Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 07.06.2024

Internship Batch: LISUM34

Version:

Data intake by: Mohammad Tohin Bapari

Data intake reviewer:

Data storage location: https://github.com/iamtohin/Data-Analytst-Internship-at-Data-Glacier/tree/main/Week%202

Tabular data details

Cab_data Dataset

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	19.2+ MB

City Dataset

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	608.0+ bytes

Customer_ID Dataset

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	1.5+ MB

Transaction_ID Dataset

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	10.1+ MB

Proposed Approach and findings:

- **There are no null values in all 4 datasets.**
- **All the transactions are unique.**
- **Spits the both cab company into two data frame.**
- **We can see that dataset uses a excel serial date format. We change it to 'yyyy-mm-dd' format.**