Data Preparation Using Stock Market Data

For this project, there were many new concepts to learn in the course and many assignments along the way that helped to facilitate the progress for the course project. I was able to process different data sources, perform data wrangling on the datasets, and was able to create visualizations for them.

The datasets that I used were the following:

CSV/Excel/PDF or another flat file source

<https://www.kaggle.com/dgawlik/nyse> - I was able to process the different files from this dataset and created dataframes in Python to hold the data. I performed various cleanup activities as well.

Website you want to pull data from

<https://finance.yahoo.com/quote/AAPL> - I was able to scrape the data from the webpages for each of the companies and stored the data in Python dataframes. This was a little challenging as it was necessary to be intimately familiar with the webpage structure in order to get the correct data.

API you will pull data from

<https://financialmodelingprep.com/api/v3/company/profile/AAPL> - This was relatively a straightforward task as the API returned a JSON object that could be easily processed in Python. Additionally there was no security or login requirements to use the API. I also looped through similar to web scraping to get the API data for each of the companies.

Unfortunately, I was only able to provide three visualizations of the data due to time constraints. It appears that I need to do additional data wrangling on the actual data that I wanted to use to link the different data sources together. I was able to load data into the sqlite database but unfortunately ran into errors when querying the databases. I hope to be able to focus more on visualizations in future courses. I believe the data that I was able to attain from the different sources are useable but needs a little bit more work.