

# Data Intake Report

Name: G2M

Report date: 08/12/2022

Internship Batch: LISUM16

Version: 1.0

Data intake by: Gabija Vaisvilaite

Data intake reviewer: N/A

Data storage location: <https://github.com/DataGlacier/DataSets>

## Tabular data details:

<b>Total number of observations</b>	359,392
<b>Total number of files</b>	1: Cab_Data.csv
<b>Total number of features</b>	7
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	20,633KB

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

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

<b>Total number of observations</b>	440,098
<b>Total number of files</b>	1: Transaction_ID.csv
<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:

- Identify null values and how clean is the data: there are no null values and the data is clean. The “Date of Travel” column has incorrect type of date values in the Cab\_Data.csv file.
- Assume the data is correct, data will be used to perform explanatory data analysis.