DS-GA 1015 - Text as Data - Course Project
Sentiment Analysis of FOMC documents during the 2008-2009 Financial Crisis and post-2022 Russia-Ukraine War and its correlation with economic metrics

Rodrigo Kreis de Paula
rk4197@nyu.edu

Sarvesh Patki
ssp6603@nyu.edu

Word count: XXX

1 Introduction

We wish to use this project as an opportunity to apply text analysis techniques to real-world problems. We have conducted sentiment analysis on Monetary Policy statements issued by the Federal Reserve(Fed) during two significant historical events - the 2008-2009 Global Financial Crisis(GFC) and the Russia-Ukraine War(post-2022). The goals are to study the correlation between the sentiment of the FOMC statements and meeting notes with economic metrics such as GDP, inflation rates, and unemployment rates during corresponding periods. Since a recession is predicted to be right around the corner, we also attempt to assess the main differences and similarities between these two remarkable periods so that preparations can be made accordingly.

The 2008-2009 Global Financial Crisis was a major economic event that had a significant impact on global financial markets. The Fed played a critical role in addressing the crisis through its monetary policy decisions, including several rounds of quantitative easing and interest rate cuts. By analyzing the sentiment of the FOMC statements during this period, we can gain insights into the Fed's assessment of the crisis and its potential impact on the economy and the markets.

We also study the correlation between sentiment and economic metrics during the post-2022 Russia-Ukraine War. Inflation has been on the rise due to the quantitative easing policy of the Fed during the global Covid-19 pandemic. Soon after that the Russia-Ukraine war in 2022 disrupted supply chains and energy sources. By analyzing the sentiment of the FOMC statements in this context and studying how markets have been reacting to it, we can gain a better understanding of the issue and attempt to predict the upcoming recession period.

Overall, the motivation for this project is driven by both academic and personal interests, and we hope that the findings of this study can contribute to a better understanding how the abundant textual data is perceived by the markets and how it influences them.

2 Literature

The FOMC statement is an important source of information for financial analysts and investors as it provides insights into the Federal Reserve's outlook on the economy and monetary policy. The FOMC statement is released after each meeting of the Federal Open Market Committee (FOMC) and provides an assessment of the current economic situation and the Fed's policy stance. The

FOMC meeting minutes provide more detailed information on the discussions and deliberations that took place during the meeting.

Several studies have used sentiment analysis to analyze the sentiment of the FOMC statement and its impact on financial markets. For example, a study by Tetlock et al.(2008) [1] analyzed the sentiment of the FOMC statements from 1994 to 2004 and found that more positive statements were associated with higher stock returns. Another study by Luo and Wu(2019) [2] used sentiment analysis to analyze the impact of the FOMC statement on Treasury yields and found that positive sentiment in the FOMC statement was associated with lower yields.

In addition, there is a growing body of literature on the use of sentiment analysis to analyze the impact of geopolitical events on financial markets. For example, a study by Baker and Wurgler(2006) [3] analyzed the sentiment of newspaper articles on World War II and found that sentiment was a significant predictor of stock returns during the war. Another study by Brogaard et al.(2014) [4] used sentiment analysis to analyze the impact of the 2011 Japanese earthquake on financial markets and found that negative sentiment was associated with lower stock prices.

In this project, we will use sentiment analysis to analyze the FOMC statements and meeting minutes during the 2008-2009 Financial Crisis and the post-2022 Russia-Ukraine War period and study the correlation with economic metrics. Our study contributes to the literature by analyzing the sentiment of the FOMC statements during these significant historical events, how they share similarities or exhibit differences, and understanding its impact on the broader economic landscape.

3 Data

The data used for this project has been gathered by us and was not a ready-to-use dataset available online. Hence, pre-processing and extraction were critical steps involved in this project.

The data used for this process are the statements and meeting minutes released by the Federal Open Market Committee (FOMC)through their official website. As per the Fed, the economic downturn lasting from December 2007 to June 2009 was the longest since World War II. Thus we have used FOMC statements and meeting minutes for the period from October 2007 to December 2009.

Russia invaded Ukraine on February 24, 2022 which had sudden geopolitical and economic implications. Thus, we have used FOMC statements and meeting notes immeediately following the war i.e. from March 2022 and we conclude the study at March 2023 because data for some econometric parameters is not available for April 2023, as of the day when this project was being undertaken.

Our 'data' csv file has 27 rows and 9 columns which have been described in the table below. Each row corresponds to the official FOMC statement released by the Fed.

Column Name	Description
date	Date on which FOMC statement was released
text	FOMC Statement
CPI	Consumer Price Index data based upon a 1982 base of 100 LINK
SP500	% change in S&P500 index next day after FOMC statement was released
decision	Increase or Decrease in interest rate by Fed in bps
sentiment_afinn	Sentiment score using AFINN technique
sentiment_bing	Sentiment score using bing technique
sentiment_nrc	Sentiment score using nrc technique
sentiment_syuzhet	Sentiment score using syuzhet technique

4 Methodology

Sentiment analysis is a widely used technique in natural language processing (NLP) and has gained significant attention in financial analysis. It involves the use of computational methods to extract and analyze the sentiment, emotion, or opinion expressed in a piece of text. Sentiment analysis has been applied to various fields, including social media analysis, customer feedback analysis, and financial analysis.

In financial analysis, sentiment analysis can be used to analyze the impact of news and events on the financial markets. Researchers have used sentiment analysis to predict stock prices, forecast economic indicators, and understand the relationship between sentiment and market volatility. In particular, sentiment analysis has been used to analyze the sentiment of financial news and corporate announcements.

To perform sentiment analysis on the FOMC documents, we used the techniques we have learned over this semester to extract and quantify the sentiment of each statement and meeting minutes. We employed the R programming language and its related packages, such as tidytext, dplyr, and tidyr, for this purpose. These packages allow us to manipulate and analyze the text data efficiently, making it easier to conduct sentiment analysis on large volumes of text. In the following subsections, we'll describe all the steps and present their respective results.

We have tried four different methods of sentiment analysis, namely, AFINN, Bing, NRC, and Syuzhet. The monetary policy decisions by Fed have a correlation with the markets and the economy. An increase in interest rates slows down the economy and also the financial markets. Naturally, a negative sentiment is associated when the Fed hikes the interest rates. We have used the CPI and S&P 500 as metrics that demonstrate the effect of monetary policy decisions on the economy and the financial markets respectively. We have computed correlations of the sentiments by different techniques with the above mentioned econometric parameters as shown in Figure 4 below. The FED Decision row captures the increase/decrease in the interest rates by the Fed for each corresponding statement.

There are several approaches to perform sentiment analysis, and among them, four models stand out: Afinn, Bing, NRC, and Syuzhet. The Afinn model assigns a sentiment score to words based on a pre-built dictionary, while the Bing model uses a predefined list of positive and negative words to classify the sentiment of a text. The NRC model identifies the emotional tone of a text by using a combination of lexicons and machine learning algorithms, and the Syuzhet model is based on the idea that sentiment is represented by patterns of change in the emotional valence of the text over time. Each model has its strengths and weaknesses, and choosing the best model for a particular task depends on the specific context and requirements of the analysis. Each method has been summarized below. This project was a great opportunity to go beyond the syllabus of the course and learn about various techniques for sentiment analysis.

The AFINN (Affective Norms for English Words) is a simple and effective lexicon-based method for sentiment analysis. It assigns a score to each word based on its polarity, ranging from -5 to +5, with negative scores indicating negative sentiment and positive scores indicating positive sentiment. AFINN has been widely used in financial document analysis, including for predicting stock market returns, assessing market sentiment, and detecting fraudulent financial reporting.

The Bing lexicon is another popular sentiment analysis technique that was developed by Microsoft. It categorizes words as either positive or negative, with each word assigned a binary score of 1 or -1, respectively. Bing has been applied to financial document analysis for tasks such as identifying the sentiment of news articles, analyzing tweets about stocks, and predicting corporate earnings surprises.

The NRC (National Research Council) lexicon is a comprehensive sentiment analysis resource that includes words with both positive and negative sentiment as well as a range of emotions such as anger, fear, joy, and sadness. It is commonly used in financial document analysis for tasks such as detecting financial events, predicting stock price movements, and analyzing earnings call transcripts.

Syuzhet is a novel sentiment analysis technique that uses narrative arcs to identify the emotional trajectory of a text. It analyzes the sentiment of a text by identifying patterns in the sequence of positive and negative emotions over time. Syuzhet has been applied to financial document analysis for tasks such as identifying changes in market sentiment, predicting stock price movements, and analyzing company announcements.

Scatter plots for the sentiment analysis score by each of the above methods with respect to the fed's rate changes have been shown in Figures 6, 7, 8 and 9. Figure 10 shows them in a single plot for a comparative analysis. Scatter plots show that AFINN is the best indicator because of a clear inverse correlation with Fed's policies - rate hikes are associated with neg

5 Results

As interest rates increase we expect the economy to slow down and thus the financial markets acting bearish. Thus, a statement mentioning a rate hike is associate with a negative sentiment and we expect inverse correlation with econometric factors. This relationship can be seen by comparing Figures 2 and 3 with corresponding time periods in Figure 5. The AFINN method best captures the reaction of the markets and economy to the Fed's statements as is evident from the graphs and the correlation computed in Figure 4.

6 Conclusion

The AFINN method is particularly useful to conduct sentiment analysis on financial texts especially in the context of the economy and the financial markets' reaction to the Fed's interest rate policies. Through this study we also wanted to attempt to predict the period for the upcoming recession. During the GFC, interest rates peaked at 5.25% after which the fed lowered them. We have reached a similar stage and it can be seen that inflation has been under control for the past month and economy has slowed down along with a spike in the unemployment numbers. This indicates the beginning of a recession. Historically speaking recessions have lasted 12-18 months and the same is expected to occur this time as well. Drawing conclusions from the data gathered for this study, it seems that interest rates will go back to 0-0.25 % in around 18 months and inflation rate would be maintained around the target of 2%.

7 Appendix

7.1 Figures

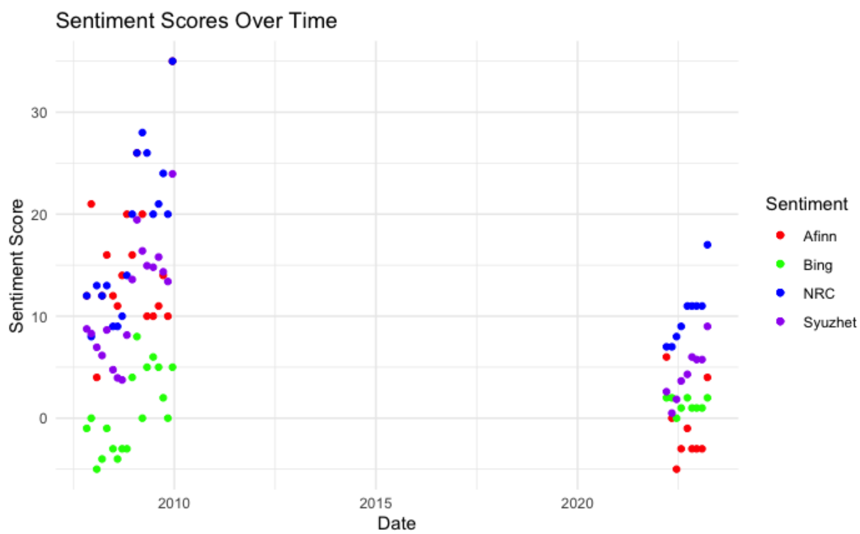


Figure 1: Sentiment Scores per Method for the two periods

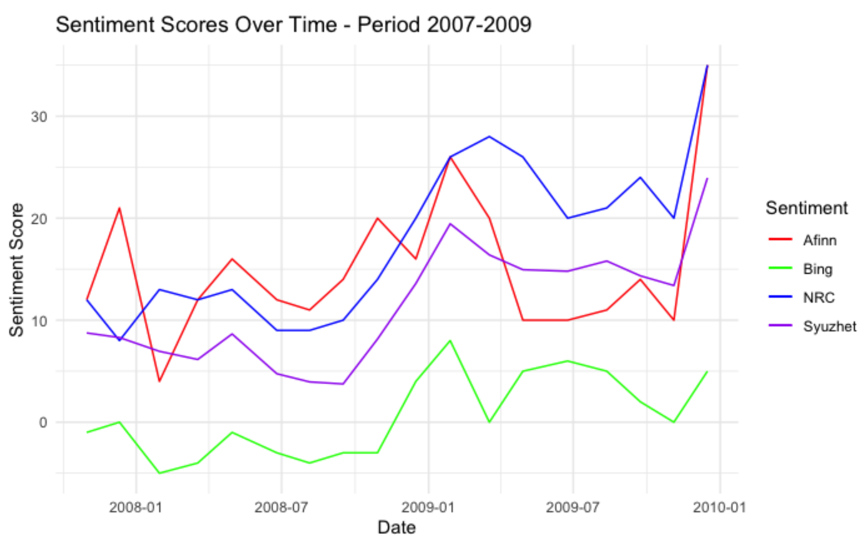


Figure 2: Sentiment Scores per Method for the period 2007-2009

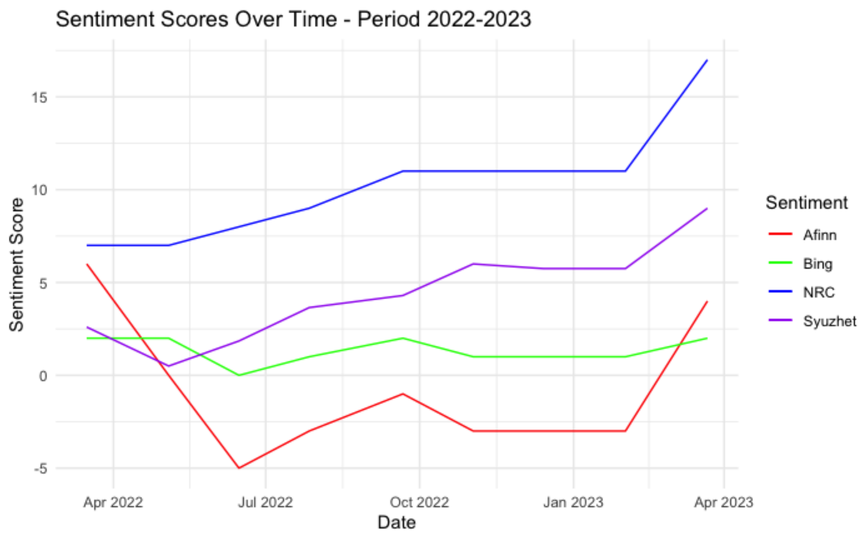


Figure 3: Sentiment Scores per Method for the period 2022-2023

Variable	sentiment_afinn	sentiment_bing	sentiment_nrc	sentiment_syuzhet
SP500	0.242426	0.06983754	0.176645	0.1354841
FED Decision	−0.6330461	0.2475876	−0.2625865	−0.3563229
CPI	−0.7643138	0.08489507	−0.4346044	−0.5286485

Figure 4: Correlations between sentiment variables and economic metrics

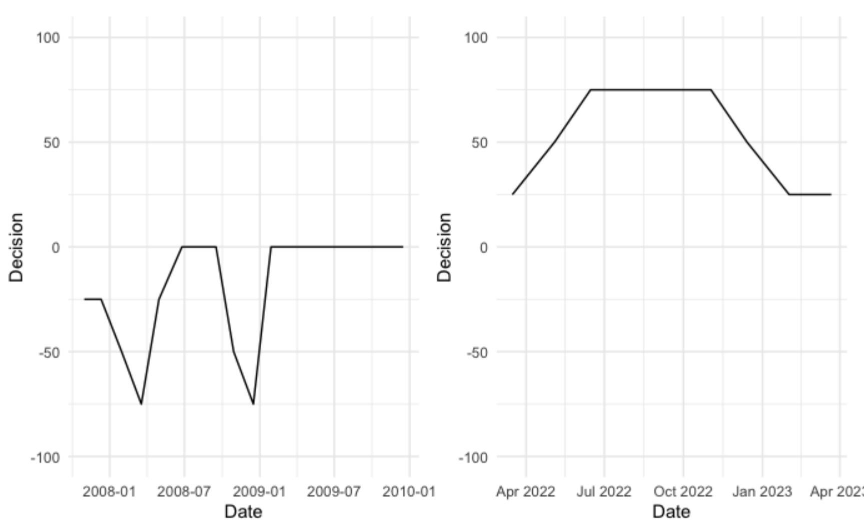


Figure 5: FED Decisions over time for both periods

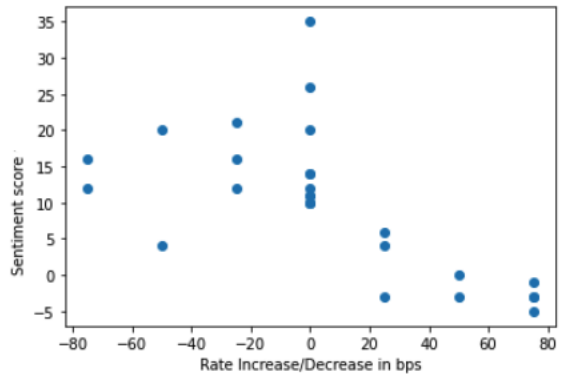


Figure 6: Sentiment Score with Afinn vs. FED Decision

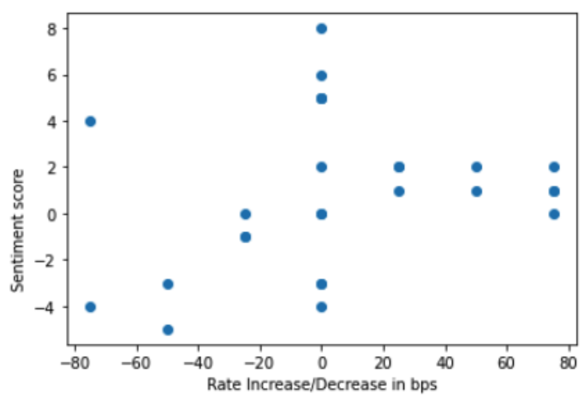


Figure 7: Sentiment Score with Bing vs. FED Decision

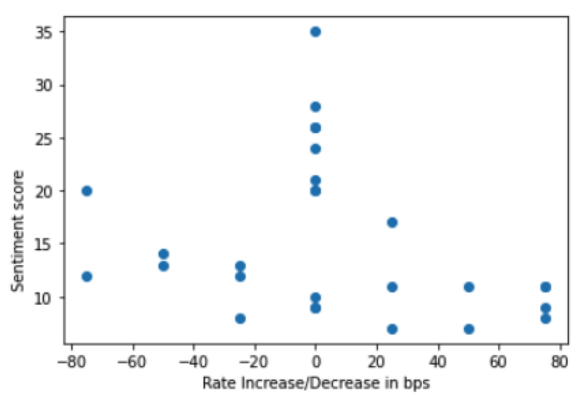


Figure 8: Sentiment Score with nrc vs. FED Decision

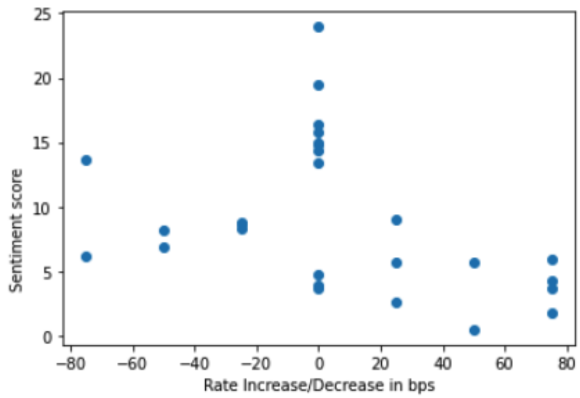


Figure 9: Sentiment Score with syuzhet vs. FED Decision

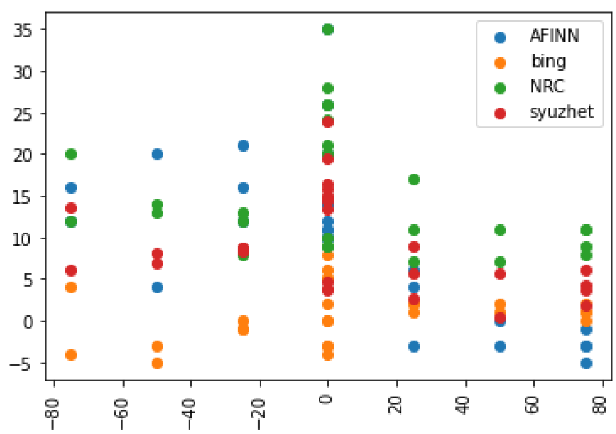


Figure 10: Sentiment Score with all methods vs. FED Decision

7.2 Code

Please find the PDF file from R Studio containing the full code at the end of this Final Report.

References

[1] Tetlock, P. C., Saar-Tsechansky, M., and Macskassy, S. (2008). More Than Words: Quantifying Language to Measure Firms' Fundamentals. Journal of Finance, 63(3), 1437-1467.

[2] Luo, Y., and Wu, H. (2019). The Effects of FOMC Communication on Treasury Yields: Evidence from a High-Dimensional VAR. Journal of Banking and Finance, 101, 98-112.

[3] Baker, M., and Wurgler, J. (2006). Investor Sentiment and the Cross-Section of Stock Returns. Journal of Finance, 61(4), 1645-1680.

[4] Brogaard, J., Detzel, A., and Sreejith, A. (2014). Geolocation and Sentiment in Foreign News and the Stock Market. Journal of Financial and Quantitative Analysis, 49(4), 971-1002.

Course Project - Text as Data

May 08 2023

Importing and Reading Data

```
# Cleaning the enviroment
rm(list = ls())

# Loading packages
pacman::p_load(ldatuning,
               topicmodels,
               ggplot2,
               dplyr,
               rjson,
               quanteda,
               tidytext,
               stringi,
               tidyr,
               lubridate,
               parallel,
               doParallel,
               tm,
               textmineR,
               lsa,
               lda,
               proxy,
               text2vec,
               LDAvis,
               stm,
               geometry,
               Rtsne,
               rsvg,
               readtext,
               quanteda.textmodels,
               quanteda.textplots,
               stringr,
               bursts,
               syuzhet,
               SentimentAnalysis,
               textdata,
               tm,
               scales,
               patchwork,
               hrbrthemes)
```

```r
# Visualizing the files
files <- list.files("./data/FOMC/", full.names=TRUE)

# Reading in the files
fomc <- lapply(files, readLines)
```

```
## Warning in FUN(X[[i]], ...): incomplete final line found on
## './data/FOMC//01_28_2009.txt'

## Warning in FUN(X[[i]], ...): incomplete final line found on
## './data/FOMC//01_30_2008.txt'

## Warning in FUN(X[[i]], ...): incomplete final line found on
## './data/FOMC//02_01_2023.txt'

## Warning in FUN(X[[i]], ...): incomplete final line found on
## './data/FOMC//03_16_2022.txt'

## Warning in FUN(X[[i]], ...): incomplete final line found on
## './data/FOMC//03_18_2008.txt'

## Warning in FUN(X[[i]], ...): incomplete final line found on
## './data/FOMC//03_18_2009.txt'

## Warning in FUN(X[[i]], ...): incomplete final line found on
## './data/FOMC//03_22_2023.txt'

## Warning in FUN(X[[i]], ...): incomplete final line found on
## './data/FOMC//04_29_2009.txt'

## Warning in FUN(X[[i]], ...): incomplete final line found on
## './data/FOMC//04_30_2008.txt'

## Warning in FUN(X[[i]], ...): incomplete final line found on
## './data/FOMC//05_04_2022.txt'

## Warning in FUN(X[[i]], ...): incomplete final line found on
## './data/FOMC//06_15_2022.txt'

## Warning in FUN(X[[i]], ...): incomplete final line found on
## './data/FOMC//06_24_2009.txt'

## Warning in FUN(X[[i]], ...): incomplete final line found on
## './data/FOMC//06_25_2008.txt'

## Warning in FUN(X[[i]], ...): incomplete final line found on
## './data/FOMC//07_27_2022.txt'

## Warning in FUN(X[[i]], ...): incomplete final line found on
## './data/FOMC//08_05_2008.txt'
```

```
## Warning in FUN(X[[i]], ...): incomplete final line found on
## './data/FOMC//08_12_2009.txt'

## Warning in FUN(X[[i]], ...): incomplete final line found on
## './data/FOMC//09_16_2008.txt'

## Warning in FUN(X[[i]], ...): incomplete final line found on
## './data/FOMC//09_21_2022.txt'

## Warning in FUN(X[[i]], ...): incomplete final line found on
## './data/FOMC//09_23_2009.txt'

## Warning in FUN(X[[i]], ...): incomplete final line found on
## './data/FOMC//10_29_2008.txt'

## Warning in FUN(X[[i]], ...): incomplete final line found on
## './data/FOMC//10_31_2007.txt'

## Warning in FUN(X[[i]], ...): incomplete final line found on
## './data/FOMC//11_02_2022.txt'

## Warning in FUN(X[[i]], ...): incomplete final line found on
## './data/FOMC//11_04_2009.txt'

## Warning in FUN(X[[i]], ...): incomplete final line found on
## './data/FOMC//12_11_2007.txt'

## Warning in FUN(X[[i]], ...): incomplete final line found on
## './data/FOMC//12_14_2022.txt'

## Warning in FUN(X[[i]], ...): incomplete final line found on
## './data/FOMC//12_16_2008.txt'

## Warning in FUN(X[[i]], ...): incomplete final line found on
## './data/FOMC//12_16_2009.txt'
```

```r
fomc <- sapply(fomc, function(x) paste(x, collapse = " "))

# Extracting the date from the filename and store it in a vector
dates <- lapply(files, function(file) {
  filename <- tools::file_path_sans_ext(basename(file))  # remove the extension from the filename
  substr(filename, nchar(filename) - 9, nchar(filename))  # extract the date part of the filename
}) %>% unlist()

# Converting the vector dates to the date format
dates <- as.Date(dates, format = "%m_%d_%Y")

# Creating the dataframe
fomc_df <- data.frame(date = dates, text = fomc, stringsAsFactors = FALSE)
```

Inserting the CPI values on the dataframe in the same order they appear

```r
# Creating a vector of CPI values that corresponds to each row in fomc_df
cpi_values <- c(211.143, 211.1, 300.84, 264.877, 213.5, 212.709, 301.836,
                213.24, 214.8, 269.195, 271.696, 215.693, 218.8, 273.003,
                219.086, 215.834, 218.783, 274.31, 215.969, 216.573, 208.9,
                277.948, 216.33, 210.0, 278.802, 210.228, 215.949)

# Adding the CPI column to fomc_df
fomc_df$CPI <- cpi_values
```

Inserting the SP500 daily variation on the dataframe in the same order they appear

```r
# Creating a vector of CPI values that corresponds to each row in fomc_df
sp500_values <- c(2.74, -0.75, 2.18, 1.33, 4.21, 2.82, -1.07, 2.3, -0.38, 2.13,
                  -0.95, 0.35, 0.13, 1.89, 2.32, 1.19, 1.85, -2.3, -0.94, -0.01,
                  0.89, -3.11, 0.02, -1.92, -1.42, 4.2, -0.2)

# Adding the SP500 column to fomc_df
fomc_df$SP500 <- sp500_values
```

Inserting the Fed Decision on the dataframe in the same order they appear

```r
# Creating a vector of CPI values that corresponds to each row in fomc_df
decision_values <- c(0, -50, 25, 25, -75, 0, 25, 0, -25, 50, 75, 0, 0, 75, 0, 0,
                     0, 75, 0, -50, -25, 75, 0, -25, 50, -75, 0)

# Adding the SP500 column to fomc_df
fomc_df$decision <- decision_values
```

Creating the Sentiment Analysis

```r
# Tokenizing the text into individual words
fomc_words <- fomc_df %>%
  unnest_tokens(word, text)
```

Calculating the Score. Method #1: Afinn

```r
# Calculating the sentiment score of each word
sentiment_scores_afinn <- get_sentiment(fomc_words$word, method = "afinn")

# Group the sentiment scores by the original text
sentiment_by_text_afinn <- fomc_words %>%
  mutate(sentiment_score = sentiment_scores_afinn) %>%
  group_by(date) %>%
  summarize(sentiment_afinn = sum(sentiment_score))

fomc_df_with_sentiment <- merge(fomc_df, sentiment_by_text_afinn, by = "date", all.x = TRUE)
```

Calculating the Score. Method #2: Bing

```r
# Calculating the sentiment score of each word
sentiment_scores_bing <- get_sentiment(fomc_words$word, method = "bing")

# Group the sentiment scores by the original text
sentiment_by_text_bing <- fomc_words %>%
  mutate(sentiment_score = sentiment_scores_bing) %>%
  group_by(date) %>%
  summarize(sentiment_bing = sum(sentiment_score))

fomc_df_with_sentiment <- merge(fomc_df_with_sentiment, sentiment_by_text_bing, by = "date", all.x = TRU
```

Calculating the Score. Method #3: nrc

```r
# Calculating the sentiment score of each word
sentiment_scores_nrc <- get_sentiment(fomc_words$word, method = "nrc")

# Group the sentiment scores by the original text
sentiment_by_text_nrc <- fomc_words %>%
  mutate(sentiment_score = sentiment_scores_nrc) %>%
  group_by(date) %>%
  summarize(sentiment_nrc = sum(sentiment_score))

fomc_df_with_sentiment <- merge(fomc_df_with_sentiment, sentiment_by_text_nrc, by = "date", all.x = TRUE
```

Calculating the Score. Method #4: syuzhet

```r
# Calculating the sentiment score of each word
sentiment_scores_syuzhet <- get_sentiment(fomc_words$word, method = "syuzhet")
```

```
# Group the sentiment scores by the original text
sentiment_by_text_syuzhet <- fomc_words %>%
  mutate(sentiment_score = sentiment_scores_syuzhet) %>%
  group_by(date) %>%
  summarize(sentiment_syuzhet = sum(sentiment_score))

fomc_df_with_sentiment <- merge(fomc_df_with_sentiment, sentiment_by_text_syuzhet, by = "date", all.x =
```

Calculating the correlations between each method and CPI, SP500 and Fed

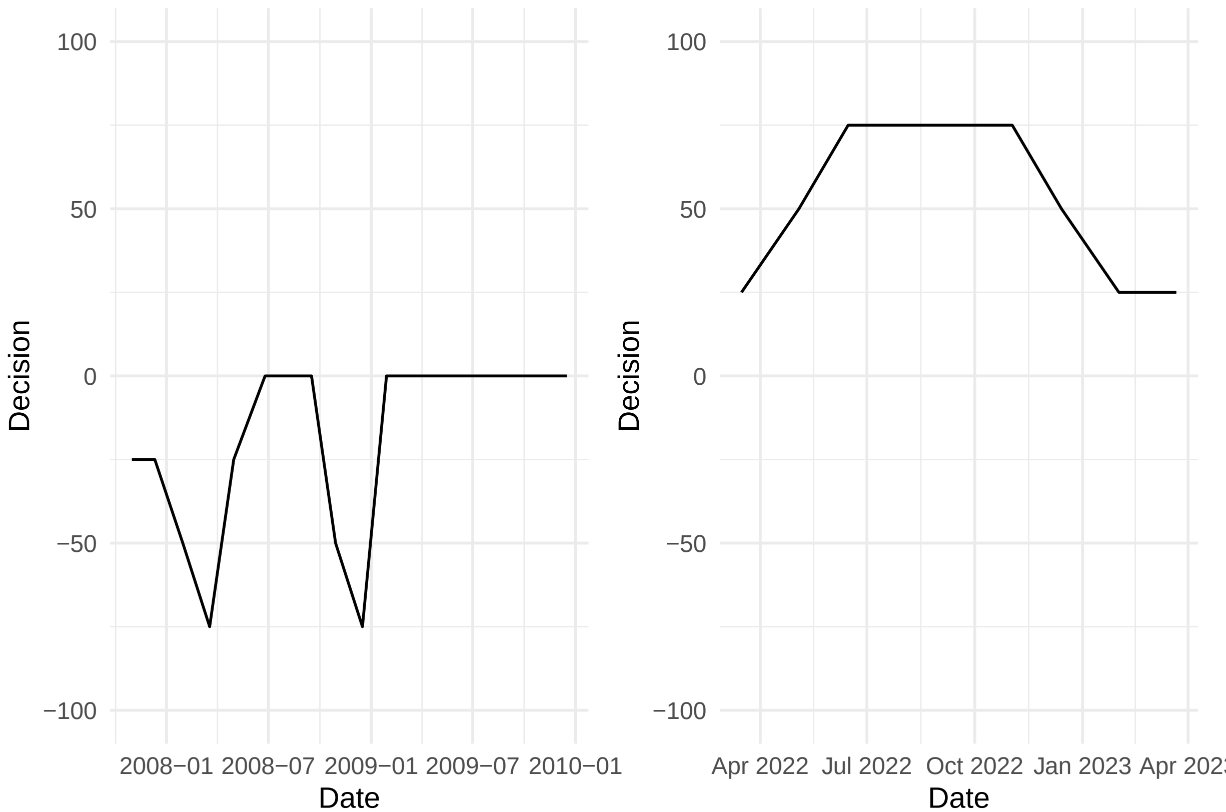
Decision

```
# calculate correlations between CPI and sentiment variables
correlations_cpi <- cor(fomc_df_with_sentiment[, c("CPI", "sentiment_afinn", "sentiment_bing",
                                   "sentiment_nrc", "sentiment_syuzhet")])

# calculate correlations between SP500 and sentiment variables
correlations_sp500 <- cor(fomc_df_with_sentiment[, c("SP500", "sentiment_afinn", "sentiment_bing",
                                     "sentiment_nrc", "sentiment_syuzhet")])

# calculate correlations between decision and sentiment variables
correlations_decision <- cor(fomc_df_with_sentiment[, c("decision", "sentiment_afinn", "sentiment_bing"
                                         "sentiment_nrc", "sentiment_syuzhet")])
```

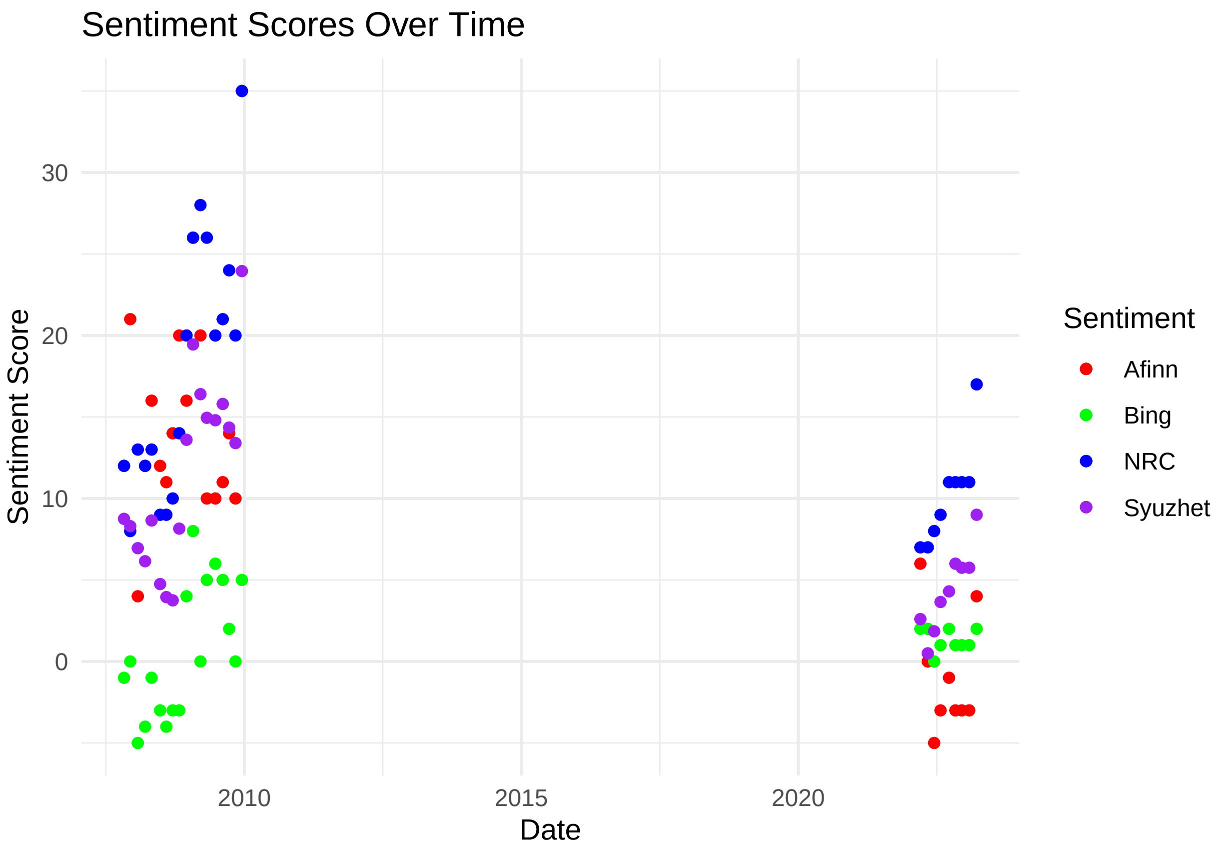
Plotting the graphs to help visualize

```
# Subsetting the periods
subsetted_data1 <- subset(fomc_df_with_sentiment, date >= as.Date("2007-01-01") & date <= as.Date("2009
subsetted_data2 <- subset(fomc_df_with_sentiment, date >= as.Date("2022-01-01") & date <= as.Date("2023

# Plot 1: FED Decision over time
# Period 1
p1 <- ggplot(subsetted_data1, aes(x = date, y = decision)) +
  geom_line() +
  labs(x = "Date", y = "Decision") +
  theme_minimal() +
  scale_y_continuous(limits = c(-100, 100))

# Period 2
p2 <- ggplot(subsetted_data2, aes(x = date, y = decision)) +
  geom_line() +
  labs(x = "Date", y = "Decision") +
  theme_minimal() +
  scale_y_continuous(limits = c(-100, 100))
```

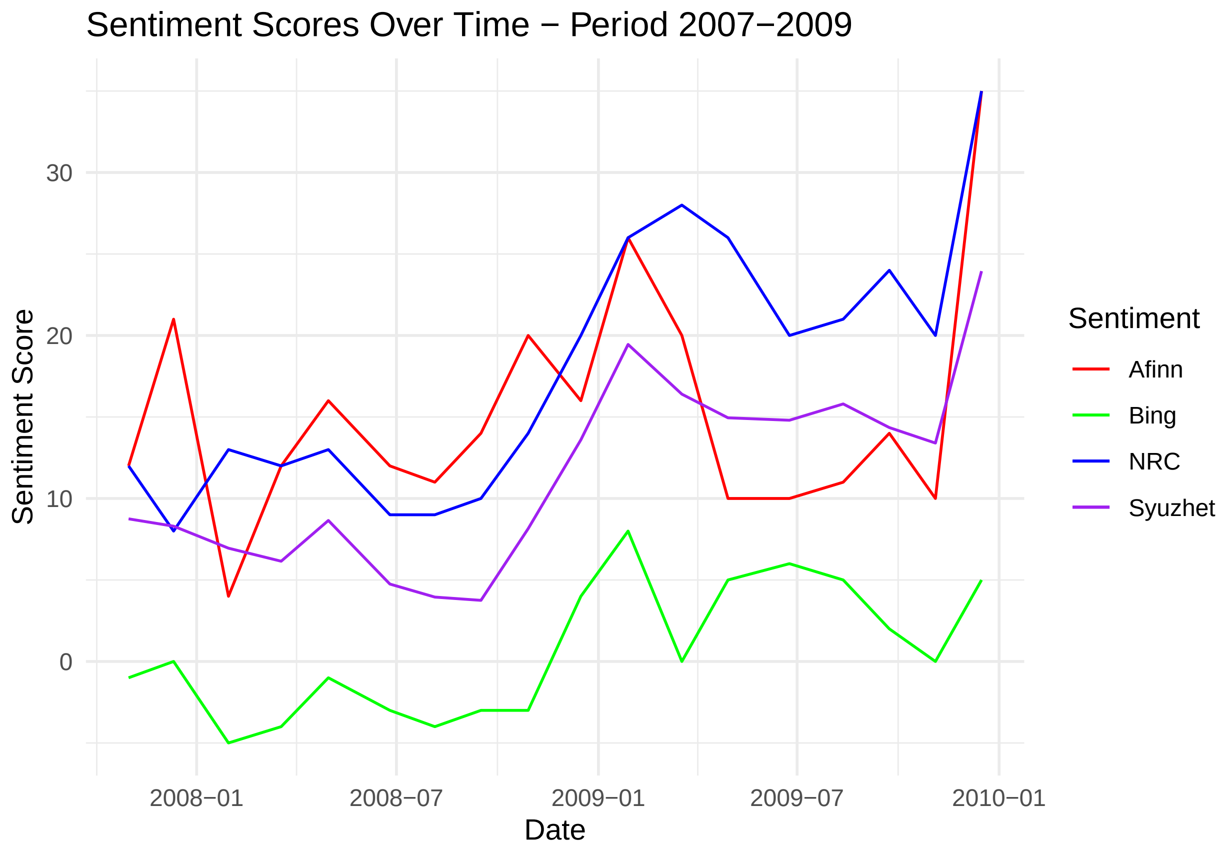
```r
# Priting Plot 1
p1 + p2
```



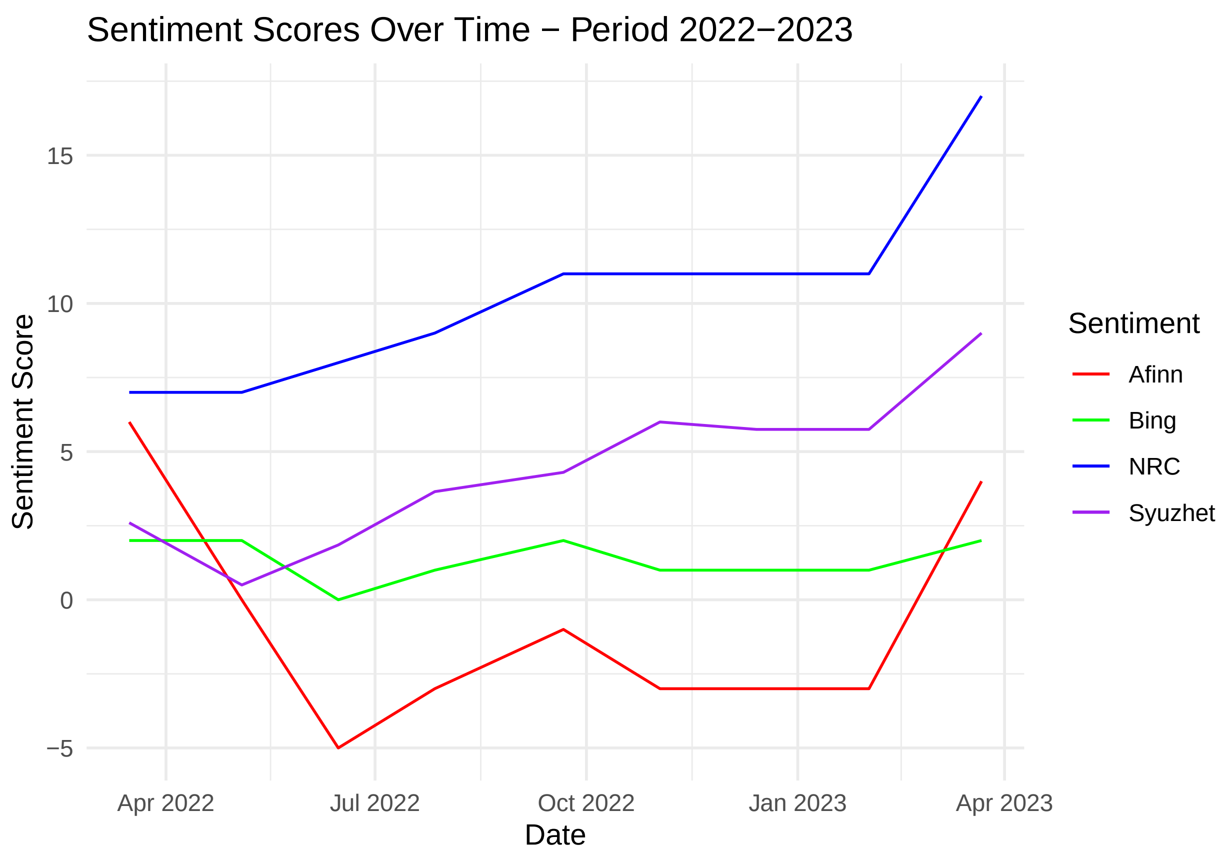

```r
# Plot 2: Sentiment Scores per method for the two periods
ggplot(fomc_df_with_sentiment, aes(x = date)) +
  geom_point(aes(y = sentiment_afinn, color = "Afinn")) +
  geom_point(aes(y = sentiment_bing, color = "Bing")) +
  geom_point(aes(y = sentiment_nrc, color = "NRC")) +
  geom_point(aes(y = sentiment_syuzhet, color = "Syuzhet")) +
  scale_color_manual(name = "Sentiment",
                     values = c("Afinn" = "red",
                                "Bing" = "green",
                                "NRC" = "blue",
                                "Syuzhet" = "purple")) +
  labs(x = "Date", y = "Sentiment Score",
       title = "Sentiment Scores Over Time") +
  theme_minimal()
```




```r
# Plot 3: Sentiment Scores per method for the Period 2007-2009
ggplot(subsetted_data1, aes(x = date)) +
  geom_line(aes(y = sentiment_afinn, color = "Afinn")) +
  geom_line(aes(y = sentiment_bing, color = "Bing")) +
  geom_line(aes(y = sentiment_nrc, color = "NRC")) +
  geom_line(aes(y = sentiment_syuzhet, color = "Syuzhet")) +
  scale_color_manual(name = "Sentiment",
                     values = c("Afinn" = "red",
                                "Bing" = "green",
                                "NRC" = "blue",
                                "Syuzhet" = "purple")) +
  labs(x = "Date", y = "Sentiment Score",
       title = "Sentiment Scores Over Time - Period 2007-2009") +
  theme_minimal()
```




```r
# Plot 4: Sentiment Scores per method for the Period 2007-2009
ggplot(subsetted_data2, aes(x = date)) +
  geom_line(aes(y = sentiment_afinn, color = "Afinn")) +
  geom_line(aes(y = sentiment_bing, color = "Bing")) +
  geom_line(aes(y = sentiment_nrc, color = "NRC")) +
  geom_line(aes(y = sentiment_syuzhet, color = "Syuzhet")) +
  scale_color_manual(name = "Sentiment",
                     values = c("Afinn" = "red",
                                "Bing" = "green",
                                "NRC" = "blue",
                                "Syuzhet" = "purple")) +
  labs(x = "Date", y = "Sentiment Score",
       title = "Sentiment Scores Over Time - Period 2022-2023") +
  theme_minimal()
```



Exporting the CSV file to do more visualizations using Python

```r
# Exporting the "fomc_df_with_sentiment" dataframe to a csv file
write.csv(fomc_df_with_sentiment, file = "fomc_df_with_sentiment.csv", row.names = FALSE)
```

Creating the DFM

```r
# Creating the DFM
fomc_dfm <- tokens(fomc_df$text,
                                 remove_punct = T,
                                 remove_numbers = T,
                                 remove_symbols = T) %>%
  dfm(tolower = T) %>%
  dfm_remove(c(stopwords("english"), "http","https","rt", "t.co")) %>%
  dfm_trim(min_docfreq = 5)
```