Data intake Report

Name: G2M insight for Cab Investment firm (Must for all Specialization)
Report Date: 05/Sep/2021
Internship Batch: LISUM03
By: Ilyas Nayle
Location: https://github.com/FutureCoderme/DataSets

The Problem

XYZ is a private firm in US. Due to remarkable growth in the Cab Industry in last few years and multiple key players in the market, it is planning for an investment in Cab industry and as per their Go-to-Market(G2M) strategy they want to understand the market before taking final decision.

Data Set: 4 individual data sets. Time period of data is from **31/01/2016** to **31/12/2018**. Below are the Data sets information.

Cab_Data.csv – this file includes details of transaction for 2 cab companies

Customer_ID.csv – this is a mapping table that contains a unique identifier which links the customer's demographic details

Transaction_ID.csv – this is a mapping table that contains transaction to customer mapping and payment mode

City.csv – this file contains list of US cities, their population and number of cab users

Total data points: 359392

Tabular Data Details

Cab Data

Total number of observations	359392
Total number of features	7
Total number of files	1
Size of data	19.2+MB
File format	.csv

City

Total number of observations	20
Total number of features	3
Total number of files	1
Size of data	608.0+bytes
File format	.csv

Customer ID

Total number of observations	49171
Total number of features	4
Total number of files	1
Size of data	1.5+MB
File format	.csv

Transaction ID

Total number of observations	440098
Total number of features	3
Total number of files	1
Size of data	10.1MB
File format	.csv

EDA:

- ➔ Show statistical for each column
- ➔ Graph
- ➔ Find null
- ➔ Remove duplicate values

Hypothesis:

- ➔ Price charged
- ➔ Age
- ➔ Income
- ➔ Profit
- ➔ KM travelled