

Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 18.06.2021

Internship Batch: LISUM01

Version:1.0

Data intake by: Thrinesh Duvvuru

Data intake reviewer:<intern who reviewed the report>

Data storage location: <https://github.com/Thrinesh1200/G2M-insight>

Tabular data details:

1.Cab_data

Total number of observations	359392
Total number of files	4
Total number of features	7
Base format of the file	.csv
Size of the data	21.2 MB

2.Customer_ID

Total number of observations	49171
Total number of files	4
Total number of features	4
Base format of the file	.csv
Size of the data	1.05 MB

3.Transaction_ID

Total number of observations	440098
Total number of files	4
Total number of features	3
Base format of the file	.csv
Size of the data	9 MB

4.City

Total number of observations	20
Total number of files	4
Total number of features	3
Base format of the file	.csv
Size of the data	759 B

Proposed Approach:

- All 4 files are in .csv format.
- Data cleaning and Exploratory data analysis has been done in Jupyter Notebook.
- All four files are merged according to their references.
- The feature “Date of Travel” in Cab_data.csv was in Epoch format and has been converted into “Date time formant”
- It is assumed that Time period of data is from 31/01/2016 to 31/12/2018 and analysis has been done to recommend XYZ firm to invest on best cab company.