Data Intake Report

Name: G2M insight for Cab Investment firm
Report date: 19 April 2024
Internship Batch: LISUM32
Data intake by: Madhvik Bhalani
Data storage location: https://github.com/Madhvik-Bhalani/VC/tree/main/week%202/Dataset

Tabular data details:

File Name	Cab_Data
Total number of observations	359392
Total number of features	7
Base format of the file	.csv
Size of the data	20.66 MB

File Name	Transaction_ID
Total number of observations	440098
Total number of features	3
Base format of the file	.csv
Size of the data	8.79 MB

File Name	customer_ID
Total number of observations	49171
Total number of features	4
Base format of the file	.csv
Size of the data	1.02 MB

File Name	city
Total number of observations	20
Total number of features	3
Base format of the file	.csv
Size of the data	1 KB

Proposed Approach:

1. **Cab_Data**
   Change Description:
   Converted "Company" column values to numerical representation:
   Pink Cab (0), Yellow Cab (1).

   Reason for Change:
   To enable numerical analysis and modeling for improved efficiency and comparability.

   Impact Analysis:
   Updated "Company" column values to numerical codes.
   No structural changes to the dataset.
   Ensured data consistency and integrity.