

# Data Intake Report

Name: G2M Insight for Cab Investment Firm

Report date: 13/07/2024

Internship Batch: LISUM35

Version:1.0

Data intake by: Sainad Reddy Naini

Data intake reviewer:

Data storage location:

## 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:**

1. Exploratory Data Analysis (EDA)
  - Checked for null values, missing data, data types, and duplicates.
2. Data Integration
  - Merged multiple datasets using appropriate foreign keys (city, customerid, transactionid).
3. Feature Engineering
  - Enhanced the dataset by adding relevant columns to facilitate visualizations and trend analysis.
4. Hypothesis Testing
  - Conducted hypothesis testing to validate assumptions and draw statistical conclusions.