Data Intake Report

Name: G2M insight for Cab Investment firm
Report date: 14/09/2024
Internship Batch LISUM37
Version:<1.0>
Data intake by: Mark Maina
Data intake reviewer: Mark Maina
Data storage location:

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: 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	0.759 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 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

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

Proposed Approach:

- The files we found not to have any null values
- Type conversion was performed of the city dataset to convert the figures from string to integer
- The date of travel column in Cab dataset was converted to datetime type
- The resultant dataframes of transaction data, customer data and cab data were merged together to form one complete dataset that was used to perform the EDA