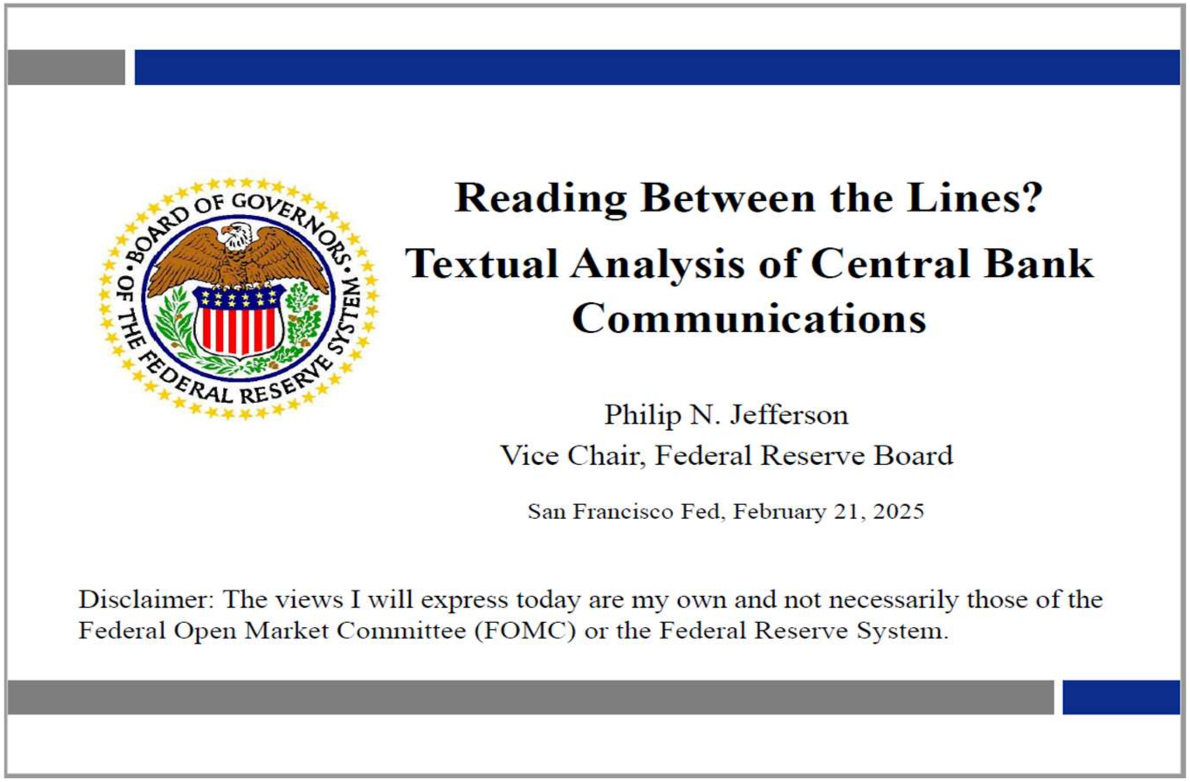
Decoding the Fed: NLP Analysis of Federal Reserve Communications (2015-2025)

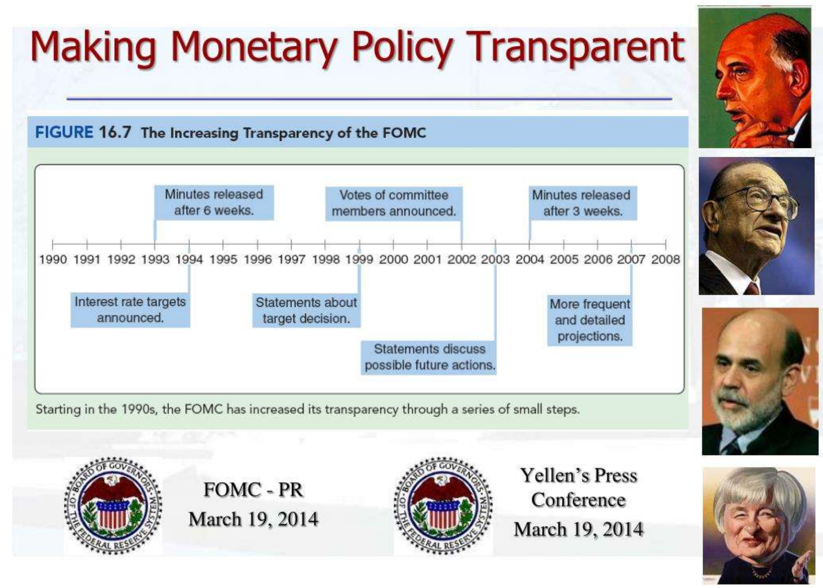
Background:

- ❑ Natural Language Processing (NLP) plays a crucial role in extracting insights from institutional communications, particularly in the financial sector.
- ❑ The Federal Reserve's meeting minutes contain valuable information about monetary policy, economic outlook, and market sentiment.
- ❑ Financial analysts, economists, policymakers, and investors use NLP to analyze trends in central bank communications to anticipate policy changes and economic shifts.

Krishna Khandelwal



https://www.federalreserve.gov/newsevents/speech/jefferson20250221a.htm



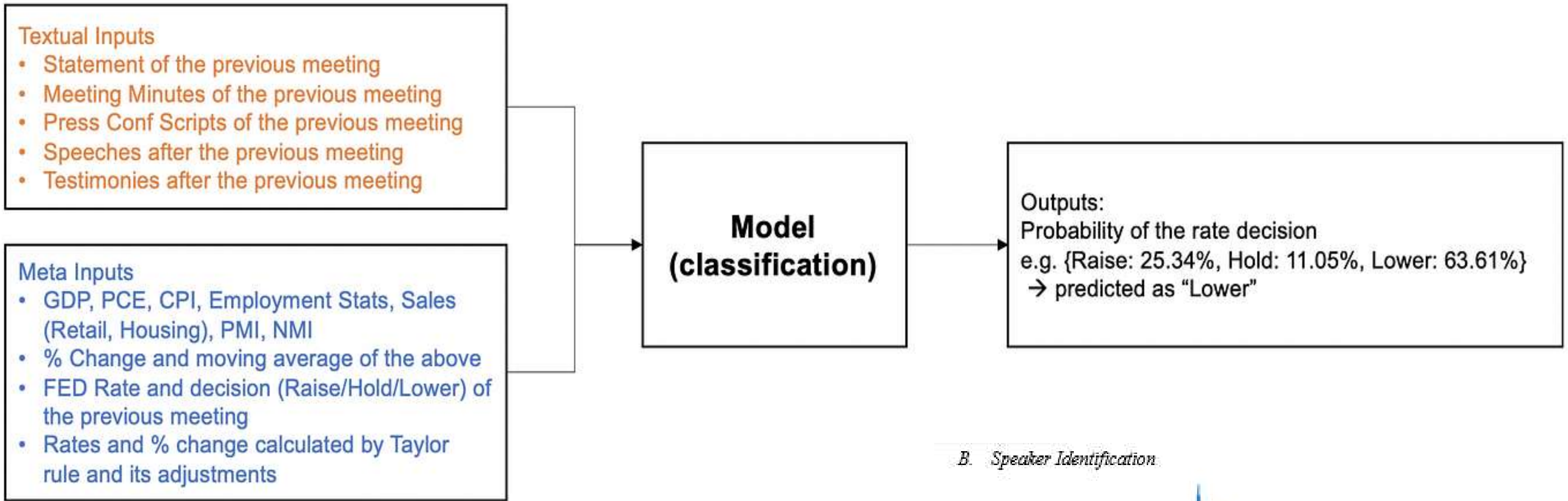
https://www.slideserve.com/louvain/the-federal-reserve-system

Importance of central bank communication analysis:

Central banks, especially the U.S. Federal Reserve, are not just economic institutions — they are narrative powerhouses. Their statements, meeting minutes, and press releases influence financial markets, investor expectations, and even employment trends long before policy is enacted.

Why Analyze Their Language?

- Small shifts in tone or emphasis can signal interest rate changes, inflation concerns, or economic outlooks.
- Markets react not just to what is said, but how it's said.
- Understanding these signals offers predictive economic insight, helping investors, policymakers, and researchers stay ahead of real-world decisions.



https://medium.com/data-science/fedspeak-how-to-build-a-nlp-pipeline-to-predict-central-bank-policy-changes-a2f157ca0434

B. Speaker Identification

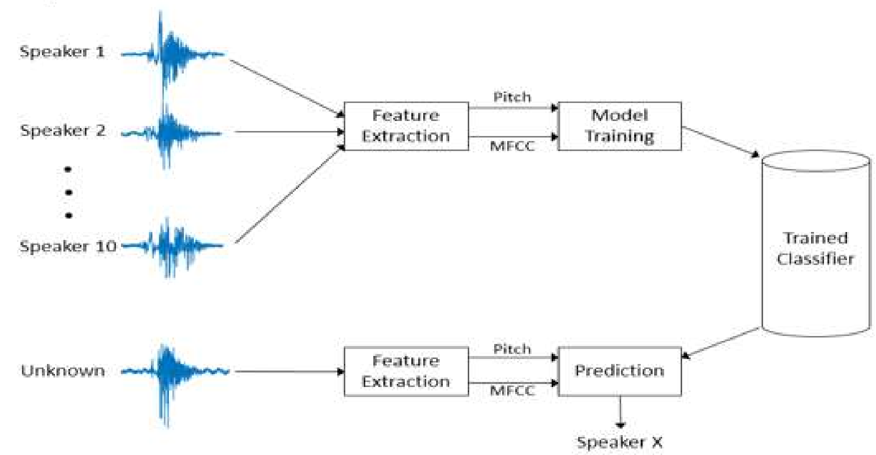


Fig. Unique audio recognition for speaker identification

https://www.ijraset.com/research-paper/automated-minutes-of-meeting-using-a-multimodal-approach

Data Collection – Source, methodology, dataset overview

```
1 df.sample(10)
```

	URL	Date	Year	Month	Day	Source	Content
29	https://www.federalreserve.gov/monetarypolicy/...	2023-11-01	2023	11	1	Main Calendar	HomeMonetary PolicyFederal Open Market Committ...
124	https://www.federalreserve.gov/monetarypolicy/...	2015-04-29	2015	4	29	Historical Archive	HomeMonetary PolicyFederal Open Market Committ...
21	https://www.federalreserve.gov/monetarypolicy/...	2024-03-20	2024	3	20	Main Calendar	HomeMonetary PolicyFederal Open Market Committ...
110	https://www.federalreserve.gov/monetarypolicy/...	2016-06-15	2016	6	15	Historical Archive	HomeMonetary PolicyFederal Open Market Committ...
78	https://www.federalreserve.gov/monetarypolicy/...	2019-01-30	2019	1	30	Historical Archive	HomeMonetary PolicyFederal Open Market Committ...
7	https://www.federalreserve.gov/monetarypolicy/...	2024-12-18	2024	12	18	Main Calendar	HomeMonetary PolicyFederal Open Market Committ...
52	https://www.federalreserve.gov/monetarypolicy/...	2022-09-21	2022	9	21	Main Calendar	HomeMonetary PolicyFederal Open Market Committ...
17	https://www.federalreserve.gov/monetarypolicy/...	2024-06-12	2024	6	12	Main Calendar	HomeMonetary PolicyFederal Open Market Committ...
51	https://www.federalreserve.gov/monetarypolicy/...	2022-09-21	2022	9	21	Main Calendar	HomeMonetary PolicyFederal Open Market Committ...
38	https://www.federalreserve.gov/monetarypolicy/...	2023-05-03	2023	5	3	Main Calendar	HomeMonetary PolicyFederal Open Market Committ...

Entries show successful extraction of structured meeting metadata, including URL, date breakdown, source origin, and raw content.
Both recent and historical FOMC meetings are included, ensuring longitudinal analysis."

The scraper retrieved **128 Federal Reserve** meeting transcripts, structured across **7 key** metadata fields including date, source, and content.

```
[ ]    1 df.shape
(128, 7)
```

Length of the articles

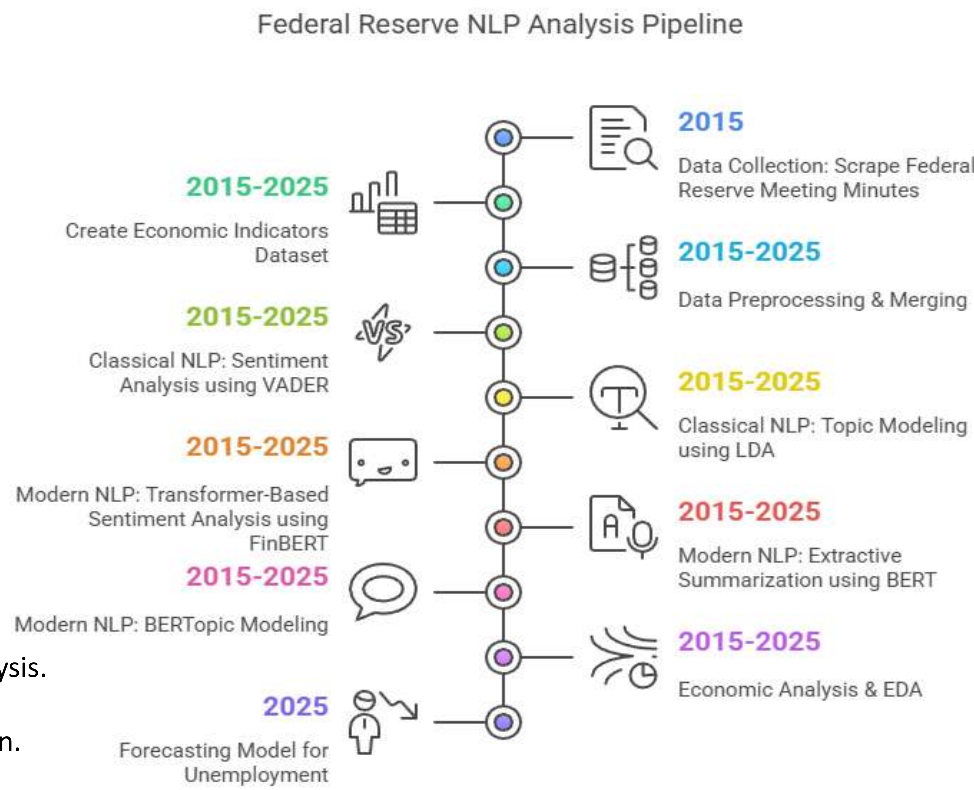
```
[ ]    1 len(minutes_data)
128
```

- **64 transcripts** from the **Main Calendar** (recent meetings)
- **64 transcripts** from the **Historical Archive** (older meetings)

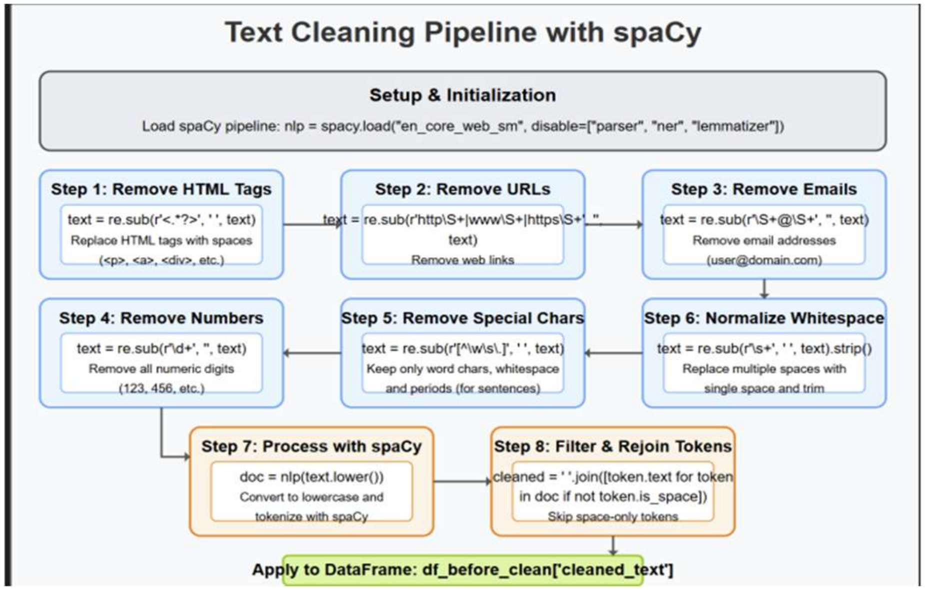
Source	
Main Calendar	64
Historical Archive	64

System Architecture

1. Data Collection (2015)
Scraped FOMC meeting minutes from 2015–2025 using web automation.
2. Data Preprocessing & Merging (2015–2025)
Cleaned, tokenized, and combined raw text with date, source, and economic indicators.
3. Classical NLP – Sentiment (2015–2025)
Applied VADER to extract basic sentiment scores from Fed statements.
4. Classical NLP – Topic Modeling (2015–2025)
Used LDA to identify recurring themes in central bank language over time.
5. Modern NLP – Sentiment with FinBERT (2015–2025)
Leveraged FinBERT for more accurate, domain-specific financial sentiment analysis.
6. Modern NLP – BERTopic (2015–2025)
Performed advanced topic modeling using transformer embeddings + UMAP.
7. Modern NLP – Extractive Summarization (2015–2025)
Used BERT-based models to auto-summarize long Fed transcripts.
8. **Create Economic Indicator Dataset (2015–2025)**
Built datasets on unemployment, inflation, etc., for correlation with text analysis.
9. **Economic Analysis & EDA (2015–2025)**
Visualized sentiment and topics alongside economic data for insight extraction.
10. Forecasting Model (2025)
Designed predictive models to forecast unemployment using textual sentiment + topics.



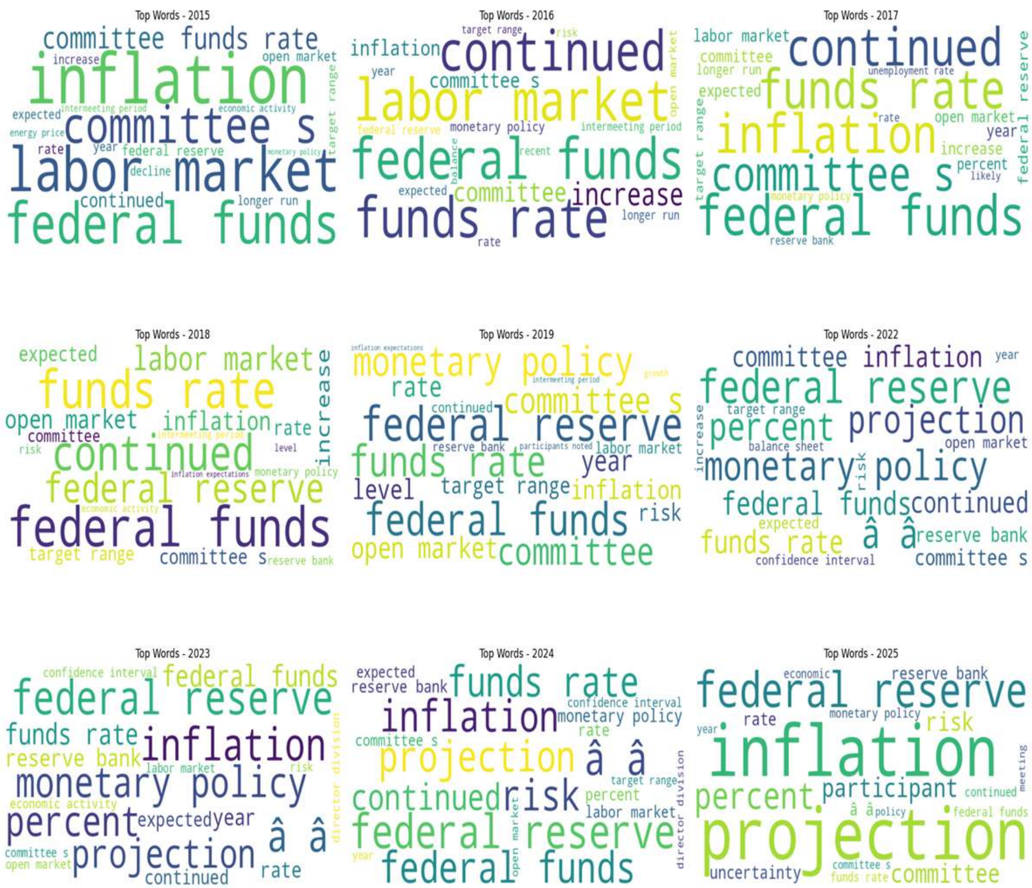
Text Preprocessing Pipeline (spaCy + RegEx)



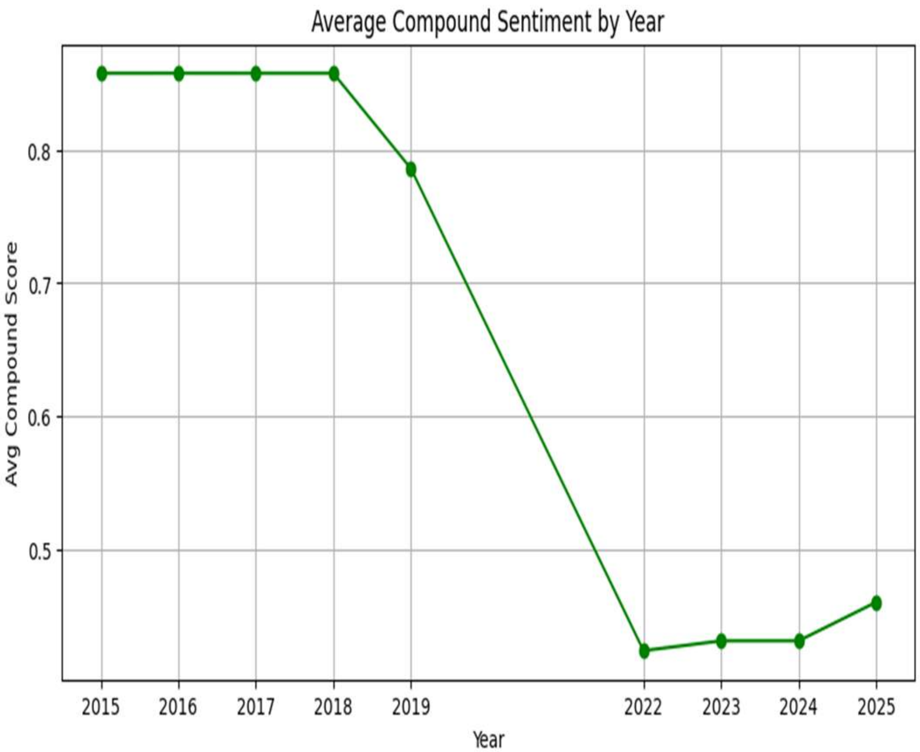

```python
import re
import spacy

nlp = spacy.load("en_core_web_sm", disable=["parser", "ner", "lemmatizer"])

def clean_text(text):
    text = re.sub(r"<.*?>", '', text)  # remove HTML
    text = re.sub(r"http\S+|www\S+|https\S+", '', text)  # remove URLs
    text = re.sub(r'\S+@\S+', '', text)  # remove emails
    text = re.sub(r'\d+', '', text)  # remove numbers
    text = re.sub(r"[^\w\s.]", ' ', text)  # remove special characters
    text = re.sub(r'\s+', ' ', text).strip()  # normalize whitespace
    doc = nlp(text.lower())
    return ' '.join([token.text for token in doc if not token.is_space])

df['cleaned_text'] = df['Content'].apply(clean_text)
```

- **Remove HTML tags**
  re.sub(r"<.*?>", ", text) — strips any embedded markup from transcripts.
- **Remove URLs**
  Deletes links like http://... or www... to reduce noise.
- **Remove email addresses**
  Eliminates unnecessary references or contact info.
- **Remove numeric values**
  Keeps the text linguistically focused by removing digits.
- **Remove special characters**
  Only retains letters and basic spacing.
- **Normalize whitespace**
  Converts multiple spaces/tabs into single spaces.
- **Lowercasing + Tokenization (spaCy)**
  Transforms text to lowercase and splits it into clean word tokens, filtering out whitespace tokens.

Top Words in Fed Communications (2015–2025)



- ✓ Visualizes the most frequent terms used in FOMC meeting minutes for each year.
- ✓ Word clouds are generated from cleaned and tokenized text data.
- ✓ Common high-frequency terms include:
  "federal funds," "committee," "rate" — consistently central across all years.
- ✓ "Inflation" becomes a dominant term in 2022–2025, reflecting real-world economic concerns.
- ✓ New terms like "projection," "percent," and "risk" gain prominence in later years.
- ✓ Earlier years focus more on structural policy terms like "target range" and "intermeeting period".
- ✓ Provides an intuitive overview of how Federal Reserve priorities evolved over a decade.

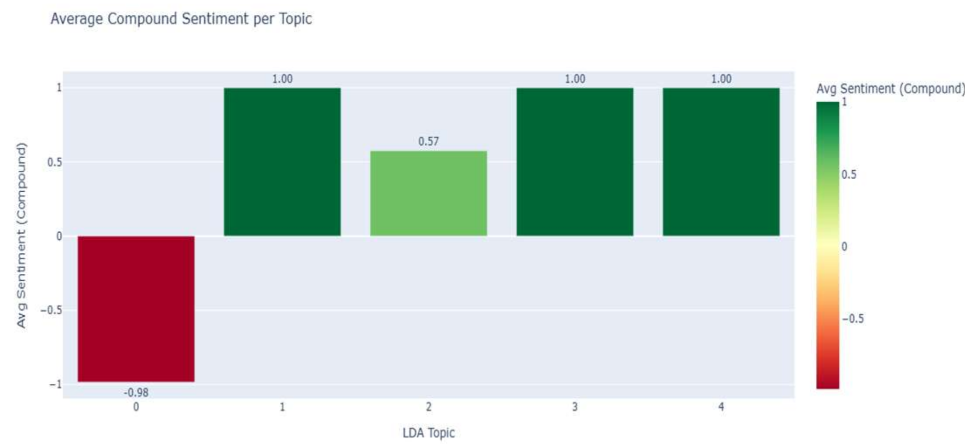


Average Compound Sentiment by Year (VADER Analysis):

- ✓ The chart shows the average compound sentiment score of FOMC meeting minutes from 2015 to 2025.
- ✓ Scores were generated using VADER, a rule-based sentiment analysis tool.
- ✓ 2015–2018: Sentiment remains consistently positive (above 0.85), reflecting stable economic outlook.
- ✓ 2019: Slight decline in sentiment suggests early signals of uncertainty.
- ✓ 2020–2022: Sharp drop in sentiment, hitting a low around 0.43 in 2022, possibly due to:
  - ✓ COVID-19 pandemic
  - ✓ Inflationary pressures
  - ✓ Global economic disruptions
- ✓ 2023–2025: Minor recovery in sentiment, though still significantly below pre-2019 levels.
- ✓ Indicates a shift in tone from confidence to caution and risk awareness in Fed communications.

Topic Trends Over Time (2015–2025):

- ✓ Visualizes how LDA-generated topics shift in prominence across years.
- ✓ Each line represents a different topic (0–4), plotted by document share.
- ✓ Notable trends:
    - ✓ Topic 2 grows steadily post-2019 — may indicate sustained policy focus.
    - ✓ Topic 4 fades after 2020 — likely short-lived relevance.
    - ✓ Topic 1 peaks around 2017 then vanishes — linked to temporary themes.
- ✓ Highlights changing focus of FOMC communications in response to global economic events.

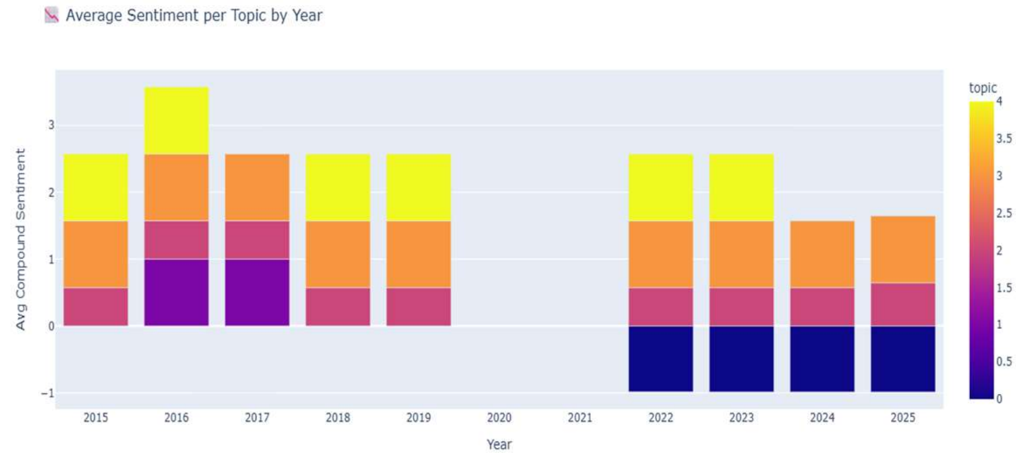


Average Sentiment per Topic:

- ✓ Shows average compound sentiment for each topic (0–4), based on VADER analysis.
- ✓ Topics 1, 3, and 4 are strongly positive (score = 1.00).
- ✓ Topic 2 is moderately positive (score ≈ 0.57).
- ✓ Topic 0 is highly negative (score = -0.98), likely representing risk- or crisis-related content.
- ✓ Useful for linking emotional tone to semantic themes in monetary policy.

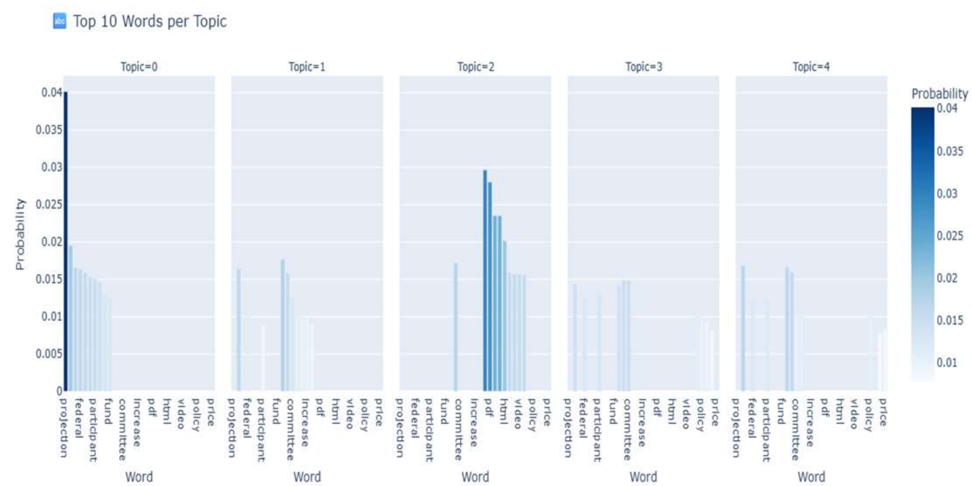
Average Sentiment per Topic by Year:

- ✓ Combines sentiment and topic modeling to analyze year-wise emotional trends.
- ✓ Stacked bars display how each topic's sentiment varies annually.
- ✓ Topic 0 disappears in early years, but dominates from 2022–2025, contributing negative tone.
- ✓ Sentiment composition changes from optimistic multi-topic mix to risk-heavy tone after 2022.
- ✓ Reveals the emotional shift in Fed language alongside economic uncertainty.



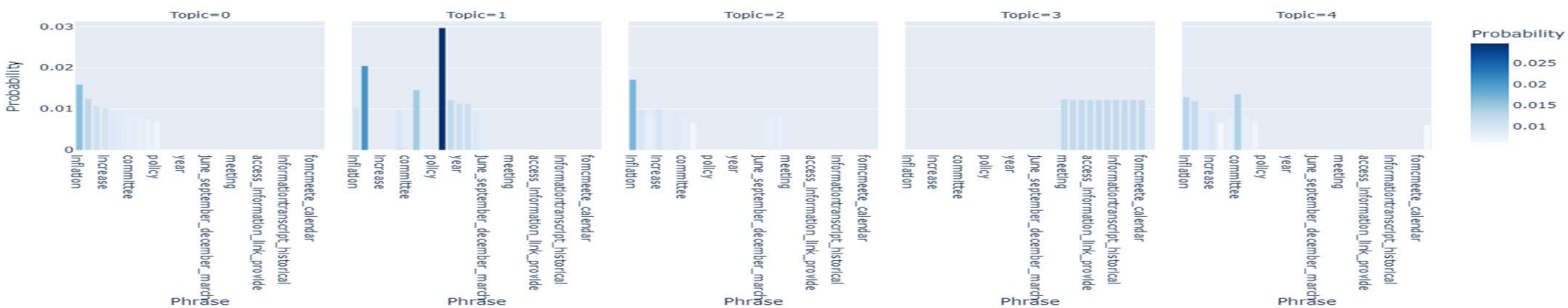
Top 10 Words per Topic – LDA (Analysis):

- ✓ Clear topic-word associations indicate effective topic separation.
- ✓ Higher word probabilities show strong topic coherence.
- ✓ Some noise terms like *html*, *pdf*, and *video* suggest minor data cleaning issues.
- ✓ Distribution reflects dominant economic themes across discussions.
- ✓ Useful for mapping recurring policy language in FOMC communications.



Phrase Popularity Rank Shift Over Time (Top 10 Trigrams)

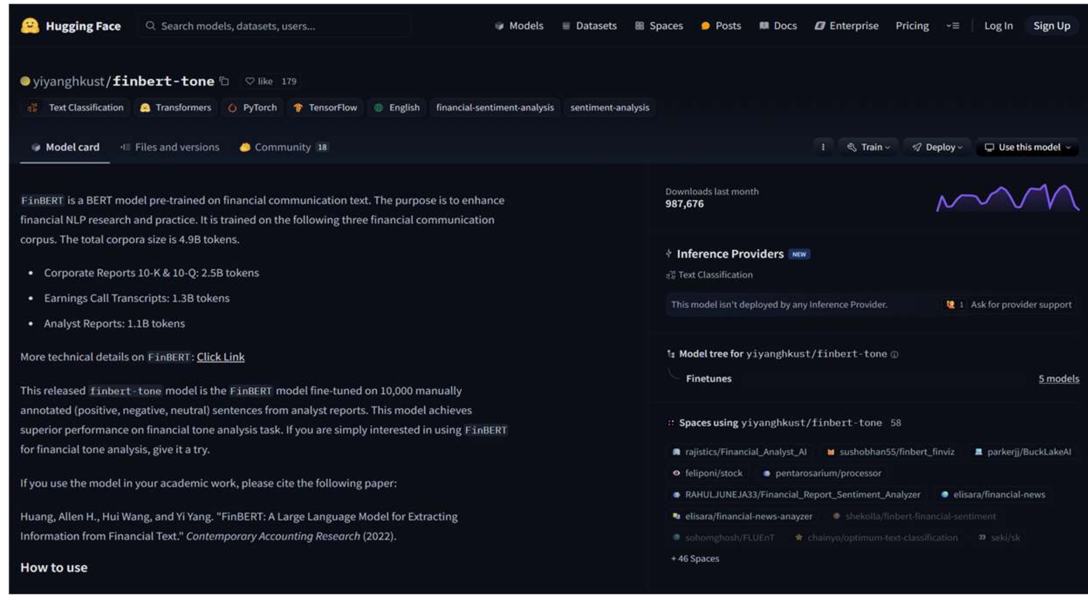
Top Bigrams/Trigrams per Topic (LDA with N-Grams)



Why Use FinBERT-Tone?

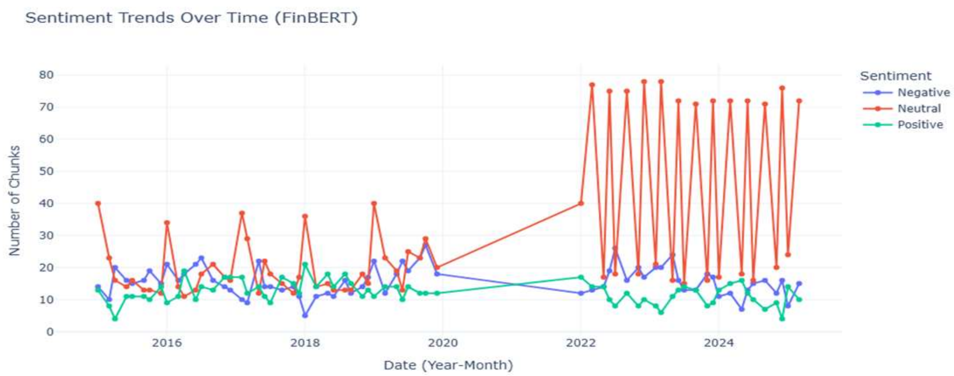
- ✓ **Domain-Specific:** Pre-trained exclusively on financial text (10-Ks, earnings calls, analyst reports).
- ✓ **Finetuned for Sentiment:** Trained on 10,000+ annotated financial sentences (positive, negative, neutral).
- ✓ **Accurate Tone Detection:** Outperforms generic sentiment models like VADER on financial documents.
- ✓ **Trusted & Popular:** Widely used in finance-related NLP projects (nearly 1M downloads/month).

https://huggingface.co/yiyanghkust/finbert-tone




```python
# Load the FinBERT model from Hugging Face
from transformers import AutoTokenizer, AutoModelForSequenceClassification

model_name = "yiyanghkust/finbert-tone"

# Load tokenizer and model
tokenizer = AutoTokenizer.from_pretrained(model_name)
model = AutoModelForSequenceClassification.from_pretrained(model_name)

# Create sentiment pipeline
sentiment_pipeline = pipeline("sentiment-analysis", model=model, tokenizer=tokenizer)

```

- ✓ Loaded **FinBERT** (yiyanghkust/finbert-tone) — a transformer model fine-tuned for financial sentiment analysis.
- ✓ Initialized both **tokenizer** and **model** using Hugging Face's transformers library.
- ✓ Built a **custom sentiment pipeline** to classify the tone (positive, negative, neutral) in Federal Reserve statements.

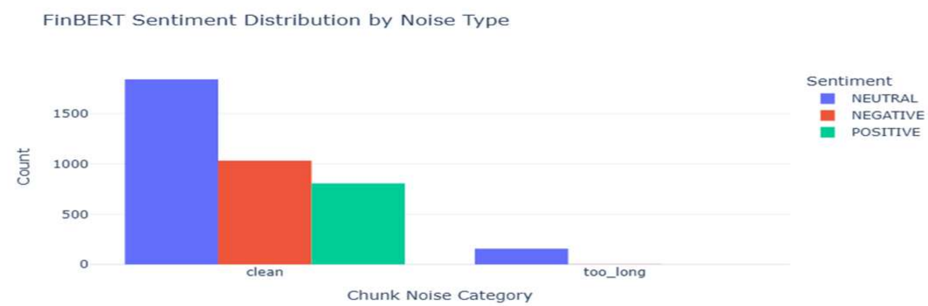


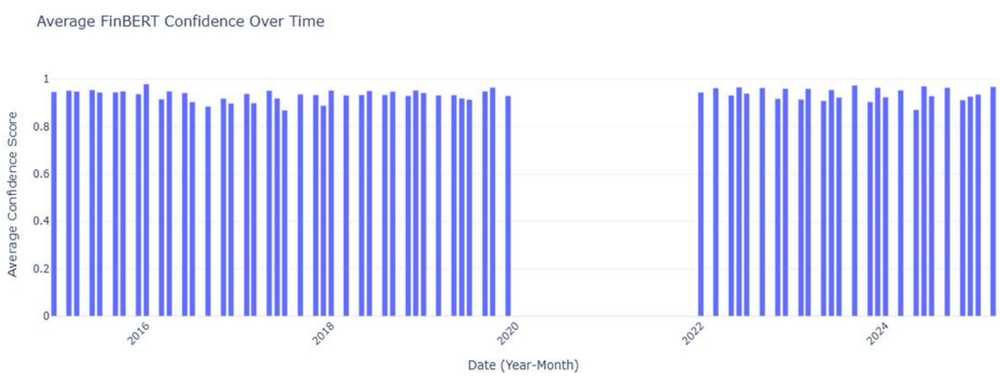
Sentiment Trends Over Time (FinBERT):

- ✓ Neutral sentiment dominates FOMC minutes, especially after 2022.
- ✓ Positive and negative sentiments show relatively stable but lower presence.
- ✓ The spike in neutral sentiment may reflect cautious and measured Fed communication post-pandemic.

Sentiment Distribution by Chunk Type:

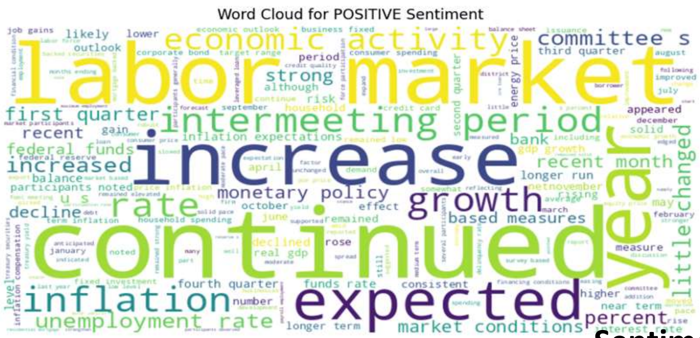
- ✓ Most clean text chunks are classified as neutral.
- ✓ Positive and negative chunks are fewer, with noise mainly in long texts.
- ✓ Indicates FinBERT performs reliably on well-preprocessed content.





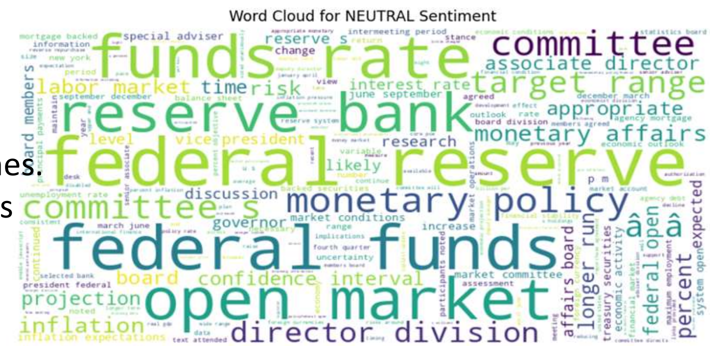
FinBERT Confidence Over Time:

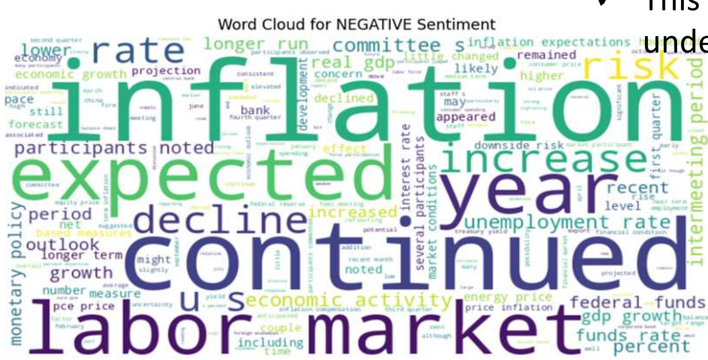
- ✓ Confidence scores remain high (close to 1) across all years.
- ✓ Confirms FinBERT's consistent reliability in analyzing sentiment in financial texts.
- ✓ Occasional dips may indicate more ambiguous or complex communications.



Sentiment Word Cloud Analysis (FinBERT):

- ✓ Word clouds were generated separately for positive, negative, and neutral sentiments.
- ✓ This helped visualize the focus of central bank communication under different sentiment tones.
- ✓ Each cloud aggregates common phrases across all documents based on FinBERT's sentiment classification.
- ✓ It provides a quick, interpretable overview of sentiment-specific language trends.
- ✓ This supports deeper economic narrative understanding without manual text review.





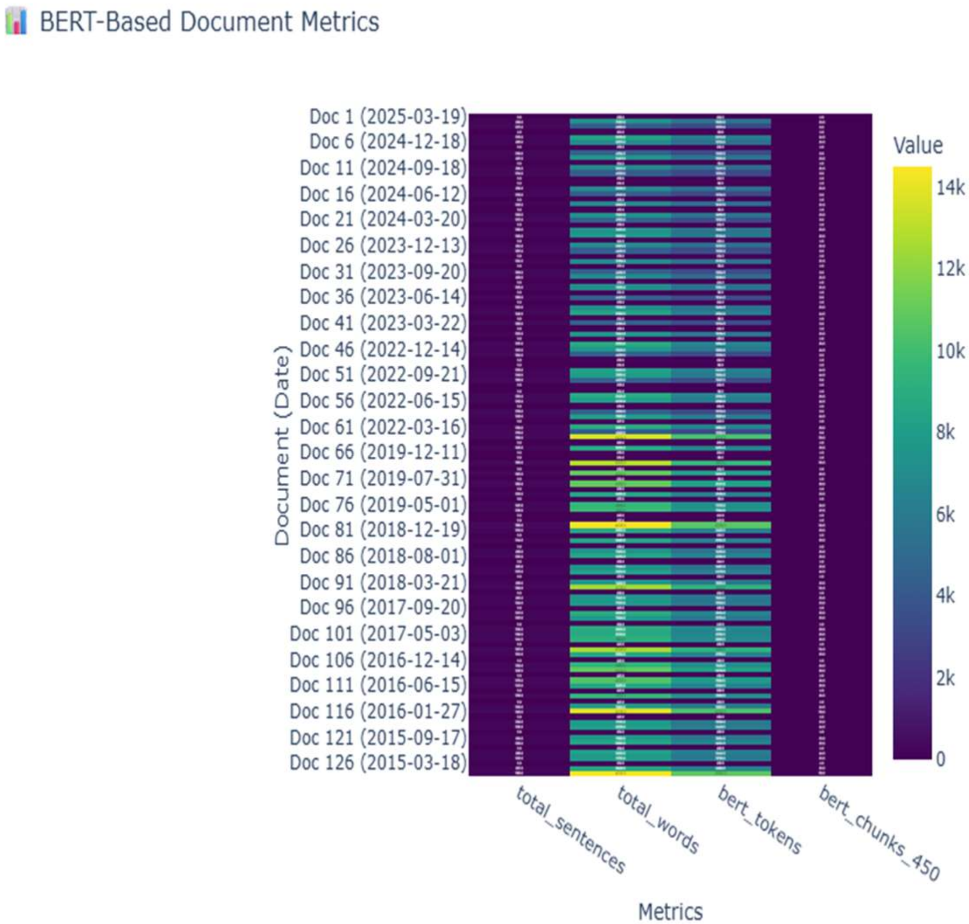
```
Top words for POSITIVE sentiment:
inflation: 1073
market: 907
remained: 875
participants: 835
economic: 789
rate: 782
continued: 707
labor: 652
quarter: 635
growth: 581
recent: 521
conditions: 518
percent: 512
credit: 478
year: 451

Top words for NEGATIVE sentiment:
inflation: 1828
participants: 1628
economic: 1092
market: 1081
rate: 1063
growth: 936
prices: 763
continued: 752
remained: 748
labor: 732
policy: 711
would: 604
financial: 599
percent: 588
year: 583

Top words for NEUTRAL sentiment:
federal: 3681
committee: 3292
inflation: 2531
rate: 2464
market: 2380
board: 2208
policy: 2194
reserve: 2049
monetary: 1843
participants: 1803
division: 1696
would: 1613
economic: 1531
funds: 1525
percent: 1490
```

Top Word Frequency by Sentiment (FinBERT):

- ✓ Frequency analysis was conducted separately for positive, negative, and neutral sentiment classes.
- ✓ This allowed identification of high-occurrence terms within each sentiment group.
- ✓ The output supports deeper investigation into tone-shaping language across FOMC statements.
- ✓ It complements topic modeling and sentiment scoring with interpretable lexical patterns.
- ✓ Helps validate consistency of FinBERT sentiment tagging with economic language tone.

This table monitors sentiment shifts between consecutive FOMC meeting statements.

	Date	finbert_label	prev_sentiment	sentiment_changed
0	2015-01-28	NEUTRAL	None	True
1	2015-03-18	NEUTRAL	NEUTRAL	False
2	2015-04-29	NEGATIVE	NEUTRAL	True
3	2015-06-17	NEGATIVE	NEGATIVE	False
4	2015-07-29	NEUTRAL	NEGATIVE	True



BERT-Based Document Metrics:

- ✓ The heatmap visualizes document-level text metrics like sentence count, word count, token count, and number of BERT chunks.
- ✓ Consistency in structure across documents is evident, though some spikes show longer transcripts (e.g., in 2019 and 2016).
- ✓ These metrics help assess text complexity and ensure fair input chunking for downstream NLP models.

```python
def chunk_text(text, max_tokens=450, overlap=100):
    doc = nlp(text)
    sents = [sent.text.strip() for sent in doc.sents]

    chunks = []
    curr, curr_count = [], 0

    for sent in sents:
        toks = TOKENIZER.tokenize(sent)
        n = len(toks)
        # if adding this sentence would overflow, finalize current chunk
        if curr_count + n > max_tokens:
            chunks.append(" ".join(curr))
            # build overlap window
            overlap_sents, count = [], 0
            for prev in reversed(curr):
                ptoks = TOKENIZER.tokenize(prev)
                if count + len(ptoks) > overlap:
                    break
                overlap_sents.insert(0, prev)
                count += len(ptoks)
            curr, curr_count = overlap_sents, count

        curr.append(sent)
        curr_count += n

    if curr:
        chunks.append(" ".join(curr))
    return chunks
```

Text Chunking Strategy for BERT:

- ✓ Long FOMC documents were split into smaller token-based chunks using a custom function.
- ✓ Each chunk is limited to 450 tokens to fit BERT's input constraints.
- ✓ An overlap of 100 tokens ensures sentence continuity and contextual flow.
- ✓ This method preserves sentence boundaries and improves sentiment classification accuracy.

MMR-based selection up to ~1500 tokens

```
from sklearn.metrics.pairwise import cosine_similarity

def mmr_summary(sents, embs, doc_emb, max_tokens=1500, lam=0.6):
    selected, sel_embs, used = [], [], set()
    cur_toks = 0
    doc_sim = cosine_similarity(embs, doc_emb.reshape(1, -1)).flatten()

    while used != set(range(len(sents))) and cur_toks < max_tokens:
        scores = {}
        for idx in set(range(len(sents))) - used:
            rel = doc_sim[idx]
            div = max(cosine_similarity(
                embs[idx:idx+1],
                np.vstack(sel_embs) if sel_embs else embs[idx:idx+1]
            ).flatten()) if sel_embs else 0.0
            scores[idx] = lam * rel - (1 - lam) * div

        best = max(scores, key=scores.get)
        tok_len = len(TOKENIZER.tokenize(sents[best]))
        if cur_toks + tok_len > max_tokens:
            break

        selected.append(sents[best])
        sel_embs.append(embs[best])
        used.add(best)
        cur_toks += tok_len

    return " ".join(selected)
```

Extractive Summarization Using MMR (Maximal Marginal Relevance):

- ✓ Selected the most relevant and diverse sentences from each document using cosine similarity.
- ✓ Combined document relevance with redundancy reduction through a tunable parameter λ = 0.6.
- ✓ Limited output to ~1500 tokens, suitable for transformer input.
- ✓ Helps retain important yet non-repetitive content for summarization tasks.

```
df
```

	URL	Date	Year	Month	Day	Source	Content	cleaned_text	doc_id	original_length
0	https://www.federalreserve.gov/monetarypolicy/...	2025-03-19	2025	3	19	Main Calendar	HomeMonetary PolicyFederal Open Market Committ...	homemonetary policyfederal open market committ...	doc1	139
1	https://www.federalreserve.gov/monetarypolicy/...	2025-03-19	2025	3	19	Main Calendar	HomeMonetary PolicyFederal Open Market Committ...	homemonetary policyfederal open market committ...	doc2	7337
2	https://www.federalreserve.gov/monetarypolicy/...	2025-03-19	2025	3	19	Main Calendar	HomeMonetary PolicyFederal Open Market Committ...	homemonetary policyfederal open market committ...	doc3	4300
3	https://www.federalreserve.gov/monetarypolicy/...	2025-01-29	2025	1	29	Main Calendar	Please enable JavaScript if it is disabled in ...	please enable javascript if it is disabled in ...	doc4	121
4	https://www.federalreserve.gov/monetarypolicy/...	2025-01-29	2025	1	29	Main Calendar	HomeMonetary PolicyFederal Open Market Committ...	homemonetary policyfederal open market committ...	doc5	8229
...	...	...	...	...	...	...	...	...	...	...
123	https://www.federalreserve.gov/monetarypolicy/...	2015-06-17	2015	6	17	Historical Archive	HomeMonetary PolicyFederal Open Market Committ...	homemonetary policyfederal open market committ...	doc124	8192
124	https://www.federalreserve.gov/monetarypolicy/...	2015-04-29	2015	4	29	Historical Archive	HomeMonetary PolicyFederal Open Market Committ...	homemonetary policyfederal open market committ...	doc125	7732
125	https://www.federalreserve.gov/monetarypolicy/...	2015-03-18	2015	3	18	Historical Archive	HomeMonetary PolicyFederal Open Market Committ...	homemonetary policyfederal open market committ...	doc126	134
126	https://www.federalreserve.gov/monetarypolicy/...	2015-03-18	2015	3	18	Historical Archive	HomeMonetary PolicyFederal Open Market Committ...	homemonetary policyfederal open market committ...	doc127	8648
127	https://www.federalreserve.gov/monetarypolicy/...	2015-01-28	2015	1	28	Historical Archive	HomeMonetary PolicyFederal Open Market Committ...	homemonetary policyfederal open market committ...	doc128	14510

128 rows × 10 columns

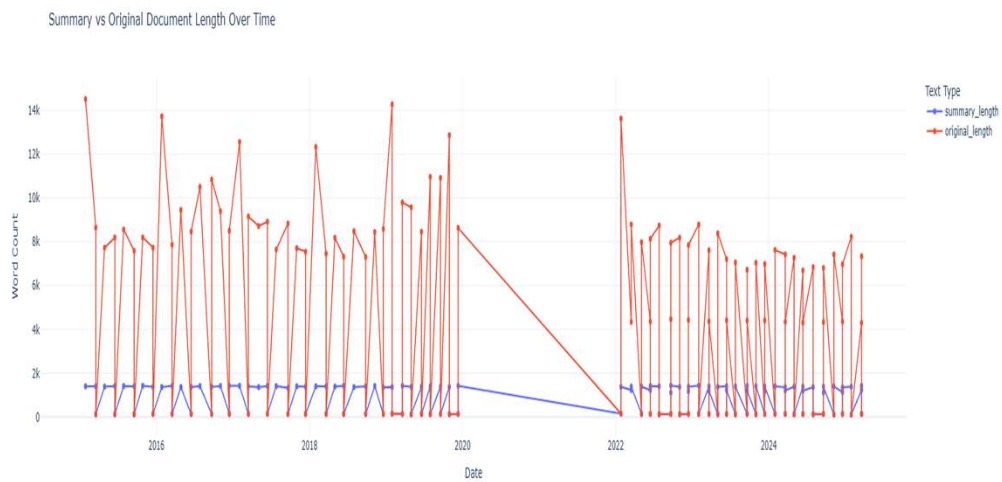
```
df
```

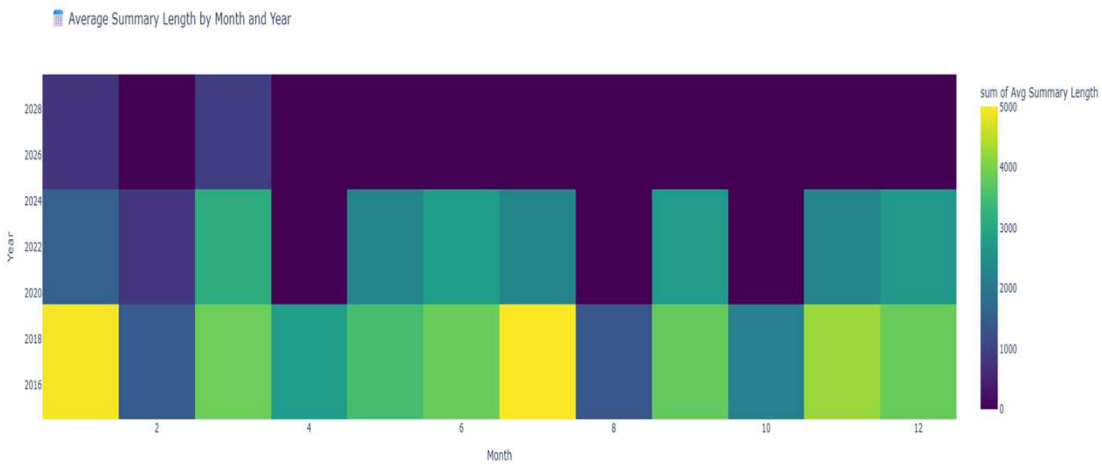
	URL	Date	Year	Month	Day	Source	Content	cleaned_text	doc_id	original_length	summary
0	https://www.federalreserve.gov/monetarypolicy/...	2025-03-19	2025	3	19	Main Calendar	HomeMonetary PolicyFederal Open Market Committ...	homemonetary policyfederal open market committ...	doc1	139	decrease volume m toggles mute on off f toggle...
1	https://www.federalreserve.gov/monetarypolicy/...	2025-03-19	2025	3	19	Main Calendar	HomeMonetary PolicyFederal Open Market Committ...	homemonetary policyfederal open market committ...	doc2	7337	as a result participants generally saw increas...
2	https://www.federalreserve.gov/monetarypolicy/...	2025-03-19	2025	3	19	Main Calendar	HomeMonetary PolicyFederal Open Market Committ...	homemonetary policyfederal open market committ...	doc3	4300	â .december projection ..... however the forec...
3	https://www.federalreserve.gov/monetarypolicy/...	2025-01-29	2025	1	29	Main Calendar	Please enable JavaScript if it is disabled in ...	please enable javascript if it is disabled in ...	doc4	121	please enable javascript if it is disabled in ...
4	https://www.federalreserve.gov/monetarypolicy/...	2025-01-29	2025	1	29	Main Calendar	HomeMonetary PolicyFederal Open Market Committ...	homemonetary policyfederal open market committ...	doc5	8229	the risks around the baseline projection for i...
...	...	...	...	...	...	...	...	...	...	...	...
123	https://www.federalreserve.gov/monetarypolicy/...	2015-06-17	2015	6	17	Historical Archive	HomeMonetary PolicyFederal Open Market Committ...	homemonetary policyfederal open market committ...	doc124	8192	there were no intervention operations in forei...
124	https://www.federalreserve.gov/monetarypolicy/...	2015-04-29	2015	4	29	Historical Archive	HomeMonetary PolicyFederal Open Market Committ...	homemonetary policyfederal open market committ...	doc125	7732	the pace of job gains had moderated and the un...
125	https://www.federalreserve.gov/monetarypolicy/...	2015-03-18	2015	3	18	Historical Archive	HomeMonetary PolicyFederal Open Market Committ...	homemonetary policyfederal open market committ...	doc126	134	decrease volume m toggles mute on off f toggle...
126	https://www.federalreserve.gov/monetarypolicy/...	2015-03-18	2015	3	18	Historical Archive	HomeMonetary PolicyFederal Open Market Committ...	homemonetary policyfederal open market committ...	doc127	8648	while the tightening of spreads was broad base...
127	https://www.federalreserve.gov/monetarypolicy/...	2015-01-28	2015	1	28	Historical Archive	HomeMonetary PolicyFederal Open Market Committ...	homemonetary policyfederal open market committ...	doc128	14510	members agreed to continue to include in the f...

128 rows × 11 columns

Extractive Summarization Output (Summary Column):

- ✓ Generated concise summaries from each FOMC document using MMR-based sentence selection.
- ✓ Each summary captures the most relevant, diverse, and non-redundant parts of the meeting minutes.
- ✓ Helps reduce document length from thousands of tokens to ~1500 while retaining key policy insights.
- ✓ Final summaries are stored in the summary column and used for downstream topic and sentiment analysis.

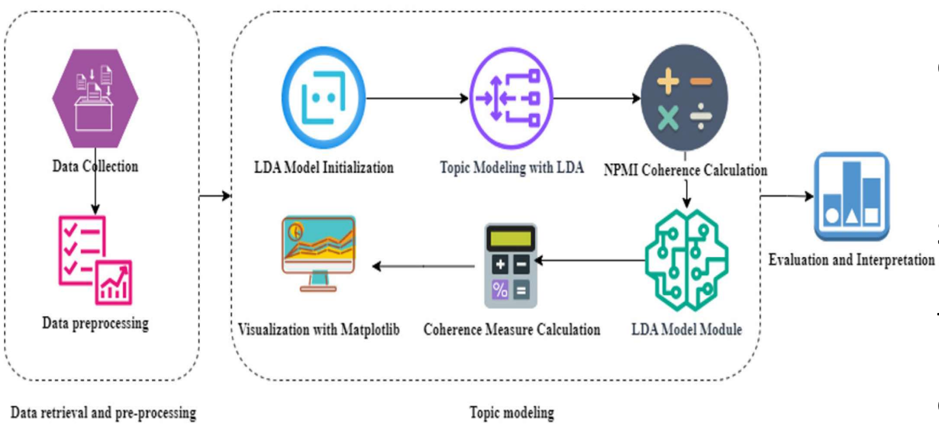




Summary Length Insights (2015–2025):

- ✓ The first chart shows a consistently **shorter summary length** compared to original document length across all years — indicating effective summarization.
- ✓ A noticeable **gap/drop occurs around 2021**, likely due to missing or failed summaries during that time.
- ✓ The heatmap reveals **2016 and 2018 had the highest average summary lengths**, particularly in months 1, 3, 6, 7, and 11.
- ✓ Post-2020, summaries became **more concise and stable**, possibly due to improved chunking or summarization methods.

Topic Modeling Pipeline



1. Data Retrieval and Preprocessing

•**Data Collection**: Text data is gathered from targeted sources (e.g., Federal Reserve meeting transcripts).

•**Preprocessing**: Includes cleaning, tokenization, stopword removal, and preparing the text for modeling.

2. Topic Modeling

•**LDA Model Initialization**: An LDA model is initialized with specified parameters (e.g., number of topics).

•**Topic Extraction**: The LDA algorithm identifies distributions of topics across documents and of words within topics.

•**NPMI Coherence Calculation**: Each topic's semantic coherence is measured using Normalized Pointwise Mutual Information (NPMI), a metric for interpretability.

3. Coherence Evaluation

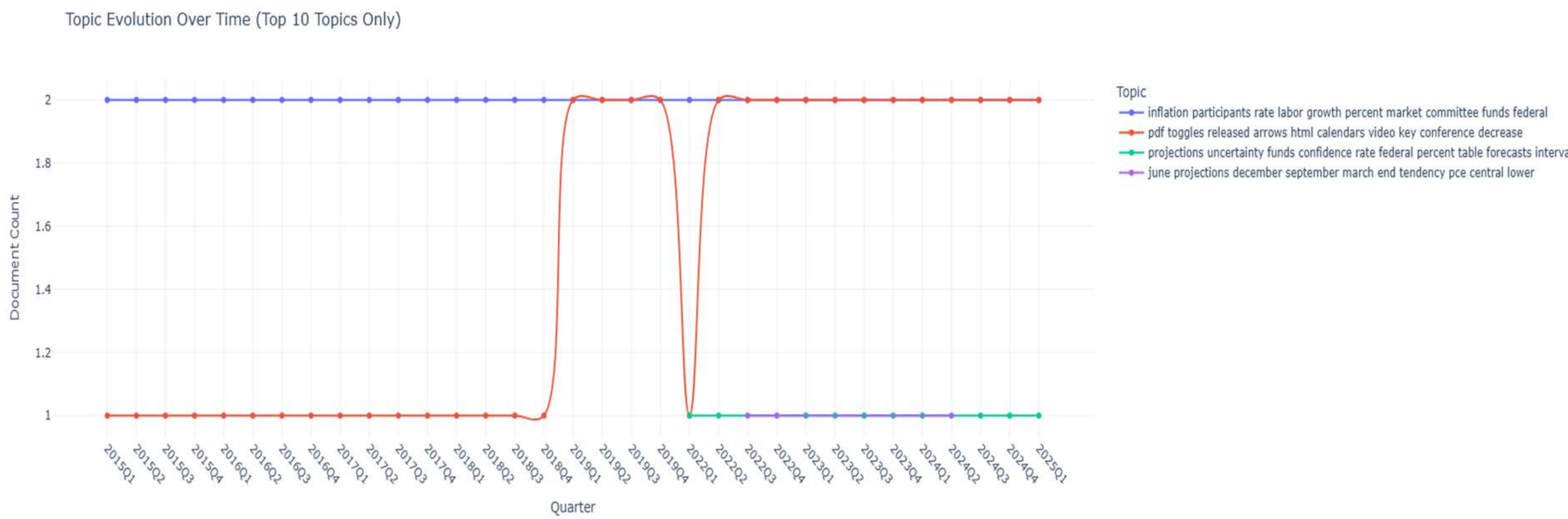
•**Coherence Measure Calculation**: A numeric score is generated to assess the quality of topic separation and relevance.

•**LDA Model Module**: Includes various model outputs and tuning steps to ensure optimal performance.

4. Visualization and Interpretation

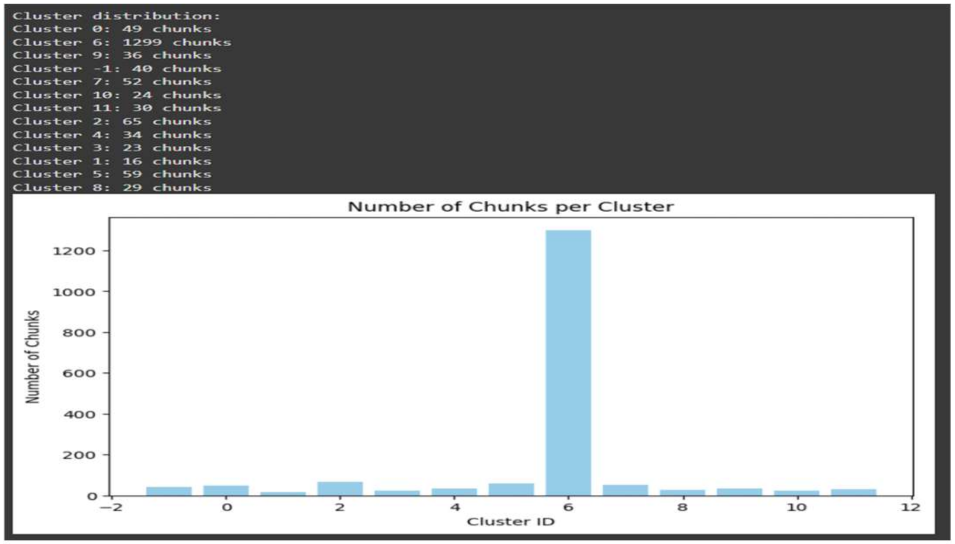
•**Matplotlib Visuals**: Charts and graphs help present topic trends, topic share over time, or top keywords.

•**Evaluation and Interpretation**: Analysts review topic quality and relevance to draw meaningful insights about theme evolution, policy focus, or sentiment trends.



Topic Evolution Over Time (Top 10 Topics Only):

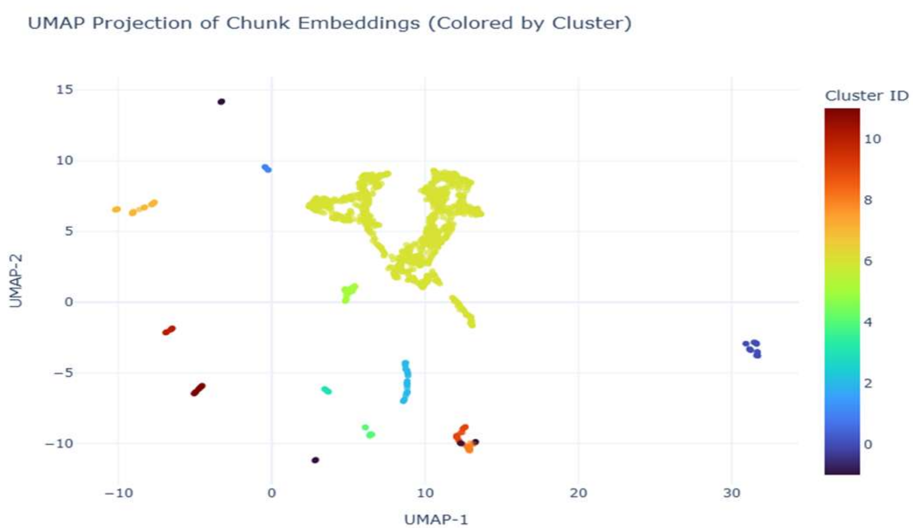
- ✓ This line chart tracks how frequently the top topics appear in FOMC meeting minutes by quarter from 2015 to 2025.
- ✓ Each line represents a topic, with the y-axis showing how many documents mention that topic prominently. This helps identify shifts in focus — such as sustained emphasis on inflation and market terms, or sudden disruptions due to data issues (e.g., missing quarters).
- ✓ The purpose of this plot is to monitor thematic trends and detect temporal topic bursts or drops, supporting interpretation of how economic communication evolves over time.



Chunk Distribution Across Clusters:

Document chunks were distributed across clusters during topic modeling or semantic clustering:

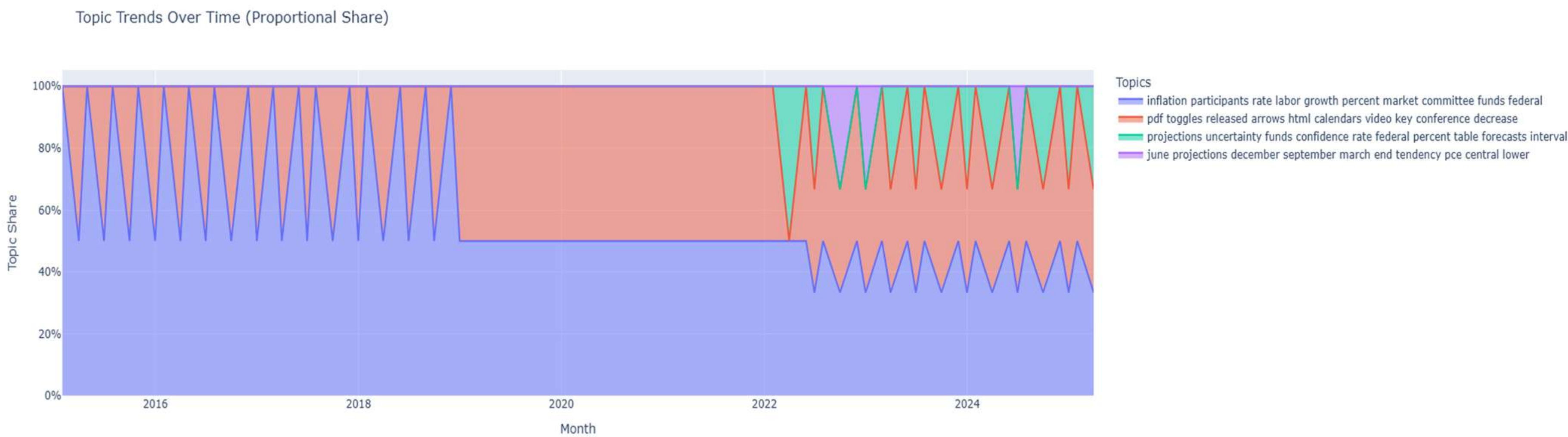
- ✓ Cluster 6 dominates with 1299 chunks, indicating one major theme or repetitive content across the corpus.
- ✓ The remaining clusters are sparsely populated, suggesting more specialized or less frequent topics.
- ✓ Cluster -1 likely represents noise or unclassified text (common in algorithms like HDBSCAN).



UMAP Clustering of Chunk Embeddings:

This 2D projection groups all chunks based on their semantic similarity using UMAP and coloring them by cluster. The aim is to:

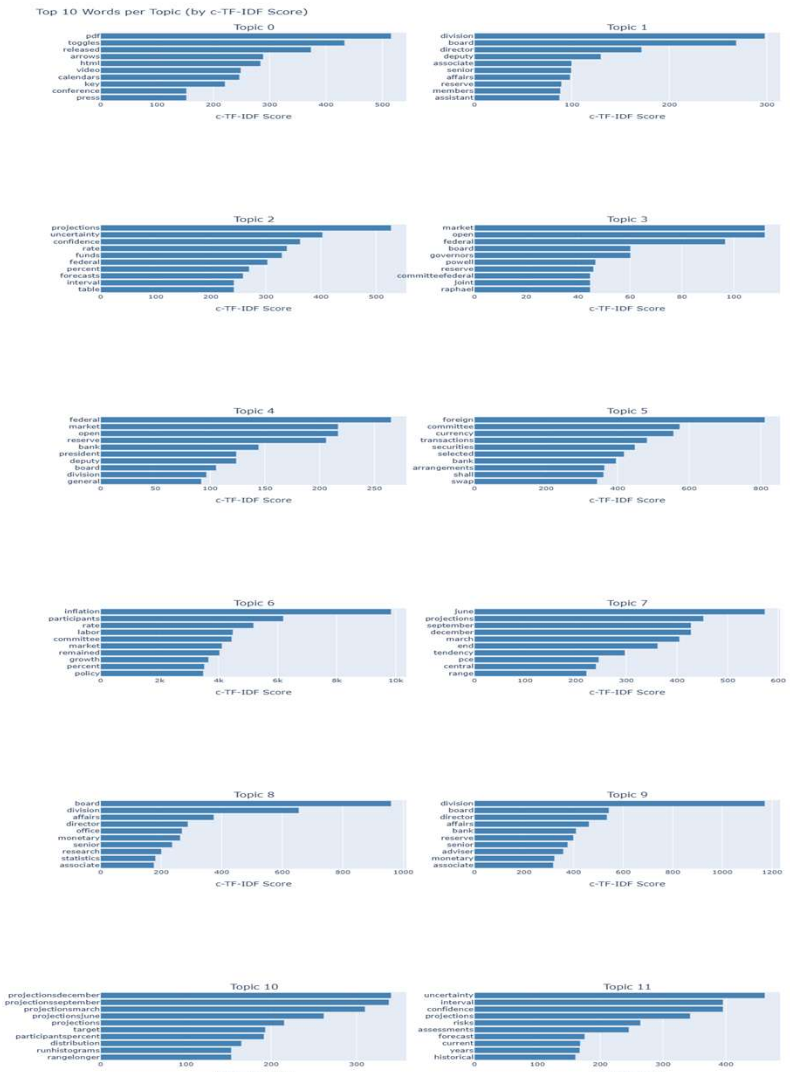
- ✓ Visualize the natural grouping of topics in the data
- ✓ Check if clusters are well-separated or overlapping
- ✓ Identify dominant or outlier themes quickly



Topic Trends Over Time (Proportional Share):

This chart helps reveal how dominant or marginal each topic has been month by month. Rather than absolute counts, it shows relative focus—highlighting how the distribution of discussion shifts over time within the same space. This gives insights into:

- ✓ Changing communication priorities of the Fed
- ✓ When new topics emerged or faded
- ✓ Structural transitions in tone or focus



This visualization highlights the top 10 words for each topic, based on c-TF-IDF scores, helping interpret the core meaning of each cluster derived from your BERTopic model.

- ✓ **Interpretability**: To see what each cluster is *actually about* by surfacing the top terms driving each topic.
- ✓ **Explainability**: c-TF-IDF highlights which words are **uniquely important** to each topic compared to the rest, not just frequent overall—making it more meaningful than plain TF-IDF.
- ✓ **Cluster Validation**: To ensure clusters aren't just technical outputs but have *semantic coherence*, which is key for reliable insights.
- ✓ **Next Steps Prep**: These keywords help in naming clusters, tracking trends over time, or linking them to external indicators.

Key Learnings & Next Steps:

This project demonstrated how NLP can be effectively used to analyze Federal Reserve meeting minutes to extract insights into monetary policy communication. Using tools like FinBERT and BERTopic, we were able to capture sentiment trends, detect shifts in tone, summarize long policy documents, and model evolving discussion topics across years of FOMC data.

Key Takeaways:

- **Sentiment Analysis:** FinBERT revealed that **neutral sentiment dominates** most Fed communications, with noticeable spikes in negative sentiment during economic uncertainty (e.g., COVID-19 period).
- **Summarization:** Extractive techniques like Maximal Marginal Relevance (MMR) enabled us to condense lengthy documents into concise summaries while preserving relevance and diversity.
- **Topic Modeling:** BERTopic uncovered recurring themes and their evolution over time, highlighting how policy focus has changed in response to economic events.
- **Clustering & Trends:** UMAP visualization and cluster distributions revealed dominant clusters in the data, validating the semantic consistency of certain policy narratives.
- **Data Structure:** Time-series analysis showed changes in document and summary lengths, offering indirect evidence of information density shifts in communication over time.

Future Improvements:

- **Topic Validation:** Incorporate expert labeling or external news alignment to better interpret topic clusters.
- **Document Quality Filtering:** Further refine preprocessing to exclude artifacts like broken HTML or script text.
- **Interactive Dashboard:** Develop a dashboard for real-time exploration of trends, sentiments, and summaries.
- **Integration with Market Events:** While economic indicators weren't included here, future work could align sentiment/topic shifts with stock market reactions or macroeconomic data.

Final Notebook's Summary & Navigation

Google Colaboratory

- ❑ Notebook1:https://colab.research.google.com/drive/1olfOac8aSDkrHcM0uK5xd2EZrYDsusSA?usp=sharing
- ❑ Notebook2:https://drive.google.com/file/d/1wElb3IYlbVUDyRdTi3hk9jMK-L6ZYDRk/view?usp=sharing
- ❑ Notebook3:https://colab.research.google.com/drive/1R3ZDrVLtzJ8vWMozuh942puBzMA8BJ4n?usp=sharing
- ❑ Notebook4:https://colab.research.google.com/drive/12a0EgEkadHbNAWZTNuP9kvjUyF9jpRPD?usp=sharing
- ❑ Notebook5:https://colab.research.google.com/drive/1JO_DbcJgEruUomSQ3NdOvQyUlsWSL1b-?usp=sharing