

Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 14-06-23

Internship Batch: LISUM22

Version:1.0

Data intake by: Himani Aryan

Data intake reviewer:

Data storage location: <https://github.com/himani-ary/week2/blob/main/week2.ipynb>

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	21.2MB

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	759bytes

Customer ID

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	1.1MB

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	9MB

Proposed Approach:

- Mention approach of dedup validation (identification):

All the provided files/datasets were merged for the purpose of analysis.