Data Intake Report

Name: Preliminary Analysis of U.S. Cab Industry
Report date: 11/12/2022
Internship Batch: LISUM15
Version: 1.0
Data intake by: Xiaoqing (Lucy) Leng
Data intake reviewer: Data Glacier Team
Data storage location: https://github.com/LucyLeng/G2M_Cab_EDA.git

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+ MB

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+ MB

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+ MB

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+ bytes

Proposed Approach:

Exploratory Data Analysis
Data Quality Analysis
Modeling (classifier)