

# Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 8/3/2021

Internship Batch: LICAN01

Version:<1.0>

Data intake by: Omar Pablo Jazouli Martínez

Data intake reviewer:

Data storage location: <<https://github.com/omarjaz/DG-Assignment>>

## Tabular data details:

<b>File name</b>	Cabs.csv
<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>	21.2 MB

<b>File name</b>	City.csv
<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>	4 KB

<b>File name</b>	Customer_ID.csv
<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>	1.1 MB

<b>File name</b>	Transaction_ID.csv
<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>	9 MB

### **Assumptions:**

Cost\_of\_Trip feature is considered as “all expenses involved in the ride” (including taxes and the part earned by the driver)

Users feature of City dataset is treated as “total number of cab users in the city at the end of 2018”(taking into account the users of all cab companies in the US)