Team member's details: **Individual**

Name: **Caleb Nyakeyo Anthony**

Email: **calebnyakeyo2018@gmail.com**

Country: **Kenya**

College/Company: **University of Embu**

Specialization : **Data Science.**

## PROBLEM DESCRIPTION.

ABC Bank wants to sell it's term deposit product to customers and before launching the product they want to develop a model which help them in understanding whether a particular customer will buy their product or not (based on customer's past interaction with bank or other Financial Institution).

## BUSINESS UNDERSTANDING

In business understand phase we basically understands the following in a bank:

### *a. The business/bank process*

**Customer onboarding:**A smooth and quick customer onboarding process wins the trust and loyalty of the customer. Legacy banking institutions depend on manual methods for gathering and verifying customer information. Modernized banks gather customer information through websites, mobile, and third-party platforms. Know your customer (KYC) verification is an important/mandatory part of the **customer onboarding process**. Owing to the rise in online frauds and financial terrorism, banks have made the KYC process **m**andatory. Anti-money laundering checks are also mandatory in customer onboarding.

**Loan processing:**Loan processing is an exhaustive process with several important steps from loan origination to disbursement. Legacy banking systems process loans manually, which entails tons of paperwork, inordinate delays in approvals and verification, unexplained bottlenecks, increased scope for fraud, and multiple rejections and re-application. Modern banking banks need to have speedy and efficient loan processing systems that can handle huge volumes of loan originations across multiple platforms.

**Credit card processing:** Credit and debit card processing is an expansive process that requires accurate capture of data across multiple banking channels and real-time updation of bank records. When done manually, credit card processing not only takes time and effort but is also prone to errors and inconsistencies. Automated credit and debit card processing cuts down significantly on the cost and time of processing and ensures 100% accuracy.

**Core Banking Operations:**Core banking operations make up all the back-end updates in the banking industry. These processes include customer data updates, KYC validations, accounting reconciliations, and overdraft protection. Speed and accuracy are paramount for core banking operations. Manual updates cannot provide the accuracy and speed required in core banking operations. Automated processing, especially robotic process automation (RPA), can increase the accuracy and speed of core banking operations.

**Risk and compliance:**Banks and financial institutions need to adhere to several policies and regulations. Updates on financial regulations and the creation of new policies is a routine feature in the banking and financial industry. Risk and compliance management is an important function in banking. Banks and financial institutions need to be audit-ready at all times by updating their data and policies. Anti-money laundering and fraud checks, KYC verification, and compliance checks can be effectively done through automation.

## b. ***Define and Frame the business problem***

ABC Bank wants to sell it's term deposit product to customers and before launching the product they want to develop a model which help them in understanding whether a particular customer will buy their product or not (based on customer's past interaction with bank or other Financial Institution).

This is a binary classification problem. Our two classes are “yes” denoting that the customer subscribed to a term deposit, and “no” denoting that the customer did not subscribe.

***About the Problem***

We are given the data of direct marketing campaigns (phone calls) of a Portuguese banking institution. The classification goal is to predict if the client will subscribe a term deposit (target variable y). This case study is inspired by this research paper where the researchers have used a very similar dataset as the one we will be using throughout this case study for determining the success of Bank Telemarketing. The researchers in their paper have mentioned that the best result they have got was a AUC score of 0.8 and a ALIFT of 0.7. So as a goal we will try to produce a similar result in our case study.

## c. ***Define the business objective***

Bank wants to use ML model to shortlist customer whose chances of buying the product is more so that their marketing channel (tele marketing, SMS/email marketing etc) can focus only to those customers whose chances of buying the product is more.

This will save resource and their time ( which is directly involved in the cost ( resource billing)).

The bank wishes to develop model without duration feature and report the performance of the model.

The classification goal is to predict if the client will subscribe a term deposit (target variable y).

## d. ***Agree on success criteria***

These stage involves the definition of the method used to solve the problem and achieve the laid down objectives.

The performance metric used for this case study is AUC ROC score also known as **AUROC** (**Area Under the Receiver Operating Characteristics**).

**Why use AUC?**
The reason we are choosing AUC over accuracy is because,as we will see in Exploratory data analysis, the dataset we are working with is an imbalanced dataset with the class “no” being the majority class. If we use accuracy as our metric, any random model can give us a very good accuracy. But at the end, it will be a random model. AUC gets over this problem by looking into both the True positive rate (TPR) and False positive rate (FPR). Only if both the TPR and FPR are well above the random line in the ROC curve, we will get a good AUC. Accuracy does not guarantee that.

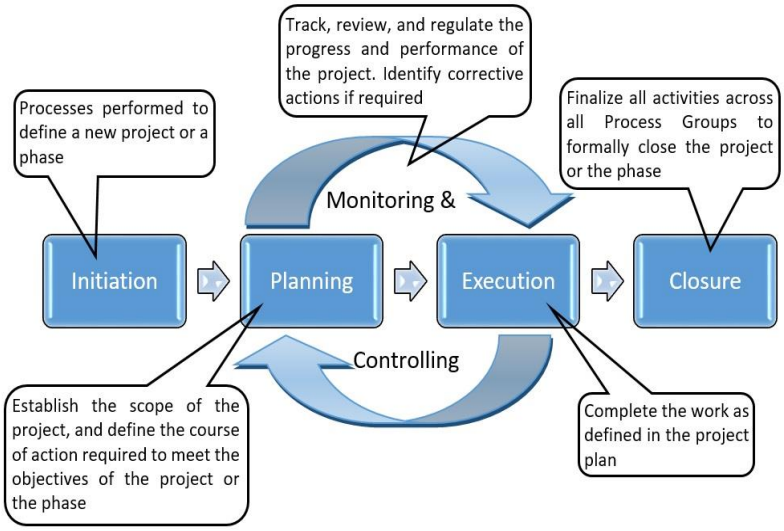
## PROJECT LIFE CYCLE

**The project initiation stage:** understand the goals, priorities, deadlines(for this project, *the deadline is on 18th December 2021*), and risks of the project

**The project planning stage:** outline the tasks and timeline required to execute on the project

**The project execution stage:** turn your plan into action and monitor project performance

**The project closure stage:** analyze results, summarize key learnings, and plan next steps




## DATA INTAKE REPORT

Name: G2M insight for Cab Investment firm
Report date: **8th November, 2021.**
Internship Batch: **LISUM04**
Version:
Data intake by: **Caleb Anthony**
Data intake reviewer:
Data storage location: https://archive.ics.uci.edu/ml/datasets/Bank+Marketing

**Tabular data details: Bank-additional-full.csv**

| | |
|---|---|
| **All Examples** | 41188 |
| **Number of Inputs** | 20 |
| **Ordered by date** | From May 2008 to November 2010 |
| **Base format of the file** | .csv |

**Tabular data details: Bank-additional.csv**

| | |
|---|---|
| **All Examples** | 4119(Randomly selected from 1) |
| **Number of Inputs** | 20 |
| **Ordered by date** | From May 2008 to November 2010 |
| **Base format of the file** | .csv |

**Tabular data details: Bank-full.csv**

| | |
|---|---|
| **All Examples** | 4119 |
| **Number of Inputs** | 17 |
| **Ordered by date** | Ordered by date(older version of this dataset with less inputs). |
| **Base format of the file** | .csv |

**Tabular data details: Bank.csv**

| | |
|---|---|
| **All Examples** | 412 |
| **Number of Inputs** | 17 |
| **Ordered by date** | randomly selected from 3 (older version of this dataset with less inputs) |
| **Base format of the file** | .csv |

The smallest datasets are provided to test more computationally demanding machine learning algorithms (e.g., SVM).