Heng Zhou, Hongyang Zheng,

Youyou Xie, Zhengqian Xu

Professor Lisa Singh

Analytics 501 Project Part1

03 October 2018

Potential Analyzes that Can Be Conducted Using Collected Data

To investigate the questions about stock market discussed in data science problem section: the factors impacting stock trading among different sectors, we need to collect both general stock market information, specific stock ask, bid price and size in a specific stock exchange website as well as general company information such as their industry and sector in the market. After collecting stock information both from a narrow and a broad scale, finding the influential factors, measuring the extent and predicting the trend when the influential factors change is our goal. Since we do not want study the trajectory of development of stock, we need to collect about 5000 different companies’ stock information for the same time point. Therefore, websites contain data about quotes for different stocks and company information are reasonable sources to scrape data.

The dataset has three part: general company information for each company, general stock market information about each stock and specific stock price as well as size in a stock exchange website. For the first part, company name and stock symbol are the basic information that should be contained in our dataset, because we may use this information to mark and analyze outliers in the future steps and we may use company name in the final report. Moreover, which industry or sector each company belongs to is also critical for our project, so that we can separate these stocks into different groups and analyze problems in a bigger scale. About general stock market information, we want to know average stock price and stock size in the market. For example, we use average total volume and latest price. With introducing these variables, we can explore differences and similarities between market average price and volume and the price and size in a specific stock exchange market. Moreover, market capitalization, risk coefficient and price change percentage of each stock are also key elements, since they can be viewed as representation of market value, reliability and fluctuation of each stock in the market respectively. As for the last part, our choice is to use the ask stock price, ask stock size, bid stock price, bid stock size in IEX stock exchange website as stock information in the narrow scope. This information contributes to describing stock trading in a specific environment, in this case, IEX website.

One direction of this project is to analyze stock price and stock volume by industry or sector and the results will show how market factors such as market capitalization or risk coefficient influence different groups by different levels. For example, our null hypothesis is that the price and volume of financials sector stocks are more vulnerable by bigger risk coefficient than stocks in other sectors. Another direction of this project is to do clustering analysis regarding to the behavior of stocks, derive stock typology and then give stock combination which has the minimum riskiness for example.