

# Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 18 April 2024

Internship Batch: LISUM32

Data intake by: Darshan Dhanani

Data storage location: <https://github.com/darshan-1611-dev/VC/tree/master/week%202/Dataset>

## Tabular data details:

<b>File Name</b>	Cab_Data
<b>Total number of observations</b>	359392
<b>Total number of features</b>	7
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	20.66 MB

<b>File Name</b>	Transaction_ID
<b>Total number of observations</b>	440098
<b>Total number of features</b>	3
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	8.79 MB

<b>File Name</b>	customer_ID_df
<b>Total number of observations</b>	49171
<b>Total number of features</b>	4
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	1.02 MB

<b>File Name</b>	city_df
<b>Total number of observations</b>	20
<b>Total number of features</b>	3
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	1 KB

## **Proposed Approach:**

### **1. Cab\_Data**

#### **Change Description:**

Converted "Company" column values to numerical representation:  
Pink Cab (0), Yellow Cab (1).

#### **Reason for Change:**

To enable numerical analysis and modeling for improved efficiency and comparability.

#### **Impact Analysis:**

Updated "Company" column values to numerical codes.  
No structural changes to the dataset.  
Ensured data consistency and integrity.