

# Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 14 Jan 2025

Internship Batch: LISUM41

Version: 1.0

Data intake by: Jameskumar Chidguwar

Data intake reviewer: Data Glacier

Data storage location: <https://github.com/chjameskumar/Week-2/tree/main/DataSet>

## Cab data details:

<b>Total number of observations</b>	359393
<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>	20.66 MB

## City data details:

<b>Total number of observations</b>	21
<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>	1 KB

## Customer ID data details:

<b>Total number of observations</b>	49172
<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>	1.02 MB

## Transaction ID data details:

<b>Total number of observations</b>	440099
<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>	8.78 MB

**Proposed Approach:**

- Importing data and preprocessing (cleaning, checking missing values, duplicates, etc).
- EDA: Univariate Analysis, Bivariate Analysis
- Hypothesis Testing