# Correlation Between Stock Prices and Tweet Sentiment.

Elias Joseph

**Introduction:**

This project will look into the possible relationship between the closing value for a company's stock prices, and the sentiment analysis of tweets about that company.  The tool developed can look at any company given a hashtag and a company name.  For the purposes of this report, I will look at Blizzard Activision, due to their recent scandal that has lowered their stock prices, and Gamestop, as their stocks have been pumped recently.

**Question:**

Does the average sentiment of tweets about a company  correlate with their stock price?

**Data:**

Yahoo finance was used to retrieve historical stock prices for different companies.  Yahoo finance offers a python api that is able to download data in a very neat dataframe.  The only preprocessing required was to fill in weekends and holidays.  The stock market is closed on the weekends, but twitter is not, so I decided to fill in the missing financial data using the ffill function in pandas, which fills in missing values with the last observed value.

Twitter data was a bit more difficult.  I pulled tweets relating to a certain hashtag using the Tweepy package.  I then extracted the date created and the text, and ran sentiment analysis using nltk's Sentiment Intensity Analyzer, and found the average sentiment for each day. I noticed that the vast majority of the tweets had a completely neutral sentiment, and examining them closer, they seemed to be tweets that were not expressing any opinion.



Distribution of Sentiment of Blizzard Tweets
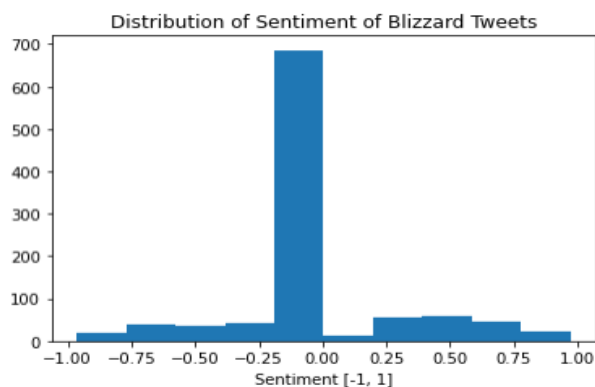
To get clearer results, I filtered out all tweets with a sentiment of exactly 0.

The next issue, which I was not able to overcome, is that twitter only stores up to 7 days worth of tweets, so I couldn't get data from more than a week previous. I set up a mongo database that holds the data that is pulled, so running this program once a week will produce a usable dataset, but I didn't have the time to do this for this project. Unfortunately, that means I will only be able to look at the trends over one week, which is too short a time to really capture all the phenomena I mentioned in the introduction.

To determine the existence of a correlation, I will use a granger causality test on the two time series, and evaluate the resulting F test.

**Program Description:**
This program starts by pulling both financial data from a company abbreviation and twitter data for a hashtag from a given time range. For yahoo finance, it uses the yahoo finance python library yfinance. For twitter, it uses Tweepy. As stated earlier, the weekend and holiday stock values are filled in using the pandas function ffill. The tweet sentiments are found using nltk Sentiment Intensity Analyzer and then averaged for each day, as to have identical timestamps to the financial data. The two data frames are then merged.

Once the data frame is processed, it is uploaded to a mongo db database. A plot is then made of the two time series datasets, and a granger causality test is run on it.

This program takes three arguments. A company abbreviation, a hashtag, and a directory to store analysis output.

**Results:**

Blizzard P Value: 0.39
Gamestop P Value: 0.78

**blizzard Stock Price vs Sentiment**



**gamestop Stock Price vs Sentiment**
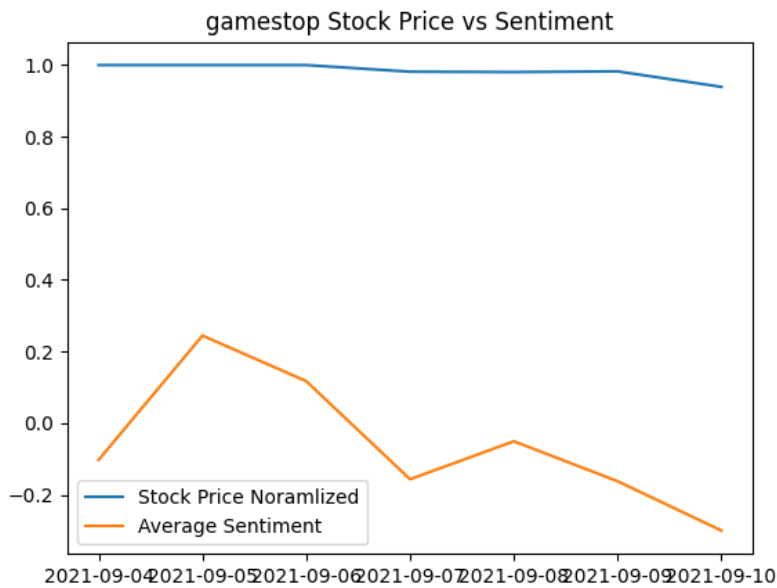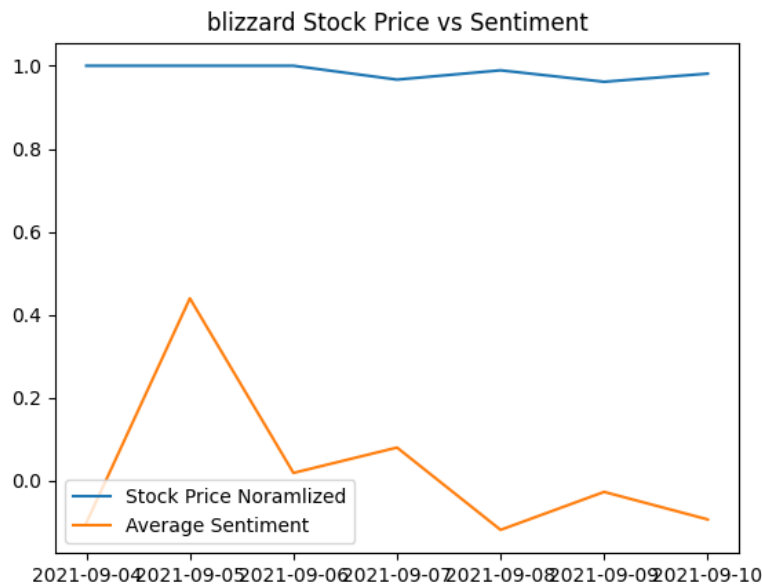
Overall, there was no significant correlation between stock price and sentiment analysis for either of the companies. I think this is largely due to the very short timeframe I was looking at and would be interested to see what would happen in a longer time frame. It might be better to use a social media platform other than twitter in the future, that does a better job archiving historical data.