

# Data Intake Report

Name: G2M Insight for Cab Investment

Report date: 07/14/2023

Internship Batch: LISUM23

Version:

Data intake by: Manisha Gayatri Damera

Data intake reviewer:

Data storage location:

## Tabular data details:

### 1. Cab Data

<b>Total number of observations</b>	359392
<b>Total number of files</b>	1
<b>Total number of features</b>	8
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	21.2 MB

### 2. City

<b>Total number of observations</b>	20
<b>Total number of files</b>	1
<b>Total number of features</b>	3
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	759 B

### 3. Customer\_ID

<b>Total number of observations</b>	49171
<b>Total number of files</b>	1
<b>Total number of features</b>	4
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	1.05 MB

### 4. Transaction\_ID

<b>Total number of observations</b>	440098
<b>Total number of files</b>	1
<b>Total number of features</b>	4
<b>Base format of the file</b>	.csv

<b>Size of the data</b>	19 MB
-------------------------	-------

**Note:** Replicate same table with file name if you have more than one file.

**Proposed Approach:**

- **No null values or duplicates**
- **A new profit margin is added assuming profit as the difference between price charged and cost of trip**
- **Correlation among different features is found after the datasets are merged**
- **Finally proposing and testing hypothesis (what we assume about the findings in the dataset)**