Luke Phillips, Austin Ritz, Lars Knutson
CSCI 4022 Final Project
May 2, 2021

Using RAPTOR Metrics to Evaluate NBA Player Performance

Introduction: We wanted to use RAPTOR scores to gain a better understanding of NBA players' impact. First, we were interested in the rankings of players regardless of year for these various statistics. We found that based on the historical RAPTOR statistics Lebron James was the most dominant player regardless of era. We also wanted to look at how prolific players progress throughout their careers, and if there are any significant results based on that. Secondly, we tried to find out if RAPTOR correlates to any specific season. More specifically, we want to see if there is a correlation between RAPTOR values between players in a specific season vs other seasons. Finally, we evaluated how players are performing with respect to their compensation. Which players are overpaid? Are any underpaid?

Data: The RAPTOR (Robust Algorithm (using) Player Tracking (and) On/Off Ratings) statistic is an advanced 'Plus-Minus' statistic that determines how many points a player contributes to his team's offense and defense per 100 possessions. In short, it helps determine a point differential when the player is on the floor vs when he is off the floor. This data set contains the offensive, defensive, and total raptor scores in the form of box RAPTOR and On/Off RAPTOR. This also contains the PREDATOR statistic which is a modified version of the RAPTOR statistic that takes into account predicted future performance. On top of that, the WAR statistic is also included which modifies the RAPTOR statistic to reflect a players value based on minutes played, pace (a teams possession length), and playoff and regular season pace with more weight placed on playoffs.

Our data set is a historical RAPTOR collection from the 1977-current NBA seasons for every player. It includes the offensive and defensive versions of these statistics as well as the combined offensive and defensive, or total, versions of RAPTOR, WAR, PREDATOR, and PACE statistics.

It is important to note that we are basing this off of a modified RAPTOR statistic that has been made to include players from the 1977-2020 seasons using the same data. This is important because in the 2013/14 season the NBA began collecting advanced statistics on players which is used in the modern version of the RAPTOR statistic. We are interested in a broader view of the NBA players so we will be using the historical data rather than the modern data sets, which only includes players from 2013 onwards.

Our data set can be found at the following link:
https://github.com/fivethirtyeight/data/tree/master/nba-raptor.

Impact:

Player metrics in the NBA have always been an important factor when teams decide who they want on the floor. In recent years, more advanced statistics have been developed that can

better determine who is really playing well and who has empty stats. The RAPTOR statistic is one of those advanced metrics that can cut the fat off of players stats and get to the meat of the problem.

The RAPTOR statistic is an advanced 'Plus-Minus' statistic that determines how many points a player is scoring and defending while he is on the floor. For example, if LeBron James were to play the first five minutes of a game and his team was ahead by 10 points after those 5 minutes, then his Plus-Minus score would be +10. On top of the basic BPM (Box Plus-Minus), RAPTOR attempts to place a value on the statistics that a player is accruing. Such as the value of a rebound, contested shot attempts, how many of their points were assisted, and other statistics that wouldn't normally show up on the box score.

We also are using the WAR statistic (Wins Above Replacement), which is a slight modification of the RAPTOR statistic. We see value in the WAR stat because it takes into account the pace of play a player has and values playoff performance over the regular season. Because playoffs are the main goal for players, this statistic is valuable when determining a player's true value/performance.

By using these advanced metrics, we can get a better understanding of which players in the NBA are truly great, by clustering them amongst their peers. On top of that, we can compare their performance to their salaries to see if we believe that they are being overpaid.

Initial Research: Our initial research consisted of finding simple patterns and players that consistently placed at the higher end of the RAPTOR score. The top three raptor scores of all time belong to Lebron James(12.58), Stephen Curry(12.48), and Michael Jordan(12.29). The top three war scores of all time belong to Michael Jordan(28.76), Lebron James(28.52), and Michael Jordan(28.01). The top three predator scores of all time belong to Stephen Curry(13.35), Michael Jordan(12.71), and James Harden(12.60).

We also found the basic stats surrounding the three data sets.

Raptor Total Stats:		**War Total Stats:**		**Predator Stats:**	
mean	-1.305130	mean	2.138467	mean	-1.492685
std	3.279292	std	3.564712	std	3.440341
min	-19.317679	min	-7.382978	min	-18.712307
25%	-3.305822	25%	-0.128583	25%	-3.699204
50%	-1.247017	50%	0.806606	50%	-1.483399
75%	0.766939	75%	3.376565	75%	0.729446
max	12.580815	max	28.762877	max	13.347925

Methods: We decided that our best option for grouping great players would be the K-Means algorithm. This algorithm fits perfectly into what we are trying to achieve. First of all, K-Means can make sense of large groups of data, and show patterns that we wouldn't see otherwise. For example, trying to compare the top 20 players of the last 5 years would be difficult, as most of their RAPTOR scores are scattered across the board. But, when using the K-Means algorithm, we can see that certain clusters stand out that have significance.

Our K-Means algorithm works in X parts. First, the number of predetermined centroids are placed randomly amongst the data. Second, every data point is assigned to its nearest cluster. Third, the centroids are adjusted to reflect the center of their clusters. Last, the cycle begins again until the clusters change under a specified tolerance or a certain number of iterations has occurred.

We plan to use our K-Means algorithm in several ways. The first of which will be clustering the top 20 players based on their RAPTOR scores over the last 5 years. Using that, we will cluster them to see how their defensive vs offensive performance has been over those 5 years. We believe that most of these players will specialize in one side of the ball over the other. The next way we will use the algorithm is to cluster cherry-picked players to compare their salaries. We believe that certain players will fall into similar clusters in terms of their production, but their salaries will differ.

Lastly, we wanted to use one type of analysis that wasn't K-Means related, so we took a look at the top 27 paid players from the 2018-2019 season and attempted to place a value on them. This value stems from a relationship between their RAPTOR scores and their salary, and their WAR scores and their salary. The equation we used was as follows:

$$\textbf{RAPTOR\_VALUE} = Player\ Salary/RAPTOR$$
$$\textbf{WAR\_VALUE} = Player\ Salary/WAR$$

Results: We evaluated two main problems in this project. The first, determining if a player was being paid fairly, is shown below in Figures 1 and 2 as well as Table 1. Our second, evaluating offensive and defensive dominance over the past five years in the league, is shown in Figure 3 as well as Tables 2 and 3.

To evaluate player salary fairness, we took a random spread across the top 70 salaries in the NBA for the 2020-2021 Season. Below, in Table 1, the five players selected along with their salaries are shown. It is worth noting that Kyle Lowry is paid much less than Stephen Curry while making slightly more than DeMar DeRozan.

Table 1. 5 Random Players - 20-21 Salary

Player Name	Salary 20-21 (USD)
Stephen Curry	$43,006,362
Kyle Lowry	$30,500,000
DeMar DeRozan	$27,739,975
Aaron Gordon	$18,136,364
Andre Drummond	$28,751,774

Below, in Figure 1, these five players were clustered into three clusters using k means. We used euclidean distance between each point to group them. Using this method, three of the six seasons represented have Lowry in the same cluster as Curry yet he is paid ~$10m less. Meanwhile, his teammate DeRozan is making close to the same amount but completely belongs to the central cluster. In Figure 2, this gap in output becomes even more apparent between DeRozan and Lowry. This suggests, when considering how close Lowry is to Curry, that Lowry is underpaid and DeMar is overpaid for what they bring to their team. Andre Drummond also seems to be overpaid slightly but contributes significantly to the team defensively. Drummond and Gordon are not reliable indicators in the WAR clustering chart because they did not have playoff appearances.

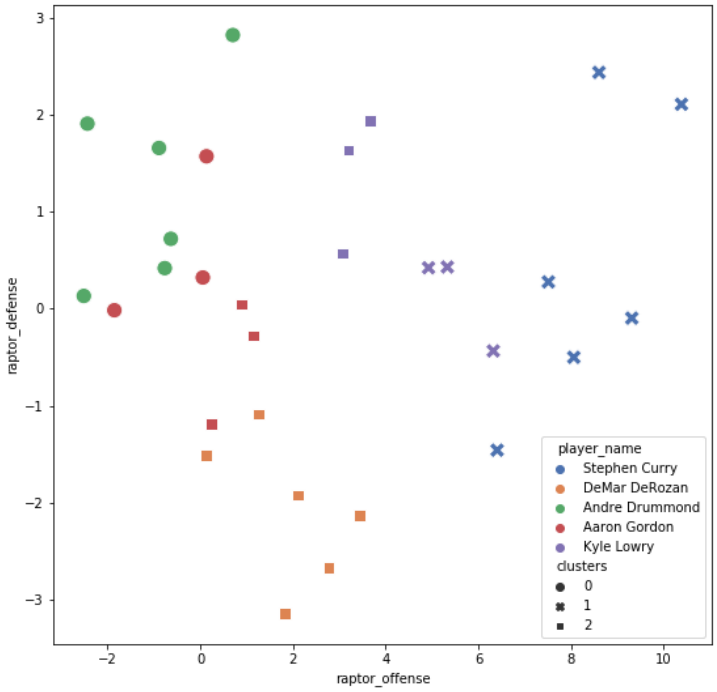

Figure 1. Raptor Offense/Defense Clustering - 2015-2021 - 5 Players

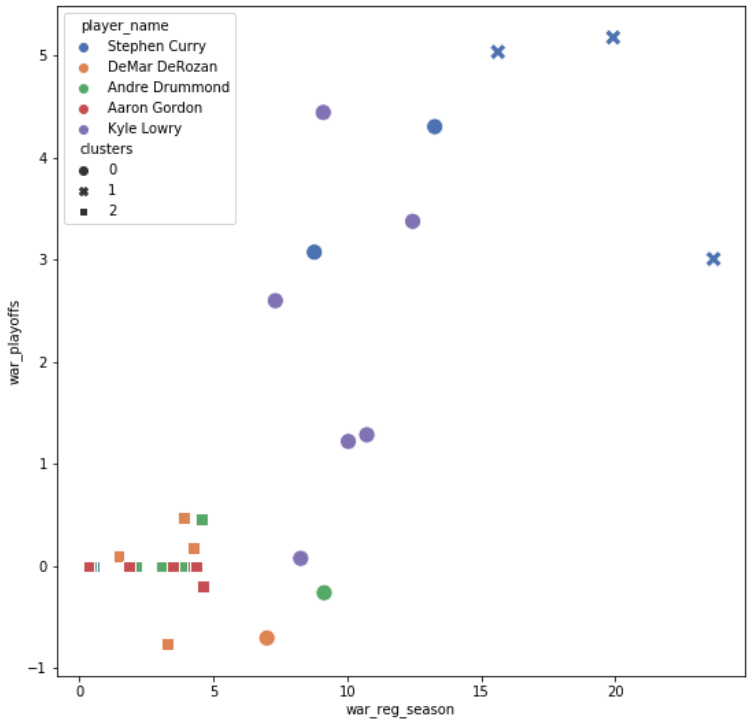

Figure 2. War Playoffs/Reg_Season Clustering - 2015-2021 - 5 Players

To evaluate player dominance over the past six years, we clustered with euclidean distance again. Comparing RAPTOR offense and RAPTOR defense for the top twenty players in the league gave us Figure 3 below. When we specified five clusters there were clear groupings in defensive, offensive, and balanced players. To determine if a player was offensively or defensively dominant, we summed how many times they appeared in each cluster and then divided that by the number of times they appeared in the plot. This means that if a newer player only appears twice that proportion is accounted for.

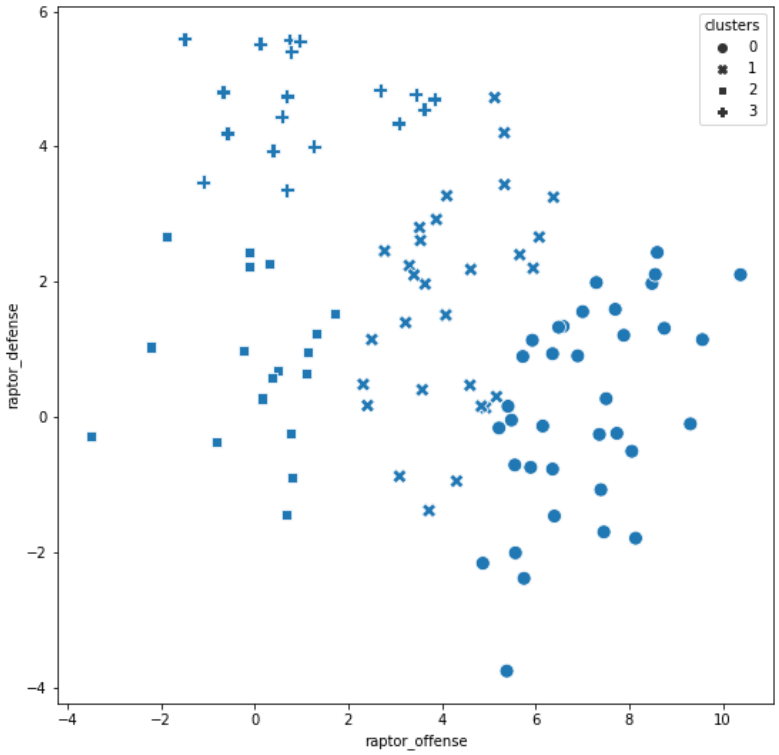

Figure 3. Raptor Offense/Defense - 2015-2021 - 20 Top Players

The results for offensive dominance held true to what we already knew about these players. LeBron James, Kevin Durant, and Stephen Curry are known to be offensive powerhouse players which was reflected below in Table 2. Meanwhile players such as Giannis Antetokounmpo and Jimmy Butler, both of whom are known to be extremely good players, did not display the offensive dominance over their past six years that someone like LeBron is known for.

Table 2. Top 20 Players Relative Offensive Dominance - Clusters 3 and 0

	player	score
0	LeBron James	66.67%
1	Kevin Durant	100.0%
2	Giannis Antetokounmpo	0.0%
3	Anthony Davis	66.67%
4	Stephen Curry	100.0%
5	Kawhi Leonard	66.67%
6	Damian Lillard	83.33%
7	Luka Doncic	50.0%
8	Nikola Jokic	40.0%
9	James Harden	100.0%
10	Jimmy Butler	0.0%
11	Jayson Tatum	0.0%
12	Paul George	16.67%
13	Chris Paul	83.33%
14	Joel Embiid	100.0%
15	Ben Simmons	0.0%
16	Rudy Gobert	100.0%
17	Bam Adebayo	0.0%
18	Pascal Siakam	0.0%
19	Bradley Beal	16.67%

We then looked at players who were offensively dominant. Immediately we see that Anthony Davis, Giannis Antetokounmpo, and Jimmy Butler have all been among the top defensive players every season over the past six years. James Harden, much to our surprise, is a terrible defensive player while above in Table 3 he has performed offensively every season.

Table 3. Top 20 Players Relative Defensive Dominance - Clusters 3 and 1

	player	score
0	LeBron James	33.33%
1	Kevin Durant	0.0%
2	Giannis Antetokounmpo	66.67%
3	Anthony Davis	83.33%
4	Stephen Curry	0.0%
5	Kawhi Leonard	66.67%
6	Damian Lillard	16.67%
7	Luka Doncic	50.0%
8	Nikola Jokic	80.0%
9	James Harden	0.0%
10	Jimmy Butler	100.0%
11	Jayson Tatum	33.33%
12	Paul George	83.33%
13	Chris Paul	16.67%
14	Joel Embiid	100.0%
15	Ben Simmons	0.0%
16	Rudy Gobert	100.0%
17	Bam Adebayo	0.0%
18	Pascal Siakam	0.0%
19	Bradley Beal	50.0%

This also begs the question of why Joel Embiid is not higher on the overall rankings. Although he doesn't consistently put up the points that a player like Luka or Steph does, he has been both offensively and defensively dominant every one of the past six years.

Conclusion: Our results provide some good insight into players' impact. In regards to RAPTOR scores over several seasons, we found that very few players are able to stay in the best clusters for very long. There were only three players that achieved this, and they are all-time greats.

In the future, we could expand our results in several ways. For one, we would benefit from a more robust value function. This function could take into account more than just one season of wages, and perhaps a league average. On top of this, we would like to expand all of our results to include more players so that the dominance of some players is more clear, and perhaps there are standout players that weren't included.

We initially intended to include NBA players from previous seasons, but the RAPTOR is based on advanced statistics that weren't recorded until 2012. There is a modified RAPTOR statistic that is used for older players that could be used, but the results could be worthless. But nevertheless we would like to see what that brings.

In terms of work still to do, we would like to change our value interpretations. The way we calculated war and raptor value were a somewhat rudimentary short-term solution. Some players such as role players will be unfairly punished for seemingly "poor" performance compared to compensation when they are actually providing good value for the team.

In the future we would also like to find a way to find patterns and predict player performance in the future. To do this we would like to use a neural network. The WAR statistic would be useful for predicting future success.

In the analysis above we only looked at select players for pay comparison as well as the top twenty for overall dominance. A future step of this analysis would be to just include more players to get a more precise gauge of how overpaid/underpaid a player was in comparison to his peers. We could also include more players than the top twenty although this would likely just confirm what we already know about the players in the top rankings.