

Data Intake Report

Name: G2M insight for Cab Investment Firm

Report date: 11-11-2023

Internship Batch: LISUM27

Version: 1.0

Data intake by: Nrusimha Saraswati Sai Teja Jampani

Data intake reviewer: Data Glacier

Data storage location: local Db

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	21.5 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	4 KB

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 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	8.58 MB

Tabular data details: US Holiday Dates

Total number of observations	440098
Total number of files	1
Total number of features	6
Base format of the file	csv
Size of the data	15 KB

Proposed Approach:

- From the preprocessing I found no duplicate values
- The data is factual with no NaN data after joining