

Data Intake Report

Report date: 19 September 2023

Internship Batch: LISUM24

Version: 1.0

Data intake by: Riwaj Neupane

Data intake reviewer: Data Glacier

Data storage location: <https://github.com/Riwaj22/Data-Glacier/tree/main/Week%207>

Tabular data details:

Bank-Full.csv

| | |
|------------------------------|-----------|
| Total Number of Observations | 45210 |
| Total Number of Files | 1 |
| Number of Features | 17 |
| Format of the file | csv |
| Size of the file | 29.2 MB |

Bank-Full-Additional.csv

| | |
|------------------------------|---------|
| Total Number of Observations | 41188 |
| Total Number of Files | 1 |
| Number of Features | 21 |
| Format of the file | csv |
| Size of the file | 30.3 MB |

Proposed Approach

- No missing data in both of the dataset
- No unique identifier like name or customer ID to identify a customer in both the dataset
- Bank-Full.csv contains no duplicated
- Bank-Full-Additional.csv contains 12 duplicated records which is negligible
- contact column contains a category as 'Unknown' which is vague, so replacement of it with more meaningful value is to be done in data cleaning stage like: mean, mode and median