Data Intake Report

Name: Cab EDA
Report date: October 13th 2022
Internship Batch: LISUM14
Version: 0.1
Data intake by: Justin Lee
Data intake reviewer:
Data storage location: https://github.com/DataGlacier/DataSets

Tabular data details: Cab_Data.csv

Total number of observations	359392
Total number of files	
Total number of features	7
Base format of the file	.csv
Size of the data	20663 KB

Tabular data details: City.csv

Total number of observations	20
Total number of files	
Total number of features	3
Base format of the file	.csv
Size of the data	1 KB

Tabular data details: Transaction_ID.csv

Total number of observations	440098
Total number of files	
Total number of features	3
Base format of the file	.csv
Size of the data	8788 KB

Tabular data details: Customer_ID.csv

Total number of observations	49171
Total number of files	
Total number of features	4
Base format of the file	.csv
Size of the data	1027 KB

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

Proposed Approach:

- After joining all of the datasets, I will use the duplicated function from pandas to perform a dedupe validation.
- Assumptions:
    - I will assume that the expenses for the trip only involves fuel charges
    - There isnt sufficient data on the internet for the base fares per year for each city in the US. This will be left out of the analysis.

Note: Convert this doc in pdf and provide the link of pdf file in your dashboard.
Please do not forget to remove this section while converting the file into pdf.