

# Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 01/11/2024

Internship Batch: LISUM29

Version: 1.0

Data intake by: Shengyu (Felix) Huang

Data intake reviewer:

Data storage location:

## **Cab\_Data.csv:**

<b>Total number of observations</b>	359,392
<b>Total number of files</b>	
<b>Total number of features</b>	7
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	20,663KB

## **City.csv:**

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

## **Customer\_ID.csv:**

<b>Total number of observations</b>	49,171
<b>Total number of files</b>	
<b>Total number of features</b>	4
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	1,027KB

## **Transaction\_ID.csv:**

<b>Total number of observations</b>	440,098
<b>Total number of files</b>	
<b>Total number of features</b>	3
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	8,788KB

## **Proposed Approach:**

- Date should be converted in Cab\_Data from an excel format to Python datetime format.
- Transaction ID and Customer ID are primary that can use to merge table for analysis.