

Data Intake Report

Name: G2M Insights for Cab Investment Firm

Report date: 06/12/2023

Internship Batch: LISUM22

Version: 1.0

Data intake by: Shuen (Kerr) Tan

Data intake reviewer: N/A

Data storage location: <https://github.com/Kerrk07/G2M-Insights/tree/main/data>

Tabular data details: Cab_Data

Total number of observations	359392
Total number of files	1
Total number of features	7
Base format of the file	csv
Size of the data	19.2+ MB

Tabular data details: Customer_ID

Total number of observations	49171
Total number of files	1
Total number of features	4
Base format of the file	csv
Size of the data	1.5+ MB

Tabular data details: Transaction_ID

Total number of observations	440098
Total number of files	1
Total number of features	3
Base format of the file	csv
Size of the data	10.1+ MB

Tabular data details: City

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	608.0+ bytes

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

Proposed Approach:

- One issue concerning data quality and trustworthiness is that 4 datasets utilized in this case have no missing values.
- Date of Travel in the Cab_Data.csv needs to be modified in the correct format of DateTime.
- Above datasets are utilized in a way of exploratory data analysis to understand the patterns, distributions, and issues within the given datasets, helping to generate business insights for the Cab industry.