Data Glacier

Your Deep Learning Partner

G2M Insight for Cab Investment Firm

Data Science Virtual Internship

Submitted By : Manisha Barse

20-Feb-2023

Outline

- Project Overview – G2M (cab industry)

- Datasets Information

- Exploratory Data Analysis

- Multiple Hypothesis Testing

- Recommendation- Investment Preference

Project Overview –G2M (cab industry)

- **Introduction:**
  XYZ, a leading private equity firm in the US, is known for its strategic investments across diverse sectors. Recognizing the substantial growth in the Cab Industry and the presence of several key players, XYZ is exploring investment opportunities in line with its Go-to-Market (G2M) strategy.

- **Objective:**
  To delve into the market dynamics of the Cab Industry and provide actionable insights to XYZ, facilitating the identification of the most promising investment opportunity.

- **Key Points:**

1. **Analyzing Datasets:** Study multiple datasets containing crucial information on two cab companies: Pink Cab and Yellow Cab.

2. **Exploratory Data Analysis (EDA):** To gain insights into market dynamics, customer behavior, and key trends within the cab industry.

3. **Multiple Hypothesis Testing:** To validate assumptions and uncover significant insights, guiding investment decisions.

4. **Conclusion:** Recommendations based on the analysis, enabling XYZ to make informed investment decisions in the cab industry.

Data Overview

- We assess four distinct datasets, each offering valuable insights into various aspects of the two cab companies (Pink and Yellow) and their operations.

- Time Period: January 31, 2016, to December 31, 2018.

- List of Datasets:

  - Cab_Data.csv: Transaction details for two cab companies (Pink and Yellow).
  - Customer_ID.csv: Unique customer identifiers mapped to demographic data.
  - Transaction_ID.csv: Mapping of transactions to customers and payment modes.
  - City.csv: Information on US cities, including population and cab user statistics.

- Data Source: Obtained from the Data Glacier GitHub repository.

Data Overview - Assumptions

- **Total Number of Observations:**
  - Cab Data   : 359,392
  - Customer Data : 49,171
  - Transaction ID   : 440,098
  - City   : 20

- **Total Number of Features:**
  - Each dataset and derived datasets consist of a total of 20 features.

- **Assumptions:**
  - External Noise: We assume that there is no significant external noise beyond the datasets provided.
  - Time Period: The data spans from January 31, 2016, to December 31, 2018.
  - Random Selection: Datasets are drawn from random selection to ensure unbiased analysis.
  - Payment Methods: Only two payment methods, cash or card, are considered for transactions.
  - Profit Calculation: Profit is calculated as the difference between the price charged and the cost of the trip.

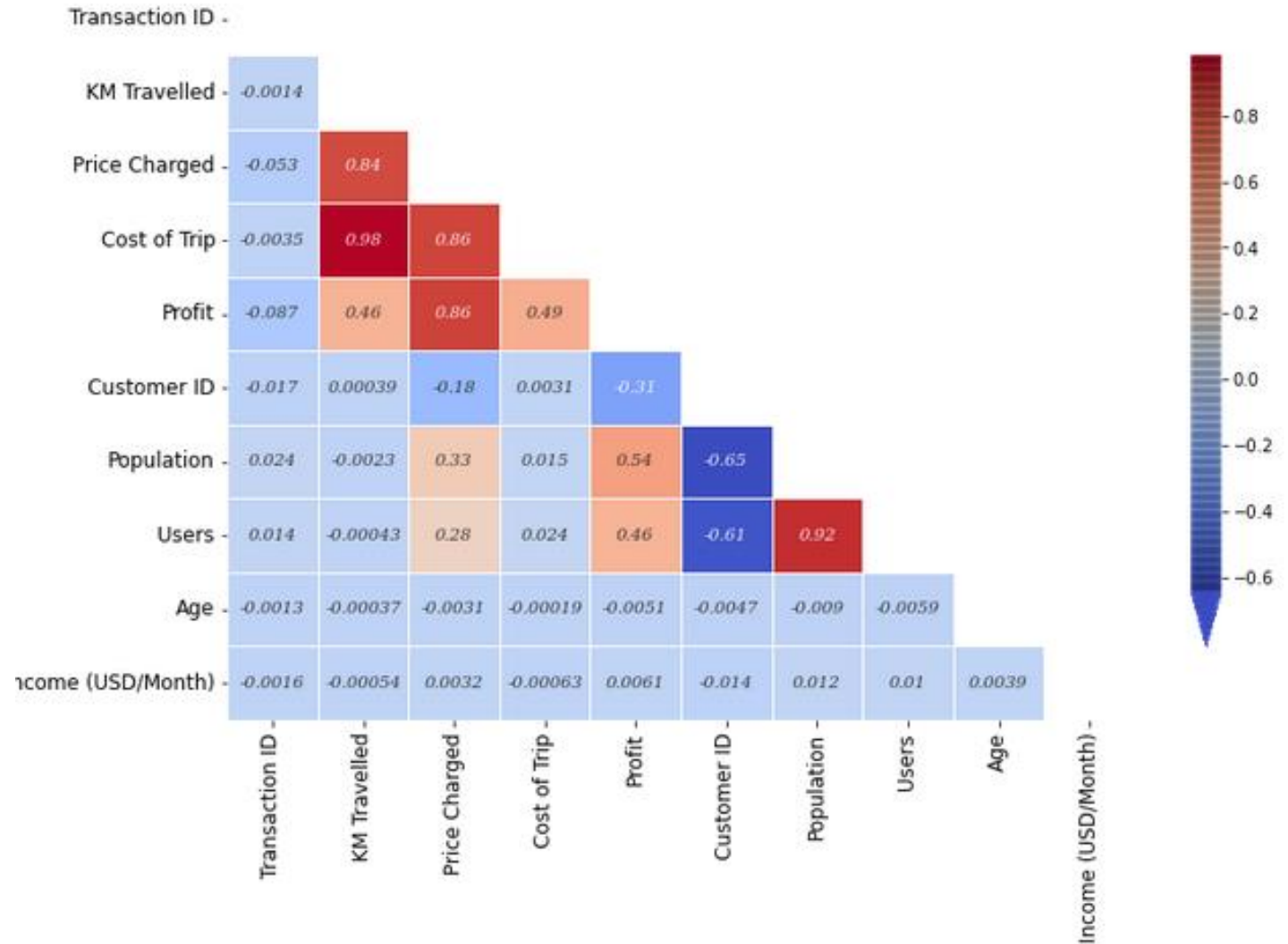- Data Source: Obtained from the Data Glacier GitHub repository.

Data Exploration
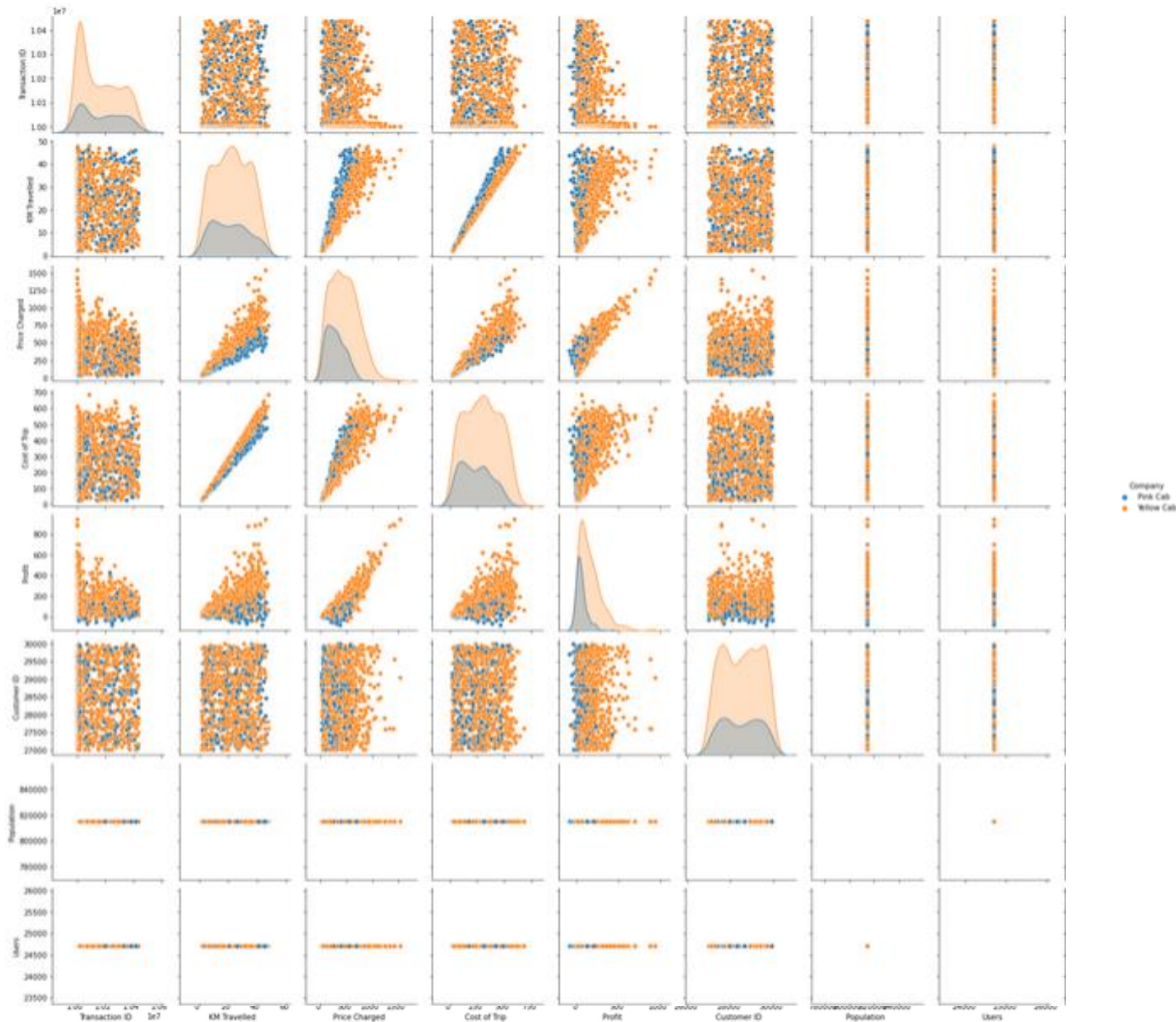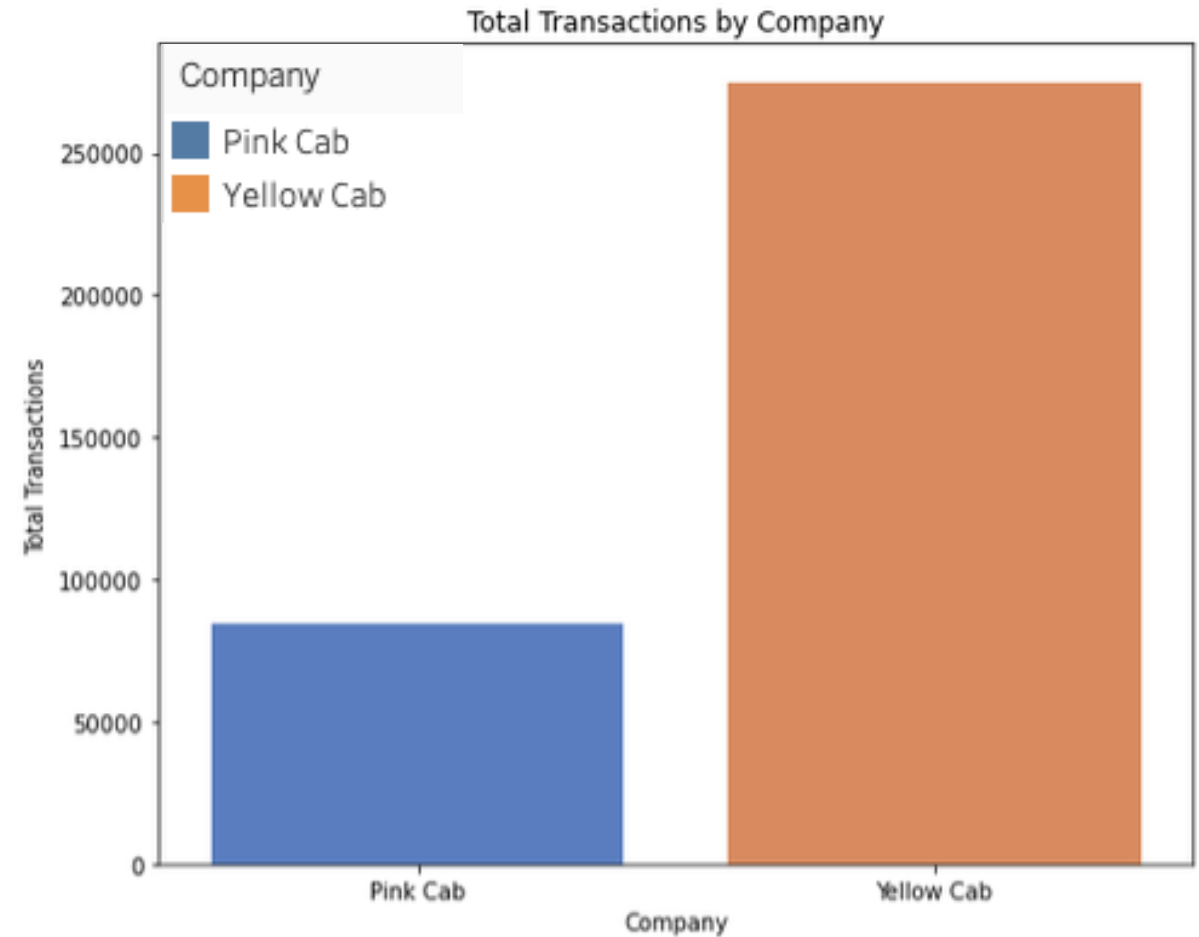
Correlation Heatmap of Master Data

Strong positive correlations between:
- Population vs Users
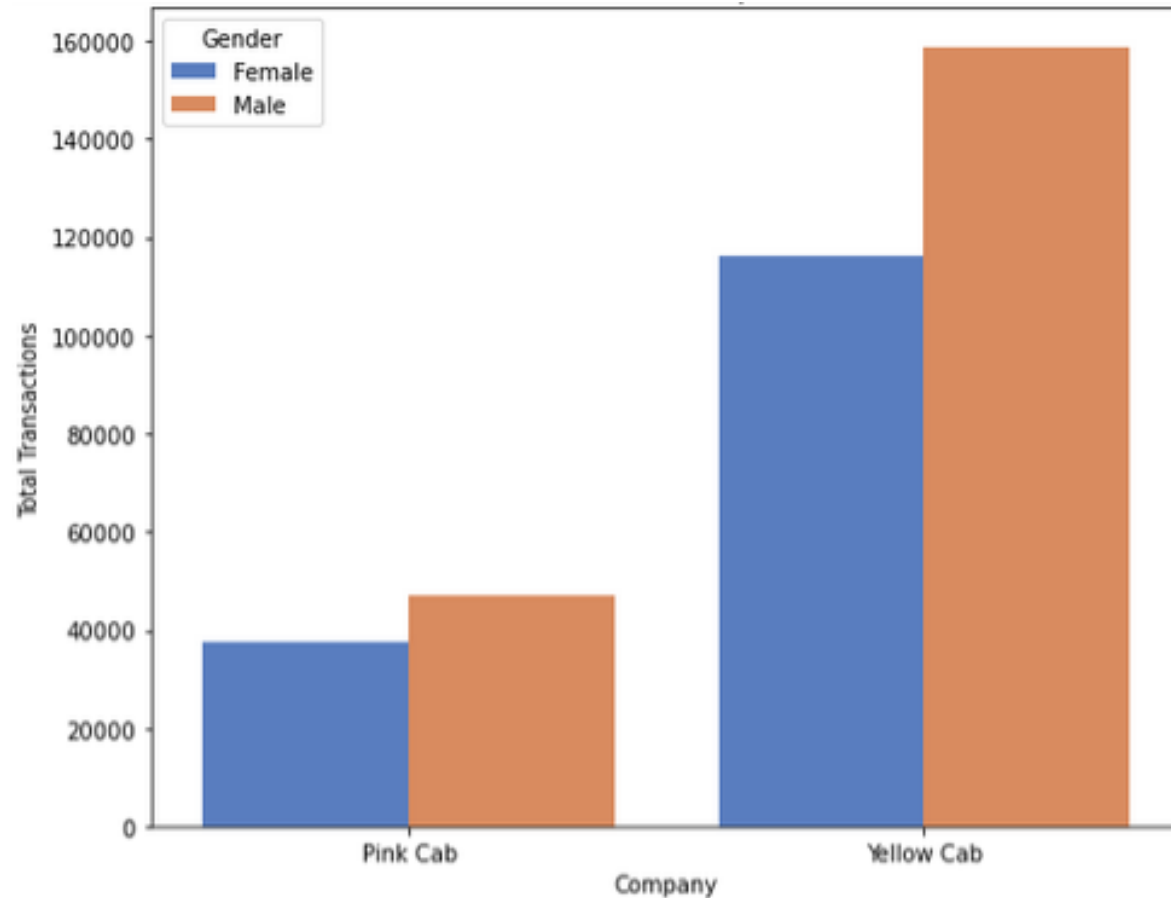- KM Travelled vs Cost of Trip

Total Transactions by Company

Yellow Cab is more popular than Pink Cab,
evidenced by the higher total number of transaction

Total Transaction Gender wise

- Across both Yellow Cab and Pink Cab, there are more transactions by males compared to females.

- However, Yellow Cab is preferred by both males and females.

State wise Income Generation by Cab Users

New York in generating the highest state income from cab users, followed by California, Illinois, and DC.
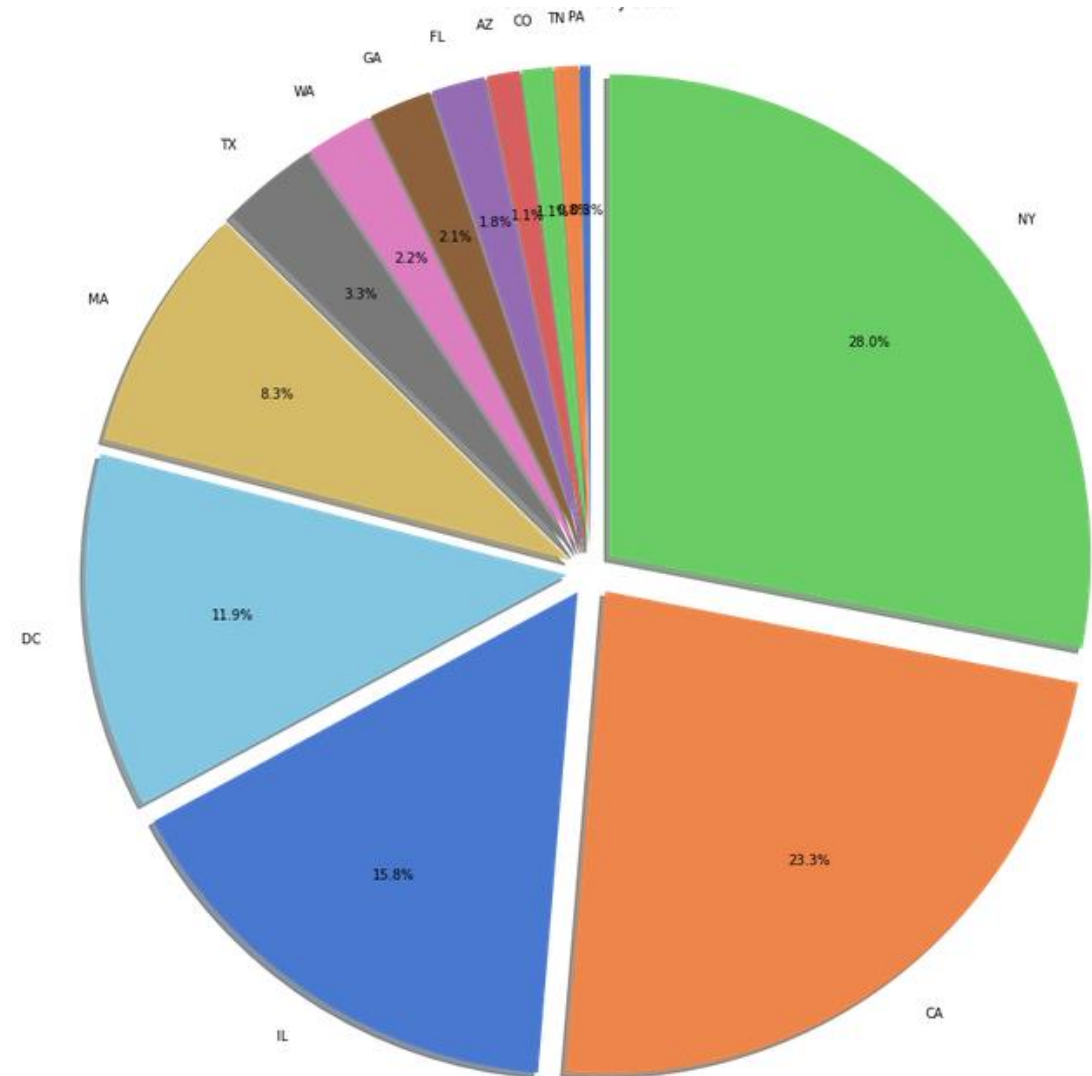
City wise Income Generation by Cab Users

- New York city generates highest income and number of cab users followed by Chicago, LA and Washington DC with yellow cab as the popular choice.

- Pink cab is preferred in 4 out of 19 cities:
  - San Diego
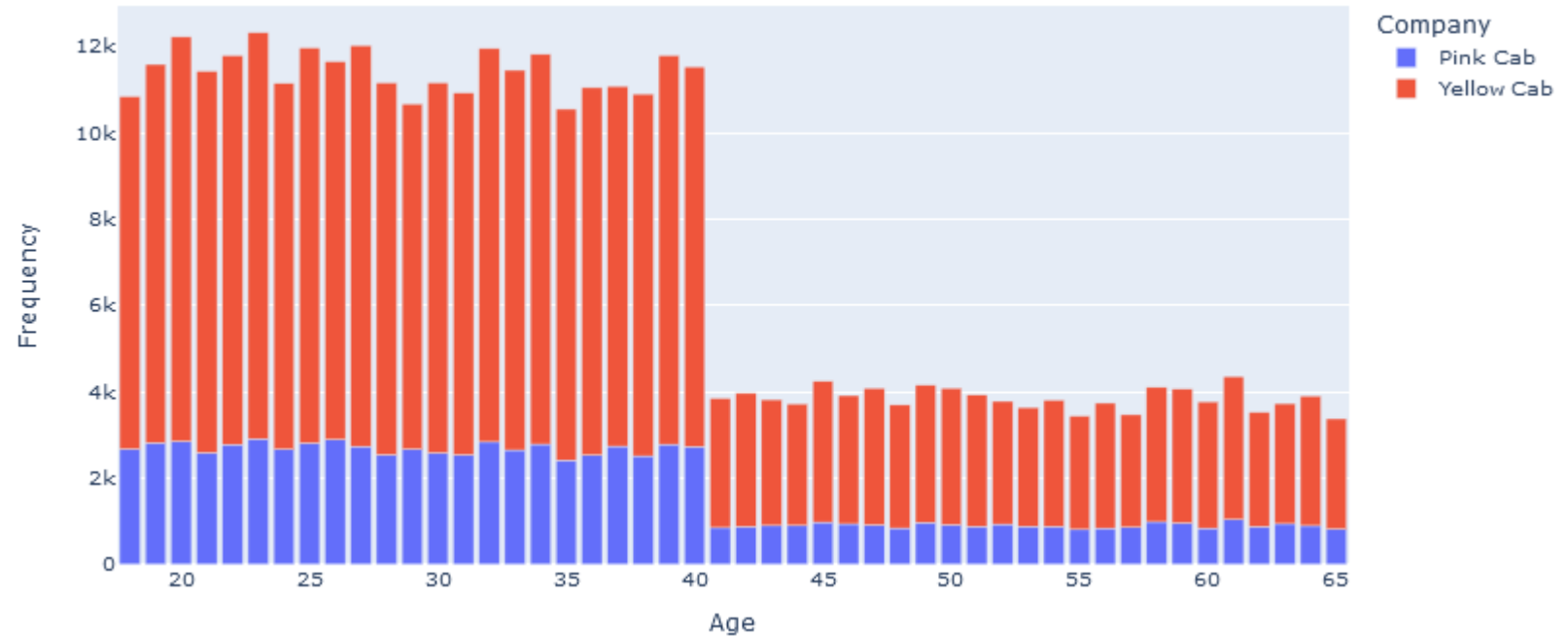 - Nashville
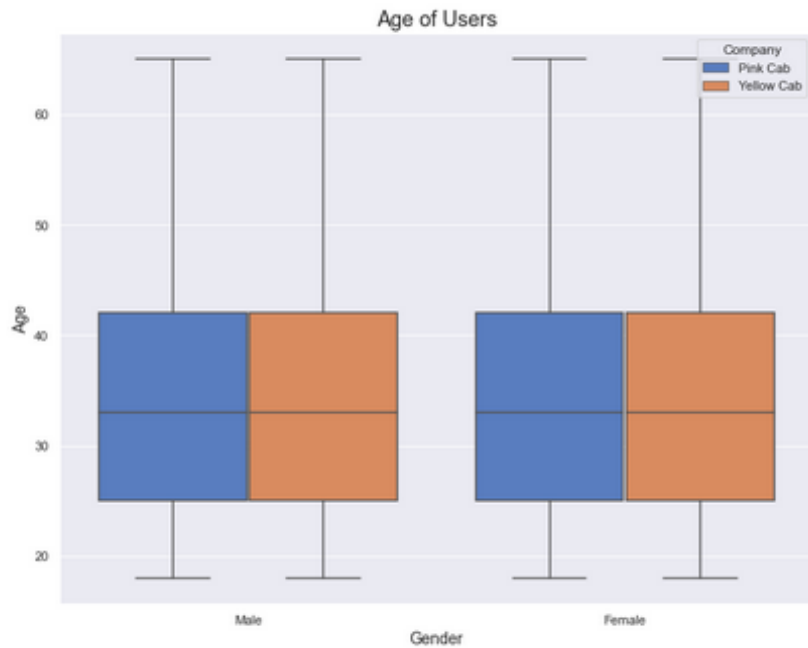 - Sacramento
  - Pittsburgh
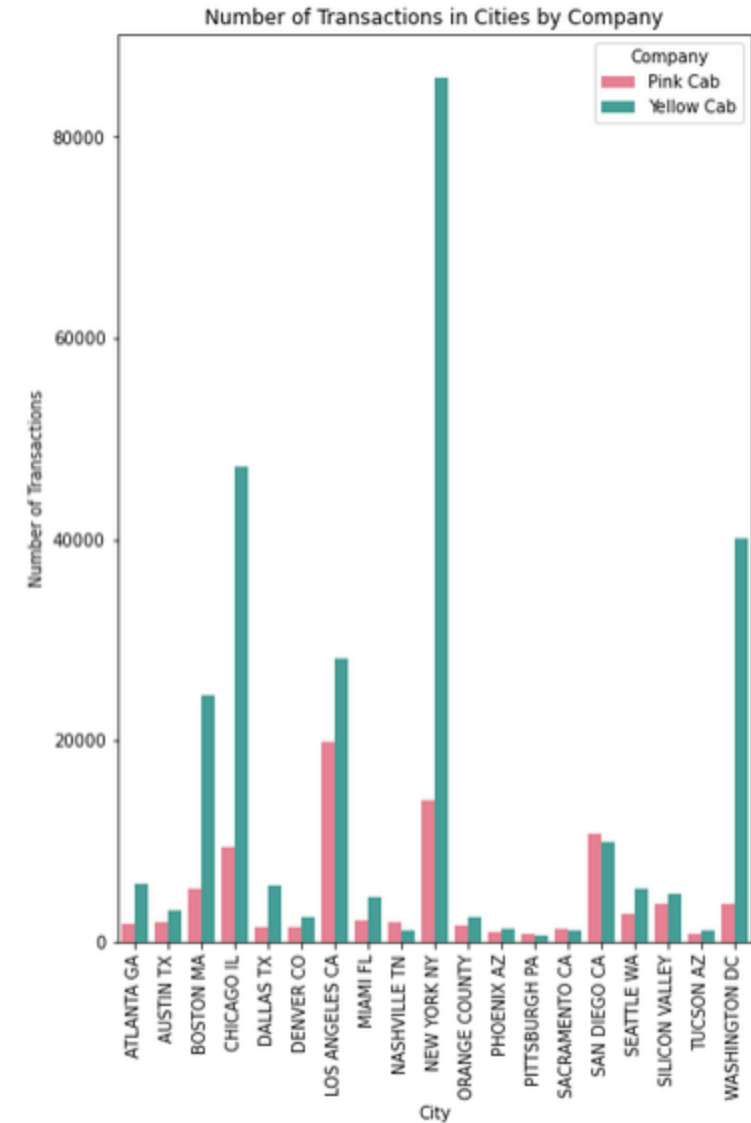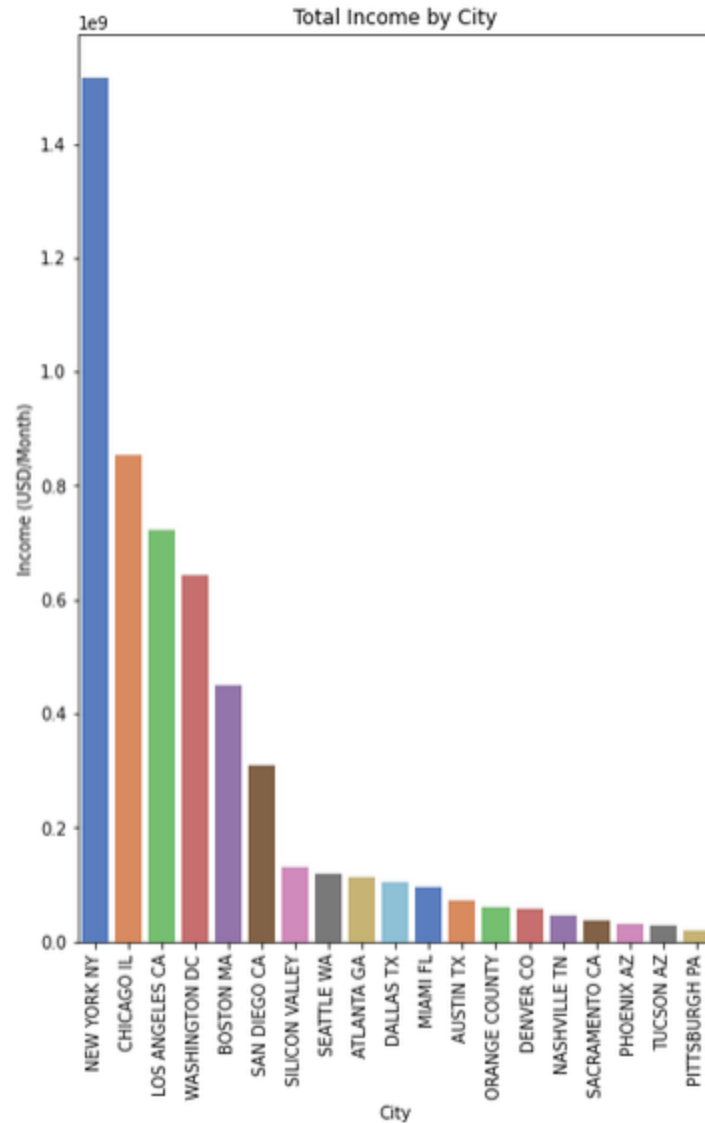
Users w.r.t Population in Cities

In cities such as San Francisco, Washington, and Boston, the cab service is utilized by over 30% of the population
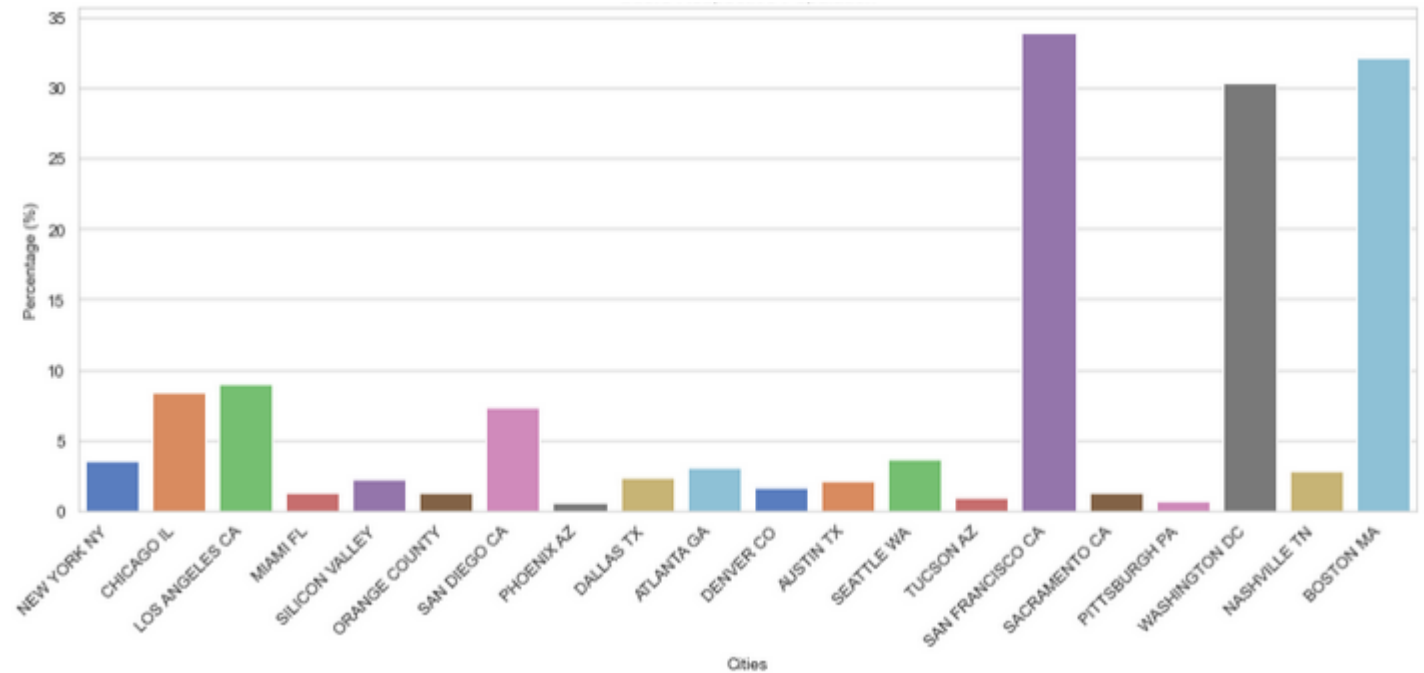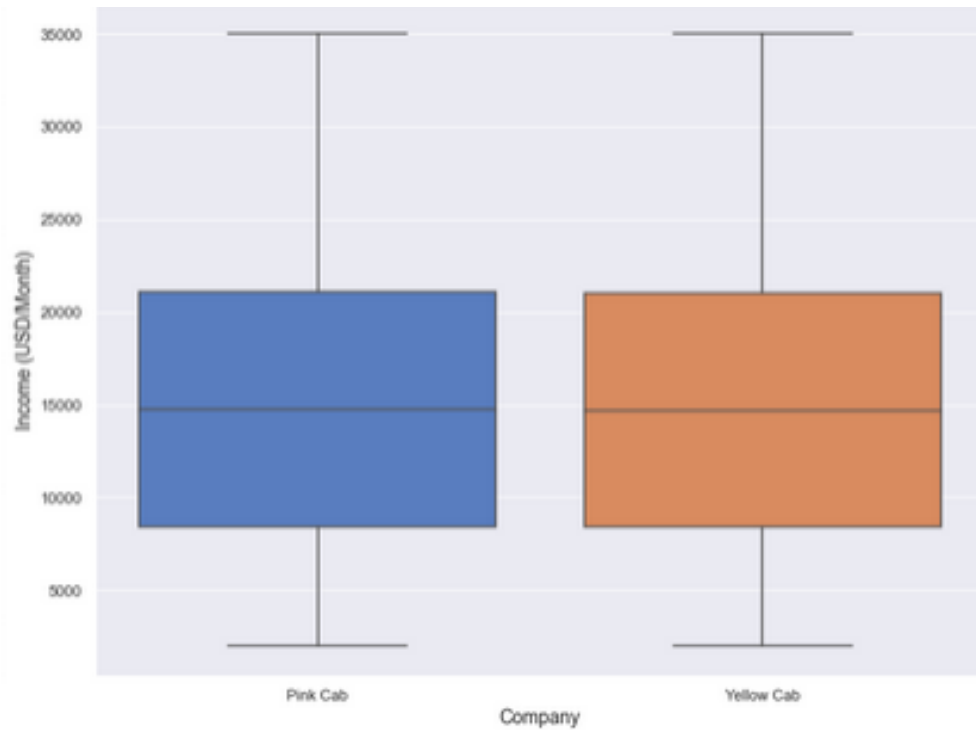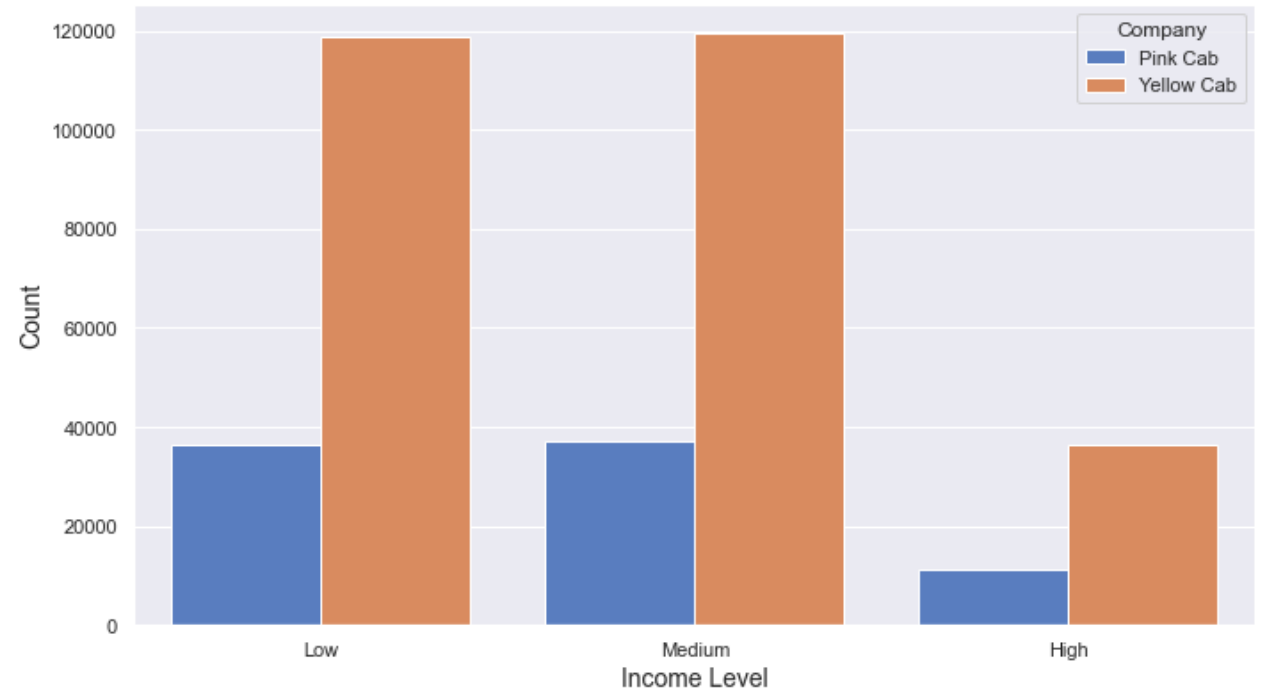
Age Wise Users Covered By Company

- Average age of users is 33 years.

- There is equal preference towards companies by the users.

- Yellow cab is more popular in all the age groups.

- Cab is more popular among users under 40 years of age.

User Income
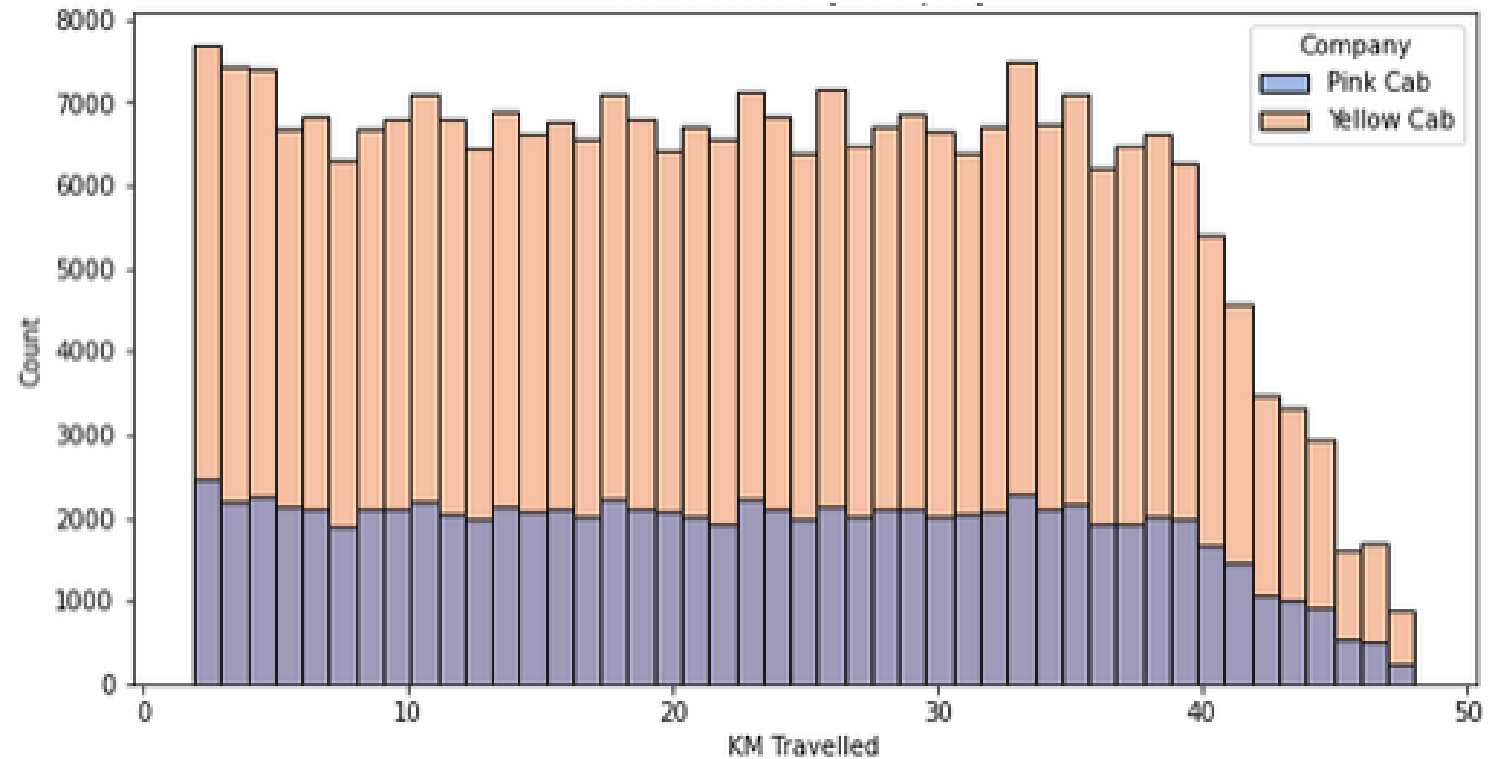

Average user income is $15000.

Yellow cab is popular among users from different income levels.

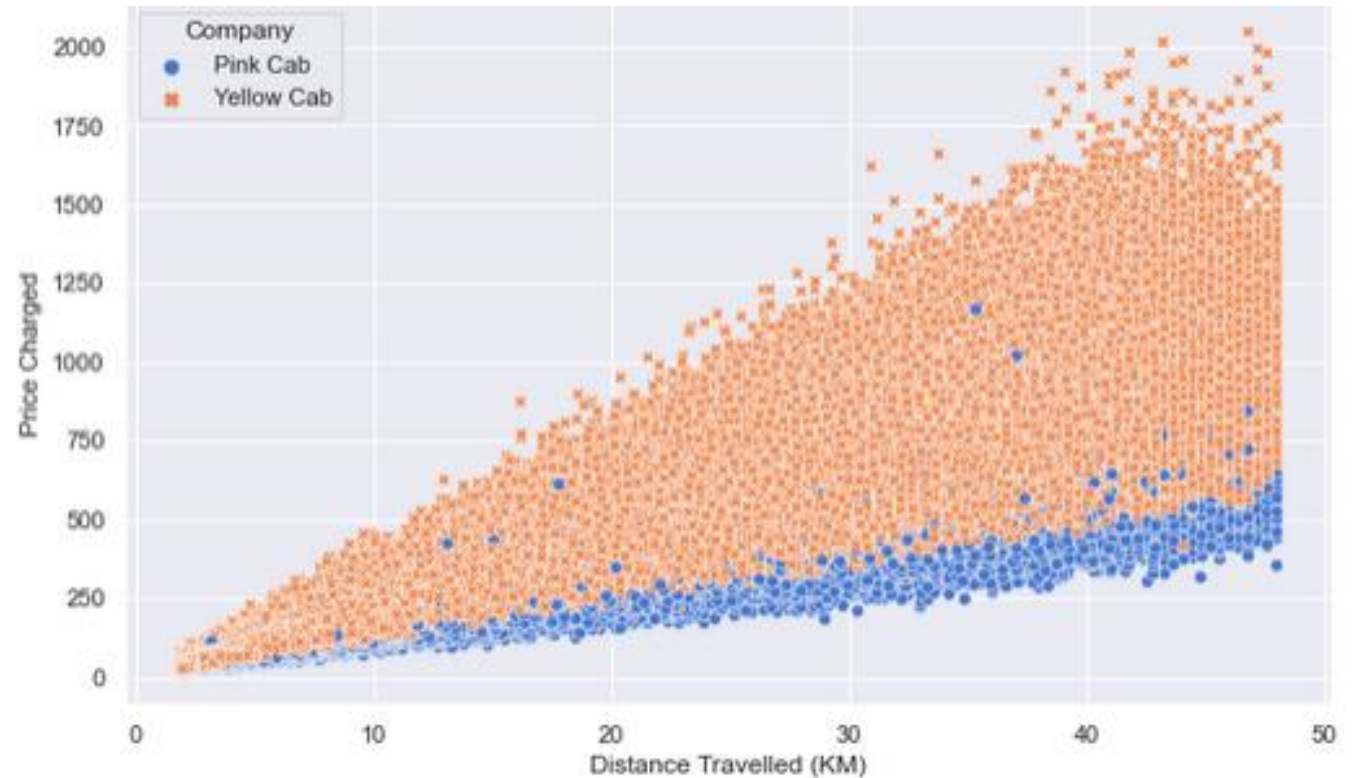# Distance Travelled by the Cab Companies

Most rides in the range from 2km to 48 km

Price Charged per Distance Travelled

- There is a linear relationship between the distance traveled (KM) and the price charged, as anticipated.
- Yellow Cab tends to have higher charges compared to Pink Cab.

Quarterly Profit by the Cab Companies
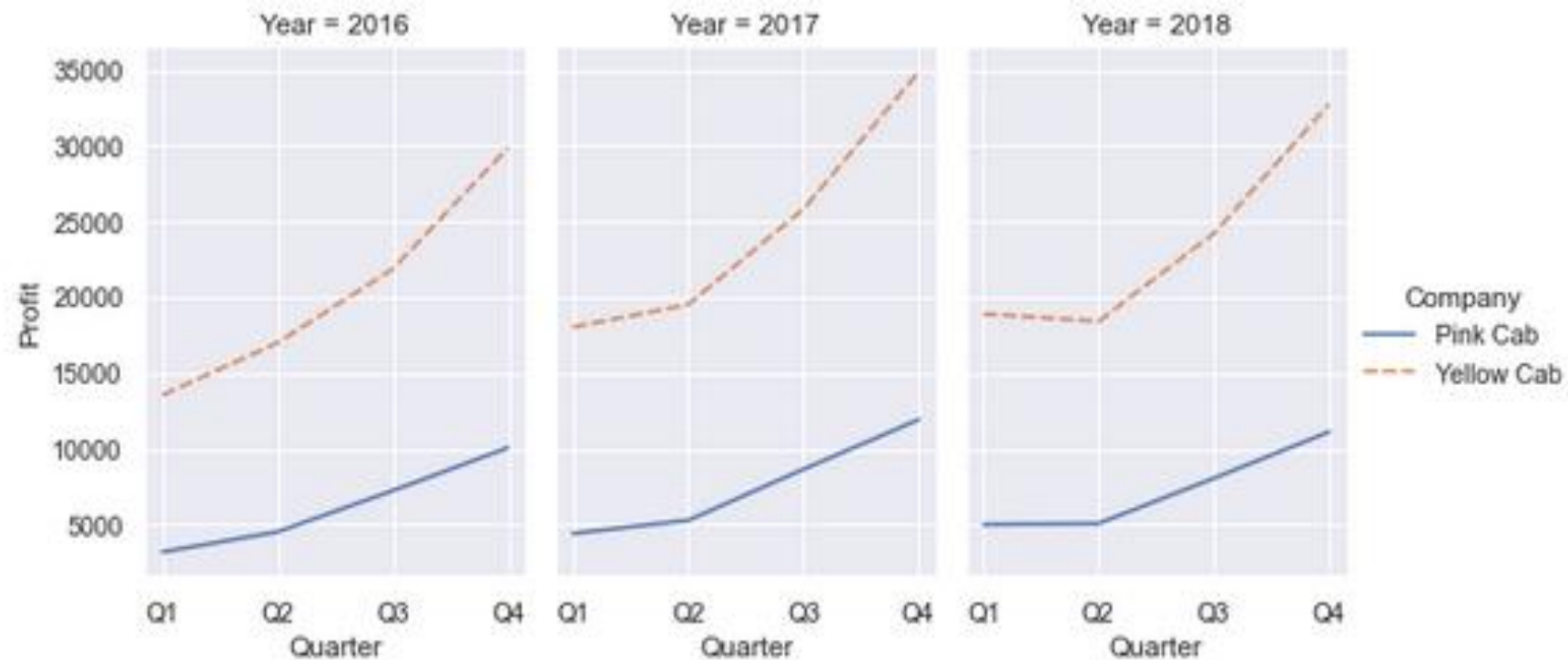
Yellow Cab has been making more profit compared to Pink Cab

Profit Margin by the Cab Companies

Profit by Years

Profit Margin and Revenue

Profit by Months

- Profit margin decreases by years for both companies.
- Profit margin varies with months.
- Overall, Yellow Cab makes thrice more profit than the Pink Cab.

Multiple Hypothesis Testing

Hypothesis 1: Is there any difference in profit regarding Gender

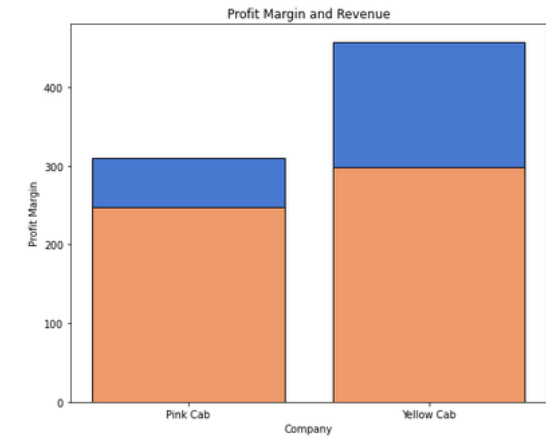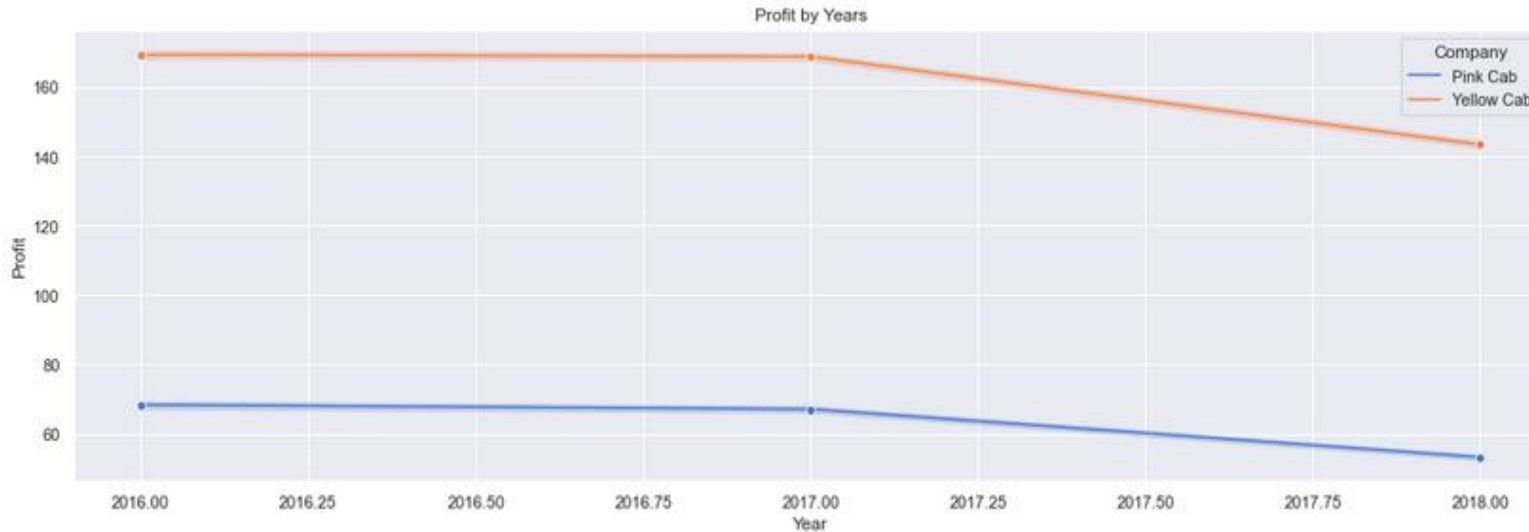H0: there no a difference in profit regarding gender

 H1: there is a difference in profit regarding gender

Cab Company	P-value	Conclusion
Pink	0.1	Fail to reject H0
Yellow	6.06e-25	Reject H0

Hypothesis 2: Is there a difference in profit between Pink Cab and Yellow Cab companies?

H0: there no a difference in profit between Pink Cab and Yellow Cab companies

H1: there is a difference in profit between Pink Cab and Yellow Cab companies

P-value	Conclusion
0.0	Reject H0

Hypothesis 3: Does the day of the week affect the number of transactions?

H0: the day of the week does not affect number of transactions

H1: the day of the week affects the number of transactions

P-value	Conclusion
1.0	Fail to reject H0

Multiple Hypothesis Testing

Hypothesis 4: Is there a significant difference in the average income of customers between Pink Cab and Yellow Cab companies?
H0: there is no significant difference
H1: there is a significant difference

P-value	Conclusion
0.67	Fail to reject H0

Hypothesis 5: Are there any differences in profit between different age groups?
H0: there are no differences in profit between different age groups
H1: there are differences in profit between different age groups

P-value	Conclusion
6.87e-09	Reject H0

Hypothesis 7: Is there a difference in profit between transactions paid with cash and those paid with card?
H0: there is no difference
H1: there is a difference

Cab Company	P-value	Conclusion
Pink	0.79	Fail to reject H0
Yellow	0.29	Fail to reject H0

Recommendation: Investment Preference

Based on the evaluation of both cab companies across various metrics, it is evident that Yellow Cab outperforms Pink Cab in several key aspects:

- **Customer Reach**: Yellow Cab demonstrates a wider customer reach, serving customers in 25 cities compared to Pink Cab's reach in only 4 cities. Additionally, Yellow Cab has shown better coverage of other cab users.

- **City Reach**: Yellow Cab is notably popular in 15 out of the 19 cities analyzed, indicating a stronger presence across a broader geographic area.

- **Age-wise Reach**: Yellow Cab caters to customers across all age groups, including the 60+ age group, showcasing a more diverse customer base compared to Pink Cab.

- **Average Profit per KM**: Yellow Cab boasts an average profit per kilometer nearly three times higher than that of Pink Cab, reflecting stronger profitability.

- **Income-wise Reach**: While both cab companies are popular among high and medium-income classes, Yellow Cab excels in providing services across all income groups—low, medium, and high.

In conclusion, based on the above observations, we highly recommend Yellow Cab as the preferred investment option.

Thank You