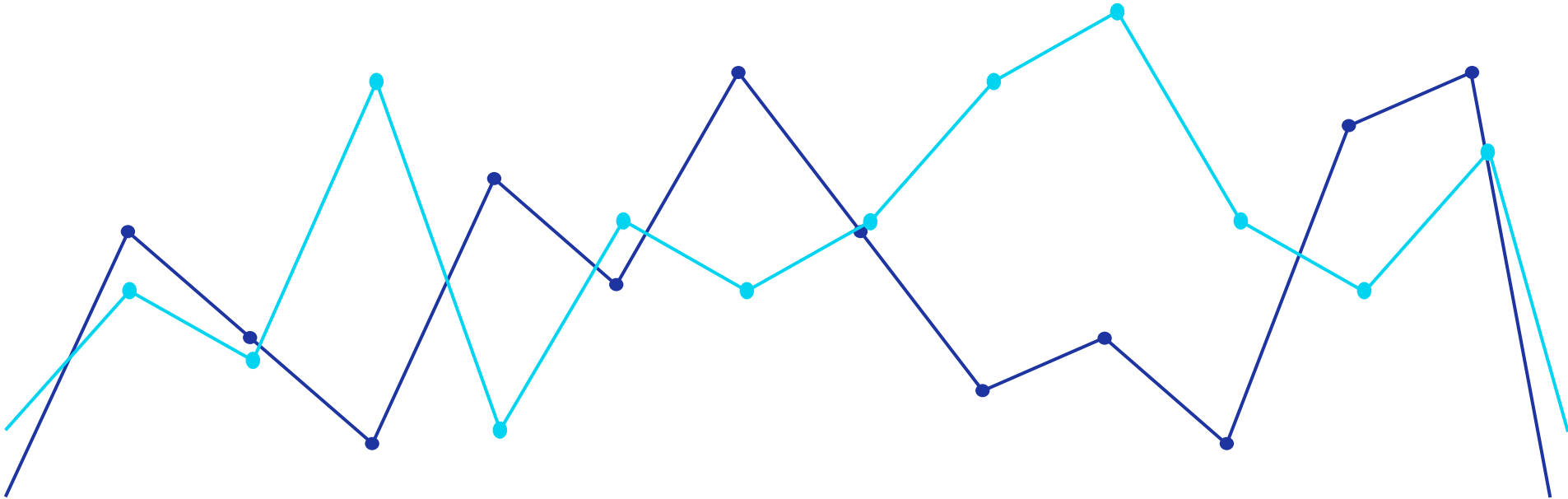
# Social Media Posts vs. Stock Price

By: Yousif Alnuaimi, Oscar Alonso, Ricardo Gomez, and Joanna Rashid



# Introduction

- Social media has enabled legions of retail investors to form online communities in which they share information and coordinate trading activity
  - Mostly notable in January 2021, where “failing” companies saw their stock price explose such as AMC Theatres, BlackBerry, GameStop, and others
  - These stocks were over shorted by hedge funds and that knowledge was widely discussed in these online communities
  - Social media communities coordinated and bought over shorted stocks to disrupt hedge funds trades
    - Example, GameStop rose 1897% in January 2021 costing hedge funds a lot of money
    - Media attributed this rise of price to a subreddit called Wall Street Bets

# Review of Current Literature

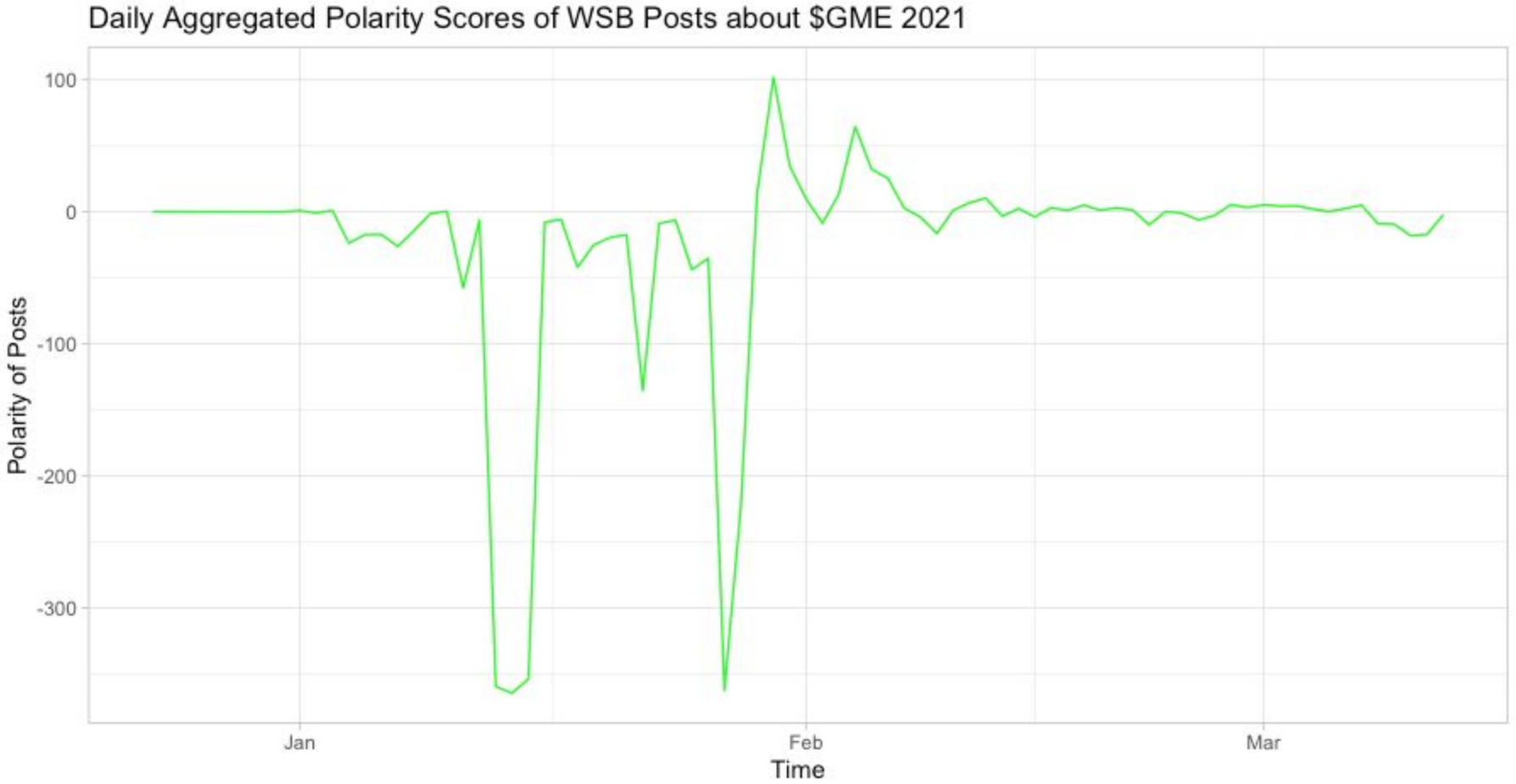
- Nobel-Prize-winning economist, Robert Shiller, has done extensive research into the power of narratives to drive major economic events
  - Argues that “studying popular stories that affect individual and collective economic behavior has the potential to vastly improve our ability to predict, prepare for, and lessen the damage of financial crises, recessions, depressions, and other major economic events.”
- New social media market metrics
  - Sentiment analysis has emerged as a widely for assigning a quantitative value to online narratives

# Data

- Kaggle was the primary source for data acquisition
  - Collected four separate datasets which were created by Kaggle users
  - Data came from the Wall Street Bets subreddit with a date range from June 2020 to April 2021
- Process of cleaning the data
  - Deconstruct datasets individually and pull relevant data
  - Combined datasets into one huge dataset

# Method

- Sentiment Analysis
  - Natural Language Processing (NLP) was used to generate a polarity score
  - A variety of module and packaged were test in Python and R but Python's module TextBlob was ultimately selected
  - Custom Lexicon was created by taking a random sample of 200 tokenized sentences
    - Sample was scored as positive or negative sentiment based on personal knowledge of the Wall Street Bets lingo
    - These values were used to train a Naive Bayes Classifier from the NLTK package
    - Despite customization, TextBlob's stock polarity algorithm outperformed the customized Naive Bayes Classifier
    - TextBlob Stock algorithm was ultimately selected to generate polarity scores

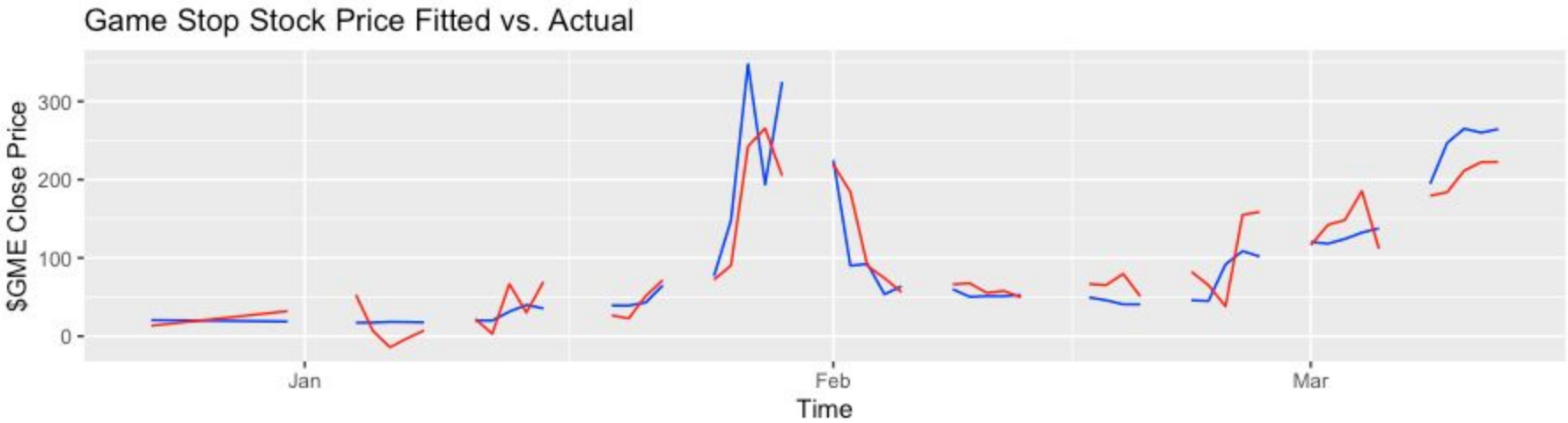

Daily Aggregated Polarity Scores of WSB Posts about $GME 2021
100
0
-100
-200
-300
Polarity of Posts
Jan
Feb
Mar
Time

# Modeling

- First attempts
  - Modeling included Multi-layer Perceptron Regressor and Multi-layer Perceptron Classifier from skLearn module with 80/20 split
  - Modeling didn’t go well, had negative accuracy scores.
    - Some literature discouraged use of such models with financial time series
- Second attempt
  - Time Series analysis was used to model the data
  - Data needed to be transformed to take care of irregular frequencies and variability
  - Auto.arima function from R was used to fit a model to the aggregated time series data
    - Data is best fit with (1,0,0) ARIMA model

# Results

- The fitted values produced by the model closely follow the same curve as the actual values, with a one or two day delay
  - Suggests that the model is not predictive
  - However, identical curve generated by the polarity scores from the sentiment analysis shows a strong relationship with trading activity



# Conclusion

- Our findings suggest that polarity scores generated by sentiment analysis of the narratives proliferating online are a significant driver of investment activity– specifically the exponential growth and decline of so-called “meme stocks.”
- Brokerages might use this technology to anticipate unusually high trading volumes in order to maintain enough net capital.
- The growth of the retail investment brokerages targeting small-portfolio, novice investors combined with the growth of online investment communities on Reddit, Discord, Twitter, and TikTok is changing the nature of the market analysis. The role of sentiment analysis of online economic narratives will certainly increase over time.