

Data Intake Report

Name: G2M Cab Investment Report

Report date: March 10, 2023

Internship Batch: LISUM19

Version:<1.0>

Data intake by: Preeti Verma

Data intake reviewer:<intern who reviewed the report>

Data storage location: Github URL: <https://github.com/vermapreeti003/DataSets>

Tabular data details:

File name: Cab_Data.csv

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

File name: 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	759bytes

File name: 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	1.1MB

File name: 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	9MB

Proposed Approach:

- Each of the datasets was read and used in the project
- Duplicate rows and null values were checked using `drop_duplicates` and `isnull`.
- A key assumption made for the analysis is that profits are calculated using profit charged and cost traveled, no other factors were accounted.