Assignment 1:

Project Scope and Plan

Bank of England Employer Project

R. Bini

O. Bowden

S. Kemper

K. Kirby

A. Mitra

C. Tachikawa

1. Background and Problem Statement

The Bank of England ensures monetary and financial stability in the UK through regulatory oversight and prudential supervision of financial institutions. Quarterly announcements from global systemically important banks (G-SIBs) provide insights into financial performance and risks but are lengthy and unstructured, making analysis challenging. Traditional data science methods struggle with these complex texts, requiring both technical and financial expertise.

This project explores whether quarterly result announcements add value to the Bank's risk assessments by addressing:

- What topics and sentiments emerge from earnings call transcripts?

- Can pre-trained language models effectively cluster and analyse them?

- How can insights be summarised for decision-makers?

By leveraging NLP and Generative AI for topic modelling, sentiment analysis, and summarization, the project aims to improve the Bank's ability to interpret unstructured data with greater accuracy and context awareness. This will enhance risk assessments, support benchmarking against peer institutions, and provide deeper insights into emerging financial trends.

---

2. Roles and Ways of Working

Our team is self-directed and cross functional, with agile methodologies at heart - focusing on speed and continuous improvement to deliver maximum value.

- **Project Lead – Sheldon Kemper:**
  Leads the project, aligning tasks with the Bank of England's objectives, managing timelines, and ensuring effective communication of insights.

- **Research Lead – Oscar Bowden:**
  Conducts research strategy, ensuring methodologies and insights align with the Bank of England's requirements.

- **Data Engineering Lead – Sheldon Kemper:**
  Collects, cleans, and prepares quarterly announcement transcripts and related data from reliable sources.

- **Data Science Lead – Chiaki Tachikawa:**
  Manages modelling workflows, ensuring robust application of sentiment analysis and topic modelling, in coordination with the NLP lead.

- **NLP Lead – Arijit Mitra:**
  Develops NLP pipelines, refining FinBERT sentiment analysis and BERTopic clustering for financial transcripts.

- **Analytics Lead – Rita Bini:**
  Manages analytical workflows and insights ensuring alignment with overall objective (sentiment, topic trends comparisons, benchmarking).

- **Reporting Lead – Kasia Kirby:**
  Compiles and communicates findings, ensuring reports and presentations meet the Bank of England's needs.

Ways of Working:

- **Communication:** Use Jira for tasks, Confluence for documentation, Slack for updates, and Google Meet for weekly discussions.

- **Meetings:** Weekly progress reviews, with ad-hoc sessions as needed.

- **Work Agreements:** Meet deadlines, engage in discussions, provide updates, and resolve issues collaboratively (escalating to the Project Lead if needed).

---

3. Project Plan

<u>Project Scope</u>

Inclusions

- Analysis of public quarterly results announcements and Q&A session transcripts.

- Topic modelling and sentiment analysis utilising pre-trained NLP models to discern emerging themes and evaluate the sentiment.

- Development of a summarisation pipeline to condense lengthy transcripts into digestible summaries grouped by categories.

- Benchmarking findings across different banks to provide a comprehensive view of risk factors and market sentiments.

- Insights on the implications of findings for financial stability.

Exclusions

- Proprietary or non-public data from G-SIBs.

- Real-time analysis of new data post-project deadlines.

Milestones

	Milestone	Deadline
1	External - Project Scope and Plan Submission	10 February 2025
2	Research Completed	12 February 2025
3	Data Cleaning and Methodology Design Completed	18 February 2025
4	External - Preliminary Solution Pitch	24 February 2025
5	NLP Build Complete	4 March 2025
6	External - Final Presentation and Report	10 March 2025
7	Individual Reflection Report	17 March 2025

Phases of the Project Life Cycle

Initiation (Week 1, 29 Jan - 4 Feb)

- Define project objectives, scope, and success criteria.

- Identify target G-SIBs for analysis (e.g., JPMorgan, Barclays) and confirm data availability.

- Finalise team roles and project roadmap.

Planning (Week 2, 5 Feb - 11 Feb)

- **Data Collection:**
  Gather quarterly announcements and transcripts from investor relations websites. Although the initial focus is on PDFs, remain open to additional formats (e.g., HTML, video transcripts).

- **Exploratory Data Analysis (EDA):**
  Assess the structure and quality of the data.

- **Pipeline Design:**
  Develop initial plans for text preprocessing, sentiment analysis, and topic modelling.

Execution (Weeks 3–5, 12 Feb - 04 Mar)

- **Data Preparation:**
  Clean and preprocess transcripts to ensure consistency (handling speaker tags, timestamps, etc.).

- **Modelling:**
  Implement NLP pipelines using FinBERT for sentiment analysis and BERTopic for topic clustering.

- **Summarisation:**
  Develop methods (using models like BART or T5) to summarise insights by for example topic, financial metric, and speaker.

- **Benchmarking:**
  Compare and contrast results against peer institutions.

Monitoring and Control (Ongoing)

- Track progress using Jira.

- Hold weekly review meetings to monitor risks, address blockers, and iteratively refine models.

- Implement quality assurance processes and peer reviews to ensure accuracy, reliability, and alignment with project objectives.

Closure (Week 6, 04 Mar - 10 Mar)

- Finalise all deliverables, including charts, reports, and code repositories.

- Rehearse and prepare the final presentation for the Bank of England.

Dependencies

- Availability of external libraries and pre-trained models.

- Timely access to accurate and complete data from investor relations websites.

- Sufficient computational resources for model training and testing.

Please see our Roadmap in the Appendix.

---

5. Risks

Risk	Impact	Likelihood	Mitigation Strategy
Delayed Data Access	High	Medium	Proactively identify reliable data sources early.
Tool Implementation Challenges	Medium	Low	Schedule team sessions to troubleshoot issues.
Stakeholder Misalignment	High	Medium	Schedule regular updates and feedback sessions.
Overload Due to Team Member Absence	High	Medium	Document workflows; cross-train team members.
Data Quality and Availability	High	Medium	Standardise data preprocessing to handle inconsistencies.
Model Performance	Medium	Medium	Fine-tune models using domain-specific financial datasets.
Time Constraints	High	Medium	Prioritise critical tasks and allocate buffer time in the timeline.

6. Evaluation

Evaluation Metrics:

- **Model Performance:** Coherence scores, sentiment accuracy, and qualitative summarisation feedback.

- **Data Quality:** Completeness and consistency of transcripts.

- **Deliverable Quality:** Usability and clarity of dashboards, reports, and presentations.

Success Criteria:

- Meaningful topic clustering and accurate sentiment classification.

- Clear, concise summaries supporting risk assessments.

- Positive stakeholder feedback in reviews and presentations.

---

7. Deliverables

- **Final Report:**
  A comprehensive document detailing methodology, analysis, findings, and recommendations.

- **Data Insights:**
  Visualisations of sentiment scores, topic distributions, and benchmarking comparisons.

- **Code Repository:**
  Well-documented code for data collection, preprocessing, modelling, and summarisation.

- **Presentation Slides:**
  A final presentation summarising the project's insights and recommendations.

---

8. Appendix

- **Project Roadmap Screenshot:**