Matthew Gottfried

DS 2002

23 September 2024

In completing this assignment, I learned quite a bit about manipulating data frames. First, I learned how to group data by columns through the questions that asked me to sort data by schools. Secondly, I learned about the loc command. This command can be used for label-based indexing. I found it helpful for sorting the minute restricted data frame. Another key skill I used in this assignment was converting data to integers and removing non-integer values. This was helpful throughout the entire assignment, as the data frame had the titles in the first row. By using this command, I was able to properly sort the data so that I could manipulate it. All these new skills are going to be very helpful in future projects for sorting and cleaning data.

The most challenging aspect of this assignment was figuring out that I had to use the loc function. I had created the new data frame easily; however, I had issues in slicing data from it. Eventually I asked one of my peers for help. He knew which function was needed and he helped me understand why it was needed. I think that he will be a good resource for me in the future.

I think that these filtering techniques can be helpful in sorting sports datasets. One application that I can think of is using those methods to sort a dataset. Last year I was trying to make a predictive model for hockey games. I was trying to sort data for line performance and goalie save percentages from different areas. I struggled to manipulate such a large dataset, but if I had the skills I do now I think I would’ve been able to better analyze the data. If I decide to create another predictive model this year, I think that I will find much more success.