Narrative

Data Mining Artifact

Zane Russell Brown

Southern New Hampshire University

CS-499 Computer Science Capstone

GitHub Pages: https://github.com/BeardedArtist

Data Mining

The data mining analysis artifact is a project that was created to help us learn and understand the field of data mining. We were tasked to analyze a database warehouse of a fictious company called the “Bubba Gump Shrimp Company”. The goal was to analyze the company’s data and determine why their online sales have been gradually falling while their on-site sales have not. A database warehouse, customer data and sales reports were provided and through these various databases, analysis began. Patterns and correlations between online stores, on-site stores and costumers were found and visual charts were created to help represent these correlations. Finally, a report was written to present all evidence that was discovered. This finalized report was designed to help the company see why sales had dropped and therefore help them make the necessary changes to improve their company.

The data analysis artifact was created in an eight-week period for the data mining course. This course taught us how to analyze data and make logical connections. These logical connections eventually helped us understand the data and allowed us to create visual charts to represent correlations, patterns and significant findings in the data.

Justify the Inclusion of Artifact Three

This artifact was the last one I chose for my ePortfolio. I decided on this artifact for two main reasons. The first reason was that I wanted to include an artifact that showcased my skills in another area in computer science other than programming. The first two artifacts, the Inventory App application and software reverse engineering artifact, focused on programming only. While this data mining artifact showcases my skills in data analysis and visual chart creation. This artifact

focuses mainly on data analysis through analyzing various databases and therefore focuses on the key computer science area, databases.

The data mining artifact is the strongest project I have created that deals with areas in computer science other than programming. The 40-page report that was created took the full eight weeks and many late 4am nights to complete. I am very proud of this artifact because I enjoyed the data analysis that took place to create the final report. This artifact also best represents my skills in analyzing data and thus I am including it in the ePortfolio.

Reflection

During the original analysis and creation of this data analysis report, most of the data that was analyzed was recorded and discussed. However, there were some patterns and correlations in the data that were not originally discovered. This was mainly due to a lack of time because of course time constraints. However, thanks to the extra time I had for this artifact, the additional data patterns and correlations were analyzed and recorded.

Enhancing this artifact showcased how important time is for data analysis. A lot of time and care need to be taken when dealing with a large amount of data. Using tools like JMP help with analyzing data quickly, but it is also very easy to overlook certain areas in the database. After taking the time to re-analyze the database, I found certain correlations in the data that I had not discovered previously. This was a welcome surprise because I was able to connect additional patterns together and add to my findings.

Because this artifact does not deal with any specific programming elements, a code review was not created. However, regardless of the lack of programming, enhancements did need to be made to the artifact. The artifact is made up of two main components. The finalized report and the

database warehouse that was used for analysis. I had to plan my enhancements for this artifact very differently than the previous two artifacts. In order to make sure that all patterns, correlations, and findings were covered when the project was created, I re-analyzed the database warehouse containing all the original provided data. After analysis, I was pleased to see that most correlations and patterns had been found. However, during the pairwise correlations and principal components analysis, I discovered a new pattern between customer purchase habits after they visited the company's restaurant. This newly discovered pattern was added to the report. Finally, because new data was discovered, the visual data charts had to be recreated in order to present the new findings. Once the new charts were created, the enhancements were complete.