Bank of England (BoE) Employer’s Project
Team 6: Final Report

Bank of England

Team members:

Alex Joseph

Manpreet Minhas

Prabha Raveendran

Rafa Abdelazim

Background/Context:

The Bank of England (BoE) holds a vital role in fostering stability and confidence in financial markets through its public speeches. The BoE endeavours to scrutinize the evolution of sentiment in these speeches, seeking correlations with significant events, economic indicators, and market trends. Key inquiries involve tracking sentiment changes over time, assessing connections with events like base rate adjustments and report releases, and gauging associations with economic factors such as GDP growth and trade balance. The analysis extends to exploring the potential of speech sentiment in predicting market behaviour, influencing stock market indices, impacting market volatility, and influencing the UK exchange rate. The resulting insights aim to offer recommendations for strengthening the BoE's ability to anticipate and respond to economic and market dynamics.

Project development process:

Matplotlib, pandas, yahoo finance, and seaborn libraries were imported on python using aliases to keep code concise. Matplotlib and seaborn enabled data visualisation. Pandas and yahoo finance assisted with data analysis and obtaining stock market data respectively.

The analysis began with loading a speeches dataset in CSV format using the pd.read_csv() function, and a comprehensive data check was conducted. This involved inspecting data types, handling missing values, removing duplicates, and ensuring correct headers and footers. The date column's data type was changed to a datetime object. The 'country' column was renamed to 'location,' and a lambda function was used to group Sweden and Switzerland into the EU. An anomalous speech from 01.01.1900 was identified and corrected to 02.10.2020.

Natural Language Processing (NLP) preparations followed, involving tokenization and sentiment analysis using NLTK and TextBlob libraries. Wordclouds were generated to visualize common words, and the 'text' column underwent further processing, including lowercasing, punctuation removal, and elimination of alphanumeric characters and stopwords. The FreqDist function was used to count the most common words, and sentiment analysis was performed using a custom function with TextBlob.

Exploratory data analysis (EDA) addressed pivotal questions through line plots segmented into three distinct periods (1996-2000, 2001-2010, 2011-2023). Employing moving averages and data filtering heightened visualization, offering insights into sentiment fluctuations over time across diverse locations.

For base rate changes, Pandas facilitated data import and processing, ensuring the appropriate data type for the date column. Dual line plots, featuring 180-day moving averages, illustrated the correlation between base rate changes and sentiment scores, enhancing chart readability and elucidating their relationship.

Analysing Bank of England (BoE) speech sentiment in correlation with the Financial Stability Report (FSR) and Monetary Policy Report (MPR) necessitated manual dataset creation, datetime conversion, and line plot generation. Dotted lines marked publication dates, and a 180-day moving average facilitated trend identification.

In addressing how speech sentiment impacts UK economic indicators and its predictive power for market behaviour, we first cleansed our dataset comprising Bank of England's speeches and financial market data from yfinance. This involved standardising speech texts for sentiment analysis and aligning them with corresponding economic indicators and FTSE100 index data. For predictive modelling, a rolling 7-day average was applied to stabilise day-to-day sentiment volatility.

Visualisation strategy was tailored for clarity and impact. Time series graphs were chosen to depict sentiment trends alongside economic indicators, using a colour scheme that distinguished between different data types for intuitive comprehension. Interactive elements were incorporated for in-depth exploration of specific data points. The layout was designed to guide the viewer logically through the narrative, ensuring that the visualisations were both accessible and informative, enhancing the overall understanding of the complex relationships between speech sentiment and market dynamics.

A custom function was created to obtain the stock market data which was then sense-checked. The df.drop() function was used to remove redundant columns from the dataframe. Year-on-year returns for each index were calculated and plotted against a 30-day moving average of speech sentiment. An average true range function measured volatility, plotted against speech sentiment for each location.

Line plots were chosen for continuous variables over time, and dual axis plots ensured appropriate scaling for easy trend identification.

The relationship between sentiment scores and trade balance (GDP scores) showed a positive correlation, while the relationship with exchange rates (USD & Euro) exhibited a negative correlation. Trade balance and exchange rate predictions were made using historical data, providing a comprehensive analysis of the interplay between speech sentiment and economic indicators.

Technical overview of the code:

The entire analysis was conducted exclusively using Python, leveraging the team's familiarity with the language.

Data Ingestion, Cleaning, and Transformation: The code proficiently manages data ingestion from diverse sources, including CSV files, Excel files, and APIs. Subsequently, the acquired data undergoes loading into Pandas DataFrames for meticulous processing, including the handling of missing values and the encoding of categorical variables.

Choice of Analysis Tools and Techniques: In the realm of exploratory data analysis, a suite of Python libraries tailored to specific tasks was employed. Matplotlib and Seaborn were specifically chosen for their prowess in data visualisation, while Pandas and Yahoo Finance played instrumental roles in data analysis and retrieving stock market data, respectively.

Python Packages: The selection of Matplotlib and Seaborn was motivated by their robust visualization capabilities, enabling the creation of diverse plots such as line plots and dual-axis plots. These visuals proved invaluable in presenting trends and relationships within the data. For text processing, the Natural Language Toolkit (NLTK) and scikit-learn were enlisted to implement machine learning models.

The project embraced a spectrum of machine learning algorithms encompassing linear regression, time series analysis, and Natural Language Processing (NLP), tailored to the specific problem at hand. As part of the data processing pipeline, word clouds were generated. Tokenization dismantles text into individual words, and subsequent vectorization transforms these words into numerical representations conducive to machine learning models.
To probe the correlation between speech sentiment and UK economic indicators and its potential to predict market behaviour, sentiment analysis was executed utilizing Natural Language Processing (NLP) libraries. These libraries facilitated the efficient parsing and interpretation of textual data extracted from economic speeches.

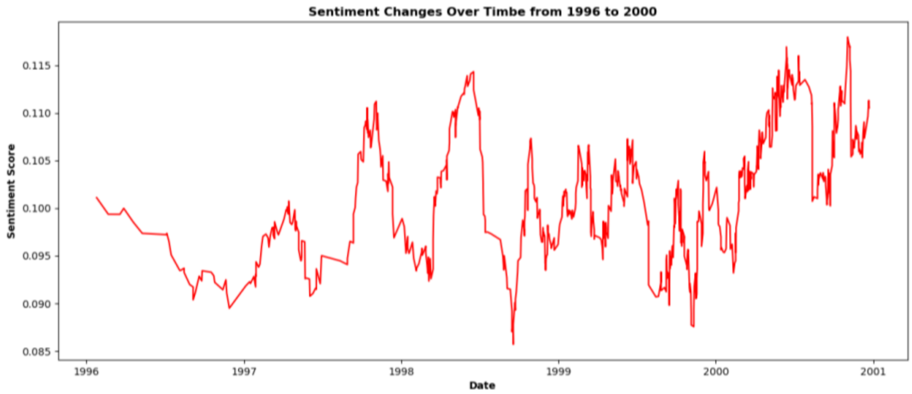
The **XGBoost** algorithm was strategically selected for its proficiency in handling intricate, non-linear data relationships, a crucial aspect for interpreting the nuanced interactions between sentiment and, more significantly, financial data. In the context of time-series analysis, pandas were leveraged to ensure precise alignment of speech sentiment with corresponding economic data points. Notably, the technical decision to employ XGBoost significantly bolstered the model's predictive accuracy but simultaneously introduced complexities in terms of tuning and interpretation. Grasping these intricacies was pivotal in extracting meaningful insights regarding the impact of economic speech sentiment on market trends.

Calculation of correlation coefficients through **Pearson correlation** was undertaken to assess the relationship between sentiment scores and trade balance (GDP scores), as well as exchange rates (USD & Euro). The prediction analysis unfolded through **regression analysis**, involving the creation of test and train data, complemented by the application of exponential smoothing techniques. This multifaceted approach facilitated a thorough exploration of the intricate dynamics linking sentiment, economic indicators, and market trends.

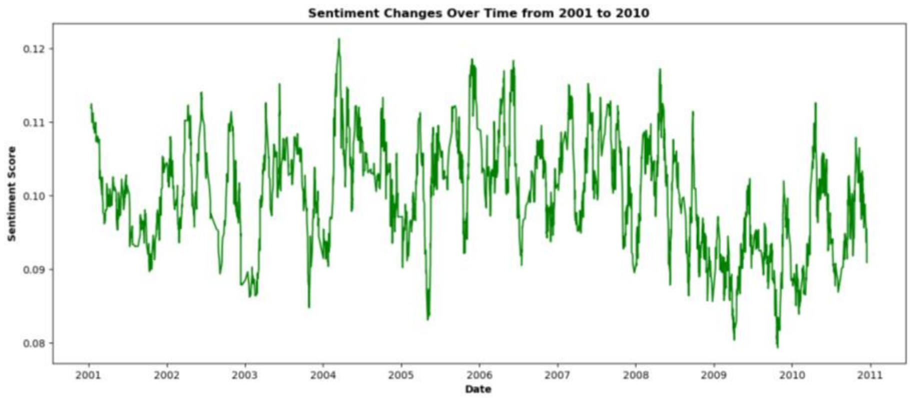
Patterns, trends, and insights

Sentiment changes over time:

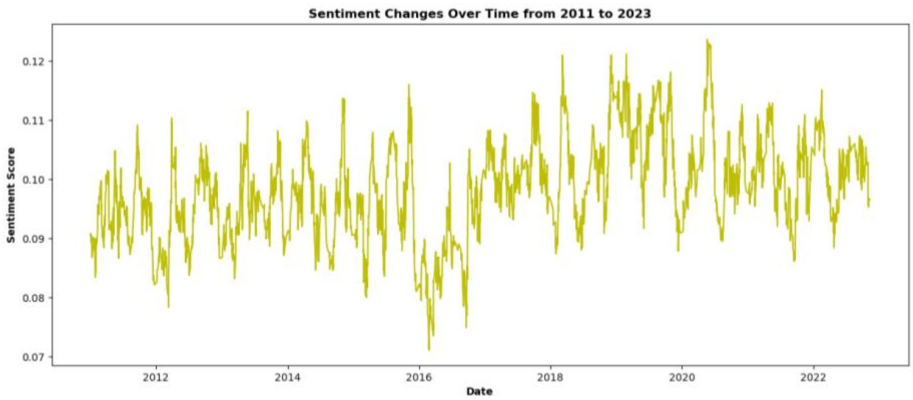
Plot 1



Plot 2



Plot 3



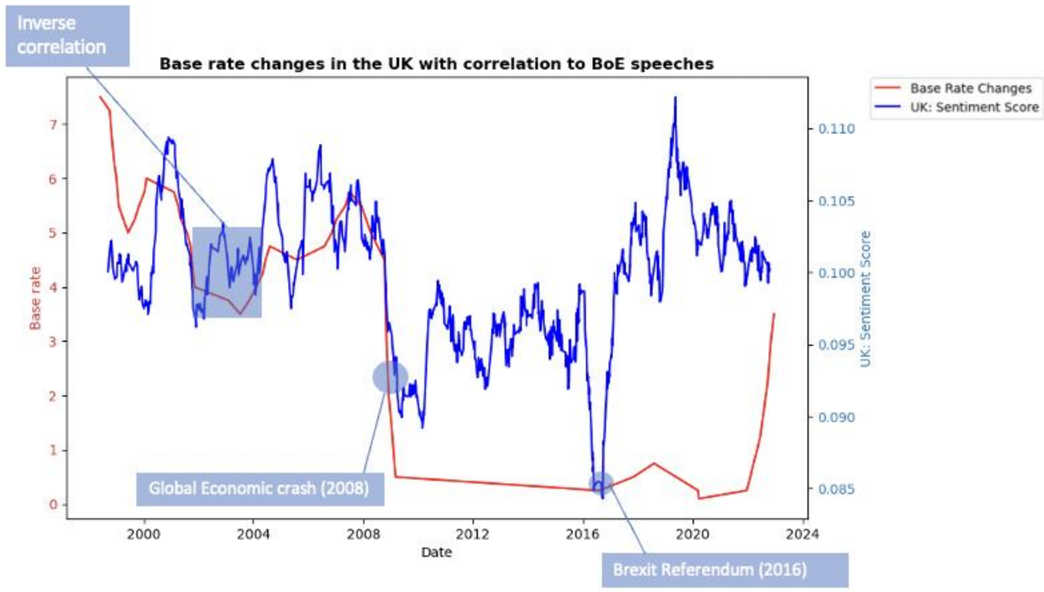
Sentiment ranged from 0.07 to 0.16, correlating with major economic events. In 2016, a low sentiment of 0.07 coincided with the Brexit referendum announcement. A word cloud from UK speeches in 2016 reflected terms like "risk," "change," and "uncertainty," aligning with the event. Another peak in 2020 may be attributed to reassuring language following the COVID-19 lockdown.

Plot 4

Correlation of BoE speech sentiment with key events:

- *Correlation with bank rate changes:*

Plot 5



After the global financial crash in 2008, sentiment continued to fluctuate, while base rate changes remained almost stable for nearly 8 years (*Plot 5*). The recovery of the base rate (2016) was interrupted by the COVID-19 pandemic. In 2022, the BRC surged for the first time in a decade to address inflation.

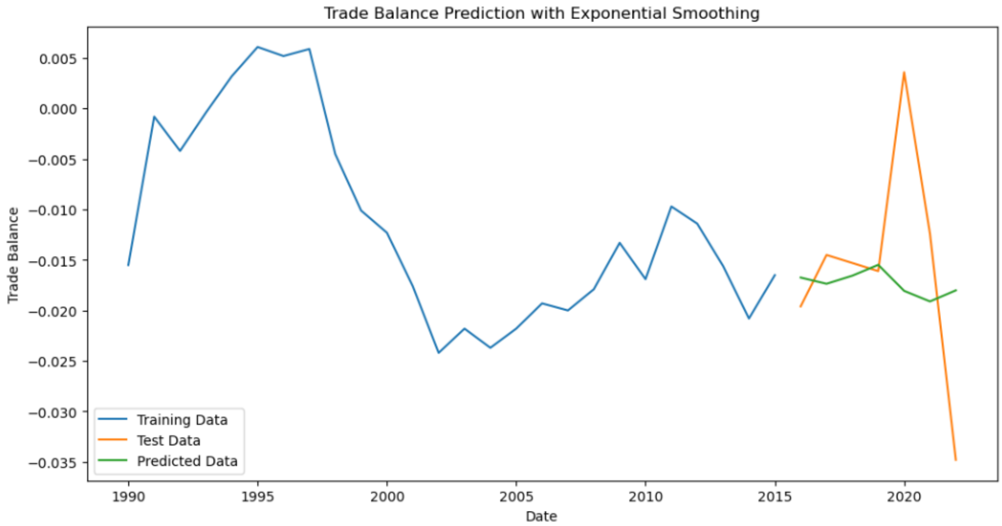
- *Correlation with MPR and FSR:*

Positive sentiment surges during Monetary Policy Report releases, signifying reassurance for the economic outlook. Conversely, Financial Stability Reports correspond to both high and low peaks in speech sentiment, indicating a proactive BoE approach to stabilizing sentiment. These reports can be strategically employed during periods of optimism and concern.

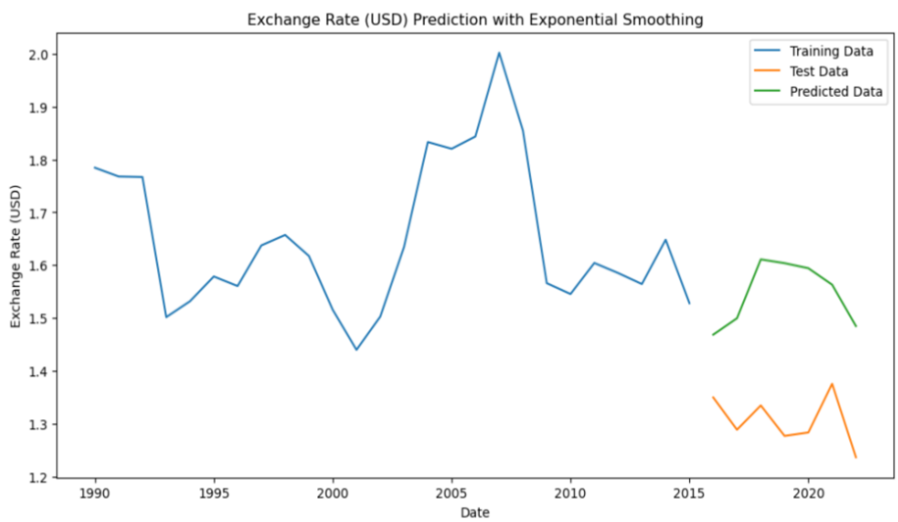
Correlation of BoE speech sentiment with key economic indicators:

In our analysis, we found a moderate positive correlation between speech sentiment and the FTSE100 index, particularly during significant events like the 2008 financial crisis and the Covid crash. Conversely, an inverse correlation emerged with unemployment, indicating negative sentiment preceding rising unemployment rates. On a finer timescale, sentiment seemed to precede trends in GDP and unemployment, hinting at its predictive potential. Our XGBoost model, scoring 614.64, outperformed the baseline. To enhance precision, we recommend exploring advanced deep learning models like Long Short-Term Memory (LSTM) networks for a nuanced understanding of sentiment's impact on economic patterns.

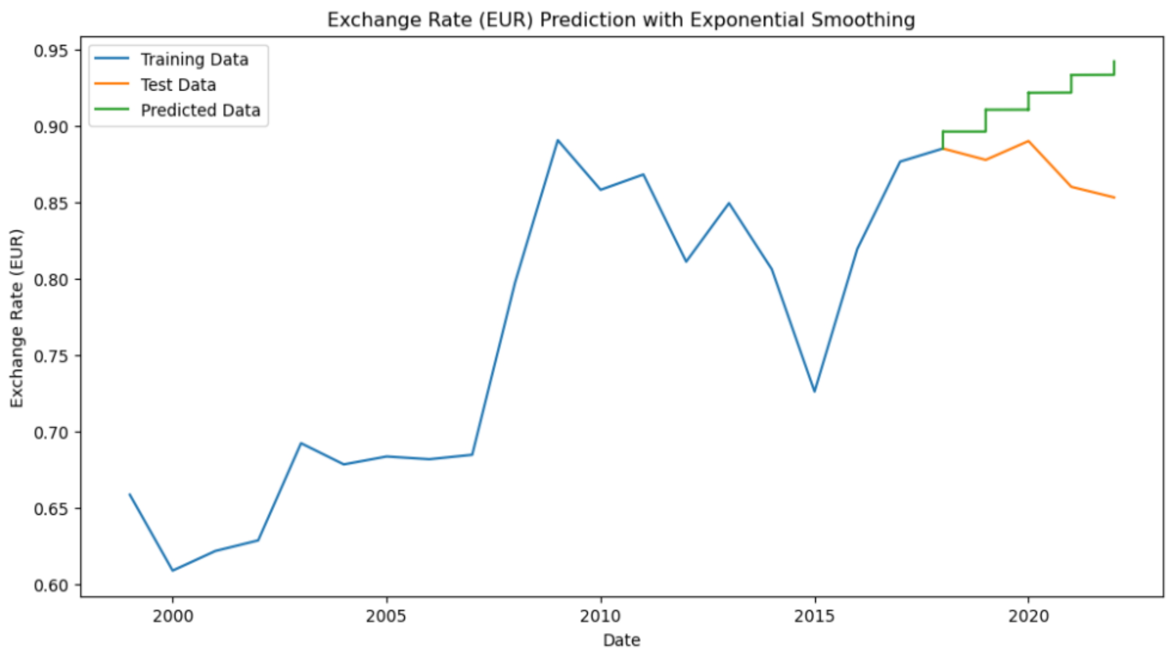
If sentiment can assist in predicting market behaviour:



Plot 6: The main hypothesis posited a positive correlation indicating public sentiment's impact on economic performance. The trade balance prediction analysis indicates a steady continuation of data without significant fluctuation.



Plot 7: The prediction analysis for exchange rate (USD) suggests that the data will continue in the same pace.



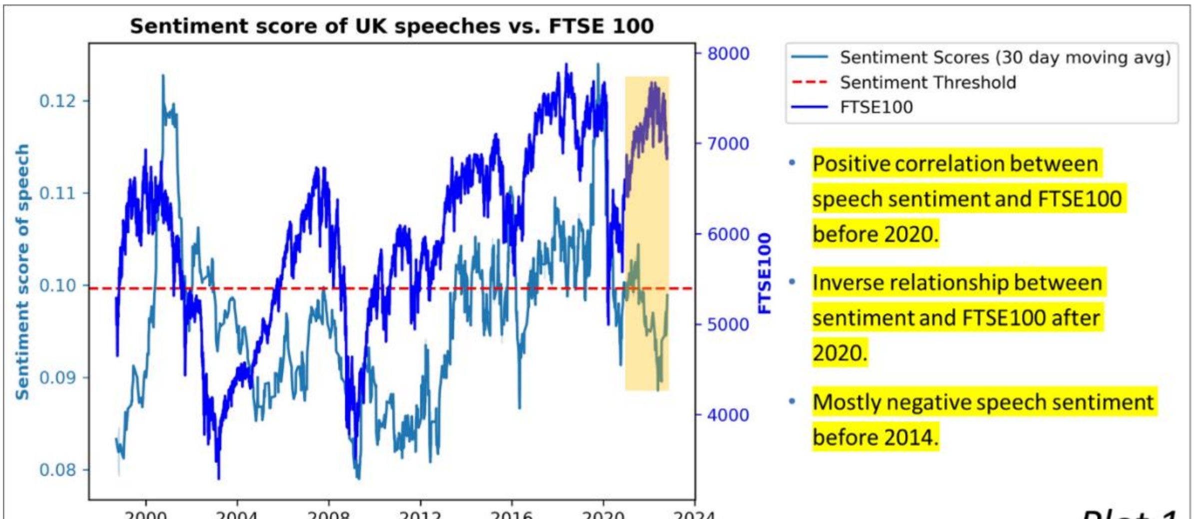
Plot 8: The prediction analysis for exchange rate (Euro) suggests that the data will continue in the upward trend.

Further analysis:

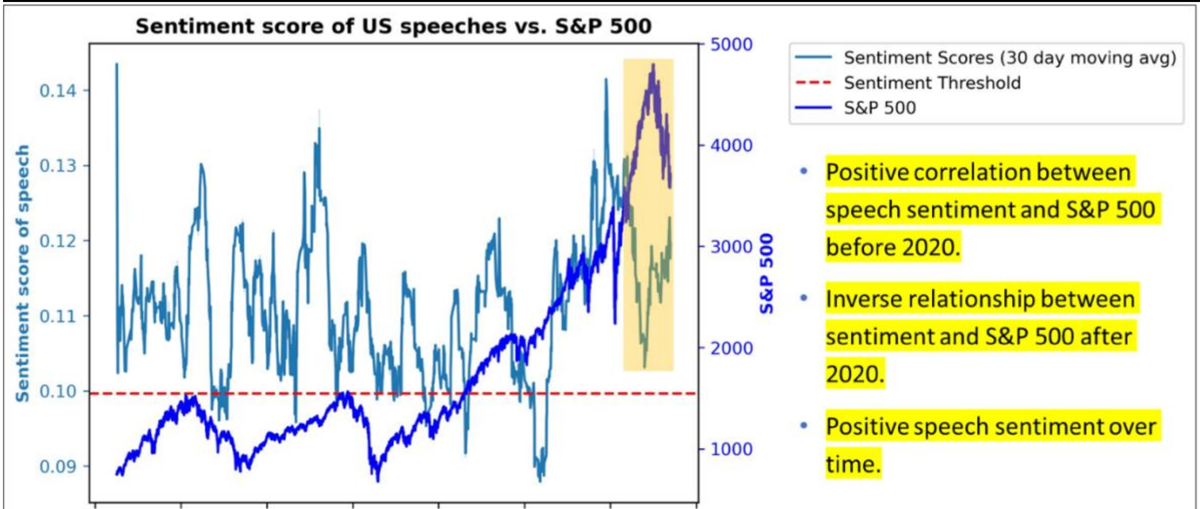
Since 2014, the UK government speakers have a more positive sentiment than non-government speakers. A recommendation is to increase the proportion of UK government speakers to non-government speakers.

Refer to the appendix for annotated plots depicting sentiment versus stock market indices, YoY returns, and sentiment versus volatility. It is recommended to experiment with delivering positive sentiment speeches during periods of volatility, assessing effectiveness by monitoring a substantial reduction in volatility post-speech. Consider a trial involving a gradual increase in speech sentiment over time to gauge its impact on the FTSE100's magnitude and direction. Additionally, explore potential trends between speech sentiment in other countries and the FTSE to enhance the Bank of England's ability to anticipate market movements.

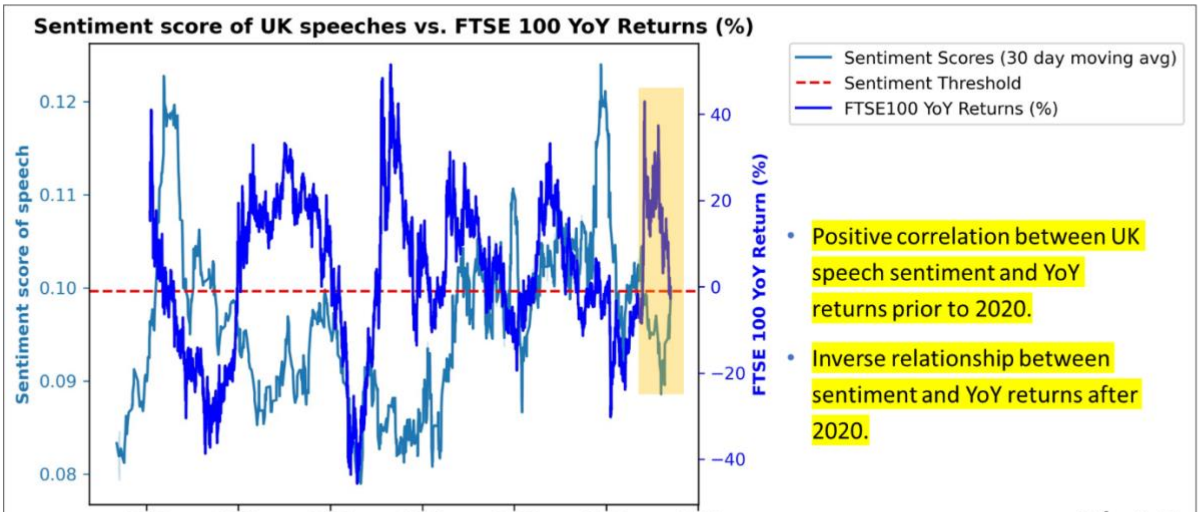
APPENDIX



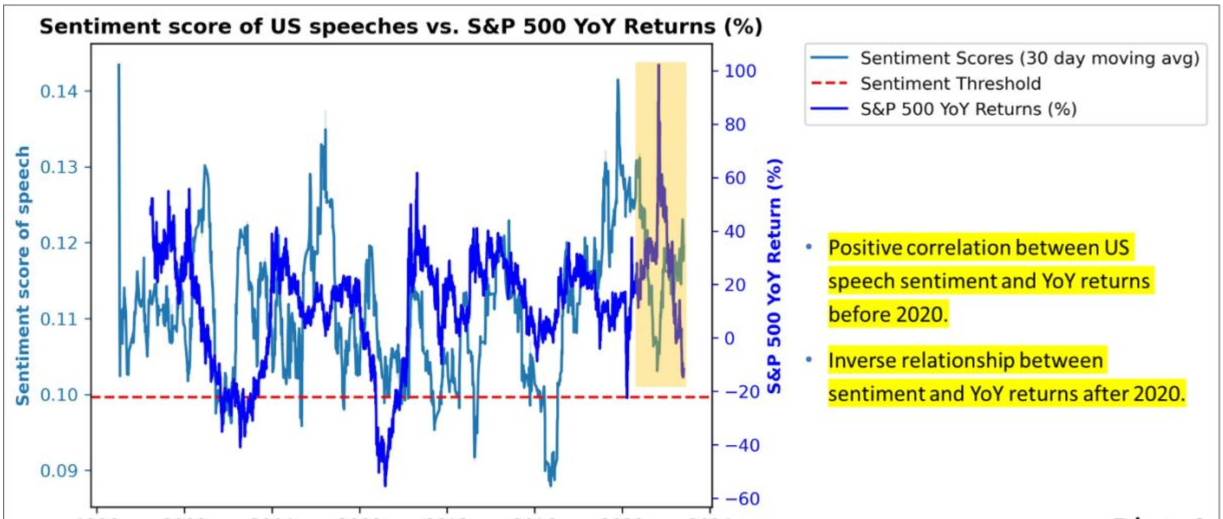
Plot 1



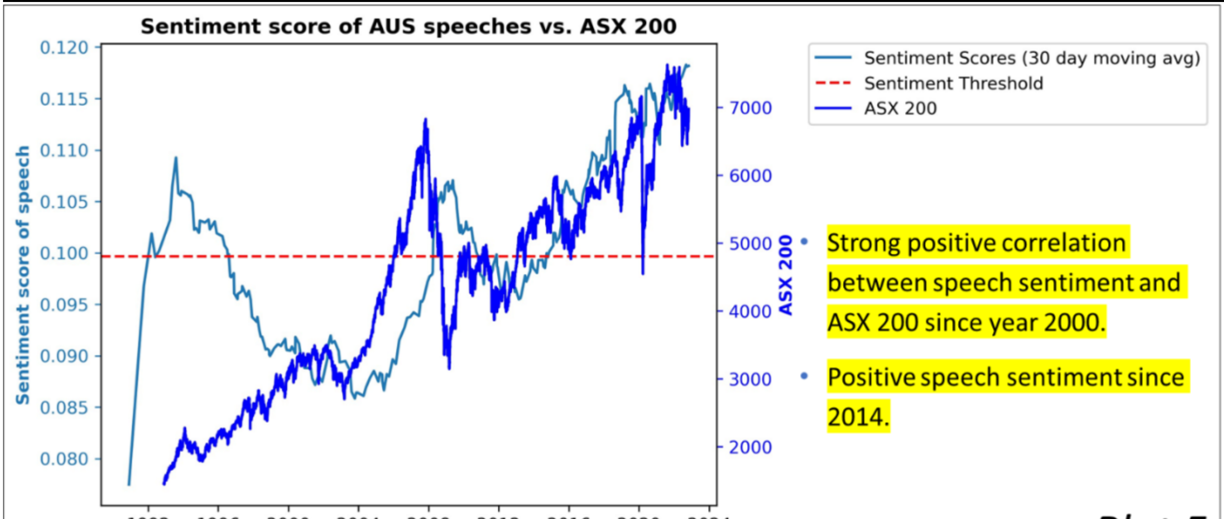
Plot 3



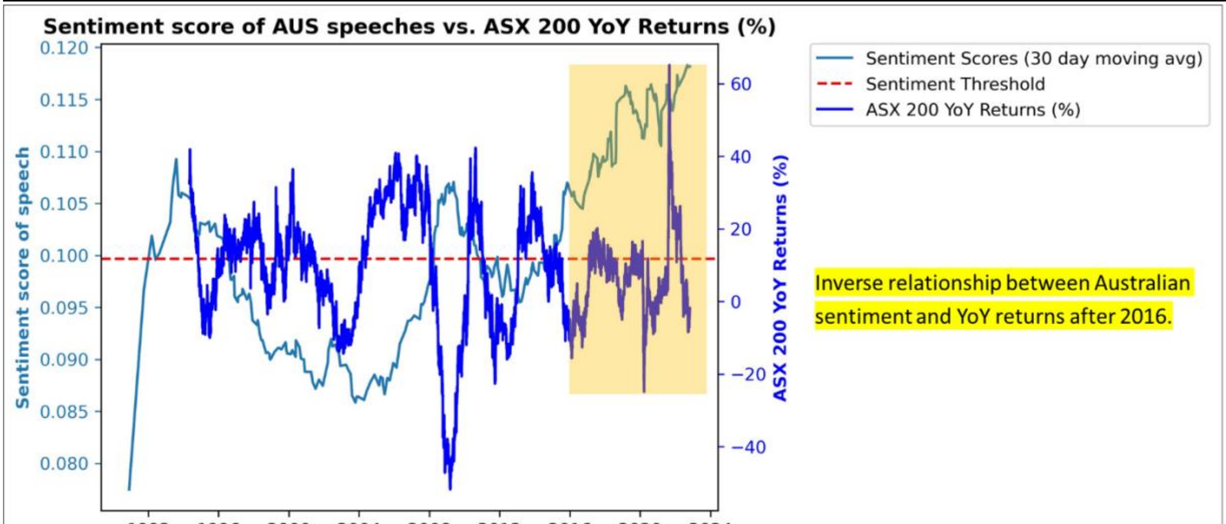
Plot 2



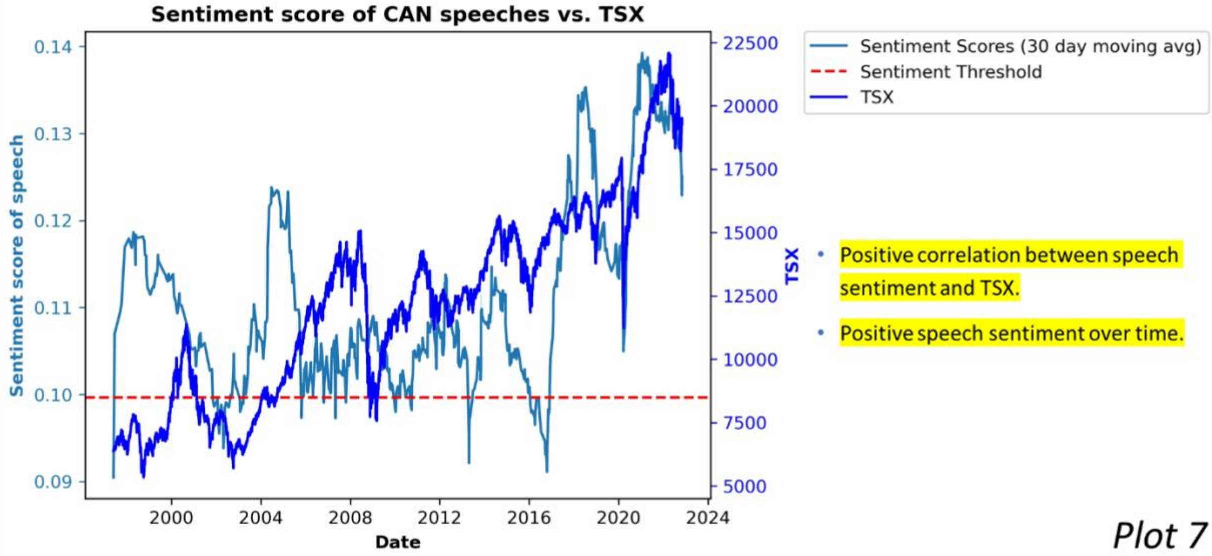
Plot 4



Plot 5

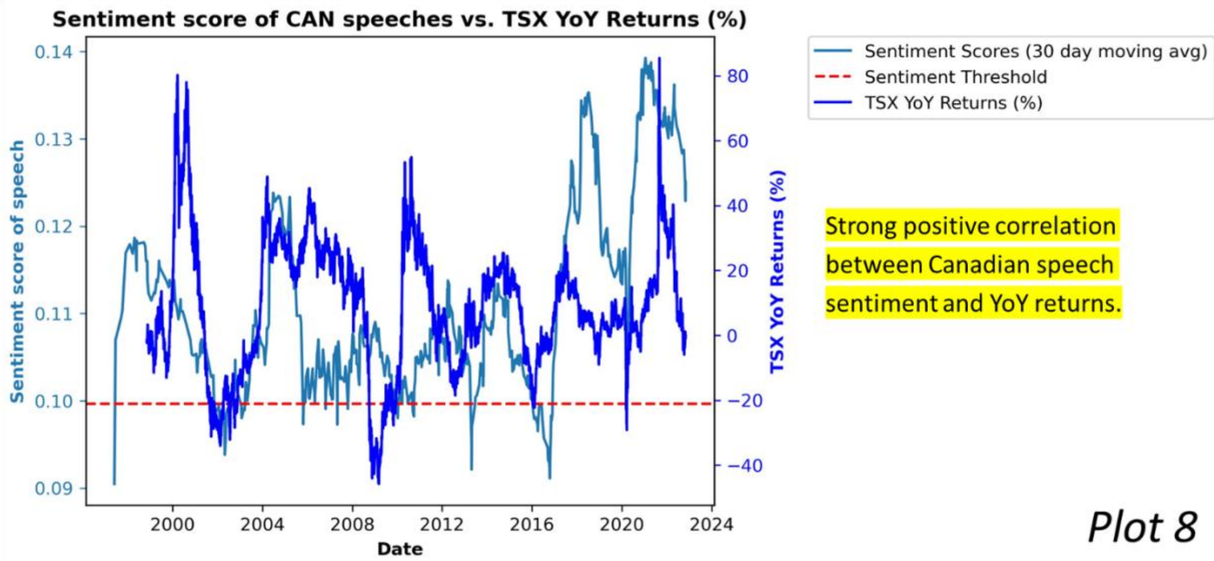


Plot 6



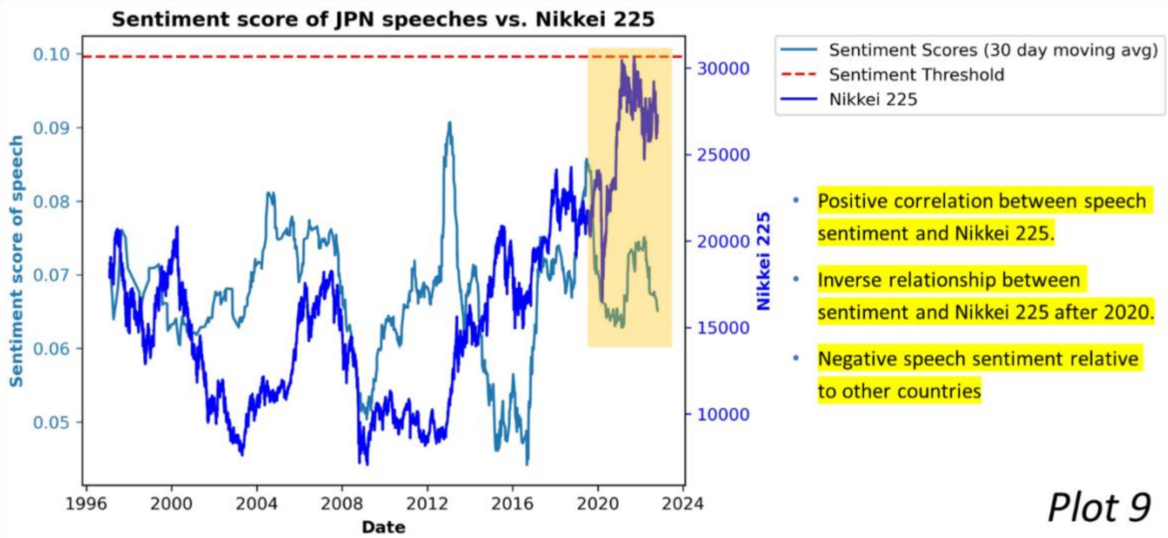
- Positive correlation between speech sentiment and TSX.
- Positive speech sentiment over time.

Plot 7



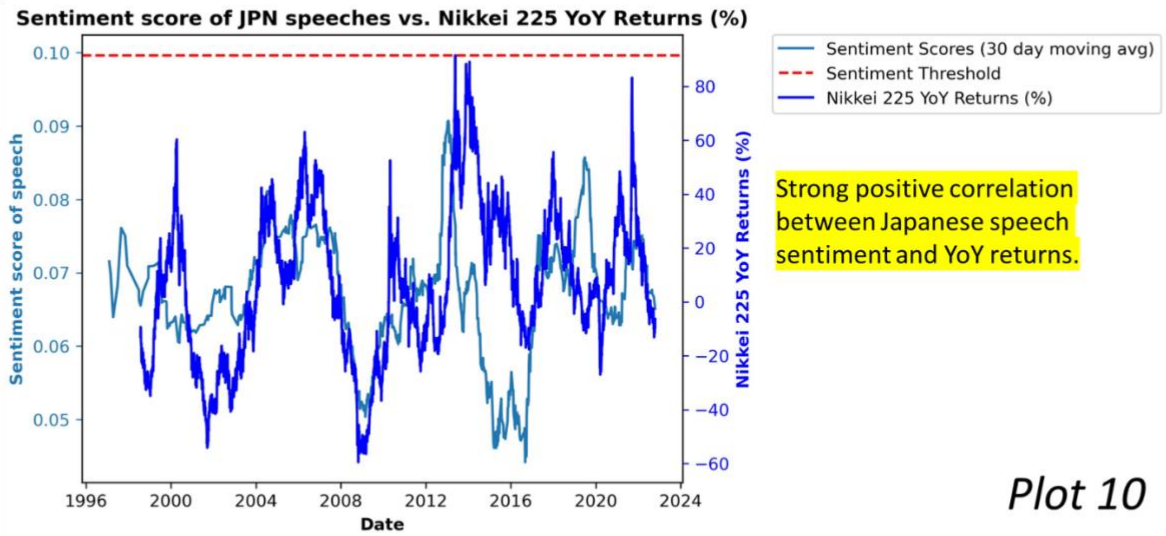
Strong positive correlation between Canadian speech sentiment and YoY returns.

Plot 8



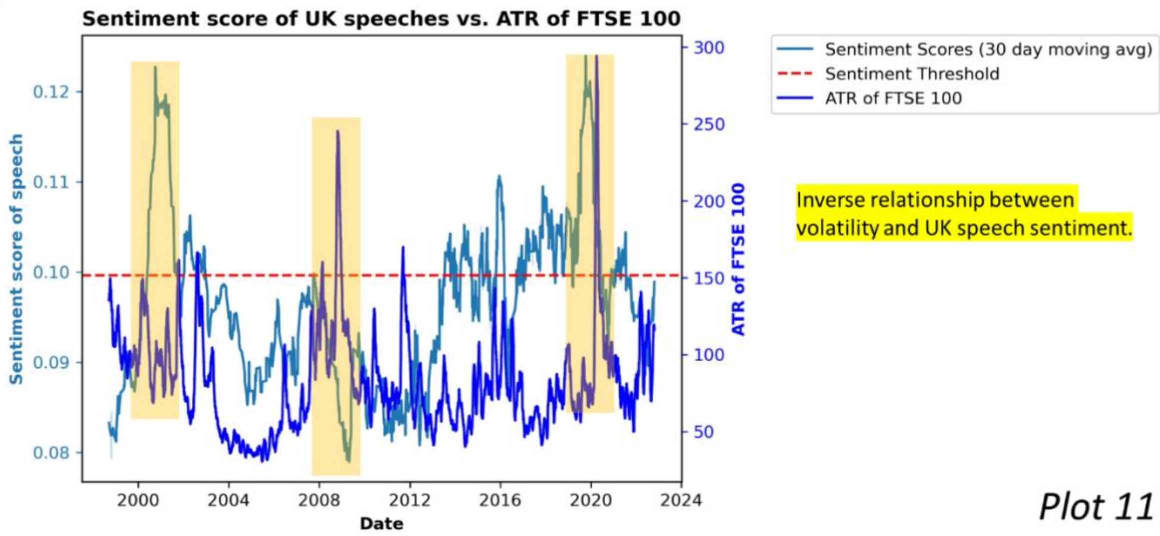
- Positive correlation between speech sentiment and Nikkei 225.
- Inverse relationship between sentiment and Nikkei 225 after 2020.
- Negative speech sentiment relative to other countries

Plot 9



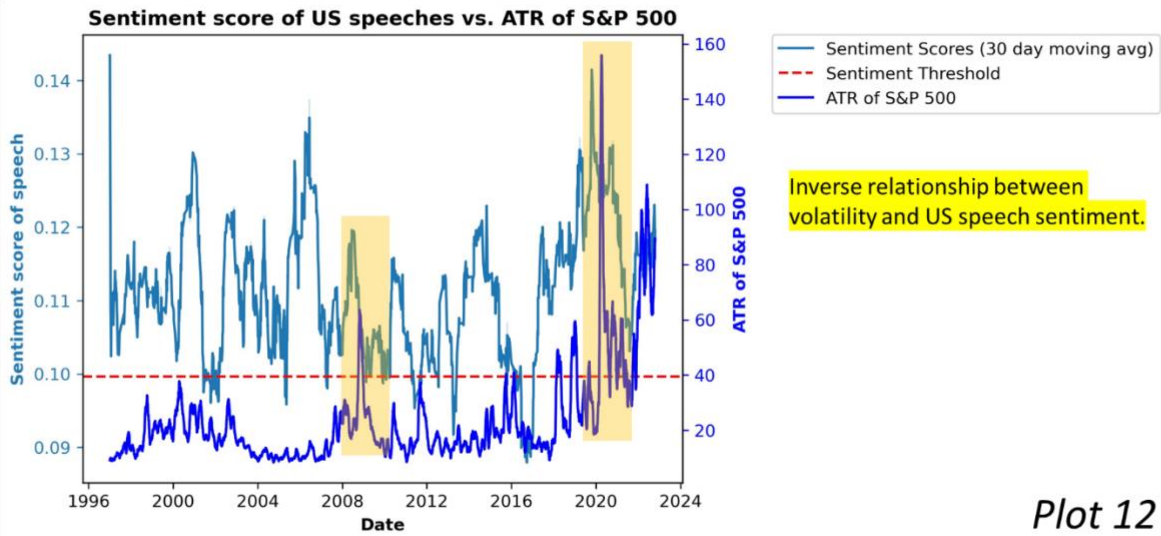
Strong positive correlation between Japanese speech sentiment and YoY returns.

Plot 10



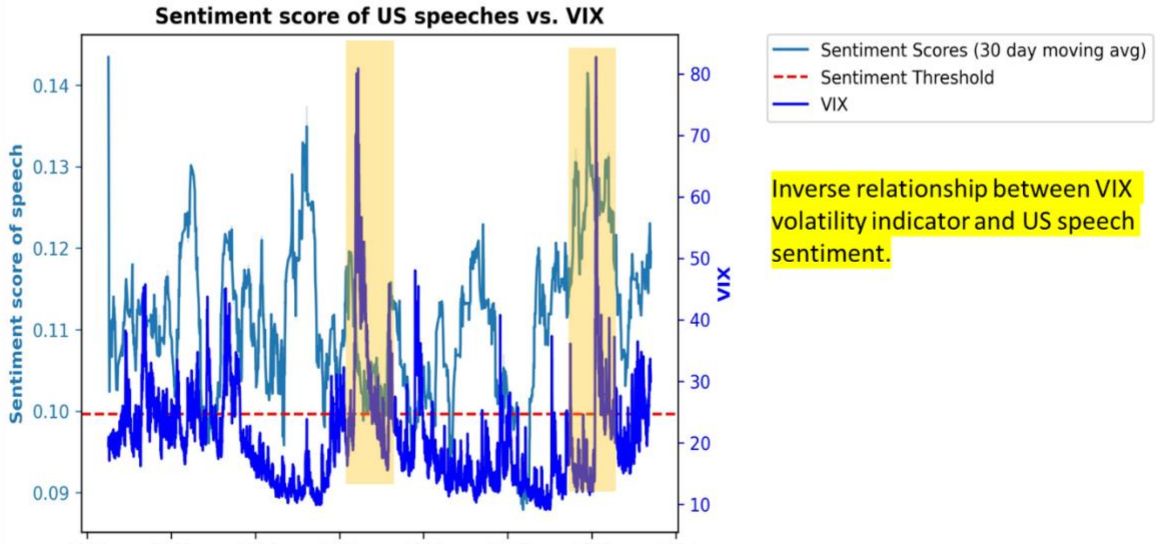
Inverse relationship between volatility and UK speech sentiment.

Plot 11

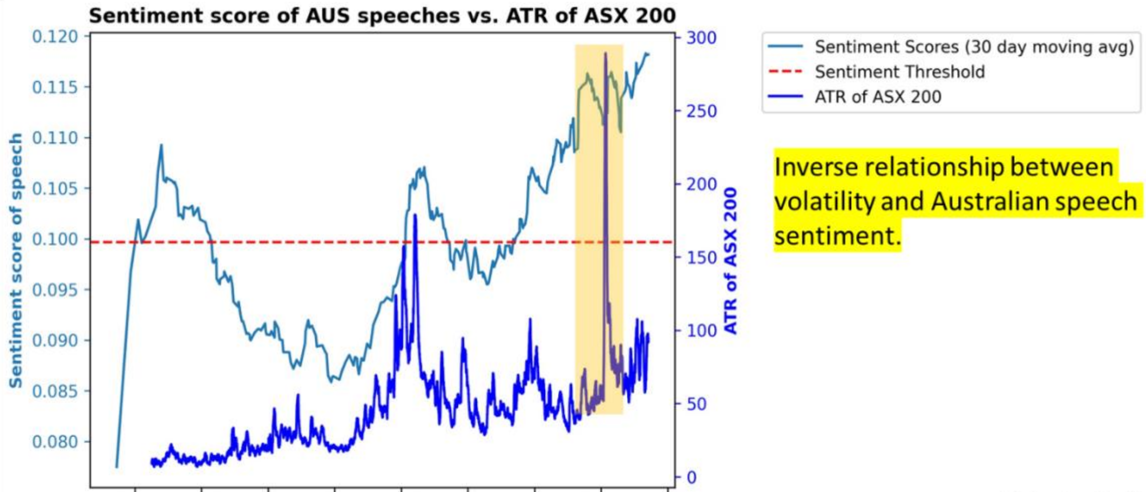


Inverse relationship between volatility and US speech sentiment.

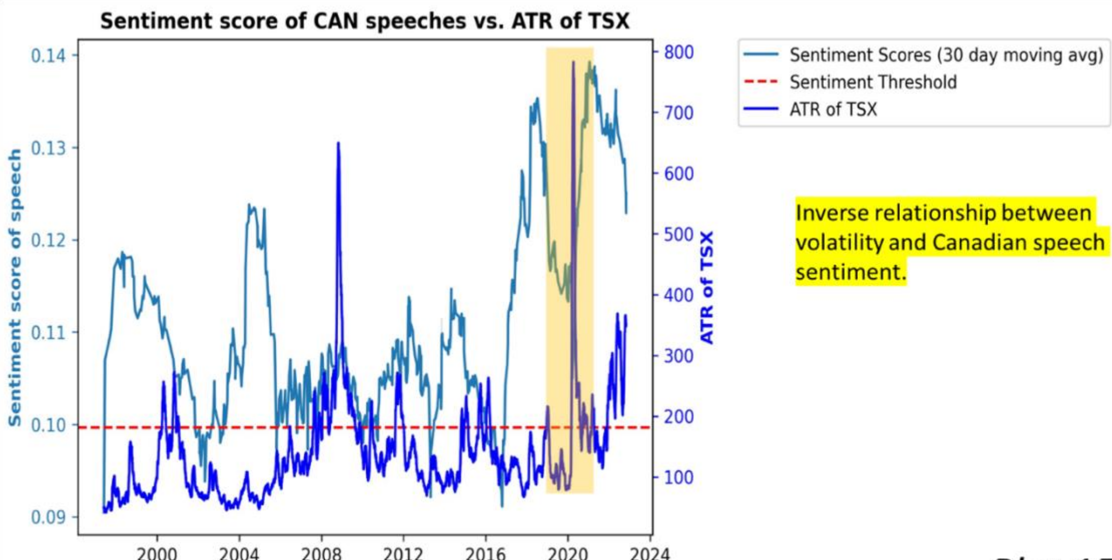
Plot 12



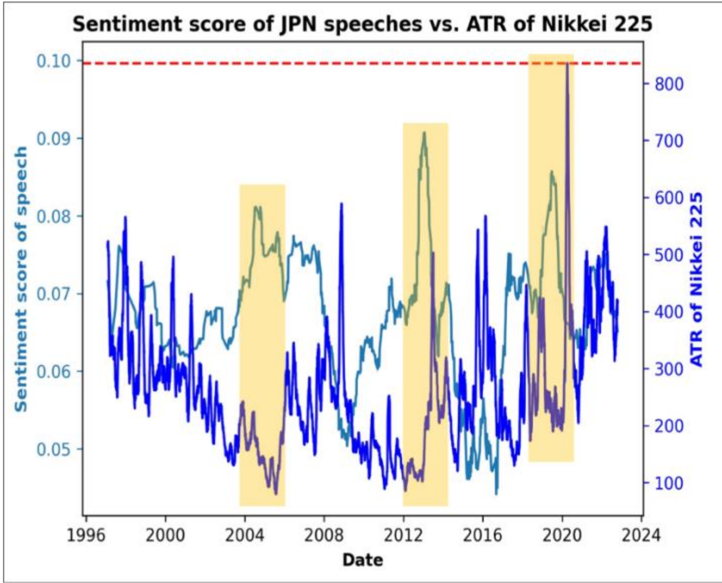
Plot 13



Plot 14

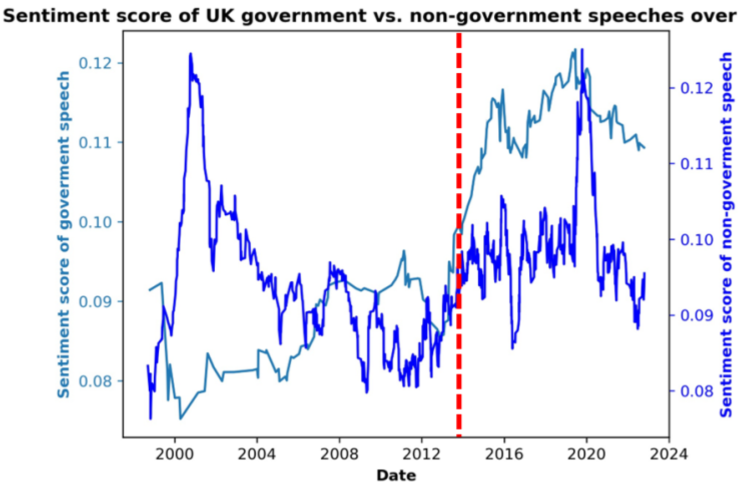


Plot 15



Inverse relationship between volatility and Japanese speech sentiment.

Plot 16



Since ~2014, UK government speech sentiment has almost always been greater than the UK non-government speech sentiment.

Plot 17