

# Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 11/16/2023

Internship Batch: LISUM25

Version:1.0

Data intake by: Farnaz Amiri

Data intake reviewer: Data Glacier

Data storage location: [DG/Week2/data at main · famiri60/DG \(github.com\)](#)

## Tabular data details: Cab\_Data

<b>Total number of observations</b>	359392
<b>Total number of files</b>	1
<b>Total number of features</b>	7
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	20.663 MB

### Proposed Approach:

- No Duplicated row, No missing values
- Feature “Date of Travel” type is not consistent with expectations, so we need to convert it to meaningful date objects.

## Tabular data details: Customer\_ID

<b>Total number of observations</b>	49171
<b>Total number of files</b>	1
<b>Total number of features</b>	4
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	1.027 MB

### Proposed Approach:

- No Duplicated row, No missing values
- Data adheres to the expected format and structure.

## Tabular data details: Transaction\_ID

<b>Total number of observations</b>	440098
<b>Total number of files</b>	1
<b>Total number of features</b>	3
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	8.788 MB

**Proposed Approach:**

- No Duplicated row, No missing values
- Data adheres to the expected format and structure.

**Tabular data details: City**

<b>Total number of observations</b>	20
<b>Total number of files</b>	1
<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:**

- No Duplicated row, No missing values
- Features “Population” and “Users” data types are not consistent with expectations, so we need to convert them to integer.