

# Data Intake Report

**Name:** Data Science: Bank Marketing (Campaign) -- Group Project

**Report Date:** 20 December 2024

**Internship Batch:** LISUM39

**Data Intake By:** Neural Nomad

**Data Intake Reviewer:** Nitish Jena

**Data Storage Location:** <https://github.com/nitishjena/Week-7/blob/main/bank-additional-full.csv>

## Tabular data details:

bank-additional.csv:

Total number of observations	4119
Total number of files	1
Total number of features	21
Base format of the file	.csv
Size of the data	570 KB

bank-additional-full.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	5.56 MB

## Proposed Approach:

After carefully reviewing the dataset, we found that it does not contain unique identifiers like customer IDs or names. Because of this, we assume that duplicate entries are either nonexistent or minimal. Additionally, there are no missing or null values in the dataset.

However, since much of the data comes from customer surveys, some individuals chose not to provide certain information due to privacy concerns. Instead of leaving these responses blank, we are initially categorizing them as "**Unknown**" or "**None**" to maintain the dataset's structure.

As we move forward with **data cleaning**, we will determine the best way to handle these non-responses. This could mean filling in missing values using statistical methods like the **mean or mode**, or even removing them if they don't add value to the analysis. Our goal is to ensure that the dataset remains reliable and well-prepared for further processing and insights.