

Corresponding Box Folder Link: <https://uofi.box.com/s/9sl8hrlq5ky2khgq3skgkg52fo2ppy5n>

Steps in Pipeline:

- Splitting videos of Federal Reserve press conferences of 2021, 2022, 2023 and 2024 (until June) into 1 minute, 5 minute and 10 minute intervals using MoviePy
- Extracting audio of each video interval using MoviePy
- Extracting speech of each interval from the audio file using OpenAI Whisper
- Get price data for the press conference dates for the same 1 minute, 5 minute and 10 minute intervals
- Calculate price change $((\text{close price} - \text{open price}) / \text{open price})$ and volatility $((\text{high price} - \text{low price}) / \text{low price})$ for each interval
- Get official Federal Reserve statements for each announcement and perform sentence tokenization on the statement texts
- Get news articles published before, on the day of the press conference announcement and perform sentence tokenization on the news article texts
- Encode the statement and news sentences using the Sentence Transformers model 'multi-qa-mpnet-base-dot-v1' and insert the vectors into the Pinecone vector database into two difference indexes – 'statement' and 'news'

Analysis:

1) Interval-Level Text and Similarity Score Extraction and Analysis:

- Create threshold true/false columns to identify intervals with absolute value of price change greater than 0.1%, 0.25%, 0.5% and 0.75% respectively.
- Calculate the price movement during the statement release on the press conference date (price movement from 2:00pm ET to 2:05pm ET)
- Query the 'statement' index in the Pinecone database, and extract 15, 20, 25 and 30 closest matching sentences and their corresponding average similarity score from the statement for each press conference interval speech text.
- Query the 'news' index in the Pinecone database, and extract 20, 25, 30 and 35 closest matching sentences and their corresponding average similarity score from news for each press conference interval speech text.
- Use GPT to analyze each interval's press conference speech text and the highest number of extracted statement sentences (30) and news sentences (35) for 3 key similar terms/phrases and 3 key different terms/phrases – **Prompt 1**
- Separate the data into two parts – press conference dates with positive price movement during statement release (from 2pm ET to 2:05 pm ET) and press conference dates with negative price movement during statement release (from 2pm ET to 2:05pm ET)
- **Result files: *combined_filtered_positive.csv* (positive statement price movement) and *combined_filtered_negative.csv* (negative statement price movement) for each of the 1 Min, 5 Min and 10 Min intervals contain the data and analysis described**

2) Price Movement Prediction using Text:

- Using data of interval speech text, extracted similar statement text, extracted similar news text and interval price movement (positive/negative) from 2023 and 2024, split the data into training and test data using a stratified split.
- Use GPT to learn how patterns of particular sentiments, terminologies and economic indicators between the interval speech text and either the statement or news text correlates to the price movement.
- Use GPT to predict the binary price movement (positive/negative) for the test data, and measure classification metrics – **Prompt 2 (Statement), Prompt 3 (News)**
- **Result File: Test Data Results on Training for Price Movement Prediction.xlsx contains classification result metrics for different configurations**

3) Price Movement Prediction using Similarity Score:

- Using data of interval speech text, average similarity score of extracted similar statement text, average similarity score of extracted similar news text and interval price movement (positive/negative), split the data into training and test data using a stratified split.
- Use GPT to learn how the similarity scores between the interval speech text and either the statement or news text correlates to the price movement.
- Use GPT to predict the binary price movement (positive/negative) for the test data, and measure classification metrics – **Prompt 4**
- **Result File: 1Min_results_price_movement_predictions.csv, 5Min_results_price_movement_predictions.csv and 10Min_results_price_movement_predictions.csv contains classification result metrics for different configurations**

4) Volatility Results:

- volatility_overall.csv: This calculation involves finding the volatility measure $((\text{high price} - \text{low price}) / \text{low price})$ of every interval from $t=0$ which is 2:30pm ET (start time of the press conference) until the end of the press conference. Then, the standard deviation (square root of sum of squares) of the volatility price measures for each t from $t=0$ until the end of press conference (maximum t) across all intervals with the given combinations of interval price movement $((\text{close price} - \text{open price}) / \text{open price})$ during interval) and statement price movement $((\text{close price} - \text{open price}) / \text{open price})$ during statement release (2:00pm ET to 2:05pm ET) is found - combinations are positive & positive, positive & negative, negative & positive, negative & negative
- For each of the intervals 1 Min, 5 Min and 10 Min:
 - i. volatility_positive.xlsx: This calculation involves finding the volatility measure $((\text{high price} - \text{low price}) / \text{low price})$ of every 1 min/5 min/10 min interval from $t-60$ minutes to end of the press conference where $t=0$ is 2:30pm ET (start time of the press conference). Then, the standard deviation (square root of sum of squares) of the volatility price measures for each t from $t-60$ to end of press conference (maximum t) across all dates where the price movement $((\text{close price} - \text{open price}) / \text{open price})$ during statement release (2:00pm ET to 2:05pm ET) is positive is found.

- ii. volatility_negative.xlsx: This calculation involves finding the volatility measure $((\text{high price} - \text{low price}) / \text{low price})$ of every 1 min/5 min/10 min interval from t-60 minutes to end of the press conference where t=0 is 2:30pm ET (start time of the press conference). Then, the standard deviation (square root of sum of squares) of the volatility price measures for each t from t-60 to end of press conference (maximum t) across all dates where the price movement $((\text{close price} - \text{open price}) / \text{open price})$ during statement release (2:00pm ET to 2:05pm ET) is negative is found.

5) Price Change Box Plots:

- For each of the intervals 1 Min, 5 Min, 10 Min:
 - i. For each combination – positive statement price movement and positive interval price movement, positive statement price movement and negative interval price movement, negative statement price movement and positive interval price movement, and negative statement price movement and negative interval price movement, constructed a box plot of the interval price change values.
 - ii. Result Files:
positive_price_change_box_plot_positive_statement_movement.svg,
positive_price_change_box_plot_negative_statement_movement.svg,
negative_price_change_box_plot_positive_statement_movement.svg,
negative_price_change_box_plot_negative_statement_movement.svg

6) Similar and Different Terms Histograms:

- For each of the intervals 1 Min, 5 Min, 10 Min:
 - i. For each combination – positive statement price movement and positive interval price movement, positive statement price movement and negative interval price movement, negative statement price movement and positive interval price movement, and negative statement price movement and negative interval price movement: constructed 6 word frequency histograms for each combination – 1 for similar terms of statement and news each, 1 for different terms of speech text for statement and news each, and 1 for different terms of baseline text (statement/news) for statement and news each.

GPT Prompts:

1) **Prompt 1:**

```
system_prompt = ""
As a financial analyst tasked with comparing two texts, your objective is
to identify similarities and differences in specific
terminology used in the two provided texts. Identify 3 similar and 3
different key terms between the first text and
the second text. Pay close attention to recurring themes and phrases that
encapsulate the
```

```

    essence of the texts, and use that to identify 3 key similar terms, and 3
    key terms that are significantly different and unique to each text.
    """

    user_prompt = f"""
    Read the following texts. Then do the following:

    1. Identify and list 3 key terms or phrases that are similar between the
    first text and the second text.
    2. Identify and list 3 key terms or phrases that are significantly
    different between the first text and the second text and unique to each text.

    First Text:
    "{transcript}"

    Second Text:
    "{text}"

    Output format: Return the output in the specified format.
    Do NOT include any additional text. Return the output in string format,
    NOT in JSON format.

    {{
        "Similar": [term/phrase 1, term/phrase 2],
        "Different": [(term/phrase 1 from first text, term/phrase 2 from
    second text), (term/phrase 3 from first text, term/phrase 4 from second text)]
    }}
    """

```

2) Prompt 2:

```

system_prompt = """
    You are an analysis assistant specialized in monetary policy and economic
    indicators. Your task is to predict future price movements based on
    detailed analyses of Federal Open Market Committee (FOMC) meeting
    statements and press conferences. The baseline text is the official summary of
    FOMC decisions. Analyze the key differences in economic outlook, inflation
    expectations, labor market conditions, and intentions for monetary
    policy adjustments, focusing on the nuances in language and emphasis.
    """

    user_prompt = f"""
    Steps:

    1. Understanding the Data:
    - Each record in the data list contains:
        - `Id`: The ID of the record.
        - `First Text`: The FOMC press conference transcript to analyze.
    """

```

- `Second Text`: The baseline text containing the official FOMC meeting statement summary.
- `Price Movement`: The price movement [Positive/Negative]

2. Detailed Analysis:

- Compare the `First Text` with the `Second Text`, focusing on:
 - Changes in emphasis on economic growth, inflation expectations, labor market conditions, and financial stability.
 - Adjustments in language related to future monetary policy direction, such as potential rate hikes, pauses, or cuts.
 - Variations in detailed economic outlook and risk assessments.
 - Overall sentiment and terminology used to describe economic conditions and policy stances.

3. Prediction Based on Analysis:

- Determine whether these textual differences and overall economic outlook would likely lead to a Positive or Negative price movement.

4. Training Data Insights:

- Learn from the provided training data to understand patterns of how particular sentiments, terminologies, and economic indicators correlate with price movements.

Training Data:

{train_prompt}

Test Data:

{test_prompt}

Predicting Test Data:

For each record in the test data list, predict the price movement (Positive/Negative).

Output Format: Return the output as a list of predictions and a summary of the insights in the specified format. Do NOT include any additional text.

Return the output in

string format, NOT in JSON format.

```
{{
  "predictions":
  [
    {{
      "Id": [Id of the test data record in integer format],
      "Price Movement": [Prediction of price movement as
Positive/Negative]
    }},
    {{
      "Id": [Id of the test data record in integer format],
```

```

        "Price Movement": [Prediction of price movement as
Positive/Negative]
    }}
],
    "insights": [text summary of the learnings of how patterns of particular
sentiments, terminologies, and economic indicators correlate with price
movements]
    }}
"""

```

3) Prompt 3:

```

system_prompt = """
    You are an analysis assistant specialized in monetary policy and economic
indicators. Your task is to predict future price movements based on
    detailed analyses of news articles related to the Federal Open Market
Committee (FOMC) meetings and press conferences. The baseline text is a
collection of
    related news articles published on the same day as and before the FOMC
meeting press conference. Analyze the key differences in economic outlook,
    inflation expectations, labor market conditions, and intentions for
monetary policy adjustments, focusing on the nuances in language and emphasis.
    """

user_prompt = f"""
Steps:

1. Understanding the Data:
- Each record in the data list contains:
    - `Id`: The ID of the record.
    - `First Text`: The FOMC press conference transcript to analyze.
    - `Second Text`: The baseline text containing related news articles.
    - `Price Movement`: The price movement [Positive/Negative]

2. Detailed Analysis:
- Compare the `First Text` with the `Second Text`, focusing on:
    - Changes in emphasis on economic growth, inflation expectations,
labor market conditions, and financial stability.
    - Adjustments in language related to future monetary policy direction,
such as potential rate hikes, pauses, or cuts.
    - Variations in detailed economic outlook and risk assessments.
    - Overall sentiment and terminology used to describe economic
conditions and policy stances.

3. Prediction Based on Analysis:
- Determine whether these textual differences and overall economic outlook
would likely lead to a Positive or Negative price movement.

```

```

4. Training Data Insights:
- Learn from the provided training data to understand patterns of how
particular sentiments, terminologies, and economic indicators
correlate with price movements.

Training Data:
{train_prompt}

Test Data:
{test_prompt}

Predicting Test Data:
For each record in the test data list, predict the price movement
(Positive/Negative).

Output Format: Return the output as a list of predictions and a summary of
the insights in the specified format. Do NOT include any additional text.
Return the output in
string format, NOT in JSON format.

{{
  "predictions":
  [
    {{
      "Id": [Id of the test data record in integer format],
      "Price Movement": [Prediction of price movement as
Positive/Negative]
    }},
    {{
      "Id": [Id of the test data record in integer format],
      "Price Movement": [Prediction of price movement as
Positive/Negative]
    }}
  ],
  "insights": [text summary of the learnings of how patterns of particular
sentiments, terminologies, and economic indicators correlate with price
movements]
}}
"""

```

4) Prompt 4:

```

system_prompt = """
You are an advanced analysis assistant with expertise in economic
indicators and monetary policy. Your primary objective is to predict future
price movements based on

```

a detailed quantitative analysis of Federal Open Market Committee (FOMC) meeting statements, news articles, and press conference transcripts. The key metric provided to you

is the average similarity score, which represents the quantitative similarity between the press conference transcript and a baseline text (either the FOMC meeting statement

or related news articles). Your task is to interpret this score and use it to accurately predict price movements.

"""

```
user_prompt = f"""
```

Steps:

1. Understand the Data:

- Each data record consists of:
 - `Id`: A unique identifier for the record.
 - `Average Similarity Score`: The computed average similarity score between the press conference transcript and the baseline text (either the FOMC meeting statement or news articles).
 - `Price Movement`: The recorded price movement [Positive/Negative].

2. Analytical Focus:

- Analyze the relationship between the `Average Similarity Score` and the `Price Movement`.
- Identify patterns or thresholds in the similarity score that are indicative of either a Positive or Negative price movement.

3. Predictive Modeling:

- Based on your analysis, predict the price movement (Positive/Negative) for each record in the test data set, using the similarity score as the primary input.

4. Insights from Training Data:

- Draw conclusions from the training data, highlighting any discovered trends or correlations between similarity scores and price movements.
- Use these insights to inform your predictions on the test data.

Training Data:

```
{train_prompt}
```

Test Data:

```
{test_prompt}
```

Output Predictions:

For each record in the test data set, predict the price movement (Positive/Negative).

Output Format: Provide your output as a structured list of predictions followed by a summary of insights. Avoid any extraneous text. The output should be in a simple string format, not JSON.

```
{
  "predictions": [
    {
      "Id": [Id of the test data record in integer format],
      "Price Movement": [Prediction of price movement as
Positive/Negative]
    },
    {
      "Id": [Id of the test data record in integer format],
      "Price Movement": [Prediction of price movement as
Positive/Negative]
    }
  ],
  "insights": [Concise summary of the relationship between the average
similarity score and price movement]
}
```