Final Project

Write 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 you felt were incorrect?
- What challenges did you face, what did you not fully understand?

For this project I decided to look at how minutes played in an NBA game affects the performance of a player. As a result, my statistical question is "how much of an impact does minutes played have on points scored?" The primary independent variable in the dataset was Minutes Played (MIN).

The variables I chose to represent performance were:

- Field Goals Made
- Field Goals Attempted
- Field Goal Percentage
- Points

In this list, 'Points' is arguably the most important variable as it is a direct indication of performance.

I do believe that the EDA was successful. It showed me how important each variable was to the overall question. Through the CDF of the FG%, the PMF of the FGM/FGA, the scatterplots of 'MIN x PTS' and 'PTS x FGM' and their corresponding Pearson correlations, I was able to understand the relationship between the variables beyond a superficial level. Furthermore, the final regression analysis literally answered my statistical question. The R-Squared of 0.731 meant that 73.1% of the variability in points scored is explained by the number of minutes played.

I don't feel like anything was missed during the analysis nor do I feel like there was a missing variable that could have helped even more in the analysis.

Also, all the assumptions I had were correct. As a matter of fact, my statistical question came about not because there wasn't something that I did not know but because there was something that I knew to be correct but wanted to confirm. So, I wasn't really surprised by any of the analysis.

This project really helped me understand the different chapters we covered in the book. The section that I really struggled on, however, was the analytical distribution section. This project made me realize that I didn't understand any of the modeling distributions. That is clearly a topic that moving forward I really need to work on understanding.