Data Intake Report

Name: G2M insight for Cab Investment firm
Report date: 3/13/2023
Internship Batch: LISUM19
Version: 1.0
Data intake by: Snigdha Chigurupati
Data intake reviewer: NA
Data storage location: NA

Tabular data details:

Cab_Data.csv

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	22.01 MB

Customer_ID.csv

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	1.027 MB

Transaction_ID.csv

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	8.788 MB

City.csv

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

Proposed Approach:

1. There are no missing values in any of the datasets.
2. All the data types are correct except for 'Date of Travel' in Cab_Data.csv which is 'object' and need to convert to 'datetime64[ns]'. I have converted 'Date of Travel' column using MS Excel. Next the 'Population' and 'Users' of 'City.csv' are object type which I converted to integer type.
3. As and required for analysis, I have merged the datasets.

Data identification/Understanding:

1. Cab_Data.csv :
   0 Transaction ID - Transaction ID of each ride
   1 Date of Travel - Date of the ride
   2 Company - Cab Company(Pink and Yellow)
   3 City - U.S cities
   4 KM Travelled - Kilometers travelled for each ride.
   5 Price Charged - Price charged to the customers.
   6 Cost of Trip - Expense incurred for the ride by the company.
2. Customer_ID.csv:
   0 Customer ID - Customer ID given by each company which is unique.
   1 Gender - Customer gender
   2 Age – Customer age
   3 Income (USD/Month) – Customer Income
3. Transaction_ID.csv:
   0 Transaction ID - Transaction ID of each ride
   1 Customer ID - Customer ID given by each company which is unique.
   2 Payment_Mode – Payment method of the trip
4. City.csv:
   0 City - U.S Cities
   1 Population - Population of the city
   2 Users – Out of the total population, the number of can service users.

Inference from Hypothesis:

1. Yellow cab company generated more revenue and profits than Pink cab company for all years.
2. In most of the U.S cities, people preferred Yellow cab company than Pink cab company.
3. The forecasting of profits was seen positively for Yellow Cab than Pink Cab company.
4. From above all, it can be suggested that investing in Yellow would be right for the XYZ firm.