MSDS 456 Basketball Assignment

Benjamin Xiao

Team: Portland Trail Blazers

1. Using the four-factor model discussed in Chapter 28 of Mathletics, determine where we ranked in the NBA last season and how we currently stack up so far this season. Briefly describe where we are deficient and where we excel.

The data is gathered from NBA.com/stats and uses effective field goal percentage, free throw attempt rate, turnover rate, and offensive rebounding rate to comprise the four-factor model. The data for this analysis only includes the current season up until the all-star break where the last games were played on February 14th, 2020. We will then get a broad look at both how our team and the opponents' teams relatively fare.

Category	Value(Difference in Offense – Defense)	Rank
EFG	0.012	10
FT rate	0.012	13
TOV rate	0.005	5
OREB rate	0.014	2

Table 1.1: Four-factors table of values and rank of the 2018-19 Portland Trail Blazers

Category	Value(Difference in Offense – Defense)	Rank
EFG	0.005	16
FT rate	-0.029	22
TOV rate	0.002	15
OREB rate	-0.029	T-26

Table 1.2: Four-Factors table of values and rank of the 2019-20 Portland Trail Blazers

Comparing to the previous season, the ranks are significantly down across the board. Granted, the team has taken major injury losses in Nurkic and Collins as well as committed to a huge roster turnover in the off-season. The differences in rebounding and free throw rate are especially stark.

Starting with effective field goal percentage, we are about even in that category, and it reflects in the rank at 16. This means we are just about average which may be an indicator that major improvement is

needed in the near future. Both the free throw and offensive rebounding rates are negative and in the lower third of the league. A difference in rebounding was expected but perhaps not to this degree. While offensive rebounding is not a priority in current basketball strategy, free throws are still crucial ways to accumulate points throughout games.

The turnover rate also sits at just average at about 0 difference per game. I think this is especially troubling because the Blazers rank 3rd best in turnover per 100 possessions at 12.4 turnovers. This is suggesting that they are not turning the ball over, but that the defense is also not putting much pressure on to force turnovers as well. This is one point I think the team can tilt defensive strategy towards to balance some weaknesses. With such a low turnover rate, we'd ideally want at least a negative difference in turnover rates.

2. Looking at page 193 in Mathletics, conduct a similar analysis that determines the relative importance of each of the four factors. Based on your findings, how many games should we have won last year? Make sure to use a training and a test set to evaluate your predictions.

This data starts at the 2013-14 season and ends at the 2017-18 season. The previous season (2018-19) will be used to make evaluations. This is the same four-factor variables used to form tables 1.1 and 1.2. The methodology uses the differences in the four variables and an ordinary least squares regression to determine the relative importance to regular season wins. The resulting equation is:

$$y = 48.143 + 400.429 \times EFG\ diff + 40.458 \times FTR\ diff + -301.824 \times TOV\ diff + 132.068 \times OREB\ diff$$

This suggests that differences in effective field goal percentage is the most telling of whether a team will win or lose. The next in intensity is the difference in turnover rate which also makes sense. EFG encapsulates both overall offensive and defensive measures. Turnover rate differences hints at which team effectively had more shots and thus, more chances to score more points. Using the time frame, this

then suggests that free throw rate is not nearly as important as offensive rebounding rate. This could easily be a shift in overall strategy with the stronger emphasis on 3-point shooting.

Using this model, the previous season was projected to win 48 games. The team instead won 53 games and overachieved under this model. Given the results of the playoffs where the team defeated both the MVP lead Oklahoma City Thunder and a rising Denver Nuggets team, this model may be antiquated for modern game.

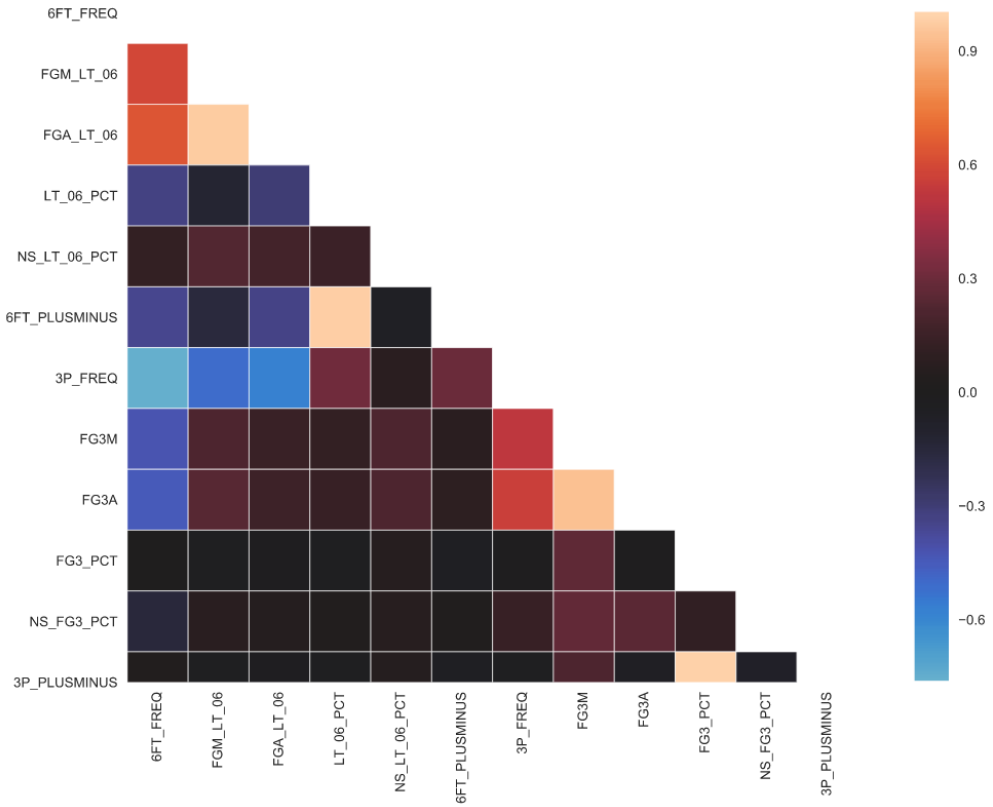
3. Using historical data, determine a new classification framework for identifying player types. Provide a description of each player type based on the metrics that you select.

As I researched where the current Trail Blazers have the most glaring issues, defense is the biggest issue that can be addressed. The Blazers currently rank 26th in defensive rating, giving up 113.3 points per 100 possessions per NBA.com/stats.

I used a dataset from NBA.com/stats to specifically categorize defensive players in this assignment and based the analysis on spatial data. The seasons range from 2013-2019, the previous complete 6 seasons. I chose these seasons because they are the most relevant to the current state of basketball. I ended up confining the data to shots defended under 6 feet of the basket and 3-pointers because current NBA strategy emphasizes these two shots. I felt it was most important to assess players operating in these two situations. The categories include frequency (% of possession) defended, defended field goals attempted, defended field goals made, field goal percentage, and the expected difference in field goal percentage.

For now, I ignored deflections, defensive rebounding percentage, blocks, and steal statistics. For future analyses, I would preferably include all these statistics except steals. Steals are a highly polarized statistic that may favor poor defensive players who suboptimally commit to steals.

I first started by looking whether any of these statistics have any strong natural correlations.



The

Figure 3.1: Correlation heat map for defensive statistics. Data ranges from the 2013-14 season to the 2018-19 season.

The glaring problems are going to be in light blue and light orange. Looking at the each individually, it made sense that the higher the frequency, the more field goals in that are a player would defend and therefore be subjected to makes and attempts. It also made sense that the more 3-pointers defended, the less shots at the rim a player defends and vice versa. Even the biggest and fastest players cannot contest two players 20 feet apart at once.

For this problem, I chose to use K-Means clustering to find the optimal defensive positions available on the floor. First, I had to scale statistics by their minimum and maximum values to a value between 0 and 1 to ensure comparable statistics by Euclidean Distance. I chose this method for the simplicity in implementation and ease of parsing out categories after model fitting. Given more time, I would explore algorithms that can make use of a non-flat geometry. Figure 3.2 shows that the frequency distributions of each variable is skewed one way or highly modal.

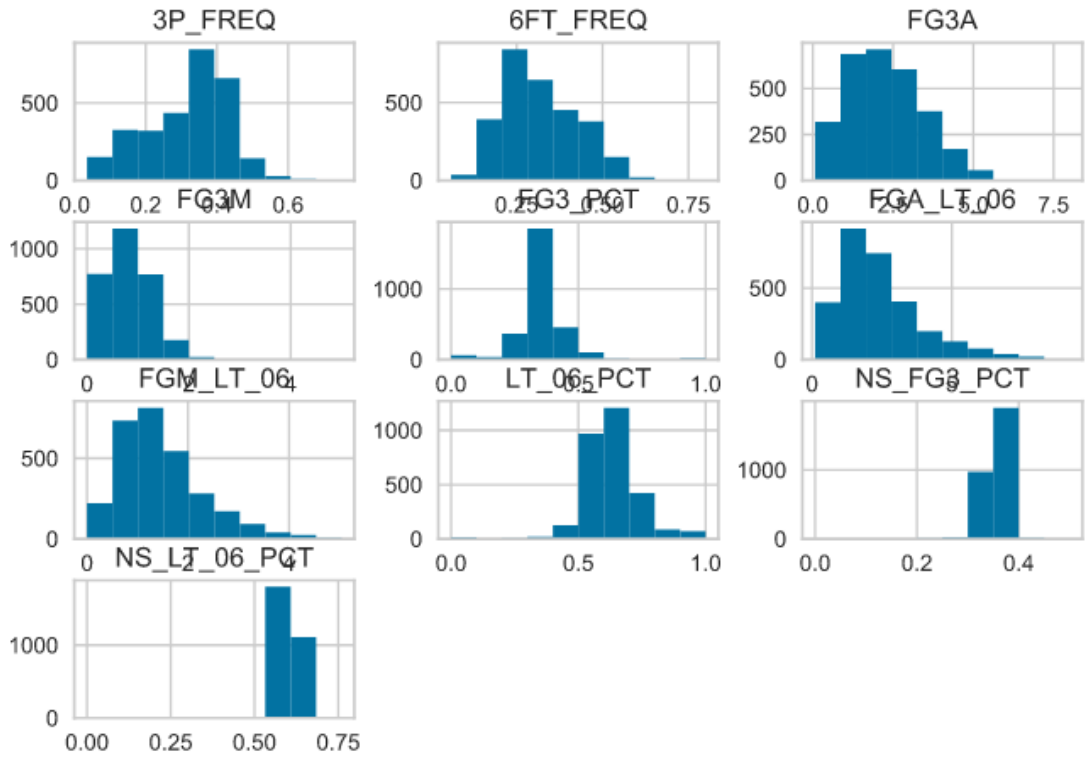


Figure 3.2: Histograms of defensive data gathered from the 2013-14 to 2018-19 NBA seasons.

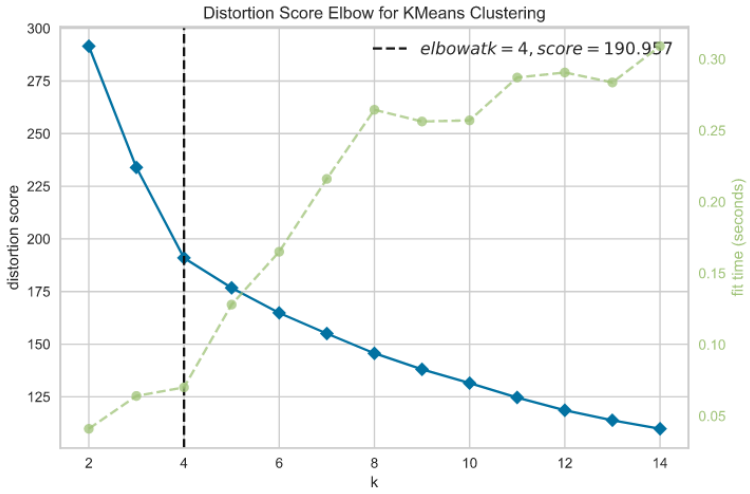


Figure 3.3: Charting fitments of KMeans clustering to the defensive statistics dataset by distortion score.

Figure 3.3 shows that the optimal number of clusters in this dataset using K-Means clustering is 4. The resulting categories are described as such:

1. Strong All-Around Defenders

These players can defend nearly every position on the floor strongly but spend a higher frequency defending perimeter shots. They result in relatively low opponent shooting percentages while either defending or dissuading shots from going up all together.

2. Primary Interior Defenders

These players are the primary defenders at the rim and are occasionally asked to switch onto wing players or close out on 3-point shooters. They defend interior shots with high frequency and 3-pointers with lower frequency.

3. Weakside Interior Defenders

Weakside interior defenders may be players such as Serge Ibaka who come from the weakside as help defenders when a primary action such as a pick-and-roll forces help from another player. These players defend wing-players with about the same frequency as primary interior defenders but are mostly asked to be help defenders within 6 feet of the basket.

4. Perimeter-only Defenders

Perimeter defenders will spend most of their time defending other wings at the 3-point line. These players often always only challenge perimeter shots and are rarely asked to defend shots within 6 feet. This could be because of a size limitation or because of a common defensive scheme.

After finishing this analysis, I believe it is not comprehensive enough to tell much besides what kinds of shots a player typically defends. I would like to later include more data and more clusters to add in a factor about how well a player defends whether that be a good perimeter defender or a poor interior defender.

4. Determine the types of players that currently make up our roster. Is there a particular type of player that we are lacking, or do we have multiple types of players that should be adjusted? What type of player(s) should we look to add this offseason either through free agency or the draft? Tie in your answers from 1 and 2 above

Player	Category

Hassan Whiteside	Primary Interior
CJ McCollum	All-Around
Carmelo Anthony	All-Around
Anfernee Simons	All-Around
Damian Lillard	All-Around
Nassir Little	Perimeter-only
Trevor Ariza	All-Around
Mario Hezonja	All-Around
Gary Trent Jr	Perimeter-only

Table 4.1: Players on the current Blazers' roster who have played more than half of the current season and their categorizations.

Table 4.1 shows a filtered list of players and their categorizations. The majority of active players are all-around defenders meaning they are asked to defend both the perimeter and rim at regular rates. The glaring missing category is a weakside interior defender. This may be because the of the defensive scheme the coaching staff has devised to hide the team's size weakness. The other significant weakness is the lack of a backup interior defender all together. This could mean that the team is taxing the only interior defender, Hassan Whiteside. During the off-season, the front office and coaching staff should work together to sure up interior defense and perhaps gather players who are stronger perimeter players to force a higher difference in turnover rate. I would also strongly advise a different categorization based on rebounding ability because that is one area where the Blazers are significantly lacking.

Sources:

https://stats.nba.com/players/defense-dash-3pt/

https://stats.nba.com/players/defense-dash-lt6/

https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html

https://stats.stackexchange.com/questions/81481/why-does-k-means-clustering-algorithm-use-only-euclidean-distance-metric

https://www.scikit-yb.org/en/latest/api/cluster/elbow.html

https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html

https://towardsdatascience.com/multicollinearity-in-data-science-c5f6c0fe6edf