

Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 14 April 2024

Internship Batch: LISUM32

Version:1.0

Data intake by: Atshaya Srinivasan

Data intake reviewer:

Data storage location: https://github.com/AtshayaSrinivasan47/DataGlacier_Project2

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: 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

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

Proposed Approach:

- There are no null values in all 4 tables.
- There are few outliers in both pink and yellow cab Price_Charged feature, so I assumed that these outliers are due to use of high-end cars.