Wall Street Bets' Effect on GameStop

Sean Hong

Background Information

- Wall Street Bets is a subreddit page on Reddit that primarily discusses stock and option trading
- GME is an abbreviation for the GameStop stock
    - GME was practically a dead stock, valued as low as 3 dollars, then suddenly spiked to 81 dollars within a month
- What inspired this project?
    - The economic phenomenon that occurred due to the social media influence on stocks like GME and AMC
    - To see if there was a way to predict the trajectory of stocks using reddit
    - To visualize just how influential the subreddit was to its astronomical rise in value

3 Major Steps

Step 1

Identifying Packages Needed

Step 2

Web Scraping the Data

Step 3

Analyzing the Data

Packages Needed

01  Pandas

Used to easily create and manipulate the data frame

02  Pushshift API

Makes the data from Reddit actually attainable

03  DateTime

Allows for the dataframe to be filtered by date as a numeric value

03  NTLK and VADER

VADER is used to calculate the sentiment scores used in analysis

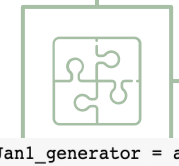
Implementation of Packages

DateTime

```
import datetime as dt
start_Jan1 = int(dt.datetime(2021, 1, 1).timestamp())
end_Jan1 = int(dt.datetime(2021, 1, 14).timestamp())
```

Pandas

```
# Creating an empty dataframe to store the new data
final_df = pd.DataFrame(columns = ["Week", "Sentiment_Value"])
GMEJan1_average = average(GMEJan1_Submissions["sentiment_score"])
final_df = final_df.append({"Week": "Jan Week 1 2021", "Sentiment_Value": GMEJan_average}, ignore_index = TRUE)
```



API

```
GMEJan1_generator = api.search_submissions(q = 'GME',
        subreddit = "WallStreetBets", score = ">3000", after = start_Jan1, before = end_Jan1)
```

NTLK/VADER

```
import nltk
nltk.download('vader_lexicon')

from nltk.sentiment.vader import SentimentIntensityAnalyzer
SIA = SentimentIntensityAnalyzer()
```

How the Packages Were Used

Pandas

Create an original df that just holds all posts

Create a new df with all averaged biweekly sentiment scores

PushShift API

Obtain all of the posts from Reddit with multiple filters

DateTime

Makes it possible to filter the API through dates by making it a numerical value

NLTK

Holds the VADER package

VADER

Obtains a sentiment value for each individual post and later averages the total

Obtaining the Data

1. Use DateTime

```
start_Feb = int(dt.datetime(2021, 2, 1).timestamp())
end_Feb = int(dt.datetime(2021, 2, 28).timestamp())
```

2. Use API/Pandas

```
GMEFeb_generator = api.search_submissions(q = 'GME', subreddit = "WallStreetBets",
                                          score = ">3000", after = start_Feb, before = end_Feb)
GMEFeb_Submissions = pd.DataFrame([submission.d_ for submission in GMEFeb_generator])
```

Dataset

3. Convert the Series into Readable String

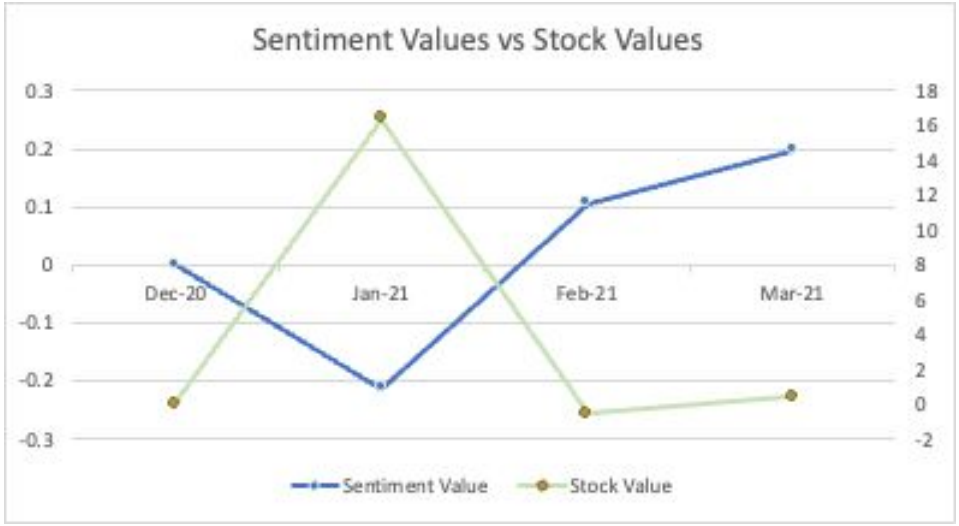
```
GMEFeb_Submissions['selftext'] = GMEFeb_Submissions['selftext'].astype(str)
```

4. Use VADER

```
GMEFeb_Submissions['sentiment_score'] = GMEFeb_Submissions['selftext'].apply(calculate_sentiment)

GMEFeb_average = average(GMEFeb_Submissions["sentiment_score"])
final_df = final_df.append({"Month": "Feb 2021", "Sentiment_Value": GMEFeb_average}, ignore_index=TRUE)
```

Graphing the Data



To standardize the two bars, percent difference was taken between the said month and the starting value (December 2020) for the stock value

The blue line represents the sentiment value while the green line depicts the actual stock value

Conclusions

Monthly Values Are Not a Good Indicator

Because a lot of data was not able to be collected, it might be better to find either weekly or biweekly values. Especially for the massive spike in January

From a Monthly Perspective...

Sentiment values are not an accurate indicator in predicting stock market values

Questions?