Team member's details:

Name	Email	Country	College/Company	Specialization
Leena Ganta	leenarganta@gmail.com	USA	North Carolina State University	NLP
Najma Abdi	abdinajma225@gmail.com	KENYA	UNIVERSITY OF NAIROBI	DATA SCIENCE
Adama	adamal.sall@gmail.com	Saudi Arabia	University Of Suffolk	BSc Honours Computer Science(AI and Data Science)

Problem description (Leena):

ABC Bank needs to predict whether a customer will subscribe to a term deposit product (target **y**: *yes/no*) using historical marketing campaign data. An accurate model will let marketing focus on high-probability prospects, reducing campaign cost and improving conversion rates.

Key Challenges

- Highly imbalanced target (~11 % "yes").
- Feature *duration* leaks post-call information; must build two models (with & without).
- Businesses must understand and trust the drivers (interpretability matters).

Success Metric (Business)

- **Lift** in conversion vs. random dialing at fixed call budget (e.g., top-k lift).
- **Cost savings** from fewer calls while maintaining current deposit sales.

Problem description:

ABC Bank is planning to launch a new term deposit product. Before investing in a large-scale marketing campaign, the bank wants to use a machine learning model to identify customers who are highly likely to subscribe, based on their previous interactions with marketing campaigns.

However, the dataset exhibits a class imbalance (most customers don't subscribe) and certain highly predictive features, such as 'duration', can't be used for real-time targeting.

The goal is to develop and compare ML models (with and without the 'duration' feature) in order to:

- Improve the precision of campaign targets.
- Reduce the operational costs of telemarketing and messaging.
- Convert ML model metrics into insights that business stakeholders can understand.

The main goal is to enable cost-efficient, data-driven campaign targeting with measurable business impact.

Business understanding (Najma):

In a fast-moving banking environment, targeting the right customer is key to staying ahead. ABC Bank wants to move beyond generic outreach and focus its efforts on customers most likely to open a term deposit—*before* any contact is made.

By leveraging machine learning to analyze past customer interactions, the bank can predict who is most likely to say "yes." This means fewer wasted calls, lower marketing costs, and better engagement.

The model will help streamline campaign efforts—allowing teams to prioritize high-potential leads and tailor their approach. While some data, like call duration, can boost prediction accuracy, it isn't available before contacting customers. So, we're also building a version of the model that works without it—ensuring practical, real-time use.

In short, this project transforms raw customer data into smarter decisions—helping ABC Bank boost term deposit conversions, reduce cost per acquisition, and make every marketing effort count.

ABC Bank's marketing team invests heavily in telemarketing and digital outreach. These efforts have been applied broadly, often resulting in the high expenditure of resources and low conversion rates.

So, now the business aims to:

- Optimise marketing costs
- Improve conversion rates.
- Enable business interpretability.
- Avoid outdated features:

KPIs (key performance indicators):

- Uplift in conversion rate compared to previous campaigns
- Reduction in cost per acquisition (CPA)
- Accuracy and precision of the ML model

Project lifecycle along with deadline (Adama):

Project lifecycle and deadline:

We plan to carry out the project in four weeks, focusing on a structured data science methodology using Cross-Industry Standard Process for Data Mining (CRISP-DM). Each week has defined deliverables and stakeholder engagement points.

Weekly activities:

Week 1:

- Business requirements validation
- Data exploration and quality checks
- Identify any missing, inconsistent or unusable fields.
- Generate an initial data intake report.
- Define modelling constraints

Week 2:

- Exploratory data analysis (EDA).
- Create visualisations and perform correlation analysis.
- Perform data pre-processing
- Document the impact of imbalance and early strategies.

Week 3:

- Train and tune models (Logistic Regression, Random Forest and XGBoost).
- Compare models with and without 'duration'.
- Implement class imbalance strategies (SMOTE and class weights).
- Evaluate model performance using machine learning metrics.

Week 4:

- Convert ML metrics to business impact metrics (e.g. ROI, CPA).
- Prepare a non-technical presentation.
- Finalise and deploy the best-performing model.
- Submit a technical report and an educational summary.

GitHub Repo Link (Leena): https://github.com/leenarganta/bank_marketing_campaign

---

Data Intake Report (Adama):

After downloading the data and uploaded it in JupyterNotebook along with “import pandas as pd” module, we have found the following:

Dataset: bank-additional-full.csv
Date Analyzed: 2025-06-19
Tool Used: Python (pandas)
Data Overview:

- Shape: 41188 rows × 21 columns
- Target Variable: 'y' (binary: yes/no)

Class Distribution (Target)

- No: 36548 (88.73%)
- Yes: 4640 (11.27%)

Conclusion: Severe imbalance detected; mitigation strategies needed.
Data Quality Checks:
Missing values (as 'unknown'):

- job: 330
- marital: 80
- education: 1731
- default: 8597
- housing: 990
- loan: 990

Duplicate rows: 12

Recommended Actions

- Apply SMOTE or weighted loss function to handle class imbalance.
- Encode categorical features using One-Hot or Label Encoding.
- Normalize/standardize numeric features if needed.
- Exclude 'duration' for deployable models.
- Continue with feature correlation analysis and EDA.