Data Intake Report

Name: G2M insight for Cab Investment firm
Report date: 10/2/2021
Internship Batch: LISUM04
Version:1.0
Data intake by: Bao Khanh Nguyen
Data intake reviewer:
Data storage location: Github link

Tabular data details: 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	20.1 MB

Tabular data details: City

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	759 bytes

Tabular data details: 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	1 MB

Tabular data details: 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	8.58 MB

Note: Replicate same table with file name if you have more than one file.

Proposed Approach:
"data" is a dataframe is created by merging mentioned 4 dataframes:
- Checking missing and duplicated data and drop it