Data Intake Report

Name: G2M insight for Cab Investment firm
Report date: 14/04/2023
Internship Batch: LISUM20
Version: 1.0
Data intake by: Yashwanth Reddy Bomma Reddy
Data intake reviewer:
Data storage location:

Tabular data details:

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

Total number of observations	20
Total number of files	
Total number of features	3
Base format of the file	.csv
Size of the data	759 B

Total number of observations	49171
Total number of files	
Total number of features	4
Base format of the file	.csv
Size of the data	1.05 MB

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

Proposed Approach:
We conducted a thorough deduplication process on the data set using exact matching methods. After manual review, we found no duplicates in the data set
Assumptions:
- The Customer ID is a unique identifier that remains consistent across different companies and systems.
- The Price charged and Cost of Trip variables are expressed in the same currency unit, namely USD.