

Data Intake Report

Name: Bank Term-Deposit Purchase Prediction

Report Date: 19/11/2024

Internship Batch: LISUM38

Version: 1.0

Data Intake By: Suyog Nemade

Data Intake Reviewer: Suyog Nemade

Data Storage Location: <https://github.com/Suyog005/Data-Glacier>

Tabular data details:

Bank-Full.csv:

Total number of observations	45211
Total number of files	1
Total number of features	17
Base format of the file	.csv
Size of the data	4547 KB

Bank-Full-Additional.csv:

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	4814 KB

Proposed Approach:

Since there are no unique incidence qualifiers (such as name or customer ID), the data is assumed to contain no duplicates or a nearly negligible number of duplicates. There is no missing or null data in the set. However, since much of the data was generated via survey response from customers, some customers declined to give information due to privacy concerns. For the initial data intake, we are treating these non-responses as their own separate category (such as "Unknown" or "None"). Whether these need to be replaced with something more suitable (such as mean, mode, or dropped all together), will be decided during the data cleaning phase.