

Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 8th May 2024

Internship Batch: LISUM33

Version: 1.0

Data intake by: Rakshith Nagaraju

Data intake reviewer: -

Data storage location: <https://github.com/DataGlacier/DataSets>

Cab_Data:

Total number of observations	359392
Total number of files	-
Total number of features	7
Base format of the file	.csv
Size of the data	20,633 KB

Customer_ID:

Total number of observations	49171
Total number of files	-
Total number of features	4
Base format of the file	.csv
Size of the data	1027 KB

Transaction_ID:

Total number of observations	440098
Total number of files	-
Total number of features	3
Base format of the file	.csv
Size of the data	8788 KB

City:

Total number of observations	20
Total number of files	-
Total number of features	3
Base format of the file	.csv
Size of the data	1 KB

Proposed Approach:

- Firstly, check Head and Tail of the dataset.
- Check shape of the data.
- Check the 'd-types' to see if anything needs conversion.
- Check for any null values in the data that might need replacement or deletion.
- Check for duplicates in the data that might skew the results.
- Get the overall description and information of the dataset.
- Begin with visualization.