Data Intake Report

Name: G2M insight for Cab Investment firm
Report date: 11/8/2022
Internship Batch: LISUM12
Version:<1.0>
Data intake by: Data Science Intern
Data intake reviewer:<intern who reviewed the report>
Data storage location:
https://github.com/Mariamali2001/Data_Glacier_virtual_internship/tree/Week2

Tabular data details:

City dataset

Total number of observations	20
Total number of files	1
Total number of features	2
Base format of the file	.csv
Size of the data	1KB

Cab 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	20,663KB

Customer 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	1,027KB

Transactions 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	8,788KB

Proposed Approach:

- Assumptions: Outliers are present in Price Charged feature but due to unavailability of trip duration details, we are not treating this as outlier and there is no other feature present in it outliers