



# **Data Science Intern at Data Glacier**

**Week 7: Report on Group Project**

**Topic: Bank Marketing (Campaign)**

**Group Name: Campaign Catalysts**

**Specialization: Data Science**

**Batch Code: LISUM19**

**Date: 19<sup>th</sup> April 2023**

**Submitted to: Data Glacier**

## Group Member Details

S.No.	Name	Email	Country	College/ Company
1.	Yash Jayesh Doshi	yashjdoshi99@gmail.com	UAE	Orpheuss LLC
2.	Anuj Singh	dsanuj21@gmail.com	India	Mumbai University
3.	Yash Jadwani	yash.jadwani1998@gmail.com	United Kingdom	Kingston University
4.	Harold Wilson	haroldwilson537@gmail.com	United Kingdom	University of Buckingham

## **1. Problem Description**

ABC Bank wants to sell its 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).

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.

## **2. Business Understanding**

In this problem statement, the marketing strategy involves promoting the term deposit product to potential customers through various channels such as tele marketing, SMS, email marketing, etc. The aim of the marketing strategy is to reach out to a large number of customers and convince them to purchase the product.

However, it can be a costly and resource-intensive process to target every customer. Therefore, ABC Bank wants to develop a machine learning model that can help them identify the customers who are more likely to buy the product. This way, the marketing efforts can be focused on these shortlisted customers, which can significantly increase the chances of success while reducing marketing costs and resources.

By leveraging the insights gained from the machine learning model, ABC Bank can optimize their marketing campaigns, tailor their marketing messages to specific customers, and improve their overall customer acquisition strategy. The bank can also use the model to identify the most effective marketing channels for each customer segment, such as email marketing for younger customers and tele marketing for older customers, for example.

Overall, the goal of the marketing strategy is to increase the customer base and revenue of ABC Bank by targeting the right customers with the right product at the right time, while maximizing the return on investment of their marketing efforts.

### 3. Project Life-cycle

S.No.	Tasks to be completed	Due Date
1.	Data Intake Report and Project Life-cycle	19 <sup>th</sup> April 2023
2.	Data understanding and data preprocessing approach	26 <sup>th</sup> April 2023
3.	Data cleaning and transformation	02 <sup>nd</sup> May 2023
4.	EDA and Final Recommendations	09 <sup>th</sup> May 2023
5.	EDA presentation and recommending ML models	13 <sup>th</sup> May 2023
6.	Model Selection and Model Building	20 <sup>th</sup> May 2023
7.	Model Deployment	25 <sup>th</sup> May 2023
7.	Final Project presentation	30 <sup>th</sup> May 2023

# Data Intake Report

Name: **Bank Marketing Campaign**

Report date: **17-Apr-2023**

Internship Batch: **LISUM19**

Version: **1.0**

Data intake by: **Campaign Catalysts**

Data intake reviewer: **Campaign Catalysts**

Data storage location: **GitHub**

## Tabular data details:

<b>File name</b>	bank.csv
<b>Total number of observations</b>	4521
<b>Total number of files</b>	1
<b>Total number of features</b>	16
<b>Base format of the file</b>	CSV
<b>Size of the data</b>	451 KB

<b>File name</b>	bank-full.csv
<b>Total number of observations</b>	45211
<b>Total number of files</b>	1
<b>Total number of features</b>	16
<b>Base format of the file</b>	CSV
<b>Size of the data</b>	4.40 MB

<b>File name</b>	bank-additional.csv
<b>Total number of observations</b>	4119
<b>Total number of files</b>	1
<b>Total number of features</b>	20
<b>Base format of the file</b>	CSV
<b>Size of the data</b>	571 KB

<b>File name</b>	bank-additional-full.csv
<b>Total number of observations</b>	41188
<b>Total number of files</b>	1
<b>Total number of features</b>	20
<b>Base format of the file</b>	CSV
<b>Size of the data</b>	5.57 MB

## Proposed Approach:

- There are some null as unknown values in **bank-additional.csv** and **bank-additional-full.csv** which can be treated as a possible class-label or we can use deletion or imputation techniques.
- There are 12 duplicate values in **bank-additional-full.csv** which we can drop during preliminary data analysis.

- The education column in **bank-additional.csv** and **bank-additional-full.csv** contains several categories of education levels. In order to ensure high data quality and avoid any distortion in the analysis, some values in this column need to be fixed. To improve the data quality, we may need to perform data cleaning operations such as correcting misspelled values or consolidating similar categories. This will help to ensure that the education data is accurate, consistent, and meaningful for analysis purposes.

### References:

- [Moro et al., 2014] S. Moro, P. Cortez and P. Rita. A Data-Driven Approach to Predict the Success of Bank Telemarketing. *Decision Support Systems*, Elsevier, 62:22-31, June 2014
- [Moro et al., 2011] S. Moro, R. Laureano and P. Cortez. Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodology. In P. Novais et al. (Eds.), *Proceedings of the European Simulation and Modelling Conference - ESM'2011*, pp. 117-121, Guimarães, Portugal, October, 2011. EUROSIS.

### GitHub Repo Link:

<https://github.com/singhanui695/Data-glacier-Group-Project>