Junhe Zhang, DATS 6401, Summer 2020

Progress Report

## Reveal and visualize the Important Factors that affect Stock Trend

The financial and Historical stock price data for Apple Inc (AAPL) was compile from finance.yahoo.com and was ready to export as csv files. To get the data from year 2005 to 2020, I created a two weeks free trial premium account. For the historical stock price data, I download the daily price data so that the available quarter financial report data can be merged with stock price data with date as identifier. There are total of three different form of financial report data, which are Income Statement, Balance Sheet and Cash Flow. Since there are a lot of accounting terminology like Operating Cash Flow, End Cash Position and so on, I put a fixed position description side bar with each important terminology explained so that users can easily find the annotation they need. In general, I will make a short description of each company that I am going to explore. Then, use a scatter plot to demonstrate the correlation between each financial data column and stock price in any fiscal year that users interested in from 2005 to 2020. Users are free to change the term of x and y coordinates and years.

Next, I plan to show users a line chart with the trend of AAPL’s stock price from 2005 to 2020 and a comparison line chart of Apple Inc’s financial data, so that users can easily find out the similarities and differences of trend between them. Finally, I will use the linear regression model or other regression models with the best accuracy that I get from using SciKitLearn GridSearchCV library which takes the financial data as independent X value to predict the dependent Y value which is stock price of that month. Since I could not know the future financial data, I will need to use the previous years’ monthly increase factor to forecast them. Then use a bar chart to show users the weight of each independent X variables in my data model, so that users can study which financial data term should be treat most important when they start investment.

I have constructed the html frame for my webpage and preprocessed the data I need. I am still testing which data model can best fit in this project. I plan to add couple more companies in the future on separate pages.