

# The Role of Social Media in Financial Decision-Making

Amogh Batwal, Rohit Rokde, Jayati Dev

Luddy School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN, USA  
 abatwal, rrokde, jdev@iu.edu

## ABSTRACT

People often refer to social media knowledge to inform personal finance decisions. Prior research has shown that sentiment analysis of Twitter and Reddit posts can play a role in predicting stock market changes. This calls for a broad analysis of available information on social media platforms that have the potential to influence financial decision-making. Our goal for this study was threefold - (a) understand the reach and user engagement of financial content on social media, (b) range of financial topics discussed, and (c) popular reaction towards these topics. In this study, we collected 112248 tweets, 75347 Reddit posts, and 12915 Youtube comments. We found that stock market events can potentially explain user activity on all three platforms. Additionally, the breadth of topics discussed is wide with a higher frequency of terms related to personal finance than global trading. Finally, sentiments expressed on different platforms are different even during the same time frame. Our work shows that diversity in information on social media platforms plays an important role in enhancing user engagement on topics pertaining to finance and perhaps, decision-making.

## CCS CONCEPTS

•Applied computing~ Law, social and behavioral sciences •Information systems~ Information systems applications~ Decision support systems~ Data analytics •Information systems~ Information systems applications~ Data mining

## KEYWORDS

Social media, data mining, finance, Reddit, Twitter, Youtube

## 1 Introduction

Social media is a powerful source of information of all kinds and finance is no exception. Social media websites like Twitter and Reddit are often used as platforms for sharing news articles related to personal finance and stock market trends which investors read for making financial decisions [1]. Prior research has found that there is a way to predict stock market fluctuations based on the analysis of people's expressed sentiments on tweets [2] and finance articles [3]. This suggests that personal financial decisions made by users are reflected in their social media interactions, especially those that are community-based, like Twitter or Reddit. These two platforms have a combined monthly user count of 660 million [4]. People also engage with information in video format over YouTube [5] on financial topics, which has over two billion users [6]. This high volume of user engagement creates a scope to study the range of topics discussed related to personal finance and their effectiveness in influencing user sentiment, and by extension, their decisions.

In this study, we measure user engagement from January to April on three social media platforms which are sources of financial news for users - Twitter, Reddit, and YouTube. Even though all three are publicly visible in terms of content, the nature of these platforms is different. Twitter unlike Reddit is not community moderated and has limited character count to 280. Reddit allows for longer, more topic-oriented posts in a discussion forum format. YouTube is specifically for video-based content where people can express their opinions through comments. Given that these platforms have several users who are in the United States, we wanted to examine the breadth of information available to users when comparing it to stock market fluctuations. Specifically, our research questions are the following:

1. How online communities like Reddit and Twitter disseminate financial knowledge?
2. To what extent can discussions on social media about financial news reflect stock market behavior?
3. How effective are financial discussions on social media in affecting people's views?

For the first research question, we used topic modeling using Latent Dirichlet Analysis to understand the different topics that were under discussion in a moderated community like Reddit and unmoderated community like Twitter. We found that there was a broad range of topics related to financial advice that people engaged in, especially credit and tax. For the second research question, in order to find how discussions on financial news reflect stock market behavior, we reported user activity (number of tweets and posts) on all three platforms over a chosen time frame (January 1, 2020 to April 24, 2020) and the corresponding stock market trend for S&P500 and NASDAQ Composite which are the two most followed stock markets in the United States. Since Twitter is used by individuals to post information often in real-time, there was a hike in the number of tweets during the stock market crash (between March 8 and March 24, 2020). Furthermore, the maximum number of YouTube comments were on March 9 and March 13, 2020, which were the days on which the stock market lost the maximum number of points<sup>1</sup>. Stock market behavior was also mirrored through Reddit forum posts `r/stockmarket` but did not affect the `r/personalfinance` subreddit. Stock market behavior reflected on Reddit forums and YouTube comments as well, with `r/stockmarket` following a similar trend as Twitter while `r/personalfinance` subreddit did not show any significant changes. For the third research question, we performed sentiment analysis using NLTK Sentiment Analyzer on all three forums to gauge how people on these platforms express their opinions on posts that pertain to finance. We found that financial discussions evoked more negative sentiments on Twitter, more positive sentiments on Reddit, and about the same levels of positive and negative sentiments on YouTube.

The findings show that diversity in information on social media platforms plays an important role in enhancing user engagement on topics pertaining to finance and consequently, decision-making.

## 2 Related Work

Social media is often considered a proxy for real-world interaction in several research studies which is known as “behavioral finance” for the finance industry [15]. Even though social media platforms like Twitter, Reddit, or YouTube might not be representative of the general population and create validity issues, it is nevertheless an invaluable source of information for human behavior. In their 1999 study, Malhotra and Galletta found that individual decision-making is often influenced by their social circles [7]. Social influence is also one of the four factors that form the basis of the Unified Theory of Acceptance and Use of Technology (UTAUT) [8] which focuses on the adoption of technology. Extending this to social media platforms, we can construct that financial information on social media platforms is also accepted by users partially due to the social influence of these platforms. Hazari and Richards interviewed 131 Twitter users for their perceptions on financial discussion topics on such social microblogging websites and found that these platforms were indeed extensively used and deemed to be beneficial for financial information [14]. Analysis of topics on such platforms also found that the discussion revolved around common financial topics that were being discussed in the news by financial professionals to address most raised queries around finance [16].

Prior research has shown that Twitter and Reddit are indeed extensively influential for financial decision making. Sprenger et al.'s work shows that users on social media, specifically Twitter, engage with tweets that provide good financial advice [9]. User activity and analysis of user networks reveal that stock market

---

<sup>1</sup> <https://www.thebalance.com/fundamentals-of-the-2020-market-crash-4799950>

activity (like price, traded volume, etc.) was highly correlated with the number of nodes in the network [10]. In other words, connection with other users as “followers” or “following” on social networks influenced stock market trading.

Furthermore, extensive sentiment analysis has also been done on social media platforms to show that events on stock markets like Dow Jones are correlated with the sentiment expressed on Twitter [11]. Sentiment analysis could even be predictive of the changes in stock market events over a period of time [2] especially when the analysis is done for specific topics on platforms like Twitter [12]. The polarity of sentiment was found to correlate with stock market returns as well [11]. In other words, negative sentiment corresponded to the fall in stock market values while positive sentiment corresponded to the rise in value. This is true not only for stock markets but also for specific retail brands when sentiment analysis is done on Twitter handles of specific retail brands [13].

However, there is a need for a more descriptive analysis of multiple platforms and their varied role in informing financial decision-making. Our approach has been the opposite. In our work, we examine three social media platforms instead of a single platform, in order to compare the breadth of user activity. We then examine the potential of stock market fluctuations to explain such activity and sentiment on these platforms.

## 3 Method

### 3.1 Data Collection and Cleaning

Reddit data was collected through the Reddit API using Python’s Reddit API Wrapper (PRAW) library. We collected data from three major subreddits related on finance, [r/personalfinance](https://www.reddit.com/r/personalfinance/)<sup>2</sup>, [r/stockmarkets](https://www.reddit.com/r/StockMarket/)<sup>3</sup>, and [r/financialplanning](https://www.reddit.com/r/FinancialPlanning/)<sup>4</sup>. Twitter data was collected using [GetOldTweets](https://github.com/Jefferson-Henrique/GetOldTweets-python)<sup>5</sup> for getting tweets in the desired range of dates (January 1, 2020 to April 24, 2020) using “personal finance” as the fixed search query. These dates also corresponded with the ongoing COVID-19 global pandemic<sup>6</sup> which created fluctuations in stock market values and must be considered while analyzing the findings. National recommendations on containing the outbreak and restrictions on incoming travel were done during the week of March 15, 2020, just before the stock market price hit their lowest (Figure 1). Since Twitter’s own API, Tweepy limits the rate of calls and cannot retrieve data beyond a week, GetOldTweets was found to be more appropriate for our dataset.

Finally, data from YouTube was collected through YouTube’s API that allowed us to search for a specific string and collect comments for specific videos. In our case, we selected the top 20 video IDs following a manual search from YouTube and parsed them through the YouTube API to get all the comments under each video.

We used libraries in Python’s Natural Language Toolkit (NLTK) to break down the text corpus from both Reddit and Twitter in word tokens. This was followed by using the `stop_words` package in NLTK to remove stop words like prepositions and conjunctions from the resulting set of words. PorterStemmer reduced the tokenized words into their root form for consistency and prepared the dataset for topic modelling using Latent Dirichlet Analysis (LDA).

---

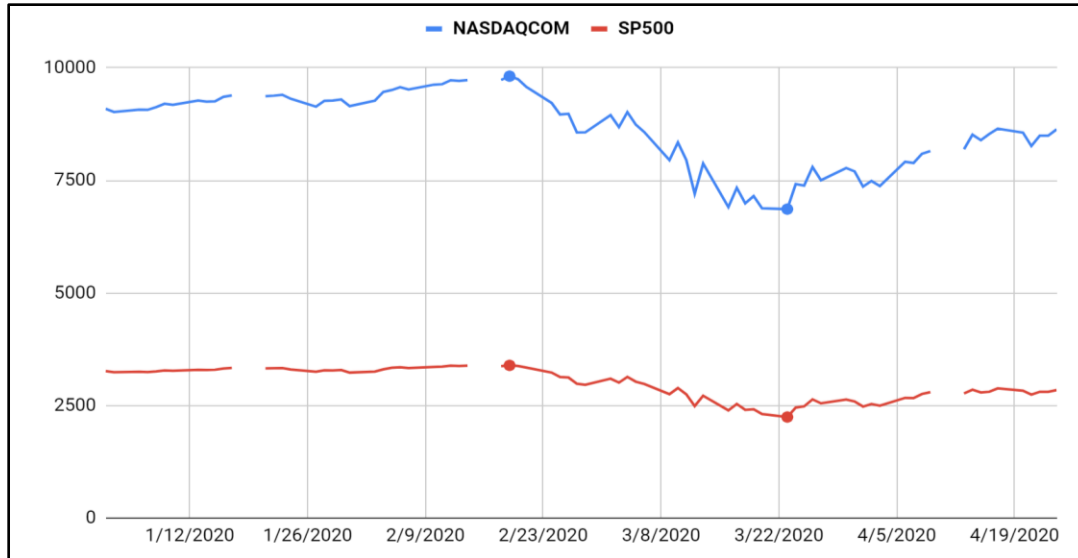
<sup>2</sup> <https://www.reddit.com/r/personalfinance/>

<sup>3</sup> <https://www.reddit.com/r/StockMarket/>

<sup>4</sup> <https://www.reddit.com/r/FinancialPlanning/>

<sup>5</sup> <https://github.com/Jefferson-Henrique/GetOldTweets-python>

<sup>6</sup> <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen>



**Figure 1: S&P500 and NASDAQ Composite values for 1/1/2020 to 4/24/2020 which are two of three highly followed stock markets in the United States, along with Dow Jones. Gaps show no records available for the timeframe. Steepest drop on 3/9/2020, 3/12/2020 and 3/16/2020. Initial date for federal tax return filing was 4/15/2020, later adjusted to 7/15/2020. There are three gaps in the graphs due to non-availability of stock market price information on Bloomberg Finance.**

## 3.2 Analysis

We performed a composite analysis on all three social media platforms due to the different characteristics of each platform. First, we reported user activity on Twitter (tweets), Reddit (forum posts), and YouTube (video comments) over a certain period (January 1 to April 24, 2020). This time frame corresponded with high market fluctuations (as can be seen in Figure 1), and we proceed to examine if user activity is influenced by stock market movement (in terms of total value only).

We then proceeded to perform LDA on the Reddit dataset. LDA is available as a set of functions in Python's `sklearn` decomposition package. We set the number of topics for LDA modelling to be 10, to capture a robust set of topics and their frequency of occurrence. We then performed sentiment analysis of Reddit posts, tweets, and YouTube comments using VADER (Valence Aware Dictionary and sEntiment Reasoner) [17] available through NLTK using `vader_lexicon`<sup>7</sup>. VADER is a commonly used tool in social media mining that uses a rule-based approach to conduct sentiment analysis using a function called `SentimentIntensityAnalyzer`. Pandas was used to convert the collected data into data frames for easy string manipulation.

## 3.3 Ethics

All data used in the study was open and publicly available for use without logging in for Reddit, Twitter, and YouTube. Thus, this did not require an Institutional Review Board (IRB) approval. We did not collect any personally identifiable information from any user across the three platforms to preserve data anonymity. Our analysis is descriptive and not experimental and will have no effect on user behavior or their relationship with these platforms, reducing the potential ethical risks of this study to minimal.

<sup>7</sup> <https://towardsdatascience.com/end-to-end-topic-modeling-in-python-latent-dirichlet-allocation-lda-35ce4ed6b3e0>

**Table 1: Data description for Reddit dataset**

| Subreddit Name    | Description from Reddit   | # Users | # Posts | # Comments |
|-------------------|---|---------|---------|------------|
| personalfinance   | A platform to learn precisely regarding budgets, getting out of considerable debt, typically investing and planning about retirements         | 14.2mi  | 67000   | 722866     |
| stockmarket       | A subreddit to get on board with stock market news, trading, long and short-term traders, technical analysis, and fundamental interpretation. | 399716  | 5719    | 64953      |
| financialplanning | A subreddit to discuss personal finance, personal income, retirement plans, investing, insurance, frugality, budgeting, and cover.            | 119536  | 2628    | 21023      |

## 4 Dataset Description

As mentioned above, Reddit posts were collected from three different subreddits, each of which is described in Table 1, generating a dataset of 75347 posts. In the case of Twitter, we collected 112248 tweets over a period of four months (January 1-April 24, 2020) according to the following keywords as the search criteria: “S&P500”, “financial planning”, and, “stock market”. The third dataset from YouTube consisted of 12915 video comments across 20 videos found using the search string “financial discussions”. Table 2 shows the list of videos collected.

**Table 2: Data description for YouTube dataset**

| Video # | Video ID    | # Views | # Comments | # Likes | # Dislikes |
|---------|-------------|---------|------------|---------|------------|
| 1       | 9qMjytwFy-g | 66,159  | 271        | 617     | 20         |
| 2       | b8fSPr2gcKQ | 574,718 | 687        | 3300    | 203        |
| 3       | J3gTeafNZc0 | 19,695  | 297        | 592     | 46         |
| 4       | n6TOlFfe-zI | 109,688 | 1014       | 1000    | 108        |
| 5       | wsk-jX0S1s0 | 69,108  | 1057       | 1400    | 143        |
| 6       | 4IoDMWCnFzM | 405,968 | 3505       | 9900    | 1400       |
| 7       | w73GeE1w_n0 | 95,060  | 16         | 2000    | 53         |
| 8       | tChL0wsMJtc | 7,696   | 16         | 161     | 8          |
| 9       | eWE2a84V13U | 48,426  | 153        | 247     | 59         |
| 10      | vuFKYvd974o | 168,049 | 213        | 4700    | 245        |
| 11      | K0LsRsuPbB4 | 64,205  | 412        | 346     | 196        |

|    |             |         |      |      |      |
|----|-------------|---------|------|------|------|
| 12 | VQqFr7_Y5NU | 672,216 | 2519 | 3600 | 1500 |
| 13 | Iqnb8u567Jo | 143,632 | 97   | 3300 | 83   |
| 14 | BgHcrBKv-Rs | 31,041  | 57   | 106  | 13   |
| 15 | tMBnt2QhQ4w | 42,856  | 216  | 856  | 35   |
| 16 | 5Ib7_XgbhYs | 92,456  | 541  | 1600 | 72   |
| 17 | WA1Ji-Hjlqo | 513,100 | 959  | 6000 | 586  |
| 18 | EzLW-cPEmy0 | 91,876  | 745  | 665  | 131  |
| 19 | bn35NNeodoc | 1,878   | 21   | 68   | 4    |
| 20 | nnY5D-maytk | 25,238  | 119  | 586  | 12   |

## 5 Findings

As aforementioned, we did three different analyses. First, a time-series analysis to show the relationship between user activity and stock market prices. This is done by plotting user activity in terms of posts (Reddit), tweets (Twitter), and comments (YouTube) over the duration discussed above. We perform a linear regression analysis between these plots with NASDAQ Composite stock market values, to find that while posts in the `r/stockmarket` subreddit are statistically significant for NASDAQ stock value, the values have no significant effect on `r/personalfinance` or `r/financialplanning`. Second, LDA on both Reddit and Twitter reveals a variety of topics, mostly related to personal finance, that dominates the financial discussions on social media platforms. Finally, we conduct sentiment analysis on all three platforms to gauge a user's expressed sentiments towards financial content on social media. A more detailed report of our analysis is presented below.

### 5.1 Activity on Stock Market Subreddit Correspond to Stock Market Trends

There is a sharp dip in Figure 2, 3, and 4 at the extreme right of the graphs possibly due to missing data for the last data of data collection (which was 3:00am, missing the posts from the following morning). Figure 2 shows the number of posts and comments on the `r/stockmarket` subreddit from January 1, 2020 to April 24, 2020. The hike in the number of posts occurs during March 8 to March 14, 2020, and subsequent hikes following till April 15, 2020. The graph plot of comments follow a similar trend. In order to measure if user activity on this particular subreddit corresponded to the stock market values, we used a simple linear regression model using NASDAQ Composite daily prices as the independent variable and number of subreddit posts as the dependent variable. Given the computed R-squared value is 0.481893, we see that the relationship between daily stock prices have a statistically significant ( $Z=0.766$ ;  $p\text{-value}=3.5e-12$ ) effect on the number of subreddit posts (See Table 3). This means that discussion on this particular subreddit is highly dependent on stock market movement.

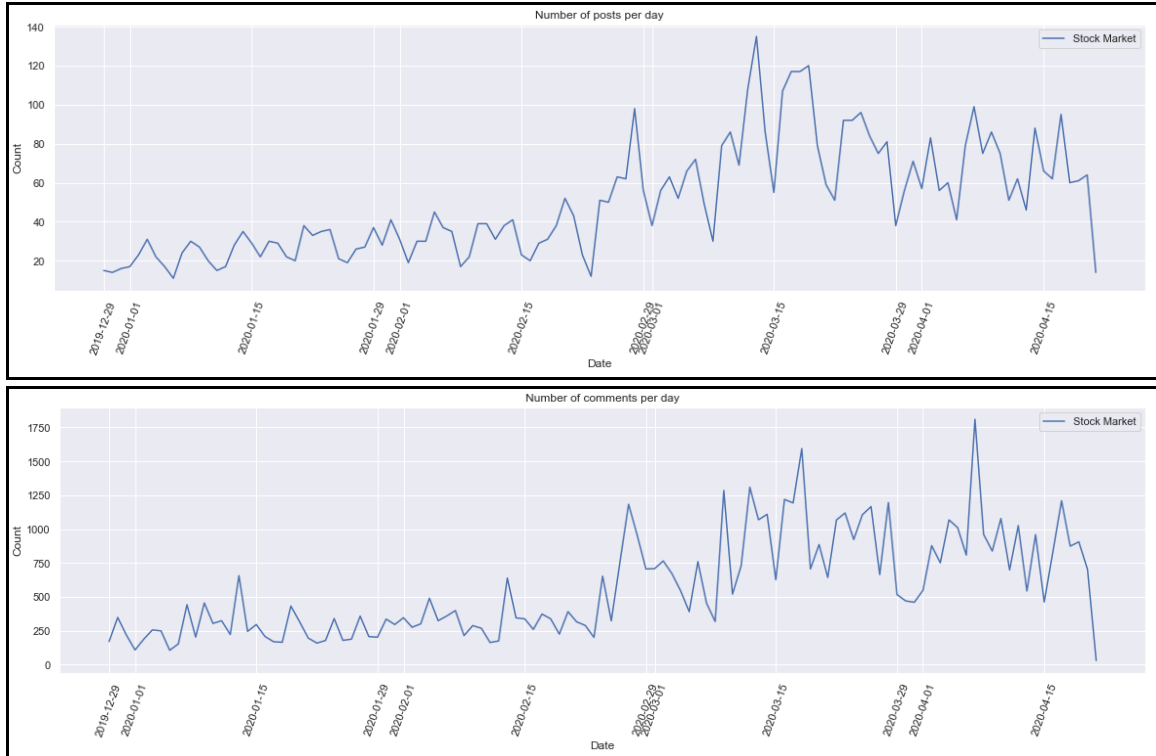


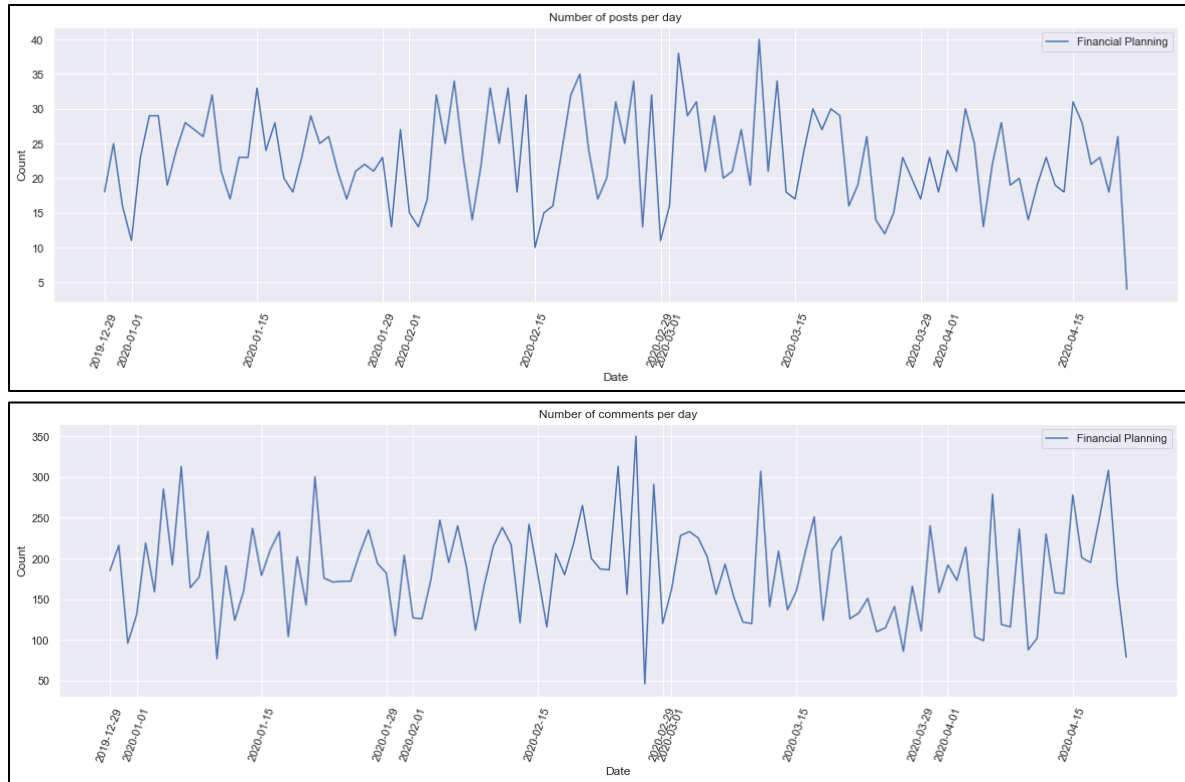
Figure 2: Number of posts and comments over time on the `/stockmarket` subreddit

Table 3: Linear regression values of number of posts on NASDAQ stock prices for `/stockmarket`

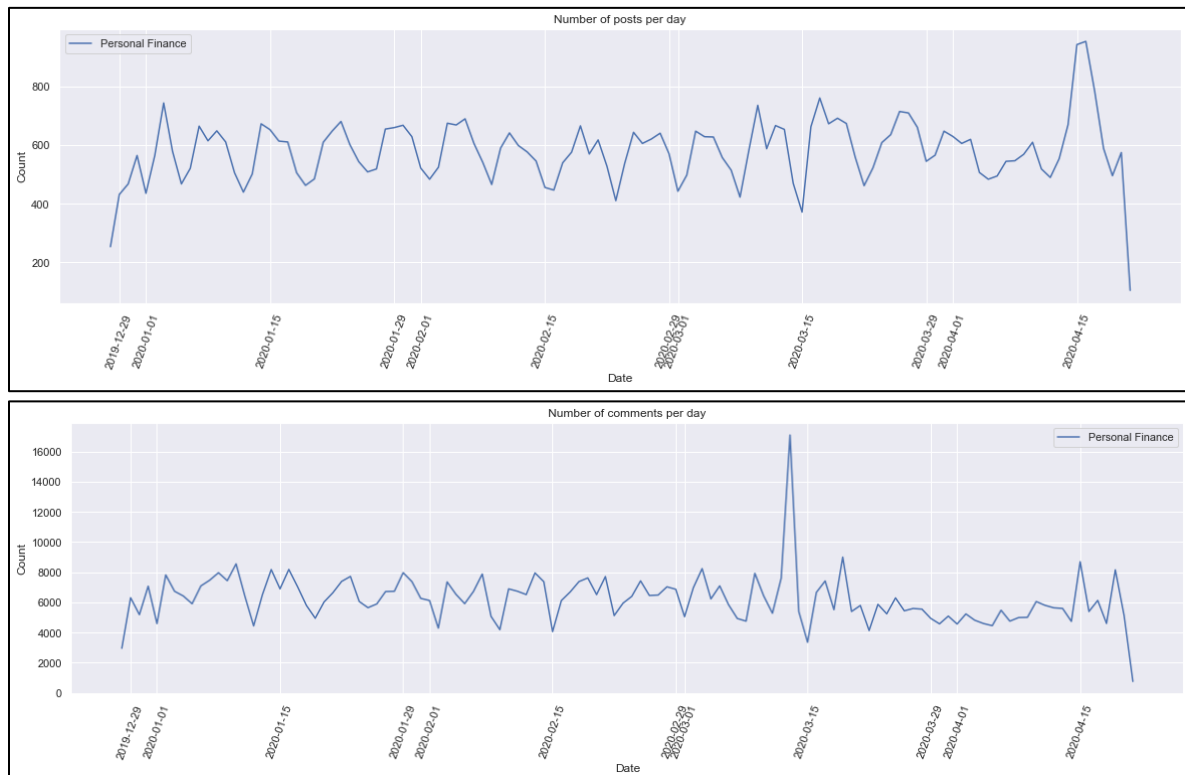
|                           | Coefficient | Std. Error | t-value  | p-value  | Lower 95% | Upper 95% |
|---------------------------|-------------|------------|----------|----------|-----------|-----------|
| <code>/stockmarket</code> | 0.766576    | 0.0924     | 8.296243 | 3.56e-12 | 0.582464  | 0.950688  |

## 5.2 Activity on Personal Finance Subreddits Do Not Correspond to Stock Market Trends

Contrary to the `/stockmarket` subreddit, user activity on `/personalfinance` and `/financialplanning` did not have a strong relationship with stock market prices. As can be seen in Figure 3 and 4, the level of user activity (posts and comments) follows a similar trend throughout the timeframe of collected data. Thus, we initially assume that stock market prices do not have an influence on user activity in these subreddits. We then perform a linear regression again to find if the number of posts on these subreddits are dependent on NASDAQ stock prices. The computer R-squared value for `/personalfinance` and `/financialplanning` were 0.000571 and 0.047267 respectively.



**Figure 3: Number of posts and comments over time on the /financialplanning subreddit**



**Figure 4: Number of posts and comments over time on the /personalfinance subreddit**



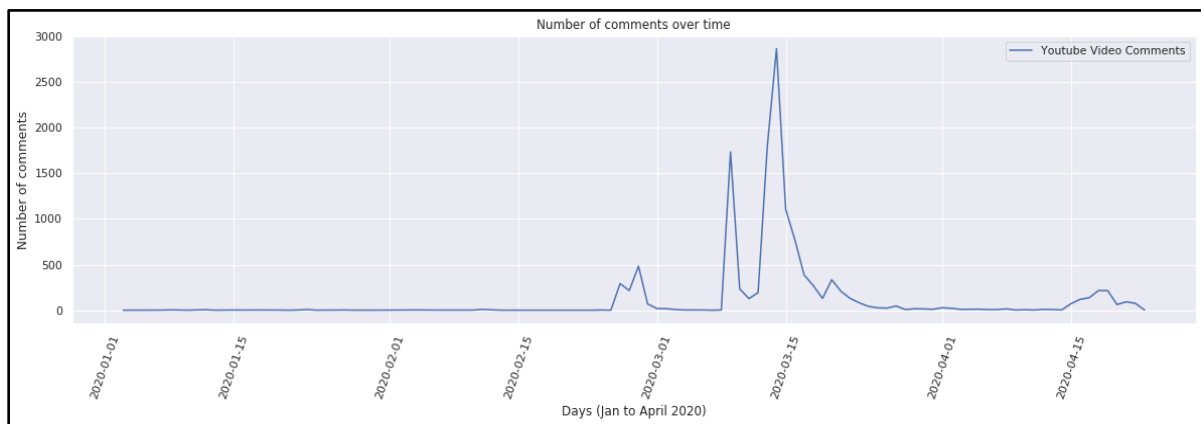
As can be seen in Table 4, the coefficient for the number of posts against NASDAQ stock market prices is not statistically significant for either /personalfinance ( $Z=0.007614$ ,  $p\text{-value}=0.8376$ ) or /financialplanning ( $Z=-1.08681$ ,  $p\text{-value}=0.0592$ ). This means that stock market prices do not affect the financial discussions that happen in these subreddits. While Figure 4 does show a peak on April 15 (which is usually Tax Day in the United States; the last day for filing taxes), it is unclear whether this was a result of tax related inquiries since federal Tax Day had been pushed to July 15 due to the ongoing COVID-19 pandemic.

**Table 4: Linear regression values of number of posts on NASDAQ stock prices for /financialplanning and /personalfinance**

|                    | Coefficients | Std Error | t-value | p-value | Lower 95% | Upper 95% |
|--------------------|--------------|-----------|---------|---------|-----------|-----------|
| /personalfinance   | 0.007614     | 0.03704   | 0.20556 | 0.8376  | -0.06619  | 0.081417  |
| /financialplanning | -1.08681     | 0.56721   | -1.9160 | 0.0592  | -2.217    | 0.043379  |

### 5.3 Twitter and YouTube Activity Correspond to Stock Market Trends

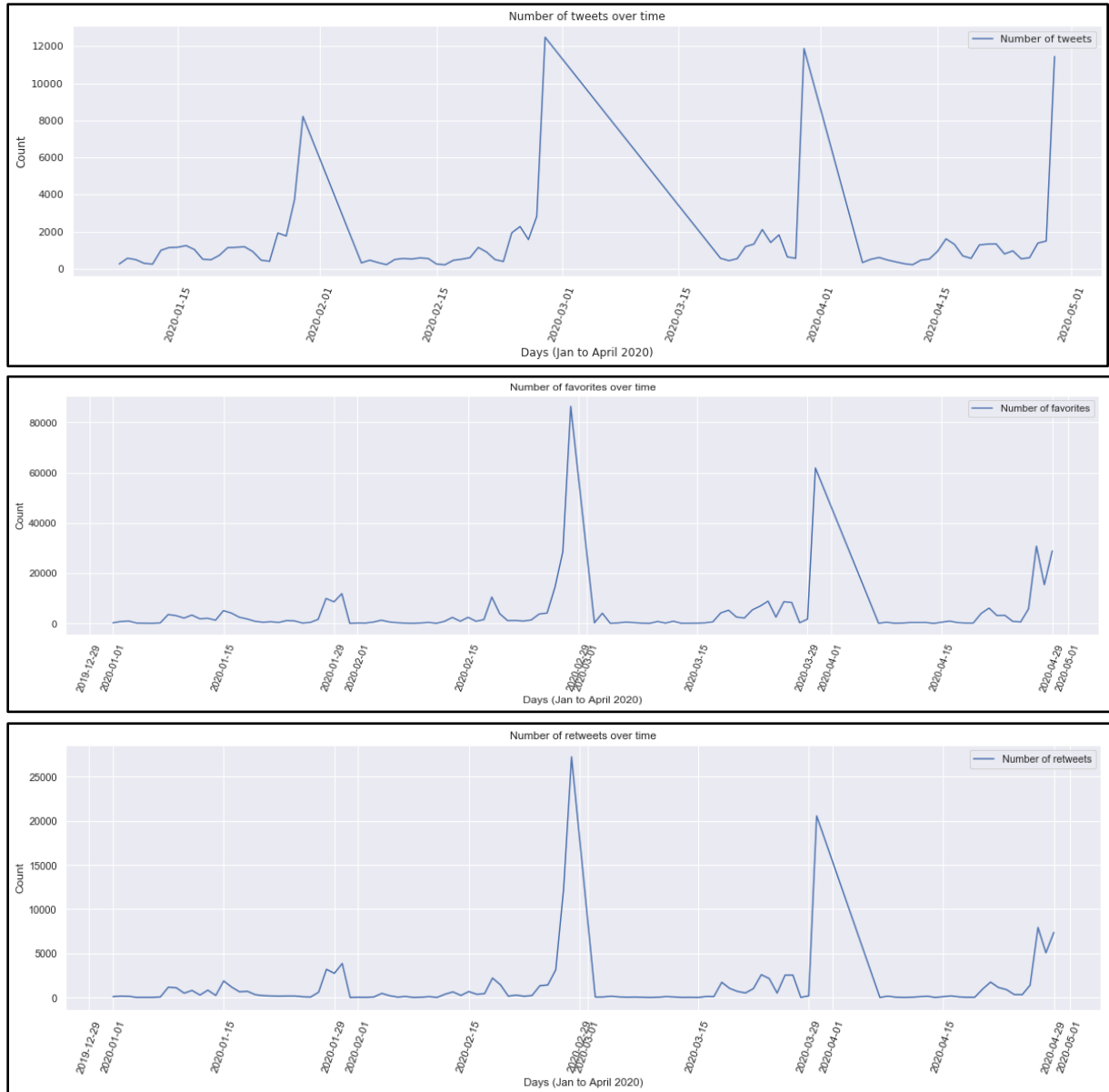
Considering the comments on videos in Table 2 from YouTube, we find that the plot of comments over time peaks during March. The maximum number of comments were on March 9, 13, and 14 which were on or around the days the stock market lost the maximum number of points (March 9 and March 12, 2020)<sup>1</sup> as can be seen in Table 5.



**Figure 5: Number of comments over time for YouTube**

**Table 5: Top three dates with maximum number of comments**

| Date | 2020-03-09 | 2020-03-13 | 2020-03-14 |
|------|------------|------------|------------|
| #    | 1732       | 1793       | 2862       |



**Figure 6: Number of tweets, retweets, and favorites over time on Twitter**

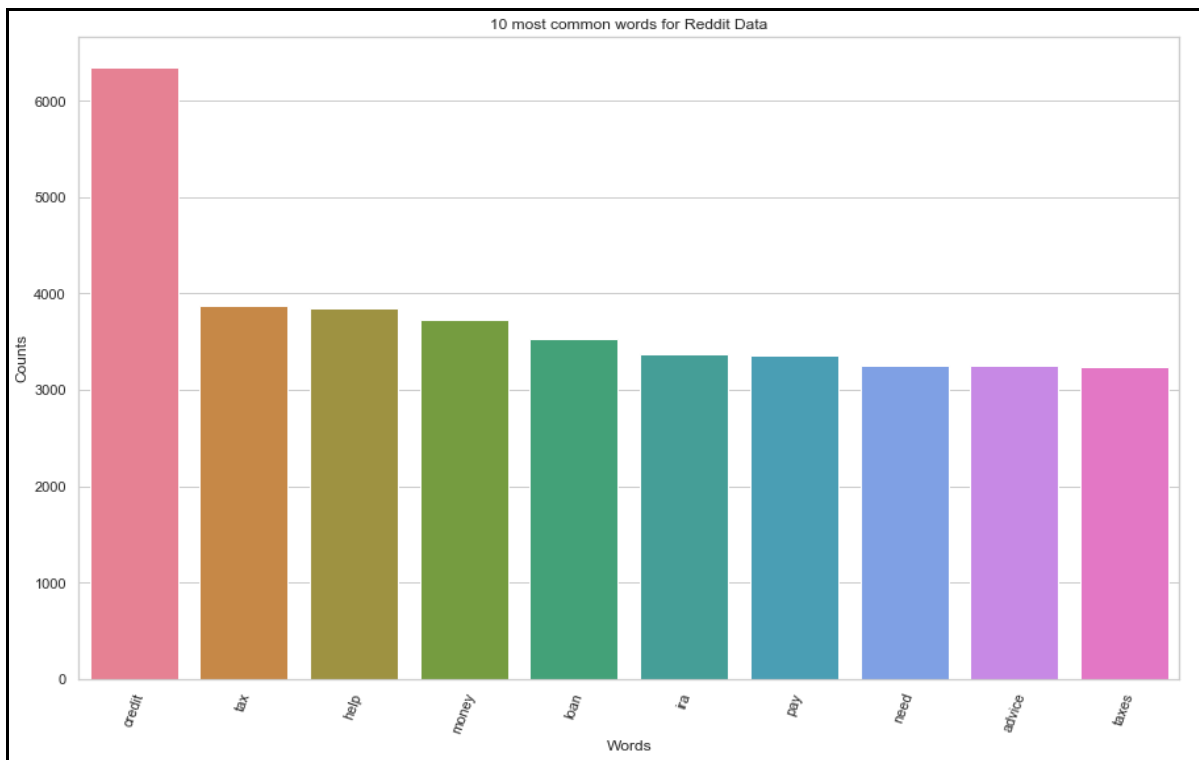
In case of Twitter, the number of user tweets seem to peak during the first week of every month (see Figure 6), with the widest stretch for 2020 being in the first two weeks of March. This is possibly because of the stock market movement as well<sup>1</sup> which had significant downward fluctuations during that period. Since our search criteria on Twitter was “S&P500” and “stock market”, it is likely that we collected more stock-related financial discussions than personal finance. The number of retweets and favorites for the tweets followed a similar trend, suggesting that the maximum amount of user engagement on tweets through in the form of retweets and marking tweets as favorite happened during the same timeframe. Table 6 shows the dates with the maximum number of tweets.

**Table 6: Top three dates with maximum number of tweets**

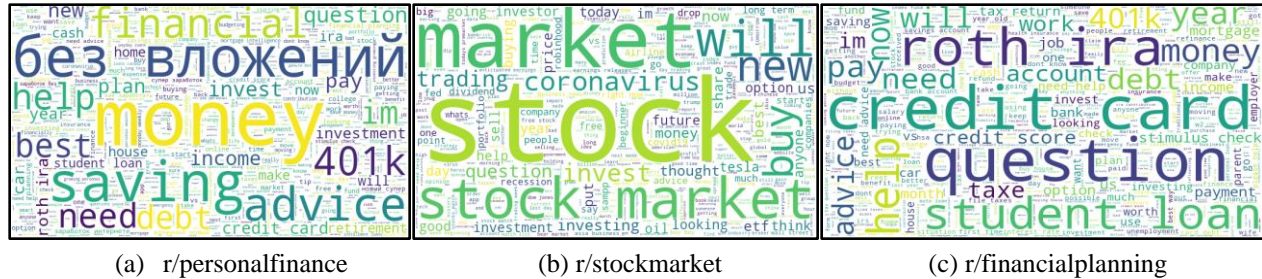
| Date | 2020-02-28 | 2020-03-30 | 2020-04-29 |
|------|------------|------------|------------|
| #    | 12493      | 11884      | 11453      |

## 5.4 Most Frequent Topics Concern Personal Finance Advice

We plotted the 10 most frequent words in the complete Reddit dataset consisting of English stop words only and initialized a count vectorizer. Credit, ira, help, and money are among the most common words occurring in the resulting plot. Comparing the plot result with the word clouds below, we observe the words occurring in the whole dataset align more with `r/financialplanning` as compared to `r/stockmarket` and `r/personalfinance`. This is an unusual result reasonably considering that the number of posts from `r/financialplanning` is the least amongst the three.

**Figure 7: Ten most common words for Reddit data overall**

We used three word clouds to represent the frequency of tokenized words in posts on the three subreddits in our study. WordCloud is a data visualization technique used for representing text data in which the size of each word typically indicates its frequency or importance. We visualized the texts from the three subreddits under study with the help of a python library called 'wordcloud.' From `r/personalfinance`, we observe words like debt, saving, advice which are closely related to that domain. The `r/stockmarket` presents market-related words like stock, market, trading, today. In `r/financialplanning` subreddit, we observe terms related to investment, retirement, loan, etc. expressing the steps one would need during a planning phase.



**Figure 8: Word clouds for most common words for Reddit for each subreddit (need to be specific which one is which)**

## 5.5 Wide Range of Financial Topics Dominate Discussion

One of our primary findings was the range of topics discussed on both Reddit and Twitter. Since we wanted a more holistic set of topics across our datasets, we performed two topic modelling analysis - one for the Reddit dataset and one for the Twitter dataset separately. The topics found using LDA on Reddit are represented in Table 7 and those for Twitter are in Table 8. We then assigned a description to each set of words based on our analysis to create a high-level topic that connects the output words from the LDA. Most topics of discussion on Reddit are found to be around personal finance discussions, possibly due to the highest user activity and high engagement in these forums. Topic 4 (for Reddit) in Table 7 is the only one that is related to stock market investment. Topic 8 (Table 7) is especially interesting because it pertains specifically to this time of the year because it relates to financial questions regarding unemployment benefits and government stimulus checks, both of which have had policy discussions recently due to the ongoing COVID-19 pandemic.

**Table 7: Topics found via LDA on Reddit**

| Topic # | Words   | Description                                    |
|---------|---|--|
| #0      | Credit Card Account Bank Score Cards Balance Transfer<br>Checking Accounts    | Personal banking                               |
| #1      | Ira Roth 401k Savings Account Retirement Vs Funds Fund<br>Question            | Retirement planning and safe investment        |
| #2      | Advice Best Financial Looking Way Money Investing<br>College Invest Whats     | Non traditional investment (college savings)   |
| #3      | Insurance Job Year Old Years Company Life Money Im<br>Employer                | Employment and long term savings               |
| #4      | Stock Market 2020 Question Business Stocks Week<br>Coronavirus Small Thoughts | Stock market discussions                       |
| #5      | Debt Personal Pay Questions Collections Finance Medical<br>Report Free Lease  | Unforeseen/pessimistic financial circumstances |
| #6      | House Time Car Buy Home Buying Money Dont Good<br>Know                        | Home and car investment                        |
| #7      | Help Need Taxes Income Advice Tax File Filing State<br>Form                   | Knowledge related to filing income tax         |

|    |  |   |
|----|--|---|
| #8 | Tax Check Stimulus Unemployment Taxes Return Work 2019 Refund Question | Tax returns and unemployment/stimulus payment questions |
| #9 | Loan Student Mortgage Pay Loans Refinance Car Rate Paying Salary       | Advice related to timely loan payments                  |

Similar to Table 7, we performed LDA for modelling the discussion topics on the Twitter dataset as well. Unlike Reddit, where discussions were more topic focused, Twitter users seem to cover a wider range of financial topics from government policy and comments on economic crisis to everyday financial necessities. Topics on Twitter seem to be more political in the way that it returns not only to financial advice, but also current policy-making in the news that pertain to finance in the United States federal government [18]. Given that a large number of Twitter users reside in the United States<sup>8</sup> (around 64%), it is unsurprising that the topics of discussion would largely be dominated by financial news in the United States. Table 8 shows the complete list of topics obtained through LDA on tweets.

We refrained from performing a topic modelling on YouTube because users tend to generally express their opinions on videos (Figure 8), more than engaging in extended topic discussions.

**Table 8: Topics found via LDA on Twitter**

| Topic # | Words   | Description   |
|---------|---|---|
| #0      | Said Like Trump Going Dont Good Mark Just Time New                | Governmental policy   |
| #1      | Paper Toilet Lol Buy Water Oil Food Going Buying Store            | Basic necessities   |
| #2      | Like Thank Hes Got Time Guys Great Right Just New                 | Positive attitude<br>*It is unclear what the specific topic would be in this case |
| #3      | Im Dont Y2k People Dave Just Going Like Youre Need                | Financial crisis in the 2000 decade   |
| #4      | People Money Dont Pay Rich Just Work Save Need Going              | Return to work to earn  |
| #5      | Flu People Virus Just Ray Like Rate China Italy World             | Public health in foreign countries  |
| #6      | God Like Right Cnn Yes Just Pointing Man White Dude               | News  |
| #7      | Time People Thanks Great Day Know Capitalism Just Bitcoin Economy | Cryptocurrency and alternate economy  |
| #8      | Market Stock Money Trump Stocks Buy Sell Years Time Markets       | Stock market impacts  |
| #9      | People Virus Dont Dave Panic Just Good Like Going Know            | Impact of pandemic on panic purchase  |

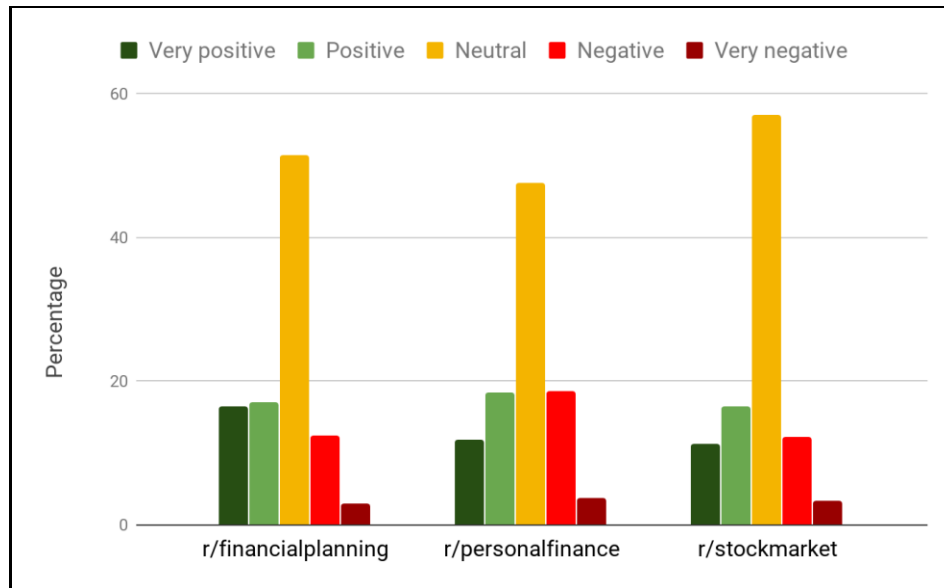
## 5.6 Sentiments Vary Across Platforms

Sentiments vary on platforms, depending on the type of social media platform. For example, as seen in Figure 9, the maximum number of posts on all three subreddits are neutral in tone, and generally lean more towards

<sup>8</sup> <https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/>

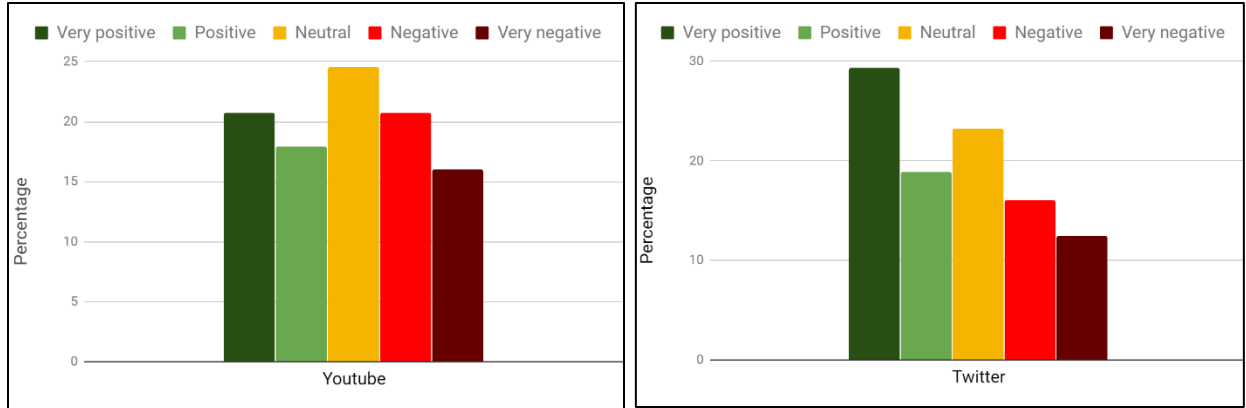
positive for the remaining posts. This is probably because Reddit is used as discussion forums where users engage in extensive discussions of articles without expressing too strong opinions. More often, posts are presented as new articles open for neutral debate rather than expressing emotions on these articles.

YouTube and Twitter on the other hand, are more open forums, used by people to express opinions and thoughts on real-time events. In the case of YouTube, these “events” are the content of the videos presented to users. As can be seen in Figure 8 (left), there is a wide distribution of expressed sentiments, spread out evenly between positive, neutral, and negative sentiments. It is surprising to note, however, that more people on YouTube tend to “like” rather than “dislike” a video. In Table 2, it can be observed that the number of likes exceed the number of dislikes for each of the 20 videos, but the sentiment of comments on YouTube are divided. This is probably because people express their thoughts on videos in comments, which can be either favorable or unfavorable, and sometimes even neutral. In the case of Twitter, there are more positive sentiments expressed in financial discussions (Figure 10; right). While it is difficult to say why this happens, we can conclude that sentiments on Twitter weigh heavily on one side, positive in case of financial tweets.



|               | r/financialplanning | r/personalfinance | r/stockmarket |
|---------------|---------------------|-------------------|---------------|
| Very positive | 433                 | 7954              | 640           |
| Positive      | 445                 | 12316             | 940           |
| Neutral       | 1352                | 31860             | 3265          |
| Negative      | 324                 | 12452             | 691           |
| Very negative | 74                  | 2418              | 183           |
| <b>Total</b>  | <b>2628</b>         | <b>67000</b>      | <b>5719</b>   |

**Figure 9: Sentiment analysis of subreddits expressed in percentages of total number of posts. The actual number of posts for each sentiment is provided in the table attached to the figure.**



|               | YouTube      | Twitter       |
|---------------|--------------|---------------|
| Very positive | 2662         | 26784         |
| Positive      | 2305         | 21931         |
| Neutral       | 3157         | 33197         |
| Negative      | 2669         | 18466         |
| Very negative | 2056         | 11870         |
| <b>Total</b>  | <b>12849</b> | <b>112248</b> |

**Figure 10: Sentiment analysis of YouTube comments (left) and tweets (right) expressed in percentages of total number of comments and tweets respectively. The actual number of posts for each sentiment is provided in the table attached to the figure.**

## 6 Discussion and Conclusion

The goal of our study was to see the effect of financial fluctuations on social media posts in terms of user activity, range of discussions, and expressed sentiments. We found that user activity was indeed correlated with stock market fluctuations for discussions related to stock markets, but personal finance discussions were not affected by it. Furthermore, social media is perhaps perceived to be useful because of the range of topics that users engage in. Finally, expressed sentiments were neutral for discussion forums, but polarized when people could present their opinions through either tweets or comments. While the reasoning behind such social network behavior is unclear, social influence likely plays a role [8]. This asserts the importance of social media discussions in being reflective of current financial news and role in financial decision making.

## 7 Limitations and Future Work

While we have explored only three social media platforms, it can be useful to see the effects of financial discussions across different social media platforms that people engage in. Furthermore, we were restricted by rate limits on all three platforms. A more comprehensive data collection in future would provide better insight into discussion topics and present a more holistic picture of expressed sentiments than our limited dataset. The comprehensive data would also benefit from extensive search criteria including more keywords pertaining to all aspects of finance. Future work could also include a longitudinal study (social media posts) over a longer period (possibly years), to see the historic effects of stock market fluctuations.

## REFERENCES

- [1] Li, X., Xie, H., Chen, L., Wang, J., & Deng, X. (2014). News impact on stock price return via sentiment analysis. *Knowledge-Based Systems*, 69, 14-23.
- [2] Bollen, J., Mao, H. and Zeng, X. Twitter mood predicts the stock market. *J. Computational Science* 2, 1 (2011), 1–8.
- [3] Schumaker, R. P., & Chen, H. (2009). Textual analysis of stock market prediction using breaking financial news: The AZFin text system. *ACM Transactions on Information Systems (TOIS)*, 27(2), 1-19.
- [4] Hutchinson, A. (2018, April). Reddit Now Has as Many Users as Twitter, and Far Higher Engagement Rates. *Social Media Today*. Available at <https://www.socialmediatoday.com/news/reddit-now-has-as-many-users-as-twitter-and-far-higher-engagement-rates/521789/> (Accessed on 05/01/2020)
- [5] Abisheva, A., Garimella, V. R. K., Garcia, D., & Weber, I. (2014, February). Who watches (and shares) what on youtube? and when? using twitter to understand youtube viewership. In *Proceedings of the 7th ACM international conference on Web search and data mining* (pp. 593-602).
- [6] Mohsin, M. (2019, November). 10 YouTube Stats Every Marketer Should Know in 2020. *Oberlo*. Available at <https://www.oberlo.com/blog/youtube-statistics> (Accessed on 05/01/2020)
- [7] Malhotra, Y., & Galletta, D. F. (1999, January). Extending the technology acceptance model to account for social influence: Theoretical bases and empirical validation. In *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences*. 1999. HICSS-32. Abstracts and CD-ROM of Full Papers (pp. 14-pp). IEEE.
- [8] Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly*, 425-478.
- [9] Sprenger, T. O., Tumasjan, A., Sandner, P. G., & Welpe, I. M. (2014). Tweets and trades: The information content of stock microblogs. *European Financial Management*, 20(5), 926-957.
- [10] Ruiz, E. J., Hristidis, V., Castillo, C., Gionis, A., & Jaimes, A. (2012, February). Correlating financial time series with micro-blogging activity. In *Proceedings of the fifth ACM international conference on Web search and data mining* (pp. 513-522).
- [11] Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., & Mozetič, I. (2015). The effects of Twitter sentiment on stock price returns. *PloS one*, 10(9).
- [12] Nguyen, T. H., Shirai, K., & Velcin, J. (2015). Sentiment analysis on social media for stock movement prediction. *Expert Systems with Applications*, 42(24), 9603-9611.
- [13] Souza, T. T. P., Kolchyna, O., Treleven, P. C., & Aste, T. (2015). Twitter sentiment analysis applied to finance: A case study in the retail industry. *arXiv preprint arXiv:1507.00784*.
- [14] Hazari, S., & Richards, A. R. A. (2011). A qualitative study of adoption of social media for personal finance investing. In *Proceedings of the Southern Association for Information Systems Conference*, Atlanta, GA. Retrieved from <http://sais.aisnet.org/2011/HazariRichards.pdf>.
- [15] Bukovina, J. (2016). Social media big data and capital markets—An overview. *Journal of Behavioral and Experimental Finance*, 11, 18-26.
- [16] Way, W. L., Wong, N., & Gibbons, D. (2011). *Online talk about money: An investigation of interactions around personal finance in social media*. Madison, WI: Center for Financial Security.
- [17] Hutto, C. J., & Gilbert, E. (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth international AAAI conference on weblogs and social media*.
- [18] Cheng, J., Skidmore, D., and Wessel, D. (April 2020). What's the Fed doing in response to the COVID-19 crisis? What more could it do?. *Brookings Institute*. Available at <https://www.brookings.edu/research/fed-response-to-covid19/> (Accessed on 05/01/2020)