

A Comparative Analysis Between 3PA and Deep Mid-Range Attempts (DMA) in the NBA
(Two Variable Statistics Analysis)

Introducing a new variable: The NBA's Deep Mid-Range Attempts (or DMA)

Definition: A DMA is a shot that is counted when a player attempts to shoot the ball between 16ft and the 3P line (23.90 ft from the top of the key and 22ft from the corners) from the basket anywhere on the court. More specifically, the ball must release a player's fingers in a natural motion, however, it is worth noting that getting blocked still counts towards a player's DMA. Similarly, if a player is fouled inside the range and the ball goes into the basket (a three-point play opportunity) the shot is still counted as an attempt. However, if a player is fouled then misses the shot, it does not count because a foul was committed before the attempt.
In figure A, the shot distances are marked based on color. The DMA is counted when shots come from 16 ft and 23.90 ft (green to light blue).

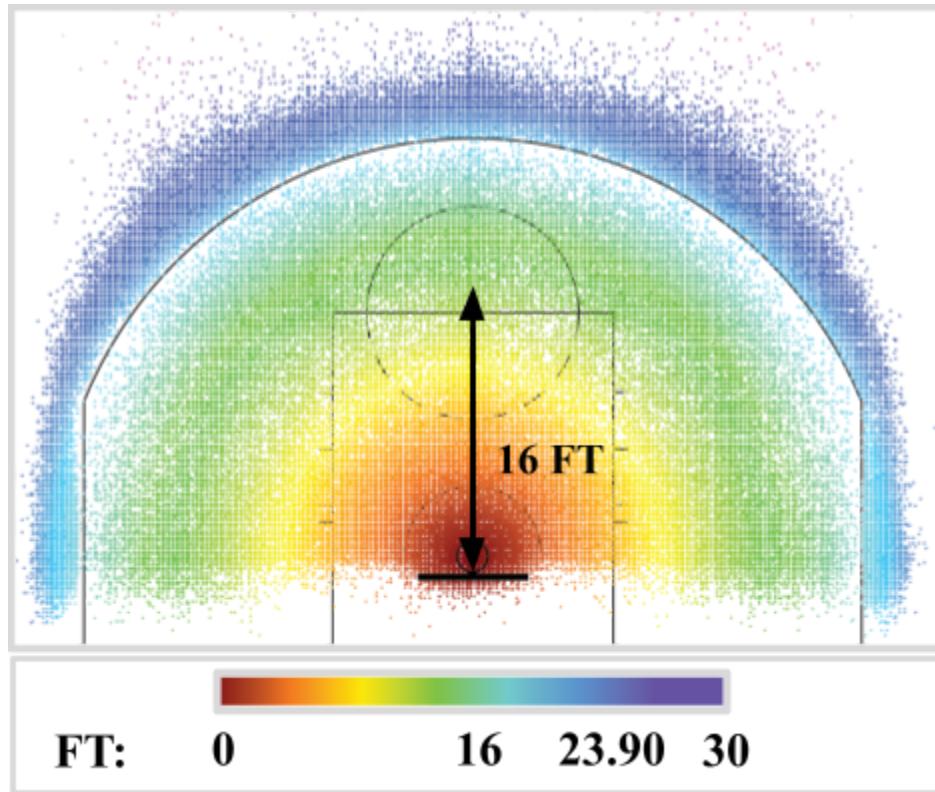


Figure A (NBA shot distance hot chart)

Data discussion:

1. Understanding our variables, their significance, sources, data type, and validity.

For my one-variable analysis, I did a comparative analysis of the 3PA variable between two different time periods of the 3P revolution (2010-2015 and 2015-2021) in the NBA. In the end, I concluded that 3PA per game since 2010 continues to rise over time in the NBA. I found that this phenomenon could be attributed to the rise of sports analytics, high pace offensive strategies, points per shot advantage, and efficiency. In this comparative analysis, I wanted to further back up my previous conclusions by looking at the correlation between two-point attempt (2PA) and 3PA across the same period. However, to strengthen previous data and conclusions found in my one variable and probability assignment, I decided to look at a more specific, complex, and interesting variable within the 2PA variables. Therefore in this report, I decided to use the deep mid-range attempt (DMA) variable. I created and calculated this variable using data from my previous data set and additional “advanced shooting stats” from basketball reference. I calculated from each team across 10+ seasons the DMA variable by multiplying a teams FGA by the % of FGA by distance from 16ft to the 3P line¹.

$$\text{DMA} = \text{FGA} * (\%_{\text{FGA}} \text{ 16ft-3P})$$

Using DMA, I want to analyze the correlation between the rise of 3PA and its contribution on deep mid-range attempts across the same time periods and population.

2. Data Classification

Variable	3PA (independent variable)	DMA (dependent variable)
Type	3PA is an average variable. In this data set, it was an average of three-point attempts per game per team, which is calculated by taking the total 3PA of a team during the regular season and dividing it by the number of games played during that season.	DMA is an average variable. In this data set. It is an average of deep mid-range shots attempts (16ft-3P) per team per game, which is calculated by multiplying a team's FGA by the % of FGA from 16ft to the 3P line during the regular season.

¹ https://www.basketball-reference.com/leagues/NBA_2021.html

Collection Method	The raw data of 3PA is population data because every single individual 3PA is counted and recorded across each game, season, and decade. In terms of collection methods, the NBA uses hundreds of cameras that use advanced A.I, such as computer vision to track data. Additionally, each game has actual scorekeepers that verify and count each stat. Therefore, in this context, every single 3PA is counted and rarely missed. If an event is missed, it will be corrected after the game using recorded footage.	The raw data behind the DMA variable is population data as well because every single individual FGA and FGA from 16ft to the 3P line is counted and recorded across each game, season, and decade. Each game ends with a whole number of deep midrange attempts that are accounted for, collected, and then later archived. In terms of collection methods, the NBA uses the same tracking technology and cameras as previously mentioned.
Classification	3PA is a quantitative discrete data type because the raw data consists of all specific numerical values that are counted. In the NBA, each arena is filled with hundreds of cameras that use computer vision to track player movement and events. Additionally, each game has actual scorekeepers that verify and count stats.	DMA is a quantitative discrete data type because the raw data consists of all specific numerical values that are counted and not measured. Moreover, each attempt is counted as a single event (whole numbers), meaning you cannot have a half shot attempt.
Impact of Definitions And Data Collection Methods	The definitions of the two variables in this project should not affect the analysis because the definitions are very clear and heavily reviewed by the NBA to account for any overlaps or loopholes. The data as mentioned before is also meticulously counted, recorded, and archived by the NBA and various 3rd party analytics firms. Moreover, even though the data is based on averages of an entire season this should not affect the strength of my conclusions and the overall trends. This is because I will be making general conclusions of entire seasons over time rather than specific players, games, or events which would not be appropriate given the nature of the data. Furthermore, because it is a population data set it will allow for a stronger conclusion compared to samples because you have every possible data point within the time period. Moreover, the data collection methods would not impact the strength of the conclusion as the scorekeepers are extremely well trained and 3PA and FGA are fairly easy for these scorekeepers to keep track of. Furthermore, the automation of most statistics allows for even fewer discrepancies. In fact, recently the NBA partnered with Second Spectrum ² to further track their players using A.I. As a result, in the early 2010s the NBA installed camera systems in every arena that track the X, Y, and Z coordinates of the ball and players at 30 frames per second. About an hour after each game ends, the league and second spectrum send this data to every team for their internal organization to analyze.	

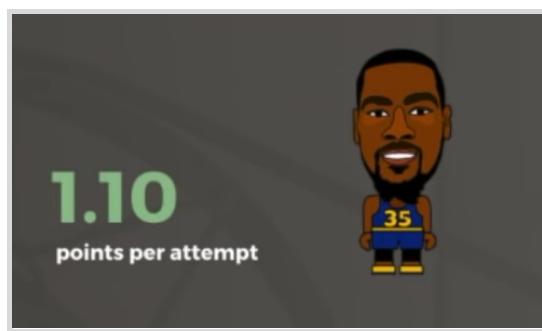
² <https://www.secondspectrum.com/index.html>

3. Hypothesis

I hypothesize that an increase in 3PA will result in a decrease in DMA. In other words, as 3-point attempts rise across time throughout the league, in general, deep mid-range attempts from 16ft to the 3P line will decrease as a result. Therefore this will cause a negative correlation between the two variables. The correlation would most likely be moderate to strong because of the obvious offensive value of taking a couple of steps back and shooting from the three-point line.

In my first probability project, I discussed the expected points per shot of the 3P shot and the mid-range shot. I found that on average the 3P shot was much more efficient because it resulted in a greater amount of points per shot. Additionally, when analyzing visually, in the last 15+ years teams have drastically changed their offensive playstyle. There are now fewer isolation plays, post moves, and contested mid-range shots, than we used to see from legends like Kobe Bryant and Michael Jordan. However, there can still be other factors that affect taking less DMA such as teams shifting their focus on taking more efficient shots, such as layups, dunks, and free throws rather than solely 3P shots. In sports, where even if measurements are precise the environment is noisy and there are often other influences that could affect how X and Y are connected. Often, R squared values of above 0.5 are good indicators in sports, especially in a dynamic sport like basketball.

Lastly, I believe that the media exaggerates the 3PA meta and labels the "2P" shot as a dying art. Throughout this project, we may find out that the league is taking fewer mid-range shots over time, however, the mid-range shot should not be discredited so easily because of the 3P shot only. Kevin Durant, LeBron James, and Kawhi Leonard are all top five players in the league right now and are all efficient mid-range shooters that shoot at a high volume, therefore, we may find a couple of outliers in our data as well per team that may indicate limitations as well.

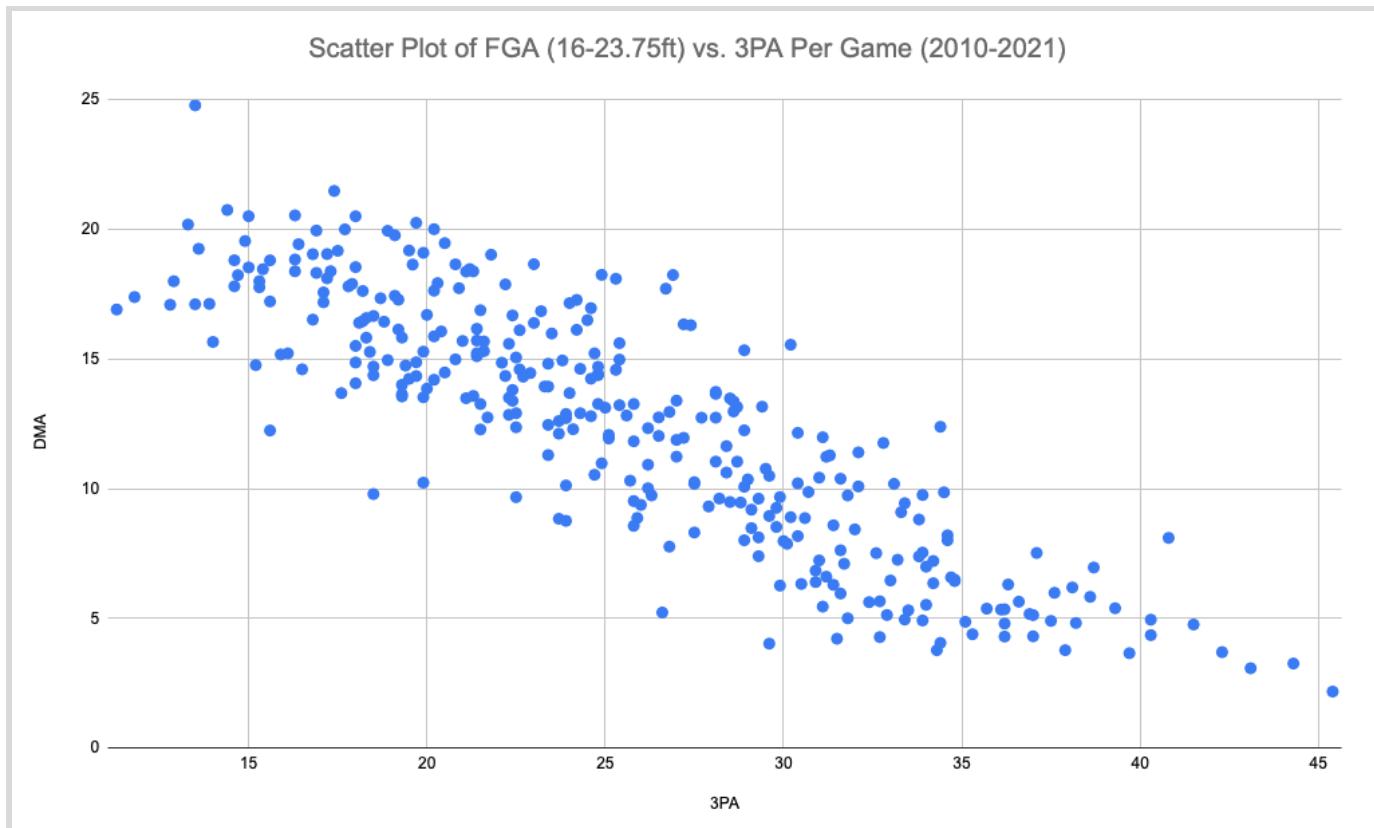


Kevin Durant's Mid-Range Stats: 1.10 points per attempt (55%) in 2018-2019.

League Average 3PA = 1.065 points per attempt or 36%.

Calculations and Results:

4. Scatter Plot (DMA & 3PA)



- Looking at the scatter plot, it is clear that there is a strong downwards trend, indicating that there is a negative correlation as predicted. The distribution seems normal and there is no distinct pattern to be worried about. We will get a much better understanding of the shape in the residual plot as well later on in this report.



5. Linear Regression (DMA & 3PA)

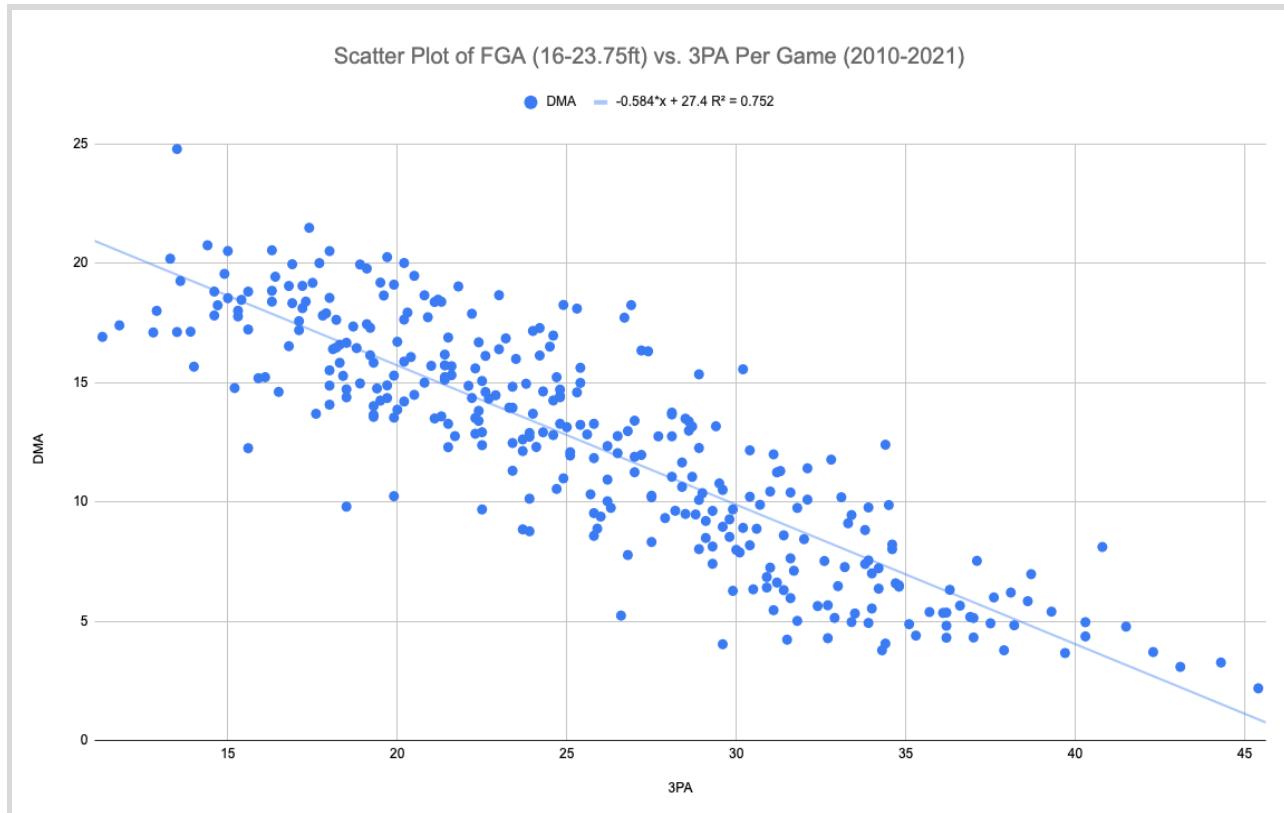
When I perform linear regression on the data above, I obtain a line of best fit with the following equation, r value, and r^2 value:

$$\hat{y} = -0.584x + 27.4$$

$$r = -0.8671108721$$

$$r^2 = 0.7518812645$$

Scatter plot (with linear regression):



- **What is the direction and strength of the correlation?**

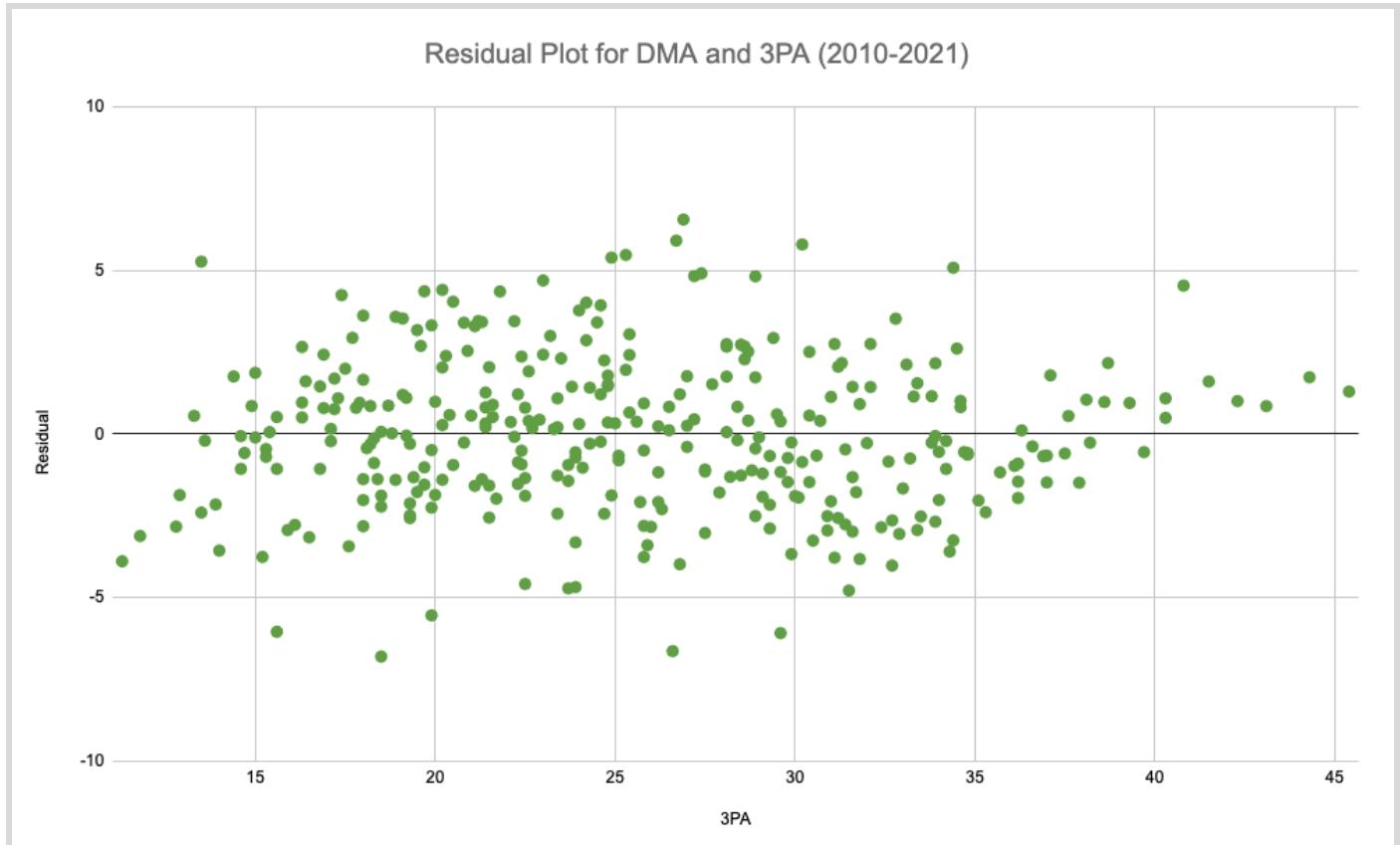
The correlation coefficient (r) is around -0.87, which indicates mathematically that there is a strong negative correlation between the per game 3PA and DMA because the value is within the range of 0.67 and 1.00. The direction of the data is negative and going downwards.

- **Computing the coefficient of determination (r^2) and understanding it in terms of the data collected.**

The coefficient of determination (r^2) is around 0.75, which signifies that about 75% of the change in the DMA is explained by the change in 3PA.

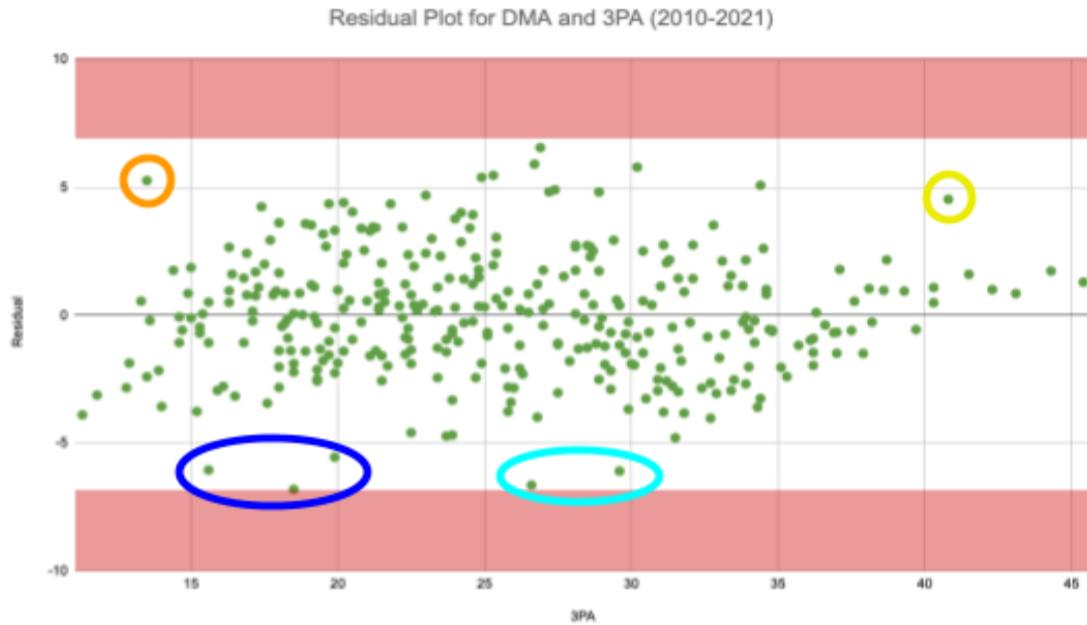
6. Examining residual values and analyzing its impact through a residual plot

- Residual Plot:



The residual plot has no visible pattern or shape, meaning that the linear regression model is appropriate for the data. Generally, the residuals are equally distributed above and below the x-axis. In this data set, above the x-axis, there are 161 points and below there are 169 points. Since there are a lot of data points already that indicate a trend it is safe to say that it would continue to be all random if more data was added, such as including data from 2000-2021. Therefore, the linear regression model above is the best method to describe the trend of the data.

7. Identifying Outliers:



- Range that contains strong possible outliers.

In general, as we look at both graphs, the data is mostly concentrated on the left side. This is because most of the teams have only recently started shooting 3P at a higher volume, which explains the fewer data points on the right side. As we look at the shape and distribution of the residual plot, it is clear that all the data points are scattered randomly. Generally, there are no major outliers that disrupt the trend overall, however, there are a few possible points that could be considered outliers that are worth mentioning.

The first set of outliers consist of possible outliers located above the x-axis and circled in orange and yellow. Although they are not in the range of strong possible outliers, they may be considered outliers because of their position compared to most data points within their 3PA range, they are quite far and inflated positively. The first point circled in orange was per game data from the Charlotte Bobcats in 2011-2012. They can be considered an outlier because their 3PA is way below league average and their DMA is much higher than their 3PA (very rare) and much higher than the league average as well. Moreover, their DMA is much higher than expected, more than 5 points above predicted for that 3PA range. After doing some extra research, I found out the Bobcats underwent massive culture shifts, rebranding, had a losing culture, inexperienced owners, and bad overall management³. In fact that year, they won an appalling 7 games only and had 59 losses. This team was fundamentally struggling to survive

³ <https://www.discussthegame.com/posts/5efe4a815514190004b4fe99>

and stay relevant. However, they fell behind and failed to follow trends due to bad coaching, a weak shooting team, and a terrible roster, making them a possible outlier to this trend.

The next possible outlier is circled in yellow above the axis and represents the Portland Trailblazers during the 2020-2021 season. They are considered outliers because their stats add noise to the r-squared value and hypothesis. These do not fit into the hypothesis because they have a very high 3PA rate (40.8 compared to the league average of 34.63) and a high DMA of around 8.11 (close to 5 points above the expected value). Therefore they are shooting a high amount of threes and deep mid-range shots as well. This can be due to the new offensive contributions of Carmelo Anthony (a high-volume mid-range shooter). Furthermore, their star duo of Damian Lillard and CJ McCollum have substantially increased their 3P shooting numbers and efficiency while only slightly decreasing in mid-range attempts. In fact, McCollum increased his 3PA by 40% and his mid-range attempts have decreased by 10% only in the last 2 years⁴.



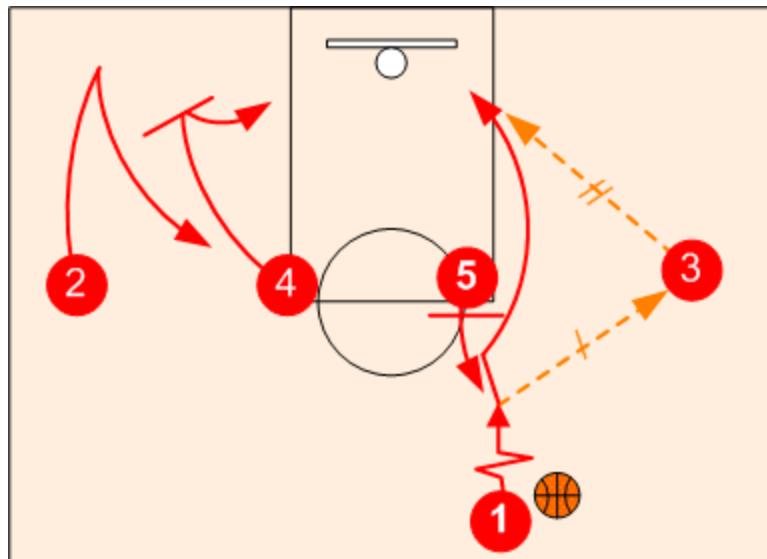
Lastly, the two other sets of outliers below the x-axis represent the Cleveland Cavaliers (2020-2021), the Milwaukee Bucks (2015-2016), the Denver Nuggets (2011-2012 and 2012-2013), and the Houston Rockets (2013-2014). The data points circled in blue represent the Nuggets and Bucks, who similarly to the Bobcats both had below league averages in 3PA. In contrast to the Bobcats, they had low DMA, which contradicts the line of best-fit prediction. This can be the result of both team's offenses not being about 3P shots or midranges. This makes sense as both team's star players and focal point of the offense, Giannis Antetokounmpo and Kenneth Faried⁵, are big men that exclusively score from the paint. The points circled in cyan are the Cavaliers and the Rockets. Specifically, the Rockets are considered outliers in this situation because they are the extremes of the 3P revolution. Their DMA (5 DMA, 2013-2014) is far below the league average for that year (15 DMA, 2013-2014) and the predicted value, which makes it seem like an outlier because it's way below most points that are concentrated near the x-axis. In contrast, the

⁴ <https://www.basketball-reference.com/players/m/mccolcj01.html>

⁵ <https://www.basketball-reference.com/players/f/farieke01.html>

Cavs are an outlier because their overall numbers are weak (extremely low 3PA and DMA) which visually makes it seem like they're an outlier⁶. After all, it is far away from the normal distribution of most points in the residual graph. This may be because they have the worst offensive efficiency rating, the second-worst record in the eastern conference, an outdated playstyle, and a young inexperienced roster.

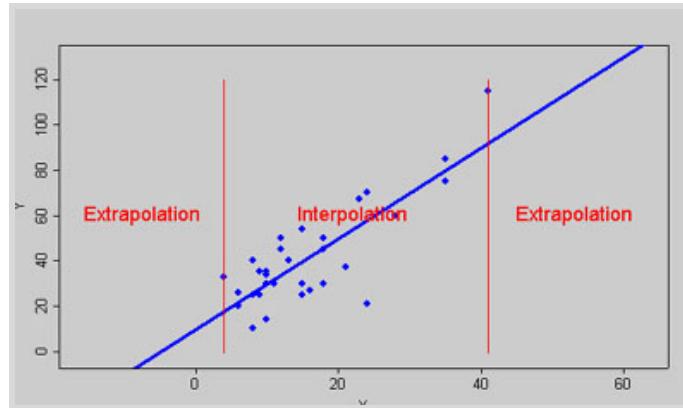
If I were to take these outliers out I would obtain an r squared value of about 78%. Although some of these points seem to be outliers, the reasoning behind the points does not approve the removal of them from the data set. These variations are perfectly natural and some teams just simply have weaker years as they adapt to new metas, coaches, or new players. Every team has different goals, ambitions, expectations, and talent, so removing the data point based on the natural progression and evolution of teams seems unjustified and would take away from the overall conclusion. In addition, this would be cherry-picking data to obtain a better linear regression model; we cannot just remove data points based on what would make the outcome look better. It would only be fair to rule out outliers if a team was completely isolated from the central concentration of data consistently throughout the years. As a result, this would drag the results of data towards that point and ultimately destroy the trend. In sports, there are always successful and unsuccessful teams. Moreover, as seen with the Nuggets, Bucks, and Cavs, not all teams hop on certain trends. This is completely fine as some teams have different strategies that work better for their roster, style, and coach.



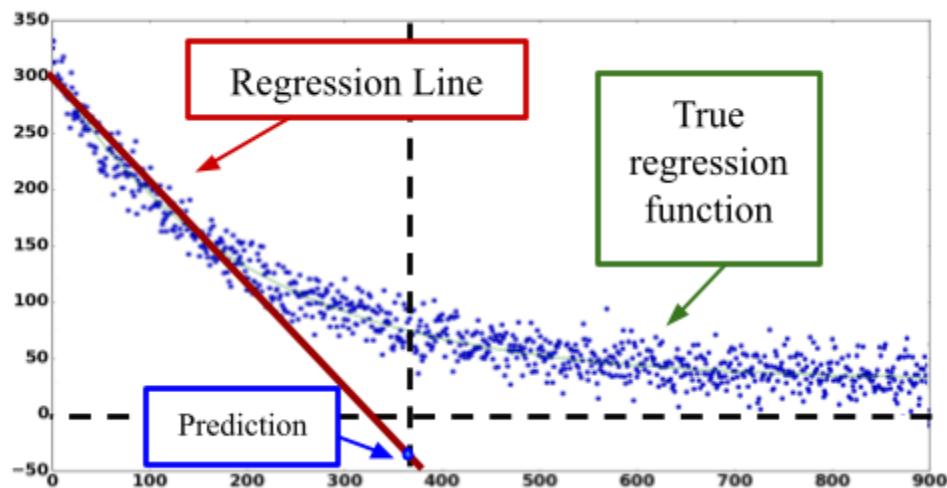
⁶ <https://hashtagbasketball.com/cleveland-cavaliers/content/collin-sexton-leading-cavaliers-midrange-barrage>

8. Making specific predictions using the line of best fit:

Predicting the DMA between 2010-2021 (interpolation) using the model above will be strongly accurate (75% accuracy which is very high in sports), however, predicting results from the past or the future can be tricky.



Extrapolation can be “dangerous” because the predicted values and logic may seem reasonable with regards to the trend of the data however it might not stay true to the actual y-value for a variety of reasons. Therefore before we can consider making such predictions, we must consider all the directions of the data, the hidden variables, and the ceiling of the data. The following graph below perfectly portrays this



The graph above is perfectly realistic to my data set because taking 3P shots doesn't necessarily equate to an exact linear decrease in DMA. The data I have so far is fairly small with regards to the big picture, and the entire history of the league. It represents a 3P bubble, however, as we

saw in the section above, players and teams don't necessarily increase and decrease their shooting habits in a linear fashion (ex: CJ had 3PA +40% and DMA -10%). Furthermore, there is an evident limitation and certain range in the function and regression above. Since teams cannot physically attempt negative deep mid ranges the range of the regression as we approach the limits must be $R = \{y \mid y \geq 0, y \in \mathbb{R}\}$. Furthermore, there are limits to the amount of 3PA (N) possible in a full game (48 minutes) and the domain must be $D = \{x \mid 0 \leq x \leq N, x \in \mathbb{R}\}$, which can complicate future predictions of 3PA. Additionally, as 3PA rises, DMA should plateau and hover around the 0 mark as we reach the x-axis limits to the right because of the range limitation.

In general, extrapolation is safer when the r and r squared values are relatively high as not many other factors would influence the response. In my case, the r values are very high and indicate a strong correlation however that may just be true for that time period collected. I believe that given what was stated above and the fact that there are too many variables that affect the both variables, such as a team's offense focusing on other aspects (not only 3PA) that extrapolation is too risky.

In contrast, interpolation should be effective because there is a clear strong trend within the current 3P revolution and a strong r^2 value of 75%. We can calculate and predict a couple of examples to illustrate the strength of the correlation.

Ex: A new team in the NBA called the Montreal Ducks are attempting on average 45 3PA per game, what is the predicted value of DMA?

$$x = 45 \text{ 3PA}$$

$$\begin{aligned}\hat{y} &= -0.584x + 27.4 \\ \hat{y} &= -0.584*(45) + 27.4 \\ \hat{y} &= 1.12\end{aligned}$$

\therefore The Ducks given a 3PA rate of 45, should shoot on average 1.12 DMA per game.

Ex: A weak shooting team in the NBA is attempting on average 10 3PA per game, what is the predicted value of DMA?

$$x = 10 \text{ 3PA}$$

$$\begin{aligned}\hat{y} &= -0.584x + 27.4 \\ \hat{y} &= -0.584*(10) + 27.4 \\ \hat{y} &= 21.56\end{aligned}$$

\therefore The team given a 3PA rate of 10, should shoot on average 21.56 DMA per game.

9. Conclusions

- **Interpretation of the Data (i.e what does my statistics mean?).**

In summary, the dataset shows there is a strong negative linear correlation between 3PA and DMA. This means that the more 3-pointers a team attempts, the lower their DMA will be per game. This idea is further reinforced with a strong r-squared value of 75%. Moreover, I can interpret from the data that correlation does not cause causation in DMA. As stated before, there are many hidden variables that can affect both variables, such as a coach's offensive plan, playsets, roster type, and team goals.

- **Correlation present (e.g. presumed cause and effect, common cause, etc.).**

Given all of the information, I believe that the 3PA and DMA variables have a common cause correlation. To have a guaranteed cause and effect relationship, an experiment would need to be held in a closed and isolated environment, where only the following variables would be analyzed. Moreover, there are too many hidden variables in this situation for the correlation to be presumed cause and effect. Hidden variables, such as a coach's offensive plan, roster type, and player skill sets. As we saw earlier, the Cleveland Cavaliers during the 2020-2021 season shot among the least in 3PA because their offensive strategy was primarily focused on many other aspects that were tailored to their star players, such as heavy pick and rolls, dunks, layups, and close mid-range attempts. No doubt that there is a strong correlation between the variables and their effects on each other, however, there are so many factors that affect both events in a basketball game. Therefore, it must be defined as a common cause because these external factors have a strong effect on both variables in basketball throughout a game, season, and possibly a playoff series in the NBA.

- **An explanation for my findings.**

The hypothesis is supported by the interpretation of the data above, the evident findings, the strong correlation, and overall conclusions. Given the scatter plot and residual plot, the following conclusions can be made.

- Throughout this report, it was interesting to see a relatively high r squared value. In general in sports, there are a lot of elements of randomness and variation. Although it is quite high it is not perfect. As discussed above, there are many more variables at play with both variables. Often when there is a correlation, many people automatically assume direct causation.

- Unfortunately, the media does this often with regards to 3P shooting. Therefore, it can be concluded that although 3P shoot has a strong influence on DMA, it is not the only factor that caused the decrease in DMA.
- Additionally, the analysis of outliers concluded that not all teams jumped on board this trend. This is natural as teams have different rosters or take time to evolve their strategy and culture. For example, the Los Angeles Clippers this season are one of the top favorites to win the finals. They shoot an average number of three, however, they also shoot above league average in DMA which is rare to see in the current NBA era. This works for them because they have elite DM shooters that are as efficient as very good 3P shooters in regards to the probabilities of expected points per shot. Situations such as these should not be discredited or removed from the data set. Rather it should be embraced because it shows the natural progress, evolution, and variations in the game of basketball.

Some possible explanations for the rise in 3PA and decrease in DMA are as follows.

- Efficiency! It's quite simple, the statistics and probabilities behind both types of shots favor the 3P shot and label the midrange as the least efficient shots in basketball. For example, a deep mid-range shot (16-23 ft away) is made on average at a rate of around 40.5%, while the 3pt shot is made 35.5% of the time. However, the 3pt shot is worth 50% more! Therefore we can rank the value of each shot:

$$\textbf{2P Shot: } 0.405 * 2 = \textcolor{red}{0.810} \text{ Points per Attempt}$$

$$\textbf{3P Shot: } 0.355 * 3 = \textcolor{green}{1.065} \text{ Points per Attempt}$$

- As a result of this and the rise of sports analytics in the last 10 years, inefficient parts of the game have slowly started to fade away with time...

Final thoughts and ideas:

As players and teams continue to work on their three-point game accuracy and attempts, I believe that the r^2 value will continue to rise in correlation to DMA. But, what should be taken from this analysis today is that the 3P revolution is not solely responsible for the diminishing popularity of DMA as the media will make you think. In fact, it's much more complex because there are many hidden variables at play that sometimes can't be easily quantified. As mentioned previously, physiological, social, and skill biases can easily affect the dynamics of a game and live shot decisions. For example, if a team is down double digits in the 4th quarter, they might try chucking up threes rather than mid-range shots, hoping they might catch up.

- **Possible limitations of my findings.**

This trend, regression, and conclusion don't necessarily apply to other basketball leagues (Euroleague, college ball, WNBA, etc) and time frames. In different leagues, there are different rules, competition levels, skill, and play styles that may affect this trend. Many other hidden variables are not factored into this analysis, which could affect the decrease in DMA as well, such as a team focusing on layups (the most efficient shot based on attempts). Therefore, an increase in 3PA doesn't necessarily mean a lower DMA in all other leagues as well!

As mentioned before, there is a limitation on the number of 3PA and DMA in a full 48-minute game which the regression does not factor in.

- You cannot shoot negative 3PA and DMA and there may be a limit in the positive direction of both variables (ex: is 120 3PA realistic and even possible?).

- **Next Steps.**

Continuing to reflect on the impact of the 3PA and DMA metrics as I watch and analyze the NBA playoffs currently. I should actively look for hidden variables, impact, and possibly a 3rd variable to reflect on as I watch the playoffs unfold.

- For example: The Utah Jazz shot the most 3PA and lowest DMA this year, and they are ranked number one in the western conference. However, a hidden variable for why they do not shoot many mid-range shots may be the fact that one of their best players on the roster is Rudy Gobert, a very weak shooter who shoots exclusively from the paint. In contrast, their best player, Donovan Mitchel is a prolific finisher and three-point shooter. In essence, their offense is not very centered around the mid-range shot but rather an open 3P shot, off-ball screens, dunks, player movement, and layups.



Analyzing other variables:

- Looking at percentages (3P% or DM%) would be interesting to analyze to see if teams have gotten better at shooting. With these variables, we could find out if weak 3P shooting teams tend to settle for mid-range shots instead.

Lastly, determining if this is a global trend across basketball:

- Examine the Euro League, the Olympics, the WNBA, and College ball (despite them having a shorter 3P line. Is it more widely adopted because of this reason? Do they attempt more 3P shots because it's closer and within the range of a deep mid-range shot in the NBA?).



Works Cited

“2020-21 NBA Season Summary.” *Basketball*,
https://www.basketball-reference.com/leagues/NBA_2021.html.

Robert, Clark. “Exponential Curves in Sport.” *It's All about the Vertical*, 11 Apr. 2019,
<https://itsallaboutthevertical.wordpress.com/tag/exponential-curves-in-sport/>.

“The 3-Point Revolution.” *ShotTracker*,
<https://shottracker.com/articles/the-3-point-revolution>.

NCAA.com, Daniel Wilco. “How the New 3-Point Line Might Affect College Basketball.”
NCAA.com, NCAA.com, 8 Nov. 2019,
<https://www.ncaa.com/news/basketball-men/article/2019-10-03/how-new-3-point-line-might-affect-college-basketball>.

Paynting, Jake. “Kevin Durant and the Art of the Mid-Range.” *Medium*, Medium, 7 May 2019,
<https://medium.com/@jakepaynting/kevin-durant-and-the-art-of-the-mid-range-83334e406969>.

“The next Way of Seeing Sports.” *Second Spectrum*,
<https://www.secondspectrum.com/index.html>.

“Qualtrics.” *Qualtrics XM*, 12 Apr. 2021,
<https://www.qualtrics.com/support/stats-iq/analyses/regression-guides/interpreting-residual-plots-improve-regression/>.