Emily Zhou

Professor Morrison

DS 2002

22 Sep 2024

Reflection

During this assignment when working with Pandas and CSV files I found myself looking back upon previous assignments and reflecting upon how there were many ways to interact with a dataframe and how to manipulate one to your needs, especially in reference to the recent in-class assignment. The libraries themselves are already extremely efficient and being familiar with the methods or at least having some exposure made this assignment much more straightforward than other ones have been. Now that we have been given this tool and are able to interact with one of the more common data types, CSV files, interacting with data is easier due to this unified sort of system and how versatile our experience with dataframes is now.

Personally, the most difficult part of this assignment was just reapplying the methods we had learned before since I couldn't remember them off the top of my head, especially the syntax of python and the libraries, yet after some research and familiarizing myself further with different methods such as nlargest when looking for the top n of something, I realized that there is a plethora of tools packed into the pandas library. Also, I had pulled out my ML knowledge and notes in order to complete the bonus coefficient portion and that was also a bit difficult. I had to recall how the scatterplots should have been formatted and referenced notes that were not from this class yet also used the matplotlib.pyplot so I was able to not only understand the relation between the two values but also reinforce my experience with the correlation coefficient.

When interacting with the ACC basketball statistics, I got the chance to practice cleaning data, retrieving, and then analyzing it. The process varied from understanding the format of the data, what the labels meant, what were key points of information or data that I should focus on, and also what the data implicated. My experience through this process was not unique to just this project as much of other real

world applications demands a thorough understanding of any data presented to you before you are able to interact with it. Once the cleaning and analysis of the structure and impact of the data is done, you are able to apply the information you have learned to something more tangible, even if it means just betting on the top scoring schools because once you've ran the data analytics you now have access to this sort of information. On a broader scale, let's say if I were a sport broadcasting channel, I would choose to focus more time or invest more attention towards higher performing schools and players and also have the statistics to back it up. These applications can be so much more varied than the ones I have listed, which is truly what I find to be the strongest suit of data science as I will proceed into more ML heavy applications which demand intensive data analysis.