A. The artifact that I chose for the Database section was the final project from my DAT-220 course, Fundamentals of Data Mining. This was created some time in 2018 and served as a tool to demonstrate how to use data analysis in order to help businesses and corporations. In order to complete the assignment, we were given a large amount of data (500 people) that we had to take, import into JMP, and run tests on to see the results and then give the company our professional opinion on what they should do to increase revenue. The only problem with this project was that everyone had to have the same data analytics – and they were the only ones that we were taught how to do. That is one of the reasons why I decided to chose this for my enhancement because I felt like there was more to do and more to learn that we were not able to in this course.

B. I decided to include this in my portfolio because I never truly felt finished with it. Since everyone used the same data mining techniques, I feel like I did not get to learn to the fullest of my ability and practice different techniques to figure out if there was a better way to connect the data together. By going back and doing these enhancements I felt like I finally finished it – I think it’s always important to put your best foot forward and show that you can go above and beyond to complete assignments. Also, another way it showcases my skills is because it shows that not only am I proficient with Excel, but I can also use JMP successfully and do a variety of data mining techniques that could come in handy if I do decide to proceed in a database career.

The artifact improvements that I decided to do was that I took data that I thought could be somewhat related and ran different data analytical techniques to see if there were any relationships and how those relationships could affect the business. There were a few issues with the data, like some of the titles of the columns did not exactly explain what that data was meant for or where it came for, or there were two sections for the zip code which could potentially be causing errors in the data from this. I did create a few new graphs that I thought would really help to show relationships between data. The first one I created was a K-Means Cluster which is similar to a Hierarchical Cluster but rather than being a vertical graph, it shows the groups as circles that overlap if they have similarities. I think that this truly helps to be able to visualize the relationships easier because the hierarchical cluster seems a bit packed and has a more reduced visibility if you don’t know what to look for – which is perfect for stakeholders and people who do not have the knowledge of a data analyst. Some of the other graphs, like the Control Chart, Distributions, and Spending Charts demonstrate the relationships between the states and the studied people’s incomes and restaurant spending. As you can see, for the most part the states and the spending are very similar between the 3 spending types, web-store, restaurant, and third. The income and states are very similar as well so it goes to say that since the income of those people is higher in those states, then their spending in those states would be higher as well. With this big of a data pool we can make generalizations off of the data and apply it on a large scale as well. With these graphs I feel like we have more options to offer the company – Do we focus on those states that have the largest spending? The spending between web, restaurant and third are all very similarly so it’s not like we can just focus making the website better or something because I do not feel like that would have the most impact.

C. The course outcomes that I had anticipated meeting during the ePortfolio Selection were: [CS-499-04] Demonstrate an ability to use well-founded and innovative techniques, skills, and tools in computing practices for the purpose of implementing computer solutions that deliver value and accomplish industry-specific goal, and [CS-499-01] Employ strategies for building collaborative environments that enable diverse audiences to support organizational decision making in the field of computer science. I feel like I have accomplished these. For CS-499-04, I have developed a multitude of graphs and diagrams that will hopefully show this specific company more information on what they should do to generate revenue, which is the industry-specific goal for this project. Next, CS-499-01 is completed because when working on a team assignment like this one would have been, had I had a team, there would have been a diverse audience – There would have been project stakeholders, a project manager, other team members, a project sponsor, many people with many different knowledge levels would have been involved and maybe could have brought something to the table that I, as the data analyst, may have not thought of or expected. But there would have been talking and even simply having to present my findings would have made it a collaborative environment because I would be sharing my feedback with them, and they would be asking what I thought would be the best route to continue on down.

D. The process of enhancing this database artifact was actually really fun. When I first took this course, I fell in love with data analytics because it is so focused on data and you get to create diagrams, organize data, display it in various formats. It’s just super technical and I loved that about it and realized that I really still do love it and it may end up being a career to consider. One of the challenges was even getting the data. It is stored in a JMP file, but unfortunately you have to pay to use JMP – I was able to fix this by getting access to the Virtual Desktop that Southern New Hampshire offers for a multitude of courses. The next issue that I had was that I did not have access to the actual course anymore which meant that I could not access the data set that we used for the final project. I had to go through the Technology Help Desk, who told me they could not help me and to speak with my advisor. When I did that, they were luckily able to get into the course and email me the dataset that I needed so that I could even begin my enhancements. I also had to change and update a lot of the data – For example, one column was just named “Restaurant”, but did that mean it was the restaurant location number or how much that person spent when they went to the restaurant in their last visit? It was all very confusing, so I updated it to be restaurant spending because that made the most sense when looking at how high some of the numbers were in similar areas.