

Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 04/13/2025

Internship Batch LISUM44

Version: <1.0>

Data intake Thanh Tran

Data intake reviewer

Data storage location: <https://github.com/thanizmy/VC.git>

Tabular data details:

Cab_Data.csv

Total number of observations	359392
Total number of files	
Total number of features	7
Base format of the file	.csv
Size of the data	20,663KB

City.csv

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

Customer_ID.csv

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

Transaction_ID.csv

Total number of observations	440098
Total number of files	
Total number of features	3
Base format of the file	.csv
Size of the data	8,788KB

Proposed Approach:

- The dataset contains unique identifiers in each dataset
- Null or missing values in key fields are ignored
- There are no duplicated data points in the dataset.
- No external data sources will be used during statistical analysis.

