G2M insight for Cab Investment firm

Data Analyst Virtual Internship

submitted by : Ishwarya Arulvel

Date :11-oct-2022

Agenda

- Description
- Data Information
- Data Analysis
- EDA
- EDA Summary
- Hypothesis Testing
- Recommendations

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,
  - Recommendation.

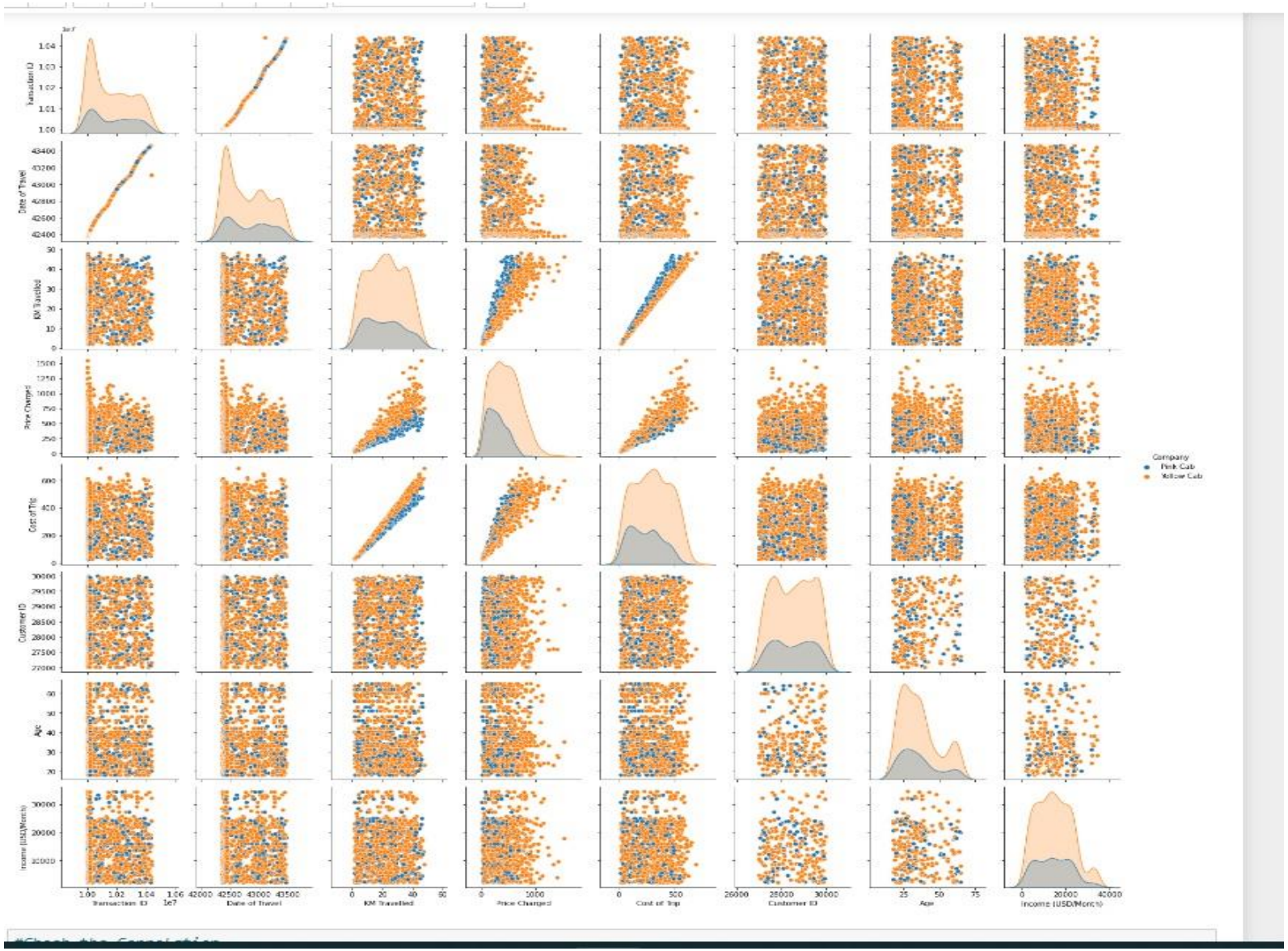
Data Information:

There are 4 datasets:

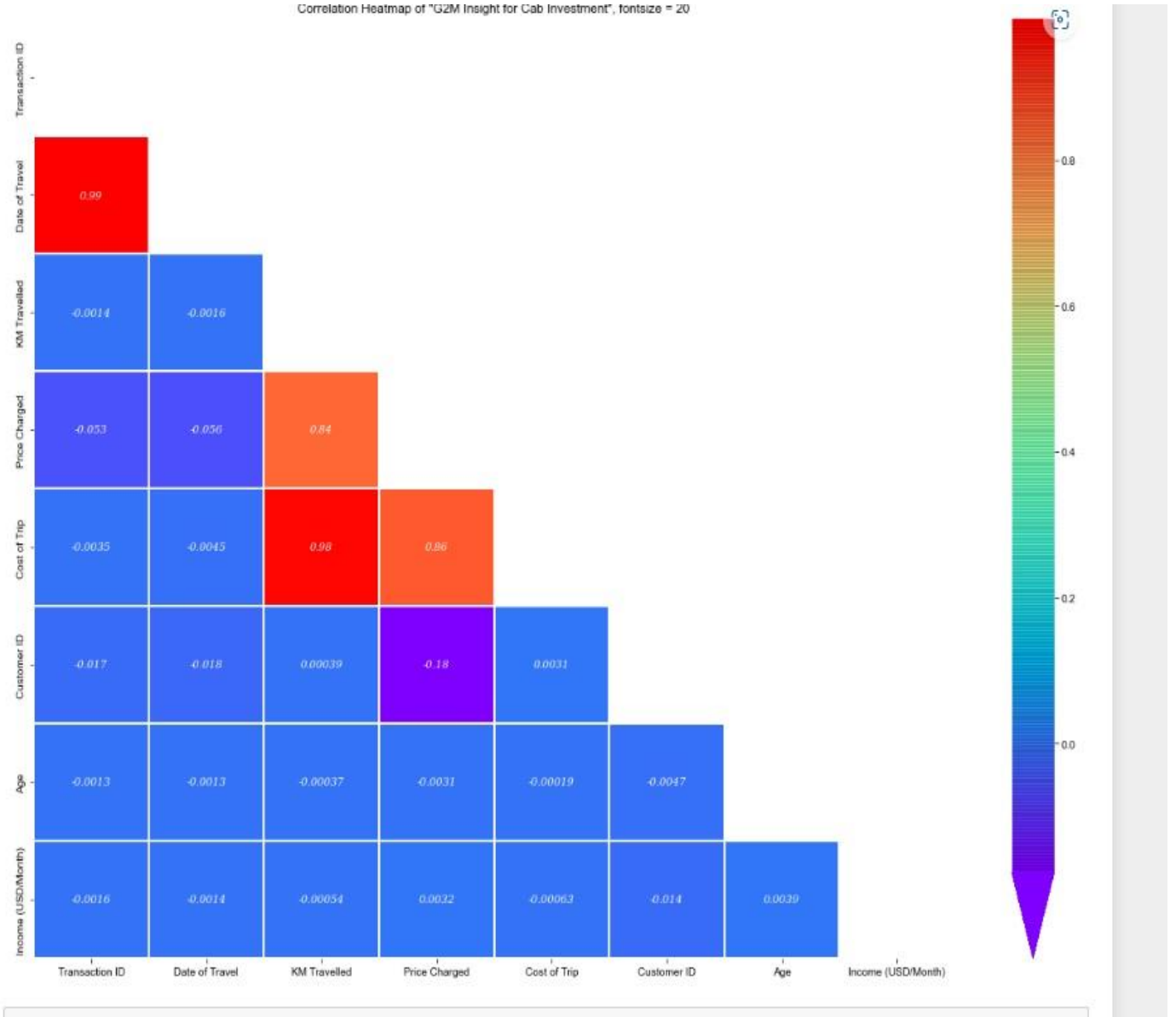
- Cab_Data.csv – this file includes details of transaction for 2 cab companies.
- Customer_ID.csv – this is a mapping table that contains a unique identifier which links the customer's demographic details.
- Transaction_ID.csv – this is a mapping table that contains transaction to customer mapping and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

DATA ANALYSIS

Relationships Between Variables

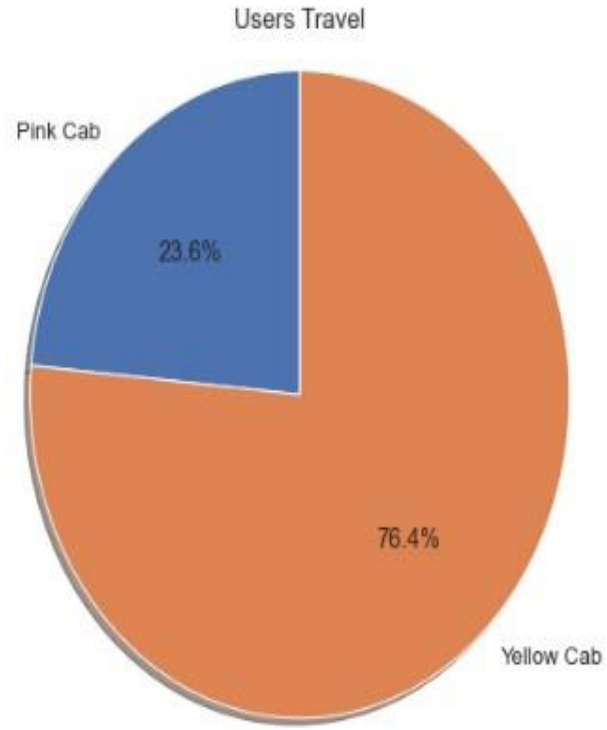


Correlation Between Variables

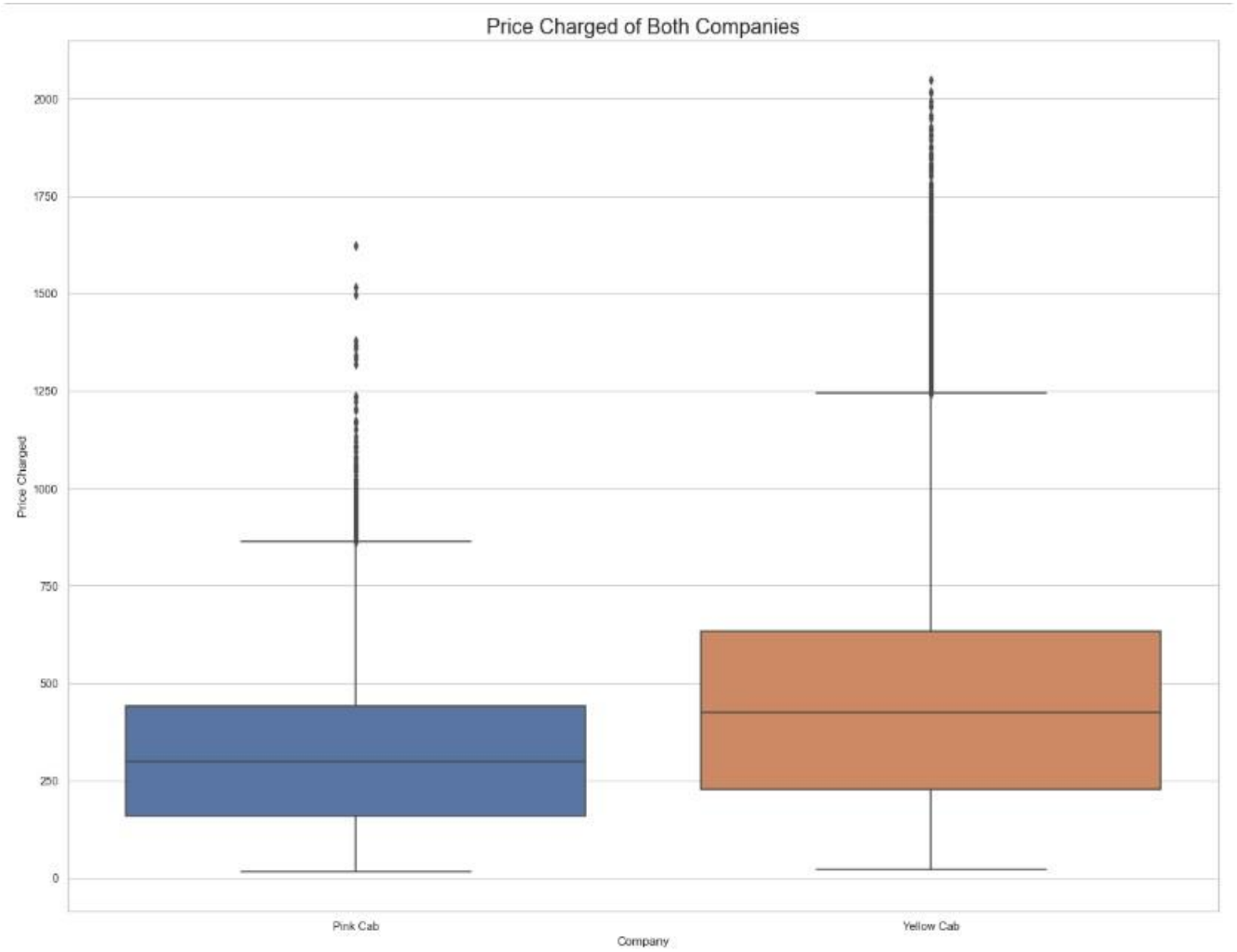


EDA

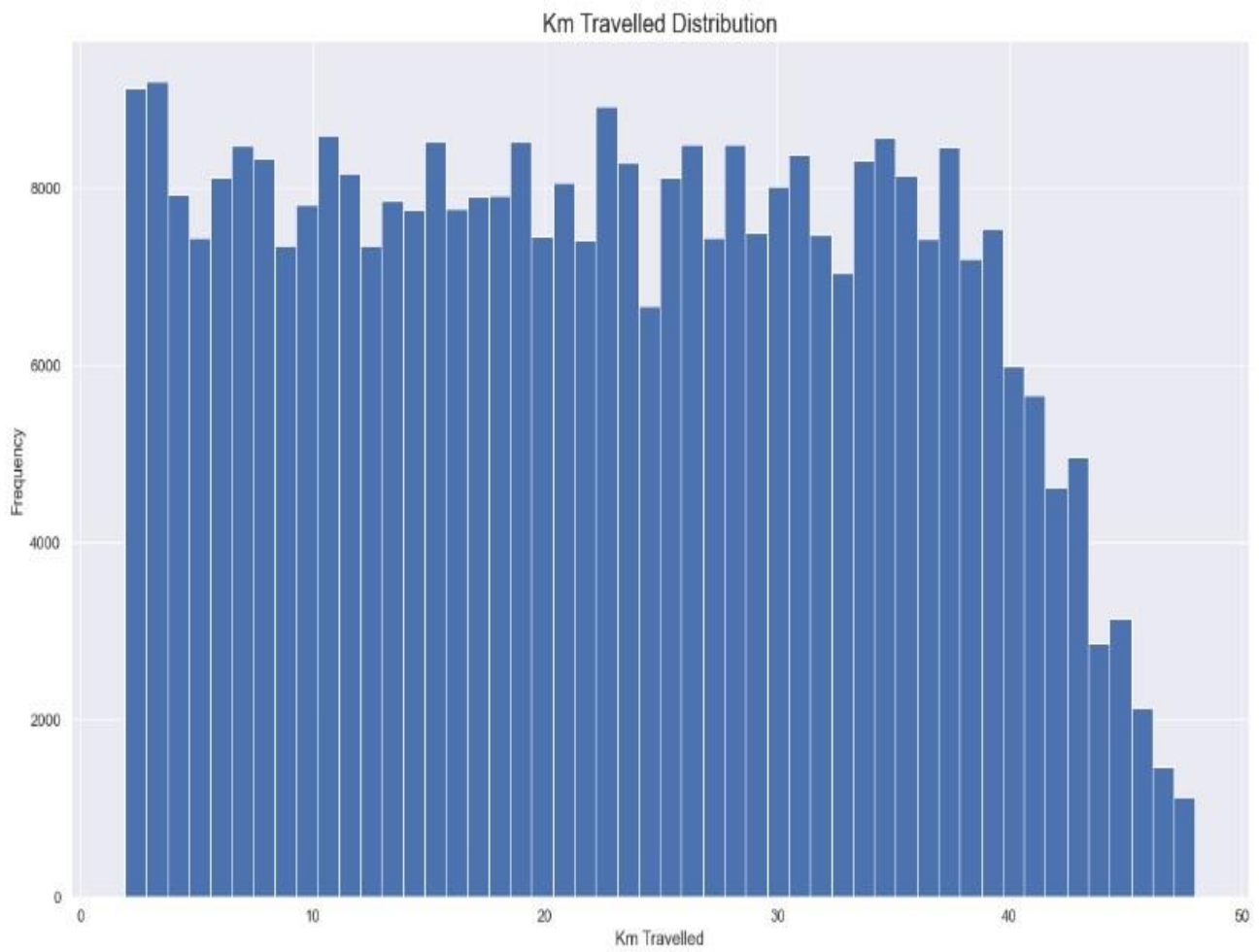
As we can see users like to ride on Yellow cab more as compared to Pink Cab



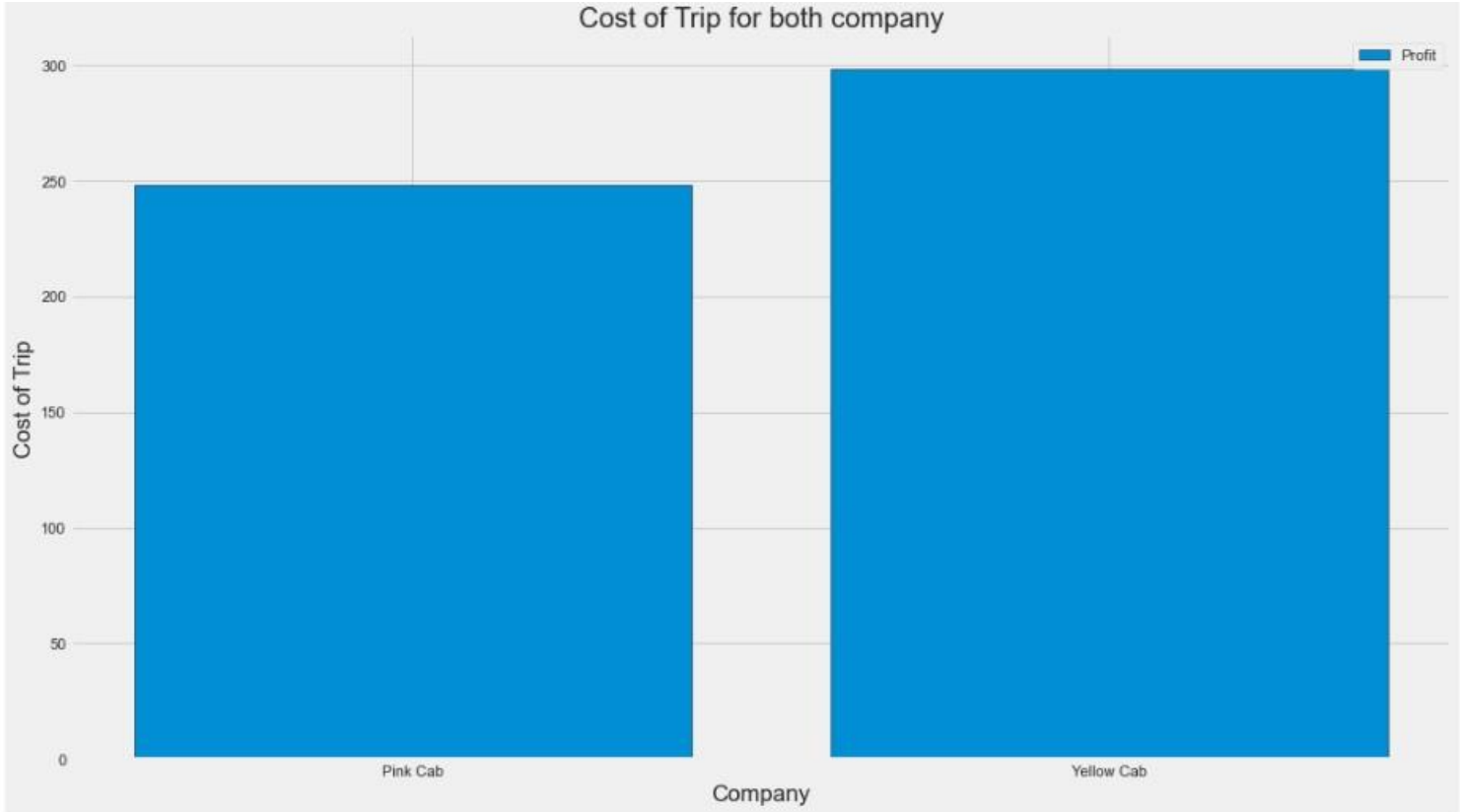
As we can see Price Charged for Yellow Cab is highest as compared to Pink Cab



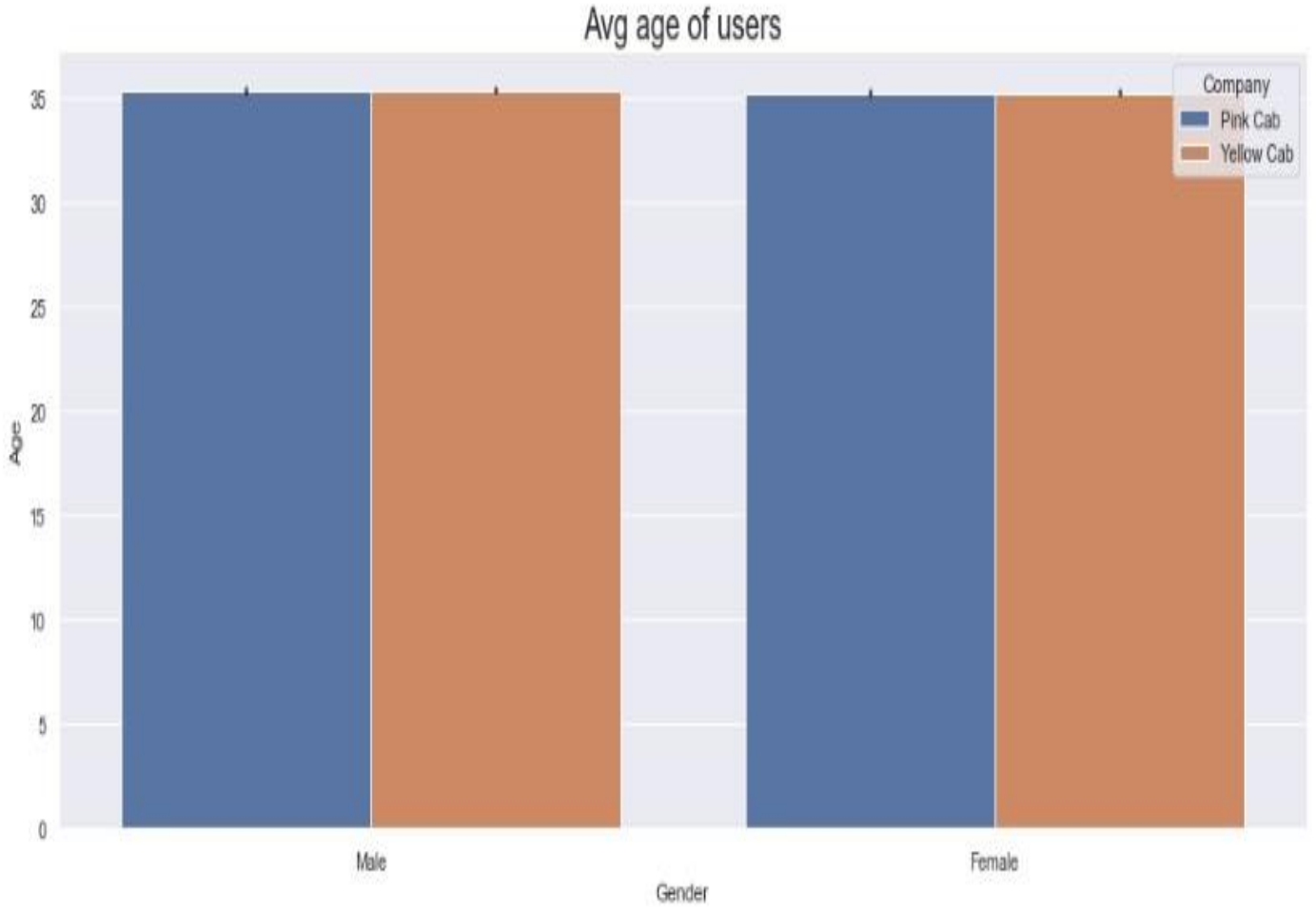
Most of the rides varies from 2 to 48 KM.



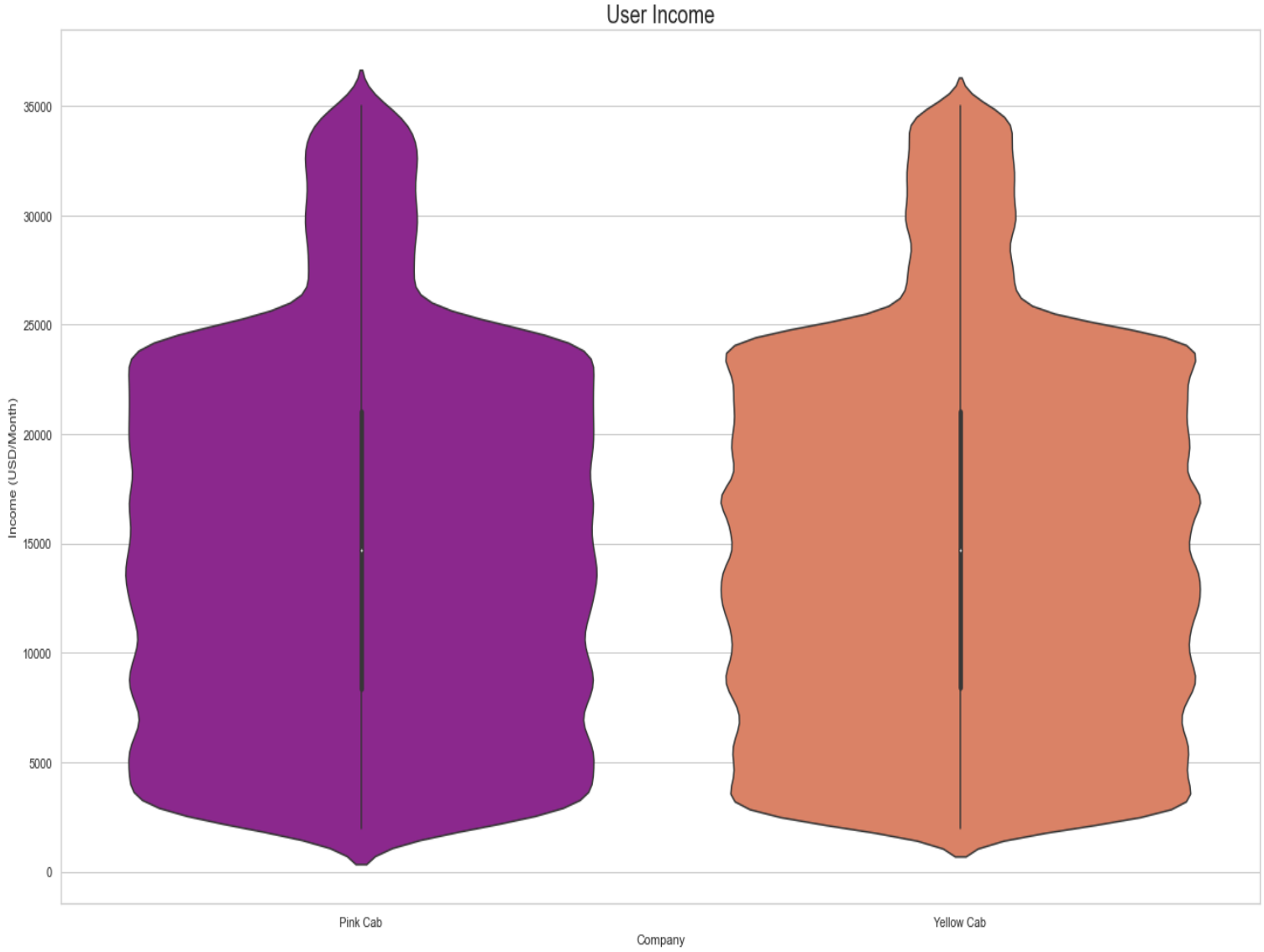
Yellow Cab has higher cost of trip compared to Pink Cab



Average Age per users: From the graph it shows 35 Average age of female and male who use cab service.



From the graph it shows Average income is around $35000 who use cab service



EDA SUMMARY

Pink Cab

- Rides are in the range of approximately 2 to 48 KM.
- Price Charge range from 150 to 450 dollars.
- The pink has 23.6% of users.
- The maximum cost of Trip around 250 dollars.

Yellow Cab

- Rides are in the range of approximately 2 to 48 KM.
- Price Charge range from 250 to 600 dollars.
- The pink has 76.4% of users.
- The maximum cost of Trip around 300dollars.

Hypothesis Testing

Hypothesis 1: There is no difference gender in pink cab

- Pink cab: We accept hypothesis that there is no difference regarding gender for pink cab

```
print('P value is ', p_value)
```

- Pink Cab Result:
- 47231 37480
- P value is 0.11515305900425798
- P>0.05
- Therefore ,accept hypothesis that there is no difference regarding gender for Pink Cab

Hypothesis 2: There is difference gender in yellow cab

- Yellow Cab : We accept hypothesis that there is a difference regarding gender for yellow cab

```
print('P value is ', p_value)
```

- Yellow cab Result:
- 158681 116000
- P value is 6.060473042494144e-25
- P<0.05
- So , we accept hypothesis that there is a difference regarding gender for Yellow Cab

Hypothesis 3: There is no difference regarding age for pink cab

- Pink cab: We accept hypothesis that there is no difference regarding age for Pink Cab

```
print('P value is ', p_value)
```

- Pink Cab:
- 71228 14405
- P value is 0.3905344473325604
- P>0.05
- We accept hypothesis that there is no difference regarding age for Pink Cab

Hypothesis 4:There is difference regarding age for yellow Cab

- Yellow Cab: We accept alternative hypothesis that there is a difference regarding age for Yellow Cab

```
print('P value is ', p_value)
```

- Yellow cab:
- 231480 46368
- P value is 3.893841736668538e-05
- P<0.05
- We accept hypothesis that there is a difference regarding age for Yellow Cab

Hypothesis 5: There is no difference in payment mode for Pink Cab

- Pink Cab: We accept hypothesis that there is no difference in payment mode for Pink Cab

```
print('P value is ', p_value)
```

- Pink Cab:
- P value is 0.7900465828793288
- P>0.05
- We accept hypothesis that there is no difference in payment mode for Pink Cab

Hypothesis 6: There is no difference in payment mode for Yellow Cab

- Yellow Cab: We accept hypothesis that there is no difference in payment mode for Yellow cab

```
print('P value is ', p_value)
```

- Yellow Cab Result:
- P value is 0.2933060638298729
- P>0.05
- We accept hypothesis that there is no difference in payment mode for Yellow Cab

Recommendation

Yellow Cab Company is better than Pink Cab Company Because:

- **Profit**
- **More Users**

Thank You