Data Intake Report

Project name: Bank Marketing (Campaign) -- Group Project

Report date: 30th August 2022

Internship Batch: LISUM10

Version:1.0
Data intake by: Mohini Kalbandhe , Kashish Joshipura, Amir Shahcheraghian,
Mohammed Maqsood

Data intake reviewer:

Data storage location: https://github.com/amohini099/Banco-de-portugal-
marketing/tree/main/Week%207

Tabular data details:

Total number of observations	45211
Total number of files	1
Total number of features	16
Base format of the file	.csv
Size of the data	3.80 MB

Name	bank-names
Total number of observations	-
Total number of files	1
Total number of features	-
Base format of the file	.txt
Size of the data	4 KB

Name	bank
Total number of observations	4521
Total number of files	1
Total number of features	17
Base format of the file	.csv
Size of the data	461KB

Name	bank-additional-full
Total number of observations	41188
Total number of files	1
Total number of features	21
Base format of the file	.csv

Size of the data	5.8 MB

Proposed Approach:
- Unbalanced for target and outliers, overall data looks clean.
- Looking towards few of numeric feature in the dataset, it looks skewed and need to be transformed.
- Few techniques like IQR score and Information value will be helpful to deal with outliers and skewness.
- Imbalanced dataset can be handled by eliminating unknown values and replacing them with mean.