

Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 14 June 2023

Internship Batch: LISUM22

Version:<1.0>

Data intake by: Maria Contractor

Data intake reviewer: Data Glacier

Data storage location: <https://github.com/mcont55/Data-Glacier-Intenship/tree/main/Week%202>

Tabular data details: Cab_Data.csv

Total number of observations	359393
Total number of files	1
Total number of features	7
Base format of the file	.csv
Size of the data	20.1 MB

Tabular data details: City.csv

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

Tabular data details: Customer_ID.csv

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

Tabular data details: Transaction_ID.csv

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

Proposed Approach:

- There are no missing or null values in any of the four datasets
- I propose to analyze the data through merging and comparing the datasets with each other to create visualizations
- I plan to use Python and Jupyter Notebook