

# Data Intake Report

Name: G2M insight for Cab Investment Firm

Report date: 13-09-2023

Internship Batch: LISUM25

Version: 0.1

Data intake by: Ishwarya Rajaiya

Data intake reviewer:

Data storage location:

## Tabular data details:

<b>File Name</b>	Cab Data
<b>Total number of observations</b>	359392
<b>Total number of files</b>	
<b>Total number of features</b>	7
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	22.96 MB

<b>File Name</b>	City Data
<b>Total number of observations</b>	20
<b>Total number of files</b>	
<b>Total number of features</b>	3
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	759 B

<b>File Name</b>	Customer ID
<b>Total number of observations</b>	49171
<b>Total number of files</b>	
<b>Total number of features</b>	4
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	1.05 MB

<b>File Name</b>	Transaction ID
<b>Total number of observations</b>	440098
<b>Total number of files</b>	
<b>Total number of features</b>	3
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	9 MB

**Proposed Approach:**  
**Column data types:**

CITY DATA	
Columns	Data type
City	Object
Population	Object
Users	object

CAB DATA	
Columns	Data type
Transaction ID	int64
Date of Travel	Object
Company	object
City	Object
KM Travelled	float64
Price Charged	float64
Cost of Trip	float64

CUSTOMER ID	
Columns	Data type
Customer ID	int64
Age	Object
Gender	int64
Income (USD/Month)	int64

TRANSACTION ID	
Columns	Data type
Transaction ID	int64
Customer ID	int64
Payment mode	Object

**Proposed Approach**

- The dataset has no duplicates, so deduplication approach is not necessary.
- Also, there are no missing values in the dataset indicating complete data.
- Gender data shows the gender of customers, with 'Male' being predominant.
- The data appears to be of good quality, with no apparent errors or inconsistencies.
- Since the city dataset has only 20 observations, we can use it later for visualizations.
- I will begin the data analysis by looking into the below questions which will give enough insight to come to conclusion about which cab company to invest in:
  - ✓ DOES GENDER HAS SIGNIFICANT EFFECT ON DISTANCE OF TRIPS TAKEN BY CUSTOMERS?
  - ✓ DOES CHOICE OF CAB COMPANY AFFECT THE PRICE CHARGED?
  - ✓ DOES INCOME AFFECT THE NUMBER OF CUSTOMERS?
  - ✓ MARKET SHARE OF CAB COMPANIES IN EACH CITY
  - ✓ MARGIN PROPORTIONALITY RELATIONSHIP WITH NUMBER OF CUSTOMERS
  - ✓ MAXIMUM CAB USERS AT PARTICULAR TIME PERIOD