Artifact Two Narrative

Daniel Dennington

Southern New Hampshire University

CS – 499

The enhancement to my second artifact is a Python program that creates a dataset full of dummy data of several different metrics that simulates 100,000 users I created in late January. It then writes out all of this data into a csv file. The intention of this program is to create a dataset for some simple data mining practice.

The inclusion of this artifact into my ePortfolio came after several attempts to create a Python program that could properly showcase my abilities to use data structures and algorithms in order to create a program completely from scratch. At first, I was going to have a program do both reading and writing from a file, but I couldn’t really justify the usefulness or scalability of that program.

So, instead of creating a program which operates from the terminal of a computer that incorporates user inputs to manipulate a dataset- like the original project did from my CS- 260 class. The plan is to split the original project up into separate parts, one to use data structures and algorithms to create the dummy dataset and to write that data out to a csv file. By doing so, it sets up my third and final artifact, which will mine and comb through the dataset to find ‘meaningful insights’ to mimic a simple data science project.

I honestly believe I exceeded my original expectations that I held when I first began to make the changes to the original project. For instance, I planned to mimic the original project and only make changes to the scale of the data. Instead, I not only increased its scale, but also found a way to structure the enhancement to incorporate data science into the project as well.

I also included a generalized version of an advanced data structure that is similar to a Binary Search Tree, it is known as an AVL tree. An AVL tree can be considered to be a self-balancing Binary Search Tree that has a measured ‘Balance Factor’. It achieves this balance by ensuring it maintains a low Balance Factor of less than one. It achieves this through the rotation of its nodes. The reason I chose to include this along with my python script is because I wanted to utilize the generated CSV file and parse it to eventually manipulate and store some of the data from this CSV file sometime in the near future.

In my reflection of this enhancement, I learned that I could create a computer program to just create one for the sake of it, but I feel that it’s necessary to give real purpose to the creation behind a computer program. By doing so, it gives a project a real sense of direction and a clearer goal. However, I can still utilize the knowledge from my failures when creating, such as the CSV file not having any real analyzation purposes, and reshape the scope of the project to further my knowledge of data structures and algorithms. And when it comes to challenges that I faced while developing this program, they came more in the form of structural and design issues than any programming ones, since the Python program is rather small and simply made.