Data Intake Report

Name: G2M insight for Cab Investment firm
Report date: 12/03/2023
Internship Batch: LISUM19
Version: 1.0
Data intake by: Thanuja Modiboina
Data intake reviewer: (Thanuja Modiboina)
Data storage location: https://github.com/DataGlacier/DataSets.git

Tabular data details:
Dataset1: 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	20663 kb

Dataset2: 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	1 kb

Dataset3: 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	1027 kb

Dataset4: 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	8788 kb

Note: Replicate the same table with the file name if you have multiple files.

Proposed Approach:

- Mention the approach of dedup validation (identification)
- Mention your assumptions (if you assume any other thing for data quality analysis)
- Here, I used an interquartile range to find outliers. This is used to determine the variability in a dataset by subtracting the 25^{th} percentile (Q1) from the 75^{th} percentile (Q3).
- From this dataset, the profit is calculated from the 'Price Charged' and 'Cost of Trip' columns.

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