What I Learned in this Course

Rachel Nelson

DSC530-T303 Data Exploration and Analysis (2211-1)

This course covered how to perform statistical analysis and exploratory data analysis using Python. Many of the techniques and scripts that were included in our textbook can be used in the real world by just attaching a new data set and fixing parameters. I liked how the course material used consistent data sets and explained the problem along with the solution.

Occasionally, the programming assignments were pretty daunting. I have never been the strongest python programmer. **One thing that I found that helped me the most was using Jupyter Notebooks.** This allowed me to work though the problem by slicing it into smaller chunks. It also helped me be successfully (most of the time) in putting together code that runs without errors. I think switching to use Jupyter Notebooks instead of trying to keep my code in a python file has helped me the most in this class.

Areas where I will need to spend more time understanding is understanding the different analytical methods and when and where to use them as well as how to interpret them. There are so many methods out there! I am comfortable with the p-value for hypothesis testing, r-squared for evaluating a regression model and Pearson's correlation coefficient for determining the strength of the correlation between two variables. I also need to spend more time increasing my personal GitHub repository so that I have quick access to different techniques.

I have always found statistics interesting. Some of the areas that I had not used prior (or just don’t remember) were things like Probability Mass Functions, Cumulative Distribution Functions, and Probability Density Functions. I was also surprised how the course pain points changed for me from at the beginning, it was writing the essay, but at the end, the essay was the easy part and the programming was painful. Because this course was so interesting, I would have greatly benefitted from an in-person class for this particular class with lectures and more examples. Maybe even supplemental text. The time it took me to complete the coursework was above what any other class has taken me in both my undergraduate as well as my graduate classes thus far. I think that has mostly to do with my programming struggles. For some reason, Python is so much more difficult for me then R, but I need to get comfortable with Python since it is so widely used.