



Virtual Internship Data Science Data Intake Report

Group Name: LISUM19: Data science Group 1

Members:

No	Name	Email	Country	Specialization
1	Preeti Verma	vermapreeti.dataanalyst@gmail.com	Canada	Data Science
2	Thanuja	thanujayadav953@gmail.com	UK	Data Science
3	Abishek James	abishekjames1998@gmail.com	Ireland	Data Science

Name: Bank Marketing (Campaign)

Report date: 13-04-2023

Internship Batch: LISUM19

Data intake by: Preeti Verma

Data intake reviewer: Data Glacier

Data storage location: https://github.com/vermapreeti003/Bank_Campaign_Marketing

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 helps in understanding whether a particular customer will buy the product or not (based on the customer's past interaction with the bank or other Financial Institution). This is an application of the company's marketing data.

Business Understanding:

The goal is to build a Machine Learning model that helps in predicting the outcomes of each customer's marketing campaign and analyzing which features impact the outcomes will help the company understand how to make the campaign more effective. Additionally, categorizing the customer group that subscribed to the term deposit helps to determine who is more likely to purchase the product in the future, thereby developing more targeted marketing campaigns.

This can be accomplished by using an ML model that shortlists the customers with a higher possibility of purchasing the product. So marketing such as telemarketing, SMS or email marketing can concentrate only on those customers. It will save time and resources by doing this.

Project Lifecycle

Deadline (Date/week)	Plan and Deliverables
19 April 2023(Week 7)	<ul style="list-style-type: none">• Problem statement• Business understanding• Dataset collection
26 April 2023(Week 8)	<ul style="list-style-type: none">• Data understanding• Data analysis - finding null values, and outliers.• Data processing
2 May 2023(Week 9)	Data cleaning and transformation
9 May 2023(Week 10)	EDA and Model Recommendation
16 May 2023(Week 11)	EDA Presentation and Proposed Modeling Technique

23 May 2023(Week 12)	Model Selection and Building the Model
30 May 2023(Week 13)	Final project report and code submission

Tabular data details:

File 1: bank_additional_full.csv

Total number of observations	41189
Total number of files	2
Total number of features	21
Base format of the file	.CSV
Size of the data	5836800(5.56MB)

File 2: bank_additional.csv

Total number of observations	4120
Total number of files	2
Total number of features	21

Base format of the file	.CSV
Size of the data	583898(572KB)

Exploratory Data Analysis

1. The data cover a period from May 2008 to November 2010.
2. There are 2 datasets, the second dataset is a sample of the first dataset.
3. There are 10 integers and 11 categorical variables.
4. The missing values in both datasets are presented by an "unknown" string. We changed it to NaN.
5. There are missing values in six variables namely, job, marital status, education, default, housing, and loan. This will be imputed using various methods.
6. There are 12 duplicates in the first dataset and no duplicates in the sample dataset, this will be dropped since they are minimal and will not affect our analysis

Assumptions

We assume the data provided is correct and up to date