***Correlation of Stock Market and Sentiment of Discussion on Reddit***

**Motivation and Overview**

There have been instances where an online discussion group can impact the real world. Large hedge fund companies who had taken the short position have bearer the large sum of loss. For example, the reddit group had raised the stock price of the “game stop” by asking the community to buy shares. So, with the help of this study, the sentiment of the reddit discussion is correlated with stock market movement.

**Related Work**

There are studies which show a correlation between the crypto price and the Elon musk’s tweet. We as a team are fascinated with how social media or online platforms affect our social and business lives. So, with some research on google we concluded to work on how online group discussions affect the stock market. Regarding this we researched on twitter, reddit, Quora etc., and looked for data sets to work for our project. We found a data set related to Game Stop on Kaggle website. So, we decided to work on this topic as our semester project for CS-5610 course.

**Initial Questions**

We designed our project in such a way that it has to collect data from reddit, collect stock market data. It also has to perform correlation analysis. We explored different NLP tools of R and performed NLP on reddit data. We should also find any hidden relationship between the sentiments and market movement.

**Data**

We collected data from many sources. But we are listing out our many data sources below,

Reddit data (online discussion data set):

<https://www.kaggle.com/gpreda/reddit-wallstreetsbets-posts>

Stock market data (company’s share data set):

<https://finance.yahoo.com/>

**Exploratory Data Analysis**

For this project, we used R studio as a medium to run statistical and computational analysis with the data sets that we have collected. With the help of R programming, we created a Shiny app to correlate the online discussion and the stock price of Game Stop. We used histogram, scatter plots, time series plots so that our project may understand and communicate better.

**Data Analysis**

Our main aim for this project is to find correlation between online discussion and the stock price of Game Stop. We calculated correlation coefficient to find the relationship between online discussion and the stock price of the Game Stop. We also used linear regression and p-test for hypothesis testing. With the help of visualization tools, we made our project in a feasible way to understand easily.

**Narrative and Summary**

Reddit sentiments and Stock price are linearly dependent and there is significant evidence of after using p-test. So, this answers that, yes, there is some relationship between the Reddit sentiment and the game stop stock price. We have p-values which we got after doing hypothesis testing to justify the correlation. The only limitation we observed, only a single event had been observed with the correlation of reddit sentiment and the stock price, so we cannot say that Reddit sentiment influences the stock value for all scenarios.

**Project Website (github)**

<https://github.com/geethikach/R-project>