

# Data Intake Report

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

Report date: 21 July 2022

Internship Batch: **LISUM11**

Version: 1.0

Data intake by: Hussein Sabrawi

Data intake reviewer:

Data storage location: <https://github.com/sabrawihussein/DataGlacier/tree/main/Week2>

## Tabular data details: Cab\_Data

<b>Total number of observations</b>	359393
<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>	21.2 MB

## Tabular data details: City

<b>Total number of observations</b>	21
<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>	759 bytes

## Tabular data details: Customer\_ID

<b>Total number of observations</b>	49172
<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>	1.1 MB

## Tabular data details: Transaction\_ID

<b>Total number of observations</b>	440099
<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>	9.0 MB

## Proposed Approach:

- Merged all data sets into one main data set
- All data was clean, there was no need for data cleansing, only renamed a few columns

