

# Data Science - Bank Marketing Campaign

## Table of Contents

<i>1. Group Information .....</i>	<i>2</i>
<i>2. Problem description .....</i>	<i>3</i>
2.1. Introduction.....	3
<i>3. Business understanding .....</i>	<i>4</i>
3.1. Business Problem:.....	4
3.2. The Data: .....	4
3.3. Columns Description:.....	4
<i>4. Project life cycle along with the deadline.....</i>	<i>7</i>
<i>5. Data intake report .....</i>	<i>8</i>
5.1. Bank-additional-full data.....	8
5.2. Bank-additional data .....	8
5.3. Bank-additional-name data.....	9
<i>6. Github repository.....</i>	<i>9</i>

## 1. Group Information

Group Name: Datalux

Group Members: 3

Name	Email	Country	Uni/Company
Huu Thien Nguyen	<a href="mailto:nguyenuuthien27296@gmail.com">nguyenuuthien27296@gmail.com</a>	Sweden	Skövde University
Asmaa Alqurashi	asmaa.idk@gmail.com	Saudi Arabia	Taif University
Deepak Rawat	deepakrawat68@gmail.com	Ireland	Dublin Business School

Specialisation: Data Science

Submitted to: Data Glacier canvas platform

Internship Batch: LISUM09

## 2. Problem description

### 2.1. Introduction

A common approach for increasing business is to run marketing and selling campaigns. Companies use direct marketing to reach certain categories of clients to achieve a specific goal. Customer distant interactions may be centralised in a contact centre, making campaign administration easier. Technology allows us to reimagine marketing by optimising customer lifetime value through analysing accessible data and customer KPIs, allowing us to develop longer and closer relationships in line with company needs.

The ABC Bank wants to market its term deposit product to clients in this project. Before doing so, they want to construct a machine learning model that will assist them in determining whether a particular consumer would buy their product based on the customer's previous interactions with the bank or other financial institution.

To solve the problem mentioned above, the bank wants to use machine learning modelling to identify the customers who are more reluctant to buy their services so that their marketing channels will only focus on these customers, which in turn will save the time and resources and finally leads to optimised cost for this campaign.

### 3. Business understanding

#### 3.1. Business Problem:

ABC Bank wants to sell its term deposit product to customers. Before launching the product, they want to develop a model which helps them in understanding whether a particular customer will buy their product or not (based on the customer's past interaction with the bank or other Financial Institution). Bank wants to use the ML model to shortlist customer whose chances of buying the product is more so that their marketing channel (telemarketing, SMS/email marketing etc.) can focus only to those customers whose chances of purchasing the product is more.

#### 3.2. The Data:

The data is related to direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls. Often, more than one contact with the same client was required to assess if the product (bank term deposit) would be ('yes') or not ('no') subscribed.

**Size:** 41188 records, 20 explanatory variables, and one binary response variable

#### 3.3. Columns Description:

- Customer Information:

1 - age (numeric): customers' age which ranges between 17 and 98 in this dataset

2 - job (categorical): customers' job

- admin
- blue-collar
- technician
- services
- management
- retired
- entrepreneur
- self-employed
- housemaid
- unemployed
- student
- unknown

3 - marital (categorical): marital status (categorical)

- married
- single
- divorced
- unknown

4 - education (categorical)

- university.degree
- high.school
- basic.9y
- professional.course
- basic.4y
- basic.6y
- unknown
- illiterate

5 - default (categorical): Does the customer has credit in default?

- yes
- no
- unknown

6 - housing (categorical): Does the customer has a housing loan? (categorical)

- yes
- no
- unknown

7 - loan (categorical): Does the customer have a personal loan? (categorical)

- yes
- no
- unknown

- Communication Information:

8 - contact (categorical): communication type

- cellular
- telephone

9 - month (categorical): last contact month of the year

10 - day\_of\_week (categorical): last contact day of the week (in working days)

11 - duration (numeric): last contact duration, in seconds

- Campaign Information:

12 - campaign (numeric): number of contacts performed during this campaign and for this client

13 - pdays (numeric): number of days that passed by after the client was last contacted from a previous campaign

14 - previous (numeric): number of contacts performed before this campaign and for this client

15 - poutcome (categorical): outcome of the last campaign marketing

- nonexistent
- failure
- success

- social and economic context attributes:

16 - emp.var.rate (numeric): employment variation rate - quarterly indicator

17 - cons.price.idx (numeric): consumer price index - monthly indicator

18 - cons.conf.idx (numeric): consumer confidence index - monthly indicator

19 - euribor3m (numeric): Euro Interbank Offered 3-month rate - daily indicator

20 - nr.employed (numeric): number of employees - quarterly indicator

- **Target:**

21 - y - has the client subscribed to a term deposit? (binary: 'yes','no')

#### 4. Project life cycle along with the deadline

The project's general view, along with the deadline, is described in the table below. The deadline for week 7 is 18 June 2022. The following week is added up accordingly.

Task Name	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12
Data understanding						
Data exploration						
Data preprocessing / Feature engineering						
EDA (Exploratory Data Analysis)						
Presentation						
Modeling						

Table 1: The project's timeline

## 5. Data intake report

The dataset was downloaded from UC Irvine Machine Learning Repository (<https://archive.ics.uci.edu/ml/datasets/Bank+Marketing>) and stored on the group GitHub repository. The link online for the group project can be found in the section “Github repository”.

Tabular data details:

The data folder contains three separate files, which offer different objectives for the users. The detailed usage of each file is described in the table below.

### 5.1. Bank-additional-full data

Total number of observations	41188
Total number of files	1
Total number of features	20
Base format of the file	.csv
Size of the data	5.8 MB
Usage	Contains all observations of the bank data, which is the main file used for this project. The file is sorted by date (from May 2008 to November 2010)

### 5.2. Bank-additional data

Total number of observations	4119
Total number of files	1
Total number of features	20
Base format of the file	.csv
Size of the data	586 KB



<b>Usage</b>	Contains randomly selected observations of the full bank data, which is used for testing if the computing cost is too high. The file is sorted by date (from May 2008 to November 2010)
--------------	---

### 5.3. Bank-additional-name data

<b>Total number of files</b>	1
<b>Base format of the file</b>	.txt
<b>Size of the data</b>	8 KB
<b>Usage</b>	Contains the description of features in the bank data. It also provides additional information, for instance, citations, requests.

## 6. Github repository

The link for GitHub: <https://github.com/AndrewNguyen27296/DataGlacier>