Data Intake Report

Name: G2M insight for Cab Investment firm
Report date: March 5th, 2021
Internship Batch: LISP01
Version:<1.0>
Data intake by: Sebastián J. Castro
Data intake reviewer:<intern who reviewed the report>
Data storage location: https://github.com/sebacastrocba/dataglacier-internship

Tabular data details:

Cab_Data.csv

Total number of observations	359392
Total number of features	7
Base format of the file	.csv
Size of the data	21.2 MB

City.csv

Total number of observations	20
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 features	4
Base format of the file	.csv
Size of the data	1.1 MB

Transaction_ID.csv

Total number of observations	440098
Total number of features	3
Base format of the file	.csv
Size of the data	9 MB

US_Holidays_Day.txt

Total number of observations	90
Total number of features	2
Base format of the file	.txt

Size of the data	2.9 kB

Proposed Approach:
- Mention approach of dedup validation (identification)
- Mention your assumptions (if you assume any other thing for data quality analysis)