

Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 04-May-2023

Internship Batch: LISUM21

Version: 1.0

Data intake by: Nilesh Rathi

Data intake reviewer:

Data storage location: <https://github.com/nileshrathi99/DataGlacier/tree/main/week2/>

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	22.6MB

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	4KB

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	1.1MB

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	9MB

Proposed Approach:

- There are no missing values in any of the data sets.
- Assumptions made:
 - Complete: since data has no missing values.
 - Consistent: data is uniform in its format, units of measurement and data types.
 - Relevant: data is totally related to the research question being studied.
 - Unbiased: data is not influenced by any personal or external factors