Final Project Progress

Jeremy Tsai

**Changes:**

I have had to slightly change the focus of my project due to difficulty in acquiring recent data regarding food expenditures, oil consumption, and overall consumption. Thus, I have shifted the focus to the changes in the most popular stock market indices: Dow Jones, S&P500, and NASDAQ Composite. I will continue to do sentiment analysis of tweets related to COVID-19 by week and see if there is any correlation between the positivity/negativity of the tweets in a week and the rise/fall of the opening prices of the three aforementioned indices. If there is a correlation, then it could signify that the tweets can be used to accurately judge the public’s perception and that these tweets have enough social influence to impact the stock market.

**Progress:**

For the data regarding the stock market indices, I have found the opening prices of the Dow Jones, S&P500, and NASDAQ Composite by week from February to April. The links are below.

Dow Jones: <https://finance.yahoo.com/quote/%5EDJI/history?period1=1580515200&period2=1588636800&interval=1wk&filter=history&frequency=1wk>

NASDAQ Composite: <https://finance.yahoo.com/quote/%5EIXIC/history?period1=1580428800&period2=1588636800&interval=1wk&filter=history&frequency=1wk>

S&P500: <https://finance.yahoo.com/quote/%5EGSPC/history?period1=1580515200&period2=1588636800&interval=1wk&filter=history&frequency=1wk>

For the data regarding tweets, I will use the February and March tweets datasets that were provided in Homework 2.

I also have begun researching ways to train a model to conduct sentiment analysis that will classify tweets as positive or negative. For each week, I want to run the sentiment analysis model through all the tweets and mark if a week is majority positive or negative. I will make this statistic range between 0-1, where 1 means all tweets were positive and 0 means all tweets were negative. I also plan to use WordCloud to visualize the most frequent words in the tweets for each week. This will help validate how accurate the sentiment analysis for the tweets for each week is and provide insight on which words were seen most for overall positive/negative weeks.

To combine the data regarding the opening prices of the three indices and the positivity/negativity of tweets for the weeks between February and April, I want to first plot the normalized stock market index opening price versus week for all three indices. Then, for each respective index, I will plot the data of positivity/negativity of each week over the normalized data points for the opening prices for each week. From this, we will be able to see if the positivity/negativity of the tweets for a week track the opening price of the index at that week. I would expect that when the tweets for a week are more negative, then the opening price will be lower versus another week that had more positive tweets. Furthermore, I will plot the opening price versus the positivity/negativity statistic for each week and conduct statistical tests to see if there is any correlation.