 A 250-500-word paper summarizing the following: Statistical/Hypothetical Question

* Outcome of your EDA
* What do you feel was missed during the analysis?
* Were there any variables you felt could have helped in the analysis?
* Were there any assumptions made you felt were incorrect?
* What challenges did you face, what did you not fully understand?

In the NBA, there is no question that the team which scores the most points in a game wins, an age-old answer to Sports. In this analysis we will be observing how the PPG (Points Per Game) relates to other statistics within an NBA game. To take a closer look at these variables we used an array of Python libraries to observe the relationship between multiple variables. Looking at the distribution of PPG we see that the distribution is skewed mostly to the left with a mean of 9.38 and a max of 32.9 PPG. Looking at this distribution is important, as players who have higher scoring totals are more valuable as they can contribute to higher chances of winning for their team. Looking at a probability mass function of different positions and scoring averages, we did not find any significant differences visually. This indicates that the PPG stat is not dependent on the position of the player, but I do feel like I made this assumption and could have found a better way to display this information outside of a probability mass function comparing positions of players. Using a Cumulative distribution function, I was able to verify that it is rarer for a player to score more points than less, this verifies that higher scorers are more valuable. I compared all my variables using a scatter plot matrix and found that shooting percentages are weakly correlated with the points per game statistic, this omitted these variables from my final regression. Using a Multivariate Regression Formula, I narrowed my PPG as my dependent variable and minutes played, Field Goal Attempts, and Free throw Attempts as my independent variables. Running a regression analysis, I found that the more minutes a player plays, and the more shots they get up, the higher their PPG stat goes up. So, players who stay in the game longer and have more shot attempts regardless of shooting percentage or position will score more points than players who do not have the same opportunities, this may provide a bias as to which players are better shooters if looking at the PPG category itself and may be misleading in scenarios when deciding on what players to draft or sign to a team.

While I did get an answer out of my final analysis, there is more that I wish I would have done given the information once I have finished my project. I would like to push the next step of my analysis on finding the highest impact shooters which would consider the shooting percentages and efficiency of players across the board along with their scoring averages. This would be the next step in my EDA and would have helped provide a clearer picture on narrowing down better players. I did face challenges in regards to providing some of the analysis provided within this project such as the Cumulative Distribution Function as I thought there could be better ways to establish the same thoughts, but overall It was helpful to relearn these topics from earlier in the term.