CIS4120 NLP Project Group #1 Final Report

Sahiti Kovvuri, Amaris Efthimiou, Yarema Ganushchak, Taohid Shadat, Muhammad Megahed, Andrea Ramirez

---

FED VS THE MARKET

---

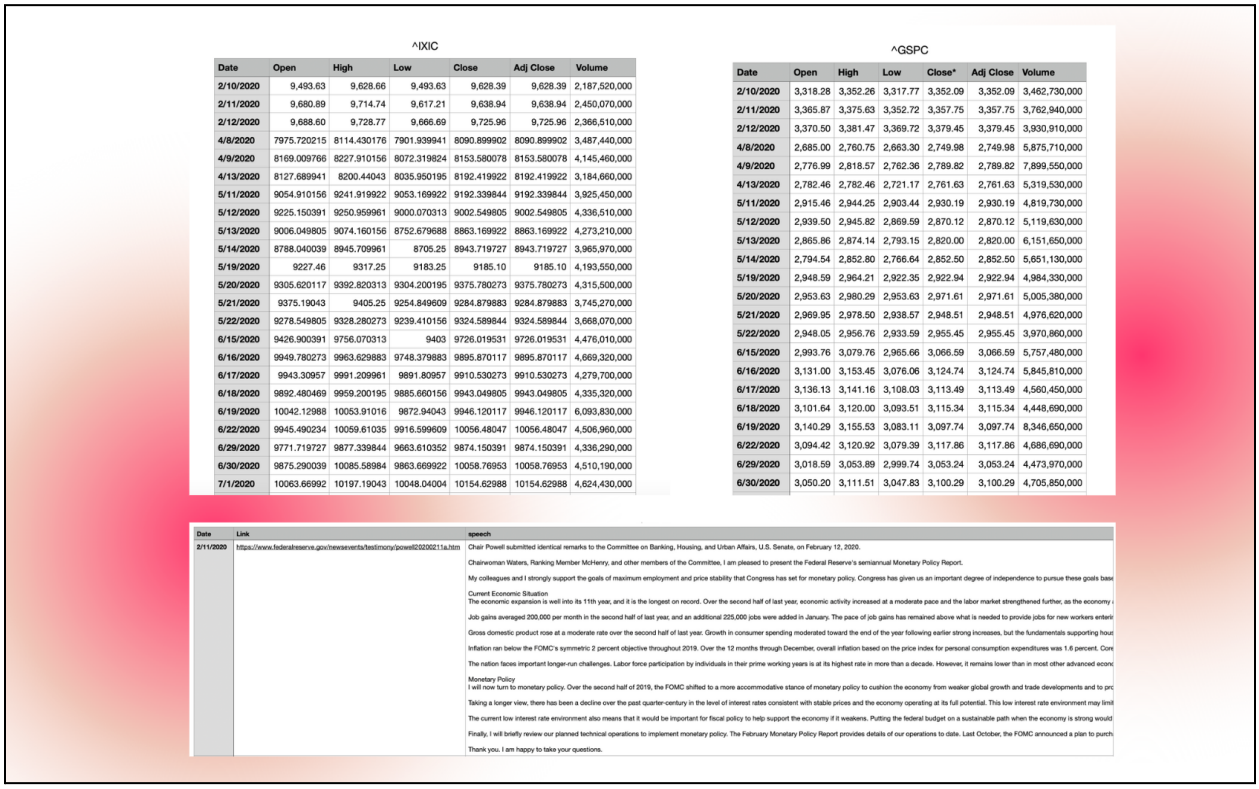
(1) Business Question/Problem

The Federal Reserve is the central bank of the United States and has an immense influence on the economy due to the responsibilities it upholds, including: promoting financial stability, supervising/regulating banks, and conducting monetary policy. It communicates policies, actions, and economic outlooks through various channels, including speeches by policymakers, reports, and press releases. Some speeches by the Federal Reserve Chair have the potential to impact financial markets, including the stock market. The stock market is a measure of economic health, and its performance is influenced by a variety of factors, including economic indicators, corporate earnings, and political events. In our project, we conducted a sentiment analysis between the speeches made by Jerome Powell, and stock prices from 2020. Some important measures of the entire stock market can be indexes, and for our project, we chose to analyze the S&P500 and NASDAQ. Specifically, the S&P500 represents the performance of 500 large publicly traded companies listed on stock exchanges, capturing about 80% coverage of the total US equity market. The NASDAQ, however, is an electronic stock exchange in the United States and primarily focuses on listing technology companies, but it also includes companies from various other sectors. Both indexes are ever-changing based on a committee's opinion on what is driving the U.S. economy, and the biggest weights in the indexes are usually the biggest or most influential stocks.

As students fascinated by the influence of global events on financial markets, we were particularly drawn to the historically significant year of 2020. This period, noted by significant market volatility due to the COVID-19 pandemic, presented a compelling case for study. With an interest in how programming techniques could help us with future career paths, we dove into the sentiment analysis of this period. By doing so, we aimed to understand the impact of Federal Reserve communications on market sentiment during such critical times, and discovered how this sentiment evolved in response to major events. This exploration not only provided valuable insights into the financial market dynamics but also demonstrated the significance of NLP techniques in understanding these effects.

(2) Data Set

The datasets we used included the S&P500 and NASDAQ[1] indices with the prices corresponding to the open, high, low, close, adjusted close, and trading volume for the day of the speech, as well as the days preceding and following the speech. Additionally, we used the transcripts of 14 speeches and testimonies[2] delivered by Jerome Powell between February and December 2020 which consisted of the date, link, and transcript of the speeches given by the Federal Open Market Committee (FOMC) in 2020.



^IXIC

Date	Open	High	Low	Close	Adj Close	Volume
2/10/2020	9,493.63	9,628.66	9,493.63	9,628.39	9,628.39	2,187,520,000
2/11/2020	9,680.89	9,714.74	9,617.21	9,638.94	9,638.94	2,450,070,000
2/12/2020	9,688.60	9,728.77	9,666.69	9,725.96	9,725.96	2,366,510,000
4/8/2020	7975.720215	8114.430176	7901.939941	8090.899902	8090.899902	3,487,440,000
4/9/2020	8169.009766	8227.910156	8072.319824	8153.580078	8153.580078	4,145,460,000
4/13/2020	8127.689941	8200.44043	8035.950195	8192.419922	8192.419922	3,184,660,000
5/11/2020	9054.910156	9241.919922	9053.169922	9192.339844	9192.339844	3,925,450,000
5/12/2020	9225.150391	9250.959961	9000.070313	9002.549805	9002.549805	4,336,510,000
5/13/2020	9006.049805	9074.160156	8752.679688	8863.169922	8863.169922	4,273,210,000
5/14/2020	8788.040039	8945.709961	8705.25	8943.719727	8943.719727	3,965,970,000
5/19/2020	9227.46	9317.25	9183.25	9185.10	9185.10	4,193,550,000
5/20/2020	9305.620117	9392.820313	9304.200195	9375.780273	9375.780273	4,315,500,000
5/21/2020	9375.19043	9405.25	9254.849609	9284.879883	9284.879883	3,745,270,000
5/22/2020	9278.549805	9328.280273	9239.410156	9324.589844	9324.589844	3,668,070,000
6/15/2020	9426.900391	9756.070313	9403	9726.019531	9726.019531	4,476,010,000
6/16/2020	9949.780273	9963.629883	9748.379883	9895.870117	9895.870117	4,669,320,000
6/17/2020	9943.30957	9991.209961	9891.80957	9910.530273	9910.530273	4,279,700,000
6/18/2020	9892.480469	9959.200195	9885.660156	9943.049805	9943.049805	4,335,320,000
6/19/2020	10042.12988	10053.91016	9872.94043	9946.120117	9946.120117	6,093,830,000
6/22/2020	9945.490234	10059.61035	9916.599609	10056.48047	10056.48047	4,506,960,000
6/29/2020	9771.719727	9877.339844	9663.610352	9874.150391	9874.150391	4,336,290,000
6/30/2020	9875.290039	10085.58984	9863.669922	10058.76953	10058.76953	4,510,190,000
7/1/2020	10063.66992	10197.19043	10048.04004	10154.62988	10154.62988	4,624,430,000

^GSPC

Date	Open	High	Low	Close*	Adj Close	Volume
2/10/2020	3,318.28	3,352.26	3,317.77	3,352.09	3,352.09	3,462,730,000
2/11/2020	3,365.87	3,375.63	3,352.72	3,357.75	3,357.75	3,762,940,000
2/12/2020	3,370.50	3,381.47	3,369.72	3,379.45	3,379.45	3,930,910,000
4/8/2020	2,685.00	2,760.75	2,663.30	2,749.98	2,749.98	5,875,710,000
4/9/2020	2,776.99	2,818.57	2,762.36	2,789.82	2,789.82	7,899,550,000
4/13/2020	2,782.46	2,782.46	2,721.17	2,761.63	2,761.63	5,319,530,000
5/11/2020	2,915.46	2,944.25	2,903.44	2,930.19	2,930.19	4,819,730,000
5/12/2020	2,939.50	2,945.82	2,869.59	2,870.12	2,870.12	5,119,630,000
5/13/2020	2,865.86	2,874.14	2,793.15	2,820.00	2,820.00	6,151,650,000
5/14/2020	2,794.54	2,852.80	2,766.64	2,852.50	2,852.50	5,651,130,000
5/19/2020	2,948.59	2,964.21	2,922.35	2,922.94	2,922.94	4,984,330,000
5/20/2020	2,953.63	2,980.29	2,953.63	2,971.61	2,971.61	5,005,380,000
5/21/2020	2,969.95	2,978.50	2,938.57	2,948.51	2,948.51	4,976,620,000
5/22/2020	2,948.05	2,956.76	2,933.59	2,955.45	2,955.45	3,970,860,000
6/15/2020	2,993.76	3,079.76	2,965.66	3,066.59	3,066.59	5,757,480,000
6/16/2020	3,131.00	3,153.45	3,076.06	3,124.74	3,124.74	5,845,810,000
6/17/2020	3,136.13	3,141.16	3,108.03	3,113.49	3,113.49	4,560,450,000
6/18/2020	3,101.64	3,120.00	3,093.51	3,115.34	3,115.34	4,448,690,000
6/19/2020	3,140.29	3,155.53	3,083.11	3,097.74	3,097.74	8,346,650,000
6/22/2020	3,094.42	3,120.92	3,079.39	3,117.86	3,117.86	4,686,690,000
6/29/2020	3,018.59	3,053.89	2,999.74	3,053.24	3,053.24	4,473,970,000
6/30/2020	3,050.20	3,111.51	3,047.83	3,100.29	3,100.29	4,705,850,000

Date	Link	speech
2/11/2020	https://www.federalreserve.gov/newsevents/testimony/powell20200211a.htm	Chair Powell submitted identical remarks to the Committee on Banking, Housing, and Urban Affairs, U.S. Senate, on February 12, 2020. Chairwoman Waters, Ranking Member McHenry, and other members of the Committee, I am pleased to present the Federal Reserve's semiannual Monetary Policy Report. My colleagues and I strongly support the goals of maximum employment and price stability that Congress has set for monetary policy. Congress has given us an important degree of independence to pursue these goals base Current Economic Situation The economic expansion is well into its 11th year, and it is the longest on record. Over the second half of last year, economic activity increased at a moderate pace and the labor market strengthened further, as the economy Job gains averaged 200,000 per month in the second half of last year, and an additional 225,000 jobs were added in January. The pace of job gains has remained above what is needed to provide jobs for new workers enteri Gross domestic product rose at a moderate rate over the second half of last year. Growth in consumer spending moderated toward the end of the year following earlier strong increases, but the fundamentals supporting hous Inflation ran below the FOMC's symmetric 2 percent objective throughout 2019. Over the 12 months through December, overall inflation based on the price index for personal consumption expenditures was 1.6 percent. Core The nation faces important longer-run challenges. Labor force participation by individuals in their prime working years is at its highest rate in more than a decade. However, it remains lower than in most other advanced econ Monetary Policy I will now turn to monetary policy. Over the second half of 2019, the FOMC shifted to a more accommodative stance of monetary policy to cushion the economy from weaker global growth and trade developments and to pr Taking a longer view, there has been a decline over the past quarter-century in the level of interest rates consistent with stable prices and the economy operating at its full potential. This low interest rate environment may limit The current low interest rate environment also means that it would be important for fiscal policy to help support the economy if it weakens. Putting the federal budget on a sustainable path when the economy is strong would Finally, I will briefly review our planned technical operations to implement monetary policy. The February Monetary Policy Report provides details of our operations to date. Last October, the FOMC announced a plan to purch Thank you. I am happy to take your questions.

The algorithms/models employed to address our business question include vader sentiment analysis, sumy, and various nltk packages. After creating our dataframes, we preprocessed the data, and then calculated elements of our data (such as the percent change of

[1] NASDAQ and S&P500 data were retrieved from Yahoo Finance using their "Historical Data" section
[2] Speeches collected from https://www.federalreserve.gov

Open and Close for each speech date). We then used Vader Sentiment Analysis to generate sentiment scores for the speeches. We then found the most negative and positive sentences in each speech by splitting the speeches into individual sentences, and running each through vader sentiment analysis. Next, we used Sumy to generate a summary of each speech. We then ran the summary through vader sentiment analysis to generate a sentiment score for the summary. After all of these steps, we combined all of the data together into a single dataframe. This allowed us to generate visualizations using matplotlib and better understand our data.

Snippet of Code

```
# function to get sentiment of speech
def get_sentiment(speech):
    scores = analyzer.polarity_scores(speech)
    score = scores['compound']
    if score >= 0.5:
        return 'Positive'
    elif score < 0.5 and score >= -0.5:
        return 'Neutral'
    else:
        return 'Negative'

speeches['Speech_Sentiment'] = speeches['speech'].apply(get_sentiment)
```

```
# define function to find the most positive and negative sentences in a speech
def find_most_positive_and_negative_sentences(speech):
    # split the speech into sentences
    sentences = nltk.sent_tokenize(speech)

    # initialize variables to store the most positive and negative sentences
    most_positive = {'sentence': '', 'score': -1}
    most_negative = {'sentence': '', 'score': 1}

    # iterate over each sentence and compute the sentiment score
    for sentence in sentences:
        # compute the sentiment score using VADER
        score = analyzer.polarity_scores(sentence)['compound']

        # check if the current sentence has the most positive or negative score
        if score > most_positive['score']:
            most_positive['score'] = score
            most_positive['sentence'] = sentence
        elif score < most_negative['score']:
            most_negative['score'] = score
            most_negative['sentence'] = sentence

    # return the most positive and negative sentences
    return most_positive['sentence'], most_negative['sentence']

speeches[['Most positive sentence', 'Most negative sentence']] = speeches['speech'].apply(
                        find_most_positive_and_negative_sentences).apply(pd.Series)
```

```
# using sumy to summarize the speech to understand it better
#!pip install sumy
from sumy.parsers.plaintext import PlaintextParser
from sumy.nlp.tokenizers import Tokenizer
from sumy.summarizers.text_rank import TextRankSummarizer
```

```
def summarize_speech(text):
    # Set the number of sentences in the summary
    num_sentences = 1

    # Initialize the TextRank summarizer
    summarizer = TextRankSummarizer()

    # Parse the text and tokenize it
    parser = PlaintextParser.from_string(text, Tokenizer("english"))

    # Generate the summary and join the sentences
    summary = " ".join([str(sentence) for sentence in summarizer(parser.document, num_sentences)])

    # return the summary
    return summary
```

(3) Analysis and Results

After finishing the programming portion of the project, we were able to begin our analysis. Out of the 14 speeches, only 3 (or 29%) of the sentiment scores accurately matched the negative or positive movement of the NASDAQ's % change on the day of the speech. Similarly, 6/14 or 43% of the sentiment scores accurately matched the movement of the S&P's percent change on the day of the speech.

Final Dataframe Created

	Date	Link	speech	Speech_Scores	Speech_Sentiment	Most positive sentence	Most negative sentence	Summary	Summary_Sentiment	nasdaq_Pct_Change	sp500_Pct_Change
0	2020-02-11	https://www.federalreserve.gov/newsevents/test...	Chair Powell submitted identical remarks to th...	{'neg': 0.051, 'neu': 0.796, 'pos': 0.153, 'co...	Positive	This low interest rate environment may limit t...	Residential investment turned up in the second...	Over the second half of 2019, the FOMC shifted...	Neutral	-0.433328	-0.241245
1	2020-04-09	https://www.federalreserve.gov/newsevents/test...	Good morning. The challenge we face today is d...	{'neg': 0.08, 'neu': 0.747, 'pos': 0.172, 'com...	Positive	Even more importantly, we have acted to safegu...	The coronavirus has spread quickly around the ...	Many of the programs we are undertaking to sup...	Positive	-0.188881	0.462011
2	2020-05-12	https://www.federalreserve.gov/newsevents/test...	Chairman Crapo, Ranking Member Brown, members ...	{'neg': 0.106, 'neu': 0.745, 'pos': 0.149, 'co...	Positive	Over the past two months, the Federal Reserve ...	More than a decade ago, U.S. banking organizat...	We advised institutions that working construct...	Positive	-2.412975	-2.360265
3	2020-05-13	https://www.federalreserve.gov/newsevents/spee...	The coronavirus has left a devastating human a...	{'neg': 0.126, 'neu': 0.754, 'pos': 0.12, 'com...	Negative	And fourth, temporary regulatory adjustments t...	The scope and speed of this downturn are witho...	Long stretches of unemployment can damage or e...	Negative	-1.586488	-1.600218
4	2020-05-19	https://www.federalreserve.gov/newsevents/test...	Chairman Crapo, Ranking Member Brown, and othe...	{'neg': 0.078, 'neu': 0.756, 'pos': 0.166, 'co...	Positive	In addition to monetary policy, we took forcef...	By these measures and many others, the scope a...	In addition to monetary policy, we took forcef...	Positive	-0.459065	-0.869907
5	2020-05-21	https://www.federalreserve.gov/newsevents/spee...	Good afternoon. I just want to say a few words...	{'neg': 0.034, 'neu': 0.844, 'pos': 0.122, 'co...	Positive	But all of us have our own decisions to make a...	And while the burden is widespread, it is not ...	From an economic perspective, we hope to learn...	Neutral	-0.963293	-0.721898
6	2020-06-16	https://www.federalreserve.gov/newsevents/test...	Chair Powell submitted identical remarks to th...	{'neg': 0.064, 'neu': 0.782, 'pos': 0.154, 'co...	Positive	To support the small business sector, we estab...	I want to end by acknowledging the tragic even...	To provide stability to the financial system a...	Positive	-0.541823	-0.199936
7	2020-06-19	https://www.federalreserve.gov/newsevents/spee...	Thank you, President Mester and Treye Johnson,...	{'neg': 0.109, 'neu': 0.759, 'pos': 0.132, 'co...	Positive	But given the opportunity, I'll always bet on ...	A particular cruelty of the pandemic has been ...	And employers' input has influenced work acros...	Neutral	-0.956070	-1.354970
8	2020-06-30	https://www.federalreserve.gov/newsevents/test...	Chairwoman Waters, Ranking Member McHenry, and...	{'neg': 0.059, 'neu': 0.765, 'pos': 0.176, 'co...	Positive	In addition to these steps, we took forceful m...	In contrast to the 2008 crisis when banks pull...	In addition to these steps, we took forceful m...	Positive	1.857966	1.642187
9	2020-08-27	https://www.federalreserve.gov/newsevents/spee...	Thank you, Esther, for that introduction, and ...	{'neg': 0.059, 'neu': 0.826, 'pos': 0.115, 'co...	Positive	This emphasis on transparency reflected what w...	Having declined significantly in the five year...	For the past year and a half, my colleagues an...	Neutral	-0.537727	-0.016929
10	2020-09-22	https://www.federalreserve.gov/newsevents/test...	Chair Powell submitted identical remarks to th...	{'neg': 0.029, 'neu': 0.825, 'pos': 0.145, 'co...	Positive	With funding from the CARES Act (Coronavirus A...	Main Street loans have a five-year maturity, n...	The Main Street Lending Program The Federal Re...	Positive	0.830841	0.601381
11	2020-10-06	https://www.federalreserve.gov/newsevents/spee...	Good morning. It has been just eight months si...	{'neg': 0.072, 'neu': 0.817, 'pos': 0.111, 'co...	Positive	On a more positive note, we have seen that the...	By the end of the month, many important market...	We identified three ways that our tools could ...	Neutral	-1.413498	-1.401398
12	2020-11-10	https://www.federalreserve.gov/newsevents/test...	Vice Chair for Supervision Quarles submitted i...	{'neg': 0.068, 'neu': 0.788, 'pos': 0.144, 'co...	Positive	Since then, working with our colleagues in oth...	But in light of continuing uncertainty, we too...	We published a set of key principles to guide ...	Positive	-0.590066	0.064065
13	2020-12-01	https://www.federalreserve.gov/newsevents/test...	Chair Powell submitted identical remarks to th...	{'neg': 0.027, 'neu': 0.845, 'pos': 0.128, 'co...	Positive	With funding from the Coronavirus Aid, Relief,...	In contrast, spending on services remains low ...	The Primary Market Corporate Credit Facility T...	Positive	0.339063	0.454761

A possible reason for the higher correlation between the movement of the S&P 500 and the sentiment scores may be a consequence of the differing characteristics and compositions of the S&P 500 and the NASDAQ. The NASDAQ mainly consists of tech companies, while the S&P 500 is composed of the 500 largest publicly traded companies in the US, representing an array of sectors.

Due to this, the NASDAQ may be more reactive to changes in sentiment specifically related to the technology sector, whilst the S&P 500 may be governed by a larger distribution of economic factors and sentiment related to the market as a whole. Hence, the sentiment expressed

in Jerome Powell's speeches may have had a stronger impact on the more diversified and paradigmatic index, the S&P 500, than on the NASDAQ.

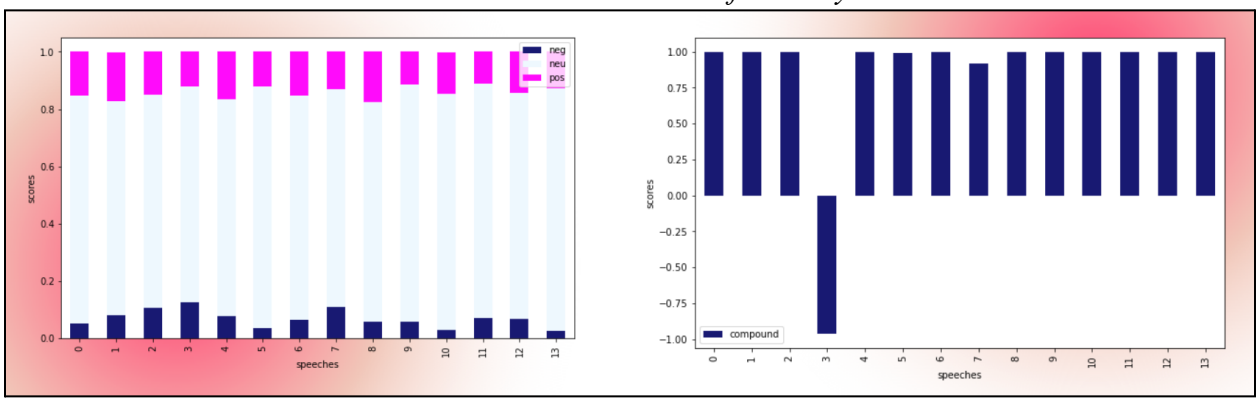
To understand why we may have gotten these results, we were cognizant of these four factors:

1. **Market inefficiencies:** Due to the stock market being a complex and dynamic system influenced by economic indicators, corporate earnings, and political events, the market may not respond rationally or predictably to news or events.
2. **Timing:** As a consequence of the volatility of the stock market, quick reactions to news or events are expected.
3. **Limited scope:** Analyzing the sentiment of a single speaker's speeches may not provide a complete picture of the factors influencing the stock market. Other factors, such as macroeconomic indicators, corporate earnings reports, and geopolitical events, can also play a significant role in stock market movements.
4. **Model limitations:** Sentiment analysis models are not perfect and can have limitations. Some models may not capture the complexity of language, the nuances of sentiment, or the context of the speech. It's essential to evaluate the model's accuracy and limitations before relying on its results.

To summarize, analyzing the sentiment of speeches by the Chair of the Federal Reserve may provide insights into market reactions, but it may not be the only factor driving stock market movements. It's important to consider other factors that may be influencing the market and evaluate the accuracy and limitations of the sentiment analysis model. One significant factor worth noting is that out of the 10,800,000,000 trading volume present in 2020, 19.5% were retail investors. Retail investors can make the stock market, including the S&P 500 and NASDAQ,

more unpredictable due to their trading behavior. This is due to the fact that retail investors tend to have less access to information and resources than institutional investors, so they may make decisions based on emotion or incomplete information. However, this information must be taken with a grain of salt; the impact of retail investors on the overall market can vary depending on the size and liquidity of the stocks they trade. If the retail investors were a factor in the unpredictability of the market, it would have been a very limited role.

Data Visualizations Created for Analysis



(4) Implications

The sentiment analysis we did on the Federal Reserve's impact on stock prices from S&P500 and NASDAQ in 2020 has the potential to generate value and provide insights for solving existing and future business problems and improving decision-making in several ways:

1. **Understanding the Impact of Federal Reserve Communications:** By analyzing the sentiment of speeches delivered by Jerome Powell and correlating them with stock market performance, businesses and investors can gain a thorough understanding of how the communications of the Federal Reserve impact financial markets. This knowledge can help them make more informed and refined decisions regarding investments, portfolio management, and risk assessment.
2. **Assessing Market Sentiment during Significant Events:** Analyzing sentiment during major events such as the COVID-19 pandemic provides valuable insights into how market sentiment evolved and responded to those events. Businesses can use this information to assess the impact of such events on their industry, identify potential risks, opportunities, and adjust their strategies accordingly.
3. **Identifying Market Volatility and Trends:** The analysis of stock prices from S&P500 and NASDAQ can help businesses and investors identify market volatility, recognize trends, and uncover patterns. This information can be leveraged to make informed decisions related to timing investments, adjusting portfolios, and managing risk.
4. **Enhancing Financial Forecasting and Risk Management:** The correlation between Federal Reserve communications, sentiment analysis, and stock market

performance can contribute to improving financial forecasting and risk management models. Businesses can incorporate sentiment analysis as an additional factor in their models to better predict market movements and assess potential risks.

5. **Supporting Policy and Decision-Making:** The insights generated from this analysis can also be valuable for policymakers and government agencies. Understanding the relationship between Federal Reserve communications and stock market performance can assist in formulating effective monetary policies, promoting financial stability, and making informed decisions that impact the overall economy. As well as putting more effort into formulating and rehearsing their communications and speeches, to not cause negative or unintentional reactions.

In conclusion, the analysis of sentiment within speeches by the Federal Reserve Chair might illuminate potential market reactions. However, it is crucial to acknowledge that this is not the lone driver of stock market dynamics. A comprehensive understanding of the market demands consideration of a diverse array of influencing factors and necessitates an evaluation of the precision and inherent limitations of the sentiment analysis model. Overall, the results of this project can provide valuable information and insights that can be utilized by businesses, investors, policymakers, and decision-makers to improve their understanding of the financial markets, improve decision-making processes, and mitigate risks.