**Overview**

**Project Title: Reddit GameStop Analysis -** Does Reddit Message Traffic affect GameStop Pricing

**Team Members:** Carter Kioski, Claire Wong, Ed Ober

**Project Description / Outline**

A group of Reddit users have driven up the price of the GameStop stock over the past year, collaborating through the Reddit application and purchasing stock. Our theory is that message volumes on Reddit would precede movements of the GameStop stock price. Our approach will be to create a model of price by day against the message volume, lagged by a range of 1-3 days.

**Research Questions to Answer:**

**Does Reddit message volume affect GameStop the magnitude of stock price changes within a 1-3-day period?**

- Message volume will be in the WallStreetBets subreddit

- Messages mention GameStop (GME)

**Hypotheses:**

**(H0 ): Null Hypotheses -** If Reddit message volume is not related to the magnitude of the GameStop stock price within 1-3 days, then a random sampling of 50 days from the top message volume days will show no effect on stock price.

**(Ha): Alternate Hypotheses** - If Reddit message volume is related to the magnitude of the GameStop stock price within 1-3 days, then a random sampling of 50 days from the top message volume days will have an effect on stock price.

**Appropriate Statistical Test: T-test**

**Acceptable significance value:** since this is a **business setting, particularly the stock market, a p-value of .05? .1?**

**Approach**

**Datasets to be used**

* **Messaging Volume: (Reddit API):** To provide message volume by day for the specific group, and for those messages containing a variant of $GME, GameStop, or potentially other names / tags
* **Yahoo Stock Market History: (Kaggle or Yahoo file):** stock prices and volume by day for the GameStop stocks

**Breakdown of tasks**

* Identify and validate data sources to be used
  + Messaging volume: Reddit API
    - Data for messages in the WallStreetBets subreddit from February 2020 forward.
    - Assume messages can be from or identifiable to the Reddit Group associated with the activity
    - Text is indexed or can be scanned for keywords for GameStop
  + Stock Market history: Yahoo, Kaggle
    - Data by Date
    - For GameStop - $GME
    - Pricing for the period: Open, Close, High, Low
    - Volume for the period
* Outline of the analysis
  + Overview of the “story”
  + Data analysis steps
    - ETL, particularly of the message data
    - Merging and verification
    - Modeling / graphing
    - Analysis of results – T Value
    - Insights
  + Define set of graphs and statistics showing the story
    - Time Plots of
      * Reddit Message Volume by day
      * GameStop Stock Price by day
      * GameStop Price Closing Change absolute value by day
      * GameStop Volume by day
    - Scatter Plots with regression and r-value
      * Message Volume Versus Price Change absolute value, multiple day lags (1,2,3)
      * Stock Volume versus Message Volume, multiple daily lags (1,2,3)
    - T-Test / p-value
  + Documentation
  + Integrate graphs and documentation
    - Highlights / lowlights
    - Results

**Reference Links:**

**http://reliawiki.org/index.php/Simple\_Linear\_Regression\_Analysis**

<https://www.investopedia.com/articles/active-trading/092214/hypothesis-testing-finance-concept-examples.asp>

<https://www.datacamp.com/community/tutorials/stocks-significance-testing-p-hacking>