Data Intake Report

Name: G2M insight for Cab Investment firm
Report date: 17-06-2022
Internship Batch: LISUM10: 30
Version:1.0
Data intake by: Anju Paul
Data intake reviewer: Anju Paul
Data storage location: https://github.com/DataGlacier/DataSets.git

Tabular data details:

Cab_Data

Total number of observations	359392
Total number of files	1
Total number of features	14
Base format of the file	.csv
Size of the data	20663 KB

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

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

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

Proposed Approach:
- Mention approach of dedup validation (identification)

- Mention your assumptions (if you assume any other thing for data quality analysis)

  - Outliers are calculated with Inter Quartile Range (25 <IQR <75)
  - December is holiday month