Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 03/05/2021

Internship Batch: LISP01

Version: 1.0

Data intake by: Walquer X. Valles Ruiz

Data intake reviewer:

Data storage location: https://github.com/DataGlacier/DataSets

**Tabular data details:**

1. Cab\_Data

|  |  |
| --- | --- |
| **Total number of observations** | 359,392 |
| **Total number of files** | 1 |
| **Total number of features** | 7 |
| **Base format of the file** | .csv |
| **Size of the data** | 20.1 MB |

1. City

|  |  |
| --- | --- |
| **Total number of observations** | 20 |
| **Total number of files** | 1 |
| **Total number of features** | 3 |
| **Base format of the file** | .csv |
| **Size of the data** | 759 bytes |

1. Customer\_ID

|  |  |
| --- | --- |
| **Total number of observations** | 49171 |
| **Total number of files** | 1 |
| **Total number of features** | 4 |
| **Base format of the file** | .csv |
| **Size of the data** | 1.00 MB |

1. Transaction\_ID

|  |  |
| --- | --- |
| **Total number of observations** | 440098 |
| **Total number of files** | 1 |
| **Total number of features** | 3 |
| **Base format of the file** | .csv |
| **Size of the data** | 8.58 MB |

* Data deduplication approach: Transaction\_ID feature will be used as a mean to secure the uniqueness of the records for each trip in the dataset. So, it will be used as an index in data joining process.