Data Intake Report

Name: G2M Insight for Cab Investment Firm

Report Date: 22/07/2022

Internship Batch: LISUM12

Version: 1.0

Data Intake by: Rohit Sunku

Data Intake Reviewer: Data Glacier

Data Storage Location:

Cab Data Intake Report

Tabular Details: Cab_Data

Total Number of Observations	359392
Total Number of Files	1 (4 in data schema)
Total Number of Variables	7
Total Number of Numeric Features	5
Total Number of Categorical Features	2
Total Number of Null Values	0
Total Number of Duplicate Values	0
Base Format of the Data	.csv
Size of the Data	19.2 MiB

Transaction Data Intake Report

Tabular Details: Transaction_ID

Total Number of Observations	440098
Total Number of Files	1 (4 in data schema)
Total Number of Variables	3
Total Number of Numeric Features	2
Total Number of Categorical Features	1
Total Number of Null Values	0
Total Number of Duplicate Values	0
Base Format of the Data	.csv
Size of the Data	10.1 MiB

City Data Intake Report

Tabular Details: City

Total Number of Observations	20
Total Number of Files	1 (4 in data schema)
Total Number of Variables	3
Total Number of Numeric Features	0
Total Number of Categorical Features	3
Total Number of Null Values	0
Total Number of Duplicate Values	0
Base Format of the Data	.csv
Size of the Data	604.0 B

Customer Data Intake Report

Tabular Details: Customer_Data

Total Number of Observations	49171
Total Number of Files	1 (4 in data schema)
Total Number of Variables	4
Total Number of Numeric Features	3
Total Number of Categorical Features	1
Total Number of Null Values	0
Total Number of Duplicate Values	0
Base Format of the Data	.csv
Size of the Data	1.5 MiB

Proposed Approach

- There are no missing or duplicate values in any of the datasets.
- We merge all the datasets together to create a master data frame.
- The data can be represented in the from of a schema shown here:

Cab Data
- Transaction_ID
- Date of Travel
- Company
- City
- KM Travelled
- Price Charged
- Cost of Trip

City Data
- City
- Population
- Users

Customer Data
- Customer ID
- Gender
- Age
- Income (USD/ Month)

Transaction Data
- Transaction_ID
- Customer ID
- Payment Mode