

Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 21/03/2024

Internship Batch: LISUM31

Version:1.0

Data intake by: Tarusmita Boro

Data intake reviewer:<intern who reviewed the report>

Data storage location:

https://github.com/TaruIndia/G2M/blob/main/G2M_Cab_InvestmentWeek2.ipynb

Tabular data details: Cab_Data

Total number of observations	359392
Total number of files	1
Total number of features	8
Base format of the file	.csv
Size of the data	20.2 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	759 Bytes

Tabular data details: Customer_ID

Total number of observations	49172
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: 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

Proposed Approach:

- No duplicate records.
- All the tables are joined using inner join. Some derived columns generated.
- Assumptions: No outside data used. Considered the data provided was sufficient for the work.