

Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 21-10-2023

DataGlacier Internship: LISUM26

Data intake by: Okeoma Ihunwo

Data intake reviewer:

Data storage location: <https://github.com/Okeoma/Week-2>

Cab_data Dataset details:

Total number of observations	359392
Total number of files	1
Total number of features	7
Base format of the file	Comma-separated values (.csv)
Size of the data	20.1 MB

Transaction_ID Dataset details:

Total number of observations	440098
Total number of files	1
Total number of features	3
Base format of the file	Comma-separated values (.csv)
Size of the data	8.58 MB

City_data Dataset details

Total number of observations	20
Total number of files	1
Total number of features	3
Base format of the file	Comma-separated values (.csv)
Size of the data	759 bytes

Customer_ID Dataset details

Total number of observations	49171
Total number of files	1
Total number of features	4
Base format of the file	Comma-separated values (.csv)
Size of the data	1.00 MB

Proposed Data Preparation Approach: (Data Understanding and Data Cleaning)

- There are no null values in all the 4 datasets.
- There are no duplicated rows.