Dylan W. Sanderson

CS-499-T4088 Computer Science Capstone 21EW4

Databases Narrative

The initial artifact that I chose for the Databases enhancement was the final project from the course DAT-220-J2600 Fundamentals of Data Mining 19EW2. It was initially completed in November 2019. The project itself was less traditional coding and more processing and interpreting data, with us utilizing Excel and various mathematical skills to seek trends in a 500 point data set.

I selected this item because it showcases how well rounded my skill set has become. Being able to provide a portfolio with only various types of code demonstrates only that, an aptitude for coding. The computer science field is much larger than that and, with data mining becoming more prevalent, I thought being able to show not only that I understood the original project, but was able to continue down the logical path to try to provide actionable data would be valuable to future employers. As such, I performed additional linear regressions to continue to seek trends, and interpreted said data in an added section at the end of the document, which has been provided.

The question as to if I met my objective from Module One is somewhat difficult to answer. One of the chief issues with this project was the limited size of the data set, as well as the means of obtaining it, which potentially let to skewed information that is not necessarily representative of the whole. However, using what was available, I ended up seeking linear regressions of a particular subset of data, which yielded fruitful information. At this point, I could continue to chase the information by attacking a different subset, namely, those who visited the website, rather than those who just made a purchase (which decreased my data set to 191 points).

I learned that continued data mining and analysis are key, and that you need to approach a problem multiple ways when seeking a solution to a complex problem. However, the biggest obstacle I ran into in this project was a knowledge gap issue. I had initially performed the analysis using Microsoft Excel, which has wonderful analytic capabilities built in, that automate a huge amount of the process for you. However, I was using WPS Office this time, and found that, while many of the capabilities were still present, it was in a more raw, underdeveloped form, that required notably more work from me to reach the same end.