Data Intake Report

Name: G2M insight for Cab Investment firm
Report date: 03/21/2023
Internship Batch:<Enter your batch code from Canvas course>
Version:<1.0>
Data intake by: Mandar Bhushan Parkar
Data intake reviewer:<intern who reviewed the report>
Data storage location: https://github.com/MandarParkar/Cabs-G2M-Strategy

Tabular data details:

City

Total number of observations	20
Total number of files	4
Total number of features	3
Base format of the file	csv
Size of the data	1 KB

Cab_Data

Total number of observations	359392
Total number of files	4
Total number of features	7
Base format of the file	csv
Size of the data	20663 KB

Customer_ID

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

Transaction_ID

Total number of observations	440098
Total number of files	4
Total number of features	3
Base format of the file	csv
Size of the data	8788 KB

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

Proposed Approach:

- Mention approach of dedup validation (identification)
  There were no duplicate rows in the datasets and no null values as well.
- Mention your assumptions (if you assume any other thing for data quality analysis)
  Some assumptions such as KM_travelled, Price charged and cost of trip are considered to be unbounded and therefore, data type of these columns are not changed.

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.