**Quantifying Momentum in an NBA Game**

Ansh Rao

Department of Mathematics, Arizona State University

Barrett Honors Thesis

Professors Laurence Schneider and Daniel McIntosh

April 12, 2022

**Abstract**

In the basketball world, perhaps one of the most sought-after feelings is that of momentum. Basketball players, coaches, analysts, and fans alike are all too familiar with the idea that a “team has momentum” during a stretch of time, or that the team needs to do something to “generate their own momentum”. In a game that appears to be an accumulation of independent possessions, what exactly does momentum really mean? My goal was to see if there is a way to quantify momentum in an NBA game, particularly by looking at the Phoenix Suns 2021-2022 NBA season.

Ultimately, I wanted to look at strings of offensive possessions to see if there was evidence of momentum across a sequence of possessions. I grouped the data into sets of three shots by implementing a lag on the dataset, which give us 8 total outcomes as to how the sequences of shots can go in terms of misses and makes. Naturally, the best possible outcome in a set of 3 shots is three consecutive makes, so I defined momentum as the groupings of three shots where all three were made baskets, and the goal was to see what factors contribute to this specific sequence. I collected data from the offensive possessions in the 4th quarter from the first 35 games of the Suns season. I kept track of data related to each shot taken by a Suns player in the 4th quarter by watching the games, and these variables included both user-defined variables as well as NBA box score variables. Some variables included the player shooting it and his role on the team, the distance of the shot and how contested it was, whether the shot was assisted or not, turnovers before the shot, and stoppages after the shot, to name a few.

I needed to figure out what specific variables I would want to use for a model, so I started out by looking at the three consecutive make sequences (denoted as make make make) to see if anything was noteworthy. I found that the make make make sequences did not necessarily have high assist levels, and while open shots did comprise the majority of the shots in the sequences, it stuck out that the sequences usually involved some sort of contest over the three shot span. I decided to use a logistic regression to model a binary variable comparing make make make sequences to all others in the subset. Ultimately, the variables that proved to be significant indicators of make make make sequences were assists, contest levels, and average distance. What resulted was the following first order model: Make make make = 1/(1+e−(−0.6387+1.2228x1)) + 1/(1+e−(−0.6387−0.6184x2)) + 1/(1+e−(−0.6387−0.23x3)), where x1 is the sum of the assists, x2 is the sum of the contest levels, and x3 is the average distance of the shot, all over the three-shot sequence.

**Introduction**

At first glance, a basketball game can look like several independent possessions, unrelated to each other. But anyone who has played the game at any level understands that the game takes on a much different feel when constantly going back and forth. Now, in an NBA setting, when you add thousands of people in the stands watching the game, in addition to potentially millions watching at home, it is natural that the players get caught up in the flow of a game as opposed to the individual possessions. As such, when a player or a team feels this flow positively, it would prove useful to quantify this momentum. As I started to try and define what I could use as an indicator of offensive momentum in particular, I knew that it would be important to look at strings of possessions and how the shots accumulated offensively.

I also knew that momentum could be different when it comes to a player’s momentum versus the overall team’s momentum. For example, a single player making several shots in a row might not necessarily mean that the team feels this idea of momentum. I wanted to keep this in mind moving forward, because my goal was to identify team momentum.

I needed to pick a team that I could analyze, and the Phoenix Suns proved to be a worthy choice. Coming off a 2020-2021 season in which the team made an NBA Finals appearance, they were poised to have yet another good season as serious contenders. The Suns have two superstar-level players in Chris Paul and Devin Booker who can make difficult shots. Around these two superstars, the team has high level starters and role players such as Mikal Bridges, Deandre Ayton, and Cameron Johnson, who are also capable of making timely, difficult shots. The Suns truly embody team basketball and I chose them because I felt they would be a good case study in team momentum.

Looking back at how the Suns performed in the 2021-2022 regular season, it became abundantly clear that the team was a perfect candidate for quantifying momentum. The Suns finished this season as the top overall seed in the NBA by a wide margin, with the #2 ranked offense and the #1 clutch offense in the league. The team had an NBA-best 18 game win streak, and multiple other 10 game win streaks. Despite the various injuries the Suns faced, the depth of the roster allowed for the team to play at a high level, finishing with 64 wins and 18 losses on the year.

**Literature Review**

**The Hot Hand: A New Approach to an Old “Fallacy”**

This paper examines the idea of whether a “hot hand” exists in basketball. The hot hand existing implies that shots are not necessarily independent events, but instead that a player could be affected positively by making consecutive shots before the recorded shot attempt. There have been several studies on the hot hand, showing mixed results. What this study found interesting was that shot difficulty seemed to increase as the player felt as though he was “hot”. The shot distance would increase with heat while the defender distance would decrease. Additionally, the player seems to buy into the idea that he is “hot”, given that he is more likely to shoot. However, defining heat becomes the key in this situation – a player making 3 or 4 layups in a row is not the equivalent of a player making 3 or 4 three pointers in a row.

**Experts’ Perceptions of Autocorrelation: The Hot Hand Fallacy Among Professional Basketball Players**

This paper looked at the Los Angeles Lakers and analyzed each of the players when they were on hot streaks, and categorized them into three categories: sub-optimal performers (performance worsens after getting hot), consistent performers (not significantly affected by hot streaks), and benign responders (respond significantly but shoot 3 pointers which provide more returns). Interestingly, the star player for that Lakers teams, Kobe Bryant, was categorized as a sub-optimal performer, but a player’s role can affect this. Kobe Bryant took significantly more shots than any of his teammates and so he burdened more of the offensive responsibility, and the level of difficulty on the shots he was taking led to a decline in points. Role players might be consistent or benign performers because their shot difficulty does not change significantly.

**The Hot Hand Myth In Professional Basketball**

This paper looked at the NBA 3 point contest as the best example of being able to examine if there is a hot hand, because of the idea of shooting consecutive open shots over and over. This study found no significant evidence of a hot hand existing, and advise against the notion that coaches and players should use past performance as indicators of future performance, if it comes from an uncharacteristic hot streak by a player.

**Summary**

The literature review provided me with some context around player versus team momentum. The literature showed that this idea of hot hand exists, but it took a look at the player momentum aspect. A player’s shot difficulty may increase as he feels hot, but this may not necessarily translate to overall team momentum, especially if it is simply the same player shooting over and over. I wanted to keep this in mind, and as a result of the emphasis on the difficulty of shot, I knew that I needed to have variables that measured shot difficulty in my data collection.

**Data Collection**

**Rationale**

I was creating my own dataset based on the Phoenix Suns’ offensive possessions in the 4th quarters for the first 35 games of the season. To gather this data, I used Google Forms to input several variables on every shot attempt taken. Most of these variables required looking at each game’s log recorded on the NBA website, but there were some variables that were user-defined. Because of this, I would rewatch every game the day after and simultaneously go through the game log and the game film to record the data I needed. This process took quite a lot of time, and given I was working on this data collection alone, I would not be able to collect data for the whole game. The 4th quarter made the most sense to watch because the intensity of the game picks up, and teams will play their best available players for a majority of the quarter. Additionally, when the game is within 5 points and there are less than 5 minutes, the game is considered to be in clutch minutes. These moments are the most intense moments of a game and momentum can matter so much more in these minutes. I wanted to track several variables that I thought may be of significance to each shot taken by the Suns, and once the data was collected, I would be able to lag the observations, such that I could look at sequences of three consecutive shots. Statistics available in the box score were taken from the NBA website. I used Google Forms and submitted a form for each shot taken by a Suns player, and from there I compiled the data in an excel file. After a couple of trial runs with older games, I determined what variables I needed to record. The variables I recorded were as follows.

**Elementary Data: Game Number, Time of Shot, Opponent Record, Location**

I tracked the game number and time of each shot, both taken from the NBA game log, just as a way to index each data point in the set. I kept track of the opponent record and location of the game, because I felt as though there could exist differences in home and away games, as well as the success of the opponent the Suns faced.

**Player Data: Shooting Player and Player Role**

I tracked which player took the shot, as well as whether this player was a starter or came off the bench for the specific game. Because of all the injuries that occurred for the Suns this season, several different players had the opportunity to start for the team, so from game to game a player’s role could change. I had kept track of player data because I had hoped to analyze potential individual player momentum, but I was not able to gather a sufficient number of observations of a player shooting three consecutive shots.

**Shot Data: Shot Success, Number of Points, Difference After Shot**

I tracked if the shot was made or missed, and the number of points each shot yielded, ranging from 0 to 3 points. I also tracked what the difference in the game score was after the shot was attempted, with negative numbers indicating a deficit for the Suns and positive numbers indicating a lead.

**Shot Difficulty Data: Shot Type, Shot Distance, Assisted Shots, Contest Levels**

I categorized each shot into one of 6 different shot types: dunks, layups, free throws, short jumpers/floaters, long jumpers, and three pointers. Dunks, layups, free throws, and three pointers are defined per the NBA definition, but I distinguished between the short jumper and long jumper category. I defined a short jumper/floater as any jump shot taken within 10 feet of the rim or a jump shot taken with at least one foot in the paint, and I also included any floater in this category, in which a player “floats” the ball toward the rim as opposed to the typical jump shooting motion. Long jumpers were then defined as two point jump shots that had a shot distance of more than 10 feet and outside the paint.

I included the shot distance that the NBA game log listed for each shot, as well as if the shot was assisted or unassisted (all the misses were listed as unassisted shots).

Contest levels were not in the game log, but I felt as though it was an important factor in the difficulty of a shot. Each shot was categorized as open, lightly contested, or heavily contested. To define these, I split up a player’s shot into two phases: when they gather the ball to jump for their shot, and when they actually release the ball. I categorized a shot as open if at both the time of the gather and the release, there was no defender within arm’s length of the shooting Suns player. A shot was also considered open if a defender did not attempt to contest the shot, by not putting his hands up in front of the shooter. I categorized a shot as lightly contested if a defender was within arm’s length for either the gather or the release, but not both. Most often, lightly contested shots occurred when the shooting Suns player gathered the ball with no defender around him, but as he released the ball a defender closed out on the shot and came within arm’s length. Occasionally, a defender would be present at the time of the gather, but as the Suns shooter would release the ball, the defender would bail out on the contest attempt. Lastly, I categorized a shot as heavily contested if there was a defender present within arm’s length of a Suns shooter throughout the gather and release process.

**Non-Shot Data: Stoppages, Turnovers**

I also wanted to make sure that certain factors in between shots were taken into account.

Stoppages in the game can disrupt the flow of a team. Opposing coaches typically take timeouts when a team is playing well, in hopes of changing their strategy and disrupting the team. As a result, I wanted to keep track of whether a stoppage occurred after a recorded shot. Mostly, these were opponent timeouts, but other stoppages included official timeouts from the referees, or perhaps possible challenges coming from either team to dispute a decision made by the referees.

I also felt it was imperative to keep track of Suns turnovers, because those are losses of possession that could harm the team. I kept track of whether a turnover occurred before the recorded shot, and if so, I wanted to know how many turnovers occurred in between shots.

**Totals of Variables in Sequences**

Given that I was looking at sequences of shots, it made sense for me to create variables that summed the totals over the course of a sequence. I summed the number of points, total assist levels, and total turnovers in a sequence. I also created a numerical variable for contest levels, with 0 being open shots, 1 denoting lightly contested shots, and 2 denoting heavily contested shots, so I could sum total contest levels in a sequence. I then created a variable for the average distance of shot in a sequence, and I also looked at how the difference in the score of the game changed from shot 1 to shot 3 in a sequence.

**Methodology**

I wanted to look at strings of possessions, so I decided that I would lag the data twice over to create three shot sequences. This would create a total of 8 possible outcomes, and naturally, any sequence that consisted of three straight makes would be the best possible outcome. This would then prove to be the best indicator of what I could quantify as momentum, so I would define these sequences as the momentum sequences. From here, I needed to determine what variables I should collect for my dataset, and then figure out the best way to create a model.

After collecting the data and compiling it in an excel sheet, I was able to use SAS to lag the data so that I could look at sequences of three consecutive shots. After lagging the variables, I was able to look at different subsets of the data based on how the sequences went, from three consecutive makes to three consecutive misses, including everything in between.

From here, I used SAS to first attempt to find what variables would best fit in a model. I quickly realized that a linear model would not be the best way to model the data, but rather to make a logistic regression using a binary variable comparing the make make make sequences to all other sequences.

**Exploratory Data Analysis**

The goal of the exploratory data analysis was to find out if anything stood out regarding the make make make subset of data in particular.

**Player Momentum**

I was going to look at individual player momentum in the make make make sequence, but with only 14 data points for this occurrence, I felt as though it would not be reasonable to make any valid claims about individual momentum.

**Shot Types**

I was going to accumulate the shot types in a make make make sequence, but looking at the shot types did not provide me with any takeaways. The modern NBA is geared towards more three point shots taken, and the data tends to reflect this, but knowing this does not provide any value to the idea of momentum for the team. Given that the shot types were self-defined, I would need to compare the Suns’ shot type charts to the rest of the league’s shot types, which I did not have the capacity to do in this project.

|  |  |  |
| --- | --- | --- |
| **Sum of Assists** | **Frequency** | **Percentage** |
| 0 | 35 | 22.58% |
| 1 | 63 | 40.65% |
| 2 | 45 | 29.03% |
| 3 | 12 | 7.74% |

|  |  |  |
| --- | --- | --- |
| **Sum of Assists** | **Frequency** | **Percentage** |
| 0 | 12 | 9.09% |
| 1 | 63 | 47.73% |
| 2 | 45 | 34.09% |
| 3 | 12 | 9.09% |

**Assist Frequencies in Make Make Make Sequence**

*Table 1: Sum of Assists Including 3 Consecutive Made FTs Table 2: Sum of Assists Without 3 Consecutive Made FTs*

I wanted to look at the sum of the assists that happened in a make make make sequence, meaning the minimum would be 0 assists and the maximum would be 3 assists. Initially, I was surprised to see that 63% of the time, there was only a maximum of 1 assist that occurred in the three shots. I then felt as though sequences of three consecutive made free throws would skew the data, because it only yielded three total points in the sequence, yet accounted for three unassisted, wide open made shots. After dropping the three consecutive made free throws from the data set, I saw that 56.8% of make make make sequences had either 0 or 1 total assists in the three shots. This follows the logic of the existing literature, in that players tend to take more difficult shots when they have the “hot hand”, so they may end up taking more difficult, unassisted shots.

**Contest Levels in Make Make Make Sequence**

|  |  |  |
| --- | --- | --- |
| **Contest - Shot 1** | **Frequency** | **Percentage** |
| Open | 63 | 47.73% |
| Lightly Contested | 29 | 21.97% |
| Heavily Contested | 40 | 30.30% |
| **Contest - Shot 2** | **Frequency** | **Percentage** |
| Open | 70 | 53.03% |
| Lightly Contested | 25 | 18.94% |
| Heavily Contested | 37 | 28.03% |
| **Contest - Shot 3** | **Frequency** | **Percentage** |
| Open | 68 | 51.52% |
| Lightly Contested | 25 | 18.94% |
| Heavily Contested | 39 | 29.55% |

*Table 3: Contest Levels for Each Shot in Make Make Make Sequence*

For the contest levels, I also dropped the three consecutive made free throws from the dataset. I summed the total contest levels in the make make make sequence. I made the contest variable numerical with 0 indicating open shots, 1 indicating lightly contested shots, and 2 being heavily contested shots.

I looked at the contest levels of each made shot in the sequence separately as well as the sum of the total in the sequence. As is to be expected, the majority of shots made in the make make make sequence were open shots, but it is important to note that heavily contested shots also occurred with frequencies hovering around 30% for the first, second, and third shots in the sequence. When looking at the sum of the contest levels in a make make make sequence, it jumps out that nearly 84% of the time, there was some sort of contest that occurred.

|  |  |  |
| --- | --- | --- |
| **Sum of Turnovers** | **Frequency** | **Percentage** |
| 0 | 53 | 58.89% |
| 1 | 31 | 34.44% |
| 2 | 6 | 6.67% |
| 3 | 0 | 0.00% |

**Turnover Levels in Make Make Make Sequence**

|  |  |  |
| --- | --- | --- |
| **Sum of Turnovers** | **Frequency** | **Percentage** |
| 0 | 87 | 65.91% |
| 1 | 39 | 29.55% |
| 2 | 5 | 3.79% |
| 3 | 1 | 0.76% |

*Table 4: Sum of Turnovers in a Make Make Make Sequence Table 5: Sum of Turnovers in a Miss Miss Miss Sequence*

I looked at the sum of the turnovers in the set of three shots, but I compared the make make make sequence to the sequences of three straight misses. I found that in the make make make sequence, there were 0 turnovers roughly 66% of the time, which is about 8% higher than the rate at which 0 turnovers occurred during sequences of 3 missed shots. Turnovers can disrupt a team’s perceived flow, so the make make make sequence being interrupted by turnovers at the lowest rate makes sense.

It became quite clear that the best way to move forward in the modeling process was to model the make make make subset. Ultimately, I decided the best way to model the data would be a logistic regression comparing the make make make subset to all the outcomes where make make make was not the result, and I would look at what proved to be significant indicators of the three make sequence. The EDA also helps contextualize the final model, because it helps characterize the Suns as a team. The Suns showed that they have the ability to make contested and unassisted shots even in three make sequences, and they showed that they tend to keep their turnovers down in make make make sequences.

**Results**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Sum of Points | Sum of Assists | Contest Levels | Shot Types | Sum of Turnovers | Average Distance | Difference in Game |
| Sum of Points | 1 | 0.58604 | 0.43683 | 0.68233 | 0.05522 | 0.31774 | -0.14951 |
| Sum of Assists | 0.58604 | 1 | 0.23403 | 0.09064 | 0.02817 | -0.1852 | -0.10048 |
| Contest Levels | 0.43683 | 0.23403 | 1 | 0.20936 | -0.0006 | -0.19036 | -0.07784 |
| Shot Types | 0.68233 | 0.09064 | 0.20936 | 1 | -0.02868 | 0.78703 | -0.0655 |
| Sum of Turnovers | 0.05522 | 0.02817 | -0.0006 | -0.02868 | 1 | -0.01363 | 0.20492 |
| Average Distance | 0.31774 | -0.1852 | -0.19036 | 0.78703 | -0.01363 | 1 | -0.00409 |
| Difference in Game | -0.14951 | -0.10048 | -0.07784 | -0.0655 | 0.20492 | -0.00409 | 1 |

**Variable Selection and Filtering Through Appropriate Models**

*Table 6: Check of the Correlation Matrix*

|  |  |
| --- | --- |
| **Variable** | **Variance Inflation** |
| Intercept | 0 |
| Sum of Points | 2.18904 |
| Difference | 1.24235 |
| Sum of Assists | 1.95455 |
| Sum of Contests | 1.09716 |
| Dunk | 1.3122 |
| Layup | 1.91699 |
| Free Throw | 2.17216 |
| Short Jumper | 1.75566 |
| Three Pointer | 2.29739 |
| Sum of Turnovers | 1.05065 |
| Average Distance | 1.01995 |

*Table 7: Check of Variance Inflation Statistic for Multicollinearity*

The first thing I wanted to do was check for the correlations before I attempted to look at significant variables in a model. I looked at the sum of the points in a three shot sequence, how much the difference in the game changed over the sequence, the total assists in the sequence, the sum of the contest levels, each shot type, the total turnovers in the sequence, and the average distance of shots taken in the sequence. For the shot type in particular, I needed to choose one shot to be the base shot type and then compare all others to that. Given the hot hand literature, I selected the long jumper as the base shot type, because I expect a higher difficulty of shots as makes increase, and this shot is often the most difficult in an NBA game.

I looked at the variance inflation statistic to see if there was one independent variable that was affecting the others. While all of the Vif statistics looked fine, I did notice that including the total points might be too repetitive. The sum of the points is dependent on the other variables, because more makes goes hand-in-hand with more points. Additionally, including the sum of the points would just make the model favor three point shooting, because of the points per shot a three point make generates relative to other shots.

Initially, I wanted to try and use a linear model within the make make make subset to attempt and quantify momentum. In this attempt, I used the sum of the points as my dependent variable and attempted to find significant predictors of this outcome. I looked at the correlation matrix and tried to come up with a model for the three make sequence based on highest overall correlations in a model. However, given the issue of points, and seeing how including the sum of points was not a good indicator of momentum, I decided to expand the data I was looking at, not just to the make make make subset, but comparing the make make make sequences to all other sequences, using a binary variable in a logistic model. This allowed for me to use the make make make sequences as their own variable, which more accurately depicted the momentum as I had defined.

I then wanted to see if there existed any interactions between independent variables and if there was evidence of higher order correlations, but no significant 2nd order terms or interactions existed. As such, a first order logistic model made the most sense.

**The Final Model**

The variable I modeled was a binary variable comparing sequences ending in a make make make to all other sequences. The final model with the long jumper variable as the base shot type yields a first order model with assists, contest levels, and average distance significant. This model shows that higher assist levels, lower contest levels, and closer shots are significant indicators of sequences of three makes. The model took the following form: Make make make = 1/(1+e−(−0.6387+1.2228x1)) + 1/(1+e−(−0.6387−0.6184x2)) + 1/(1+e−(−0.6387−0.23x3)), where x1 is the sum of the assists, x2 is the sum of the contest levels, and x3 is the average distance of the shot, all over the three-shot sequence.. The area under the curve for this logistic model is 0.8080, making it a strong model for the make make make variable. The baseline area under the curve is around 0.7, so anything exceeding this can be considered a useable model.

**Conclusion**

In these 35 games, the Suns went 27-8 with the 2nd best offense in the NBA. They had the #1 clutch offense in the NBA as well, meaning that when the game was within 5 points with 5 minutes or less left in the game, nobody executed better than the Suns on the offensive end. They finished the season as the best team in the NBA by far, despite players constantly rotating in and out of the lineup due to injury. From the exploratory data analysis, I know that the team can make unassisted and contested shots. Devin Booker and Chris Paul are two superstars capable of hitting these shots, and even high level starters like Mikal Bridges and Deandre Ayton have shown their ability to make shots at a high level. The team’s depth and coaching allows for the Suns to continue to operate their offense even when bench players have to play larger roles. These characteristics define the team, and this contextualizes the model that more assists, lower contest levels, and shorter average distance predict make make make sequences in the 4th quarter for the team. For a team that does not have the roster makeup of the Suns, a model for their momentum might look different.

When looking at the rest of the season even after the 35 games that I modeled for, the team continued on to be 37-10, while operating in the same manner as the first half of the season. Even with Chris Paul missing time, the Suns maintained their position as the best team in the NBA with the #1 clutch offense, so it seems reasonable to use this model for the team for this season.

When it comes to the use of this model, the Suns can use this to try and generate assisted, shorter, and less contested shots in their offense. One play that is common in an NBA setting is an isolation, in which one player attempts to score on another player without any screen set for them or any assistance from his teammates. The Suns would be better advised to stay away from these isolation situations, and rather continue to move the ball and run plays for closer and more open shots. In this way, it would make sense for the Suns team to run sets in which the primary option generated is an open, closer shot, as opposed to a difficult isolation. The Suns could generate team momentum in this way. Obviously, the Suns’ opponents will try to stop them from getting the looks they want, but the Suns’ ability to counter with shots that are still assisted or open will result in a higher likelihood of the team generating momentum.

**Next Steps**

There are a few ways that I can improve on this model. With more resources and time, I could get even more detailed with the variables I measured offensively, going beyond just box score statistics and perhaps even looking at advanced statistics. I think the biggest thing to recognize in this model and where I could incorporate more data is the fact that I completely ignored the defensive end of the court, where the Suns also had the #2 defense in the NBA. The Suns were the only team in the NBA with a top 5 offense and top 5 defense, so the Suns’ confidence on the other end may play a factor in quantifying momentum. Simply put, if a team cannot get a stop on one end of the floor, the number of makes they get on the other side simply may not matter. Generalizing a model across the league would require analysis of other teams, and I would be curious to see what models for other good teams look like in particular.

I am also curious to see how player momentum is different from team momentum. I would hypothesize that the shot type bears much higher significance in an individual’s momentum, as the literature described. A player may not feel momentum simply because he has made three consecutive layups, but if the overall team generates three consecutive layups with multiple people involved, they may feel a stronger sense of team momentum.

**Appendix**

**Google Form for Data Collection**

A picture containing table

Description automatically generated

**Code for Creating Variables for Dataset and Lagging Variables**

data shot2;

set shot;

if GameNumber\_1<>GameNumber\_2<>GameNumber\_3 then delete;

if Location = 'Home' then Loc = 1;

else Loc = 0;

if Role = 'Starter' then Rol = 1;

else Rol = 0;

if Shot\_Success = 'Make' then Shot\_Succ = 1;

else Shot\_Succ = 0;

if Assist = 'Assisted' then Assi=1;

else Assi = 0;

if Contest = 'Open' then Cont = 0;

else if Contest = 'Lightly Contested' then Cont = 1;

else Cont = 2;

if Shot\_Type = 'Long Jumper' then Type = 1;

if Shot\_Type = 'Dunk' then TypeA = 1;

else TypeA=0;

if Shot\_Type = 'Layup' then TypeB = 1;

else TypeB=0;

if Shot\_Type = 'Free Throw' then TypeC=1;

else TypeC=0;

if Shot\_Type = 'Short Jumper' then TypeD=1;

else TypeD=0;

if Shot\_Type = 'Three Pointer' then TypeE = 1;

else TypeE=0;

if Stoppage = 'No' then Stop = 1;

else Stop = 0;

if Turnovers\_before\_shot = 'No' then Turn = 1;

else Turn = 0;

GameNumber\_1=lag1(Game\_Number);

GameNumber\_2=lag2(Game\_Number);

GameNumber\_3=lag3(Game\_Number);

Succ\_1=lag(Shot\_Succ);

Succ\_2=lag2(Shot\_Succ);

Succ\_3=lag3(Shot\_Succ);

Wins\_1=lag1(Opponent\_Wins);

Wins\_2=lag2(Opponent\_Wins);

Wins\_3=lag3(Opponent\_Wins);

Losses\_1=lag1(Opponent\_Losses);

Losses\_2=lag2(Opponent\_Losses);

Losses\_3=lag3(Opponent\_Losses);

Loc\_1=lag1(Loc);

Loc\_2=lag2(Loc);

Loc\_3=lag3(Loc);

Time\_1=lag1(Time\_of\_Shot);

Time\_2=lag2(Time\_of\_Shot);

Time\_3=lag3(Time\_of\_Shot);

Player\_1=lag1(Player);

Player\_2=lag2(Player);

Player\_3=lag3(Player);

Rol\_1=lag1(Rol);

Rol\_2=lag2(Rol);

Rol\_3=lag3(Rol);

Points\_1=lag1(Points);

Points\_2=lag2(Points);

Points\_3=lag3(Points);

SumPoints=Points\_1+Points\_2+Points\_3;

Distance\_1=lag1(Shot\_Distance\_\_ft\_);

Distance\_2=lag2(Shot\_Distance\_\_ft\_);

Distance\_3=lag3(Shot\_Distance\_\_ft\_);

avgDistance = (Distance\_1 + Distance\_2 + Distance\_3)/3;

if avgDistance <=5 THEN avgDist=0;

else if avgDistance>5 AND avgDistance<=10 THEN avgDist = 1;

else if avgDistance>10 AND avgDistance<=15 THEN avgDist = 2;

else if avgDistance>15 AND avgDistance<=20 THEN avgDist = 3;

else if avgDistance>20 AND avgDistance<=25 THEN avgDist = 4;

else avgDist = 5;

Assi\_1=lag1(Assi);

Assi\_2=lag2(Assi);

Assi\_3=lag3(Assi);

SumAssist=Assi\_1+Assi\_2+Assi\_3;

Cont\_1=lag1(Cont);

Cont\_2=lag2(Cont);

Cont\_3=lag3(Cont);

SumContest=Cont\_1+Cont\_2+Cont\_3;

Type\_1 = lag1(Type);

Type\_2 = lag2(Type);

Type\_3 = lag3(Type);

sumType=Type\_1+Type\_2+Type\_3;

Difference\_1=lag1(Difference);

Difference\_2=lag2(Difference);

Difference\_3=lag3(Difference);

Diff=Difference\_3 - Difference\_1;

Stop\_1=lag1(Stop);

Stop\_2=lag2(Stop);

Stop\_3=lag3(Stop);

Turn\_1=lag1(Turn);

Turn\_2=lag2(Turn);

Turn\_3=lag3(Turn);

NumberTOs\_1=lag1(Number\_of\_TOs\_Before\_Shot);

NumberTOs\_2=lag2(Number\_of\_TOs\_Before\_Shot);

NumberTOs\_3=lag3(Number\_of\_TOs\_Before\_Shot);

SumTurns=NumberTOs\_1+NumberTOs\_2+NumberTOs\_3;

if Succ\_1 = 1 AND Succ\_2 = 1 AND Succ\_3 = 1

then a='MaMaMa';

else if Succ\_1 = 1 AND Succ\_2 = 1 AND Succ\_3 = 0

then a='MaMaMi';

else if Succ\_1 = 1 AND Succ\_2 = 0 AND Succ\_3 = 1

then a='MaMiMa';

else if Succ\_1 = 1 AND Succ\_2 = 0 AND Succ\_3 = 0

then a='MaMiMi';

else if Succ\_1 = 0 AND Succ\_2 = 0 AND Succ\_3 = 0

then a='MiMiMi';

else if Succ\_1 = 0 AND Succ\_2=0 AND Succ\_3 = 1

then a='MiMiMa';

else if Succ\_1 = 0 AND Succ\_2 = 1 AND Succ\_3 = 0

then a='MiMaMi';

else if Succ\_1 = 0 AND Succ\_2 = 1 AND Succ\_3 = 1

then a='MiMaMa';

**Code for EDA (Tables 1-5)**

data shot3;

set shot2;

where a='MaMaMa';

proc freq data = shot3;

tables Shot\_Succ Loc Rol Cont Stop Turn Type SumPoints SumAssist SumTurns;

run;

data shot4;

set shot2;

where a='MiMiMi';

proc freq data = shot4;

tables Shot\_Succ Loc Rol Cont Stop Turn Type SumPoints SumAssist SumTurns;

run;

**Code to Create Binary Variable and Drop Three Consecutive FTs**

data shot3;

set shot2;

if a='MaMaMa' then MaMaMa = 1;

else MaMaMa=0;

if Type\_1 = 2 AND Type\_2 = 2 AND Type\_3 = 2 THEN delete;

run;

**Correlation Matrix Coding (Table 6)**

proc corr data = shot3;

var MaMaMa sumPoints Diff sumAssist sumContest sumType sumTurns avgDist;

run;

**Multicollinearity Check Coding (Table 7)**

proc reg data = shot3;

model MaMaMa =sumPoints Diff sumAssist sumContest Type TypeA TypeB TypeC TypeD TypeE sumType sumTurns avgDist/vif;

run;

**Code to Model Logistic Regression**

Proc Logistic data=shot13 plots(only)=(roc influence orplot) descending;

model MaMaMa = Diff sumAssist sumContest TypeA TypeB TypeC TypeD TypeE sumTurns avgDist AsCoDist;

run;

**Code and Output for Final Model**

Proc Logistic data=shot13 plots(only)=(roc influence orplot) descending;

model MaMaMa = sumAssist sumContest avgDist;

run;

Table

Description automatically generated Table

Description automatically generated

Chart, line chart

Description automatically generated

**References**

*The hot hand: A new approach to an old fallacy”* . (n.d.). Retrieved February 2, 2022, from https://global-uploads.webflow.com/5f1af76ed86d6771ad48324b/5fee084d4df428821899d2fc\_2014\_SSAC\_The-Hot-Hand-A-New-Approach.pdf

Justin, R. M. (n.d.). *Experts' perceptions of autocorrelation: The hot hand ...* Retrieved February 2, 2022, from https://www.researchgate.net/publication/228781130\_Experts'\_Perceptions\_of\_Autocorrelation\_The\_Hot\_Hand\_Fallacy\_Among\_Professional\_Basketball\_Players

Koehler, J. J., & Conley, C. (2009, September 7). *The 'hot hand' myth in professional basketball*. SSRN. Retrieved February 2, 2022, from https://deliverypdf.ssrn.com/delivery.php?ID=028106081081004099102000002003022103027078085060074049089003076112004079126105097074045057059045033007034068081031028080005106013055005029015121119114017115086074123077083050118077081118067006119075091077074030121004103097067079029094093088007098031025&EXT=pdf&INDEX=TRUE

“Official NBA Stats.” *NBA Stats*, https://www.nba.com/stats/.