Data Intake Report

Name: <G2M insight for Cab Investment firm>
Report date: <20-06-2021>
Internship Batch:<LISUM01>
Version:<1.0>
Data intake by:<Vincent Obiri>
Data intake reviewer:
Data storage location: <https://github.com/Vince689/DataSets.git>

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	`<19.2+ MB>`

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	`<1.5+ MB>`

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	`<10.1+ MB>`

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	`<608.0+ bytes>`

US Holiday Dates

Total number of observations	`<342>`
Total number of files	`<1>`
Total number of features	`<6>`
Base format of the file	`<.csv>`
Size of the data	`<1.5+ MB>`

Proposed Approach:

- There are no duplicated in the datasets provided. A dedupe check conducted confirms this.