

Data Intake Report

Datasets:

- Cab_Data.csv: This dataset contains transaction details for 2 cab companies, which includes Transaction ID, Date of Travel, Company, City, KM Travelled, Price Charged, and Cost of Trip.
- Customer_ID.csv: This dataset serves as a mapping table that contains a unique identifier linking the customer's demographic details, such as Customer ID, Gender, Age, and Income (USD/Month).
- Transaction_ID.csv: This dataset is a mapping table containing transaction-to-customer mapping and payment mode.
- City.csv: This dataset contains a list of US cities, their population, and number of cab users.

Data Preparation:

- Datasets have been merged on appropriate keys to create a comprehensive view of the information.
- Duplicates were removed from the final merged dataset.
- Missing values were filled using forward-fill method.
- New features were engineered including "Profit" and "Income Group".

Data Overview:

- The merged dataset provides a detailed view of transactions for 2 cab companies over the period from 31/01/2016 to 31/12/2018.
- It contains information about the customers who made the transactions and the specifics of each transaction such as the date, distance travelled, price charged, cost of trip, and profit.
- Additional customer demographic details are included such as gender, age, and income.

- The dataset also includes city-level data such as population and number of cab users.

Insights Obtained:

- Insights about total revenue, number of rides, and profit for each company were derived.
- Customer segmentation was done based on income and age, and patterns were observed in their cab usage.
- Market share analysis was performed to compare the presence of the two companies in different cities.
- Time-based analysis was carried out to observe the revenue trend by the day of the week.