The Impact of Subreddit Comments on Daily Return and Volume

Joey Wong

The Impact of Subreddit Comments on Daily Return and Volume

Joey Wong

Finance

Thesis Advisor: Matthew Morey

Presentation Date: April 28, 2021

Graduation Date: May 17, 2021

Abstract

This paper examines the subreddit WallStreetBets and its online weekday posts discussing stock market trades. By collecting data from the subreddit and Yahoo Finance, linear regression models are done using Rstudio to determine the relationship and impact that the number of comments has on percentage change in return and volume. This study concludes that when it comes to the relationship between returns and the number of comments, it is insignificant. However, we find that trading volume is positively and significantly related to the number of comments.

Table of Contents

Introduction

Because of the COVID-19 Pandemic, the world economy came to a screeching halt. This in turn led to necessary, unprecedented, stimulus recovery packages. Variations of stimulus packages were passed by governments around the globe, giving citizens the cash needed to provide for their families and/or keep their small businesses running. In April 2020, the U.S. government provided those that met the requirements, a $1,200 stimulus check, along with an additional $600 to unemployment checks (Bayer, 2020, pg. 1). It was almost as if the government was throwing out free cash to everybody. With the U.S. Federal Reserve printing money in order to sustain the economy through these stimulus packages, many people successfully got their checks (Bayer, 2020, pg. 1). However, a staggering amount turned to the stock market with them in hand. In the past few months, with the combination of fiscal and monetary stimulus, as well as pandemic lockdowns, people have become interested in investing and trading, primarily on Robinhood (Barber, 2020, pg. 3). With a substantial amount of cash but no market or trading experience, many people turned to social media to find groups and forums where people discuss market trades and ideas. One of the larger, more notable groups is Reddit's subreddit, WallStreetBets. This singular group was chosen because of the international attention it had received due to the GameStop frenzy that it fueled. As reported by The New York Times, GameStop's market value increased to "$24 billion from $2 billion in a matter of days" (Phillips, 2021). "Its shares have risen over 1,700 percent since December". Between two days, the market value rose over $10 billion. WallStreetBets, with already a million users at the start of the pandemic, had proven to be a go-to group for new investors, or for those who want quick, big payoffs in exchange for larger risk (Bolyston, 2021, pg. 4). With constant activity and daily posts

titled "What Are Your Moves Tomorrow", WallStreetBets seems like not only the ideal place for new traders to come and learn something, but also for those who are unsure of what to trade and want to jump in with a crowd.

For the purpose of this study, we will be focusing mainly on WallStreetBet's daily "What Are Your Moves Tomorrow" weekday posts. In addition to that, we want our time frame to focus on the beginning of 2020 before the pandemic, to right before the start of the GameStop frenzy that WallStreetBets caused. This will prevent the data from being skewed and will allow us to look at activity shortly before and during the pandemic. Aside from the attention WallStreetBets had received, the group was chosen because it had the most amount of users for a trading and investing community, as well as daily threads/posts on trade ideas. It is not private either and anyone, you don't even need to be a member, can browse through the website. Reddit's platform also made it considerably easy to collect data while utilizing Rstudio.

Overall, this study takes a look to see if the government stimulus, accessibility of Robinhood, and the sheer number of people on WallStreetBets caused any sway in the stock markets. If this research produces any significant results, we will be able to better understand how social media, large groups, and monetary stimulus plays a role in financial markets.

<u>Literature Review</u>

For this literature review, we will focus first on explaining what the WallStreetBets community actually is, giving a better sense of who they are as a group. Next, we will explain establishing a common understanding of Robinhood. After that, we will go into defining specific terms needed for this topic. It will be more focused on recent trends in 2020. Next, we'll take a

closer look at the Robinhood crowd and their effects. Lastly, we will discuss market movements based on retail investors.

What Is WallStreetBets?

WallStreetBets is a subreddit on Reddit that focuses on investment and trading discussions. The group gained notoriety during the GameStop frenzy in late January 2021. During this time, GameStop, as well as some other companies, experienced a surge in popularity, stock price, and market value. GameStop's market value went from $2 billion to $24 billion in days (Philips, 2020). BlackBerry and AMC stock have also seen a surge in price, increasing 280 percent and even 840 percent respectively. This event has been called a "David and Goliath" type scenario: "pitting plucky upstarts against well-heeled Wall Streeters" and beating them at their own game. Prior to the frenzy and international attention that it garnered, WallStreetBets was a very niche community. They have their own slang and term words, constructively berate each other, and have strong loyalty. According to Boylston et al. (2021) the community is quite chaotic, offensive, yet with "one of the most loyal user bases on the platform." Around back in March 2020, the community had 1.1 million users (Boylston et al., 2021, pg. 3). Understandably, the number has grown tremendously, now with about 9 million users. The demographics are not surprising. In Boylston et al.'s paper, 13 members of WallStreetBets were interviewed. All 13 of them identified as male and are considered to be Millennials and Gen Z. Their ages ranged from 21 to 38, with the average being around mid-20s. The majority of these users came from the United States, with a small fraction being residents of other countries. A notable point was that many of these 13 users work or have formerly worked in corporate jobs. Some identified as engineers, programmers, workers in the financial industry, and entrepreneurs. Another user who

isn't working a corporate job is one of the younger users at 24 and works as a grocery store employee.

It's a chaotic community that loves to trade extremely risky options, primarily on Robinhood, a commission-free trading app. To clarify, options are contracts that give the "owner the right to buy or sell an asset at an agreed upon price within a specified time period" (Boylston, 2021, pg. 3). One important factor that makes option contracts risky is that they are much more sensitive to price changes. Because of this, holders and traders are able to amplify their profit… or loss. For example, "a one dollar decrease in the price of a stock could mean a $100+ loss to someone holding a call option" (Boylston et al., 2021, pg. 4). Options are considerably riskier than buying and selling shares of a company since there are many more such factors that go into options trading compared to stock trading. Many of these factors are unknown or unfamiliar to retail investors. Despite this, the WallStreetBets community loves options trading. In fact, more of the behavior that is depicted within the community more so resembles "gamblers rather than investors" (Boylston, 2021, pg. 4). Community members praise and commend big risks for big rewards. As Boylston's paper puts it, it is like buying a lottery ticket "in that they most often result in losing the cost of the contract, but sometimes payout handsomely" (Boylston, 2021, pg. 4). Oftentimes, you will see posts containing user gains and/or losses, with the numbers being quite staggering; the WallStreetBets community absolutely loves to see this. As a founder of WallStreetBets had put it, the forum is populated with users of "different levels of risk tolerance" and experiences, with some people that are "'completely new' and have 'no concept of risk whatsoever'" (Boylston, 2021, pg. 13).

As mentioned previously, the go-to platform for WallStreetBets users, as well as other beginner traders, is Robinhood. Due to the commission-free trading and accessibility, Robinhood

is a popular platform for younger people to flock to. Something to note is that since options are risky investments, some brokerages do not allow regular retail investors to use them. Robinhood has almost no such barrier. A user and member of WallStreetBets that Boylston et al. had interviewed mentioned that what attracted him to Robinhood was the commission-free *options* trading (2021, pg. 7). With the ease of accessibility and commission-free trading on even options, Robinhood had proved to be the perfect tool to match the same level of risk and carelessness WallStreetBets loves.

Understanding Robinhood

Trading in the stock market has become more and more accessible to Main Street over the years. There have been many traditional brokerage firms that helped fuel this, such as TD Ameritrade and Charles Schwab. However, online retail brokerage companies, those without the standard brick and mortar office locations, have been stealing the spotlight. Robinhood, the notorious online brokerage that is easily accessible and usable through an app on your phone, comes to most minds. Founded in 2013 in Silicon Valley, it's an app that offers "people the ability to invest in stocks, ETFs, and options through Robinhood Financial" (Welch, 2020, pg. 8). Bitcoin is also accessible through the app through "Robinhood Crypto." Robinhood Financial, LLC and Robinhood Crypto, LLC are merely subsidiaries under the Robinhood name. Trading through an app with no physical office seems unreliable and risky at the least, but Robinhood has proved that it's not some sort of scam: Robinhood is a FINRA regulated broker-dealer and is registered with the U.S. Securities and Exchange Commission (SEC). There are a few reasons as to why Robinhood has experienced enormous popularity over the years: It has a very friendly mobile user interface, easy access, a lot fewer restrictions than your typical

brokerage, Bitcoin access, no investment minimum, fractional share purchasing, and most important of all, commission-free trading. Commission-free trading was probably the biggest highlight for Robinhood since around the time of its launch, charging clients brokerage fees for trades was the industry standard. It is quite arguable that this single-handedly helped Robinhood's rapid growth and "led other brokers to abandon brokerage fees by October 2019" (Welch, 2020, pg. 9). Many people have asked the simple question: "How does Robinhood make money if it doesn't charge fees?" It's simple— their main source of revenue comes from selling order information to high-frequency traders, margin lending, and interest earned on customers' cash balances (Welch, 2020, pg. 8). Once on the app, Robinhood does an exceptional job at keeping things simple. Its presentation of information, free stock when referred by a friend, and easy-to-understand charts made it the better choice for young investors. The company had actually intended the app to "make investing more like a game", as quoted in Barber, 2020. It certainly had a game-like feel, when "new members were given a free share of stock, but only after they scratched off images that looked like a lottery ticket" (Barber, 2020, pg. 1). It is no surprise here that this mixed with WallStreetBet's gambling nature, makes a great pair. There's also the fact that Robinhood provides considerably less stock information than rival brokers. For example, TD Ameritrade provides "over 400 charting indicators for each stock, while Robinhood provides five" (Barber, 2020, pg. 1). It's quite clear how Robinhood has not only been separating itself from the competition but has also done so in a way that really caters to young novice investors. As mentioned in Barber et al. (2020), back in 1971, "an investor had to call her broker during business hours to place a trade. She paid a minimum commission of $160.50 to purchase 100 shares of a $10 stock. Today, she can buy a fractional share, commission-free, day or night with a few clicks on her smartphone" (Barber, 2020, pg. 1).

In 2020, Robinhood had an exceptional year in terms of user growth and popularity, as well as some legal troubles. Robinhood had recently been accused of misleading customers about the true cost of trading with the firm. On the bright side, "As of mid-2020, RH had attracted a clientele of over 13 million investors" (Welch, 2020, pg. 2). To put it in better context, Robinhood users grew "from one million in 2016 to 13 million in May 2020" (Barber, 2020, pg. 3). This eclipses Charles Schwab, which has 12.7 million users and E-Trade, which had 5.5 million users at the end of 2019. Interest-free trading may not be the reason for such a high increase in users this time. Covid-19 and government stimulus are really the ones adding to the fire. By September of 2020, it was reported by WSJ based on Robinhood sources that "first time investors accounted for 1.5 million of its 3 million funded accounts opened in the first four months of 2020" (Welch, 2020, pg. 2). Basically, about 50% of those joining Robinhood in just the first third of 2020 had no experience trading and most likely relied on government COVID relief checks and unemployment. In other words, these new pandemic traders were relying on quantitative easing, or Q.E.

Quantitative Easing and Government Stimulus Packages

Quantitative easing is a "non-standard monetary policy", which "refers to a broader suite of instruments including negative nominal interest rates and lending to financial institutions" (Bartkiewicz, 2020, pg. 1). More understandably, Q.E. is "the asset purchase programs [leading] to a sizable increase in central bank assets and, working through various channels, eased financial conditions, raised output and inflation" (Bartkiewicz, 2020, pg. 1). In just this year alone because of the pandemic, both the Federal Reserve and U.S. Government have announced unprecedented monetary and fiscal policies. This is all in order to help alleviate the economic

fallout we were set to face as COVID-19 was spreading worldwide. To put things in perspective, by mid-March 2020, "17 million initial claims to unemployment benefits have been filed in the U.S" (Bayer, 2020, pg. 1). In the last period of recorded unemployment benefit claims, "a typical week saw some 250,000 new filings" (Bayer, 2020, pg. 1). It was also around this time that "observers suggested that the unemployment rate may reach 30% in the second quarter of 2020". 30% is an unbelievably staggering amount, considering the fact that the unemployment rate during the Great Recession in 2008 was around 10% and during the Great Depression of the early 1930s, around 25%. It is also worth noting that the VIX, the Volatility Index which is commonly used to measure risk, had spiked during this time. "By March 21, it reached an all-time high, surpassing levels previously seen during the financial crisis in 2008" (Bayer, 2020, pg. 1).

So what was the response to the potential economic fallout? A massive record-breaking stimulus package. By late March, the Coronavirus Aid, Relief, and Economic Security (CARES) Act was signed by President Trump. "As a result, $2 trillion of federal funds are being disbursed to households and firms through various channels" (Bayer, 2020, pg. 1). These are where the one-time payment checks of $1,200 came from, for those with a gross income of $75,000 or less. In addition to these checks, an additional $600 was being added to states' weekly unemployment benefits. In comparison, the fiscal stimulus response to the 2008 Financial Crisis was "$800 billion of additional federal spending in total", which was considered an incredibly high amount back then. The Federal Reserve has also been doing its fair share of work to make sure the economy stays afloat. Because of the potential damage, "The Federal Reserve announced as many emergency programs in eight days (March 14 to March 23) as it did during all of 2008" (Mosser, 2020, pg. 1). The Federal Reserve, as well as almost every other central bank in the

world when facing a recession, "sharply cut policy interest rates" (Mosser, 2020, pg. 2). Aside from the stimulus portion of quantitative easing, the Federal Reserve's move to purchase assets was extreme. "On March 15, the FOMC (Federal Open Market Committee) announced asset purchases of $700 billion 'over the coming months', amounts that were somewhat larger than the overall pace of purchases during QE1 and QE2". In a clearer example of the scale of things, the New York Fed's open market desk "bought almost $40 billion of U.S. Treasuries in a single day on March 13". Within only 6 days, the same open market desk was "purchasing $75 billion of Treasuries per day: a pace that was maintained until April 1" (Mosser, 2020, pg. 2).

The effect of these fiscal and monetary responses was interesting, to say the least. In a paper by Narayan et al. 2020, they utilized time-series regression models and daily time series data to see how lockdowns, travel bans, and government stimulus packages had affected G7 countries' stock markets. The authors collected data from Datastream, a website that collects global financial and macro-economic data. The data spans from July 1, 2019 to April 16, 2020. Looking at the results of government stimulus (announcement and onset of said packages), it was reported that 3 of the 7 countries, Canada, the UK, and the US, had experienced positive stock returns, while the other 4 countries did not (Narayan, 2021, pg. 4). In conclusion for the paper, they had stated how lockdowns, travel bans, and government stimulus packages had all demonstrated, on aggregate, "a positive effect on the G7 country stock market excess returns" (Narayan, 2021, pg. 5). However, they had stopped short of "claiming that any specific policy was more effective".

The Robinhood Crowd

The Robinhood newcomers proved to be a force to be reckoned with. In response to the volatility and widespread in the market, it was shown that the Robinhood crowd did not panic. Instead, they actually embraced the chaos. "They added more positions four days after market increases or decreases, suggesting that they transferred funding to their [Robinhood] accounts in response to volatility" (Welch, 2020, pg. 23). Most notably during the massive drop in the market in March 2020, Robinhood users not only "increased their relative holdings in individual stocks when their stock prices increased or decreased greatly, but also increased their overall holdings". A study conducted by Welch mimicked the Robinhood crowd consensus. In this study, a 50 variable "kitchen-sink" regression containing variables like IPO date, alphabetic index of name and ticker, market caps, minimum or maximum to highs and lows, share prices and returns, financial statement variables (Welch, 2020, pg. 27) yielded interesting results. The portfolio performed "well in the cross-section, earning positive alphas with respect to the risk-free rate, the market-model, and the Fama-French 5-factor plus momentum model" (Welch, 2020, pg. 32). Simply put, these are a few of many criteria and different ways of judging how well the mimicked Robinhood portfolio performed.

In another study, by observing the impact of Robinhood's three outages, situations where the Robinhood app was down and needed repair, Barber et al. took the opportunities to see the impact the Robinhood crowd had on trading volume, compared to general retail investors, or individual, non-professional investors. Based on what they found, Robinhood users "have more concentrated buying and selling" (Barber, 2020, pg. 15). The concentrated buying was most likely attention-driven and heavily influenced by the easy and friendly Robinhood interface. But not all of the trading proved to be profitable for the Robinhood crowd. The more concentrated trading, the more "buy-side herding events that are usually followed by negative returns"

(Barber, 2020, pg. 23). The example provided showed that the "top 0.5% of stocks bought by Robinhood users each day experience negative average returns of approximately 5% over the next month". In other more extreme examples, herding events occurred and were followed by a "negative average return of approximately 9%". A great highlight from the paper by Barber et al. is that the simple focus of attention by many investors on a small number of stocks "can promote herding behavior that affects market returns and redounds to the investor's detriment".

Stock Movements by Retail Investors

In an interesting study, another one by Barber et al., they observe and analyze the movement of markets based on retail trades. In order to do this, they rely on "tick-by-tick" transaction data from different time periods and different institutions/exchanges, and to identify buyer or seller initiated trades. Institutions like The Institute for the Study of Securities Market and American Stock Exchange (AMEX) gives them data for 1983-1992. The New York Stock Exchange (NYSE) and NASDAQ shows them 1993 to 2000. Overall, 18 years of transactional data is used. The researchers identified whether trades were buyer or seller initiated "using a quote rule and a tick rule" (Barber, 2008, pg. 157). The quote rule simply identifies if buyers initiated the trade above the midpoint of the most recent bid-ask quote and if sellers initiate the trade below the midpoint. The tick rule "identifies a trade as buyer initiated if the trade price is above the last executed trade price and seller initiated if the trade price is below the last executed trade price". The study also observes trade size—categorizing size into five bins based on dollar amount. Small trades amount to $5,000 or less, acting as a proxy for retail investors, while large trades amount to $50,000 or more, acting as a proxy for institutional trades. The aim of this study was to compare retail to institutional trades by focusing on various correlations and trade

imbalances. Barber and the others concluded something interesting. Going back to the small

trades as a proxy for individual investors, "we find that the buyer-initiated (and seller-initiated)

trades of individual investors are highly correlated; that is, in any given month individual

investors systemically buy some stocks and sell others" (Barber, 2008, pg. 182). In addition to

that, individual investors were shown to buy (or sell) the same stocks one month as they did the

previous month. In other words, retail investors *are* in fact prone to herding, or following what

other investors are doing. When it comes to the big question, whether or not retail investors

move markets, it was reported to differ depending on the time lengths. For short horizons,

"evidence is consistent with noise trader models in which the buying (selling) of retail investors

push prices too high (low) leading to subsequent reversals" (Barber, 2008, pg. 182). More

specifically, weekly imbalances between buyer and sellers who initiated small trades are

"correlated with contemporaneous returns and forecast cross-sectional differences in returns for

the subsequent week" (Barber, 2008, pg. 182). Basically, retail trades indeed moved the stock

prices in the direction of their trades. For annual horizons, their evidence that retail investors

move prices was not as clear. What the researchers took note of was how in the years in which

retail investors are net buyers, small capitalization stock prices rise. When it came to medium

and large capitalization stock prices fell. They do, however, rise during "days or weeks of intense

retail investor buying", but there is a clear negative correlation between annual retail investor

buying and stock returns for medium and large stocks (Barber, 2008, pg. 182). The same could

be said vice versa: where retail investors are net sellers, small capitalization stock prices fall, and

so on. So for longer periods, retail trades have the ability to move only small capitalization

stocks in the direction of their trade. On another note, their research had highlighted that stocks

that were heavily bought underperformed stocks that were heavily sold by 4.4 percentage points

the following year. In their own words, "The quintile of stocks with the highest proportion of buyer-initiated small trades underperforms the quintile with the lowest proportion of buyer-initiated smaller trades by that amount". Barber and researchers concluded that "over both short and long horizons, retail trade imbalances forecast future returns" (Barber, 2008, pg. 183).

Research Question

Has the ubiquity of social media and the accessibility of Robinhood allowed young investors to disrupt the market? Are young, inexperienced traders on WallStreetBets impacting stock return and/or volume? The research conducted for this paper will attempt to answer these questions.

Data and Methodology

The data was collected primarily from Reddit's subreddit, WallStreetBets. Knowing that the majority of users were either new, inexperienced traders on Robinhood, we felt that this single source would be suitable. In addition to this, community activity was bustling, even before the GameStop frenzy of late January 2021. There are posts and threads ranging from all sorts of topics such as trade ideas, gain/loss returns on positions, market news, and of course, memes. Our main focus was on their daily discussions on trades. There are daily posts created by moderators that allow for members to discuss their trade ideas and moves for the next day, often titled "What Are Your Moves Tomorrow, (date)". Taking the number of comments per day proved to be enough data to compare against returns and volume. The dataset consists of 235

observations, or in other words, weekdays. Initially, we gathered 373 posts discussing trades. 250 posts were weekday discussions and 123 posts were weekend discussions. We decided to focus on weekdays because weekends typically generated substantially less attention and activity compared to discussions during the week. For clarity, weekend posts only generated about 2,000-5,000 comments, while weekday posts generated well over 10,000 comments consistently. This leads the dataset to have 250 posts. Since the dataset included activity from the GameStop (GME) frenzy, which started on the week of January 25th, 2021, it was decided to remove those days, since it completely skewed the data. Therefore, data starting on January 25, 2021 onward were removed. This resulted in having 15 posts removed, leaving 235 posts to use as the sample. The final dataset used for this thesis consists of data starting on February 20, 2020 and ending on January 25, 2021. This gives us an opportunity to see WSB and stock activity before and during the COVID pandemic, but also before WSB becomes nationally recognized for the GameStop mania and the number of comments significantly increased. Before the mania, daily trade discussions easily generated well over thousands of comments, giving the entire dataset an average of almost 20,000 comments. We felt that the data was substantial enough to use in hypothesis tests against stock returns and volume in 2020 and early 2021.

To collect our data from Reddit, we utilized an R package called "RedditExtractoR". This was our main tool for gathering data from WallStreetBets. This package includes a collection of functions and tools which allow the user to gather structured data from Reddit, specifically "subreddits", or Reddit pages for specific topics. The main functions of "RedditExtractoR" that we used were: "get_reddit", "reddit_content", and "reddit_urls". The predominant commands under these functions were: "search_terms", "subreddit", "cn_threshold", "page_threshold", and "sort_by". Understandably, these were the parameters that we were able to adjust and change to

gather the data we were looking for. We were able to input which specific subreddit we wanted to analyze, specific words and terms each post had to include, the minimum number of comments, and how many pages we wanted the R code to sift through.

Once the data from Reddit was collected, we gathered data on different indices and companies from Yahoo Finance. Using their "Historical Data" section, we were able to easily download data on open, highs and lows, closing, and volume. For this thesis, we were only focused on percent change in return, which was calculated using open and close, and volume. Percent change in return is simply how much change there was in the index or stock each day. Defined by Investopedia, Volume is "the amount an asset or security changes hands over some period of time". In terms of stock trading, volume refers to the number of shares a security has been traded within the day's open and close. To include the percent change in return for this thesis, two columns were created in addition to the data collected from Yahoo Finance: a "Change_Per" column, abbreviated for "Change in Percent" and a "Change_Poi" column, abbreviated for "Change in Points. "Change_Per" was calculated by taking the day's closing price (x), subtracting it by the previous day's closing price (y), then dividing the product by the previous day's closing price (y). Lastly, we multiplied by 100 for simplification. The "Change_Poi" column was simply used to see the change in closing price for the index or stock. For this, we took the day's closing price (x) and subtracted the previous day's closing price (y). For the linear regressions, we only used "Change_Per", or Change in Percent.

Our initial focus was the Vanguard Total Stock Market Index Fund (VTI). Because this exchange-traded fund (ETF) serves as a measure for the overall U.S. stock market, we felt that it would be great to analyze and test out first. Aside from just VTI, we felt that looking at specific stocks, especially the ones that experienced some short-lived hype or Robinhood attention, as

well as major indices, would give us an accurate idea of the effect online trade discussions had on the market. We also decided to look at the Volatility Index, commonly known as VIX, Tesla (TSLA), Nikola (NKLA), Hertz (HTZGQ), Kodak (KODK), NASDAQ, and S&P500. We chose the VIX because it is more of a metric to measure market risk and investor sentiment. Tesla was chosen because it had experienced significant growth and was extremely popular, if not the most popular stock in 2020. Nikola, Hertz, and Kodak were chosen because they received a lot of mainstream attention, whether for filing for bankruptcy or misleading investors. These seemed to be reasons for WallStreetBets users to flock to them. Also, these three were listed by Investopedia, a trusted online source for financial information, as some of the stocks that experienced a surge in popularity in 2020. Lastly, we wanted to use the NASDAQ and S&P500 because they are some of the largest and most popular indices. Just to note, we don't expect a significant impact on these two indices from the number of comments; we simply want to see what degree of impact there is, no matter how small.

When it came to cleaning the dataset, we matched Reddit posts to the following day's performance. Since Reddit posts were "What Are Your Moves *Tomorrow*", we matched a day's post and its number of comments with the market's next day results in price, return, and volume. For example, if a post was titled "What Are Your Moves Tomorrow, April 23, 2020", it would be aligned and compared to the market results of April 24, 2020, etc. This makes more sense and allows us to see the impact of these comments on the market the next day.

Our last step was to utilize the linear regression model in Rstudio using the lm() function. We took the percent change in return (dependent variable) and ran the regression model against the number of comments (independent variable). Once that step was done, we then switched out the percent change in return for volume, our other dependent variable. To reiterate, we are

running this regression model to explain the impact the number of comments a post has on certain indices and stocks.

Results and Discussion

We are interested in seeing the impact these WallStreetBets users' trade discussions had on certain indices and companies. We first take a look at the descriptive statistics on the numbers of comments. Then, we discuss findings in the linear regression model to find the impact number of comments has on percent change in return. Lastly, we discuss findings in a different linear regression model that looks at the impact number of comments has on volume.

Table 1a: Descriptive Statistics – WSB Discussion Post Comments (2/20/20 – 1/24/21)	
	Comments
Mean (Average of All Days)	19,529.31
Median	18,070
Minimum	9,200
Maximum	45,926
Standard Deviation	5,746.34
First Quantile	15,913
Third Quantile	21,952.50
Interquartile Range (IQR)	6,039.50
Sunday Average	18,424.02
Monday Average	18,733.20
Tuesday Average	20,433.38
Wednesday Average	19,078.52
Thursday Average	21,012.02

Table 1b: Descriptive Statistics – Averages		
Index / Stock	**Average Daily Return (%)**	**Average Volume**
VTI	0.09	4,941,482
S&P500	0.08	5,101,259,447
NASDAQ	0.17	4,384,305,830
^VIX	0.62	-
TSLA	0.82	67,607,380
NKLA	0.82	18,316,424
KODK	1.88	11,796,814
HTZGQ	0.75	33,900,585

On the surface level, we can take a look at the descriptive statistics in Table 1a of the data collected. To reiterate, the data consists of weekday "What Are Your Moves Tomorrow" posts from WallStreetBets. The average post generated about 19,000 comments from users, while the minimum and maximum number of comments recorded were just a little under 10,000 and over 45,000 comments, respectively. An interesting take from Table 1 is that the later days of the week generated more comments on average than days earlier in the week. It is a little surprising that Sunday comments were lowest on average despite being the day before the start of a trading week. It could be assumed that as trading sessions go on during the week, more and more users become interested in what people are saying within the community, possibly looking for reactions or comments on what has been happening. Despite this, all 5 days collected show a very concentrated comment average (see Appendix for graph displays of average number of comments, sorted by day).

In Table 1b, we are able to observe the averages of both percent return and volume. It is interesting to see that no index or stock had any huge change in percent return. The highest company that showed an average percent return turned out to be Kodak, but not by much. What is interesting to note from just looking at the average percent return column is that every single one of the seven indices and stocks listed is positive. So in other words, throughout the time of our dataset which spans from February 20, 2020 to January 25, 2021, all seven managed to average out a positive return. When it comes to the volume, it's extremely clear how indices differ from companies. The S&P500 and NASDAQ have over 5 billion and over 4 billion in volume, respectively. This is no surprise since they are literally an index that follows the performance of numerous companies. To note, this will actually impact their coefficients in the linear regression models. When it comes to the others, the volume is more reflective of what it would be for a single company. Unsurprisingly, Tesla has the most volume by far. This is most likely due to the fact that it was so popular among traders in 2020. VTI has the least in volume, most likely because it is an ETF, acting like a small index. Lastly, the VIX does not have any volume because it is just a measure of market risk and investor sentiment.
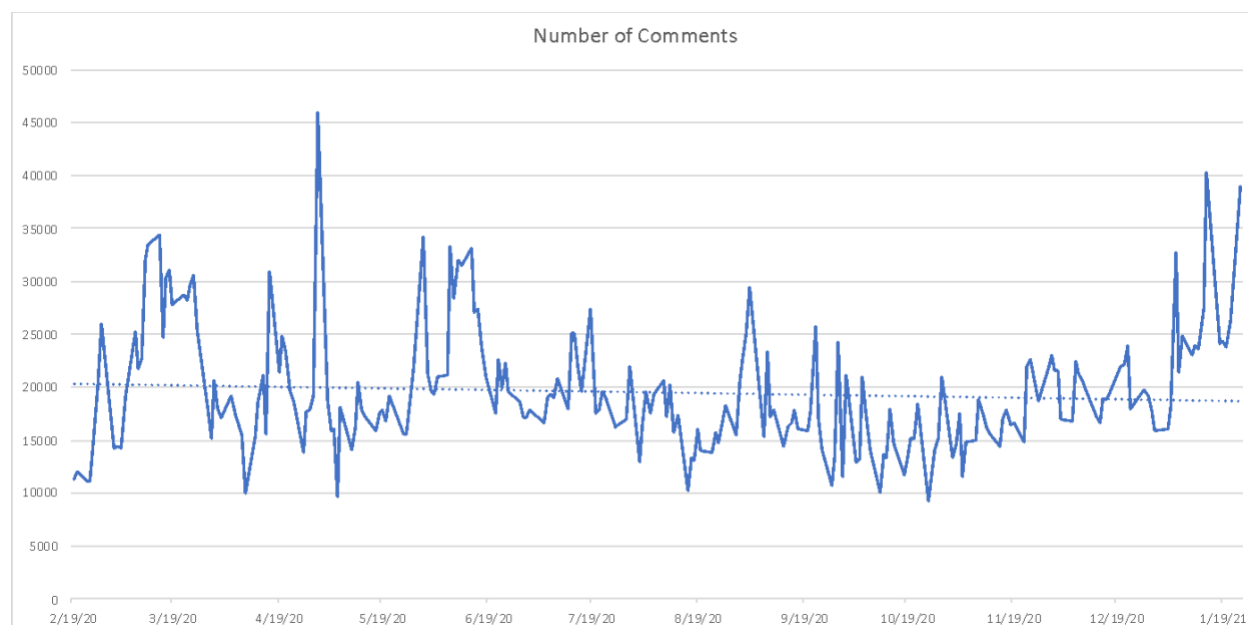
Figure 1

Figure 1 displays the number of comments throughout almost an entire year; this is our

full dataset. Included in the graph is a linear trendline, which holds steady but actually has a

negative slope. Some interesting highlights are March 2020, late April, and the beginning of

2021. In March 2020, the Covid-19 pandemic came to fruition and lockdowns were established.

It could be assumed that because of this, early users of WallStreetBets turned to the community

to discuss trades going forward. This was also the time that the stock market had dropped 10%,

"the Dow Jones' worst day since 1987", according to the Wall Street Journal (McCabe, 2020,

pg.1). In April 2020, the number of comments spiked to its highest level. This was also the time

when the first round of government stimulus checks came out, giving Americans that qualified,

$1,200 (Bayer, 2020, pg. 1). As mentioned previously, many people did turn to Robinhood and

investing as a result of getting money handed to them. The spike in comments on WallStreetBets

could possibly be a result of this as well. Lastly, there is the beginning of 2021. Although we

decided to exclude the GameStop mania, it is worth noting that attention was picking up here, as

seen with the spike in comments reaching the 40,000 mark.

Table 2 – Coefficients Relationship – ΔReturn(%) and Number of Comments
This table shows the regression relationship between a day's worth of comments (num_comments) vs the next day's **return** (change in percent). We are observing the relationships of how the number of comments posted will impact change in return. For daily return, we look at 8 different variables: Vanguard Total Stock Market Index Fund ETF (VTI), NASDAQ index, S&P500 index, Volatility Index (^VIX), Tesla (TSLA), Nikola Corp (NKLA), Eastman Kodak Company (KODK), and Hertz Global Holdings (HTZGQ).

(*** = 1% significance level, ** = 5%, * = 10%)

	VTI	NASDAQ	S&P	^VIX	TSLA	NKLA	KODK	HTZGQ
Intercept (Estimate)	0.009666	1.015	0.00909.	-2.1029	3.508	3.455	6.678	5.829
# of comments (Estimate)	-0.0000004*	-0.00004*	-0.00004*	0.0001394	-0.00014**	-0.00014	-0.00025	-0.00026
T-value of num_comments	-1.776	-1.654	-1.670	1.268	-2.242	-1.066	-0.829	-1.113
P-value	0.077	0.0996*	0.0962.	0.206	0.025*	0.287	0.408	0.267
R-Squared	0.013	0.0116	0.0118	0.0068	0.021	0.0048	0.0029	0.0053
Observations	235	235	235	235	235	235	235	235

Based on the linear regressions done on these 8 different indices/companies, it is pretty

clear that almost all of them result in a negative slope, as seen in the "# of comments (Estimate)"

column. In other words, this is an indication that the more comments a post generated, the worse

the change in return would be for that particular index or stock the next day. The only one that

did not show a negative slope in relationship to number of comments was the VIX. It could be

pondered that as more people commented on posts, they spoke out of fear and therefore felt the

need to sell, or buy more than usual. The p-values for all the linear regression models, except for

Tesla, are above the 0.05 significance level. Due to this, it is quite hard to actually state that a

relationship between the number of comments posted by WallStreetBets and a percent change in

return exists.

Table 3 – Coefficients Relationship – Volume and Number of Comments
This table shows the regression relationship between a day's worth of comments (num_comments) vs the next day's **volume** (change in percent). We are observing the relationships of how the number of comments posted will impact change in return. For daily return, we look at 8 different variables: Vanguard Total Stock Market Index Fund ETF (VTI), NASDAQ index, S&P500 index, Tesla (TSLA), Nikola Corp (NKLA), Eastman Kodak Company (KODK), and Hertz Global Holdings (HTZGQ). The Volatility Index (^VIX) is not included in this because it does not measure volume.

(* = 1% significance level, ** = 5%, * = 10%)**

	VTI	NASDAQ	S&P	TSLA	NKLA	KODK	HTZGQ
Intercept (Estimate)	796888.91	2.880e+09	3.283e+09	4.437e+07	22187577	25989784.2	2.021e+07
# of comments (Estimate)	212.22***	77040***	93120***	1190***	-198.2	-726.8*	700.8
T-value of num_comments	5.869	7.331	6.796	3.271	-0.790	-1.901	0.658
P-value	0.001	0.001	0.001	0.001	0.4304	0.0585	0.5112
R-Squared	0.1288	0.1874	0.1654	0.0439	0.00267	0.0153	0.00186
Observations	235	235	235	235	235	235	235

When it comes to the linear regressions using volume as the dependent variable, our

results are quite different from the linear regressions where the percent change is the dependent

variable. From a glance, there is more significance in the models; their p-values are much lower

than the p-values seen in regressions with percent change in return. This could be due to the fact

that volume will be impacted no matter if the discussions on WallStreetBets promote selling or

buying. Unlike percent change in return, where more negative comments will likely result in a

worse return, and vice versa, it won't be the same case with volume. If there is just generally

more discussion on a particular stock, the volume *will* be impacted.

Limitations

We did not use many of the other large social media platforms like Facebook and Twitter,

or even TikTok for this study. Although all these social media platforms are suitable sources to

see how much trading interest there is among newcomers online, we felt that Reddit's

WallStreetBets was the most effective and simplest when it came to observing dialogue and

collecting data between new traders. Not only that, but thanks to Reddit's ability to filter

upvotes, comments, and posts, analyzing the popularity and attention a post generated was

extremely convenient. Facebook has trading groups, but they are private. Even once in the group,

posts are stacked and there is no filter for posts, neither is there a way to organize posts. Twitter

may have been our second option after Reddit, but the tweeting system seemed to be

unorganized. Also, Twitter does not have specific posts/tweets dedicated to one question or

topic. TikTok may have been the least likely to be used to gather data, considering trade

discussions would have been in short video format or as comments on that video.

Another limitation we had was that the Rstudio package "RedditExtractoR" did not have

adjustable parameters on the date. Selecting specific dates for Rstudio to collect data could not

be done using code. This proved to be a problem, for if we wanted to go further back in time

with the data and posts, the code would have trouble doing so. In order to go back further in time, I was able to adjust the "page_threshold", which controls how many pages the program sifts through, according to the package's documentation, or manual, in hopes of reaching dates further in the past by asking the code for more dates. Despite doing this, the code would not always return more comments from older dates.

Conclusions

Our research showed us the results of linear regression models on the relationship and impact of the number of comments from WallStreetBet's "What Are Your Moves Tomorrow" on the percent change in return and volume for indices and stocks. To reiterate, our dataset was set up to reflect what a day's worth of comments would have on the next day in the market. Therefore, the results we see show impact.

Based on our results, we see that the relationship between the number of comments and percent change in return is insignificant, as well as negatively correlated. From our findings and data, it could be assumed that if there are more comments, it is a bad sign for the markets, since more comments equals less return. But again, the relationship proved to be insignificant. Volume, on the other hand, showed us something different. The correlation here was both strong and positive. This makes sense since if there is more talk of trade and investing, volume will increase. As mentioned previously, it does not matter whether or not comments are negative or positive towards the stock. As long as the stock or index is traded, volume will be impacted. However, when it comes to percent change in return, it does actually depend on whether or not

what was said was negative or positive. This, of course, will change what the percentage change comes out to be.

This research could lead to more findings on the relationship between WallStreetBets and social media's influence on the stock market. Understanding online communities and their view on the market could serve as potential forecasts and measures, especially after what has happened with the GameStop frenzy. For future research, a deep dive should be done into the thousands of comments themselves and determine user/trader sentiment. Utilizing a word count could also prove to be very useful. This would help with finding out which company ticker had been mentioned the most. Future studies on this topic could also branch out to other social media sites such as Facebook and Twitter, since people also discuss stock market trades on those platforms as well. Online discussions about trades and the stock market can only continue to grow, especially as we have become accustomed to staying inside and social distancing.

<div align="center">References</div>

Barber, B. M., Huang, X., Odean, T., & Schwarz, C. (2020). Attention induced trading and

    returns: evidence from robinhood users. *http://dx.doi.org/10.2139/ssrn.3715077*

Barber, B. M., Odean, T., & Zhu, N. (2009). Do retail trades move markets? *Review of Financial*

    Studies, 22(1), 151–186. https://doi-org.rlib.pace.edu/10.1093/rfs/hhn035

Bartkiewicz, P. (2020). Quantitative Easing: new normal or emergency measure? *Contemporary*

    Economics, *14*(3), 306.

Bayer, C., Born, B., Luetticke, R., & Müller, G. J. (2020). The coronavirus stimulus package:

    How large is the transfer multiplier?. CEPR Discussion Paper No.
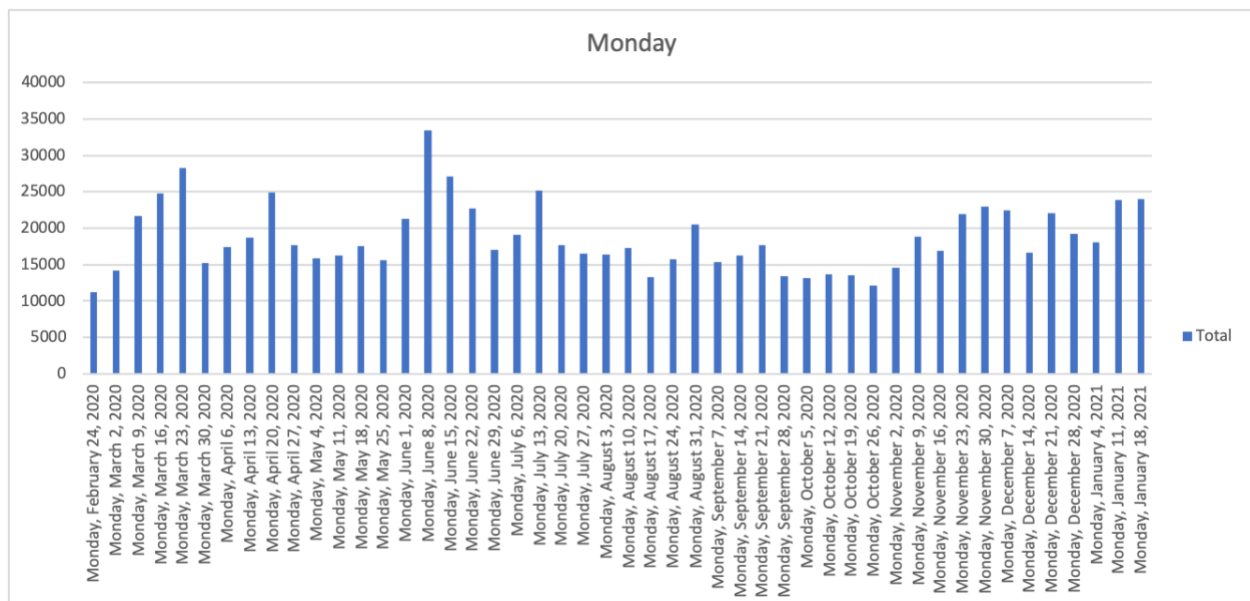
    DP14600. https://ssrn.com/abstract=3594222

Boylston, C., Palacios, B., Tassev, P., & Bruckman, A. (2021). WallStreetBets: Positions or

    ban. *arXiv preprint arXiv:2101.12110*. https://arxiv.org/abs/2101.12110

Hayes, A. (2021, April 9). *Volume Definition*. Investopedia.

    http://www.investopedia.com/terms/v/volume.asp.

McCabe, C., & Ostroff, C. (2020, March 12). *Stocks plunge 10% in DOW'S worst day since*

    1987. Wall Street Journal. https://www.wsj.com/articles/global-stocks-follow-u-s-markets-

    lower-11583975524.

Mosser, P. C. (2020). Central bank responses to COVID-19. *Business Economics*, *55*(4), 191-

    201. https://doi-org.rlib.pace.edu/10.1057/s11369-020-00189-x

Narayan, P. K., Phan, D. H. B., & Liu, G. (2021). COVID-19 lockdowns, stimulus packages,

    travel bans, and stock returns. *Finance research letters*, *38*, 101732.

    https://doi.org/10.1016/j.frl.2020.101732

Phillips, M., & Lorenz, T. (2021, January 27). *'Dumb money' is on Gamestop, and it's beating*

    Wall street at its own game. The New York Times.

    https://www.nytimes.com/2021/01/27/business/gamestop-wall-street-bets.html.

Welch, I. (2020). Retail raw: Wisdom of the Robinhood crowd and the COVID crisis (No.
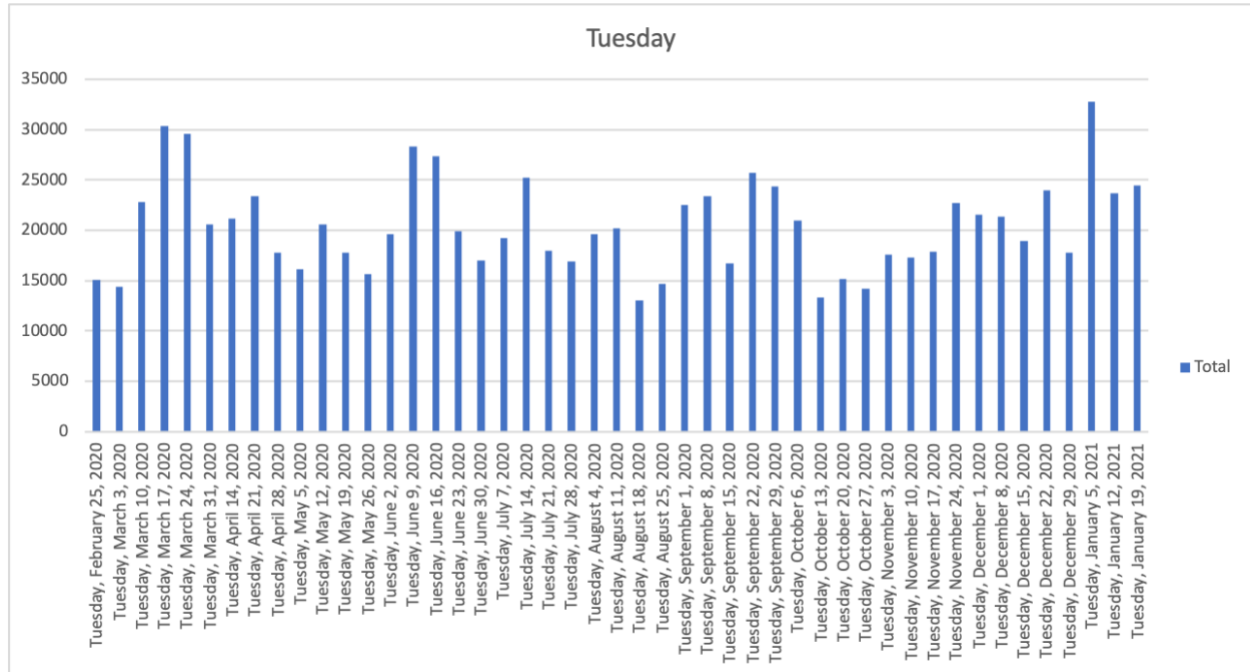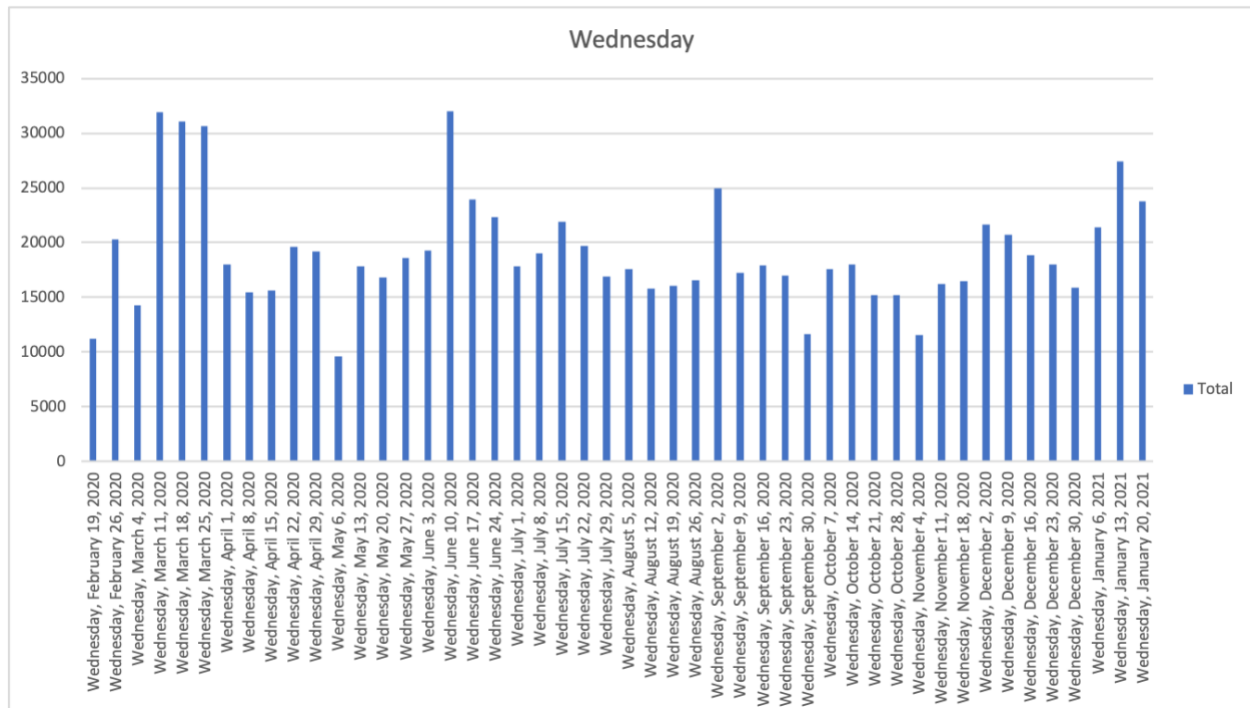
    w27866). *National Bureau of Economic Research.* http://www.nber.org/papers/w27866

Appendix

Sunday
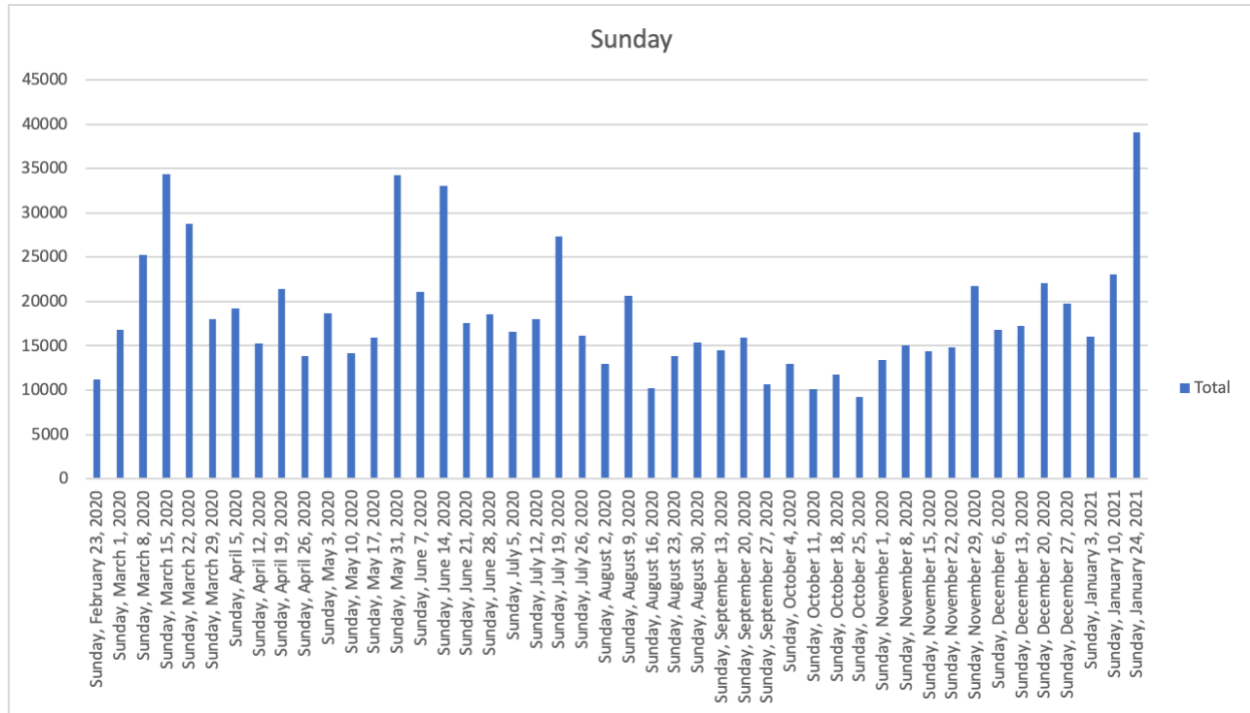
Monday

Tuesday

Wednesday

Thursday