

# Data Intake Report

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

Report date: 19<sup>th</sup> July 2022

Internship Batch: LISUM10

Version:1.0

Data intake by: Mohini Kalbandhe , Kashish Joshipura, Amir Shahcheraghian

Data intake reviewer:

Data storage location: <https://github.com/amohini099/Banco-de-portugal-marketing>

## Tabular data details:

|                                     |         |
|-------------------------------------|---------|
| <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>             | 3.80 MB |

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

|                                     |       |
|-------------------------------------|-------|
| <b>Name</b>                         | bank  |
| <b>Total number of observations</b> | 4521  |
| <b>Total number of files</b>        | 1     |
| <b>Total number of features</b>     | 17    |
| <b>Base format of the file</b>      | .csv  |
| <b>Size of the data</b>             | 461KB |

|                                     |                      |
|-------------------------------------|----------------------|
| <b>Name</b>                         | bank-additional-full |
| <b>Total number of observations</b> | 41188                |
| <b>Total number of files</b>        | 1                    |
| <b>Total number of features</b>     | 21                   |
| <b>Base format of the file</b>      | .csv                 |
| <b>Size of the data</b>             | 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.