

# Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 10-July-2022

Internship Batch: LISUM11

Version: 1.0

Data intake by: Mohamed Elmorsy

Data intake reviewer: N/A

Data storage location: <https://github.com/elmorsym1/G2M-EDA/data>

## Tabular data details:

<b>Filename</b>	Cab_Data
<b>Total number of observations</b>	359392
<b>Total number of files</b>	4
<b>Total number of features</b>	7
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	20 MB

<b>Filename</b>	Transaction_ID
<b>Total number of observations</b>	440098
<b>Total number of files</b>	4
<b>Total number of features</b>	3
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	9 MB

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

<b>Filename</b>	City
<b>Total number of observations</b>	20
<b>Total number of files</b>	4
<b>Total number of features</b>	3
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	1 KB

## Proposed Approach:

- Python codes will be used to check and validate data duplication
- Python codes will be used to perform EDA over this dataset