# Problem Definition

ABC Bank, a leading financial institution, endeavors to launch a new fixed-term deposit product, catering to potential customers seeking secure and lucrative investment options. Understanding the evolving financial landscape and customers' changing preferences, the bank embarks on a strategic initiative to leverage its vast reservoir of customer data and historical banking interactions.

By harnessing the power of data analytics and machine learning, ABC Bank aims to unravel intricate patterns within customers' past banking behaviors. This holistic approach encompasses a deep analysis of various financial interactions, encompassing deposit patterns, loan histories, investment tendencies, and overall financial preferences.

The core objective revolves around not just the introduction of a new financial product but the personalized engagement of potential clients. Through the development of a sophisticated machine learning model, derived from a wealth of historical data, the bank aspires to craft targeted marketing strategies. These strategies will be finely tuned to resonate with specific customer segments identified through comprehensive data analysis.

The ultimate goal is twofold: first, to introduce the fixed-term deposit product in a manner aligned with customer preferences, and second, to optimize marketing efforts. By targeting individuals more inclined towards secure and profitable investments, ABC Bank seeks to not only boost sales but also enhance customer satisfaction through tailored financial offerings.

This initiative aligns with the bank's commitment to fostering financial literacy and offering innovative products designed to meet the diverse investment needs of its valued clientele. With a strong focus on customer-centricity and data-driven decision-making, ABC Bank endeavors to continue its legacy of providing financial solutions tailored to individual aspirations and financial goals..

# Project Lifecycle

**Project Definition and Planning:**

- Project Objective: Increase sales of the fixed-term deposit product.
- Timeline: Completion within a 6-week period.
- Team and Resources: Comprising data scientists, the marketing team, and technical support.
- Data Collection: Involves gathering bank customer data and historical banking interaction records.

**Data Understanding: (19 Nov 2023 to 26 Nov 2023)**

- Data Sources: Acquiring customer database and historical interaction records.
- Data Inspection and Preprocessing: Includes analyzing and preprocessing data (addressing missing values, outliers, etc.).

**Exploratory Data Analysis: (26 Nov 2023 to 02 Dec 2023)**

- Data Exploration: Understanding distributions, relationships, missing elements, and feature importance.
- Visualization: Creating visual representations for feature distributions, correlations, and identifying data imbalance.
- Data Preparation:
- Feature Engineering: Selection or creation of relevant features.

- Data Splitting: Creating training, validation, and test sets.
- Handling Imbalanced Data: Applying techniques to address imbalanced datasets, such as SMOTE.

**Model Development: (09 Dec 2023 to 23 Dec 2023 )**

- Model Diversity: Building various models (e.g., Logistic Regression, RandomForestClassifier, Gradient Boosting).
- Performance Evaluation: Tuning model parameters and assessing performance using Cross-Validation.

**Model Selection and Performance Reporting: (09 Dec 2023 to 23 Dec 2023 )**

- Model Comparison: Analyzing model performances (Accuracy, F1-Score, Precision, Recall).
- Optimal Model Selection: Choosing the most suitable model and creating a performance report.

**Model Deployment and Implementation: (09 Dec 2023 to 23 Dec 2023 )**

- Model Integration: Integrating the chosen model into bank systems.
- Operationalization: Making the model operational and conducting comprehensive testing.

**Conversion to Business Metrics and Presentation Preparation: (23 Dec 2023 to 30 Dec 2023 )**

- Business Understanding: Making the model's performance comprehensible for

non-technical stakeholders.

- Presentation Preparation: Crafting a presentation for non-technical individuals, outlining the model's outcomes and potential impacts.

**Monitoring Results and Improvements: (23 Dec 2023 to 30 Dec 2023 )**

- Performance Monitoring: Tracking the model's performance and assessing real-world outcomes.
- Continuous Development: Providing suggestions for improvement and keeping the model updated with new data.

# Data Intake Report

Group Name: Banking Insights Squad

Group: Members: Canmert Demir & Josep Pang

Name: Canmert Demir - Bank Marketing (Campaign) -- Group Project
Email: canmertdemir2@gmail.com
Country: Turkey
College/Company: Msc Bartin University - Applied Mathematics / Data Glacier
Specialization: Data Science

Name: Josep Pang - Bank Marketing (Campaign) -- Group Project
Email: joseph302156@gmail.com
Country: United States
College/Company: University of California, Berkeley/ Data Glacier

Specialization: Data Science

Report date: 17.11.2023

Internship Batch: 30 Sept to 30 Dec 2023

Version: 3.11.5

Data intake by: The UC Irvine Machine Learning Repository

Data intake reviewer: Canmert Demir, Josep Pang

Data storage location: https://github.com/Canmertdemir/DataGlacierWeek7

**Tabular data details:**

| | |
|---|---|
| Total number of observations | 4521 |
| Total number of files | 1 |
| Total number of features | 17 |
| Base format of the file | bank.csv |
| Size of the data | 11,4 MB |

| | |
|---|---|
| Total number of observations | 4119 |
| Total number of files | 1 |
| Total number of features | 21 |
| Base format of the file | bank-full.csv |
| Size of the data | 4,39 MB |

| | |
|---|---|
| Total number of observations | 45211 |
| Total number of files | 1 |
| Total number of features | 17 |
| Base format of the file | bank-additional..csv |
| Size of the data | 0.556 MB |

| | |
|---|---|
| Total number of observations | 41188 |

| | |
|---|---|
| Total number of files<br>Total number of features | 1<br>21 |
| Base format of the file | bank-additional-full..csv |
| Size of the data | 5,56 MB |

**Proposed Approach:**

The dataset consists of four main components: `bank`, `bank_additional`, `bank_full`, and `bank_additional_full`. These datasets vary in the number of features and observations, and the main goal is to analyze and model the `bank_additional` dataset. Following step is used to understand data set in first place. Also, aknowledge the dataset is inbalance, recognizing its impact on subsequent analytical steps.

Data Exploration and Preprocessing:

- Dataset Overview Review the shape, statistical summary, variable types, and head/tail of each dataset using the `quick_look()` function.
- Handling Imbalance:** Note the imbalance in the dataset and address it accordingly.
- Target Variable:** Convert the `y` variable into boolean values for modeling purposes.

Correlation Analysis:

- Identify correlations between numerical features using `cor_analiz_cardinals()` function.
- Note the significant correlations, especially those related to the target variable (`y_bool`).

Variable Categorization:

- Utilize the `grab_col_names()` function to categorize variables into different types:
- Categorical Variables: Including both categorical and numerical categorical variables.
- Numerical Variables: Pure numerical variables without considering categorical representation.
- Categorical but Cardinal Variables: Categorical variables with high cardinality.
- Numerical but Categorical Variables: Numerical variables that function similarly to categorical variables.