

Data Intake Report

Name: Bank Marketing Campaign

Report date: 19 November, 2022

Internship Batch: LISUM14

Version: 1.0

Data intake by: Sebastian Bucheli

Data intake reviewer: Nolan Piloza-Hibbit

Data storage location: <https://archive.ics.uci.edu/ml/datasets/Bank+Marketing>

Tabular data details:

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	5.699 MB

Proposed Approach:

Validation:

- Schema validation using YAML file
- Using Excel functions, remove any duplicate rows

Assumptions:

- The data is imbalanced, showing more negative y values than positive, at a rate of almost 8 times greater.