

Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: April 14, 2023

Internship Batch: LISUM20

Version:1.0

Data intake by: Bindu Musham

Data intake reviewer:

Data storage location: <https://github.com/bmusham/G2M-Cab-Project>

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	20.1 MB

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	8.58 MB

Tabular data details: Customer_Data

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

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	4 KB

Proposed Approach:

- There are no missing values in all the datasets.
- There are duplicate values in these datasets.
- I have changed data types of Date of Travel, Users and Population variables.
- Created new columns for analysis.