**Final Project: LeBron James 2022-2023 Season**

**Data Analytics Engineering Program, George Mason University**

**STAT 515: Applied Statistics and Visualization for Analytics**

**Dr. Fadahunsi**

**May 12th, 2023**

**Introduction and The Data Set**

The data set chosen is LeBron James game logs for the 2022-2023 regular season. Although it is a rather small and simple data set, I wanted to look further into how his statistics may impact winning and losing. It was picked because I am huge LeBron James fan, basketball is a passion of mine, and I coach basketball part time. I wanted to learn how to start analyzing basketball statistics in hope to one day create some of my own advanced statistics and analytical website in relation to basketball statistics.

To reiterate, the data is of LeBron James game log for the 2022-2023 regular season. The data set is of all the normal statistics in a basketball game. These variables include; field goals (FG), field goals attempted (FGA), field goal percentage (FGPercent), 3 point field goals (ThreeP), 3 point field goals attempted (ThreePA), 3 point field goal percentage (ThreePPercent), free throws (FT), free throws attempted (FTA), free throw percentage (FTPercent), offensive rebounds (ORB), defensive rebounds (DRB), total rebounds (TRB), assists (AST), steals (STL), blocks (BLK), turnovers (TOV), personal fouls (PF), points (PTS), game score (GmSc), and plus minus (PlusMinus). The data was found on basketball reference, which is a website that has all the data for any NBA game, player, or team. The website allows you to pull csv files in order to further use the data. There was not much pre-processing that was needed other than deleting the rows in which LeBron did not play that game. As stated, I am analyzing how LeBron impacts winning based on more normal basketball statistics and a game in which he did not play does not matter in this case. In addition, I will look at how different aspects of scoring a point correlate to how many points he may score in a given game.

The main statistic above that shows how a player contributes to winning is the players plus minus during any given game. There is important context to each player plus minus, such as who was on the court for each team at the time, but it is a general indicator of how well the team does what that player is on the court. To begin, I ran linear regression models for some statistics against the plus minus to see which ones correlated more than others. The variables I considered were as follows: field goals (FG), field goals attempted (FGA), field goal percentage (FGPercent), 3 point field goals (3P), 3 point field goals attempted (3PA), 3 point field goal percentage (3PPercent), free throws (FT), free throws attempted (FTA), free throw percentage (FTPercent), offensive rebounds (ORB), defensive rebounds (DRB), total rebounds (TRB), assists (AST), steals (STL), blocks (BLK), turnovers (TOV), personal fouls (PF), and points (PTS).

In addition, I ran linear regression models for some statistics against the points scored to see how they may correlate to how many points LeBron would score in each game. The variables I compared against points were as follows: field goals (FG), field goals attempted (FGA), field goal percentage (FGPercent), 3 point field goals (3P), 3 point field goals attempted (3PA), 3 point field goal percentage (3PPercent), free throws (FT), free throws attempted (FTA), and free throw percentage (FTPercent). These variables were chosen because they all relate to scoring points. One would think the more field goals you attempt per game the more points you would score compared to a game where you attempted less field goals.

**Data Exploration Story**

First, I began by looking at two time series graphs of the two variables that I would be comparing the other variables two; points and plus/minus. Here the time series graph of how many points LeBron scored in each game over the course of the 2022-2023 season:

**A picture containing text, screenshot, line, plot

Description automatically generated**Here is the time series graph of how what LeBron’s plus minus was over the course of the 2022-2023 season:

**A picture containing plot, diagram, line, text

Description automatically generated**As you can see, there is no real pattern or trend in either of the time series graphs. If you knew how the Lakers played this year, it would explain why his plus/minus was so low at the beginning of the season and high at the end, but without context there is not much information to pull from the above graphs. You could infer that he was having a more consistent stretch of playing good basketball throughout the middle of the season.

Next, I began comparing all of the available statistics to the plus/minus to see if I could find a correlation between the two. There was no statistical significance found between plus/minus and any other of the statistics available. I found that strange because one would think the more points you score as an individual, would correlate to your plus/minus being higher. The plus/minus statistic does not equate to winning, although it does show how you impact winning. Even if your team lost by 20 points and you scored 35 points, you might think that your plus/minus was in the positive whereas it was in the negative while you were off the court. Since I referenced points versus plus/minus, here is the graph:

A picture containing screenshot, line, text, plot

Description automatically generated

As you can see, the scatterplot is random and there is no true trend or pattern.

Finally, I compared a few statistics LeBron’s points scored each game to see if I could determine which statistic mattered the most when seeing how many points Lebron scored. The only graph with a really strong correlation, which was no surprise, was fields goals versus points, shown below:

A picture containing line, screenshot, plot, diagram

Description automatically generated

The r-squared value for model was 0.879, signaling a strong correlation between the two.

The graphs that surprised me were points versus three pointers made and points versus free throw percentage. The points versus three pointers made r-squared value was 0.276. Although other graphs had similarly low or even lower r-squared values than the points versus three pointers made, the most surprising part about the r-squared value was that it was not higher. I would expect a player that is making more three pointers in a game would be scoring more points in that game, but in LeBron’s case, that is only partly the case. The graph is shown below:

A picture containing screenshot, text, line, plot

Description automatically generated

The points versus free throw percentage r-squared value was 0.065, which was surprising in itself. Apparently, LeBron’s free throw percentage does not correlate at all to how many points he is going to score in any given game. As someone who is a physical basketball player and goes to the free throw line often, it would make sense that these two things did correlate. It also is interesting to note that LeBron has been criticized his whole career for his poor free throw percentage, when it might not even matter.

A picture containing screenshot, line, plot, diagram

Description automatically generated

**Conclusion**

The analysis I performed on the data set did not tell a particular story or provide anything of significance, other than a few things I already knew about basketball, such as the more field goals you make the more points you score. The part of the analysis that surprised me was the story that it didn’t tell, I was expecting much more correlation between some of the variables. I think choosing different seasons of LeBron would’ve told a different story.

**References**

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). “Welcome to the tidyverse.” \_Journal of Open Source Software\_, \*4\*(43), 1686. doi:10.21105/joss.01686 <https://doi.org/10.21105/joss.01686>.

H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

Wickham H, François R, Henry L, Müller K, Vaughan D (2023). \_dplyr: A Grammar of Data Manipulation\_. R package version 1.1.1, <https://CRAN.R-project.org/package=dplyr>.

*Lebron James 2022-23 game log*. Basketball Reference. (n.d.). https://www.basketball-reference.com/players/j/jamesle01/gamelog/2023