Data Intake Report

Name: Week 7 Deliverables

Report date: March 6, 2023

Internship Batch: LISUM18

Version: 1.0

Data intake by: Zakaria Arshad

Data intake reviewer: Data Glacier

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

https://github.com/Zakaria-Arshad/Data-Science-Intern-at-Data-Glacier/tree/main/Week%207

**Tabular data details:** Bank-Additional-Full

|  |  |
| --- | --- |
| **Total number of observations** | 41188 |
| **Total number of files** | 1 |
| **Total number of features** | 20 |
| **Base format of the file** | .csv |
| **Size of the data** | 5699 KB |

**Tabular data details:** Bank-Additional

|  |  |
| --- | --- |
| **Total number of observations** | 4119 |
| **Total number of files** | 1 |
| **Total number of features** | 20 |
| **Base format of the file** | .csv |
| **Size of the data** | 571 KB |

**Tabular data details:** Bank-Full

|  |  |
| --- | --- |
| **Total number of observations** | 41188 |
| **Total number of files** | 1 |
| **Total number of features** | 17 |
| **Base format of the file** | .csv |
| **Size of the data** | 503 KB |

**Tabular data details:** Bank

|  |  |
| --- | --- |
| **Total number of observations** | 419 |
| **Total number of files** | 1 |
| **Total number of features** | 17 |
| **Base format of the file** | .csv |
| **Size of the data** | 62 KB |

**Proposed Approach:**

No null/missing values in dataset that needed to be replaced. Need to focus on dataset imbalance. Will attempt oversampling minority class or undersampling majority class.