Data Intake Report

Name: G2M Insight for Cab Investment Firm
Report date: October 14, 2023
Internship Batch: LISUM26
Version: 1.0
Data intake by: Jie Heng Yu
Data intake reviewer: Jie Heng Yu
Data storage location: https://github.com/DataGlacier/DataSets

Tabular data details:

File name	Cab_Data
Total number of observations	359,392
Total number of features	7
Base format of the file	.csv
Size of the data	21.2 MB
File name	City
Total number of observations	20
Total number of features	3
Base format of the file	.csv
Size of the data	759 bytes
File name	Customer_ID
Total number of observations	49,171
Total number of features	4
Base format of the file	.csv
Size of the data	1.1 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	9 MB
Total number of files	4

Proposed Approach:

- Rows with missing values are removed from analysis
- Duplicate rows are reduced to unique row
- Assumptions include:
    - all customers pay the cab fare after their ride, so there is realized profit
    - all variables are true at the population level
    - all variables were recorded without error
    - there was no ride-sharing -- observations for a customer is independent to that of another customer