Data Intake Report

Name: G2M insights for Cab Investment firm

Report date: 11-04-2024

Internship Batch: LISUM32

Version:1.0

Data intake by:Bisma Azeem

Data storage location: [Github](https://github.com/bisma-azeem-13/DataGlacier_Corhort_LISUM32/tree/main/Week-2%2B3)

**Tabular data details for 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** | MB |

**Tabular data details for 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** | MB |

**Tabular data details for 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** | Bytes |

**Tabular data details for 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** | MB |

**Tabular data details for Holidays\_Data:**

|  |  |
| --- | --- |
| **Total number of observations** | 342 |
| **Total number of files** | 1 |
| **Total number of features** | 6 |
| **Base format of the file** | .csv |
| **Size of the data** | KB |

**Proposed Approach:**

* Deduplication validation: I checked all the tables for duplicated rows and ensured validatity of datasets.
* Data Quality Assurance: I checked all the tables for NA values, unusual range of values and data types and ensure quality of datasets.