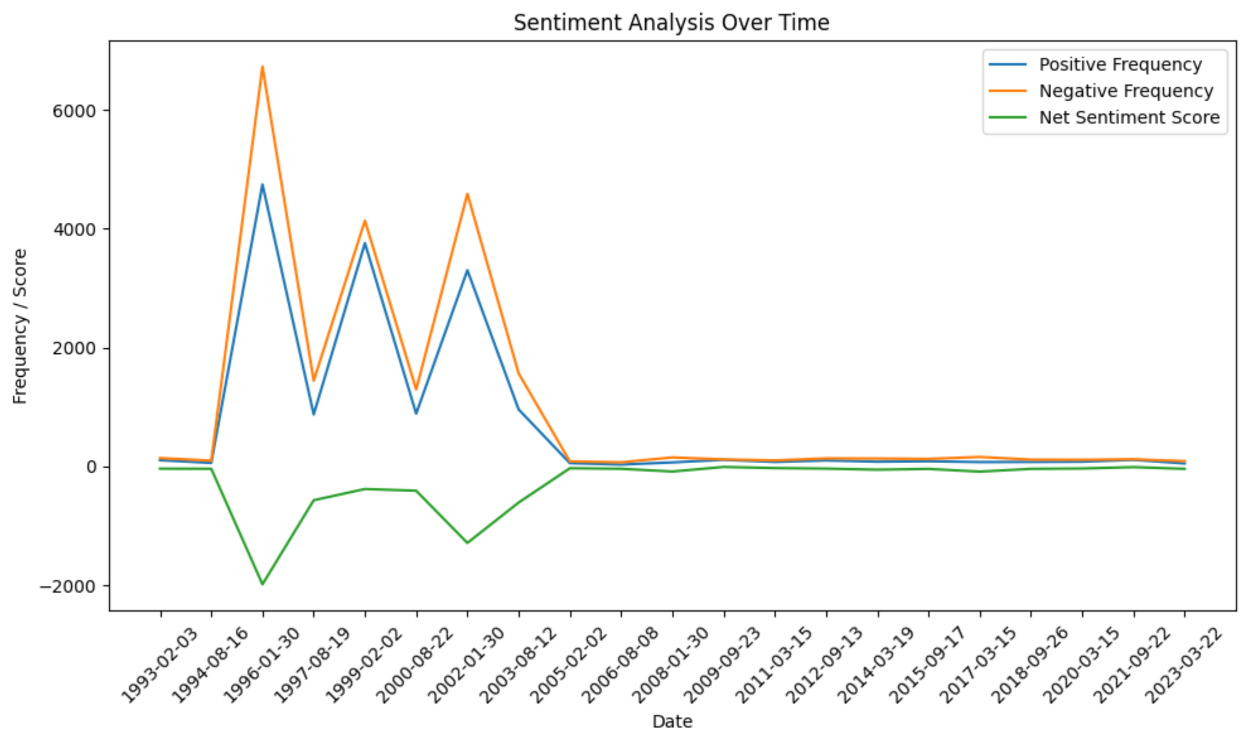
Before starting Exploratory Data Analysis, it's important to quantify the Fed Minutes that we downloaded in ingestion. First, the Loughran and McDonald Sentiment Word List is loaded. This is a list of specifically accounting and finance terms that appear in earning calls. It then attaches a positive or negative sentiment to each of these words. Joining this list with the minutes gives a total number of positive and negative words. Joining the words from this list to the text from our preprocessed text gives us a number of relevant words, their sentiment, and their frequency. The resulting number of positive and negative words and their frequency gives us a Net Sentiment Score.

**Exploratory Data Analysis – Feature Engineering**

Starting with the Net Sentiment Score that was just created, we evaluate a time series plot of the frequency of positive and negative words along with the Net Sentiment Score.




We see large skewness (something addressed later in this report) in the sentimentality of Fed minutes. There seemed to be much more "verbal activity", or many more words per minute that matched the Loughran and McDonald Sentiment Word List in the 90's and early 2000's. This will have to be addressed, as this skewness may result in a training set that is overly sentimental. That, or the data needs to be standardized in some way.

The top 10 most frequent positive and negative words also show some insight.

| Top 10 Positive Words: | Top 10 Negative Words: |
|---|---|
| stability: 10691 | risk: 19105 |
| good: 9887 | somewhat: 15138 |
| gain: 9566 | could: 14166 |
| strong: 8211 | might: 12763 |
| strength: 6654 | anticipated: 9005 |
| favorable: 5772 | suggested: 8972 |
| despite: 4858 | appeared: 8182 |
| positive: 3594 | may: 8116 |
| strengthening: 3244 | uncertainty: 8002 |
| improvement: 3028 | possible: 4184 |

Looking at the frequency, it is illuminating what kinds of words are not only used most frequently in the FOMC minutes, but also what kinds of words are given sentiment. “Somewhat”, “could”, and “might” are in the same level of frequency as “good”, “gain”, and “strong” despite having ambiguous sentimentality.

Turning back to the question of standardizing the frequency of positive and negative words, one option is to calculate the proportion of positive and negative words per minute rather than the total number, while also making data for the length of published minutes, as longer speeches may confer some sentimentality themselves. The result of all of the data cleaning after conferring sentiment to the minutes is presented thusly:

```
       Positive Frequency  Negative Frequency  Standardized Sentiment Score  \
count          241.000000          241.000000                    241.000000
mean           449.107884          652.659751                      0.752044
std            778.324288         1075.912586                      0.343809
min             33.000000           36.000000                     -1.000000
25%             73.000000          103.000000                      0.817644
50%             95.000000          139.000000                      0.907344
75%            451.000000          816.000000                      0.926072
max           5233.000000         6779.000000                      1.000000

       Proportion Positive Words  Proportion Negative Words     Word Count
count                 241.000000                 241.000000     241.000000
mean                    0.018770                   0.027514   21659.730290
std                     0.004213                   0.005540   36682.444249
min                     0.008655                   0.015183    2141.000000
25%                     0.015711                   0.023604    4247.000000
50%                     0.018689                   0.027465    5419.000000
75%                     0.021529                   0.030701   20171.000000
max                     0.030320                   0.042019  246599.000000
```

We see summary statistics in line with our previous findings: the mean negative frequency is higher than the mean positive frequency, and resulting standardized sentiment score that reflects this. The data is now ready to be analyzed across other ingested variables.

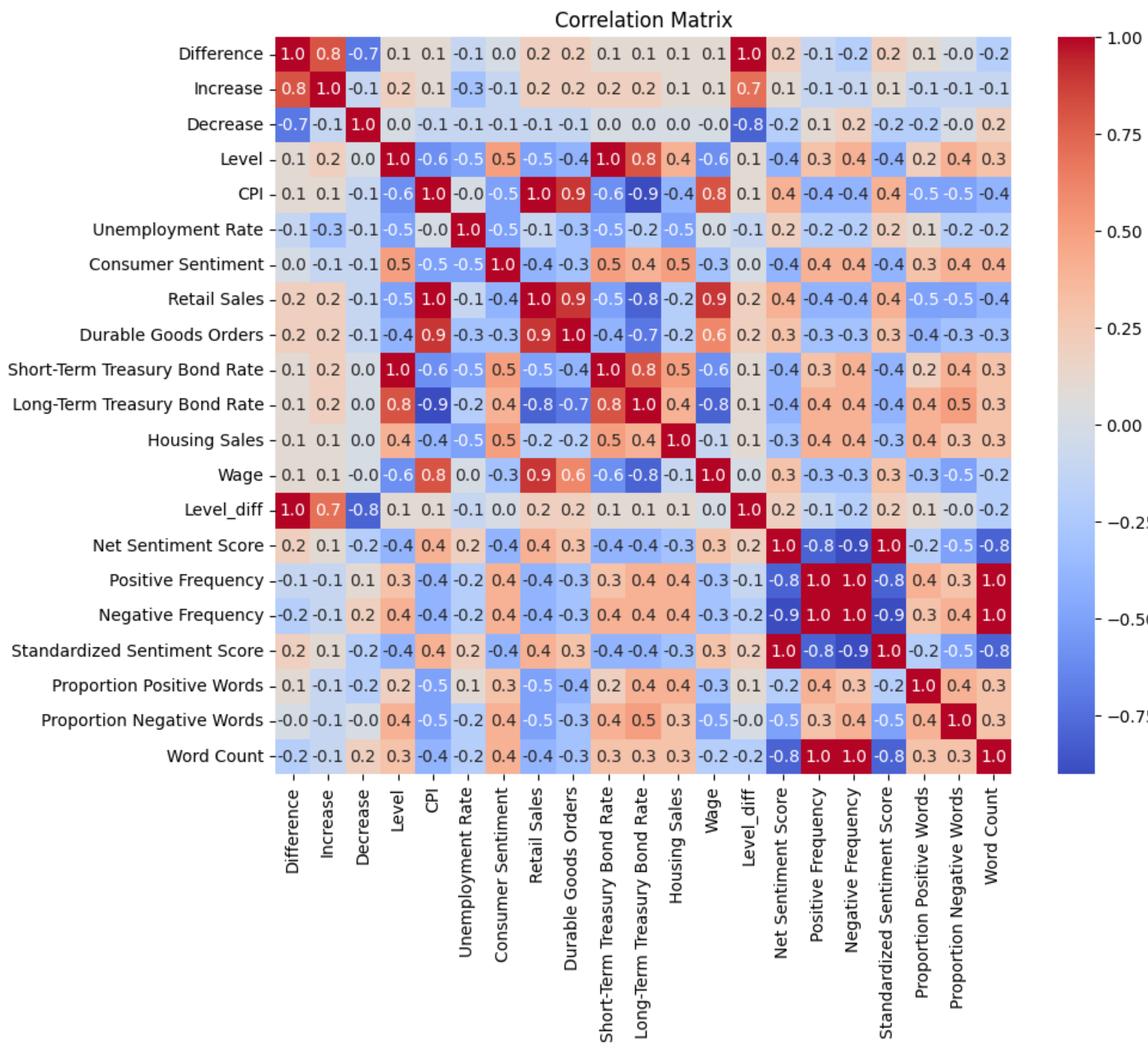
## Exploratory Data Analysis – Summary Statistics

| | Difference | Increase | Decrease | Level | CPI | Unemployment Rate | Consumer Sentiment | Retail Sales | Durable Goods Orders | Short-Term Treasury Bond Rate | Long-Term Treasury Bond Rate |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **count** | 241 | 241 | 241 | 241 | 241 | 241 | 241 | 241 | 241 | 241 | 241 |
| **mean** | 0.01971 | 0.06639 | 0.04668 | 2.510373 | 208.510365 | 5.703734 | 86.708714 | 328983.058 | 201234.739 | 2.270124 | 3.889129 |
| **std** | 0.22448 | 0.152244 | 0.144876 | 2.151646 | 40.584948 | 1.746695 | 13.348263 | 108840.128 | 35495.5388 | 2.07081 | 1.738902 |
| **min** | -1 | 0 | 0 | 0.25 | 143.1 | 3.4 | 50 | 156266 | 127233 | 0.01 | 0.65 |
| **25%** | 0 | 0 | 0 | 0.25 | 173.6 | 4.4 | 77.5 | 249845 | 173849 | 0.15 | 2.36 |
| **50%** | 0 | 0 | 0 | 1.75 | 211.445 | 5.4 | 89.3 | 321794 | 200956 | 1.71 | 3.89 |
| **75%** | 0 | 0 | 0 | 4.75 | 237.761 | 6.2 | 95.9 | 396630 | 226400 | 4.53 | 5.18 |
| **max** | 0.75 | 0.75 | 1 | 6.5 | 302.918 | 13.2 | 112 | 601983 | 297311 | 6.19 | 7.96 |

| | Housing Sales | Wage | Level_diff | Net Sentiment Score | Positive Frequency | Negative Frequency | Standardized Sentiment Score | Proportion Positive Words | Proportion Negative Words | Word Count |
|---|---|---|---|---|---|---|---|---|---|---|
| **count** | 241 | 241 | 240 | 241 | 241 | 241 | 241 | 241 | 241 | 241 |
| **mean** | 714.672199 | 337.896266 | 0.005208 | -203.551867 | 449.107884 | 652.659751 | 0.752044 | 0.01877 | 0.027514 | 21659.7303 |
| **std** | 257.430631 | 16.121467 | 0.27581 | 348.794218 | 778.324288 | 1075.91259 | 0.343809 | 0.004213 | 0.00554 | 36682.4442 |
| **min** | 270 | 312 | -1.5 | -1981 | 33 | 36 | -1 | 0.008655 | 0.015183 | 2141 |
| **25%** | 532 | 331 | 0 | -137 | 73 | 103 | 0.817644 | 0.015711 | 0.023604 | 4247 |
| **50%** | 696 | 336 | 0 | -46 | 95 | 139 | 0.907344 | 0.018689 | 0.027465 | 5419 |
| **75%** | 880 | 345 | 0 | -27 | 451 | 816 | 0.926072 | 0.021529 | 0.030701 | 20171 |
| **max** | 1389 | 393 | 0.75 | 48 | 5233 | 6779 | 1 | 0.03032 | 0.042019 | 246599 |

We see some interesting statistics immediately. Firstly, that since 1993, the maximum basis Federal Funds target rate increase has been 75 basis points while the maximum rate increase has been 100 basis points. That said, the mean for increases is higher than it is for decreases, signaling that we have increased the Federal Funds target rate more often or more severely than we have decreased it since 1993. Further analysis on the interaction between variables will follows:

**Exploratory Data Analysis – Correlation**



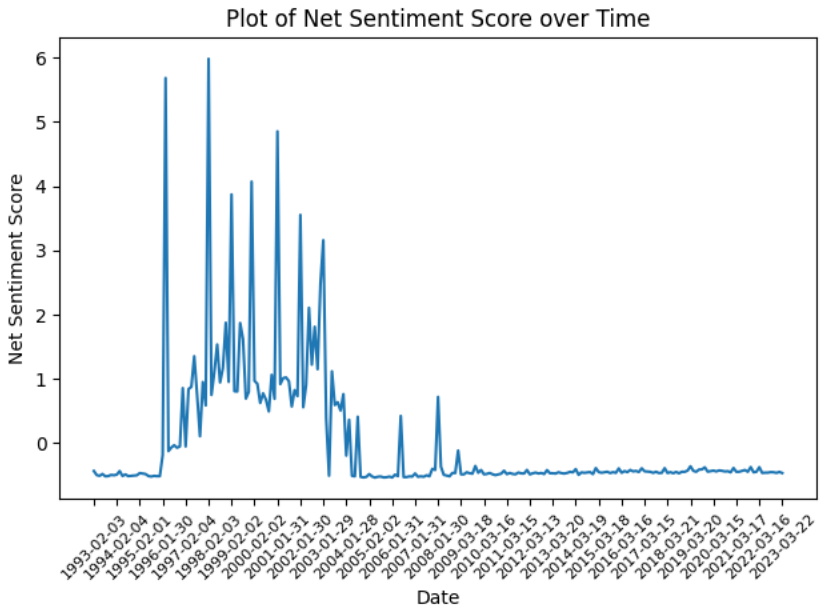

While this correlation plot is somewhat congested, we see some interesting correlations. Firstly, the Federal Funds Rate and the Short-Term Treasury Bond Rate seem to move in almost lockstep fashion. This is somewhat disconcerting for the paper, as it seems to potentially negate the need for it. If you want to know where the Federal Funds rate is going, just look at Short-Term Treasury Bonds. There is macroeconomic reasoning behind this as well. The Short-Term Bond rate is the rate at which the Treasury loans out money to corporations and banks, and is set by the market rate for these bonds. It can be thought of as similar to the "risk-free" rate, since the federal government is almost never going to default. The Federal Funds Rate is the rate at which institutional banks can trade with each other. This is instead set by the Federal Reserve through open market operations, which increases or decreases the amount of reserves in the banking system, which allows banks to balance the risk profile of trading with each other. Because the Fed's enaction of open market operations is a sign of where they think short-term rates should be, institutional investors tend to agree that Short-Term Treasury Bonds should be roughly at the same rate.

Returning to the analysis, one point of interest is that the Short-Term Treasury Bond Rate (as well as the Long-Term Treasury Bond Rate), despite being so correlated to the Federal Funds Rate, is entirely uncorrelated to Decreases, while only being marginally correlated to increases. This indicates that it is very difficult for investment banks to predict when a decrease will happen as well as the magnitude. As such, the purpose of this paper may well change direction to only predicting when the Federal Funds Rate will decrease as well as by how much.

One quick note on dimension reduction. Looking at the top correlation coefficients above 0.80, we see a few variables that may be considered superfluous. CPI x Retail Sales has a correlation coefficient of 0.984, and Durable Goods Orders x Retail Sales has a correlation coefficient of 0.884. It also shows up in Retail Sales x Wage at 0.867. It also has a correlation coefficient with Long-Term Treasury Bond Rates at 0.805. It may be smarter to just use Retail Sales and reduce my dimensionality by 4, but increased description power may be more useful depending on the model chosen. As such, a note will be made of this and revisited during model creation.

Moving to autocorrelations, there is not much to be gleaned. Most macroeconomic factors have generally consistent autocorrelations, and the last observation shows the most autocorrelation with autocorrelations tapering off. The only non-normal autocorrelation we see is in the frequency-related autocorrelations. Every lag at t=8, there seems to be a jump in autocorrelation when the rest of the lags look standard. Unfortunately, all this shows is that quarterly reports (frequency = 52) tend to be longer than non-quarterly reports.

Looking at the time series to double check these values, we see some interesting statistics. While the time series look generally sporadic and normal, the time series for Net Sentiment Score is very skewed.




We see that FOMC minutes in the 1990's were much more volatile. Looking into the z-scores for skewness and kurtosis for just this variable and not on average across the dataset, we see verification of these findings:

Z-score for skewness: 20.447

Z-score for kurtosis: 40.218

This is either true, or the calculation of Net Sentiment Score requires a new vocabulary dictionary to reassess the true nature of sentiment in Fed minutes. *Further analysis is required.*

**Exploratory Data Analysis – Skewness**

Finally, the z-score for skewness and z-score for kurtosis is analyzed for all of the data. This can be analyzed, and is in the code, but skewness is more meant for cross-sectional data than it is for time series. That said, it can show us underlying trends in the data. And indeed, looking at individual time series of all of the data, we see some that is extremely skewed. This is true of the Net Sentiment Score and the frequency of both positive and negative words. What gives them all away is the skewness of the Word Count. The FOMC minutes in the 1990's and early 2000's were many times longer than the minutes of today. As such, the decision to instead calculate the proportion of positive and negative words did much to normalize the data and show no outliers.

**Data Partitioning**

Moving on to data partitioning, we make the decision to split the data into training/validation/test sets in an 80%/10%/10% manner. This results in the number of observations as 192/24/25. This also correlates to time frames, as breaking up time series data in its logical order provides numerous benefits. Firstly, temporal dependency could occur, which shows certain correlations in certain time frames that could lead to overfitting. The point of the validation and testing sets will be to remove any bias that happens in the training set. This is especially true because the first observation of the test set happens to also be the first date of Jerome Powell's tenure as Fed Chair, and it has been said that he has a particularly open and transparent method to forward guidance that is unlike his predecessors. Testing over this time frame will be particularly important. Further, training for this test set makes for a more realistic goal: we will be with Jerome Powell for at least two more years, and a model that can accurately decipher his particular language will be that much more useful. Finally, setting hyperparameters in a validation set in chronological order provides an unbiased estimate of the model's performance on future data, which is the ultimate goal of this project.