Data Intake Report

Name: G2M insight for Cab Investment firm Report date: 13/08/2024

Internship Batch: LISUM36 Version:1.0

Data intake by: Sreedhar RONGALA

intake reviewer:

Data storage location: [**https://github.com/sreedharsiddhu/Data-Glacier/tree/main/week%202**](https://github.com/sreedharsiddhu/Data-Glacier/tree/main/week%202)

**Tabular data details:**

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

**Note: Replicate same table with file name if you have more than one file.**

|  |  |
| --- | --- |
| **Proposed Approach** | **Assumptions** |
| Cleaned and standardized customer demographic data (e.g., gender, age, income). | Assumed that demographic data provided was accurate and complete. |
| Categorized customers into different age groups and income categories. | Assumed that categorization did not oversimplify the analysis. |
| Identified and flagged repeat customers in the dataset. | Assumed that customer IDs were consistently used across transactions. |
| Addressed missing values by imputing with median or mode. | Assumed that missing values were randomly distributed. |