Final Project

Bellevue University

DSC530 – Data Exploration and Analysis

By: Tai Ngo

Date: 2/29/2020

**A Closer Look at the Growth Stocks**

Investment is a major part of our adulthood as we always look for the best way to grow our hard-earned money. It is a hot topic that has never been settled. Nobody, even the wisest investor, can say or know for certain that which stock is the best to buy. In this project, I have attempted to examine a dataset about 33 companies that are considered to be in the “growth technology stock”. I want to find the best stock in this list and obtain statistical information about it. With this information, we can compare with the average return on investment, which investors can buy from the S&P 500 index (9.8% over the last 90 years).

The hypothesis is investors should be actively looking for the best stock in the market and put their money in that stock temporarily to outperform the S&P 500 index. To make a judgement about this hypothesis, I have to find the best growth stock from the dataset, obtain statistical information and plot the data to aid my judgement.

According to the outcome of my exploratory data analysis, the company Shopify, ticker SHOP, has the best return on investment over the course of two months. Its return is 33.4%, which is 3 times the average of S&P 500 index. Another interesting insight is that it is not a very popular stock in my list because it is ranked 10th out of 33. Another good thing about this stock is that the variation between the mean of the lowest value and the highest value is only 7.9$, in which $216.5 being the mean of the lowest value and $224.2 being the mean of the highest value. So there is a very high chance that investor will obtain close to the theoretical highest percent return 33.4%.

This study will pose a lot of questions due to the missing of many values. First of all, the data only lasts for two months, which is not a long time when it comes to investing. Mistakes can always be made as time goes on. A more reliable dataset, which lasts between 6 months to 1.5 year is more ideal. More stocks should be on the list for the analysis. I have only 33 companies, which is a very small number.

For this project, I think that I have sufficient variables such as the highest/lowest prices. I can calculate the mean value. These variables are enough to make judgement regarding the hypothesis. My assumptions about the growth stock versus the S&P 500 are correct as I have been investing quite a lot of my time to observe the market. I understand that in a good market, a decent growth stock will definitely outperform any index, including the leveraged technology index such as TQQQ.

The only challenge that I face in this project is the regression. I have tried to import the statsmodels.formula.api, but it keeps giving me error. I also tried to re-install my Jupyter notebook but it still does not work. I would like to create some regression model to see if it can predict stock in the future and how accurate is the prediction. The formula for the regression model is another challenge that I have to face.