Data Intake Report

Name: G2M Insight for Cab Investment Firm
Report Date: 01/13/2024
Internship Batch: LISUM 29
Version: 1.0
Data Intake By: Rahul Korde
Data Intake Reviewer:
Data Storage Location:

Tabular data details

Cab Dataset:

Total number of observations:	359392
Total number of files	1
Total number of features	7
Base format of the features	.csv
Size of the data	21.2 MB

City Dataset:

Total number of observations:	20
Total number of files	1
Total number of features	3
Base format of the features	.csv
Size of the data	4 KB

Customer ID Dataset:

Total number of observations:	49171
Total number of files	1
Total number of features	4
Base format of the features	.csv
Size of the data	1.1 MB

Transaction ID Dataset:

Total number of observations:	440098
Total number of files	1
Total number of features	3
Base format of the features	.csv
Size of the data	9 MB

Proposed Approach:

Duplicate Rows
There are no missing values in the given dataset. The operation ".dro_duplicates( )" was used on all datasets to remove any and all duplicate values in the datasets

Missing Values
The dataset was performed with the ".isnull.sum( )" to check with null values and NA in the data. There were none in the dataset.

Assumptions:
The data for the cars is between 2016 and 2018