War on Wall Street: Analysis of Today's Stock Market vs. Social Media Influence

Contact	Matt Gilgo, George Padavick msgilgo@bu.edu, georgep4@bu.edu 513-630-2794, 614-565-2809
Organization	N/A
Organization Description	N/A
Project Type	Data Science - NLP, Finance
Project Description	An analysis of stock market trends across various markets (NYSE, OTC, etc.) versus sentiment measured through NLP of tweets on Twitter. Stocks mentioned by high profile "Fintwit" pages will be measured by a stock's price increase/decrease combined with frequency and positive/negative tweet verbiage of said stock over a set timeframe.
Data Sets	● Twitter API ● Yahoo Finance API ● Alpha Vantage API ● Finnhub API

Suggested Steps	- Feature engineering: Build off Twitter-Monitor repo to determine positive or negative momentum for a stock via sentiment analysis. - Determine the source of the stock momentum - typically high-profile Fintwit accounts - Identify indexes to compare against in a similar timeframe as a baseline of market performance - Investigate correlations between stock price and twitter sentiment - Create visualizations to highlight any trends, indicators, etc. that occur in these stocks
Questions to be answered in Analysis	- Can we classify specific market behavior as a "pump"? - What are some of the largest pumps this year through social media schemes, and what was their relative impact on price? - What is the impact to new investors, positive or negative, that trade in reaction to these tweets? - Are there indicators to help warn new investors of what stocks are being sold as "pump and dumps"? - Can characteristics of predatory trading behavior be identified to increase fairness in the market?
Additional Information	Retail investors now make up 25% of the stock market. This may introduce additional volatility as well as potentially harmful financial behavior (Source: BusinessInsider)

Deliverable 0
- Per Lance's comment we will be starting with analysis on the following stocks:
    - **XSPA**
    - **GNUS**
    - **OPES**
    - **IBIO**
    - **GME**

- Using the API's listed above we will gather price movement data along with general twitter sentiment for the stocks above to determine if there are correlations between stock price and twitter activity. Understanding the correlation between twitter activity and price will be a key insight in determining how a "pumped" stock can be categorized.
- Overall, the scope of the project remains similar to the objectives listed above in the project proposal. The main goal is to understand and categorize stocks that may be overinflated by pure hype. In these cases the stock price is driven to levels well beyond what most analysts would categorize as the true price.
- Time permitting, we would like to analyze the impact of the following twitter accounts on stock price. These twitter accounts are known to provide stock tips that may initiate a "pumped" stock:
    - **@alexcutler247**
    - **@yatesinvesting**
    - **@MrZackMorris**
    - **@StockLizardKing**
    - **@The_RockTrading**

Approach
1. Collect price data for stocks (open, close, high, low) for specific intervals.
2. Collect twitter data related to those stocks (feature engineering).
    a. Number of tweets containing stock mention
    b. Classify tweets as positive or negative - potentially with varying degree.
    c. Investigate any trends in information propagation on twitter with specific tweets
3. Test various models in predicting stock price vs twitter activity. This exercise will evolve as we learn more about the data and understand what features we can engineer for the model.

Limitations
- Price movement is incredibly complex and can be influenced by a multitude of factors. This might make it difficult to determine underlying trends and correlations with data from Twitter. To combat this we will try to compare price data with baseline trends in the market using popular indexes.
- Twitter sentiment may be difficult to classify in certain cases. We will begin by comparing price data with overall twitter activity as a baseline before trying to classify tweets as "positive" or "negative". This should be relatively straightforward as stock tickers are already tagged with the "$" sign on twitter.

- In analyzing only a few stocks, there is a risk of initially overfitting our models. We would like to eventually expand our search into other stocks that exhibit similar behavior but to stay within scope of the project we will initially only analyze a few stocks.