[Team-Grays-Peak](https://github.com/MSBX5420/Team-Grays-Peak)

4/25/2020

Design Document

Major Goal: [Team-Grays-Peak](https://github.com/MSBX5420/Team-Grays-Peak) want to analyze the trend of both positive/negative words and the stock price, get the plots and compare their trends and try to find the relationship.

Major Goal components description:

CDC news with COVID-19 related articles since March 26

stock market data: Data table of daily values from the (i) S&P 500 Futures, (ii) Nasdaq, (iii) DowJones, and (iv) Bitcoin-USD markets/indexes

Pyspark/Python jupyter notebooks for codes.

Interface:

Our stock market data and news data are stored in the S3 bucket and our pyspark and python code will be stored in the Leeds EMR cluster and Github. We will call our data by URLs from S3.

Descriptions:

Given that most of our COVID-19 data is text data, we are using various packages such as vaderSentiment.vaderSentiment, textblob and seaborn to do the sentiment analysis. We count the number of positive and negative words and use the compounds value to show the positive and negative trends of each day.

Our stock data has the price including open, close, high and low. We also decided to create another column called “rate” to show the changes for each day which calculated by (open-close)/open. We import pyplot to show the plot of rate vs time and exceed this plot with the compound value for words in CDC news plot. It could show us the trend and relationship between words from news and the stock market.