

# Data Intake Report

Name: EDA.ipynb

Report date: 16/09/2023

Internship Batch: LISUM25

Version: 1.0

Data intake by: Boris David Baros

Data intake reviewer: Boris David Baros

Data storage location: Same repository as project and intake report

## Tabular data details:

Cab\_Data.csv

|                                     |           |
|-------------------------------------|-----------|
| <b>Total number of observations</b> | 359392    |
| <b>Total number of files</b>        | 4         |
| <b>Total number of features</b>     | 7         |
| <b>Base format of the file</b>      | .csv      |
| <b>Size of the data</b>             | 20,633 KB |

City.csv

|                                     |      |
|-------------------------------------|------|
| <b>Total number of observations</b> | 20   |
| <b>Total number of files</b>        | 4    |
| <b>Total number of features</b>     | 3    |
| <b>Base format of the file</b>      | .csv |
| <b>Size of the data</b>             | 1 KB |

Customer\_ID.csv

|                                     |          |
|-------------------------------------|----------|
| <b>Total number of observations</b> | 49171    |
| <b>Total number of files</b>        | 4        |
| <b>Total number of features</b>     | 4        |
| <b>Base format of the file</b>      | .csv     |
| <b>Size of the data</b>             | 1,027 KB |

Transaction\_ID.csv

|                                     |          |
|-------------------------------------|----------|
| <b>Total number of observations</b> | 440098   |
| <b>Total number of files</b>        | 4        |
| <b>Total number of features</b>     | 3        |
| <b>Base format of the file</b>      | .csv     |
| <b>Size of the data</b>             | 8,788 KB |

## Proposed Approach:

- We have made the assumption that the given data observation period in the brief is incorrect.
- All removal of duplicate data is done within our project.