Project: G2M Insight For Cab Investment Firm

Data Science Virtual Internship
Submitted by: Farha Jabin Oyshee
Batch: LISUM13

Problem Statement

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 an investment.

The analysis has been divided into four parts:
- Data Understanding and Visualization
- Finding the most users Cab company
- Finding the cheapest Cab company for users
- Finding the most profitable Cab company
- Multiple Hypothesis and Investigate

Data
Information

Data Information

- Cab_Data.csv – this file includes details of transactions for 2 cab companies
- Customer_ID.csv – this is a mapping table that contains a unique identifier that 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 a list of US cities, their population, and the number of cab users

Correlation Between Variables

- There is a strong correlation between
  - Population vs Users
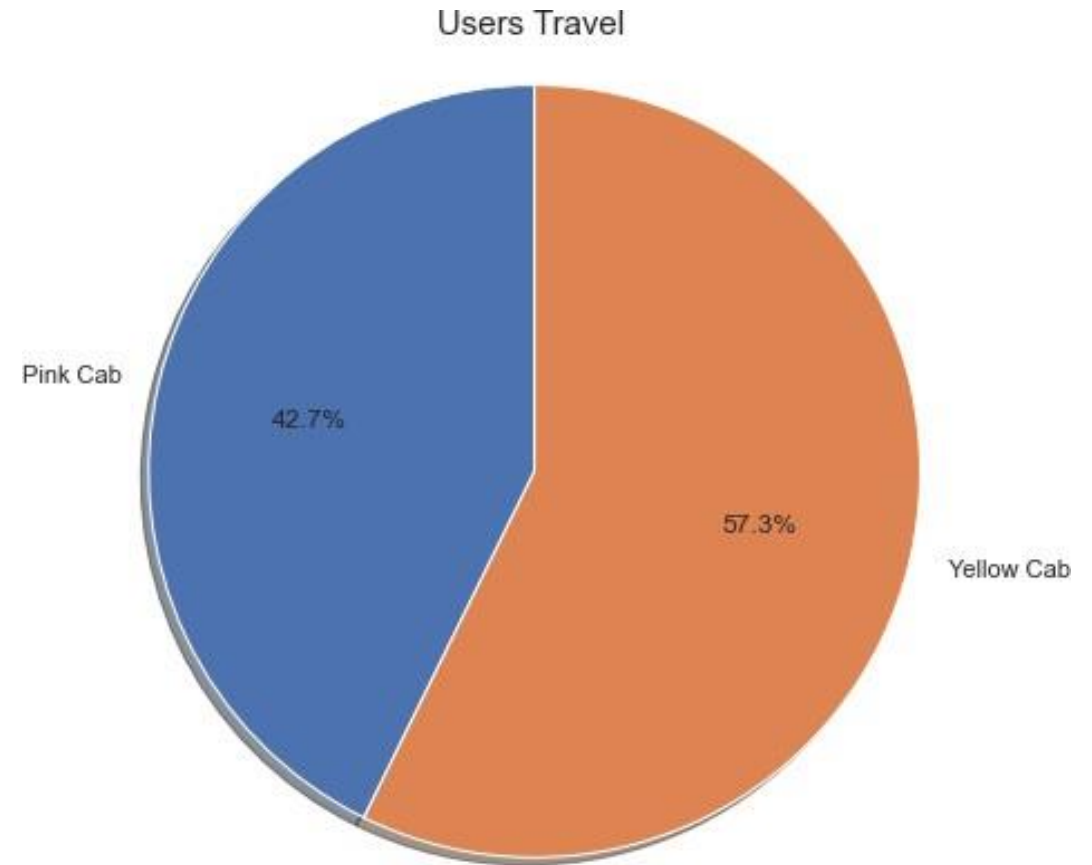 - Price Charged vs Cost of Trip vs KM Travelled

Correlation Heatmap of "G2M Insight for Cab Investment", fontsize = 20
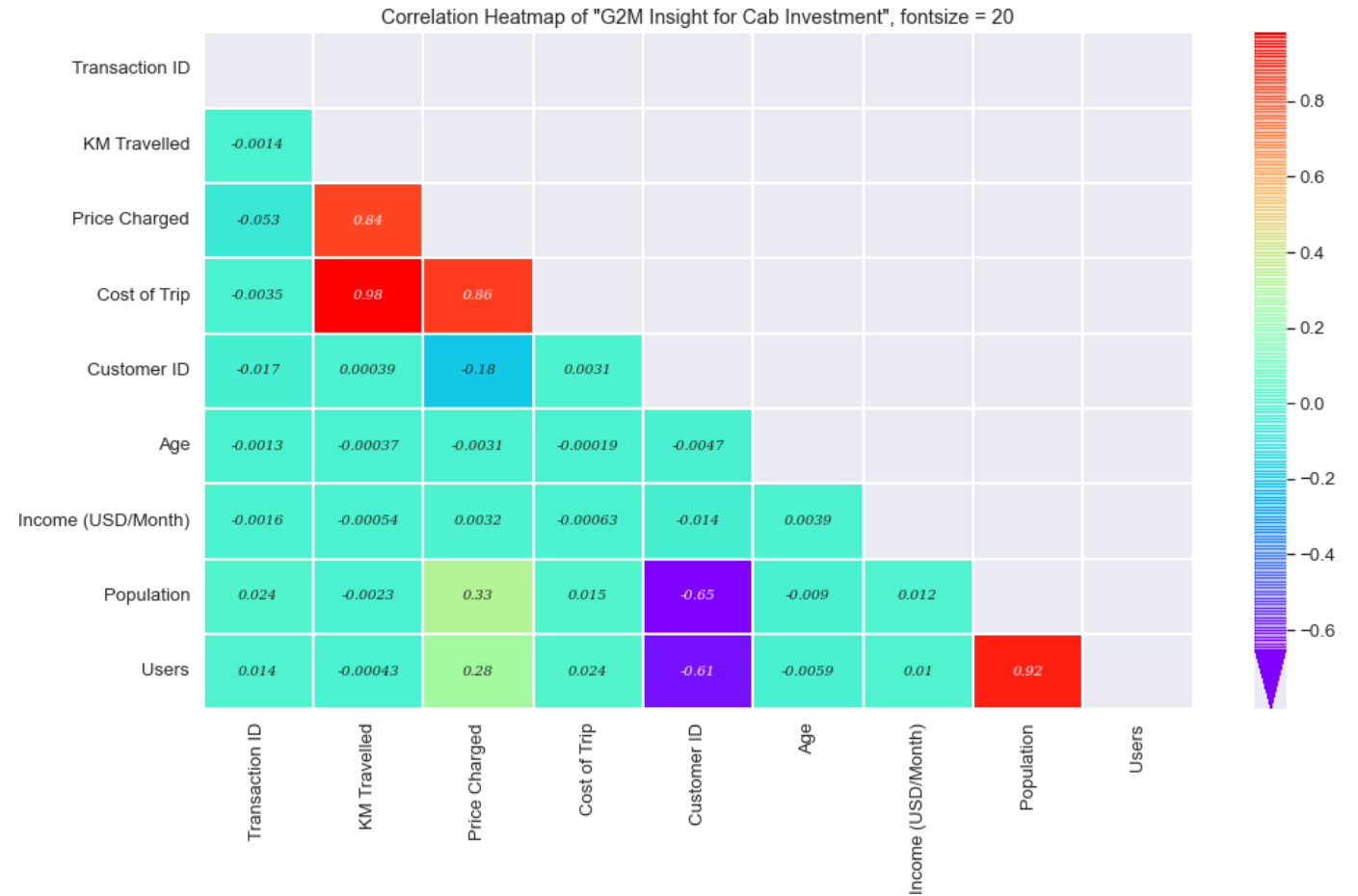
EDA Analysis

Which Company has more Users?

- User's preferable ride on Yellow cab more than to Pink Cab
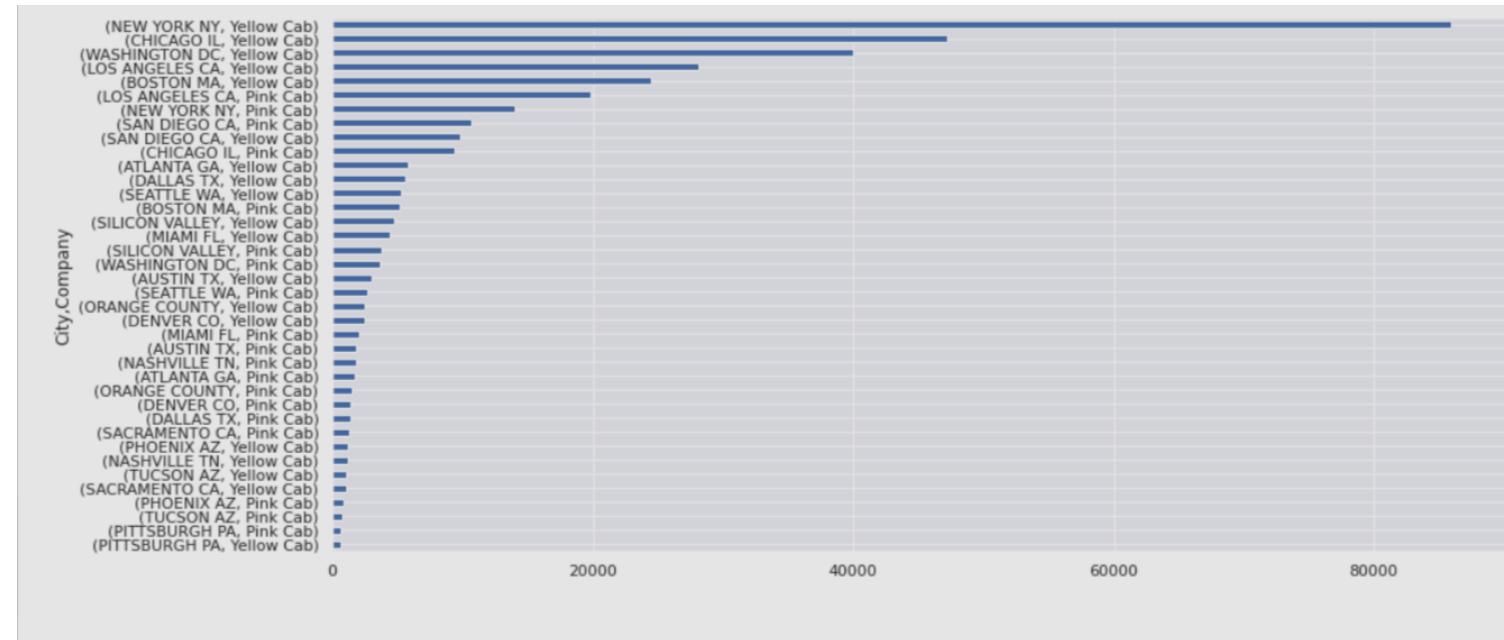
Users Travel

Pink Cab 42.7%

Yellow Cab 57.3%

Which Company has a high price charged?

- Price Charged for Yellow Cab **is higher than compared** to Pink Cab
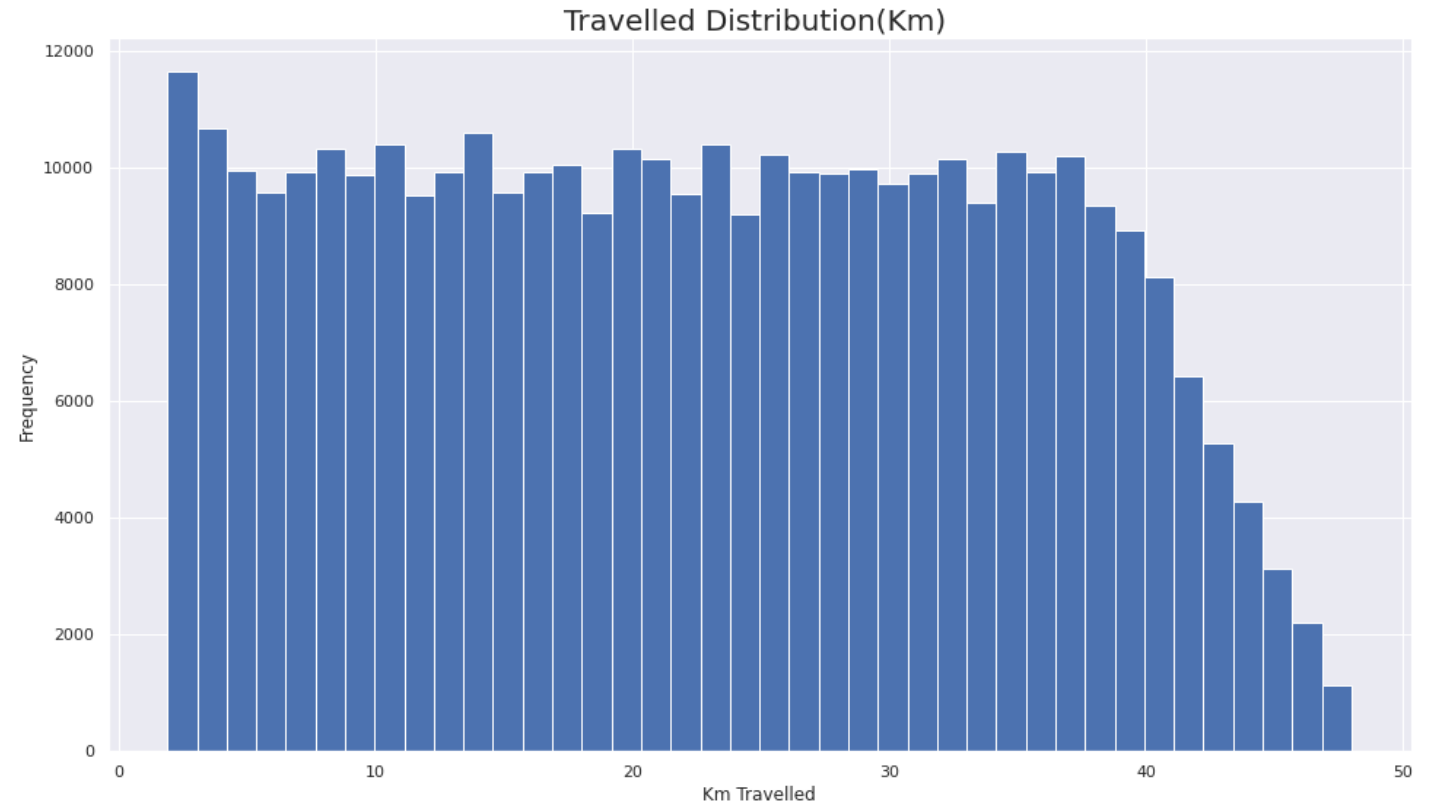
Price Charged of Both Companies
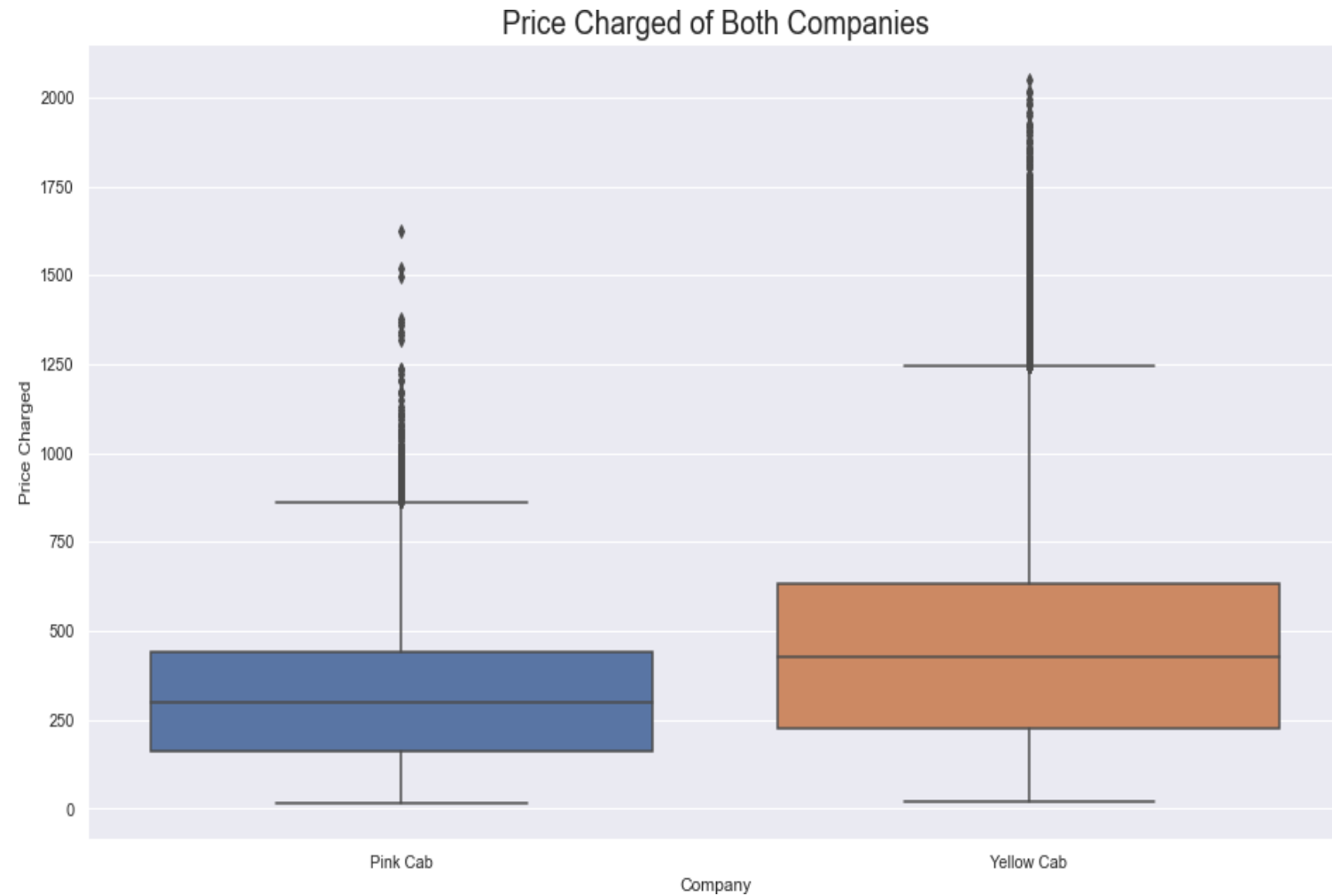
Which City has more demand ?

- New York City has more demand compared than other cities

Travelled Distribution KM

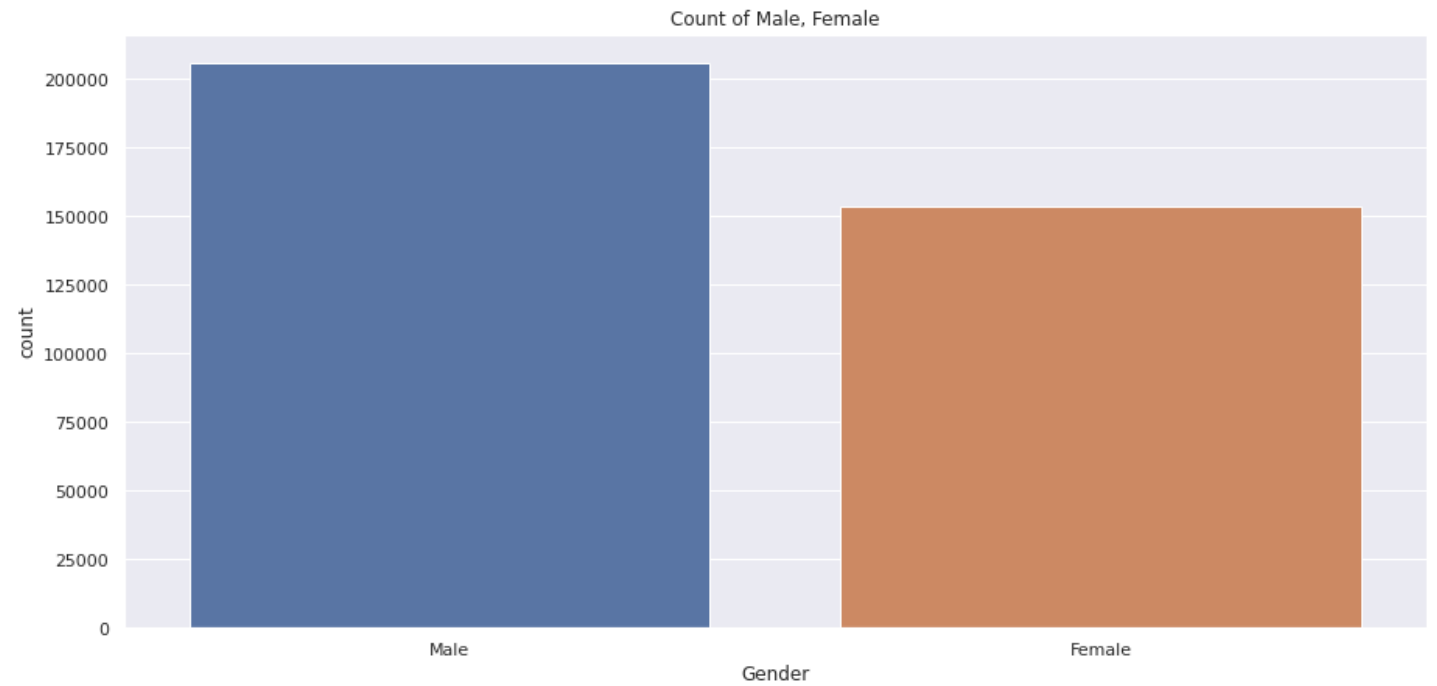- Most of the rides varies from 2 to 48 KM.

Travelled Distribution(Km)

Users based Gender

- Male users are prefer more to travel in Cab
- Users prefer to travel in Yellow Cab

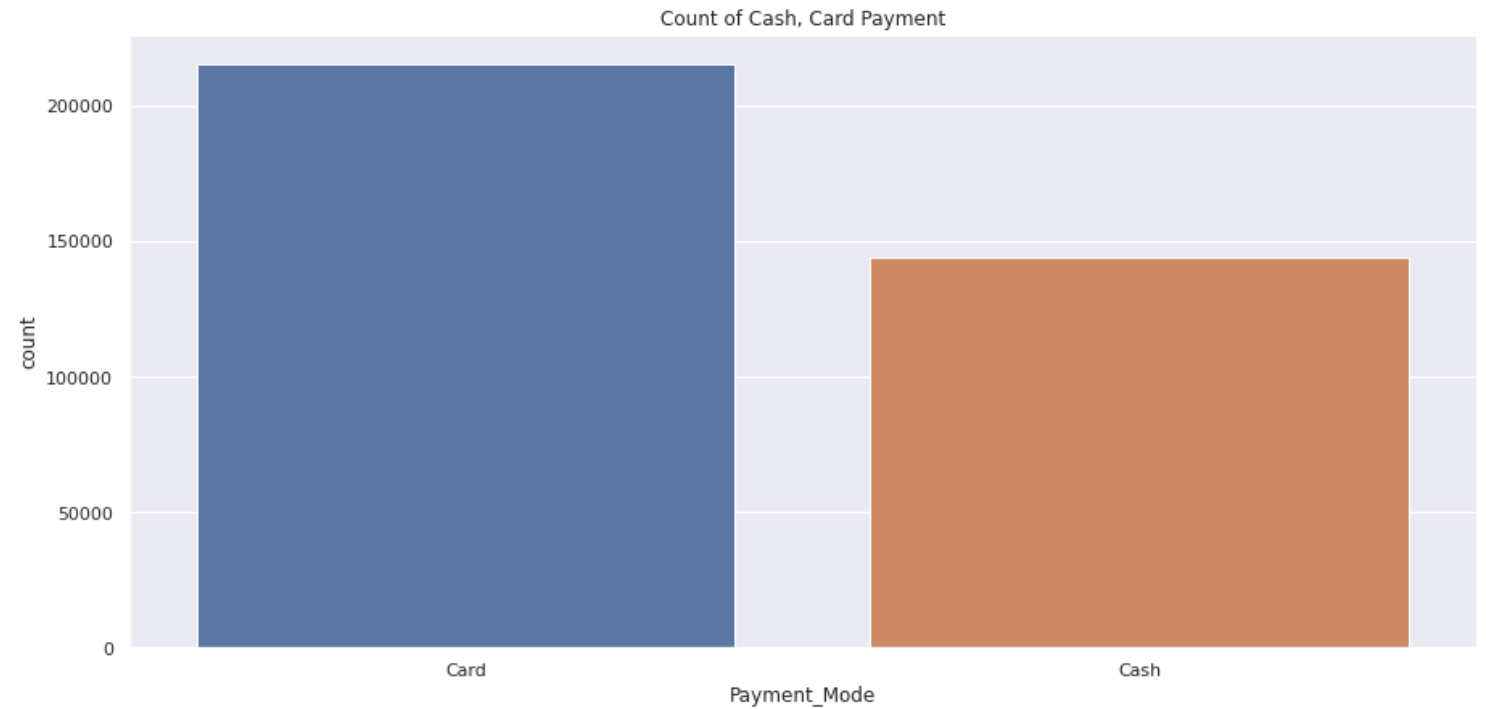Ride share per gender per cab
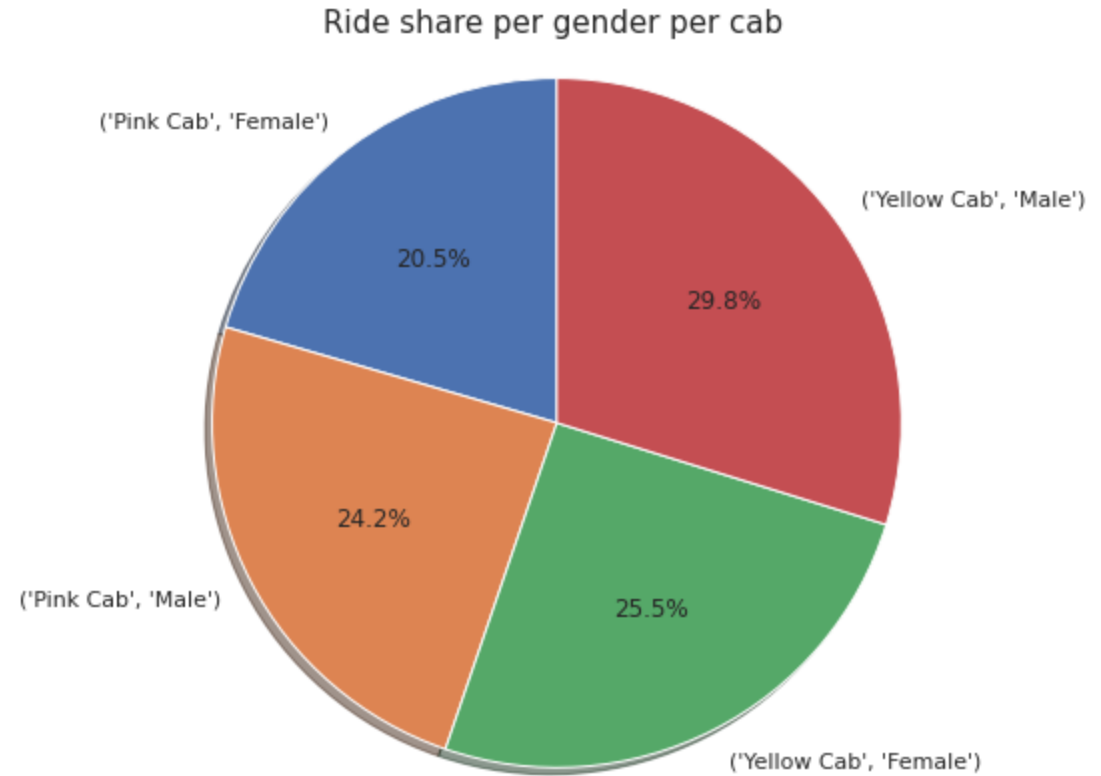
No. of Male & female user

- No. of male users are more than female users

Count of Male, Female

Payment Mode

- No. of Card payment is more than no. of cash payment
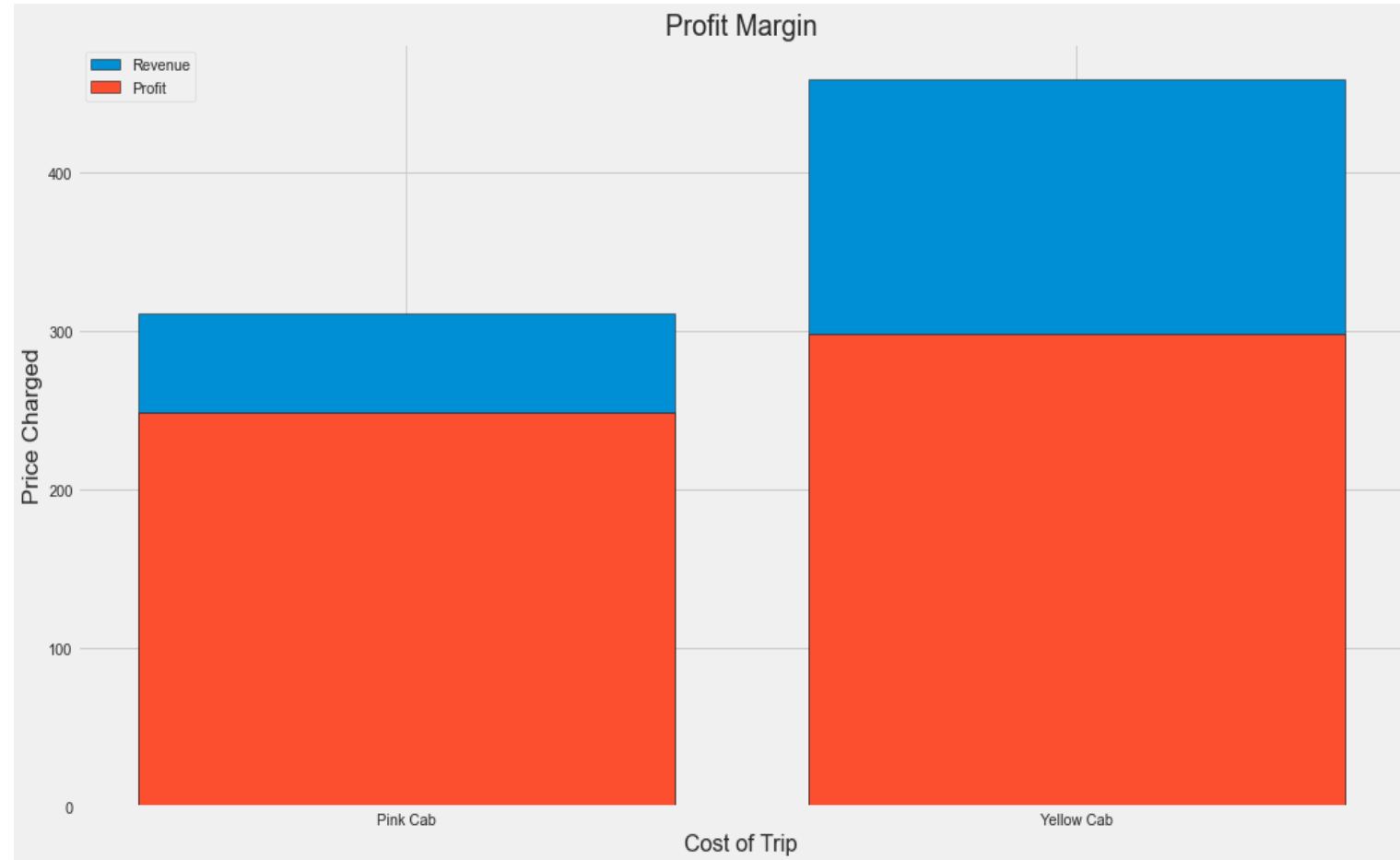

Count of Cash, Card Payment

Users per Cities

- New York City- 28%
- Chicago-16% `
- Los Angeles-13%

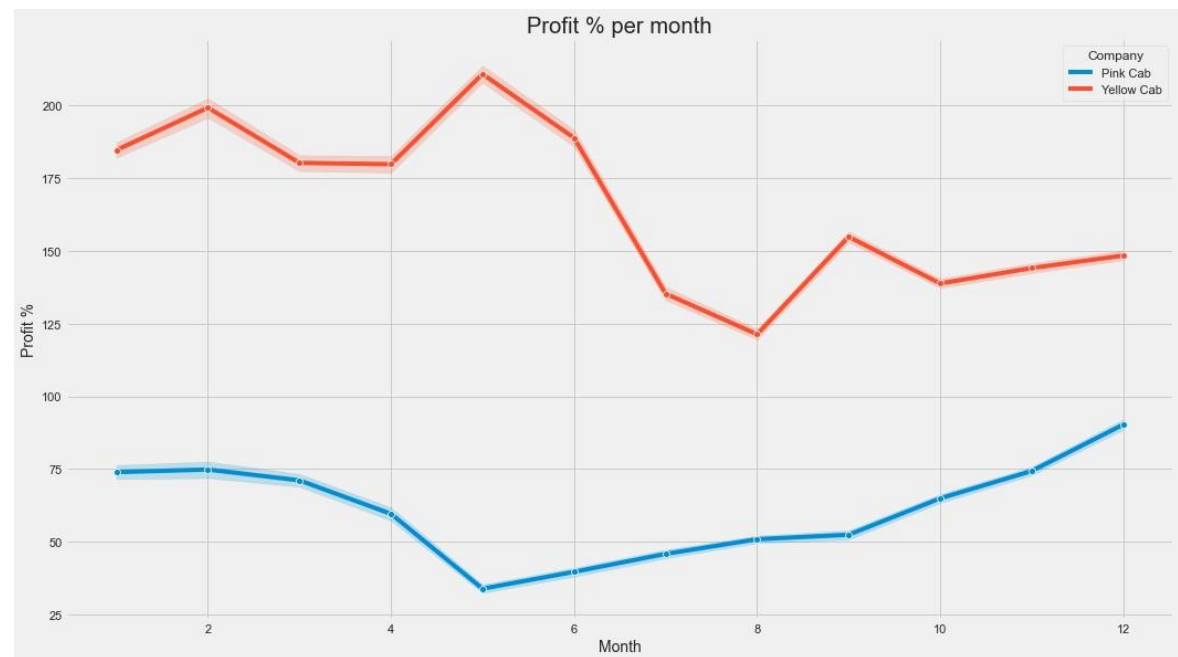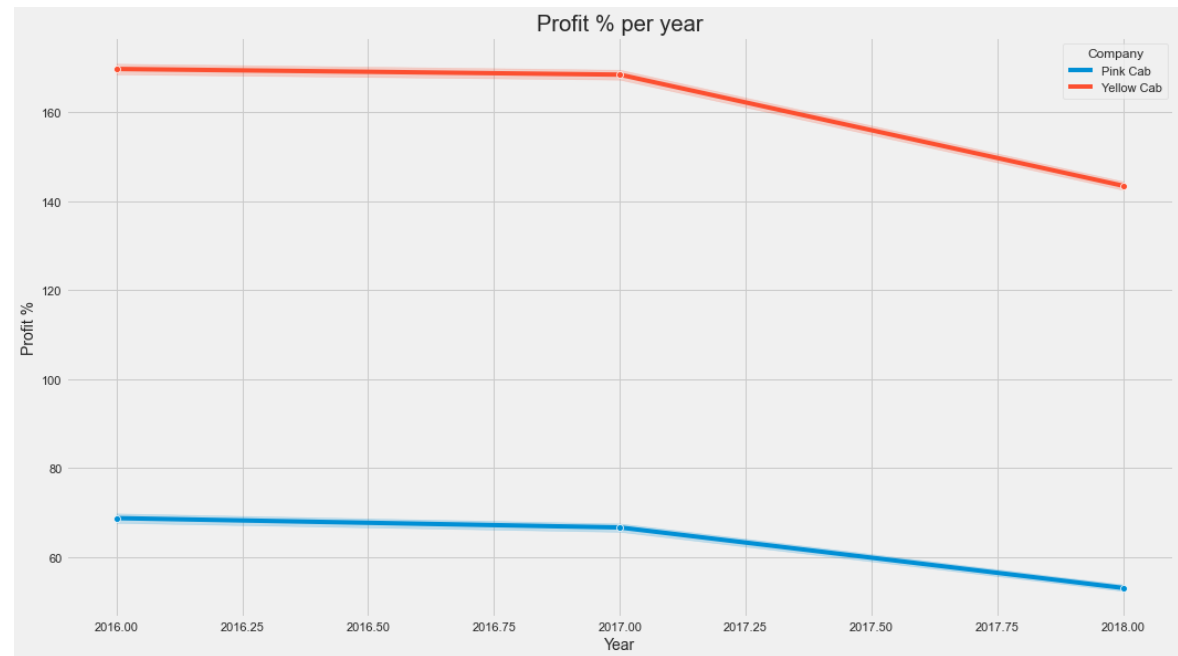Users per City
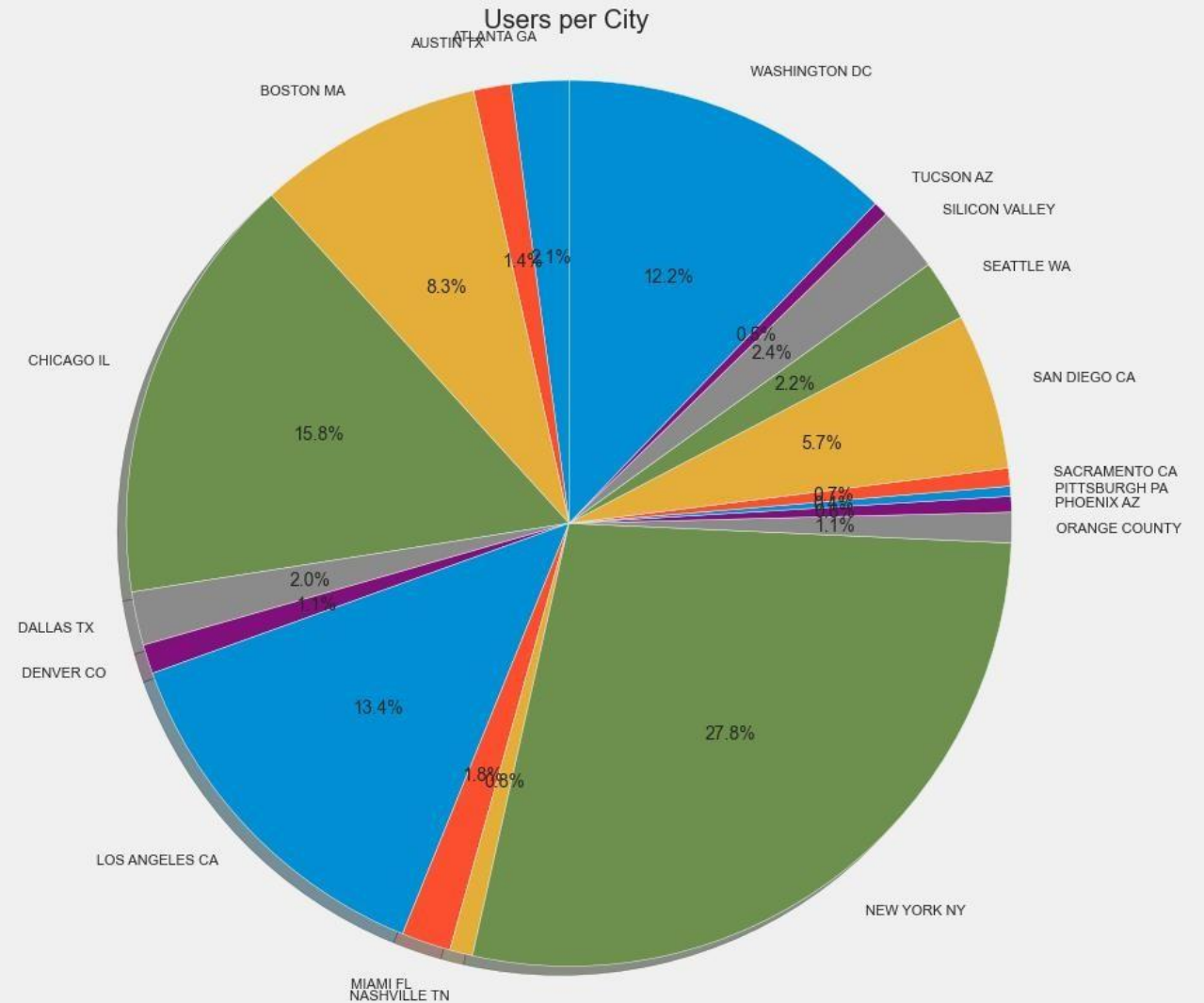
Profit & Cost Analysis

- The Yellow cab has a higher Profit Margin compared to Pink cab

Profit Per Year
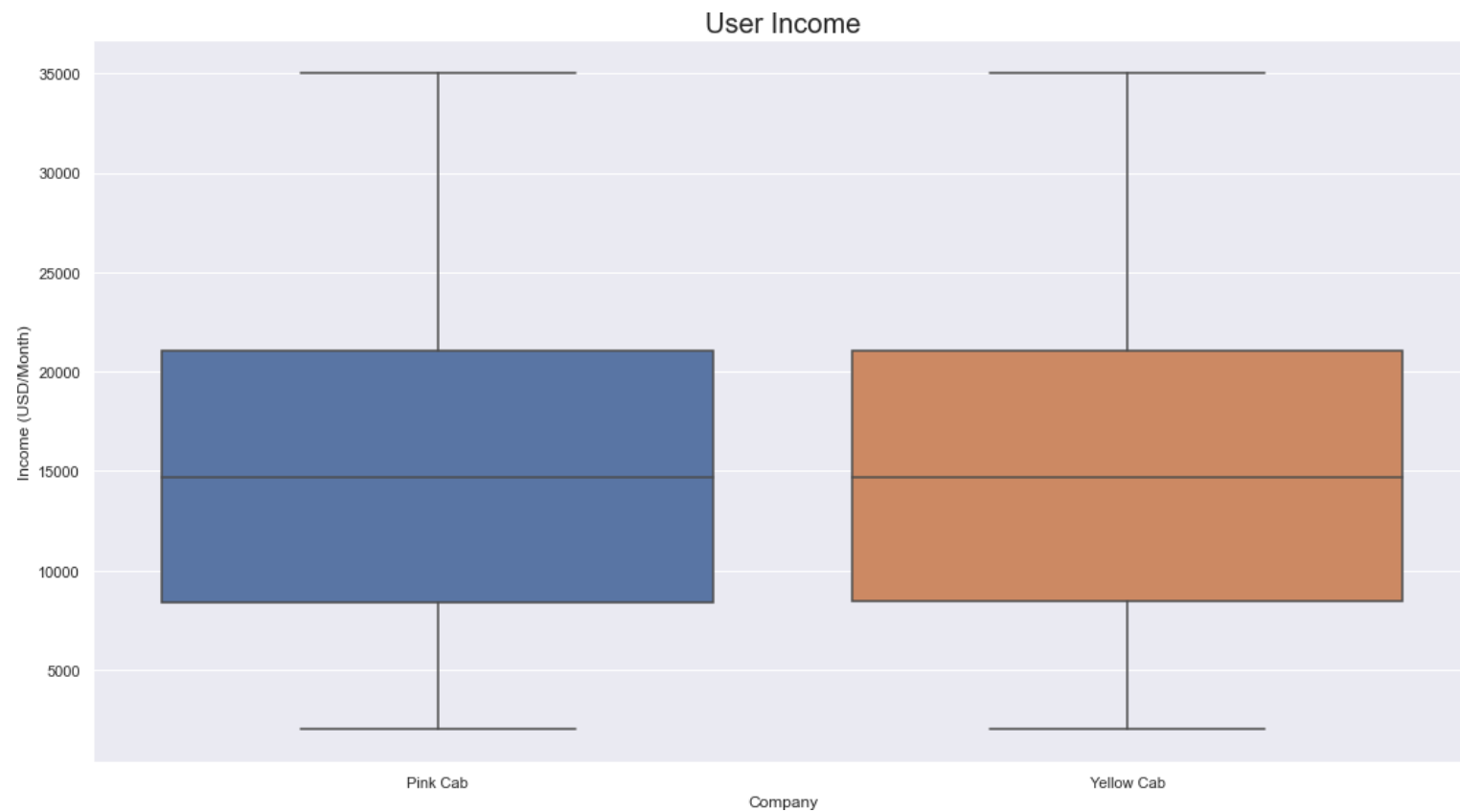
- The profit margin has decreased with respect to year

Average Income of Users

- Average income is around USD15k who use cab sevice

Hypothesis Test

Hypothesis 01: Charged
Price Same or Not

- ▶ H0 = Price charged by Pink & Yellow Cabs are same
- ▶ H1 = Price charged by Pink & Yellow Cabs are not same

```
p-value: 0.5033034260404716
We are accepting null hypothesis (H0)
```

Hypothesis 01: Charged Price Same or Not

- ▶ H0 = Price charged by Pink & Yellow Cabs are same
- ▶ H1 = Price charged by Pink & Yellow Cabs are not same

```
p-value: 0.5033034260404716
We are accepting null hypothesis (H0)
```

Conclusion

There is no difference in charged price in both cabs

Hypothesis 02: Difference Between Age

- H0 = There is no difference in Age in both cab companies.

- H1 = There is a difference in Age in both cab companies.

```
80125 5429
P value is  0.4816748536155635
We accept null hypothesis (H0) that there is no difference regarding age for Pink Cab
```

```
260356 17257
P value is  6.328485471267631e-05
We accept alternative hypothesis (H1) that there is a difference regarding age for Yellow Cab
```

Conclusion

There is a difference in Age in both cab companies.

Hypothesis 03: Profit difference for both cab services

- H0 = Profit is same for both cab services

- H1 = Profit is not same for both cab services

```
p-value: 0.0
We are rejecting null hypothesis (H0)
```

Conclusion

- There is no difference in Profit in both cab companies.

Hypothesis 04: Difference in Profit in Payment mode

- H0 = There is no difference in Payment Mode in both cab companies.

- H1 = There is a difference in Payment Mode in both cab companies.

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

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

Conclusion

There is difference in payment mode for both cab companies

Hypothesis 05: Travelled distance difference for both cab services

- H0 = There is no travelled difference in both cab companies.

- H1 = There is travelled difference in both cab companies.

p-value: 0.07571452137168612
We are accepting null hypothesis (H0)

Conclusion

There is no difference in payment mode for both cab companies

Conclusion

Conclusion

- Yellow Cab Company is better than Pink Cab Company Because:
  - ✓ Profit Margin
  - ✓ More Users
  - ✓ More transactions per Year

Thank You