Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: August 18th, 2023
Internship Batch: LISUM24
Version:1.0
Data intake by: Shreya
Data intake reviewer:
Data storage location:

Tabular data details:

Cab:

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	20663 KB

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	1 KB

Transaction:

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	8788 KB

Customer:

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	<size in GB,TB,PB,MB etc>

Proposed Approach:
- Load and explore the datasets individually to analyze and understand their features and on which features the datasets are connected.
- Next utilize the identified columns to connect the features and merge into a master data set containing details about customer ID, transaction ID, the cabs, and the city.
- Remove the non-numerical characters in any numerical columns and clean the data.
- Check for missing values or null values.
- Perform comparative analysis to get insight on which cab company is better to invest on.