Data Glacier

Your Deep Learning Partner

Exploratory Data Analysis

G2M Insight For Cab Investment Firm

20-july-2024

OBIDA ALHAMOUD

Problem Statement

- XYZ is a private equity firm in the US. Due to remarkable growth in the cab Industry in the last few years and multiple key players in the market , it is planning for an Investment in the Cab industry.

- Objective: Provide actionable insights to help XYZ firm in identifying the right company for making investment.

The analysis has been divided into four parts:

- Data Understanding and Visualizing
- Forecasting profit and number of rides for each cab type
- Finding the most profitable Cab company
- Recommendations for investment

Datasets Information

- **City.csv** – this datasets contains a list of US cities , their population, and the number of cab users.
- **Customer_ID.csv** – this dataset contains a list of customers with there unique ID's , Gender, Age and Income.
- **Transactions_df.csv** – this dataset contains a list of transaction of each customer, Transaction ID is unique and each transaction has a Payment method and its customer ID.
- **Cab_data.csv** – this dataset includes all the details of transactions for 2 cab companies

Data Cleaning and Type Conversion

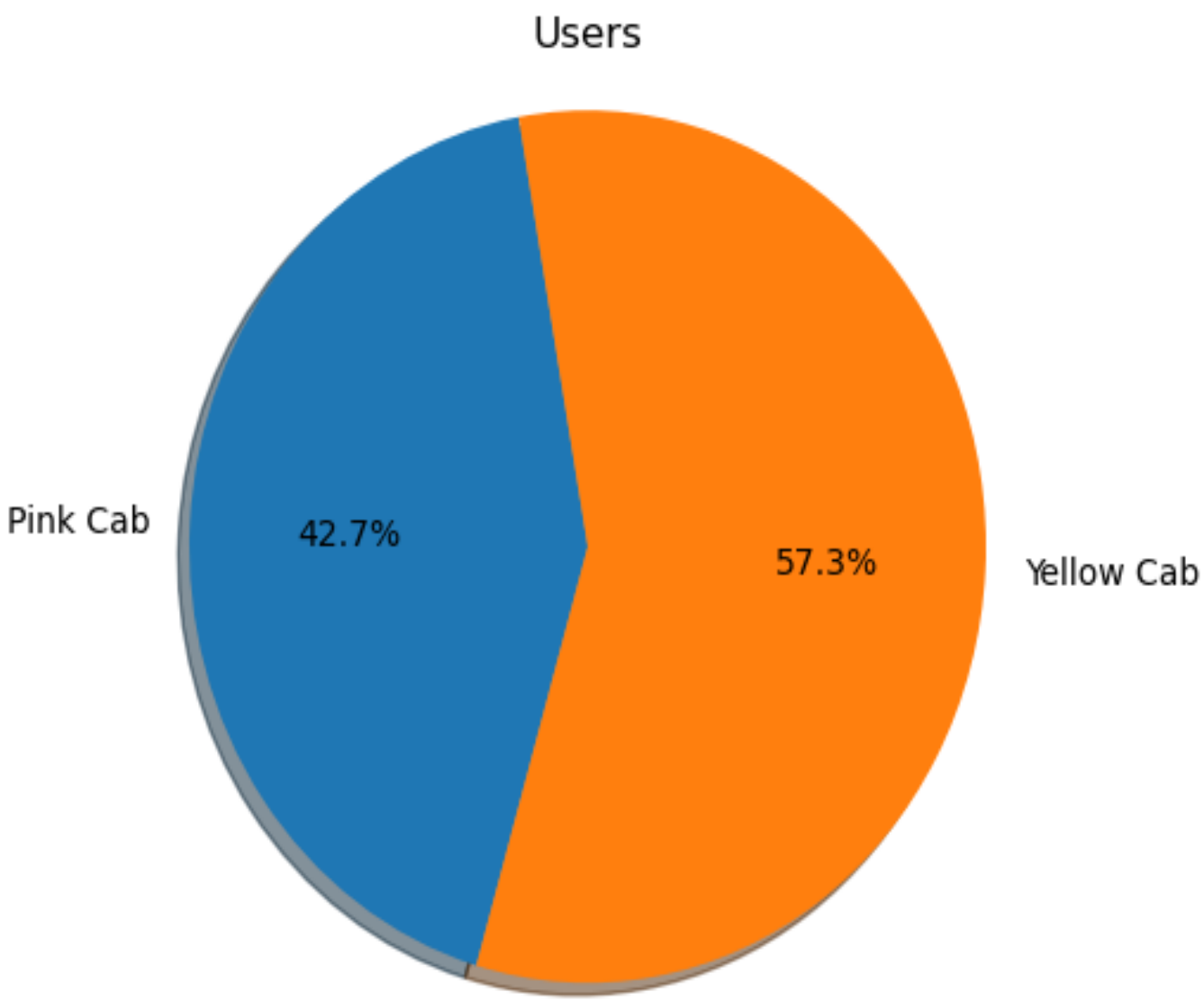
```python
import datetime

def convert_to_date(date):                          #function to convert serial dates to standart dates
    base_date = datetime.datetime(1899,12,30)
    return base_date + datetime.timedelta(days=date)

df['dates of travel'] = df['Date of Travel'].apply(convert_to_date)
```

Convert 'Price Charged' and 'Cost of trip' into integers

```python
def convert_to_integer(Number):                     ## this function converts a float to integers
    result = int(Number)
    return result

def convert_to_integerr(Number):                     ## this function converts an object to integers
    result = int(str(Number).replace(',', '').strip())
    return result



df['Population'] = df['Population'].apply(convert_to_integerr)
df['Users'] = df['Users'].apply(convert_to_integerr)
df['Price Charged'] = df['Price Charged'].apply(convert_to_integer)
df['Cost of Trip'] = df['Cost of Trip'].apply(convert_to_integer)
```

EDA

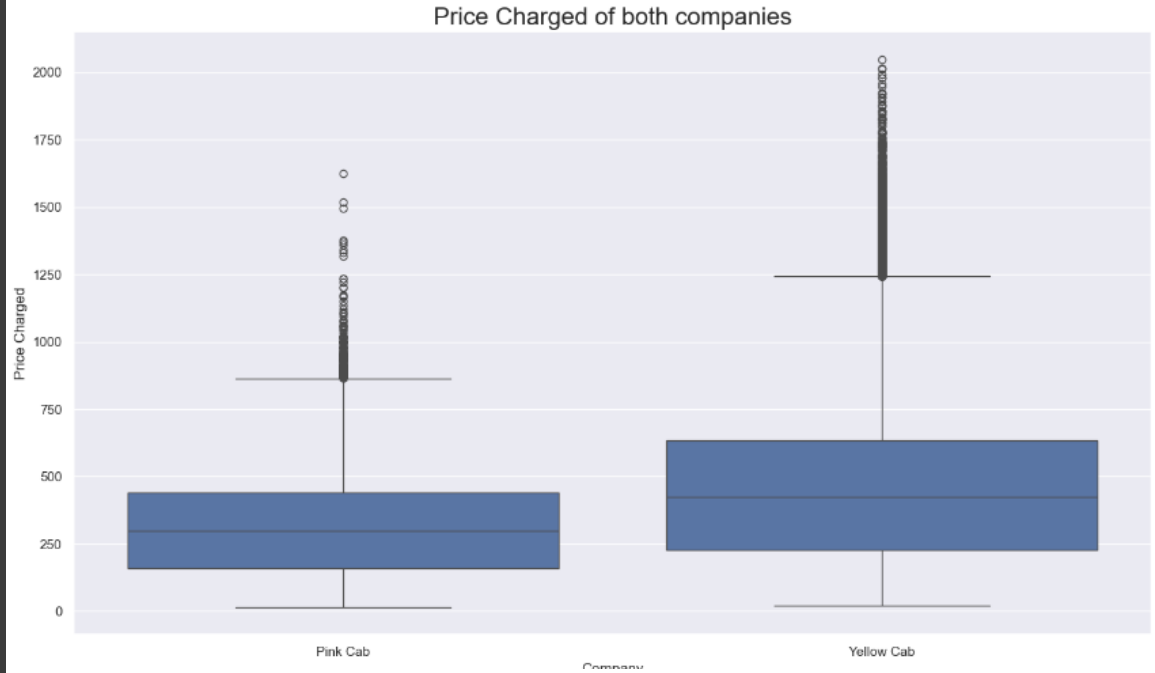
Which company has more Users?

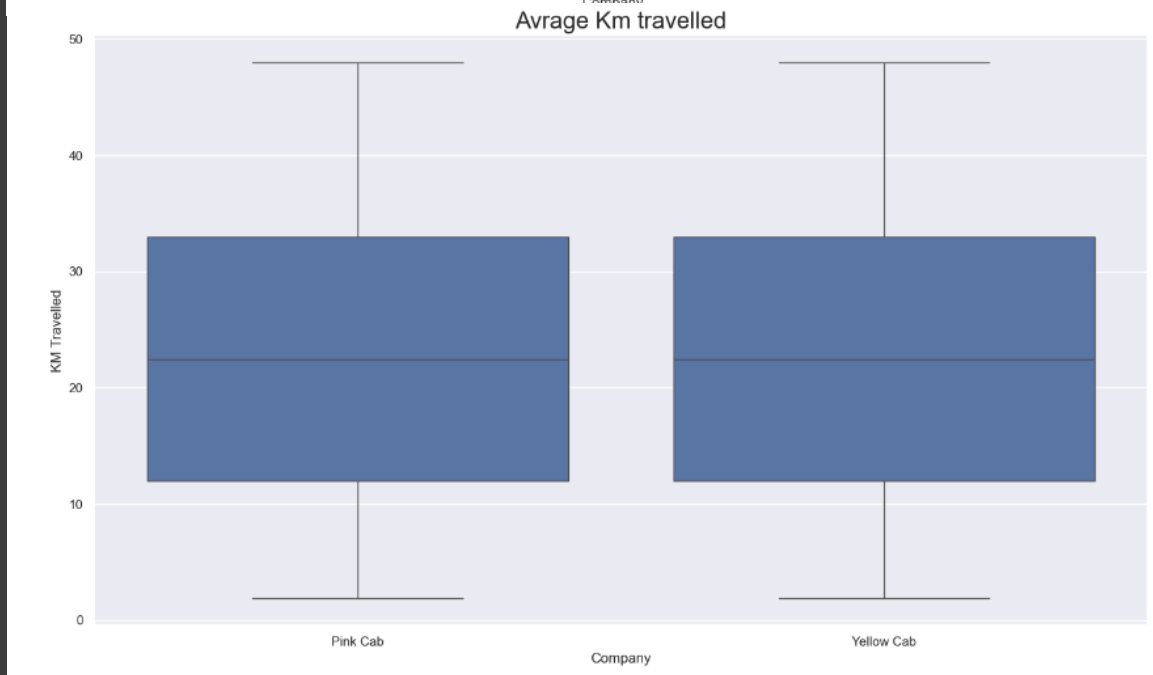


- Yellow Cab has more Users.

Data Glacier
Your Deep Learning Partner

Price Charged & KM Travelled

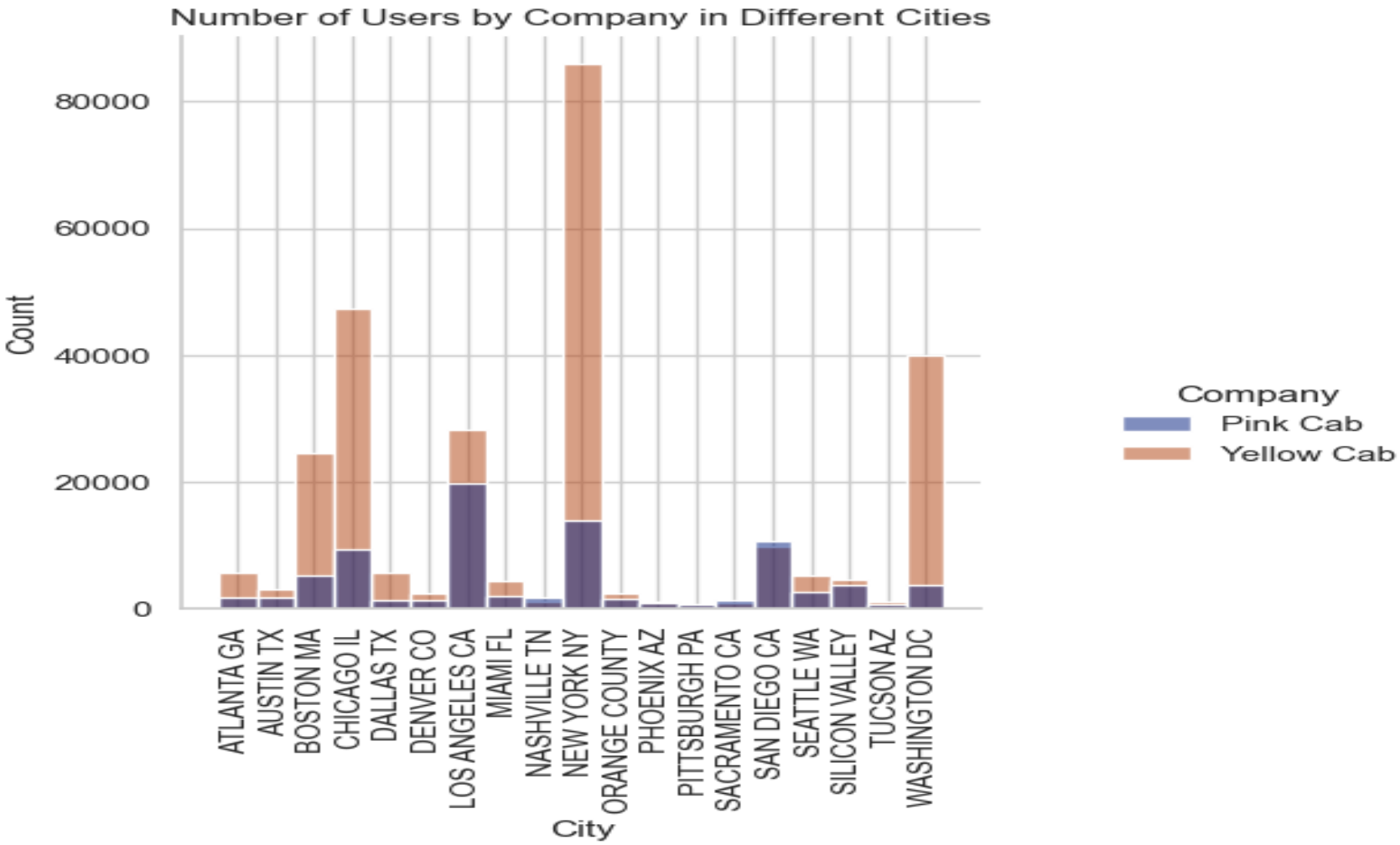




- Yellow cab charges a a little more compared to Pink Cap
- The average KM Travelled are the same

Data Glacier
Your Deep Learning Partner

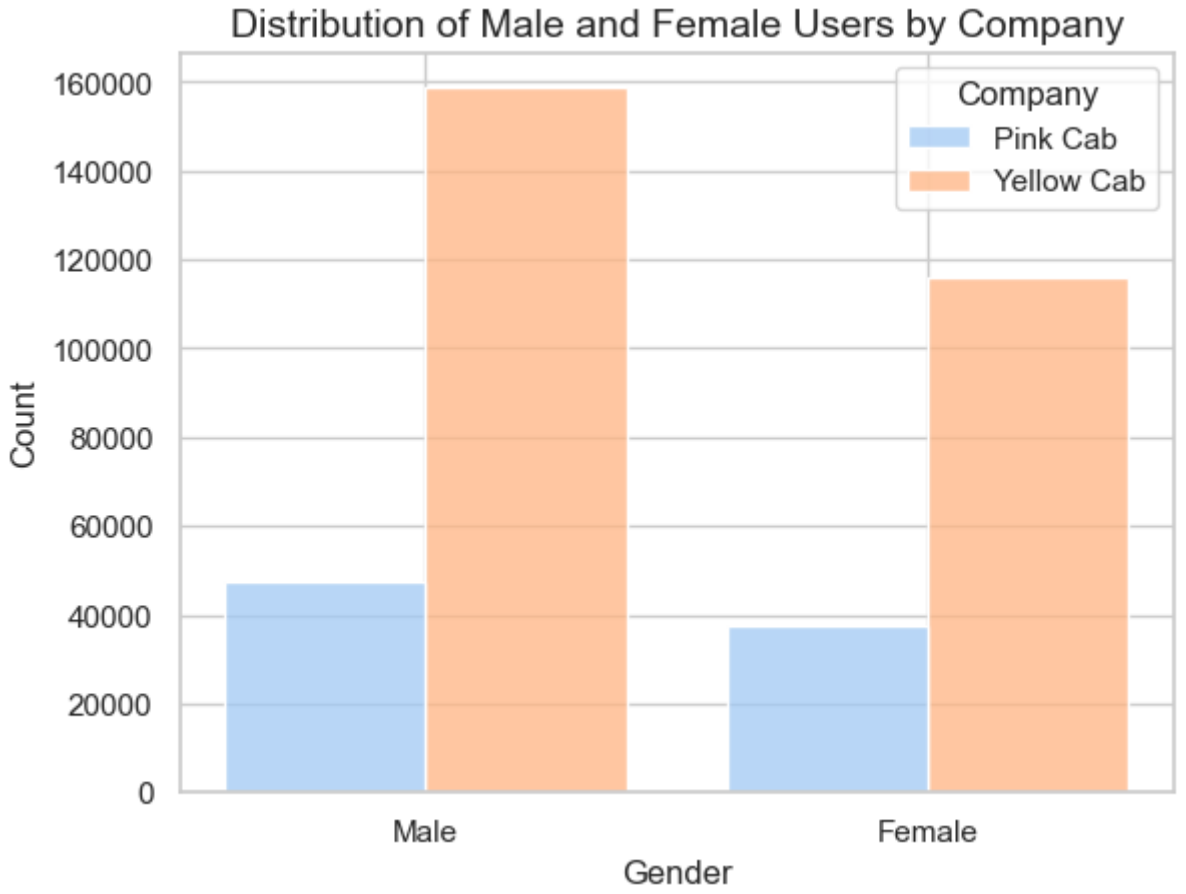
Distribution of Users by Cab Company in Various Cities



- As we can see, there is a significant difference in the number of users between each company, especially in big cities!

Data Glacier
Your Deep Learning Partner

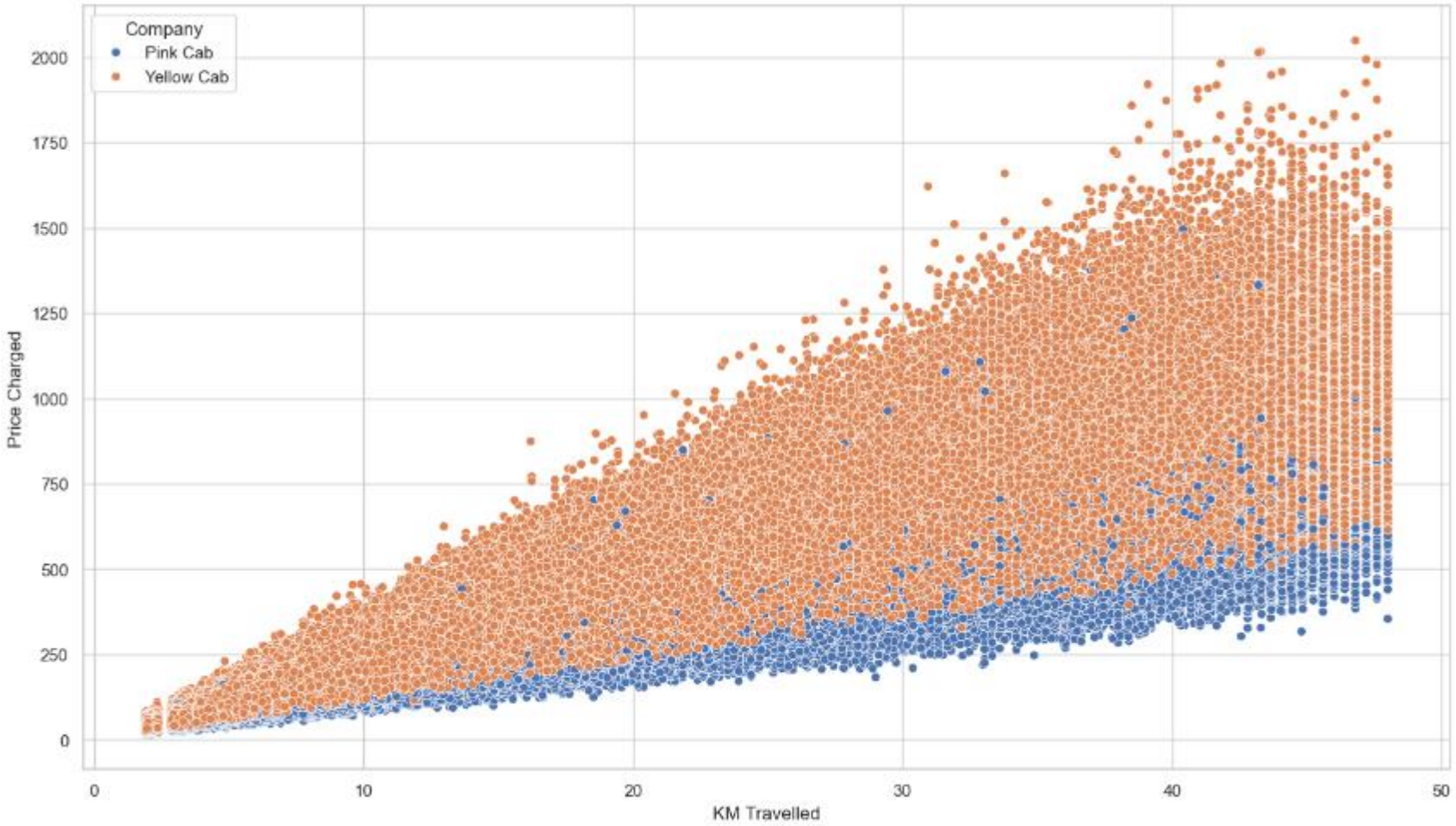
Male and Female Users



- Male users in both companies are higher than female users

Data Glacier
Your Deep Learning Partner

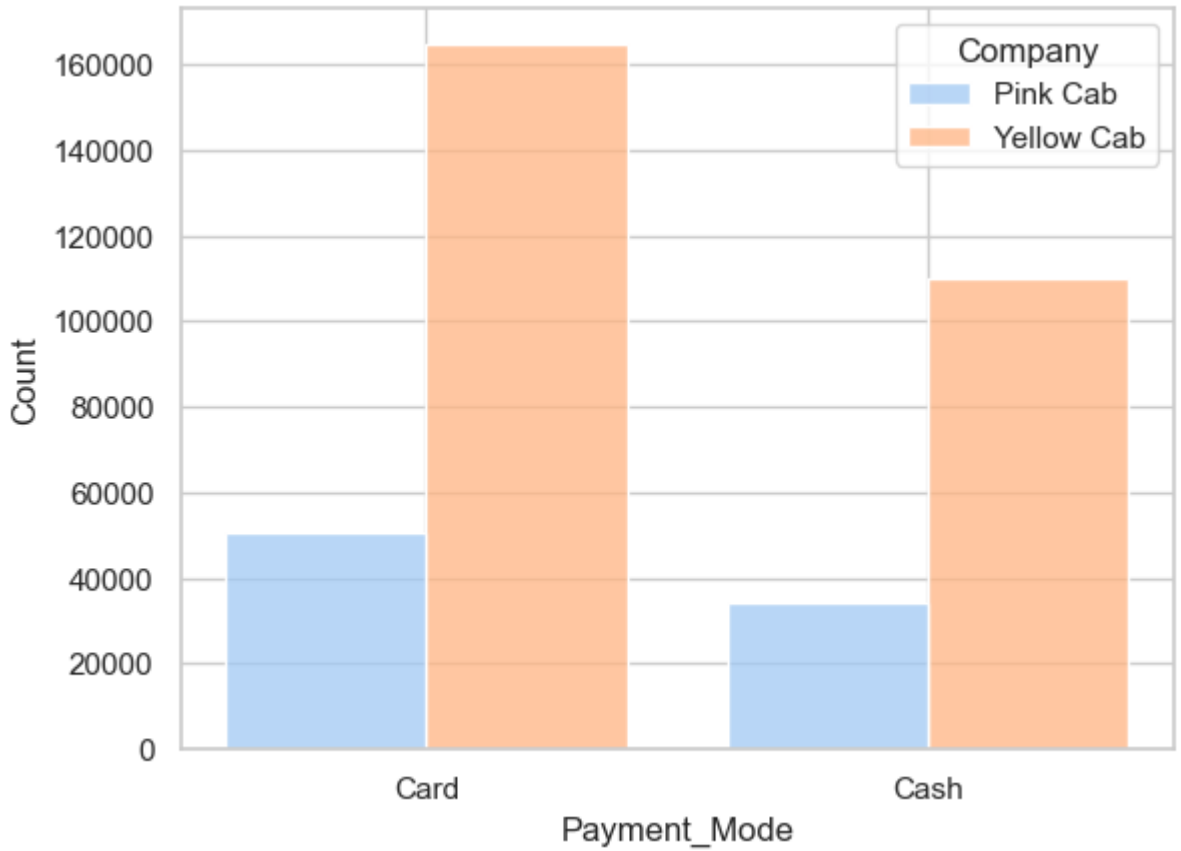
Correlation of Price Charged with Distance Travelled



- As we can see, there is a linear relationship between kilometres traveled and the price charged. However, Yellow Cab charges more per kilometer than Pink Cab

Data Glacier
Your Deep Learning Partner

Payment



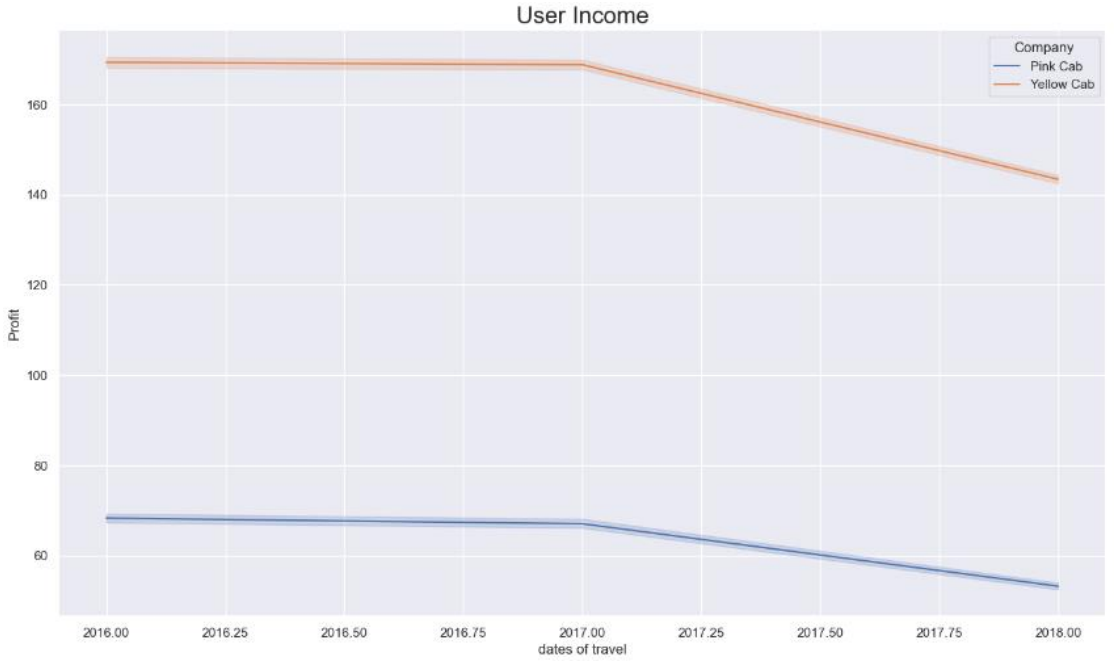
- Card Payment higher than Cash payments

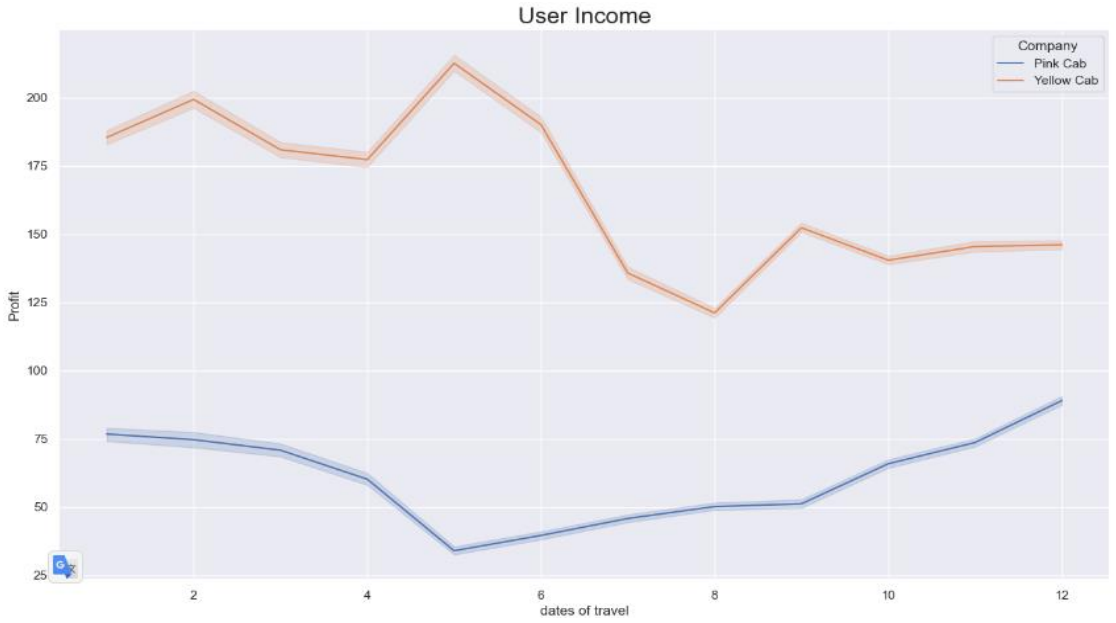
Users Average Income for both Companies

- Avg user income who use cab service is 15000$

Data Glacier
Your Deep Learning Partner

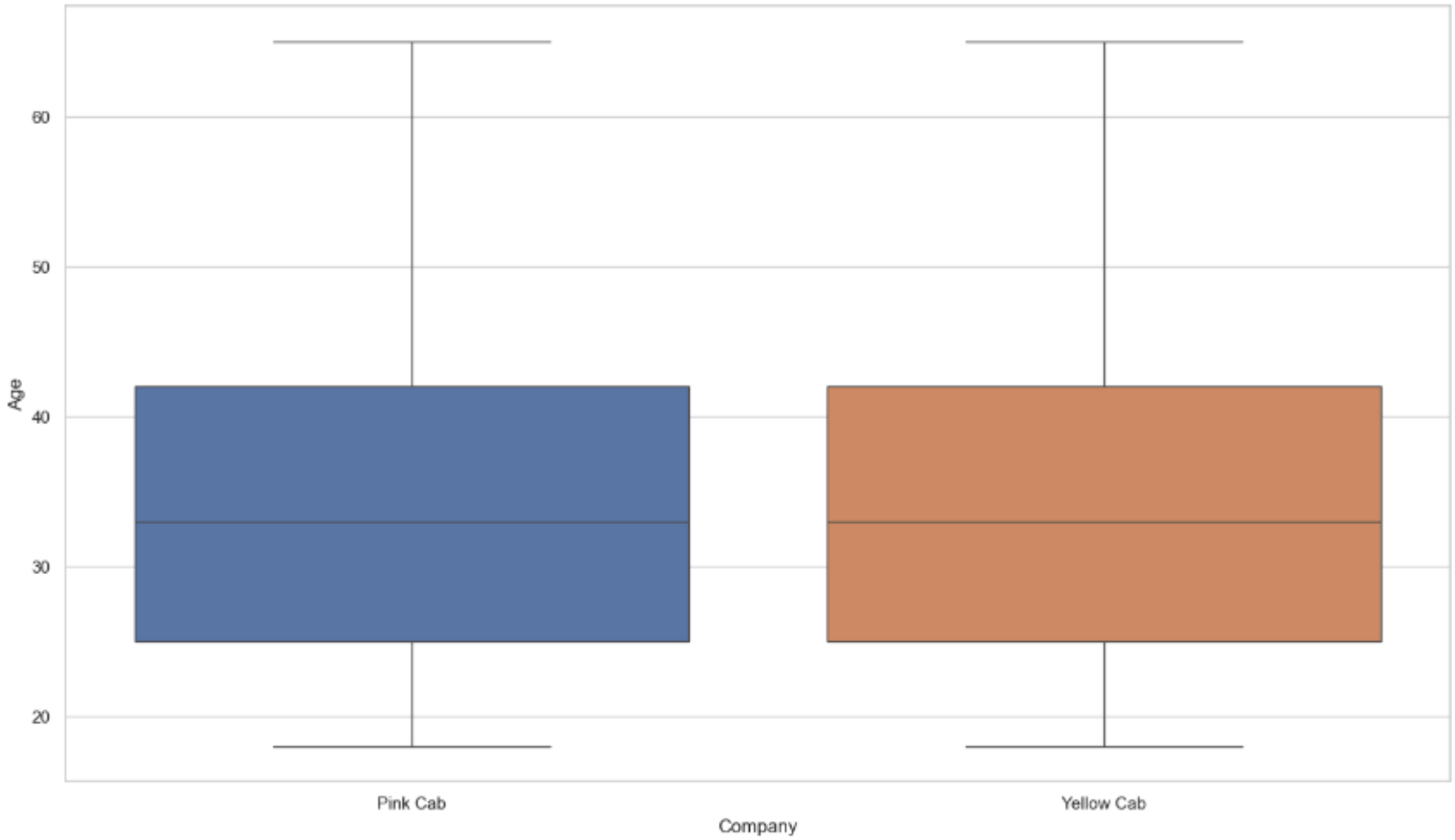
Profit Percentage Over the Years by Company



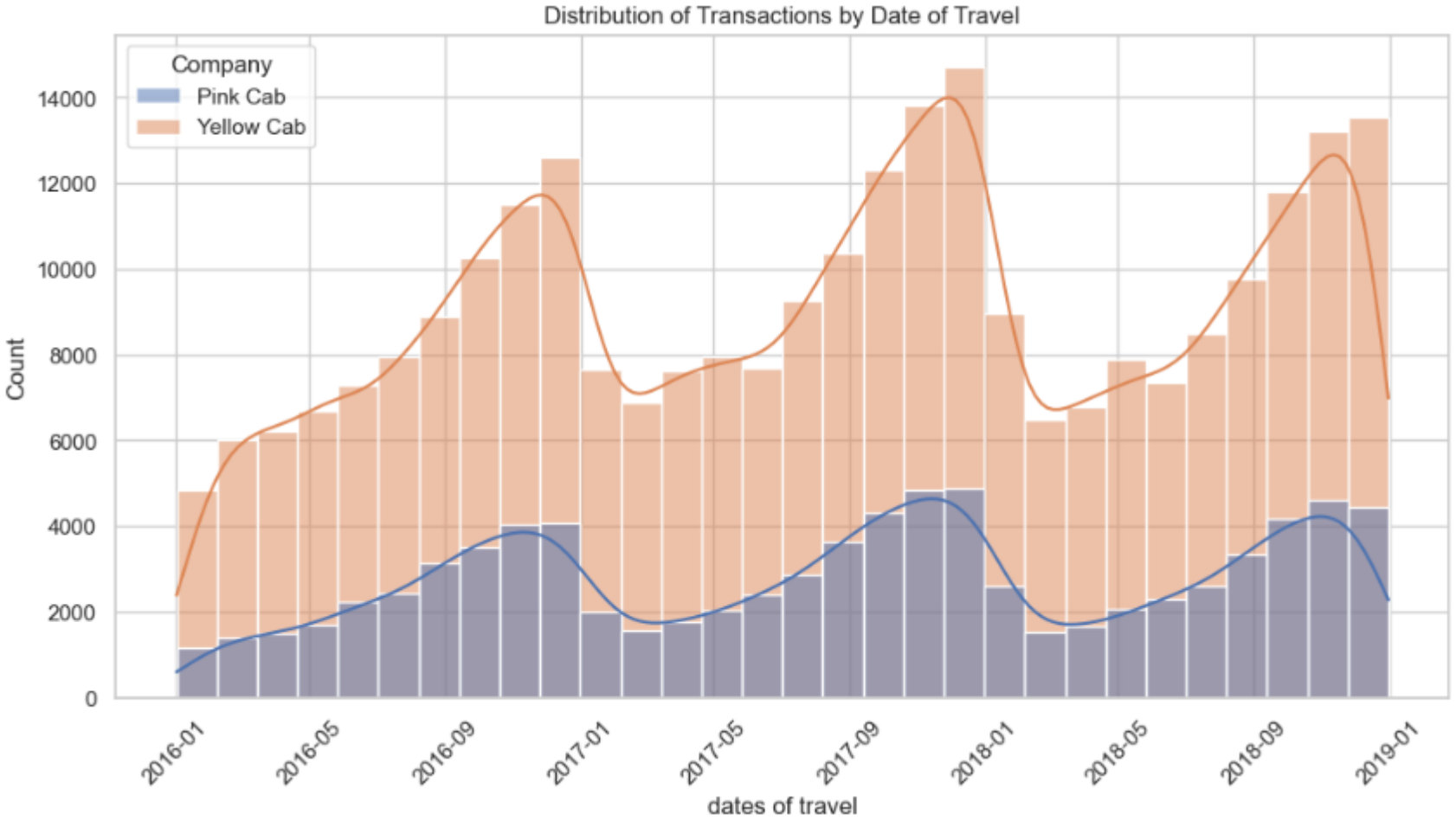


Data Glacier
Your Deep Learning Partner

Average Users Age of both Companies



Trends



- The most active time periods are consistently from **September** to **January** each year. This suggests that the demand for transactions is higher during these months, possibly due to seasonal events such as holidays, festivals, and year-end activities.

Data Glacier
Your Deep Learning Partner

Hypothesis

Hypothesis 1

Is there any difference in the average income between male and female customers.

H0 : there is no difference in the average income between male and female.

H1 : there is a difference in the average income between male and female.

```
0.5532825058546006
H0 accepted that there is no difference in the average income between male and female.
t-statistic: -0.5928491706560235, p-value: 0.5532825058546006
```

Data Glacier
Your Deep Learning Partner

Hypothesis 2

- Is there any difference in the average cost of trip between customers paying by card and those paying with cash.

- **H0** : there is no difference in the average cost of trip between customers paying by card and those paying with cash.
- **H1** : there is a difference in the average cost of trip between customers paying by card and those paying with cash.

```
0.5908527703588698
H0 accepted that there is no difference in the average cost of trip between customers paying by card and those payin
with cash.
t-statistic: -0.5376012286253666, p-value: 0.5908527703588698
```

Data Glacier
Your Deep Learning Partner

Hypothesis 3-4

Is there any difference in profit regarding Gender in both companies.

H0 : there is no difference in profit regarding gender in both companies

H1 : there is a difference in in profit regarding gender in both companies

```
p-value: 3.2360611315752597e-25
H1 accepted: There is a difference in profit regarding gender in Yellow Cab.
t-statistic: 10.375692753988766, p-value: 3.2360611315752597e-25
```

```
p-value: 0.11539907489276462
H0 accepted: There is no difference in profit regarding gender in Yellow Cab.
t-statistic: 1.5743993868244515, p-value: 0.11539907489276462
```

Hypothesis 5-6

Is there any difference in Profit regarding Age.

H0 : there is no difference in profit regarding Age.

H1 : there is a diiference in in profit regarding Age.

```
p-value: 5.662484659530008e-05
H1 accepted: There is a difference in profit regarding Age in Yellow Cap.
t-statistic: -4.027329757294125, p-value: 5.662484659530008e-05
```

```
p-value: 0.49966555840261784
H0 accepted: There is no difference in profit Age Yellow Cap.
t-statistic: -0.6750556598744547, p-value: 0.49966555840261784
```

Conclusion

- Yellow Cab Company is better Because:
  - More Users
  - More Transactions
  - Profit
  - Popularity

Thank You