Data Intake Report

Name: G2M insights for Cab Investment firm
Report date: 15-06-2022
Internship Batch: LISUM10:30
Version:<1.0>
Data intake by: Ramya Mercy Rajan
Data intake reviewer:<intern who reviewed the report>
Data storage location: [DataGlacierAssignment/DataGlaciers_Ramya.ipynb at main · ramyaabin/DataGlacierAssignment (github.com)](github.com)

Tabular data details: Cab Data

Total number of observations	359392 Rows
Total number of files	1
Total number of features	7 Columns
Base format of the file	**Cab_Data.csv**
Size of the data	20.1 MB

Tabular data details: City Data

Total number of observations	20 Rows
Total number of files	1
Total number of features	3 Columns
Base format of the file	**City.csv**
Size of the data	759 bytes

Tabular data details: Customer ID

Total number of observations	49171 Rows
Total number of files	1
Total number of features	4 Columns
Base format of the file	**Customer_ID.csv**
Size of the data	1.00 MB

Tabular data details: Transaction ID

Total number of observations	440098 Rows
Total number of files	1
Total number of features	3 Columns
Base format of the file	**Transaction_ID.csv**
Size of the data	8.58 MB

Proposed Approach:

There are no Null values
There are no Duplicate values
There are only outliers for Price charged column of Cab Dataset.
It is well observed that Yellow Cab performed well as compared to Pink Cab

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

Other Transportation modes comparison in Cities is the good analysis as it gives information about the reason peoples use cab more (High Parking fees in some areas, Shortage of Parking places can cause delay in reaching to workplace or Hospitals or Airports etc. Public transport like Buses and Metro are less accessible to some areas are another reason why people prefer Cab Service.