Data Intake Report

Name: G2M insight for Cab Investment Firm
Report date: 08/14/2023
Internship Batch: LISUM19
Version:1.0
Data intake by: Kumar Mitra
Data intake reviewer:
Data storage location: https://github.com/kbmPractice/Week2

Tabular data details:

Cab_Data.csv:

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

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	759 bytes

Customer_ID.csv

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

Transaction_ID.csv

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	38.3 MB

Proposed Approach:

1. The *duplicated()* function was used to determine that there were no duplicates in the any of the files.
2. The *isnull()* was used to determine that the files didn't not have null values either.