

Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: Nov 14th 2022

Internship Batch: LISUM15

Version: 1

Data intake by: Data Glacier Virtual Internship

Data storage location: <https://github.com/Edward-Guangyan-Huang/Week-2-deliverables>

Tabular data details:

Cab_data.csv

Total number of observations	359492
Total number of files	1
Total number of features	7
Base format of the file	.csv
Size of the data	20.1 MB

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	1 kb

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 MB

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	8.7 MB

Proposed Approach:

- No outliers of any dataset will be removed, only observations with null values will be removed.