

G2M insight for Cab Investment firm Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 23/06/2022

Internship Batch:LISUM10

Version:

Data intake by:Surabhi Mahawar

Data intake reviewer:

Data storage location:

Tabular data details:

CAB DATA

Total number of observations	359392
Total number of files	<Number of files received>
Total number of features	7
Base format of the file	.csv
Size of the data	MB

CITY

Total number of observations	20
Total number of files	<Number of files received>
Total number of features	3
Base format of the file	.csv
Size of the data	KB

TRANSCATION ID

Total number of observations	440098
Total number of files	<Number of files received>
Total number of features	3
Base format of the file	.csv
Size of the data	MB

CUSTOMER ID

Total number of observations	49171
Total number of files	<Number of files received>
Total number of features	4
Base format of the file	.csv
Size of the data	MB

MASTER DATA

Total number of observations	359392
Total number of files	<Number of files received>
Total number of features	14
Base format of the file	.csv

Size of the data	MB
------------------	----

Proposed Approach:

- First understand each table very carefully their features and entries
- Identify the relationship between them
- Relationship described below:
 1. Cabdata.Transcation ID=Transcation.Transcation ID
 2. Customer.customer ID=Transcation ID.Customer ID
 3. Cabdata.city=City.city
- Merge all 4 tables into Master data which contain 359292 rows and 14 columns
- After gaining a good understanding of data,next step is Data Cleaning
- Data Cleaning: remove duplicates,spaces,missing values,null values,commas etc.
- Then perform data analysis
- Also analyze data statiscally