Data Intake Report

Name: G2M_EDA
Report date: 11/04/2023
Internship Batch: LISUM19
Version:1.0
Data intake by: Ouedraogo Ami Samyra
Data intake reviewer:
Data storage location: https://github.com/amisamyra99/G2M_EDA

Tabular data details:
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	21.9 MB

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	759B

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	1.05 MB

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	9B

Proposed Approach:

those are some data quality issues I encounter while working with the datasets mentioned below.

there is no null value
cab_data

- The date of Travel has to be in DateTime format

City

- delete the ',' and convert from population and user and convert then into integer format
- TransactionID and cab_data could be joined