

1. What did you learn about working with CSV files and pandas DataFrames in this assignment?

In this assignment, I gained a deeper understanding of how to work effectively with CSV files and Pandas DataFrames. I learned that CSV files are a versatile and commonly used data storage and transfer format, making them a practical choice for data analysis projects. Using pandas, I discovered how to read CSV files into DataFrames, which serve as a powerful structure for data manipulation.

Through this process, I became proficient in several key techniques, such as filtering data, handling missing values, and aggregating results. I learned how to apply methods like `df.head()` and other DataFrame manipulations to derive meaningful insights from raw data. These skills are crucial in future data analysis projects, as they allow me to efficiently clean, analyze, and visualize datasets, ultimately leading to more informed decision-making based on data.

2. What was the most challenging aspect of this assignment, and how did you overcome it?

The most challenging aspect of this assignment was dealing with unexpected data formatting issues, particularly leading and trailing spaces in column names and inconsistencies in data types. Initially, when I tried to access specific columns like 'PTS' and 'FG%', I encountered `KeyErrors`, which were frustrating to figure out how to solve. To overcome these challenges, I utilized pandas functions like `df.columns.str.strip()` to clean up the column names, removing any unnecessary spaces. This helped me access the data more reliably. I also was having trouble helping the code understand that the column names started on the second line. I was having error messages pop up because I didn't clarify that "header=1".

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

The insights gained from analyzing ACC basketball statistics can be widely applied to various real-world datasets across different fields. For instance, the techniques used in this assignment—such as filtering data based on certain criteria, calculating aggregates, and visualizing results—are universally applicable in domains like business, healthcare, and social sciences. In business, similar analysis techniques can help in understanding sales data, customer behavior, and market trends, ultimately guiding strategic decisions. In healthcare, analyzing patient data could reveal patterns in treatment outcomes or resource allocation, leading to improved patient care. I am currently applying to healthcare data analytics positions in the job market and am constantly asked how I can visualize data results and filter data effectively. Finally, the skills I developed in this assignment will enable me to tackle larger and more complex datasets, empowering me to extract actionable insights that can drive progress in various domains. In my research lab, we are asked to go through massive data sets and I feel like this is a good stepping stone to be able to understand how to handle big data.