Data Intake Report

Name: <G2M insight for Cab Investment firm>
Report date: <07/13/2023>
Internship Batch: LISUM23:30
Version: 1.0
Data intake by: Zhijia Xu
Data storage location: https://github.com/DataGlacier/DataSets

Tabular data details:

Cab_Data.csv

Total number of observations	359392
Total number of files	4
Total number of features	7
Base format of the file	csv
Size of the data	19.2 MiB

Customer_ID.csv

Total number of observations	49171
Total number of files	4
Total number of features	4
Base format of the file	csv
Size of the data	1.5 MiB

Transaction_ID.csv

Total number of observations	440098
Total number of files	4
Total number of features	3
Base format of the file	csv
Size of the data	10.1 MiB

City.csv

Total number of observations	20
Total number of files	4
Total number of features	3
Base format of the file	csv
Size of the data	608.0 B

Proposed Approach:

- Identify the relationship between these datasets and try to merge them into master data before conducting any analysis.
- Conduct descriptive analysis to identify abnormal features
- Conduct correlation analysis to generate some basic insights into the data.
- Conduct contextual analysis to evaluate the company's performance on a time-basis.
- Choose appropriate features to build the model and forecast the future performance.