

# Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: June 8, 2024

Internship Batch: LISUM34

Version: 1.0

Data intake by: Awwab Ahmed

Data intake reviewer: Data Glacier

Data storage location: <https://github.com/DataGlacier/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	20.2 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	759 Bytes

## 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 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	8.59 MB

**Proposed Approach:**

Approach of dedup validation (identification):

There are no duplicate or missing values in any of the datasets.

Assumptions:

The time period of the data is from 31/01/2016 to 31/12/2018.