

Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 10/13/22

Internship Batch: LISUM14

Version:<1.0>

Data intake by: Susana Munguia Hernandez

Data intake reviewer:

Data storage location: Github

Tabular data details: Cab_Data.csv

Total number of observations	359,392
Total number of files	1
Total number of features	7
Base format of the file	.csv
Size of the data	21.2 MB

Tabular data details: Customer_ID.csv

Total number of observations	49,171
Total number of files	1
Total number of features	4
Base format of the file	.csv
Size of the data	1.05 MD

Tabular data details: Transaction_ID.csv

Total number of observations	440,098
Total number of files	1
Total number of features	3
Base format of the file	.csv
Size of the data	9 MD

Tabular data details: City.csv

Total number of observations	20
Total number of files	1
Total number of features	3
Base format of the file	.csv
Size of the data	759 B

Proposed :

I merged Cab Data, Transaction ID, and Customer ID. This is because the cab data Transaction ID which could be matched up with additional information from Transaction ID. Then, Transaction ID had the feature Customer ID, which could be matched up with Customer ID.

I left City alone, because it would be repetitive information in the big file.