

# Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 05-12-2023

Internship Batch: LISUM21

Version:<1.0>

Data intake by: Mike Wang

Data intake reviewer:

Data storage location: <https://github.com/Momowangg/EDA>

## 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>	19.8 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>	608 bytes

## Tabular data details:

<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>	2 MB

## Tabular data details:

<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>	11 MB

**Proposed Approach:**

- Checked null and duplicated values and found no such data
- Assumptions were made about data collection that data are collected randomly from the same population with equal variance.