Basketball Homework Reflection
Shaveen Saadee

What did you learn about working with CSV files and pandas DataFrames in this assignment?
- Reflect on specific skills or techniques you developed, such as reading, filtering, or manipulating data, and how these skills can be applied to future data analysis projects.

In terms of working with CSV files and pandas DataFrames, I had to manipulate the data so as to properly answer the questions. For example, converting the string numbers to numeric values. Although I chose not to, I could have also renamed the columns from "Unnamed: 1," "Unnamed: 2," and so on to their intended column names. This would have made analyzing the data and coding much more efficient. Because of this, I learned well how to use .to_numeric. Reading the data was another skill I had to develop, given that I'm not too knowledgeable with basketball terms and needed to familiarize myself with the abbreviations used at the top of each column (which I believe were the intended column names). These skills can be applied to future data analysis projects, especially in new work environments where one isn't familiar with the data yet. It's important to take time to understand the data and what it is typing to represent. In this instance, a codebook would have probably been helpful, and in future projects I would reference it to ensure I'm analyzing the correct data. Additionally, the data you receive isn't always gonna be clean, so knowing how to manipulate the data in a way that is more understandable and digestible for a wider audience is an important skill.

What was the most challenging aspect of this assignment, and how did you overcome it?
- Explain the difficulties you encountered, whether it was related to coding, understanding the data, or performing specific analyses, and describe the strategies you used to work through the challenges.

I would say the most challenging aspect of this assignment was conducting the calculations. For the first question that asked for the total points scored by all players combined, I was initially receiving a list of numbers as opposed to the sum of those numbers. For example, Amaree Abram scored 34 points, Sola Adebisi scored 2 points, Prince Aligbe scored 164 points, and Abe Atiyeh scored 3 points. Rather than adding those numbers up, I was receiving a list: PTS3421643 and so on. I realized this might have been because the numbers were actually strings, so I needed to learn how to convert those strings into numeric values using .to_numeric. I ran into this issue several times throughout the assignment, until I decided to settle on converting each column to numeric values when prompted to do a calculation for that column. Doing this prevented further errors from popping up. Additionally, I had a hard time understanding the data, and had the assignment not included the abbreviations, I likely would have been more lost than I was. It was also difficult going back and forth from the table to my code, because all of the column names were "Unnamed: #," rather than the actual name itself. For this, I opened the dataset itself rather than relying on the 5 rows I displayed at the beginning.

How do you think the insights gained from analyzing ACC basketball statistics could be applied to other real-world datasets?

- Consider the broader implications of using data analysis techniques on different datasets, whether in sports, business, healthcare, or other fields, and discuss how this experience prepares you for future data science challenges.

Having data analysis techniques is a necessary skill for analyzing different kinds of datasets, regardless of the field for which you are analyzing data. In this assignment, we were tasked with analyzing player and school performance data, which can be applicable to other industries like healthcare. Similar to data analyzed for the ACC basketball statistics, in healthcare it would be valuable information to know the number of successful or unsuccessful operations a surgeon has performed, patient outcomes for a given treatment or surgical procedure, clinical trial outcomes, number of deaths per hospital, and many others. This provides valuable information in terms of helping patients decide where to seek treatment and from who, or for physicians to understand the statistics of a specific condition. Additionally, if a player were to change their technique, and that was reflected on their performance, we could see the impact that change of technique had on the player's outcomes. This could be similar to public health, in that if a policy were adopted, we could determine the effectiveness of impact of that policy on its intended issue. Overall, this assignment helps prepare for future data science challenges by allowing me to test/code certain questions, gain insight on players' performances, and understanding how such analyzation techniques are valuable for future datasets, no matter what field that dataset may come from.