Data Intake Report

Name: G2M_insight_for_Cab_Investment_firm
Report date: 07/14/23
Internship Batch: LISUM23
Version:<1.0>
Data intake by: Yuheng Chen
Data intake reviewer: Yuheng Chen
Data storage location: https://github.com/DataGlacier/DataSets/tree/main

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	21.2MB

Tabular data details: 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	1.05MB

Tabular data details: 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	9MB

Tabular data details: 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	739B

Note: Replicate same table with file name if you have more than one file.

Proposed Approach:

1. **Data preprocessing:**
   a. **Data types: such as date time or integer/string types**
   b. **Check null values**
   c. **Removing duplicates**
   d. **Merge datasets together**
2. **EDA:**
   a. **Add new column: profit(cost- price charged)**
   b. **Add 3 new columns: year, month, date(Extract from the column called Date of Travel.**
   c. **Add one more column: User_% (percentage of cab users for each city)**
3. **Visualizations**

Note: Convert this doc in pdf and provide the link of pdf file in your dashboard.
Please do not forget to remove this section while converting the file into pdf.

The Client

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.

Project delivery:

You have been provided with multiple data sets that contains information on 2 cab companies. Each file (data set) provided represents different aspects of the customer profile. XYZ is interested in using your actionable insights to help them identify the right company to make their investment.

The outcome of your delivery will be a presentation to XYZ's Executive team. This presentation will be judged based on the visuals provided, the quality of your analysis and the value of your recommendations and insights.