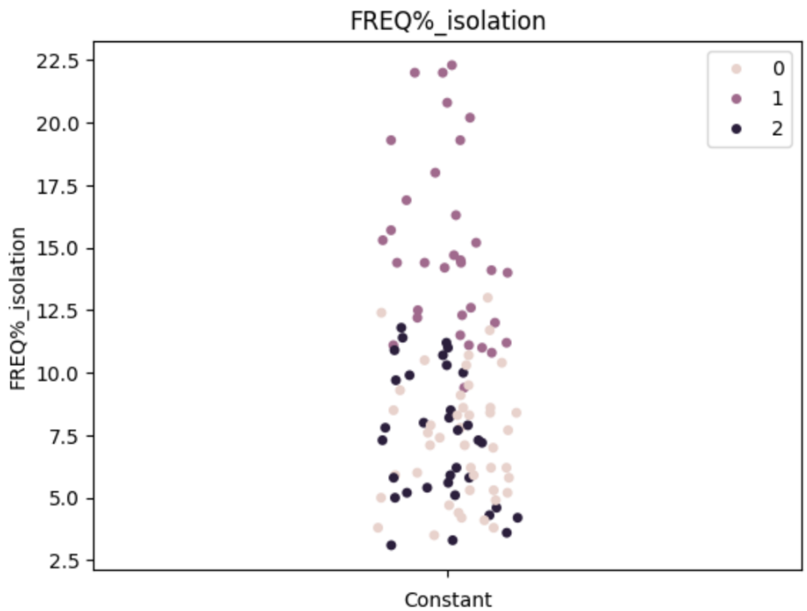
The axes are the two PCA components. It seems like the clustering algorithm is working moderately well for isolation plays.

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



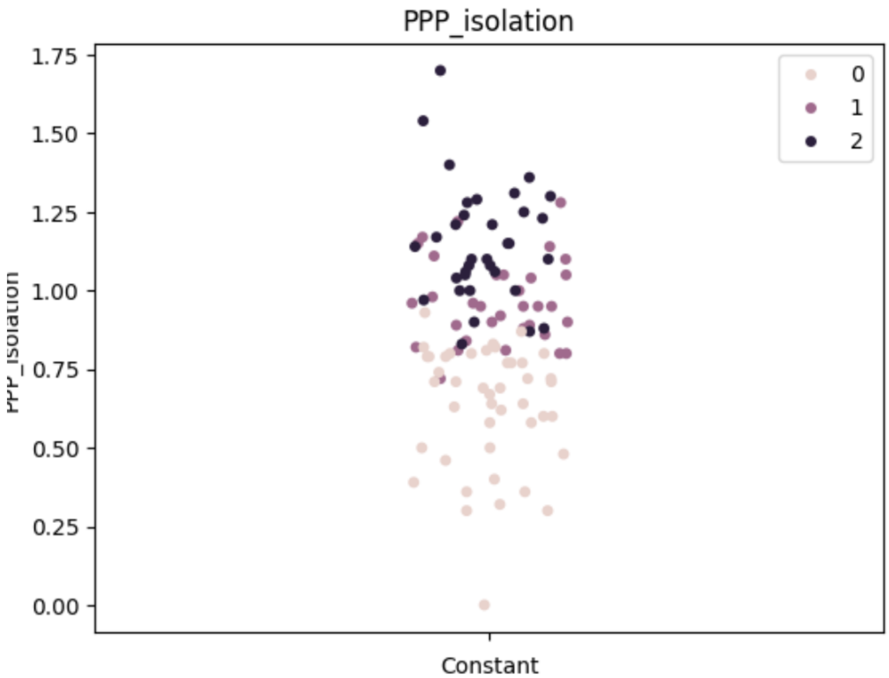
The strip plot doesn't show a very strong clustering (meaning that the model is clustering more heavily over other variables). Cluster 0 and 2 players average between 0 and 2.5 isolation possessions per game, while cluster 1 players average between 2.5 and 7 isolation possessions per game.

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



There isn't great clustering here either, but the trend is clearly, 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.

We can start to generate a broad picture of each cluster. 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 shooting bigs like Nikola Jokic and Chet Holmgren, who can shoot over any defender and score very efficiently. Cluster 1 includes players like Jayson Tatum and LeBron, who both back down defenders frequently. 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.

3.2 Cuts

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

Next, I want to create some sort of holistic visualization that takes players and shows the proficiencies/deficiencies associated with their clusters.