LSE_DA401_Employer_Project_Assignment_1

Employer Project Team 5

Bank of England (BOE)
Sentiment Analysis of Central Bank Speeches

Assignment 1: Project Plan

Ryan Lin, Bassam Alabsi, Gianni Xue, James Bhardwaj, Chizoba Itodo, Sean Meade

Word Count: 1128

Assignment 1: Project Scope and Plan

1. Introduction and Problem Statement

Context: In this fast-evolving contemporary digital world, understanding how to effectively tackle data is paramount, particularly for financial institutions like the Bank of England. Their decisions, backed by data driven insights, influences not just the UK's monetary and financial stability, but also the global economic system. One major tool the Bank uses to conveys its stance and views is through speeches delivered by its representatives at various public events.

The Problem: The evolving sentiment within these speeches can provide valuable insights into the bank's current and future strategies. However, a plethora of questions remain unanswered:

- Has the sentiment of central bank speeches changed over time, and if so, in what ways?
- Are there discernible trends in the sentiment expressed in these speeches?
- How does the sentiment correlate with major economic indicators like the UK's GDP growth, inflation, unemployment, and wages?
- Furthermore, does the sentiment expressed in these speeches have the power to predict market behaviours?

Significance: Understanding these sentiments can play a pivotal role in predicting market trends and aligning the bank's communication to stabilise financial markets. Moreover, as the Bank of England aims to steer the economy through its policies, its communication becomes a key instrument in achieving those policy objectives. Prior Natural Language Processing (NLP) and Machine Learning techniques (Extreme Gradient Boosting model) on central bank speeches[1] and Large Language Models (LLMs) using GPT-3.5 on Bank of England Monetary Policy Committee (MPC) member speeches have demonstrated predictive abilities for financial markets response. [2]

Objectives: Ultimately, the goal is to refine the Bank of England's communication strategies by leveraging sentiment analysis. This will help in making more informed data-driven decisions, understanding the key impacts, gauging financial indicators, and tailoring communications to maintain stable markets.

---

[1] [Can central bank speeches predict financial market turbulence?](#)

[2] [Can ChatGPT Predict Future Interest Rate Decisions?](#)

2. Team Roles and Ways of Working

Our team comprises six diverse individuals, each contributing a unique set of skills and experiences. While we will assign roles based on inherent strengths, we also encourage members to take on responsibilities they feel confident about:

Ryan: Designated as our team leader, Ryan is receptive to others' ideas and excels in analysis. He will interpret data to derive the relevant conclusions.

Gianni: Gianni's specialised knowledge in banking and finance are invaluable to the group, given the project's context. He will provide deeper insights, collect, and assist in data analysis.

Chi: Chi offers excellent communication skills, an accounting background, and a keen attention to detail. These strengths will be pivotal in data cleaning and in providing insights into trends and analysis.

James: James' knowledge of and interest in government and history will infuse context into the speeches. He will facilitate a nuanced analysis incorporating historical trends.

Sean: Sean's role as a data analyst, combined with his background in accounting and finance, ensures he provides invaluable insights into data, trends, and analysis. His proficiency in Python adds another layer of expertise to our team.

Bassam: Bassam's skills in data cleaning and wrangling are crucial to our team. Additionally, his experience in the capital markets enhances our ability for comprehensive trend analysis.

External Stakeholders: We will work closely with our success manager and course facilitator, meeting weekly to review our progress and ensure we are on the right track with our analysis. We also hope to utilise any opportunity to speak with available stakeholders at the BOE to clarify information and ask questions we may have.

Our primary mode of communication is WhatsApp, ensuring real-time interaction and updates. We've instituted a weekly Google Meet session to discuss progress and address challenges, with provisions for ad-hoc meetings when urgent matters arise. Contingency plans are in place for seamless workflow continuity, ensuring we maintain momentum even during unforeseen challenges.

3. Project plan:
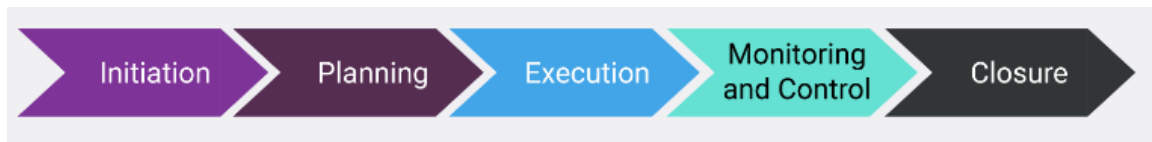
1. Initiation:

All team members conducted a preliminary feasibility study to gauge the project's viability and likelihood of successful completion. Our aim was to ensure all stakeholders understood the project's potential risks and rewards.

Initially, Rule-based Sentiment Analysis, supported by BOE's lexicon list, was considered. However, its limitations—such as not considering entire sentences and requiring frequent updates—made us lean towards Automated or ML sentiment Analysis (or a hybrid approach). Where the success of this approach depends on the quality of training data set and chosen algorithm.

2. Research Methods:

We identified the methods we'd employ and the required data. This also involved preparing the necessary research tools.

Additional datasets from the Office for National Statistics (ONS) will be downloaded to be used for the analysis.

After deciding to use Python, we will import the dataset, perform exploratory data analysis, determine if data wrangling is necessary, and create basic visualisations.

A roadmap detailing task assignments (see Appendix) and work distribution will guide our analytical approach. We're employing Monday.com to ensure the key milestones in the project are tracked and work is equally distributed amongst team members.

3. Planning:

Data importation, cleaning, and sense-checking are prioritised.

A basic outline of our approach can be seen below:

1. **Feature extraction:** tokenisation, lemmatisation, stop word removal and vectorisation. It's essential that negated words are correctly accounted for. We'll calculate a sentiment score for each text, presented as a percentage.
    a. **Using Contextual Semantic** Search to query the data, filter all contextual similar sections that relate to a given concept.
    b. **Latent Dirichet Allocation** to identify topics discussed in large corpus of speech data, to uncover any patterns, themes and relationships.
2. **Training & Prediction:** We will examine the differences between the produced sentiment-labelled training set and the lexicon list provided to us by BOE.
3. **Predictions:** Feed new text into the model, the model predicts labels for unseen data. See how this compares to the provided lexicon list.
4. **Investigate relationships:** Investigate relationships to the questions posed by BOE:
    a. How sentiment has change over time
    b. How sentiment correlates with other events
    c. How sentiment correlates with economic indicators (GDP, Inflation, labour statistics)

4. Execution:

We're considering analysing specific segments—potentially January or June of each year or perhaps on a quarterly basis. This ensures consistent time segment analysis.

Considering the likelihood of scope creep, only If time and resources allow, we will explore deep learning using an existing pre-trained transformer model. This would enable the most accurate and efficient sentiment analysis.

5. Monitoring and Control

We'll consistently refer back to our project scope and objectives, ensuring we're not expanding beyond our initial plan.

In case of unforeseen delays, we'll reassess our priorities and redistribute tasks as needed. Using the roadmap/ Gant chart, each team member will be set goals/ tasks with corresponding deadlines that everyone can keep track of.

6. Closure

Stakeholders have received final deliverables and are happy with the project outcome.

4. Analytical approach

Collaboration & Tools:

GitHub will be our primary collaboration platform for our technical analysis, where appropriate using push and pull requests. When working on shared documents we will utilise google docs and slides. We're contemplating using a live shared Jupyter Notebook using Jupyter lab-collaborative or Visual Studio Code Live Share for real-time collaboration.

Our Analytical Approach:

Week	Approach Area	Description	Key Milestone
2	**Foundation on GitHub**	Set up our collaborative space for version control and contribution integration.	A ready-to-use GitHub repository with initial data imported.
2	**Data Manipulation & Analysis**	Utilise Python in Jupyter Notebook for data handling. Our main goal is sentiment extraction using NLP tools like NLTK and TextBlob.	dataset with extracted sentiments.
2	**Exploratory Data Analysis**	Dive deep into the speeches to discern their structure and possible challenges. Tools: Python, Anaconda, Tableau, Jupyter Notebook.	preliminary visualisations and identified data challenges.
2	**Brainstorming Session**	Team alignment on the data, project phases, and timelines.	a defined project roadmap.
2/3	**Phase One - Data Cleaning**	Remove duplicates, transform texts to lowercase, and standardise formatting.	cleaned dataset ready for analysis
3/4	**Phase Two - Text Pre-processing**	Refine speeches, tokenise, and find correlations between speeches and market trends.	a processed dataset with correlations identified.
5	**Phase Three - Presentation**	Use Tableau to visualise sentiments over time.	a comprehensive presentation with visual insights.

5. Appendix: Project Roadmap (See excel file for full details, we are only able to share the roadmap on Monday.com through email)
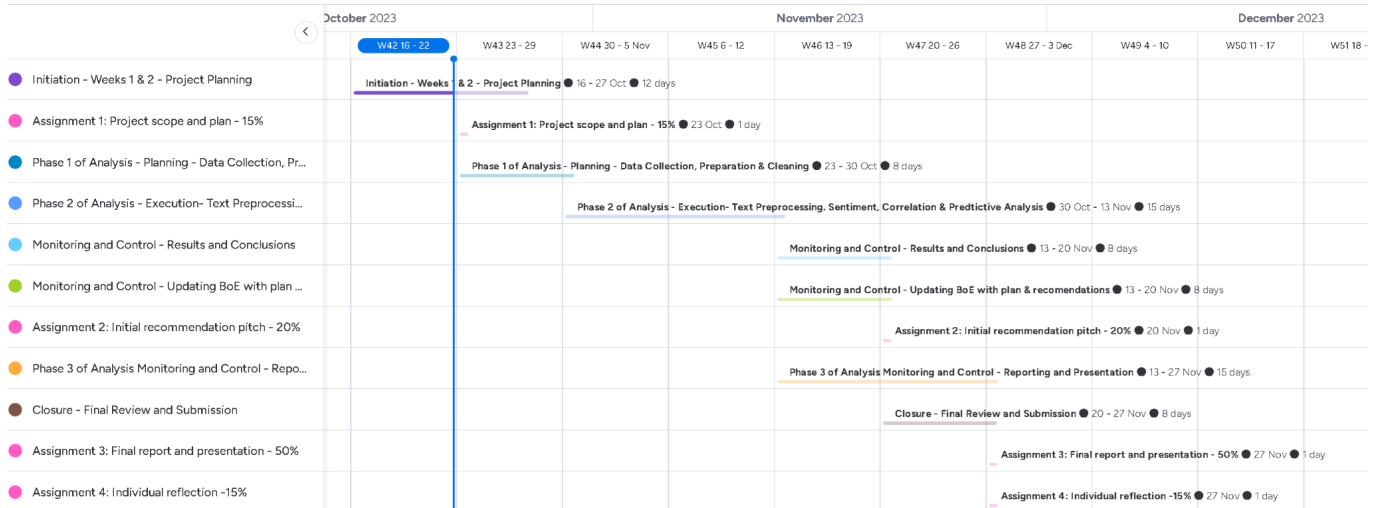

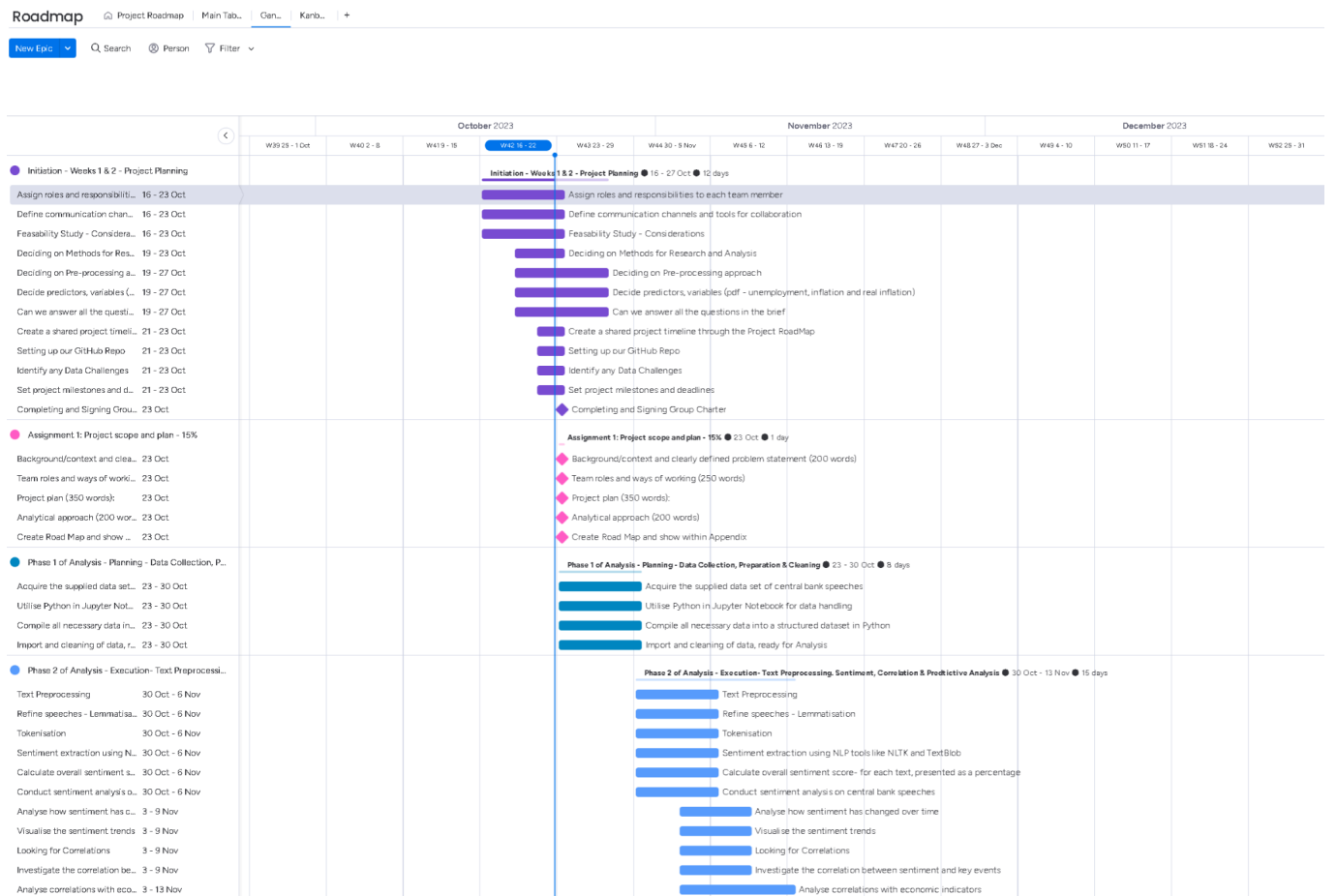**Figure 1 - Project Roadmap - Summary**

Figure 2(i) - Project Roadmap - Detailed - Part 1

Assignment 1: Project Scope and Plan
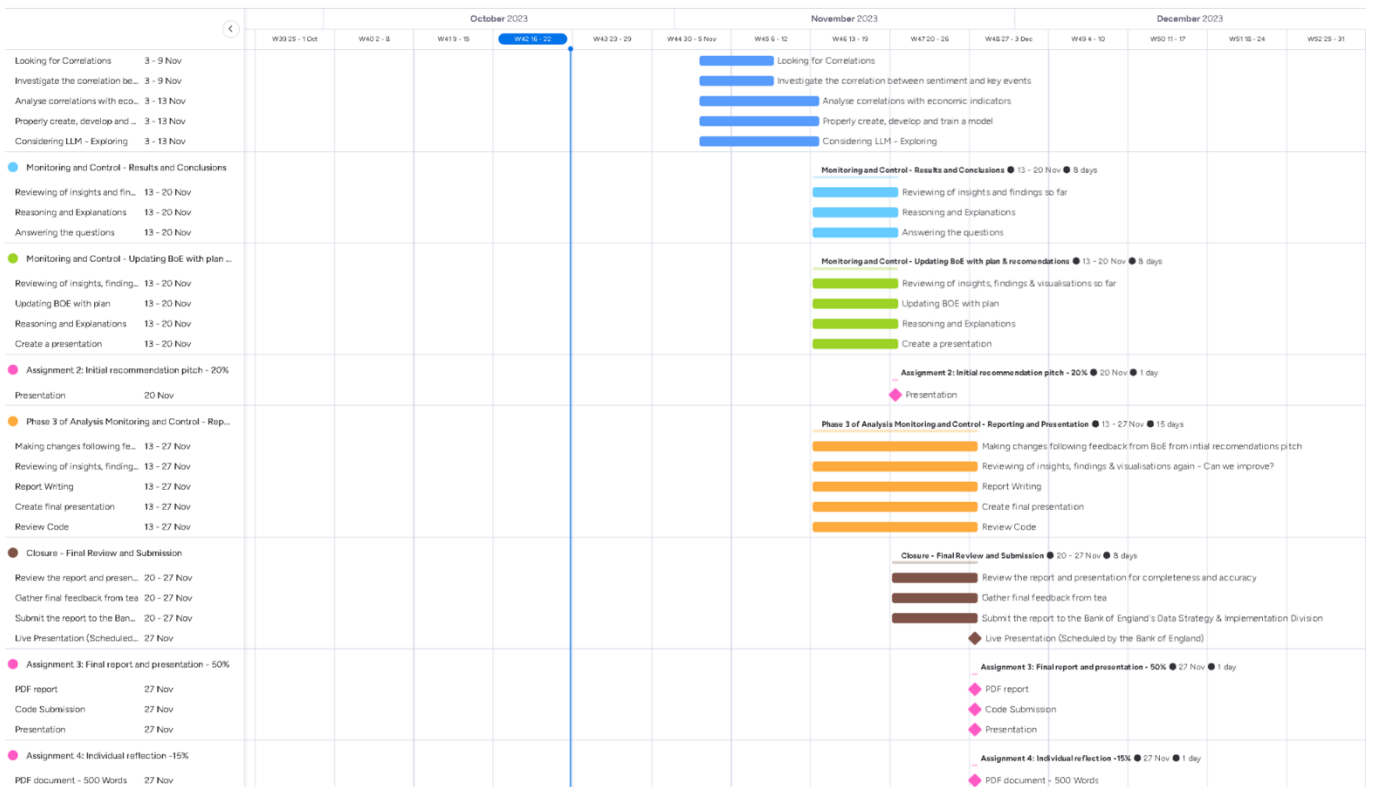

Figure 2(ii) - Project Roadmap - Detailed - Part 2

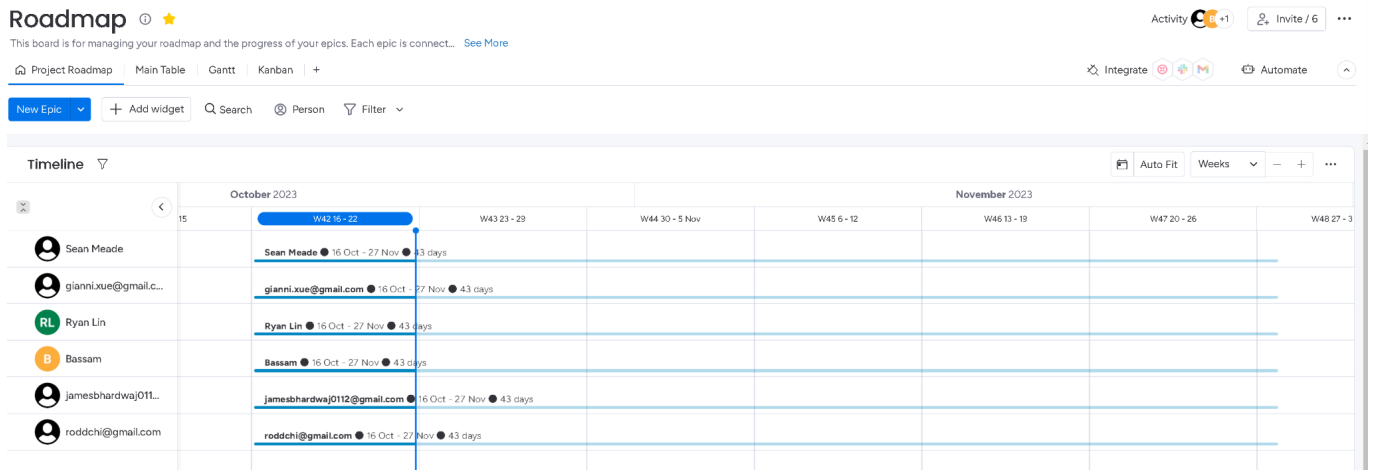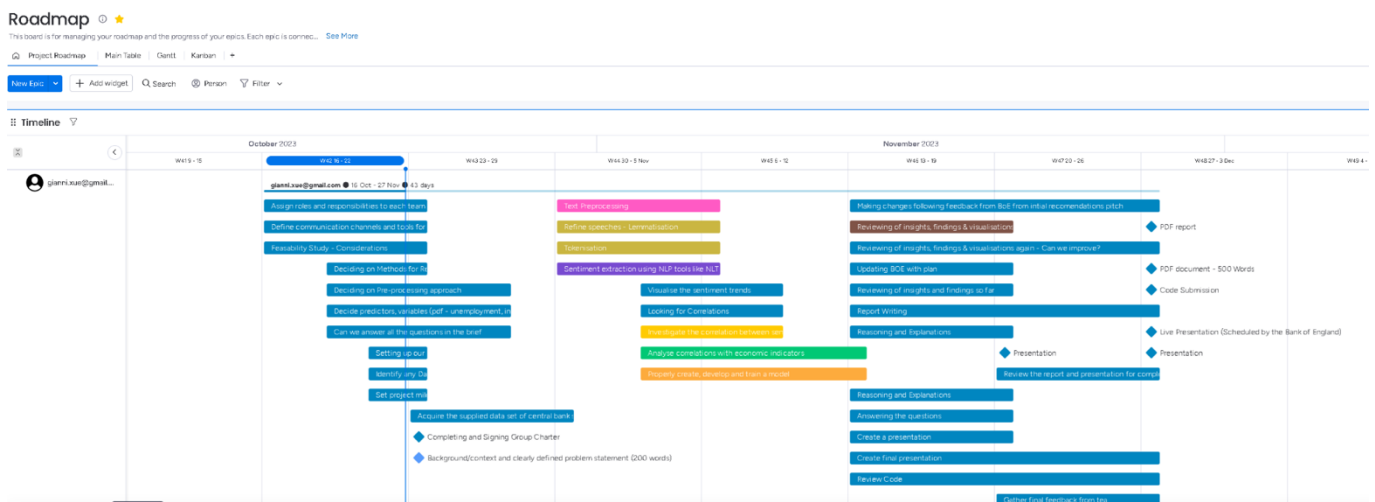**Figure 2(iii) - Project Roadmap - Detailed - Part 3**

Figure 3 - Project Roadmap - Assigned Tasks Summary
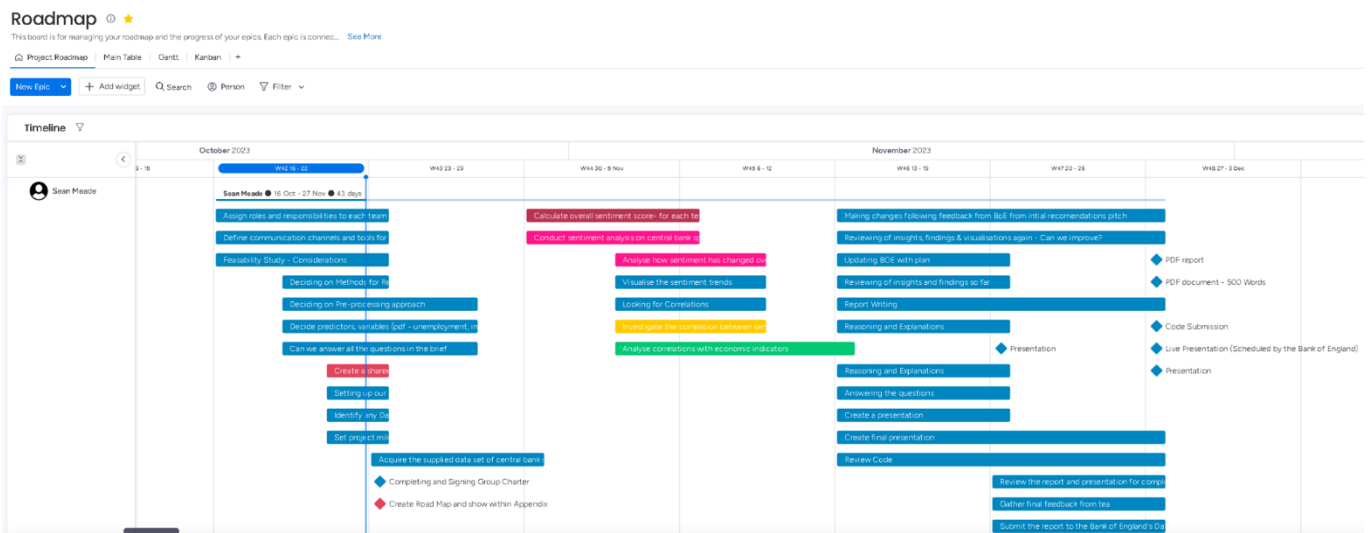
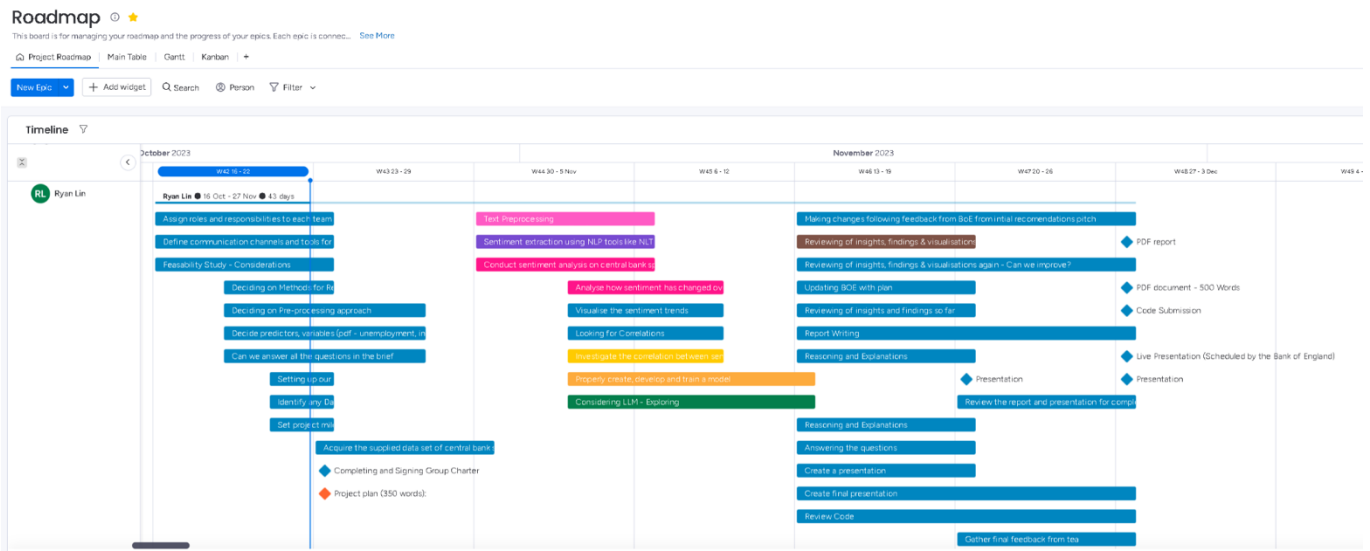**Figure 4(i) - Project Roadmap - Assigned Tasks Detailed - Part 1**

Figure 4(ii) - Project Roadmap - Assigned Tasks Detailed - Part 2

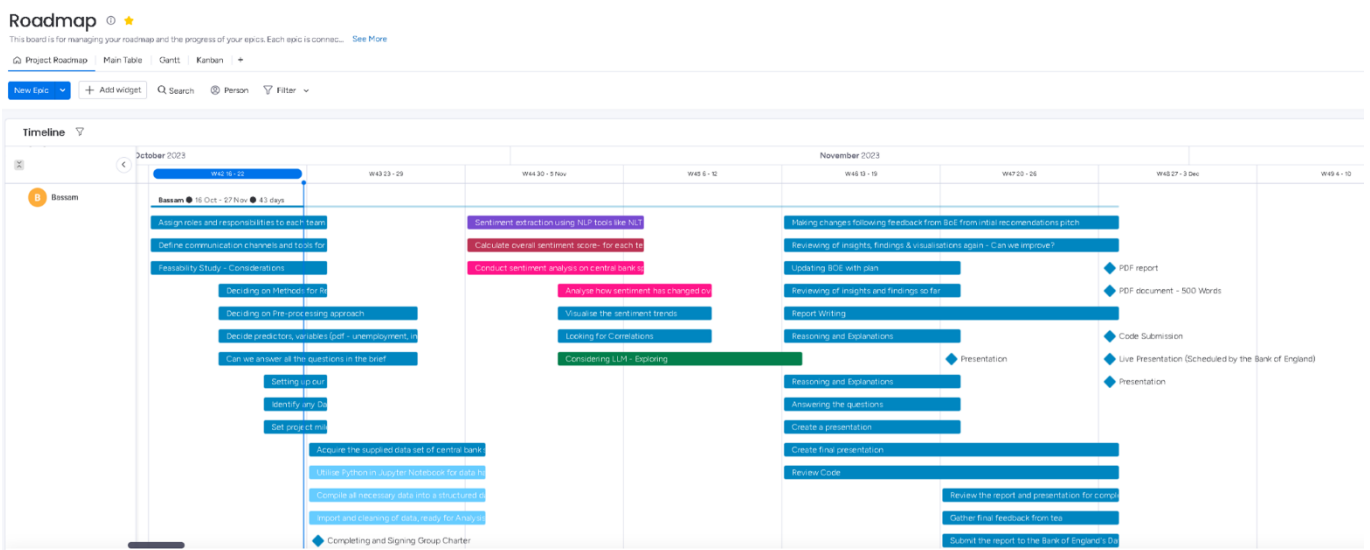**Figure 4(iii) - Project Roadmap - Assigned Tasks Detailed - Part 3**

Figure 4(iv) - Project Roadmap - Assigned Tasks Detailed - Part 4

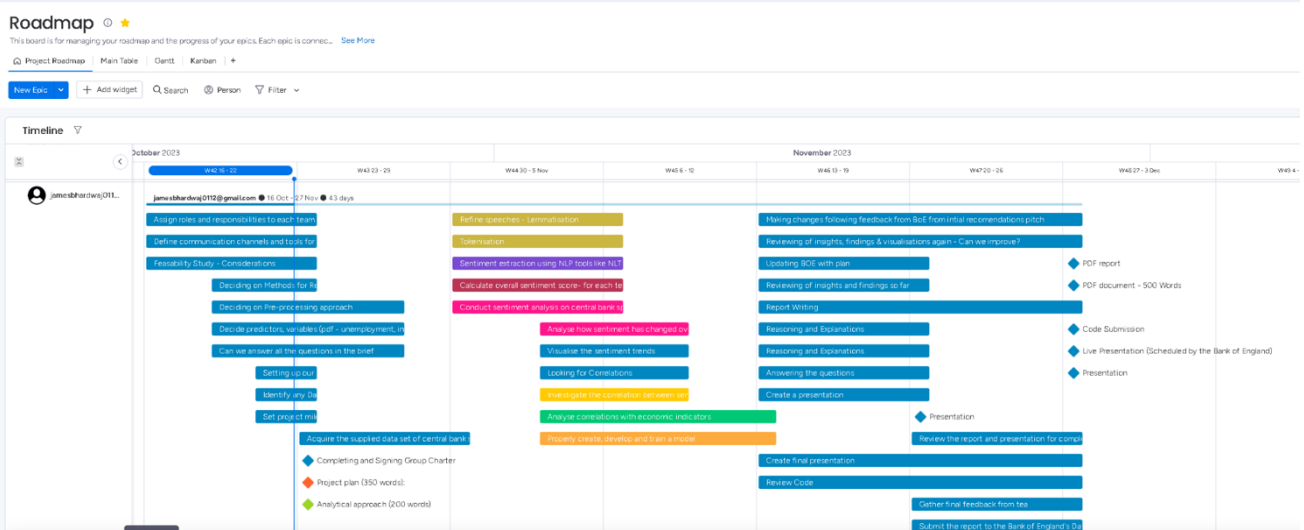**Figure 4(v) - Project Roadmap - Assigned Tasks Detailed - Part 5**

Figure 4(vi) - Project Roadmap - Assigned Tasks Detailed - Part 6