

# Data Intake Report

Name: G2M Insight for Cab Investment Firm

Report date: Wednesday February 17, 2021

Internship Batch:<Enter your batch code from Canvas course>

Version:<1.0>

Data intake by: Laith Adi

Data intake reviewer:<intern who reviewed the report>

Data storage location: <location URL eg: github, cloud>

## Tabular data details

Note: The information on the tables are on the original datasets. Before any data processing (joining of tables, filtering of attributes/features, etc.) was done.

Table: **Cab\_Data.csv**

<b>Total number of observations</b>	359,392
<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>	19.2+ MB

Table: **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>	608.0+ bytes

Table: **Customer\_ID.csv**

<b>Total number of observations</b>	49,171
<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>	1.5+ MB

Table: **Transaction\_ID.csv**

<b>Total number of observations</b>	440,098
<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>	10.1+ MB

Table: **Income.csv**

<b>Total number of observations</b>	39,030
<b>Total number of files</b>	1
<b>Total number of features</b>	80
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	23.8+ MB

### Attributes for Each Table

<b>Table</b>	<b>Attributes</b>
Cab_Data.csv (cab_data)	Transaction ID Date of Travel Company City KM Travelled Price Charged Cost of Trip
City.csv (city)	City Population Users
Customer_ID.csv (customer)	Customer ID Gender Age Income (USD/Month)
Transaction_ID.csv (transaction)	Transaction ID Customer ID Payment_Mode

Income.csv (income)	State state_ab city family_mean *there much more but we do not care about the others*
------------------------	--

## **Final Data**

Table: **final\_data**

<b>Total number of observations</b>	359,392
<b>Total number of files</b>	--
<b>Total number of features</b>	16
<b>Base format of the file</b>	Stored as Pandas Dataframes
<b>Size of the data</b>	--