JP Meyer – kxn4mn DS 2002

Beginning this assignment, I immediately recognized a key factor in data analysis: it is very important to inspect the raw data file before working with the data. As I went to calculate the total points scored, Colab did not recognize the “PTS” column as existing. What I did not realize until I took a look at the CSV file was that the top row was making a few distinctions as to what types of stats were being recorded (“Shooting,” and “Advanced”). By dropping this row, the column names accurately reflected which statistics were being recorded and allowed for the data to be manipulated correctly. If I had not inspected the raw data sheet, I may have been left confused and with limited capabilities throughout the assignment. I also learned the skill of creating a new dataframe from the existing dataframe. Doing so allows for more careful examination of the things that “matter” (in this situation, we may only want to look at players who got real playing time for their teams). In the future, I could use this skill to filter out non-important rows of data that may skew my findings.

The biggest challenge I encountered during this assignment was finding the correct pandas commands to help me filter the data and perform analyses. As someone new to pandas, it was unclear as to what the capabilities were and the language that it uses. For example, I did not have a hard time finding the max assists in the dataset, but had to do some digging to find out how to display which player aligned with that number of minutes. This information was not intuitive because coding languages all have different languages, formats, and rules. While these may become automatic after some practice, just as learning a spoken language requires research, practice, and understanding, so do coding languages. I spent some time searching the internet and forums to investigate how to make the distinction that was causing me trouble. Artificial intelligence can also be very helpful in these situations. While I prefer to investigate for myself and try to make educated guesses for some time before resorting to AI, it can be helpful to find the correct language and format for a specific question using Colab’s built-in AI capabilities. In the future, I will continue to try to solve the problem myself, using coding skills that I understand and that we have practiced to solve problems that arise.

Using data analysis techniques on datasets such as the ACC basketball statistics can provide high-level insights that are not easily calculated by hand. The school-based analysis is a great example as to how data analysis can be helpful. Given data on only players and their statistics, it would take a long time to add up the total stats for each player from each school, but data science makes it very easy to write a few lines of code and sort by schools, providing the totals and insights within seconds. This can be readily applied in other places across industries. Take, for example, government census data. This data comes from individuals and households across the country. It is used to identify important information about the population, but contains millions of data points. This data would be essentially useless if we could not perform quick data analysis with it, like grouping by location or personal demographics. By doing so, we can reach thousands of different conclusions about our population that would be impossible to understand otherwise.