Data Intake Report

Name: G2M Insight
Report date: 09/20/2025
Internship Batch: LISUM49
Version:<1.0>
Data intake by: Tahsin Azad
Data intake reviewer: Tahsin Azad
Data storage location: Github, Jupyter

Tabular data details:

City.csv

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	0.7412109375 kb

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	20662.7050 kb

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	1026.58 KB

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	8787.30 KB

Proposed Approach:
- Duplicates will be dropped
- Columns deemed unnecessary will discarded

- Some assumptions are that all the cities are in America, are major cities, and no spelling mistakes
- Also some columns between the different tables have similar names, allowing for Joins to be made