



Data Glacier

Your Deep Learning Partner

G2M insight for Cab Investment firm

Company Name : XYZ

Location: The US

Team: Data Science

Date: 21-June-2023

Agenda

Executive Summary

Problem Statement

Approach

EDA

Hypothesis Creation

Recommendation

Description:

- ❑ **XYZ is a private equity 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.**
- ❑ **Provide actionable insights to help XYZ firm in identifying the right company for making investment.**
- ❑ **Cab Companies:**
 - **Yellow Cab**
 - **Pink Cab**
- ❑ **The Analysis include :**
 - **Data Understanding,**
 - **Data Visualization,**
 - **Creating multiple hypothesis,**
 - **Building models and finding the best fit model based on Accuracy.**

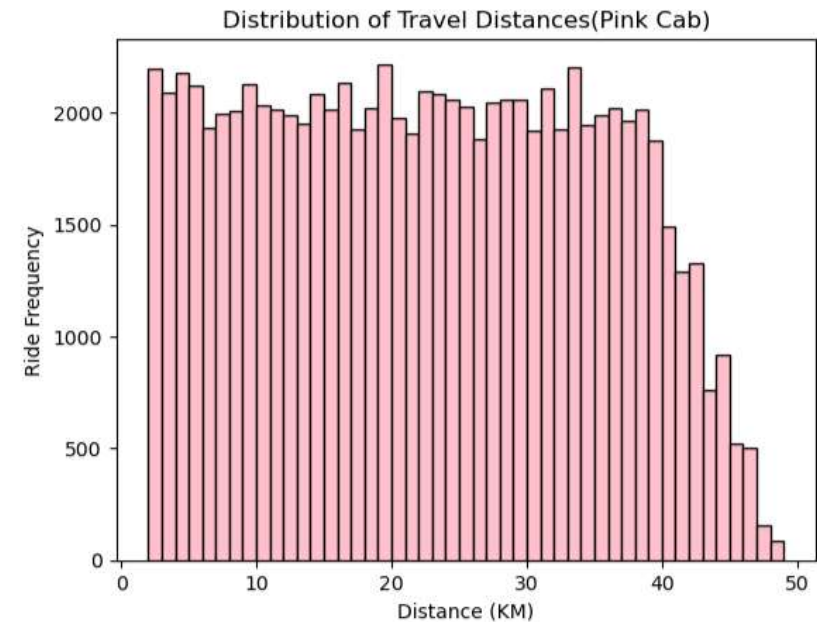
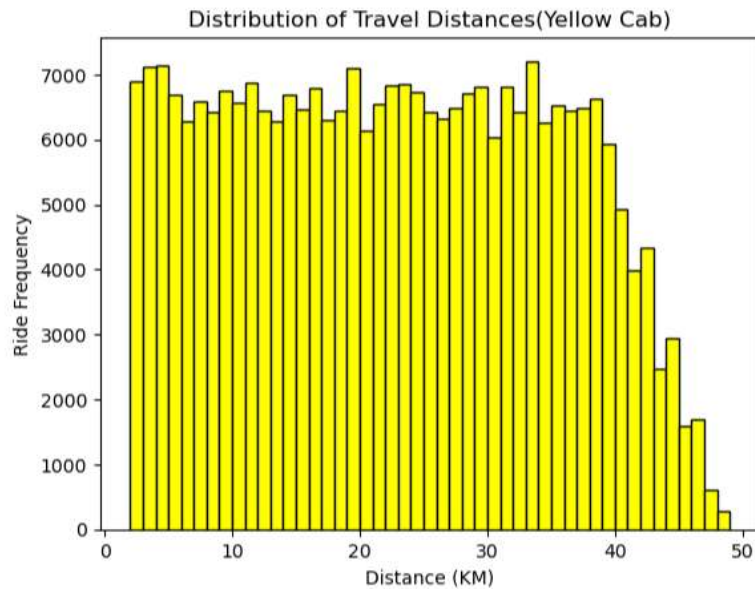
Data Preparation:

There are 5 datasets:

- **Cab_Data.csv** – this file includes details of transaction for 2 cab companies.
- **Customer_ID.csv** – this is a mapping table that contains a unique identifier which links the customer's demographic details.
- **Transaction_ID.csv** – this is a mapping table that contains transaction to customer mapping and payment mode.
- **City.csv** – this file contains list of US cities, their population and number of cab users.
- **US_Holidays.csv** – this file includes details of holiday dates in the US.

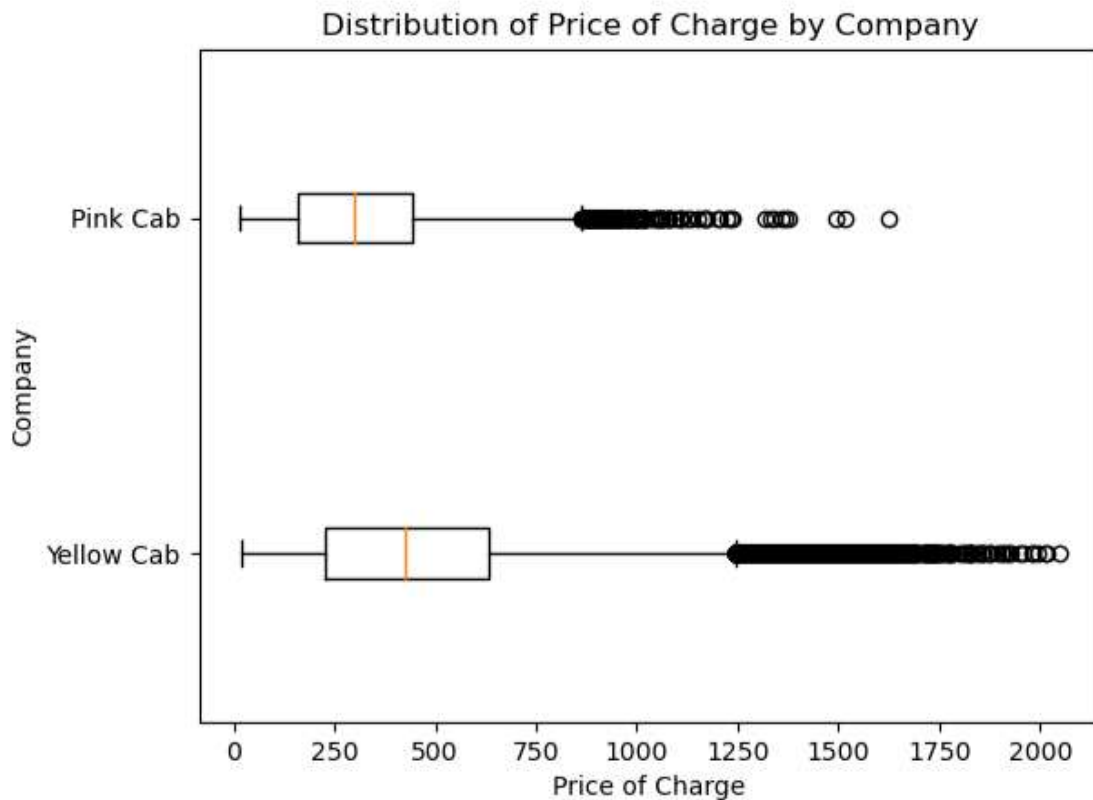
EXPLORATORY DATA ANALYSIS

Distribution of Travel Distances (KM)



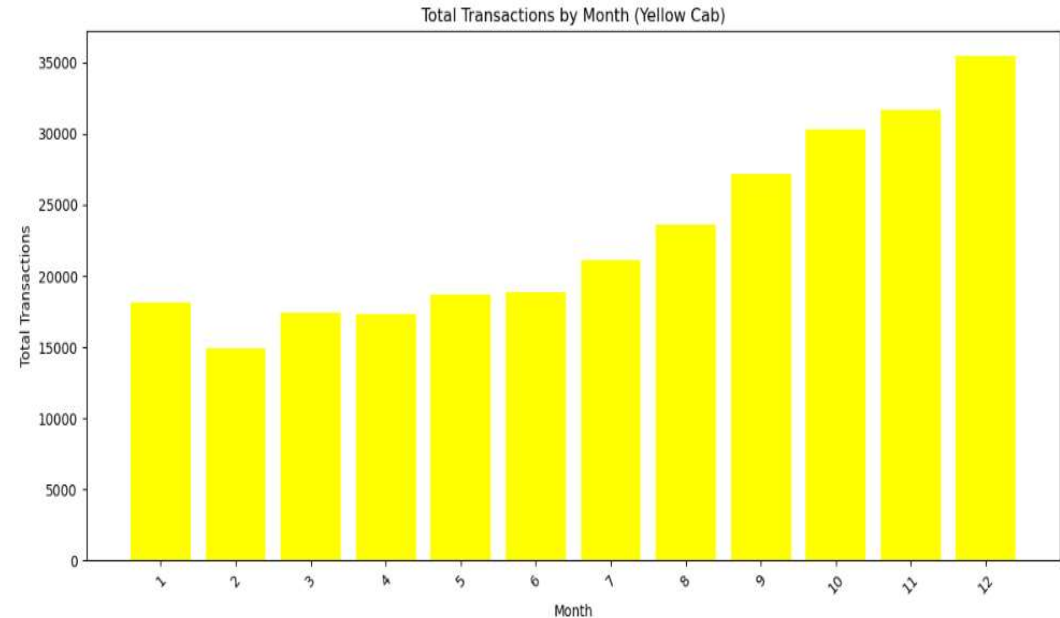
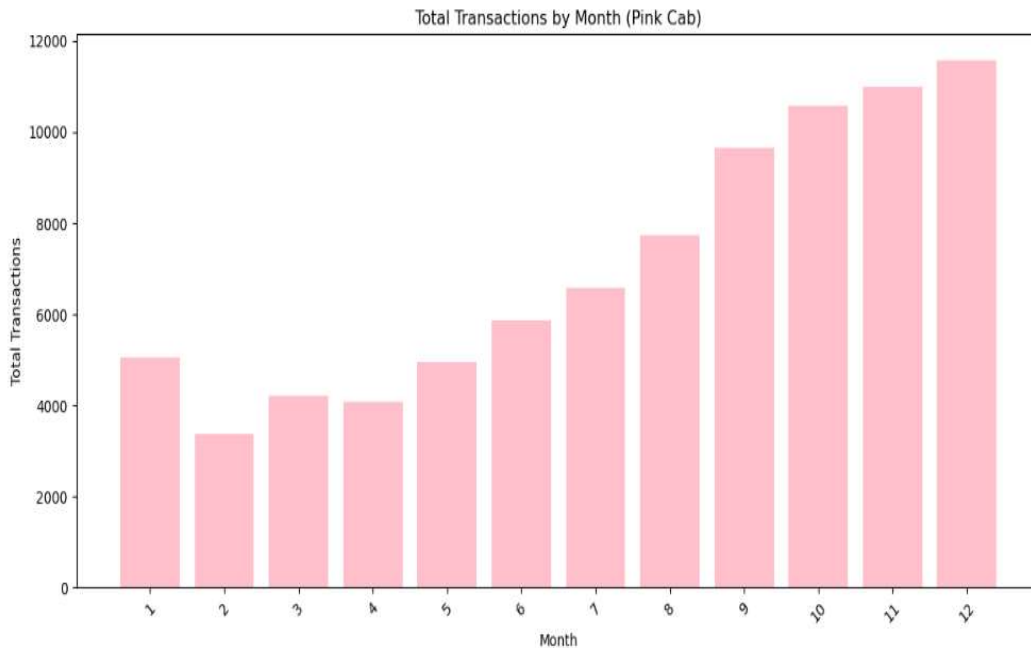
- ❑ Based on the information on these graphs, we can say that both companies (Pink and Yellow Cab) performs their rides approximately between 2 - 48 KM.

Distribution of Price Charged



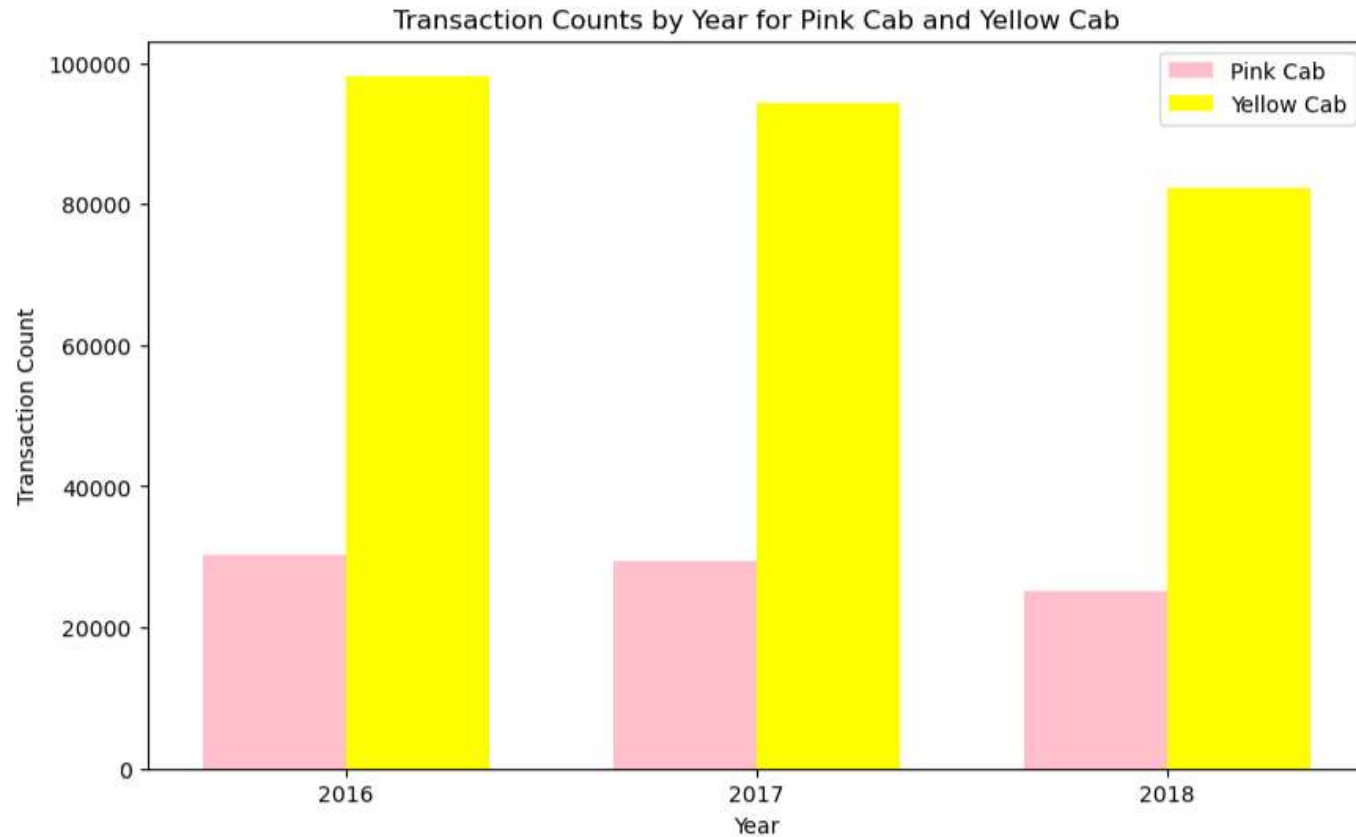
- ❑ The Price Charge range for Yellow cab is more than the Pink cab.
- ❑ The outliers are mostly due to use of high-end cars.

Travel Frequency per Month:



❑ **Yellow Cab performs more transactions (35000) in December December which is the holiday season compared to Pink Cab (11000).**

Transaction per Year



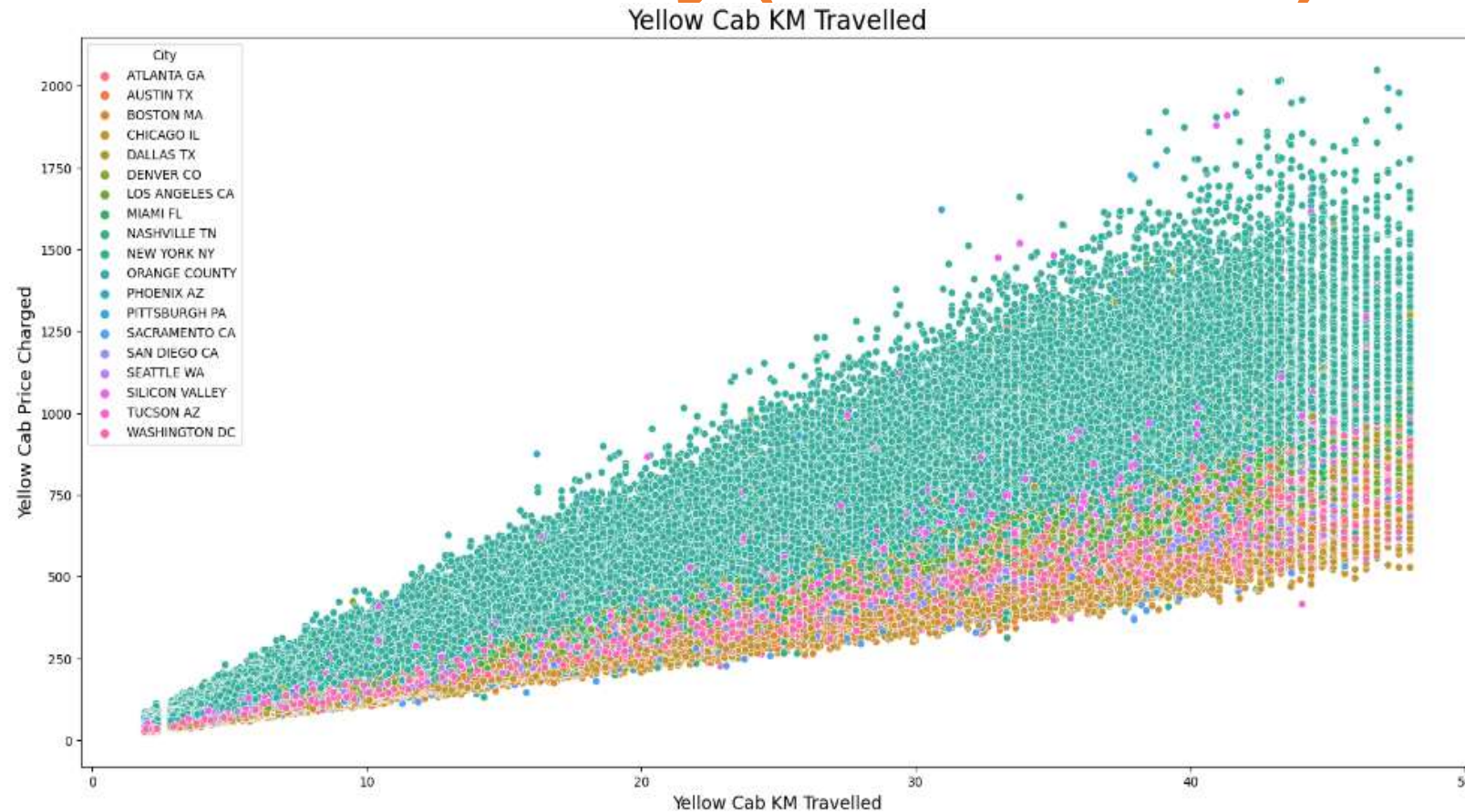
❑ Based on years, number of yearly transactions for Yellow cab is higher than Pink cab.

Price Charged based on Travel Distances per City (Pink Cab)



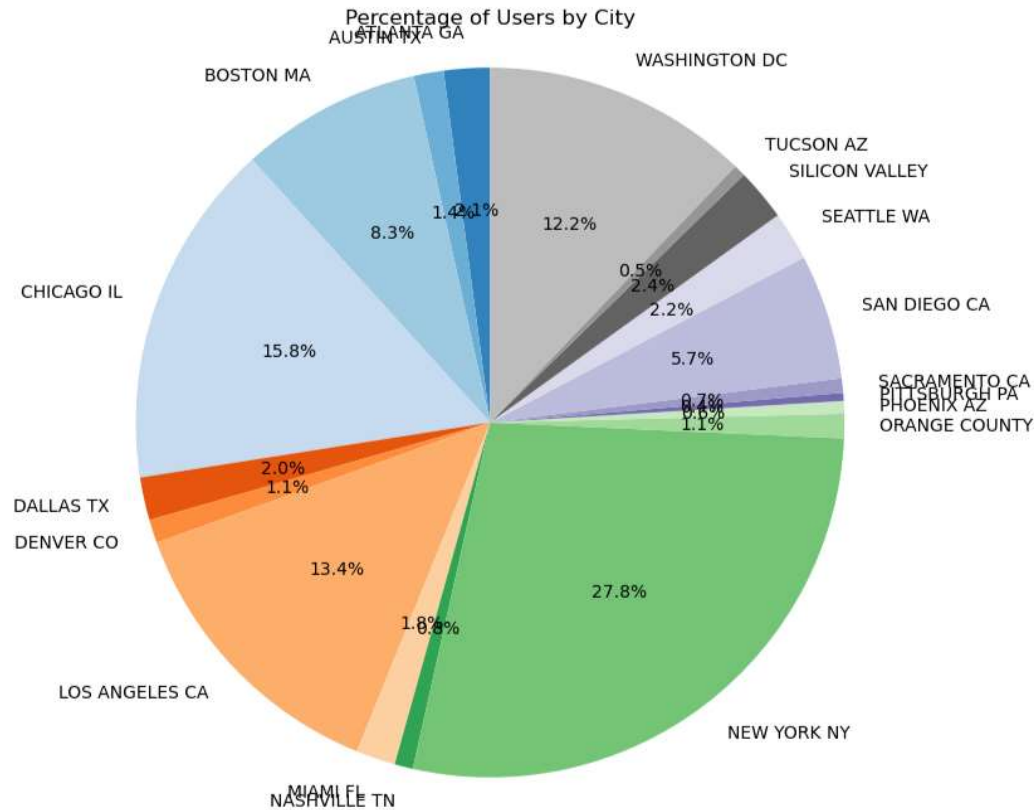
- ❑ All cities have the same increase in price depending on the increase in distance.

Price Charged based on Travel Distances per City (Yellow Cab)



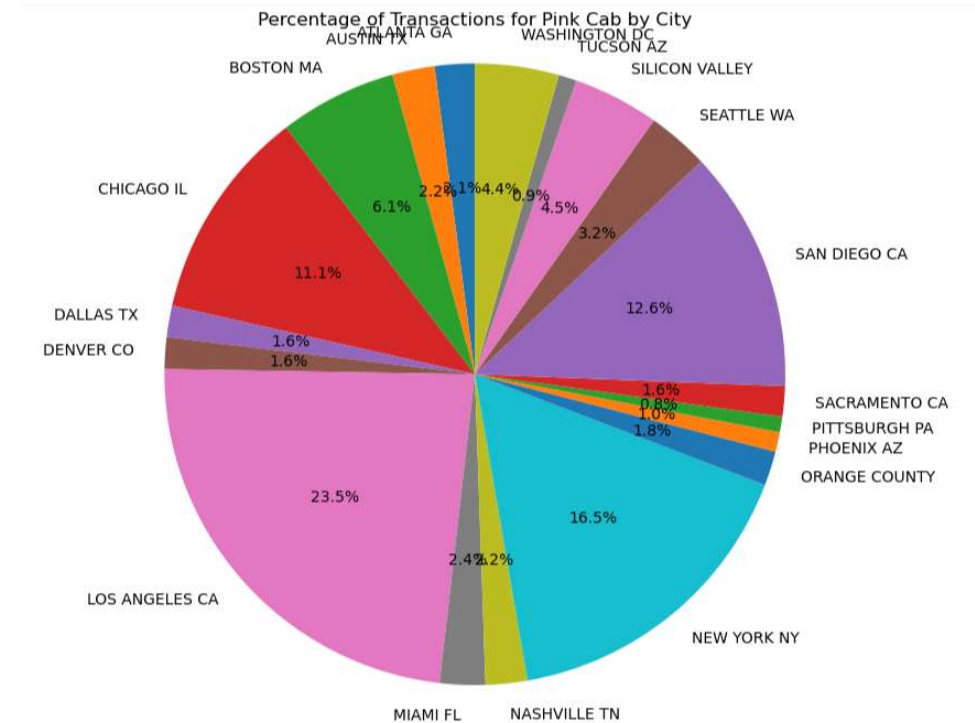
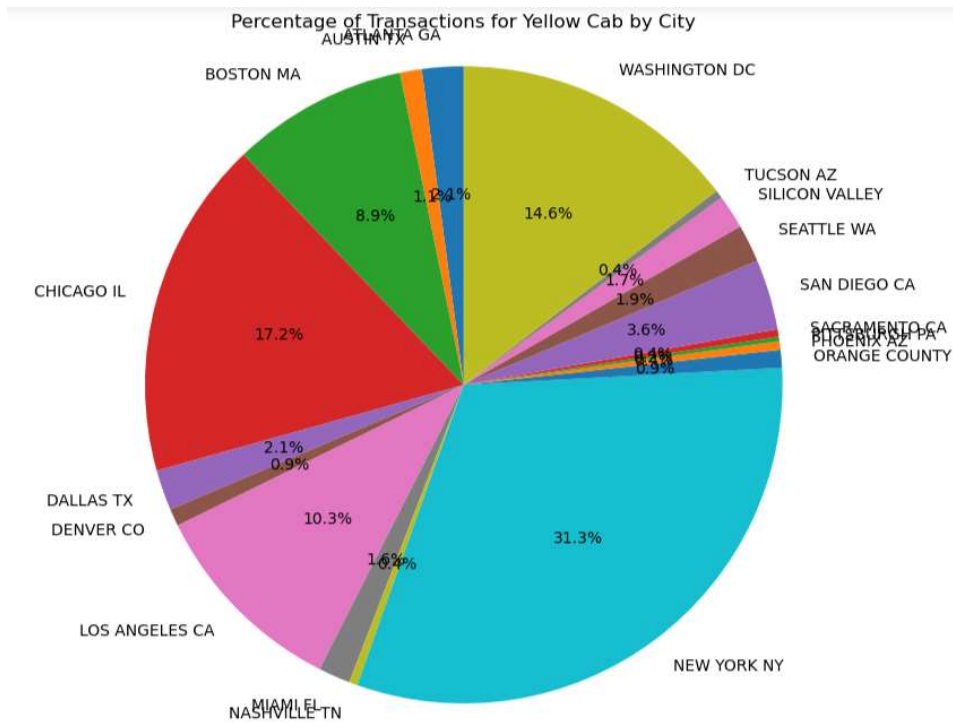
❑ In New York City the Price charged is more than other cities for Yellow Cab.

Cab User Distribution Per City



❑ New York City has the highest Cab users with 28%. 2. city with the highest user is Chicago by 16% followed by Los Angeles with 13%.

Transaction per City for both Cabs

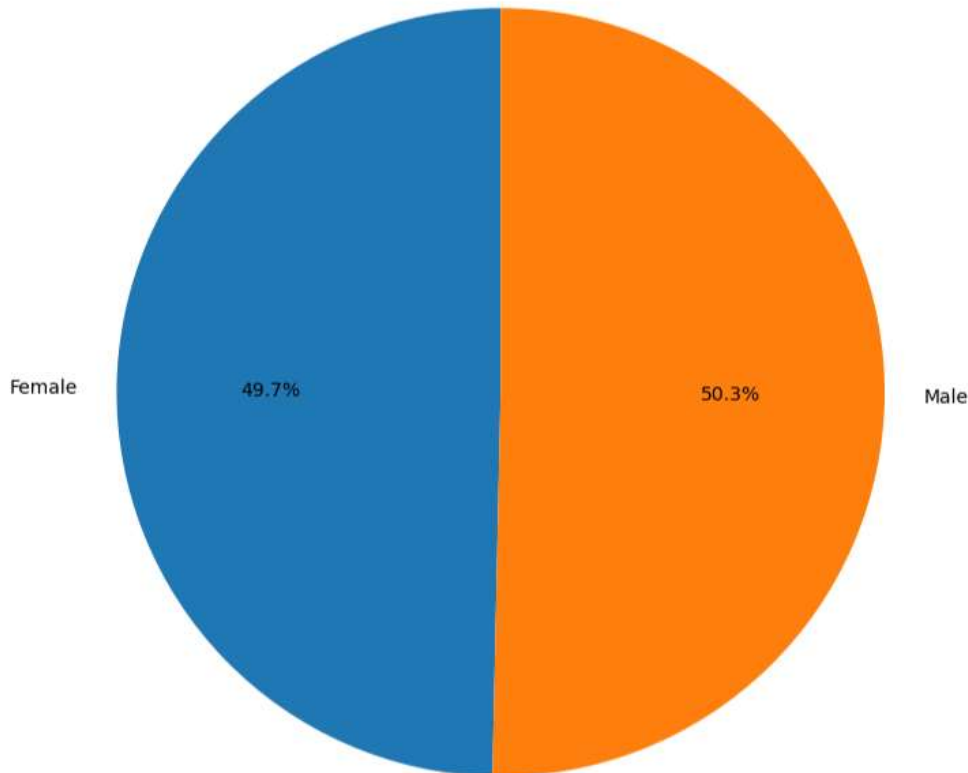


❑ For Yellow Cab transaction count is highest in New York City(31%), we shall recall that New York is the city that has the most Yellow Cab user.

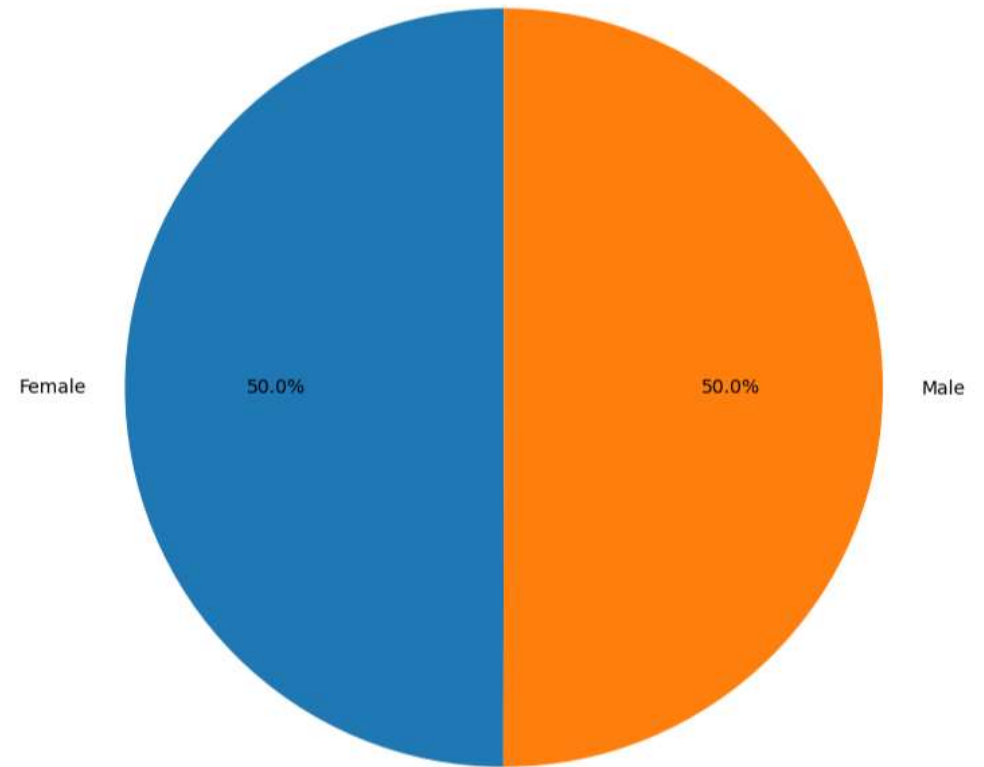
❑ For Pink Cab transaction count is highest in Los Angeles City.

Price Charged Based on Gender

Average Price Distribution by Gender for Yellow Cab



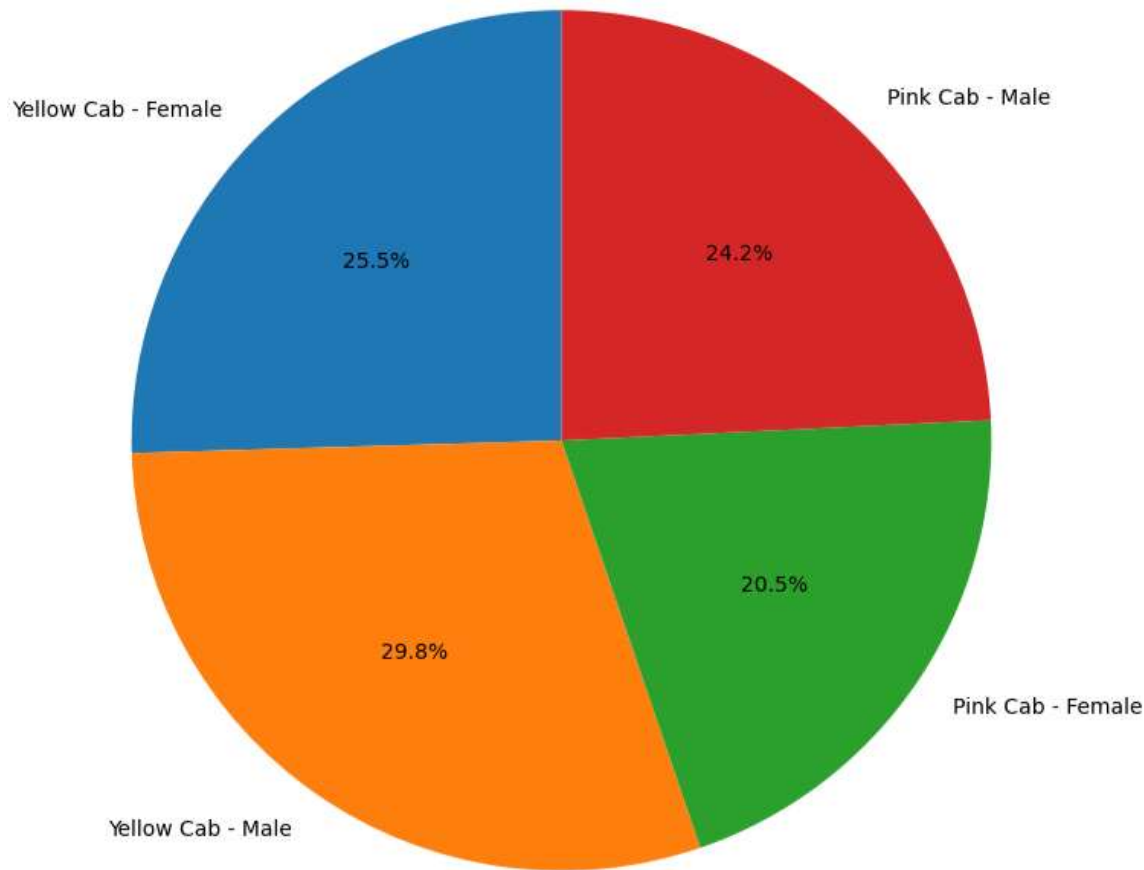
Average Price Distribution by Gender for Pink Cab



❑ Yellow Cab charge less price for Female Customers while Pink Cab charges same price for both genders.

Customer Distribituons Based on Gender for Both Companies

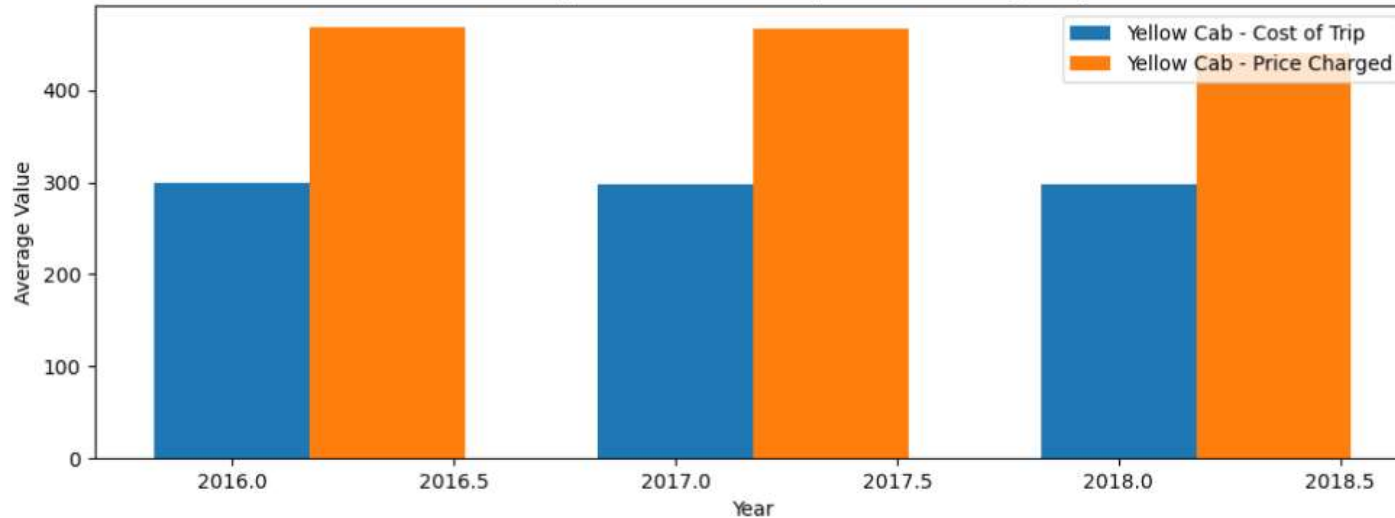
Gender Distribution for Yellow Cab and Pink Cab



- ❑ **Number of Female Customers in Yellow Cab(25.5%) is higher than Pink Cab (20.5%).**
- ❑ **We shall recall that yellow cab charges less price from females.**

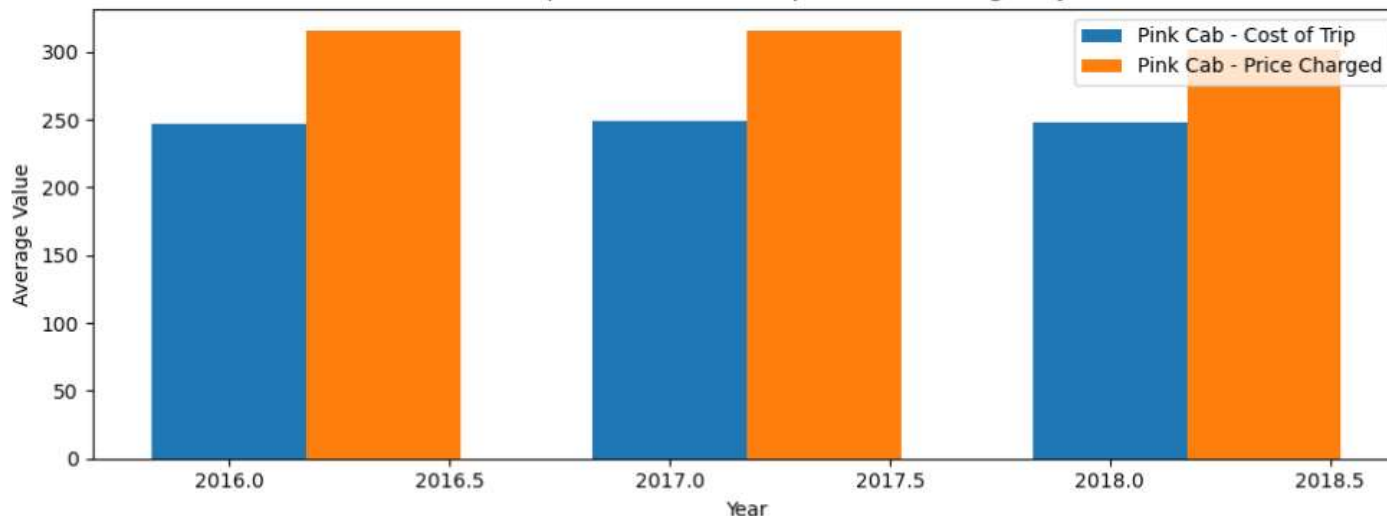
Profit Margin per year for Both Companies

Yellow Cab - Comparison of Cost of Trip and Price Charged by Year



❑ Yellow cab has higher Profit Margin than that of the Pink cab.

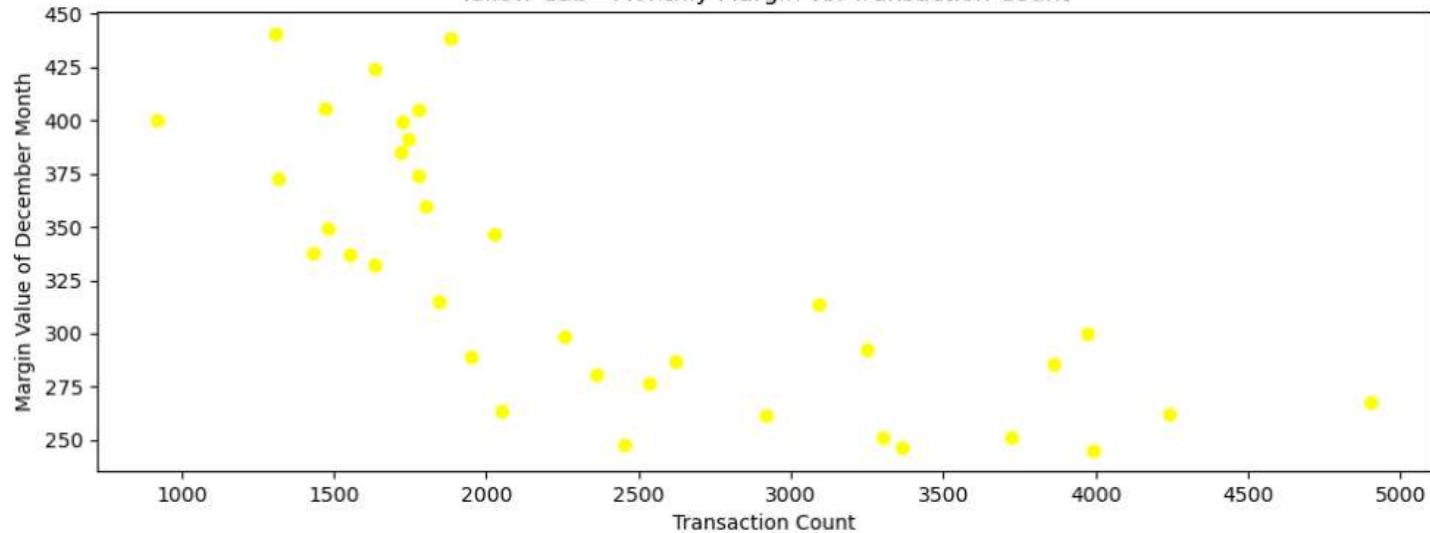
Pink Cab - Comparison of Cost of Trip and Price Charged by Year



❑ Profit Margin = Price Charged – Cost of the Trip

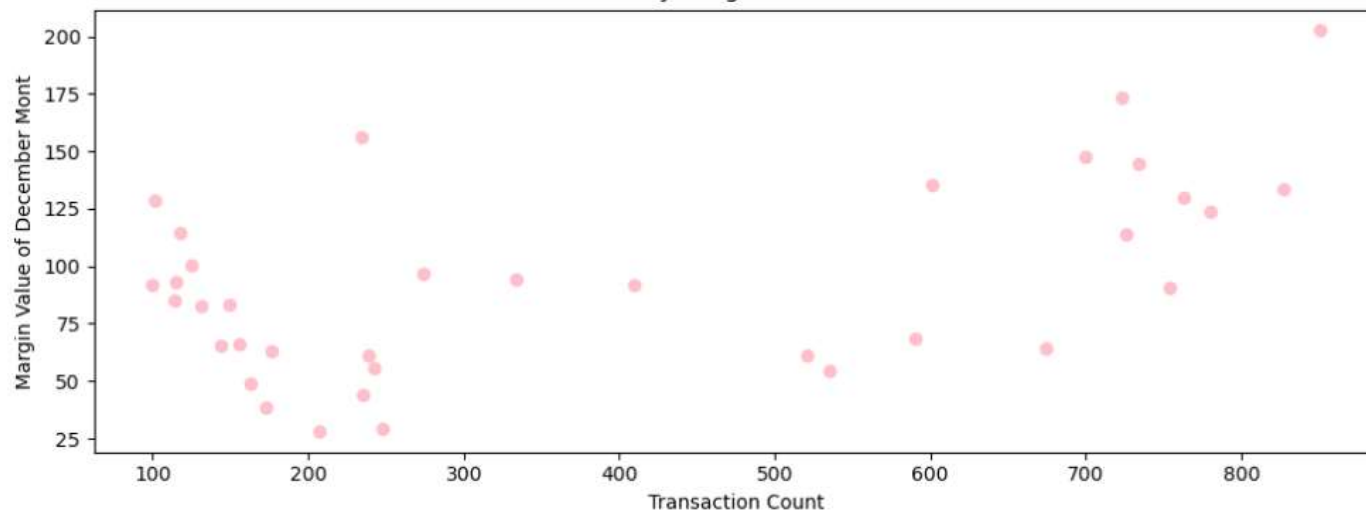
Margins per Transactions

Yellow Cab - Monthly Margin vs. Transaction Count



☐ **Pink Cab company shows increasement in margins as the number of transactions increases.**

Pink Cab - Monthly Margin vs. Transaction Count



☐ **Yellow Cab company shows decreasement int Margins as the number of transactions increases.**

Correlation



❑ **According to the graph →**

□ There is a positive correlation between Margin and Price Charged

Hypothesis Creation

Hypothesis → Average margin does not change depending on the city for both pink cab and yellow cab companies

Reject the null hypothesis.

There is a significant difference in margins depending on the city for Yellow Cab.

The P value is: 0.0

```
#For yellow cab, there is a significant difference in margin values depending on the city of the transaction.
```

Reject the null hypothesis.

There is a significant difference in margins depending on the city for Pink Cab.

The P value is: 0.0

```
#For pink cab, there is a significant difference in margins depending on the city of the transaction
```

Hypothesis → Average margin does not change depending on the payment method for both pink and yellow cab companies.

Fail to reject the null hypothesis.

There is no significant difference in margins depending on the payment method for Yellow Cab.

The P value is: 0.2933060638298729

```
# For yellow cab, there is no significant difference depending on the payment method of the customer.
```

Fail to reject the null hypothesis.

There is no significant difference in margins depending on the payment method for Pink Cab.

The P value is: 0.7900465828793286

```
# For pink cab, there is no significant difference in average margin values depending on the payment method of the customer.
```

Hypothesis → Average margin does not change depending on the gender for both pink cab and yellow cab companies

Reject the null hypothesis.

There is a significant difference in margins between genders for Yellow Cab.

The P value is: 6.060473042494056e-25

#For yellow cab company, there is a significant difference in margin value between the genders.

Fail to reject the null hypothesis.

There is no significant difference in margins between genders for Pink Cab.

The P value is: 0.11515305900425798

#For pink cab company, there is no significant difference in margin value between the genders.

Hypothesis → Average margin does not change depending on the age for both pink cab and yellow cab companies

Reject the null hypothesis.

There is a significant difference in margins depending on the Age of customers for Yellow Cab.

The P value is: 6.4942568177993685e-09

#For yellow cab, there is a significant difference in margin values depending on the age of the customers.

Fail to reject the null hypothesis.

There is no significant difference in margins depending on the Age of customers for Pink Cab.

The P value is: 0.32817487547980695

#For pink cab, there is no significant difference in margin values depending on the age of the customers.

Hypothesis → Average margin does not change depending on in which month transaction is performed for both yellow and pink cab company

Reject the null hypothesis.

There is a significant difference in margins depending on the month of the transaction for Yellow Cab.
The P value is: 1.993753103466109e-183

#For yellow cab, there is a significant difference in margins depending on the period of the year.

Reject the null hypothesis.

There is a significant difference in margins depending on the month of the transaction for Pink Cab.
The P value is: 1.9616227830629913e-293

#For pink cab, there is a significant difference in margins depending on the period of the year.

On the basis of analyzes done and results of hypotheses;

Yellow Cab Company is recommended for Investment.

Thank You



Data Glacier

Your Deep Learning Partner