

Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 12 October 2022

Internship Batch: LISUM14

Version: 1.0

Data intake by: Jiahao Yu

Data intake reviewer:

Data storage location:

Tabular data details:

Cab_Data.csv

Total number of observations	359392
Total number of files	
Total number of features	7
Base format of the file	.csv
Size of the data	20.2 MB

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

Customer_ID.csv

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 MB

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	8.58MB

Note: Replicate same table with file name if you have more than one file.

Proposed Approach:

- Cab_Data.csv is the main data set.
- The data sets can be connected (merged) by:
 - Cab_Data.csv and City.csv with the column "City".
 - Cab_Data.csv and Transaction_ID.csv with the column "Transaction ID".
 - Transaction_ID.csv and Customer_ID.csv with the column "Customer ID".