Data Intake Report

**Name**: G2M insight for cab investment firm

Report date: 13-07-2024

Internship Batch: LISUM35

Version: 1.0

Data intake by: OBIDA ALHAMOUD

Data intake reviewer: Data Glacier

Data storage location: github.com/Ob1ida/G2M-insight-for-Cab-Investment

**Tabular data details: Cab\_Data**

|  |  |
| --- | --- |
| **Total number of observations** | 359392 |
| **Total number of files** | 1 |
| **Total number of features** | 7 |
| **Base format of the file** | csv |
| **Size of the data** | 20 MB |

**Tabular data details: 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** | 8.5 MB |

**Tabular data details: 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 MB |

**Tabular data details: 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** | 4 KB |

**Proposed Approach:**

* There is no missing values in all datasets

**Assumptions for Data Quality Analysis:**

* Assume that each `Transaction ID` is unique and correctly represents a single transaction.
* Assume that each `Customer ID` is unique and consistently referenced across all datasets.
* Assume that the `City` dataset correctly maps city names to their respective attributes (e.g., population, users).
* Assume that there are no significant outliers unless identified during the initial inspection.