

In this assignment, I learned how to effectively work with CSV files and pandas DataFrames, which are powerful tools for data analysis. By loading the ACC basketball dataset into a DataFrame, I became comfortable reading structured data from my cloud drive, cleaning it when necessary, and performing a wide range of analyses (e.g. summing, sorting, and grouping data). Additionally, I practiced skills like filtering data to focus on specific subsets (e.g., players who played more than 500 minutes) and calculating aggregate statistics (e.g., total points by school). These techniques are highly transferable to future data analysis projects, as they form the foundation of manipulating and understanding large datasets in various formats.

The most challenging aspect of this assignment was dealing with unfamiliar data structures and ensuring that my code was accurate and efficient. One particular difficulty arose when I needed to calculate aggregate statistics for grouped data, such as total points and assists by school. Initially, I struggled with how to correctly combine the **groupby()** function with other operations like **sum()** and **sort_values()**. However, by reviewing Thursday's lab notebook trial and error, I learned how each function operated and gained a deeper understanding of how to manipulate DataFrames to obtain the desired results.

The insights gained from analyzing ACC basketball statistics can be applied to other real-world datasets in numerous fields, such as sports analytics, business, and healthcare. For example, in public policy (my future career), I could group data by categories such as school districts, income populations, neighborhoods, etc. and calculate aggregate statistics to identify disparities and target policy work that minimizes them. Additionally, the ability to visualize data, such as creating scatter plots to explore correlations, is equally important in any data-driven environment where decision-making is based on data insights. This experience prepares me for future data science challenges by equipping me with the technical skills to explore and interpret complex datasets.