Elizabeth Stopa 05/07/2024

Summary of CMP-262 Final Project

The tools that were used in this project include the IDE, “Jupyter Notebook”, powered by the “Anaconda.Navigator” GUI that provides the environment for writing programs in python. Several python packages were used including pandas, and matplotlib.py. Outside sources were also used and references for a few sections of code including StackOverflow and chatGPT. The original datasets that were used for this project were taken from Kaggle and are sited in the .README file in GitHub. The questions that were asked for this project include: “What date did Aetna Inc have the highest "high" in 2016?”, “Which company had the lowest "low" in March of 2013?”, “What was the average gross profit of General Motors between 2013 and 2016?”, “What was the average total revenue of Amazon.com Inc from 2013 to 2016?”, “What was the total liability of Netflix Inc. in 2014?”. Some insights that were discovered include: even if I go through the steps of cleaning the datasets before doing analysis on them, I often find while doing the analysis that I need to do more data cleaning and need to change the structure of the datasets to answer the questions and display the answers in a better format. I also learned that it is sometimes necessary to look for help from others with alternative ways how to manipulate the datasets to display the output in the way I want. If this information were to be presented for a client, it would be recommended to use a visual format to present the information, to be concise yet detailed enough to give an effective report/summary that the client would easily understand. It would be recommended to use a presentation that references sources that were used, shows the original and final datasets, explains if there was any information that was cleaned or omitted, and shows the results in a visual plot where the information could be easily deduced. There should also be a clear purpose and point that is explained to why this information is important to the client and how it would help them. The data set itself should also be provided if the client wants to investigate it in more detail. For the future, I would like to include not only historical data from the New York Stock Exchange from 2013-2016, but also use current data through web scraping, and create a program that takes in the user’s input of the ticker and date and output the returns for that date, or the total revenue/liability of that year.