CAB INDUSTRY

An SQL Research Project
Presented By Sourav Mondal

PROJECT VISION

Due to remarkable growth in the Cab Industry in last few years and multiple key players in the market, several investment firm of the United states are planning for an investment in Cab industry as per their Go-to-Market strategy and they want to understand the market before taking final decision. Hence, this project is an analysis of few research topics which might help in understanding the market better for future investments.

The Dataset comprises of four tables :

1. Taxi Table – which includes details of transaction for 2 cab companies

2. Customers Table– This is a mapping table that contains a unique identifier which links the customer's demographic details

3. Transactions Table – A mapping table that contains transaction to customer mapping and payment mode

4. Cities Table – This contains the list of US cities, their population and number of cab users

PROJECT MISSION

1. Creating a database to assign all the four tables, which are to be analysed.

2. Normalising the data and to create a relational database using ER diagram.

3. Data cleaning

4. Analysing certain research topics.

DATA TABLES

The dataset contains a total of 4 tables, having 14 unique features.

The Cities table contains
- City: containing the name for each city.
- Population: The population of each city.
- Users: The number of Cab users in each city

The Customers table contains
- Customer ID: Unique identifier for each customer.
- Gender: Gender of the customer.
- Age: Age of the customer.
- Income per month: The income per month of the customer.

The Transactions table contains
- Transaction ID: Unique identifier for each transaction.
- Customer ID : Links to the Customer ID in the Customers table, representing which customer made the transaction.
- Payment Mode: The mode of payment used for the transaction.

The Taxi table contains
- Transaction ID: Links to the Transaction ID in the Transactions table, indicating which transaction the taxi ride is associated with.
- Date of Travel: The date when the taxi ride occurred.
- Company: The taxi company involved in the ride.
- City : Links to the City in the Cities table, indicating the city where the taxi ride took place.
- Distance Travelled: The distance traveled during the taxi ride.
- Price Charged: The price charged for the taxi ride.
- Cost of Trip: The cost of the taxi trip.

DATA NORMALISATION

1. The Customer's table is linked to the Transactions table through the Customer ID field holding a **one-to-many** relationship.

2. The Transactions table is linked to the Taxi table through the Transaction ID field, here the Transaction ID acts as a **Foreign key** to the Taxi table.

3. The Cities table is linked to the Taxi table through the City field, again being the **Foreign key** to the Taxi table.
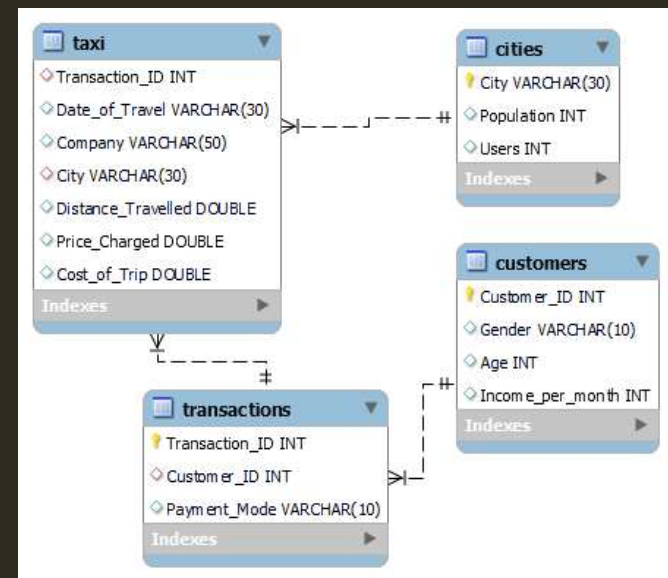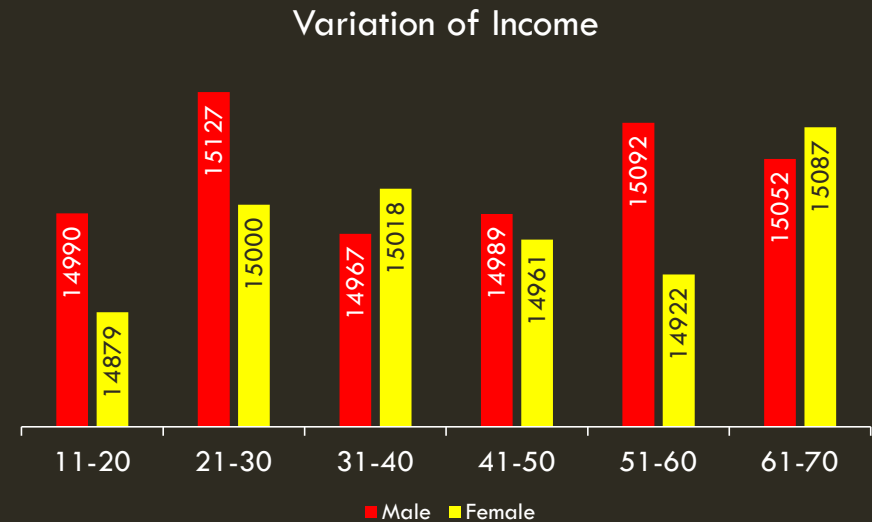
DATA CLEANING

- The initial dataset underwent thorough validation to ensure completeness and accuracy and no missing values were detected during the data validation process.

- The data column in the Taxi table, initially stored as text, was converted to a more appropriate data type using the string to date function.
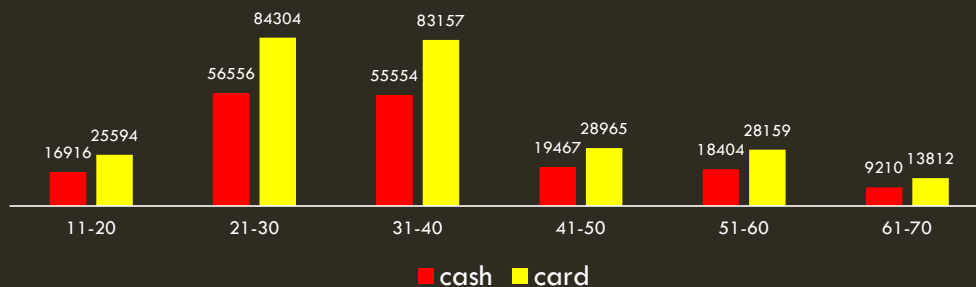
VARIATION OF INCOME AMONG CUSTOMERS

From the data chart it can be concluded that :

1. Males with age between 21-30 has the **maximum** salary, that is approximately 15,127 dollars per month on average.

2. Females with age between 61-70 has the **maximum** salary, which is approximately 15087 dollars per month on average.
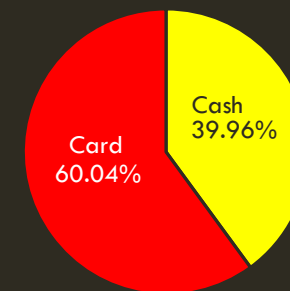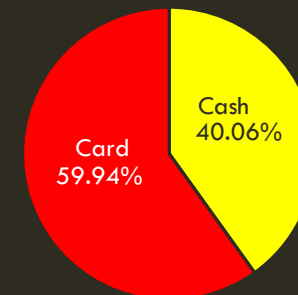
Variation of Income

14990 | 14879 | 15127 | 15000 | 14967 | 15018 | 14989 | 14961 | 15092 | 14922 | 15052 | 15087

11-20   21-30   31-40   41-50   51-60   61-70

■ Male   ■ Female

PAYMENT MODE USED BY CUSTOMERS

Payement Mode on the basis of Age

Legend: cash (red), card (yellow)

Age	cash	card
11-20	16916	25594
21-30	56556	84304
31-40	55554	83157
41-50	19467	28965
51-60	18404	28159
61-70	9210	13812

From the data table it can be stated that:

1. Customers prefer cards more, which indicates the era of cashless transactions.

2. Male and Female both customers often **prefer cards over cash approximately 60% of the time.**

Payment mode by Female Passengers

Cash 39.96%
Card 60.04%

Payment mode by Male Passengers

Cash 40.06%
Card 59.94%

TAXI TRANSACTION ON EACH CITIES

Counting the number of transactions and grouping the data based on each cites from Taxi table gives the number of transaction on each cities.

o The **maximum** Taxi transactions is seen in the New York city which is approximately **99885.**

o The **minimum** Taxi transaction is seen in Pittsburgh having a number of only **1313** transactions.

Which clearly indicates Ney York city having more population and tourist attractions. Also, New York city is a global financial and business hub with a diverse economy, attracting many business travelers and professionals who may prefer taxis for their convenience and efficiency.

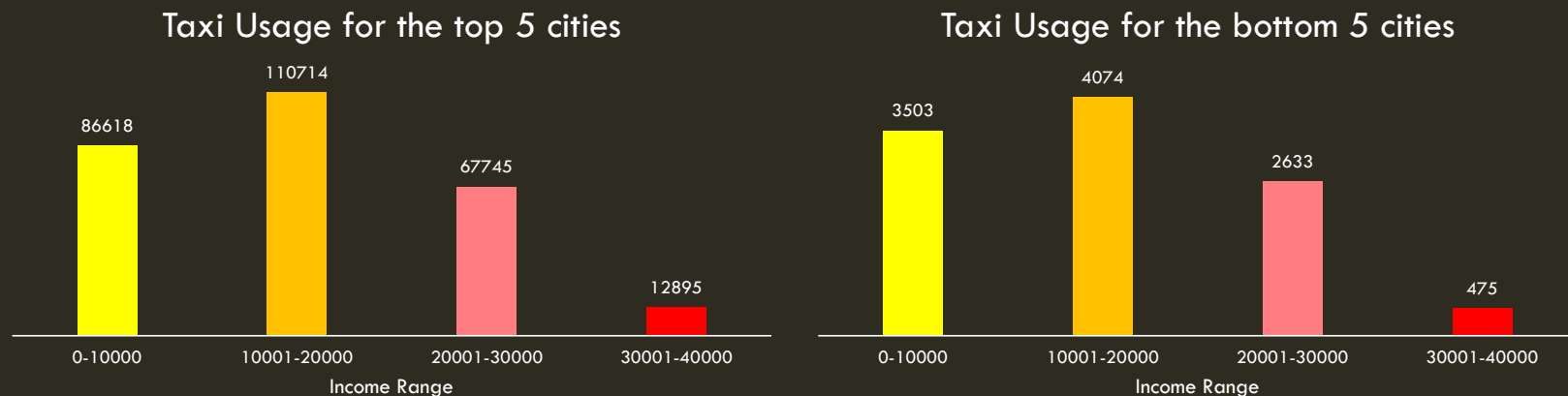# TAXI USAGE AMONG CITIES : GENDER WISE

Taxi Usage for the Top 5 Cities

Taxi Usage for the Bottom 5 Cities

Females 42%

Males 58%

Females 49%

Males 51%
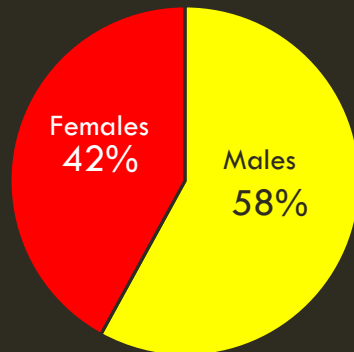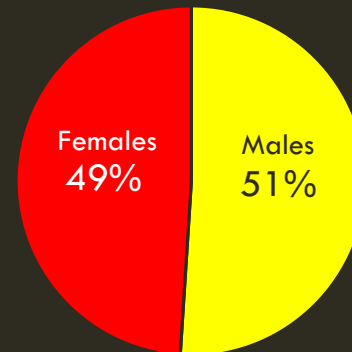
Considering the **Taxi usage** of the Top 5 and the bottom 5 cities, it is inferred that the transactions for both the genders on both the criteria is almost the same with Males taking the lead.

TAXI USAGE AMONG CITIES : INCOME WISE

Taxi Usage for the top 5 cities

86618 | 110714 | 67745 | 12895

Income Range: 0-10000 | 10001-20000 | 20001-30000 | 30001-40000

Taxi Usage for the bottom 5 cities

3503 | 4074 | 2633 | 475

Income Range: 0-10000 | 10001-20000 | 20001-30000 | 30001-40000

Among the 2 data charts which represents the **taxi transaction of the top 5 and the bottom 5 cities,** a similar pattern of taxi usage based on income range is observed. Customers having salary up to **20 thousands per month prefer cabs.** The reason why the number lessens for higher salary could be because individuals with higher salary would have their own personal car and would avail lesser cabs.

VARIATION OF PRICE CHARGED WITH DISTANCE TRAVELLED

Average Price charged Based on cities
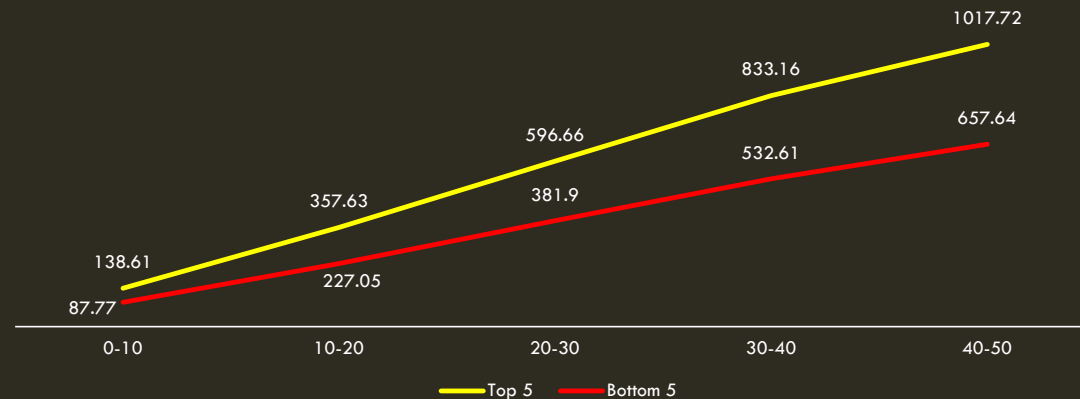
A linear growth is detected from the charts with the top 5 and bottom 5 cities based on **average price charged**, which suggests that the pricing structure follows a simple linear relationship, where the fare charged is directly proportional to the distance traveled. Also, it is clear, that the **rise in average price charged** for the top 5 cities is **steeper** than the bottom 5 cities which infers the heavy demand of cabs in the top 5 cities.

SEASONAL TRENDS IN TAXI TRANSACTIONS

Transactions trend line for each year
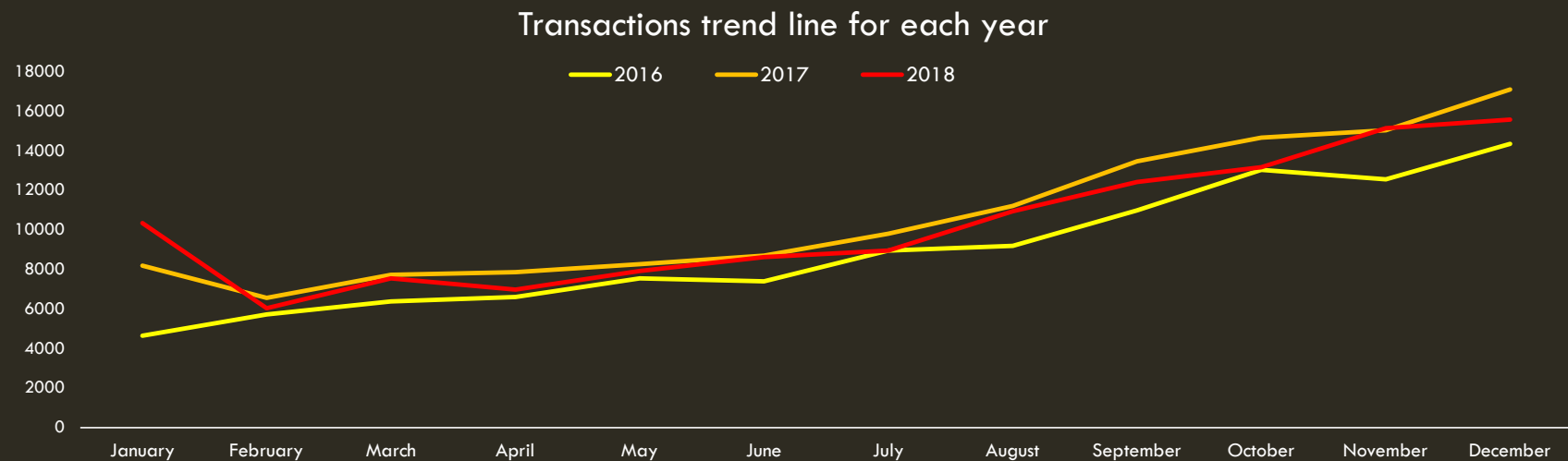
Here the Taxi transactions peaks in the months of **November and December** which reflects a combination of increased travel, tourism, shopping, social activities, and weather-related factors associated with the holiday season. It is also evident that **the number of transactions decreased all over in the year 2018 from 2017,** which indicates the concern of environmental issues and also Ride sharing services.

COMPARISON OF CAB COMPANIES
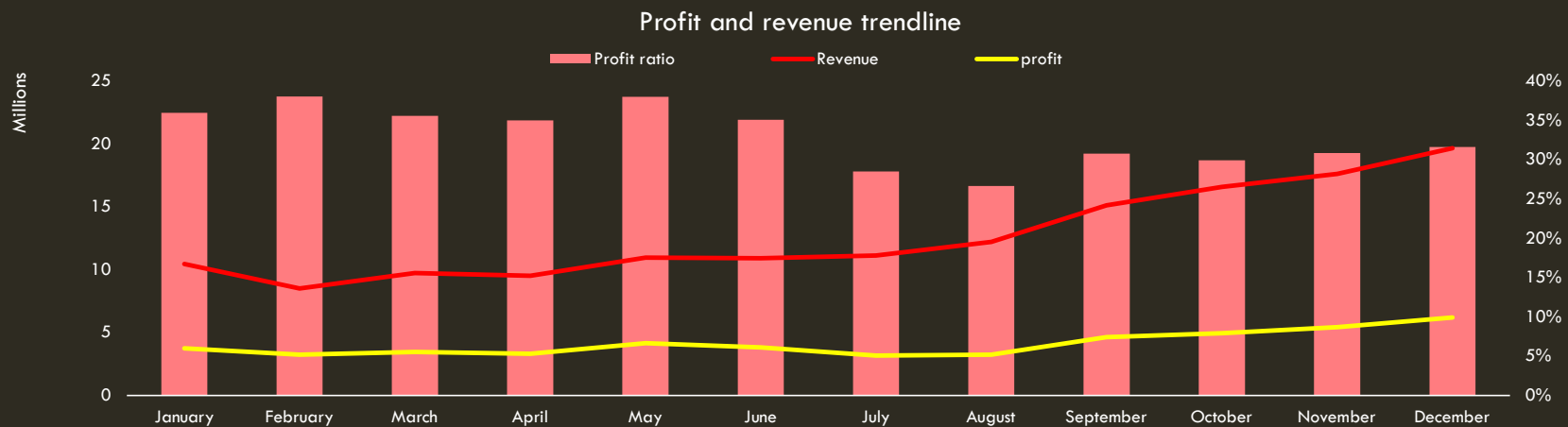
percentage of transactions

23.6%

76.4%
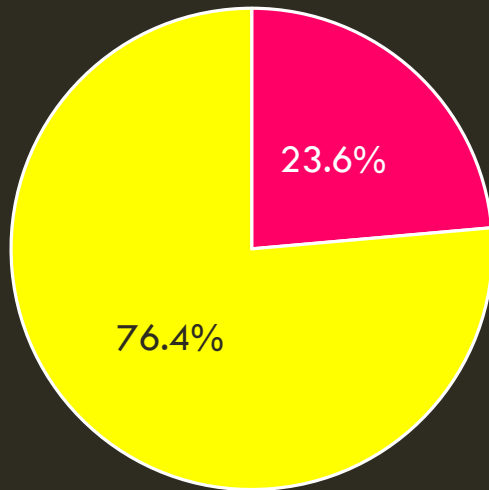
percentage of average revenue generated

17.3%

82.7%
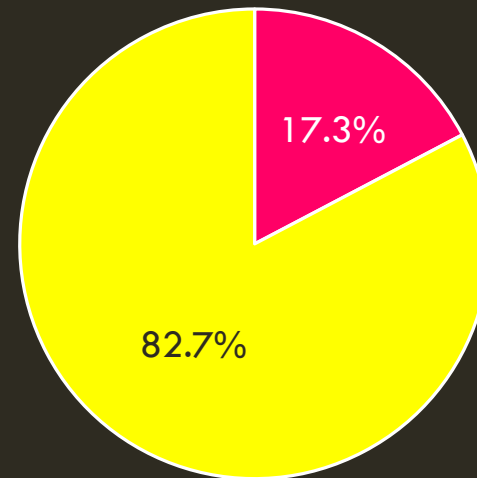
Yellow cabs have more demand than the Pink cab, as the **percentage of transaction and the average revenue generated is significantly high.**

REVENUE AND PROFIT TRENDLINE

Profit and revenue trendline

As it is quite evident from the previous analysis that the transactions peaks in the month of November and December, the revenue also increases at that point of the year, but the profit remains the same throughout the year. The profit ratios for the total year however decreases in the year ending.

THANK YOU