Artifact Three Narrative

Daniel Dennington

CS – 499

Southern New Hampshire University

For my third and final artifact, I chose to analyze and extract data from a fairly large csv file that consists of a ton of differing metrics. Initially my plan was to perform different types of data science on my very own csv file I generated before. However, I couldn’t create any meaningful or insightful correlations with all of the random data I was generating to fill my csv file from earlier. So, I instead tracked down a free-to-use CSV file that consisted of real-world data coming out of the NBA. The dataset consists of every basketball game played from the 2004-2021 seasons.

The reason I chose to include this artifact into my ePortfolio is because it highlights my competency when handling a rather large amount of data as well as my ability to use tools like Pandas, Matplot, and more importantly Python to perform data analytics on that data. By using this artifact, I was able to comb through all the data and make numerous insights into one of my favorite sports. I was able to track down specific statistics such as the fact that around 20,000 instances have occurred within that particular time frame where less than two points were scored by someone in an NBA game. Another insight I was able to make finding out who scored the most points, who had the most steals, as well as who had the most blocks.

I would definitely say that I met the course objectives I sought out to meet. Albeit, that I only met them after I made the decision to change the CSV file from my randomly generated one, to the one NBA stats file. The entire point of this artifact was to demonstrate my ability to manipulate data within a csv file using data science and I accomplished just that.

When I first started working on this artifact, I found the syntax of the Pandas and Numpy libraries to be sort of confusing to work with at first. I soon realized that most of my frustrations were coming from me just overthinking of how to work with and manipulate particular data points. However, I just had to continue referring back to the documentations of the tools and follow some basic tutorials. After all that, I was able to work out the problems I was having with the syntaxes.

The other problem was trying to find an actual correlation in my original dataset. I had to go back to my original CSV file and try to find out what was actually happening with all of the data I created previously. I figured out that creating your own large datasets from scratch is fairly easy. However, giving my data points any purposeful correlations is a topic that isn’t in my understanding as of right now.