

Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 20-June-2021

Internship Batch: LISMU01

Version: 1.0

Data intake by: Nkululeko Freedom Mqadi

Data intake reviewer: Nkululeko Freedom Mqadi

Data storage location: <https://github.com/Nkululeko353/Week-2-Submission-G2M-insight-for-Cab-Investment-firm>

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	20.1 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: 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	1.00 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	759 bytes

Proposed Approach:

- We need to check for duplicates in the data sets:

Cab_Data:

We find no duplicated rows.

Transaction_ID:

We found no duplicated rows.

Customer_ID:

We found no duplicated rows.

City:

We found no duplicated rows.

- We also need to check for null values.
We found no null values in any of these datasets.

