

# Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: April 14, 2023

Internship Batch: LISUM20

Version:1.0

Data intake by: Bindu Musham

Data intake reviewer:

Data storage location: <https://github.com/bmusham/G2M-Cab-Project>

## Tabular data details: Cab\_Data

<b>Total number of observations</b>	359392
<b>Total number of files</b>	1
<b>Total number of features</b>	7
<b>Base format of the file</b>	csv
<b>Size of the data</b>	20.1 MB

## Tabular data details: Transaction\_ID

<b>Total number of observations</b>	440098
<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>	8.58 MB

## Tabular data details: Customer\_Data

<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 MB

## Tabular data details: 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>	4 KB

**Proposed Approach:**

- There are no missing values in all the datasets.
- There are no duplicate values in these datasets.
- I have changed data types of Date of Travel, Users and Population variables.
- Created new columns for analysis.