Data Intake Report

Name: G2M insight for Cab Investment firm
Report date: October 4th, 2021
Internship Batch: LISUM04
Version: 1.0
Data intake by: Guillermo Leija Renteria
Data intake reviewer:
Data storage location: https://github.com/kaztyel/Week2.git

Tabular data details:

File name:	Cab_Data.csv
Total number of observations	359,392
Total number of files	1
Total number of features	7
Base format of the file	.csv
Size of the data	20.1 MB

File name:	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	4 KB

File name:	Customer_ID.csv
Total number of observations	49,171
Total number of files	1
Total number of features	4
Base format of the file	.csv
Size of the data	1 MB

File name:	Transaction_ID.csv
Total number of observations	440,098
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:

- For Dedup, I will use Panda function .drop_duplicates()