

Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 21 July

Internship Batch: 30 June-30 September 2022

Version:0.2

Data intake by: Alejandro Riol Triviño

Data intake reviewer:

Data storage location: <https://github.com/ART1993/Presentation-Jupyter-Canvas>

Tabular data details:

File name	City.csv
Total number of observations	20
Total number of files	
Total number of features	3
Base format of the file	.csv
Size of the data	759 bytes

File name	Cab_Data.csv
Total number of observations	35932
Total number of files	
Total number of features	8
Base format of the file	.csv
Size of the data	21.2 MB

File name	Customer_ID
Total number of observations	49171
Total number of files	
Total number of features	4
Base format of the file	.csv
Size of the data	1.1 MB

File name	Transaction_ID.csv
Total number of observations	440098
Total number of files	
Total number of features	3
Base format of the file	.csv
Size of the data	9,0 MB

City:

City: Here simply mentions the city of the user where the travel takes place.

Users: total cab users in the city

Population: Total population in the City

Cab_Data:

Transaction_ID: The identity of the transaction

Date of travel: Date where it takes place, written in serial excel.

Company: in which company it was realized

City: the same as the previous case

Km traveled: the distance traversed in the trip

Price charged: How much the client paid

Cost trip: How much it cost the company the travel

Customer ID:

Here we receive info about the client: gender age and salary.

Transaction_ID:

Here we can see who paid for each travel and how it paid for it.

Unique rows.

Found the unique rows using the method `drop_duplicates` from pandas.