

# Data Intake Report

**Name:** G2M insight for Cab Investment firm

**Report date:** 14/09/2022

**Internship Batch:** LISUM13

**Version:** 1.0

**Data intake by:** Chenyu Wang

**Data intake reviewer:** Data Glacier

**Data storage location:** [https://github.com/chenyu-wang55/Data\\_Scientist\\_Intern\\_Data\\_Glacier/tree/main/Week2/Datasets](https://github.com/chenyu-wang55/Data_Scientist_Intern_Data_Glacier/tree/main/Week2/Datasets)

## Tabular data details: Cab Data

Total number of observations	359392
Total number of files	1
Total number of features	7
Base format of the file	csv
Size of the data	21.2 MB

## Tabular data details: City

Total number of observations	20
Total number of files	1
Total number of features	3
Base format of the file	csv
Size of the data	4 KB

## Tabular data details: Customer ID

Total number of observations	49171
Total number of files	1
Total number of features	4
Base format of the file	csv
Size of the data	1.1 MB

## Tabular data details: Transaction ID

Total number of observations	440098
Total number of files	1
Total number of features	3
Base format of the file	csv
Size of the data	9 MB

## Proposed Approach:

- There are no missing values in all four datasets.
- There are no duplicated rows in all four datasets.