Elijah Wooten

April 30th, 2022

Identifying Guard Archetypes

Abstract

In the game of basketball, players have different playstyles that can indicate how well they synergize with each other. Using hierarchical clustering, we can identify the playstyle of guards to better understand the type of role they would play in a lineup. We were able to assess player archetypes based off their shooting, passing, and defensive statistics. Using hierarchical clustering, we were able to identify guards of having the archetype of off-ball scorer, slashing scorer, defensive floor general, combo scoring guard, and 2-way shooter. Upon further research we can apply this clustering method to other facets and positions in basketball.

Introduction

In basketball there are at most 15 players on a team, and only 5 players from each team can be on the court at any given time. The 5 players that are usually on the court are a point guard, shooting guard, small forward, power forward, and center. The point guard is the one who primarily dribbles the ball and sets up their team to score, the shooting guard is primarily tasked with dribbling and scoring, the small forward is used in multiple facets to both score and defend, the power forward is used to defend and rebound, and the center is often used for defense and rebounding. It is important to have players that can fill these roles, but in the past decade the game has evolved into players excelling in different skillsets. A player like center Nikola Jokic is often used as the primary playmaker and scorer on his team, and he won an NBA MVP despite not being the most defensively skilled center in the league. Or there is also a player like Stephen

Curry, who is one of the best players in the league because of his level of shooting that has never been seen before.

With the rise of players with unique skillsets, it is becoming more outdated to base a player's ability solely based off what their traditional position calls for. Coaches can experiment with different players and lineups to find out who synergizes the best with each other, not necessarily by which position they are listed as. This is where we will try and classify players based off their statistics like field goal percentage, 3-point percentage, free throw rate, etc. to identify what type of player they are and who they would play well with.

Methods

The data we will be using is private college basketball data, so this specific data analysis will not be reproducible. For the classification process, we used hierarchical clustering because it allows us to specify the number of clusters we would like. We chose 5 because it appears to be enough clusters to offer variety in different playstyles, but also not too many as to make certain playstyles feel redundant. We used hierarchical clustering in Python using the AgglomerativeClustering function that is part of the scikit-learn package. The statistics that we used to cluster the college basketball players are 3-point shooting percentage (percentage of jump shots made within the 3-point line), midrange shooting percentage (percentage of jump shots made within the 3-point line), player usage rate (percentage of team plays used by a player while he was on the floor), assist percentage (percentage of teammate field goals a player assisted while he was on the floor), at the rim percentage (percentage of shots made underneath the basket), free throw rate (how often a player scores from free throws), assisted percentage (what percent of a player's made shots are assisted), and number of total steals. These statistics were also all normalized so the clustering process would accurately assess the different scales and variances of the variables. Only guards were used in this clustering process for easier classification with relevant metrics for every player. To get a more accurate representation of how a player's shooting percentages impact their playstyles, we filtered out players using the criteria of at least 25 3-point attempts, 25 mid-range attempts, 10 assists, 25 at the rim attempts, 50 field goal attempts and at least 100 minutes played. This helps to get rid of players with limited playing time that would potentially skew the data. 25 was also used as the minimum requirement for the number of three point and mid-range attempts because it allows us to assume normality for those statistics' distributions. There were originally 7,509 observations in the dataset, but when specifying for only guards and the minimum criteria that number was reduced to 1601.

Results

After using hierarchical clustering on the player data, we found 5 clusters to be significantly identifiable on the dendrogram map in Figure 1. Figure 2 shows the averages of each group of clustered players, and from looking at the averages we could tell that they were classified often by a couple defining factors. The first cluster had an average 3-point field goal percentage (3FG%) of 34.76%, mid-range percentage (MR%) of 36.12%, usage rate (UR) of 20.80%, assist percentage (AST%) of 12.61%, at the rim percentage (ATR%) of 56.43%, free throw rate (FTR) of 19.26%, assisted percentage (ASTD%) of 51.02%, and total number of steals (STL) at 11.86. The second cluster had an average 3FG% of 31.69%, MR% of 36.67%, UR of 24.40%, AST% of 13.53%, ATR% of 55.81%, FTR of 20.38%, ASTD% of 39.79%, and STL at 5.50. The third cluster had an average 3FG% of 33.79%, MR% of 34.65%, UR of 23.29%, AST% of 20.45%, ATR% of 55.59%, FTR of 25.05%, ASTD% of 34.80%, and STL at 40.25. The fourth cluster had an average 3FG% of 35.02%, MR% of 35.96%, UR of 21.20%, AST% of 14.57%, ATR% of

54.43%, FTR of 22.76%, ASTD% of 47.10%, and STL at 17.61. The fifth cluster had an average 3FG% of 35.06%, MR% of 35.50%, UR of 21.06%, AST% of 15.13%, ATR% of 54.71%, FTR of 21.51%, ASTD% of 46.28%, and STL at 25.48.

For the first cluster, it appears the players were classified by their high 3FG%, MR%, ATR%, ASTD%, and low UR and FTR. This indicates to me that these players are good scorers because they are both good at shooting and finishing at the rim, while also not needing their ball in their hands constantly to find ways to score. They can play off a primary ballhandler and fit into an offense without needing it centered around them, so I will classify these players as “off-ball scorers”.

For the second cluster, it appears the players were classified by their high MR%, UR, ATR%, and low 3FG%, AST%, FTR, ASTD%, and STL. This type of player seems to be offensively oriented, limited mainly to scoring within the 3-point line but able to carry a heavy offensive load efficiently while creating their own shots. This cluster will be classified as “2-point scorer” for their effectiveness within the 3-point line.

For the second cluster, it appears the players were classified by their high MR%, UR, ATR%, and low 3FG%, AST%, FTR, ASTD%, and STL. This type of player seems to be offensively oriented, limited mainly to scoring within the 3-point line but able to carry a heavy offensive load efficiently while creating their own shots. This cluster will be classified as “2-point scorer” for their effectiveness within the 3-point line.

For the third cluster, it appears the players were classified by their high UR, AST%, FTR, STL, and low 3FG%, MR%, and ASTD%. This type of player looks to be a poor shooter, but still plays an important role on offense through their playmaking abilities. They are also statistically the best defenders on average, having the highest STL at 40.25. I will classify this cluster as “defensive floor general” for their ability to set up their teammates and impact the game on both sides of the ball.

The players in the fourth cluster were classified by their high 3FG%, MR%, FTR, ASTD%, and low ATR%. This player seems to be well-rounded in their skillset, not showing many inherent weaknesses but maintaining as a viable weapon on both offense and defense. This player will be classified as a “combo scoring guard”, a basketball term for a guard harnessing the ability to play both the point guard or shooting guard position and both score for himself or play make for his teammates if needed.

The last cluster of players were classified by their high 3FG%, MR%, AST%, STL, and low UR and ATR%. This cluster appears to be primarily good at shooting and defending, a skillset that is often highly valued on teams for its versatility. This cluster will be classified as a “2-way shooter”.

Discussion

This clustering process is a solid way to get a general grasp of a playstyle a player may fall into, but there are limitations associated with the clusters. For one there were not nearly as many defensive variables used as offensive variables, which does not paint the entire picture of what a quality defender looks like. Using defensive variables such as opponent field goal percentage or team on and off defensive ratings for a player could be used in addition to get a player’s more concrete defensive profile. Another limitation was that only guards were used in this study. In further research, other position players such as forwards and centers can be analyzed and used to construct a full lineup that would theoretically synergize with each other. A research interest that could be further analyzed is instead of classifying players, we could also classify teams and understand what a team’s playstyle is. This would be important for opposing coaches because they could develop a game plan for their opponent based off similar opponents they've played before, not having to create a new game plan from scratch each time.

Figure 3 shows a little more in-depth analysis of how the players were clustered together. It is a scatter plot of a player's 3FG% and UR, and can be seen that players that had a high percentage in 3FG% but low UR tended to be put into the purple and light green clusters.

Conclusion

In conclusion, classification in basketball is crucial to knowing the strengths and weaknesses of any given player. Not even specific to basketball but in many team sports, knowing your personnel is one of the key skills for a coach to possess. Classification is a tool to help with that process and uncover new opportunities to experiment with lineups, as well as putting a coaches' players in the best position to succeed. This research is ripe for potential in helping drive the use of analytics in sports and dawning a new era in the industry.

Appendix

Link to the code on GitHub - https://github.com/Coast2Coast14/Pivot-Analysis-Research-Project

Figure 1

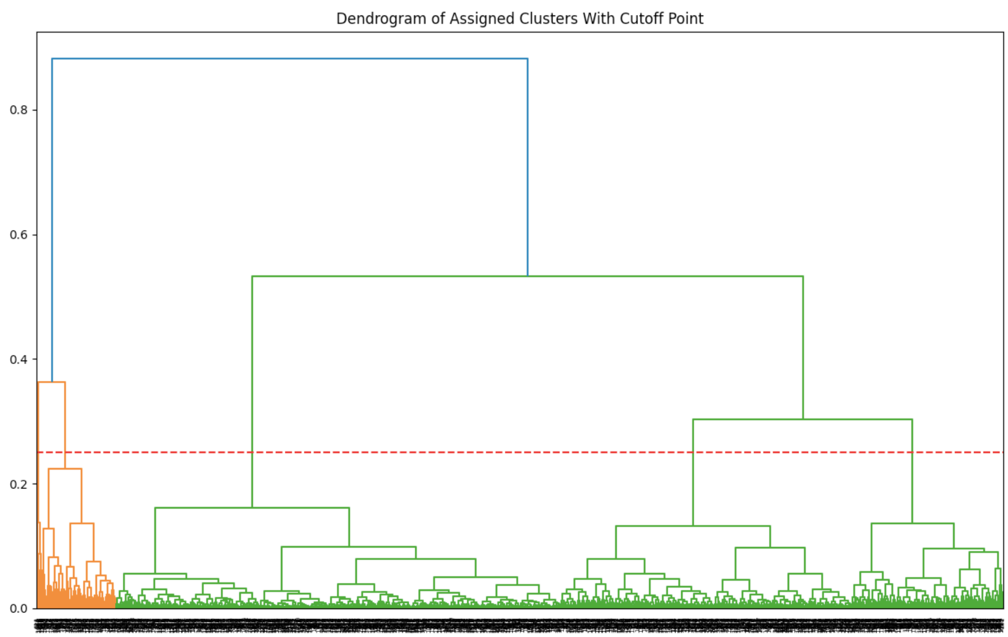
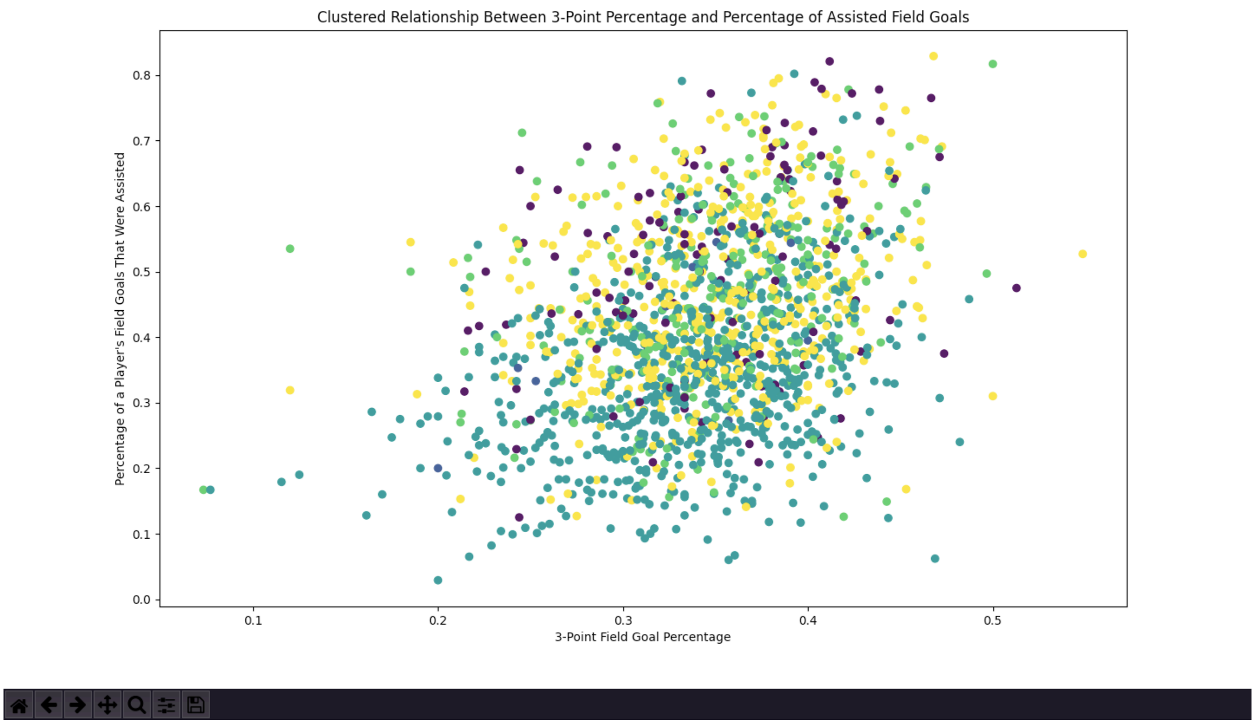


Figure 2

cluster	player_shooting_efficiency_fg3_perc	player_profile_mid_percentage	player_usage	player_assists_assist_percentage	player_profile_atr_percentage	player_four_factors_ftr	player_profile_assisted_percentage	player_boxscore_stl
0	0.3476	0.3612	0.2080	0.1261	0.5643	0.1926	0.5102	11.8607
1	0.3169	0.3667	0.2440	0.1353	0.5581	0.2038	0.3979	5.5000
2	0.3379	0.3465	0.2329	0.2045	0.5559	0.2505	0.3480	40.2514
3	0.3502	0.3596	0.2120	0.1457	0.5443	0.2276	0.4710	17.6096
4	0.3506	0.3550	0.2106	0.1513	0.5471	0.2151	0.4628	25.4751

Figure 3


Clustered Relationship Between 3-Point Percentage and Percentage of Assisted Field Goals
Percentage of a Player's Field Goals That Were Assisted
3-Point Field Goal Percentage
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.1
0.2
0.3
0.4
0.5