G2M Insight For Cab Investment Firm

JELSON LINO
Data Science Intern

09 September 2022

Agenda

- Problem Statement
- Data information
- Correlation Analysis
- EDA
- Hypothesis Test
- Recommendations

Problem Statement

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

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

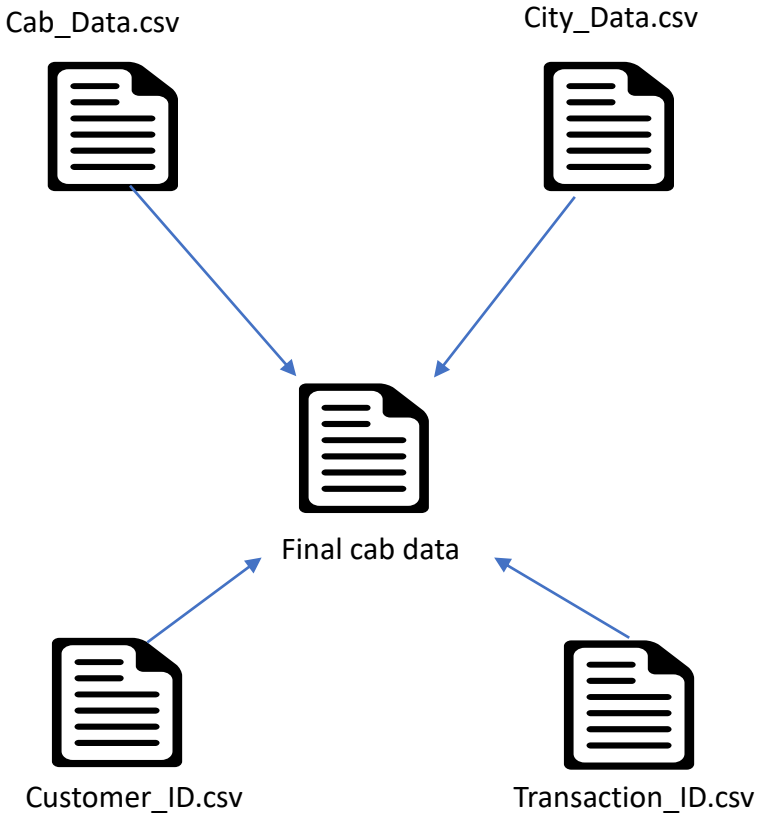
The analysis has been divided into four parts:

- Data Understanding
- Determining the company with more clients
- Determining the most affordable company for clients
- Determining the most profitable company
- Testing various hypothesis

Data Information

Data Information:

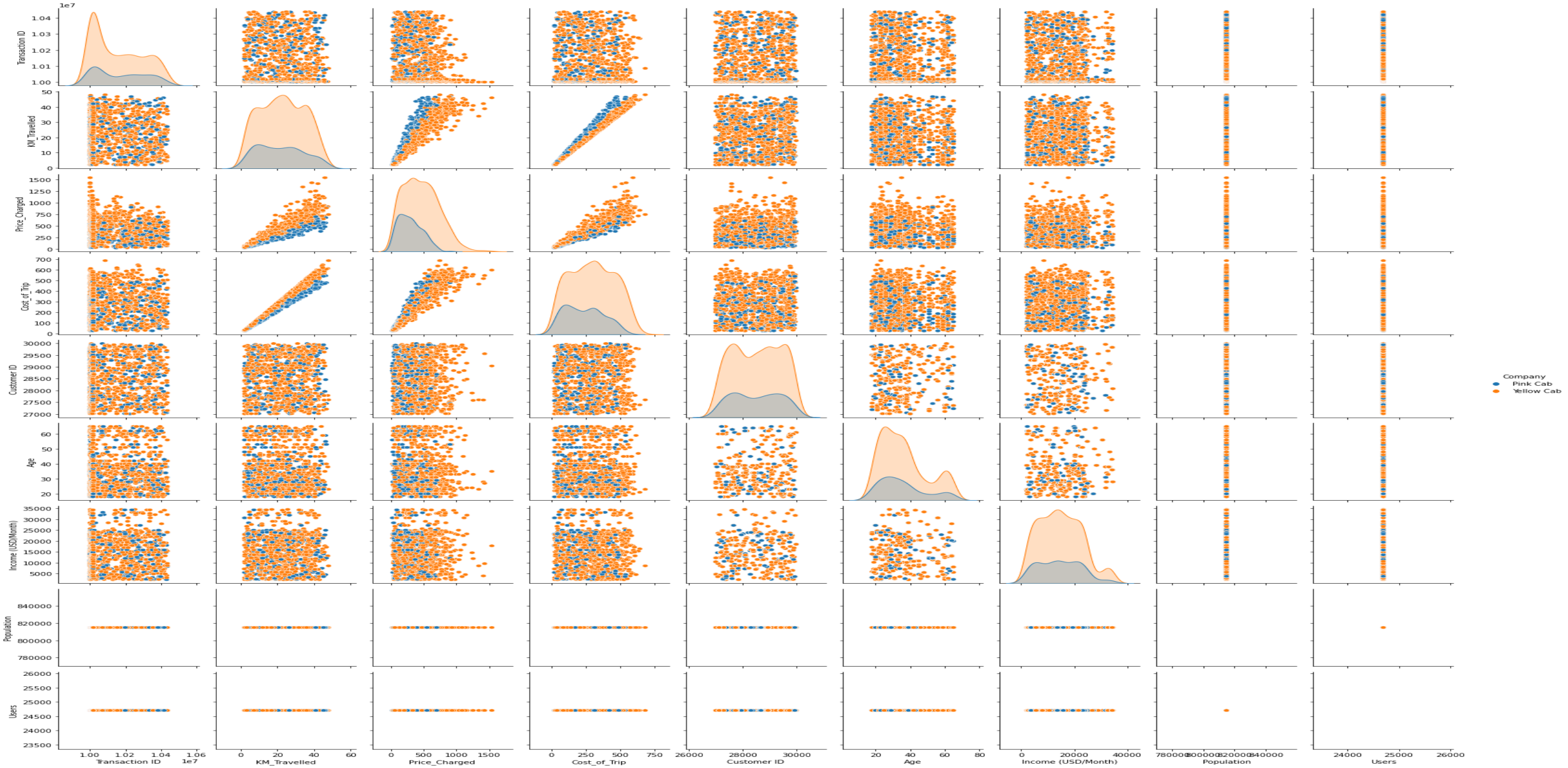
- Cab_Data.csv: This file includes information about transactions for two cab companies.
- City_Data.csv: This file contains information about some American cities' population and cab users.
- Customer_ID.csv: This file contains customers' demographic information.
- Transaction_ID.csv: This file contains information about transactions and payment methods.



Assumptions:

- There are outliers in the Price_Charged feature. However, because no information about trip duration is given, the outliers have not been treated as such.
- Profit of rides are calculated assuming that other factors are constant and only Price_Charged and Cost_of_Trip features are taken into consideration.

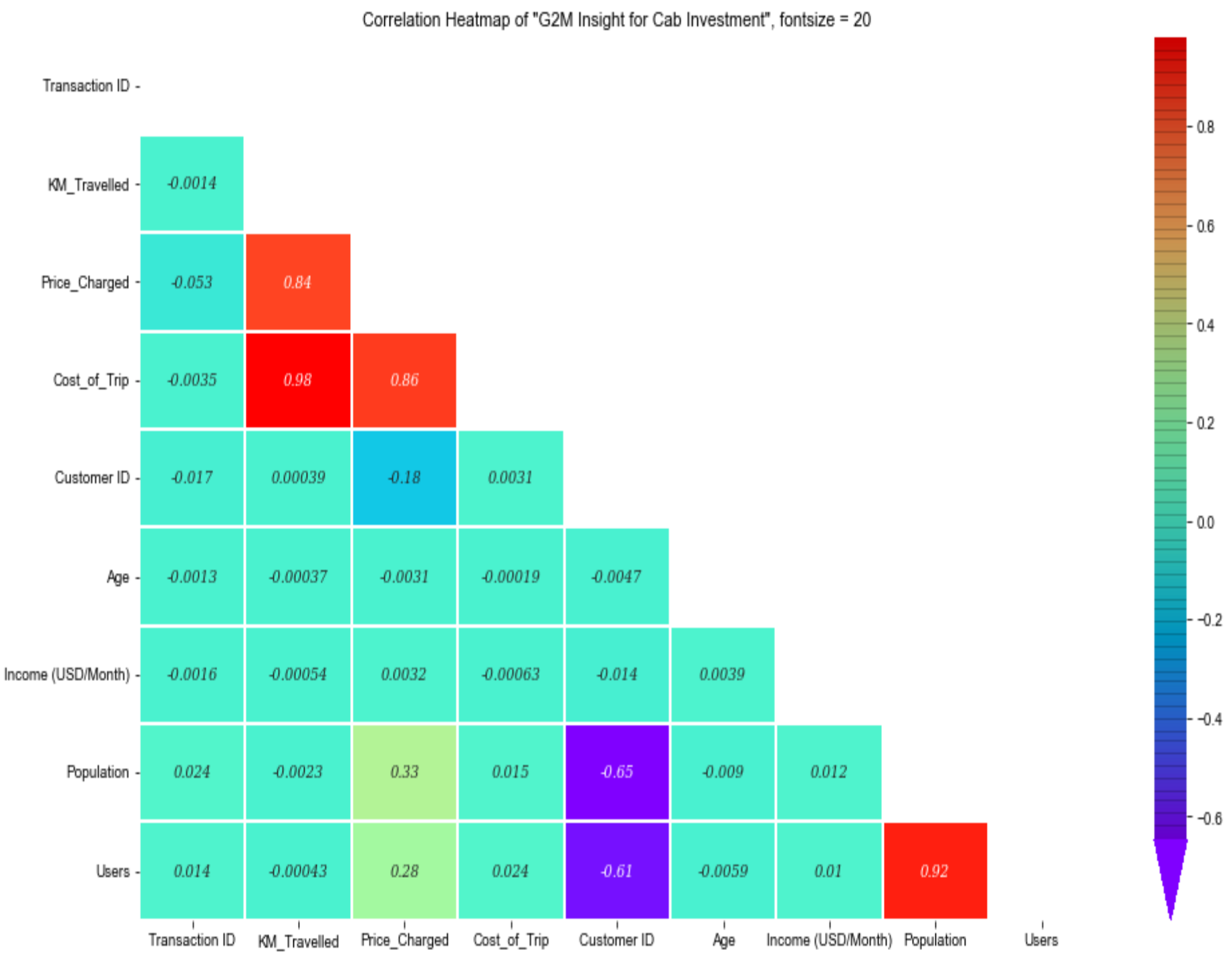
Relationships Between Features



Correlation Analysis

Insights:

- Population and Users are highly correlated
- Price_Charged, KM_Travelled, and Price_Charged are highly correlated
- Customer_ID is negatively correlated with both Population and Users



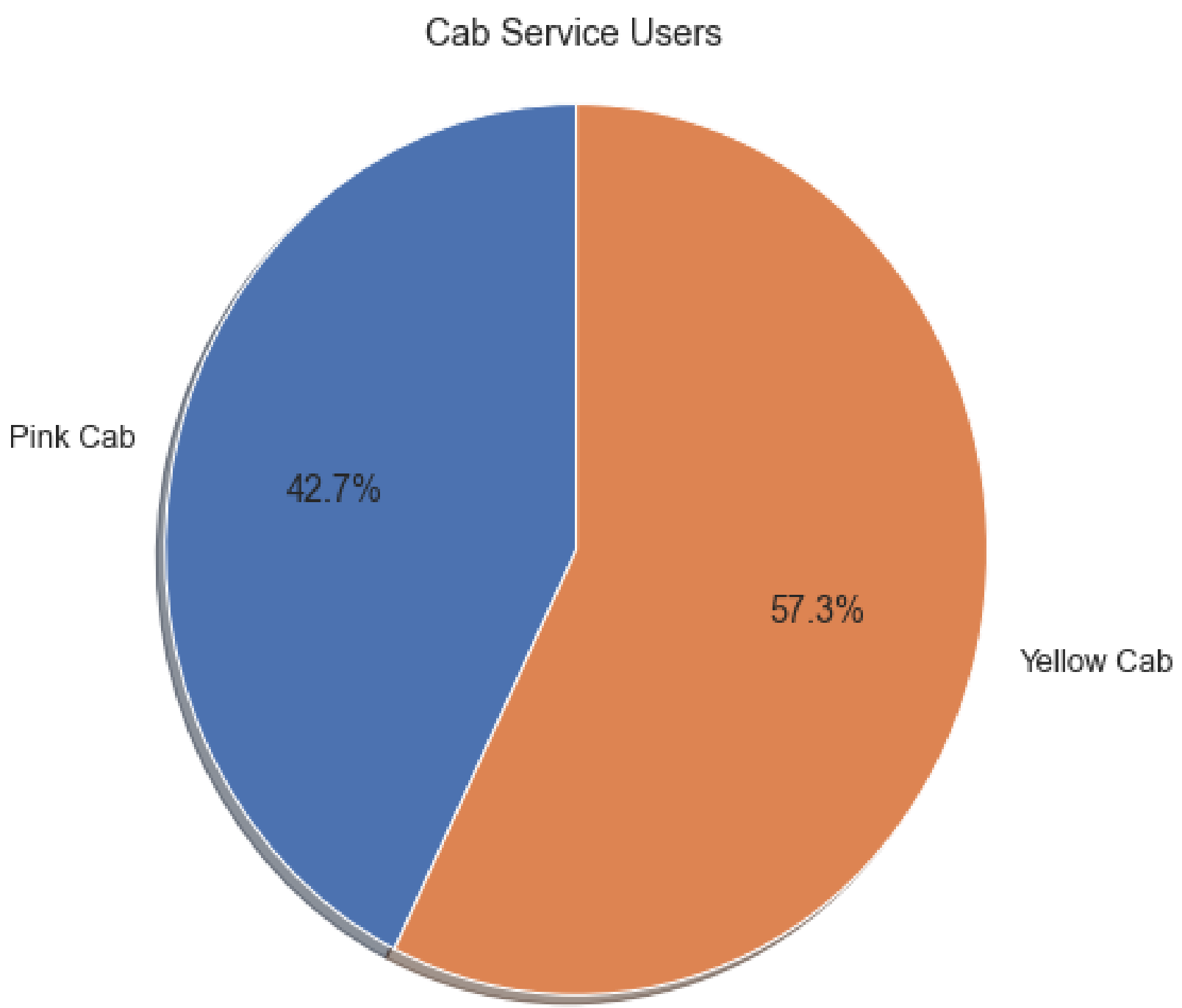
Data Glacier
Your Deep Learning Partner

EDA

MARKET SHARE

Insights:

- Yellow Cab has more clients than Pink Cab



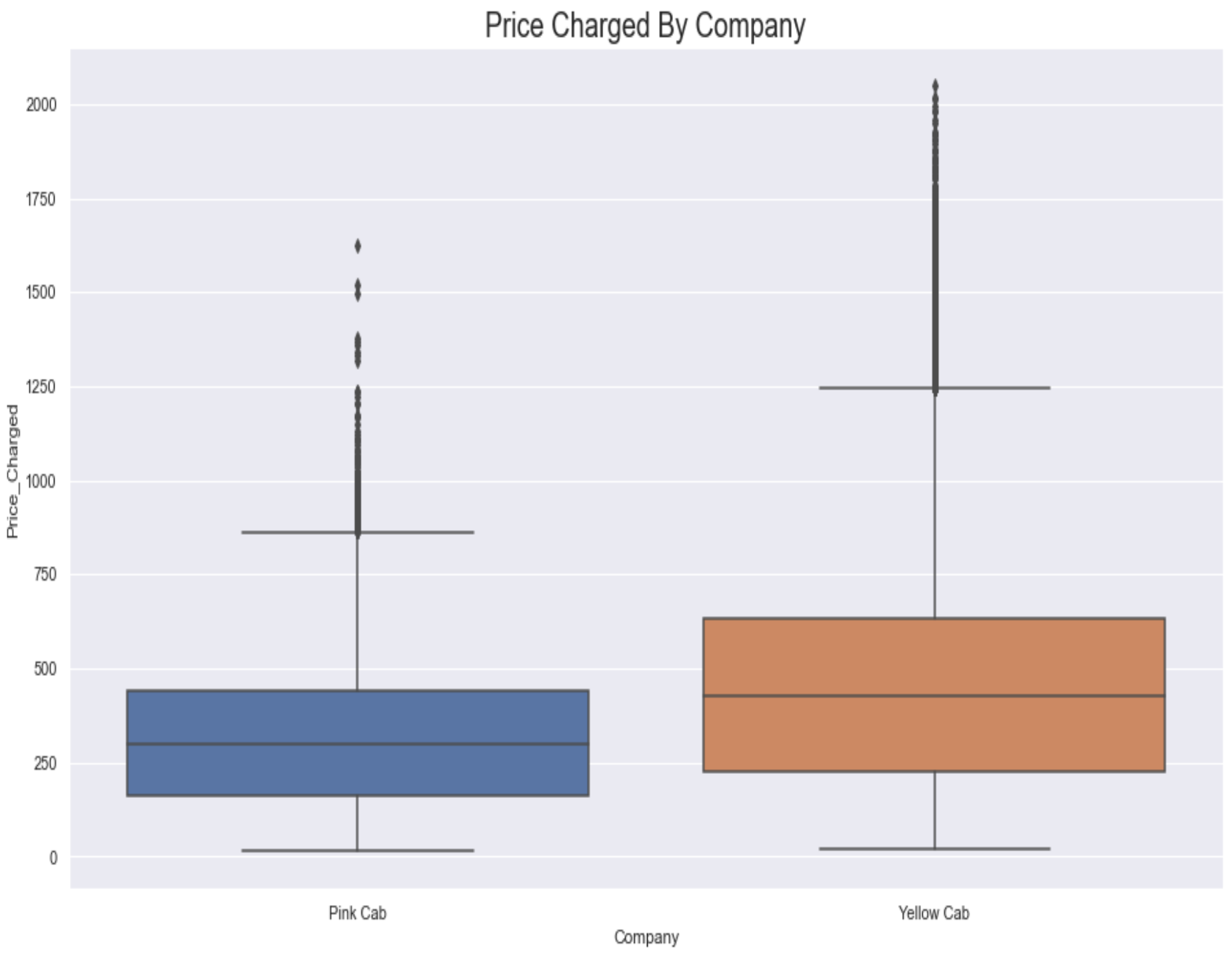
Data Glacier
Your Deep Learning Partner

EDA

Cab Service Prices

Insights:

- Yellow Cab is more expensive than Pink Cab



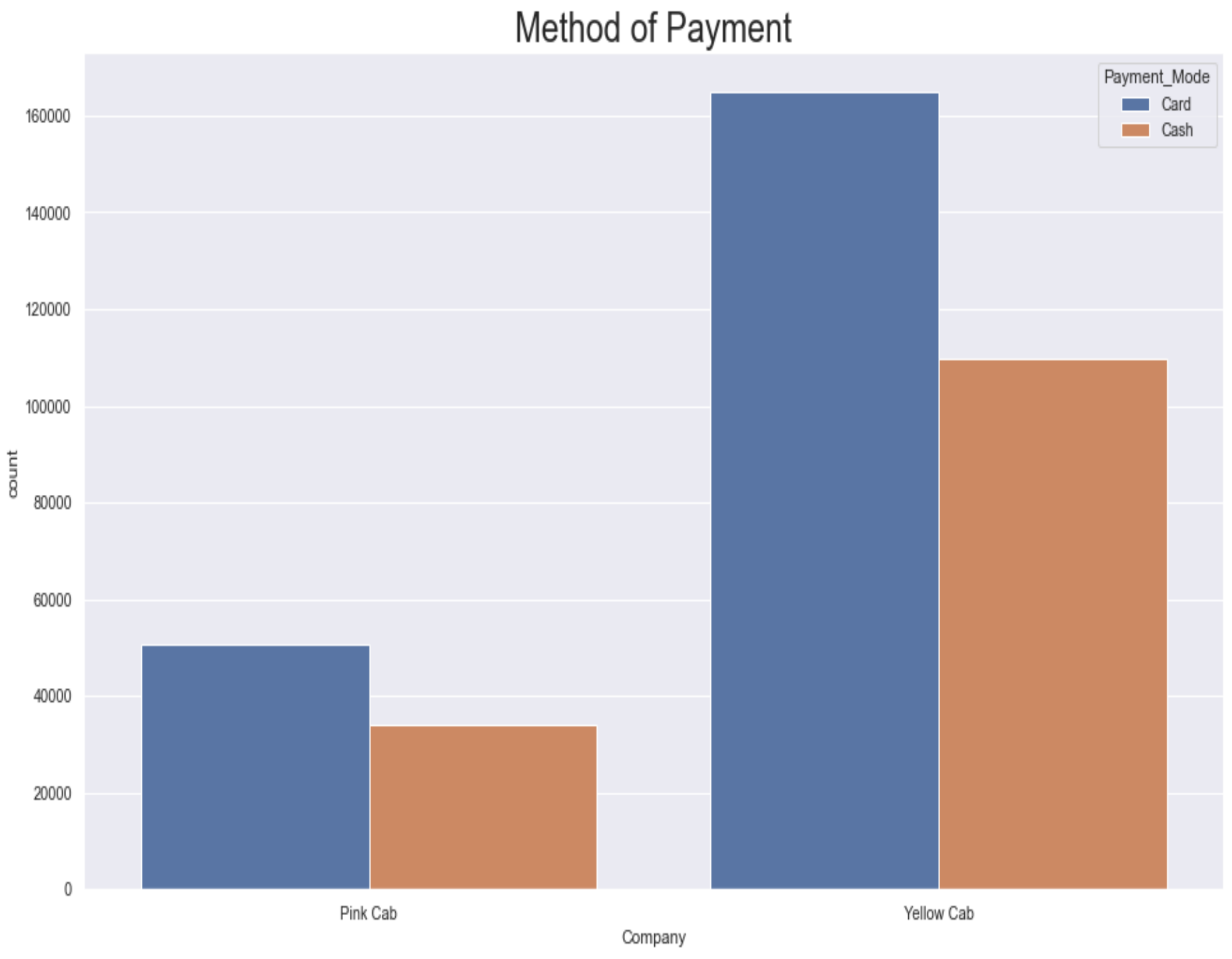
Data Glacier
Your Deep Learning Partner

EDA

How do Customers Pay For Rides?

Insights:

- Most customer pay by card



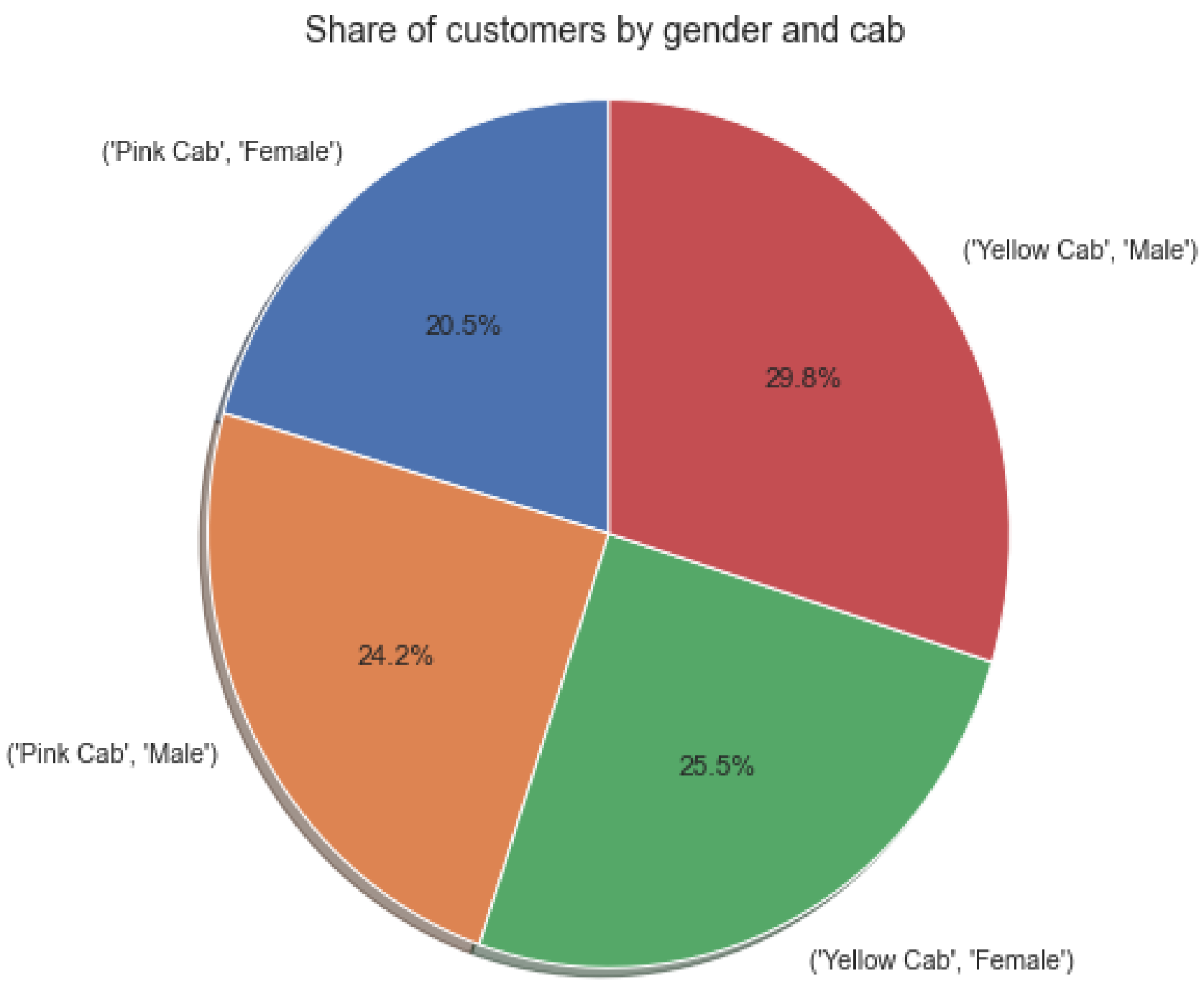
Data Glacier
Your Deep Learning Partner

EDA

Customer Distribution by Gender

Insights:

- There are more males than females using cab services



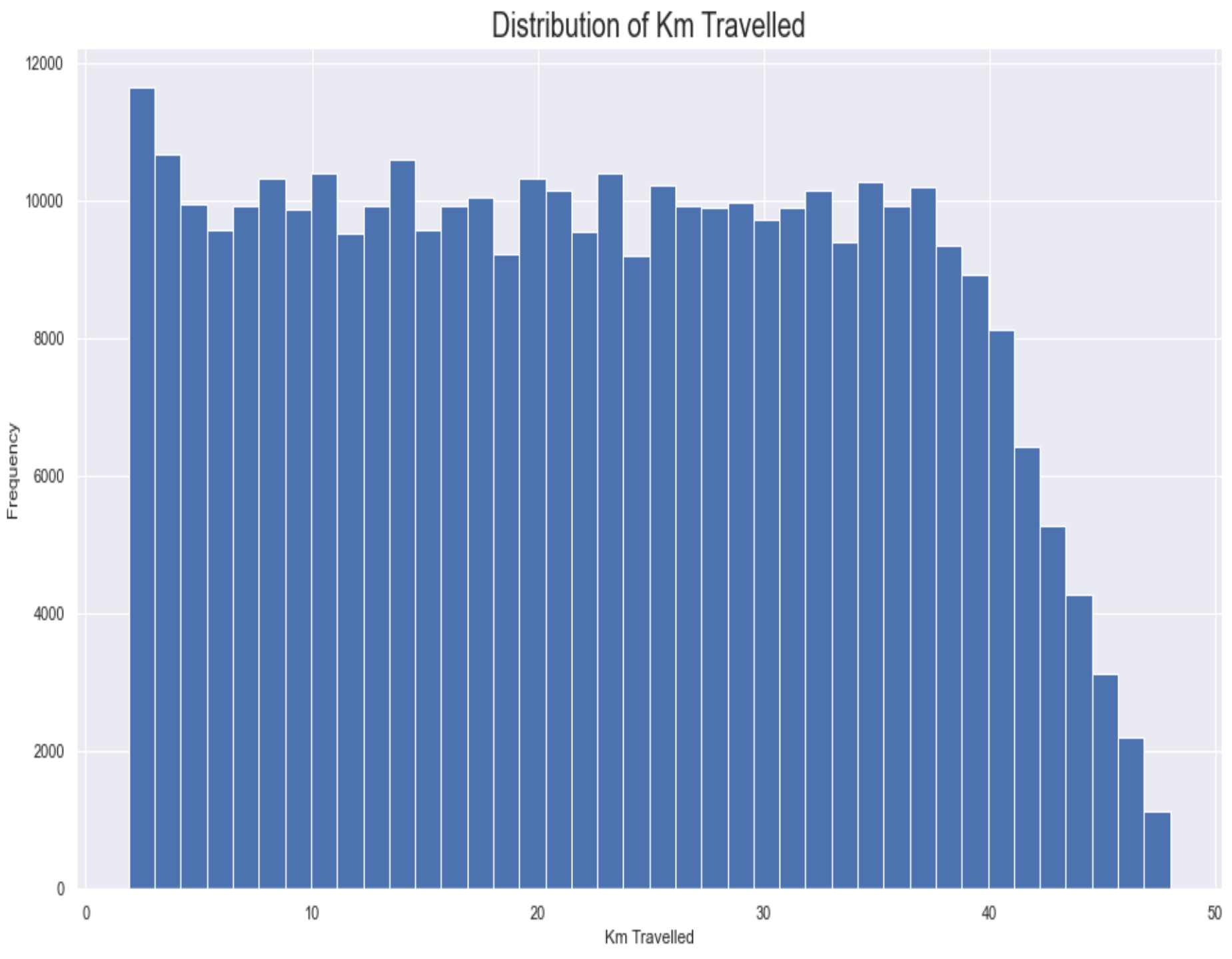
Data Glacier
Your Deep Learning Partner

EDA

Distribution of KM Travelled

Insights:

- Trips vary between 2 and 48 KM
- Most common trips are less than 40 KM



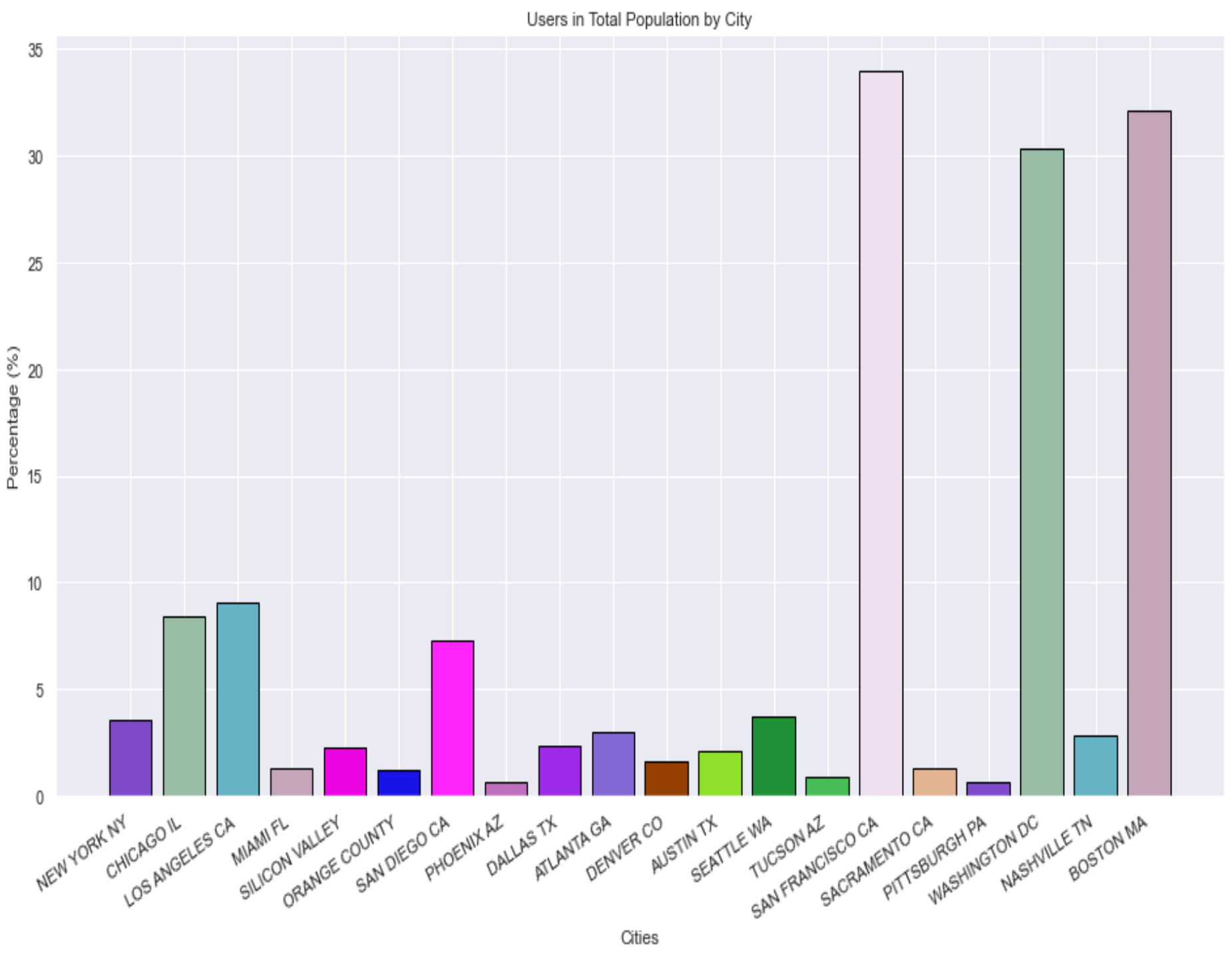
Data Glacier
Your Deep Learning Partner

EDA

Clients Per City Population

Insights:

- San Francisco, Boston, and Washington have the highest percentage of cab users per city population.
- Each of these three cities has more than one-third of their respective populations.



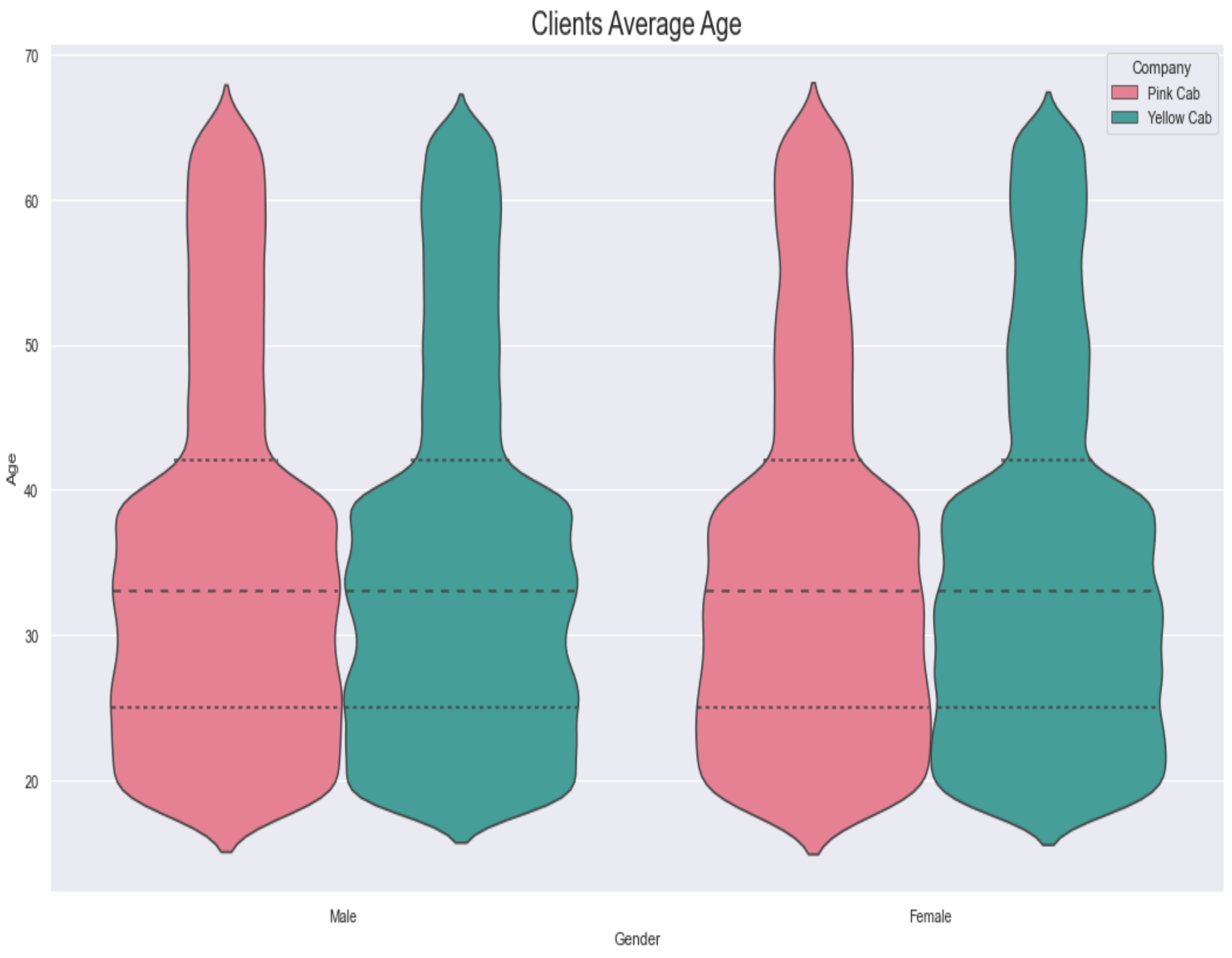
Data Glacier
Your Deep Learning Partner

EDA

Clients Average Age

Insights:

- Both male and females clients have an average age of approximately 34.



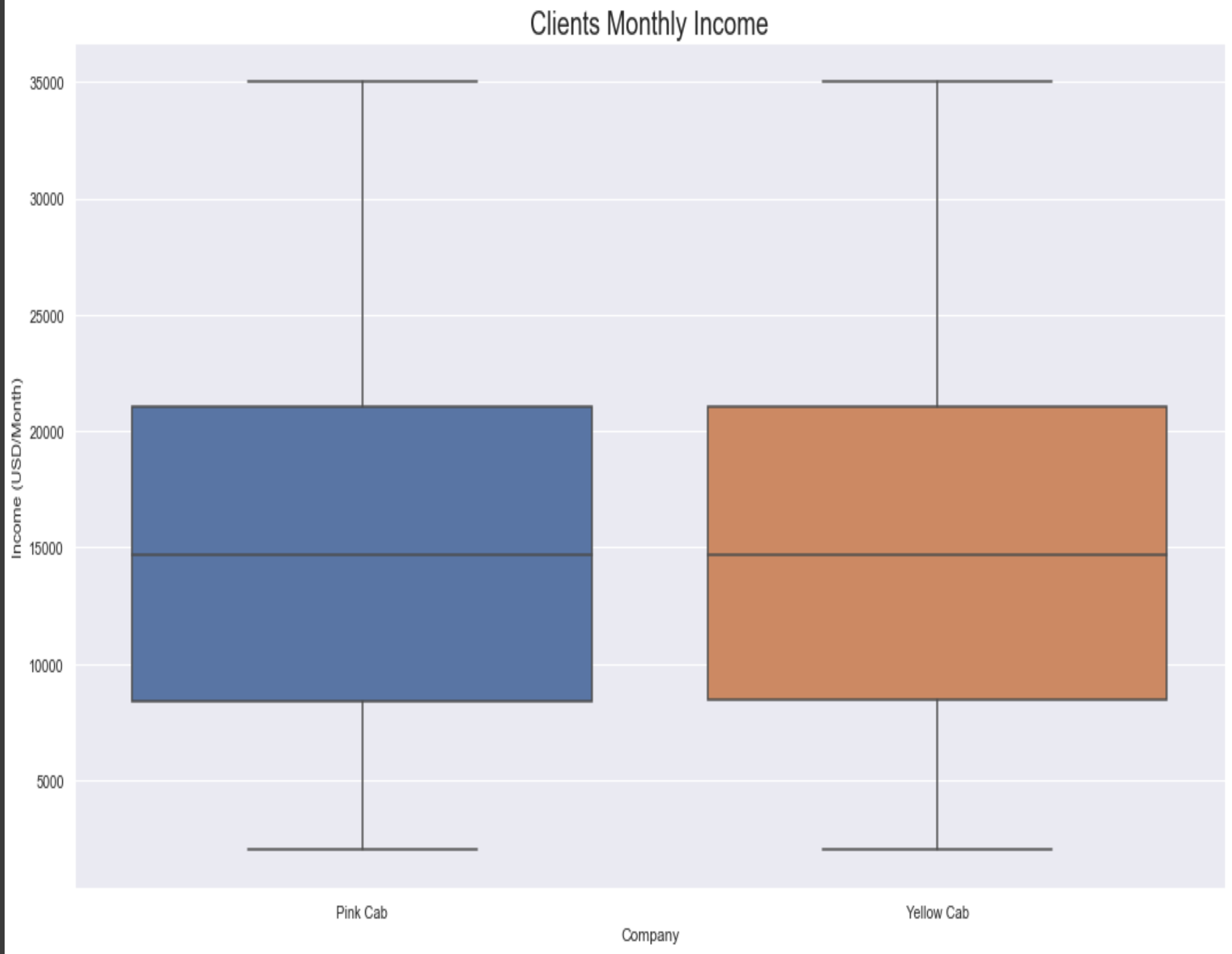
Data Glacier
Your Deep Learning Partner

EDA

Clients Average Age

Insights:

- Clients average monthly income is approximately $15000.



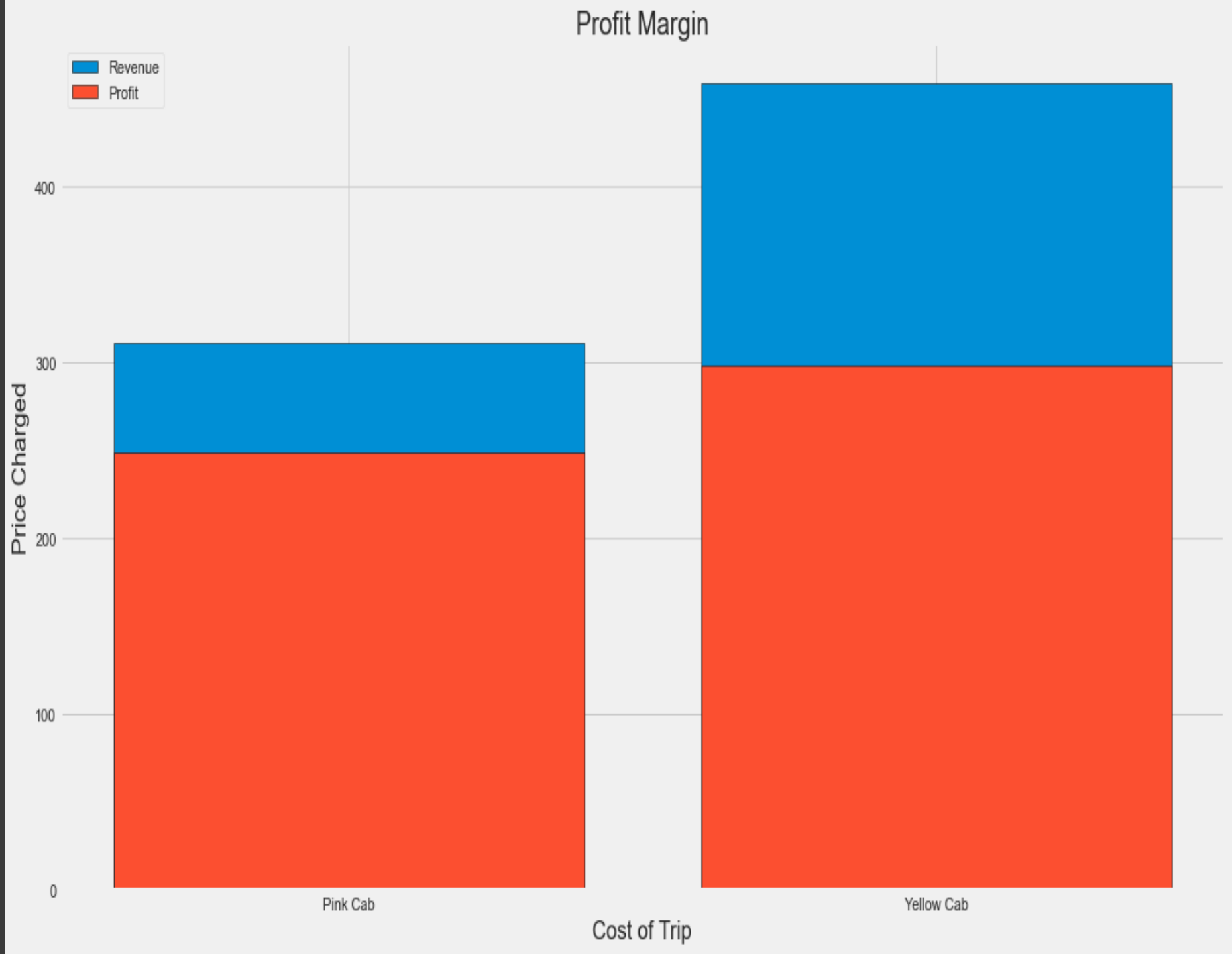
Data Glacier
Your Deep Learning Partner

EDA

Profit Margin

Insights:

- Yellow Cab profit margin is higher than that of Pink Cab.



Data Glacier
Your Deep Learning Partner

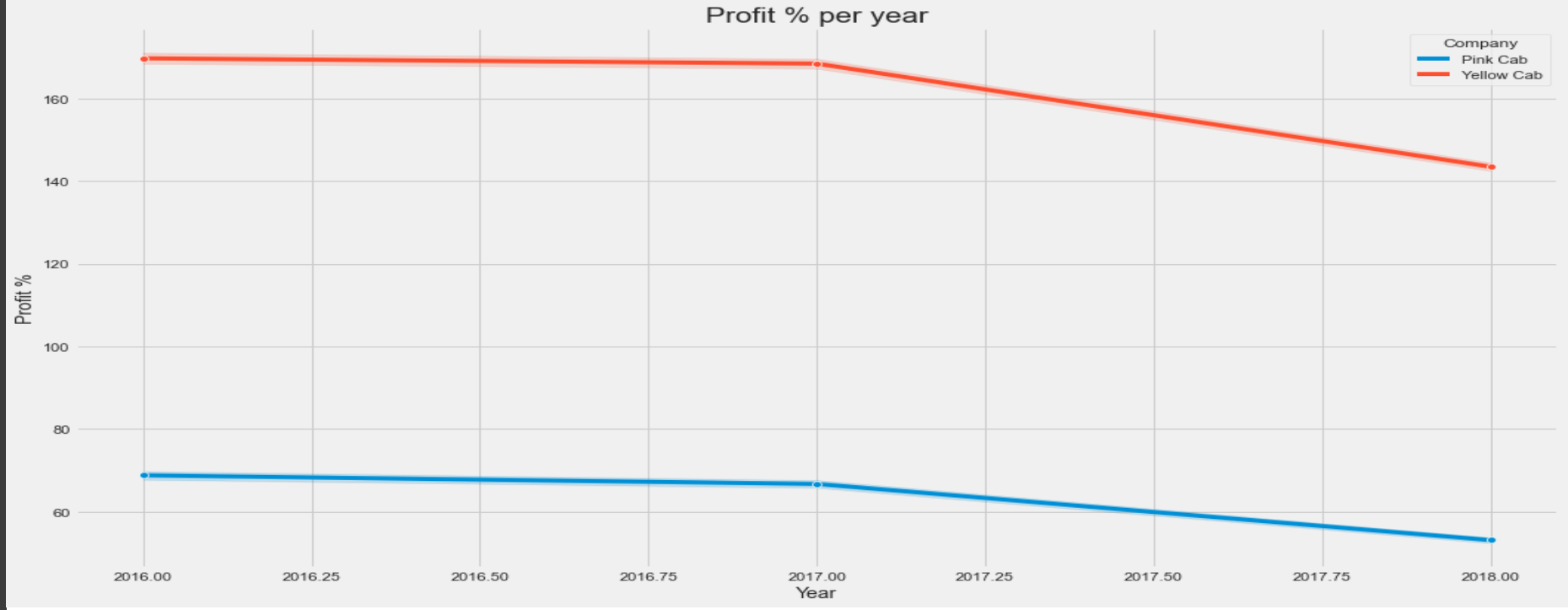
EDA

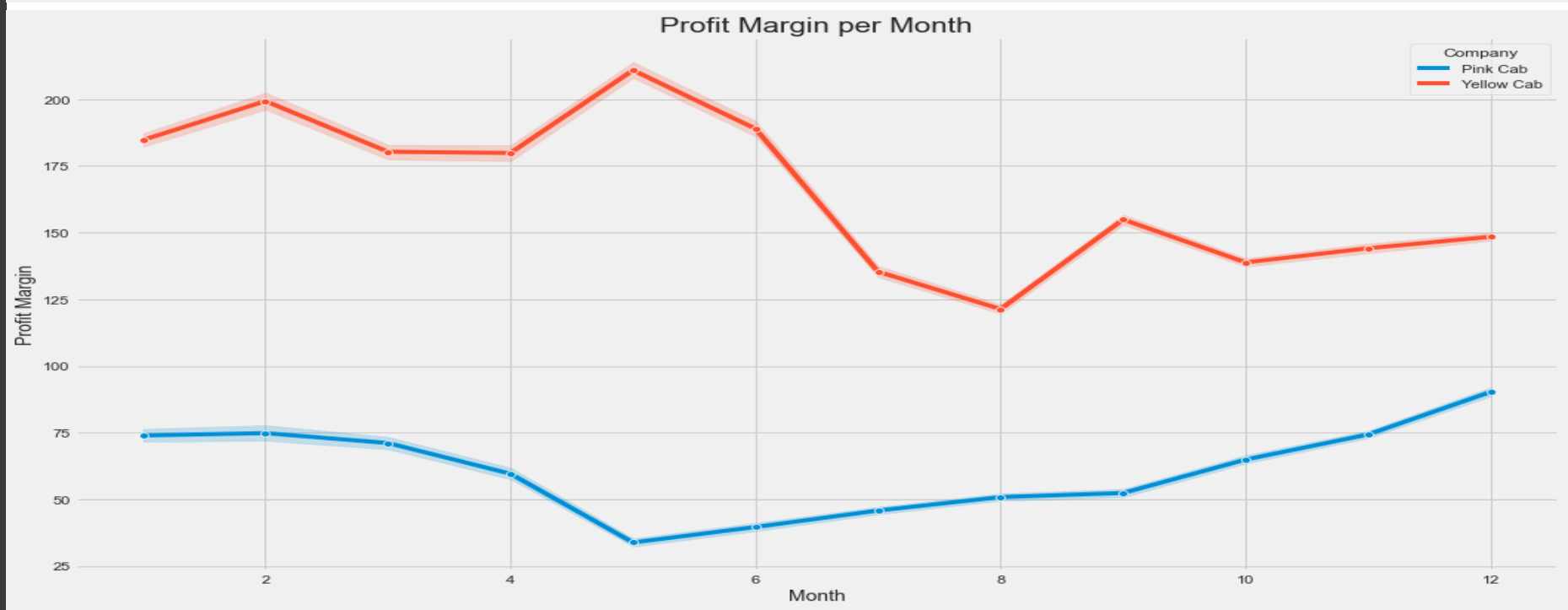
Profit Margin Variation With Time

Insights:

- Profit margin has been decreased with year.
- Profit margin varies with month.
- Pink cab has its highest profit margin at the end of the year.
- Yellow cab makes more profit in the first half of the year.

Data Glacier
Your Deep Learning Partner





EDA

Relationship Between Price and Distance

Insights:

- Price charged and distance travelled are linearly related.
- Yellow Cab price is higher than that of Pink Cab.

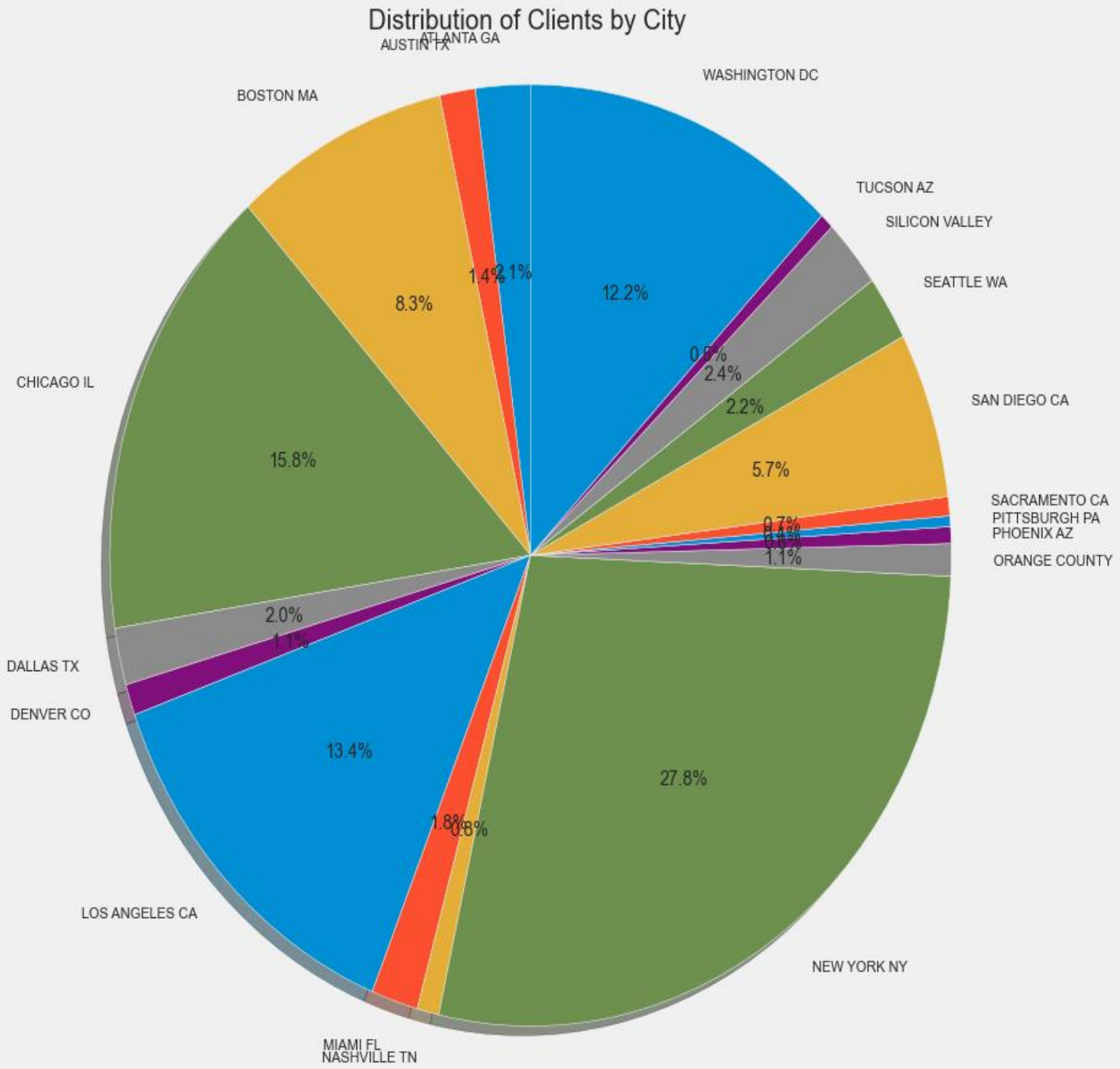
Data Glacier
Your Deep Learning Partner

EDA

Distribution of Clients by City

Insights:

- New York, Chicago, and Loas Angeles have the highest percentage of cab users.



Data Glacier
Your Deep Learning Partner

HYPOTHESIS INVETIGATION

Hypothesis 1

Is there any different in profit margin by method of payment?

H0: There is no difference in profit margin generated by different payment methods.

H0: There is no difference in profit margin generated by different payment methods.

Data Glacier
Your Deep Learning Partner

Yellow Cab

```
P value is  0.2933060638298729
We accept the null hypothesis (H0) that there is no difference in the profits by payment method for Yellow Cab
```

Pink Cab

```
P value is  0.7900465828793288
We accept null hypothesis (H0) that there is no difference in profits by payment method for Pink Cab
```

CONCLUSION

There is no difference in profit margin by payment method for both cabs

Hypothesis 2

Is there any different in profit margin by gender?

H0: There is no difference in profit margin by gender.

H0: There is no difference in profit margin by gender.

Data Glacier
Your Deep Learning Partner

Yellow Cab

```
P value is  6.060473042494144e-25
We accept the alternative hypothesis (H1) that there is a difference in profits by gender for Yellow Cab
```

Pink Cab

```
P value is  0.11515305900425798
We accept null hypothesis (H0) that there is no difference in profits by gender for Pink Cab
```

CONCLUSION

There is no difference in profit margin by gender for both cabs

Hypothesis 3

Is there any different in profit margin by age?

H0: There is no difference in profit margin by age.

H0: There is no difference in profit margin by age.

Data Glacier
Your Deep Learning Partner

Yellow Cab

```
P value is  6.328485471267631e-05
We accept the alternative hypothesis (H1) that there is a difference in profits by age for Yellow Cab
```

Pink Cab

```
P value is  0.4816748536155635
We accept the null hypothesis (H0) that there is no difference in profits by age for Pink Cab
```

CONCLUSION

There is no difference in profit margin by age for Pink cabs.
It appears that Yellow cabs offer discount to senior citizens (60 or more years of age)

RECOMMENDATIONS

Based on the facts that:
1. Yellow Cab has higher profit margin than Pink Cab
2. Yellow Cab has more clients than Pink Cab
3. Yellow Cab makes more transactions per annually

It would be better to invest in the Yellow Cab Company

Thank You