

# Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 04/13/2023

Internship Batch: LISUM20: 30 March – 30 June 2023

Version: <1.0>

Data intake by: Pooja Honneshwari Ravi

Data intake reviewer: Pooja Honneshwari Ravi

Data storage location: GitHub

## Tabular data details:

Cab\_Data.csv

<b>Total number of observations</b>	359393
<b>Total number of files</b>	1/4
<b>Total number of features</b>	7
<b>Base format of the file</b>	csv
<b>Size of the data</b>	21.2 MB

City.csv

<b>Total number of observations</b>	21
<b>Total number of files</b>	2/4
<b>Total number of features</b>	3
<b>Base format of the file</b>	csv
<b>Size of the data</b>	759 Bytes

Customer\_ID.csv

<b>Total number of observations</b>	49172
<b>Total number of files</b>	3/4
<b>Total number of features</b>	4
<b>Base format of the file</b>	csv
<b>Size of the data</b>	1.1 MB

Transaction\_ID.csv

<b>Total number of observations</b>	440099
<b>Total number of files</b>	4/4
<b>Total number of features</b>	3
<b>Base format of the file</b>	csv
<b>Size of the data</b>	9 MB

### **Proposed Approach:**

- **We collected 4 datasets namely: Cab\_Data, Customer\_ID, Transaction\_ID, and City.**
- **We performed an EDA to understand the relations between the data and to determine which company is a better choice for investment opportunity for XYZ.**
- **To better investigate, we generated few hypotheses:**
  - **Which cab is more popular in different cities?**
  - **How is the cab usage according to age?**
  - **Which cab will people with different income choose?**
  - **Which cab has travelled more distance?**
  - **What is the cost of trip of each cab?**
  - **Which cab has more profited generated?**
  - **What is the number of car users in different cities?**