

Data intake Report

The given data involved four tables namely:-

1. Cab_data.
2. Customer_ID.
3. City.
4. Transaction.

A brief summary of each data set below follows:-

1. Transaction_ID- mapped out the transactions attached to each customer. 440099 rows of data.
2. Customer_ID- mapped out a customer's unique details i.e. ID with their age, gender as well income bracket. 49172 rows of data.
3. City- mapped a list of cities being considered for this project with their respective populations and specific company cab dependence/ popularity in that area. 21 rows of data.
4. Cab_data- this mapped out the entire summary of the whole investigation. From the companies involved, transactions performed, cities involved, to trip price and distance. 359393 rows of data.

Data Discrepancies and Observations

- ❖ The city **San Francisco** does not appear in the Cab_Data data set.
- ❖ The Cab_Data data set has empty rows (last rows)
- ❖ The **date column** in the Cab_data is not in date/time format.

- ❖ The data in the Cab_data and Transaction_ID datasets is mapped according to the ascending chronological order of the **Transaction Column**.
- ❖ The City dataset has the **least inputted data**.
- ❖ The customer dataset is the only one with a data feature that directly links us to the specific Client i.e. the **Customer_ID**
- ❖ The Cab_data dataset set provides raw data of the actual financial implications of this investigation through the **Cost of Trip and Price Charged features**.

Creation of additional columns to give better insights.