



Data Glacier

Your Deep Learning Partner

Exploratory Data Analysis

G2M insight for Cab Investment firm

18 April 2023

Bindu Musham

Agenda

Executive Summary

Problem Statement

Approach

EDA

Recommendations

Conclusion

Executive Summary

- As part of this project, we have analyzed 4 datasets that contains information on 2 cab companies.
- Performed EDA to understand number of users, high profit company, company with high prices, customer demographics on company's profits.
- Based on the analysis, we have provided the actional insights to help the XYZ company to invest in the right company.

Problem Statement

XYZ is a private 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 and as per their Go-to-Market(G2M) strategy they want to understand the market before taking final decision.

The goal of the project is to provide actionable insights to help XYZ firm to identify the right company to make their investment.

Approach

- Descriptive analysis: Understand the data and distribution.
- Data transformation: change the data types based on the domain knowledge and merge the datasets.
- EDA: Perform EDA to analyze data to answer possible questions.
- Hypothesis testing
- Provide actionable insights or recommendations to help make a final decision.

EDA



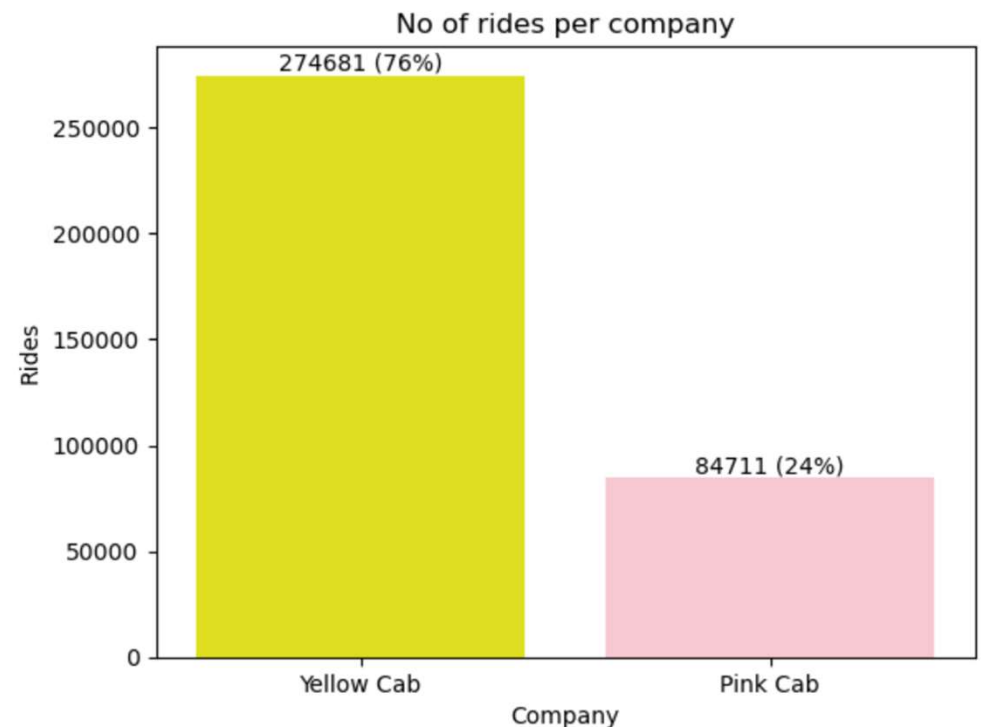
Data Glacier

Your Deep Learning Partner

Which cab company has more rides?

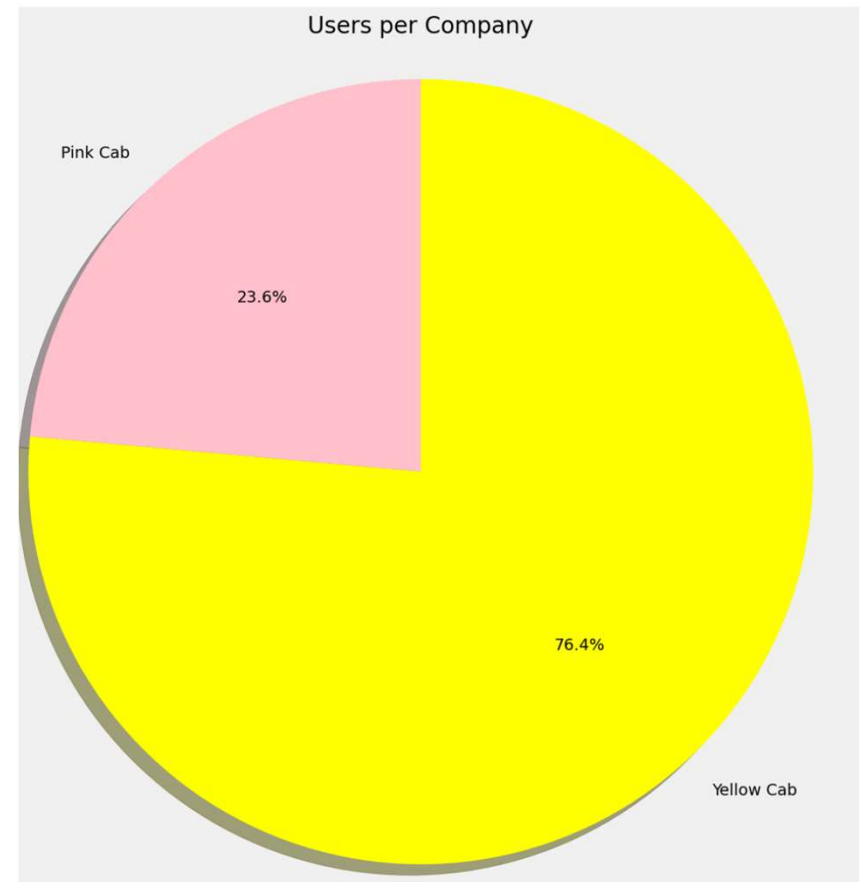
Yellow Cab company(76%) has more rides than Pink Cab company(24%).

Yellow Cab has 50% more transactions or rides compared to Pink Cab.



Which cab company has higher percentage of users?

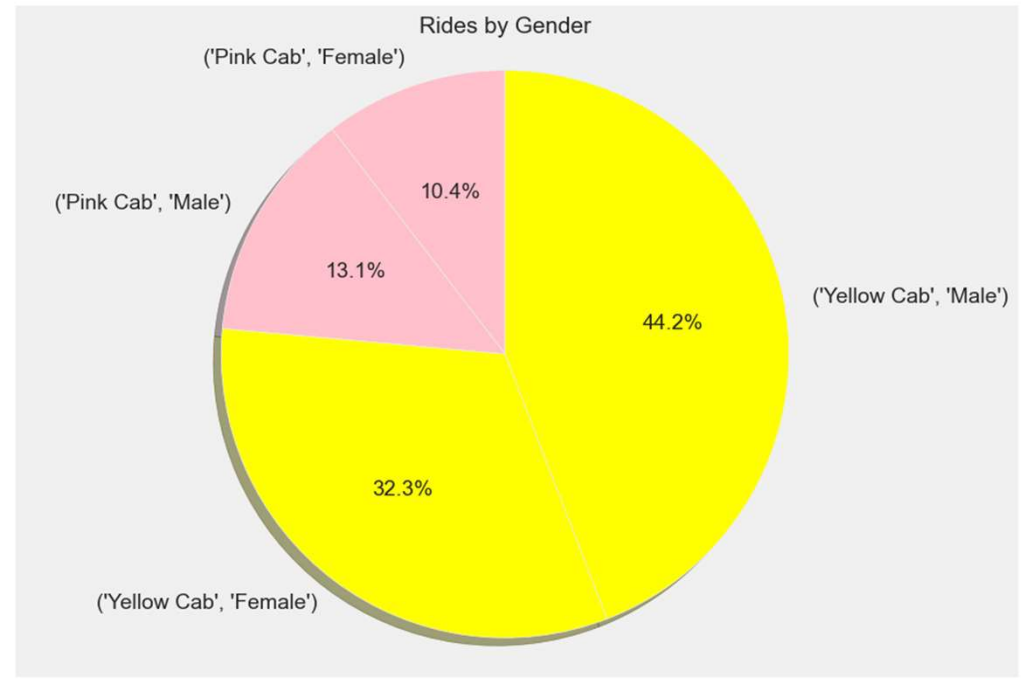
More number of users chose Yellow cabs(76.4%) compared to Pink cabs (23.6%).



Does gender have any effect on number of rides?

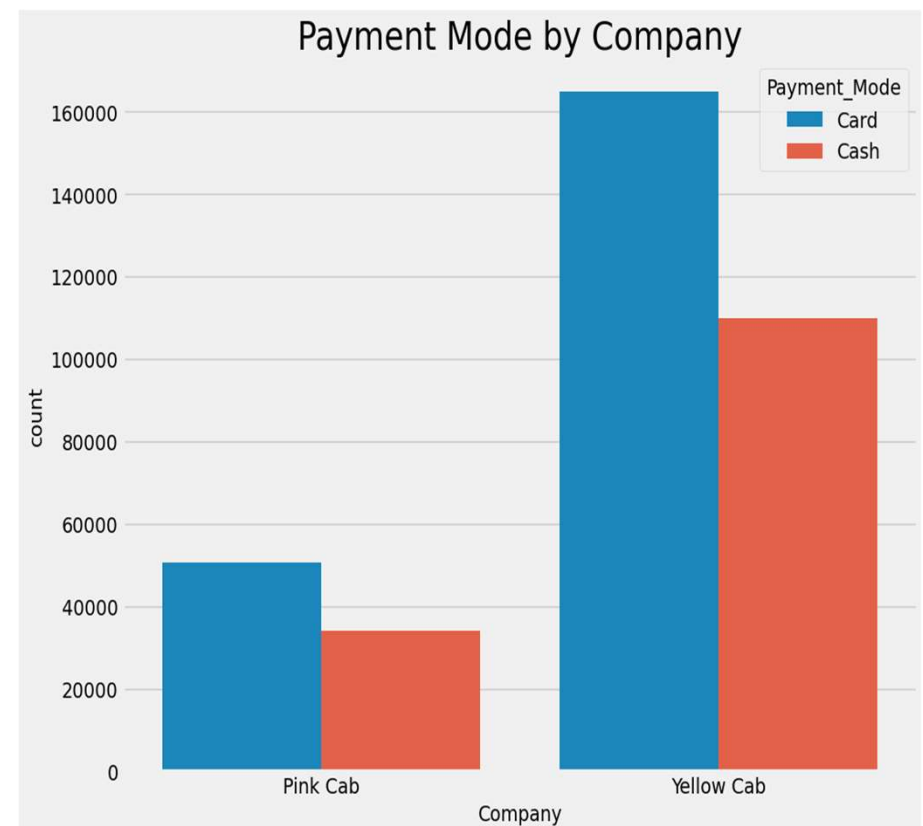
Both Cab rides are taken more by male customers than female customers.

Yellow cab have higher percentage of male customers and female customers compared to Pink Cab



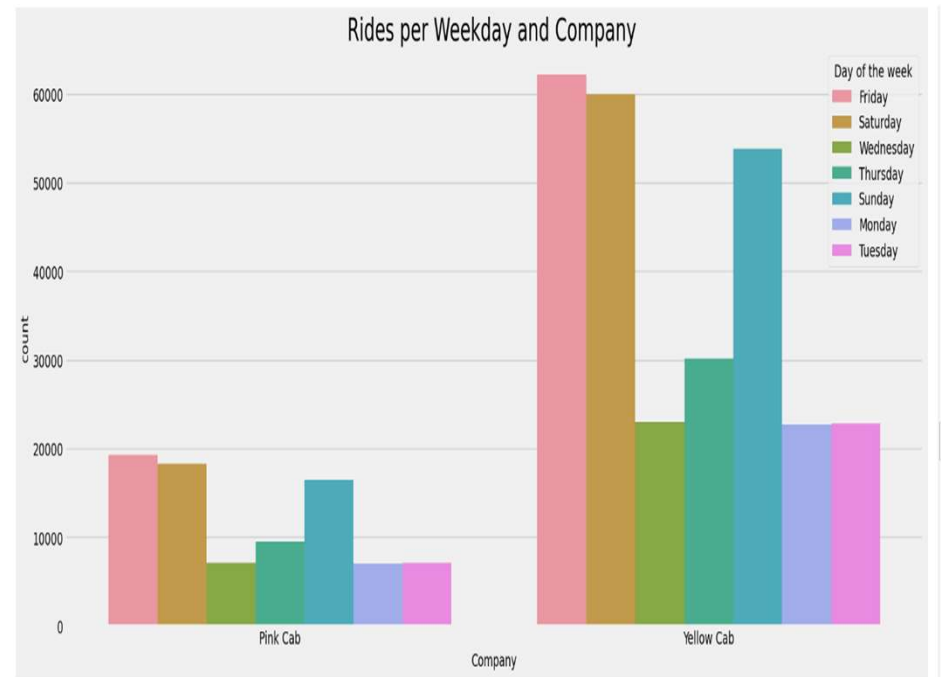
Mode of payment Vs company?

Both cab companies have higher number of card payments compared to cash payments. Users preferred to pay in card than cash payments.



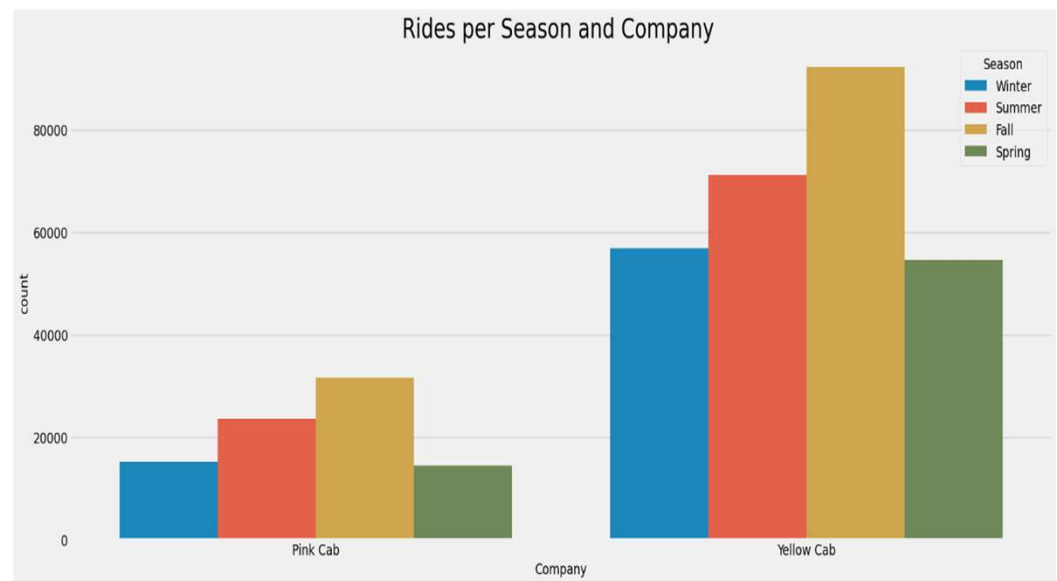
Are the number of rides depending on the day of the week?

Both have similar effect based on the day of the week. Higher cab rides are on Friday and lowest rides on Wednesday. Saturday and Sundays i.e., weekends have comparatively high rides.



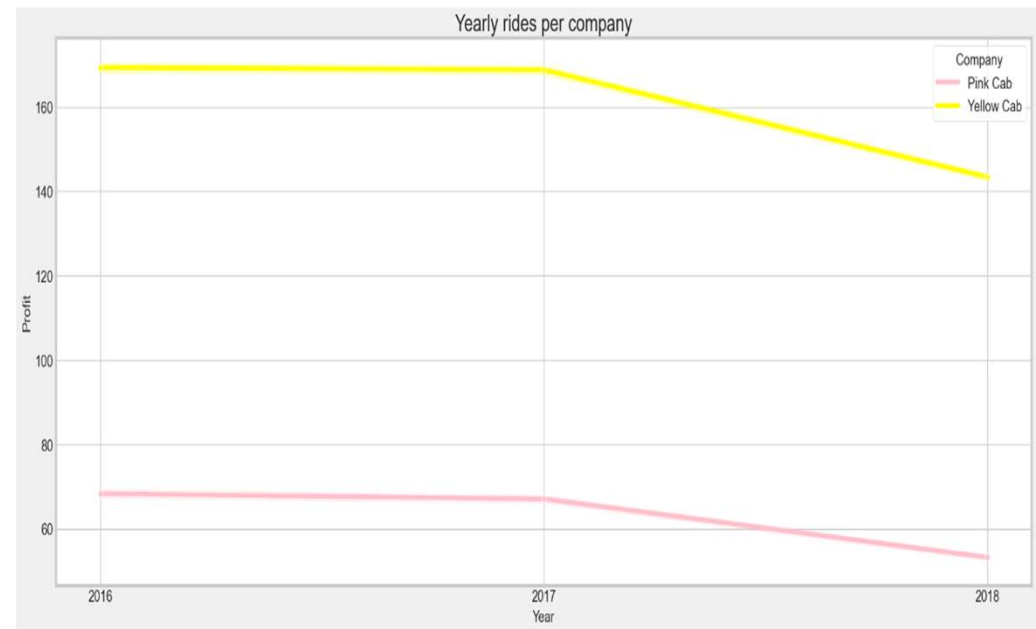
Is there any seasonality in the usage of cabs?

The number of rides for both companies have seasonality effect. Higher number of rides taken in Fall and lower in Winter.



Does profits vary across Years?

For both companies, profits were same in 2016 & 2017 and declined in 2018.

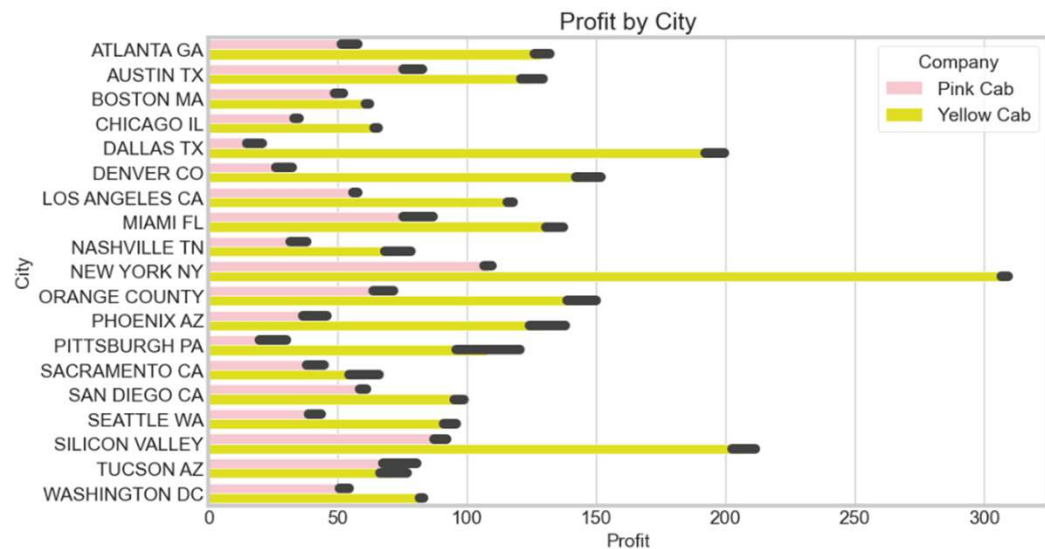


Which City have highest profits?

Pink Cab has lower profits in Dallas and higher profits in New York.

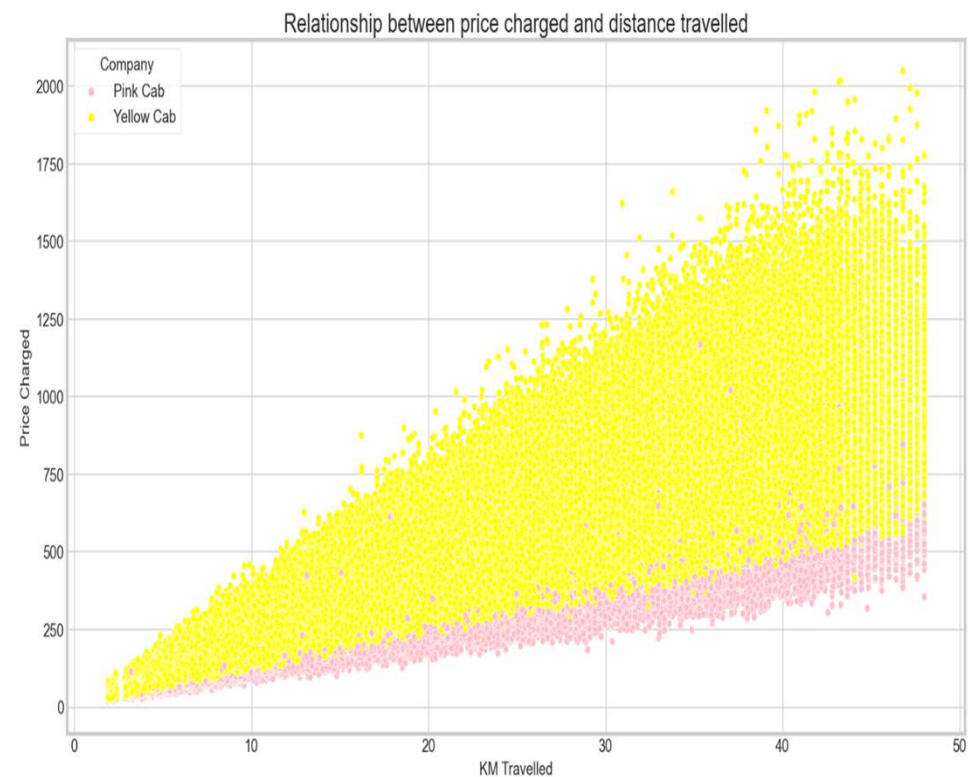
Yellow Cab have lower profits in Boston and highest profits in New York.

Overall, New York city have higher profits.



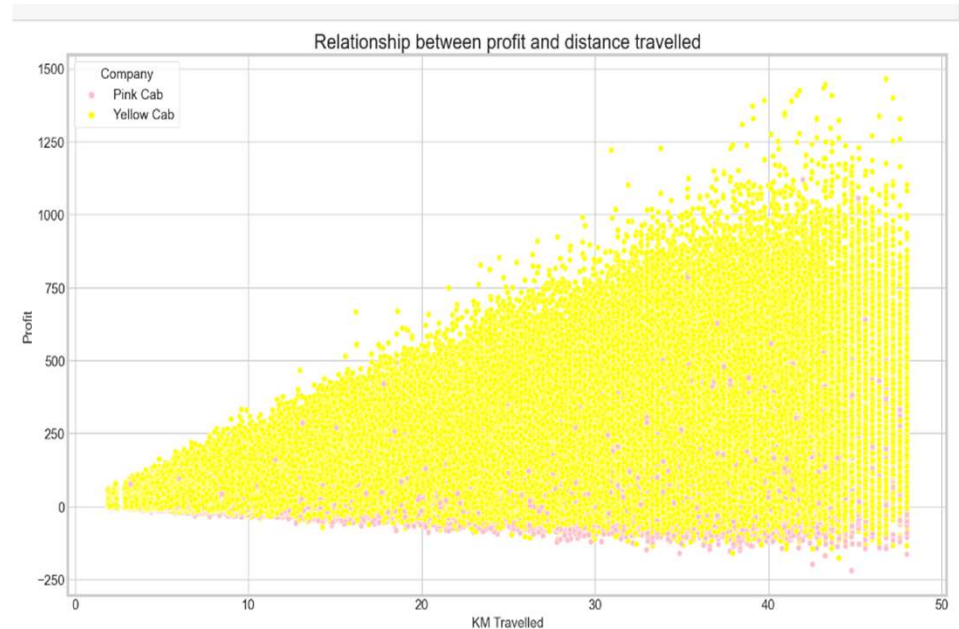
Is there any relationship between price charged and distance travelled?

We can see price charged and distance travelled are directly proportional to each and have positive linear relationship. Yellow cabs have high charges than Pink cabs.



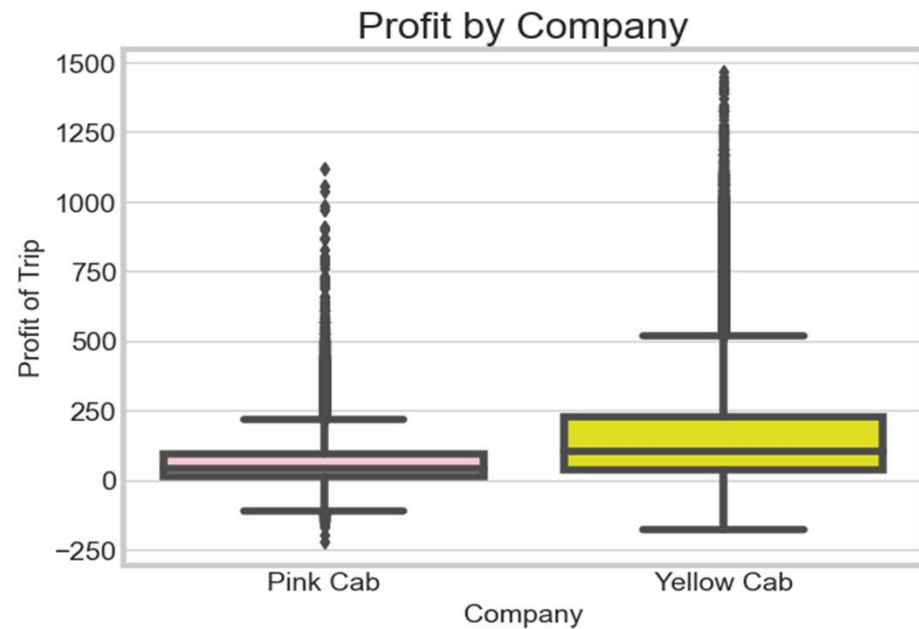
Is there any relationship between profit and distance travelled?

We can see profit and distance travelled are directly proportional to each and have positive linear relationship. Yellow cabs have higher profits than Pink cabs.



Are there any outliers in profits?

Yellow cabs have higher profits compared to Pink cabs. There are negative profits for Pink cabs.

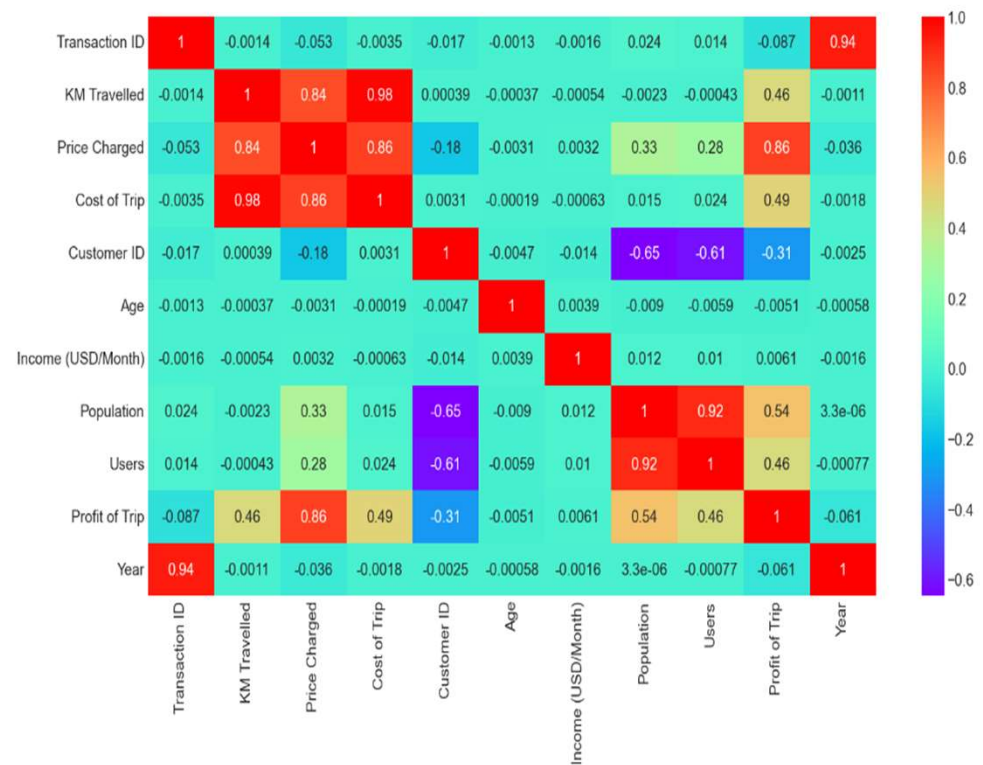


Is there any correlation between variables?

There are strong correlations between KM Travelled, Price Charged and Cost of Trip.

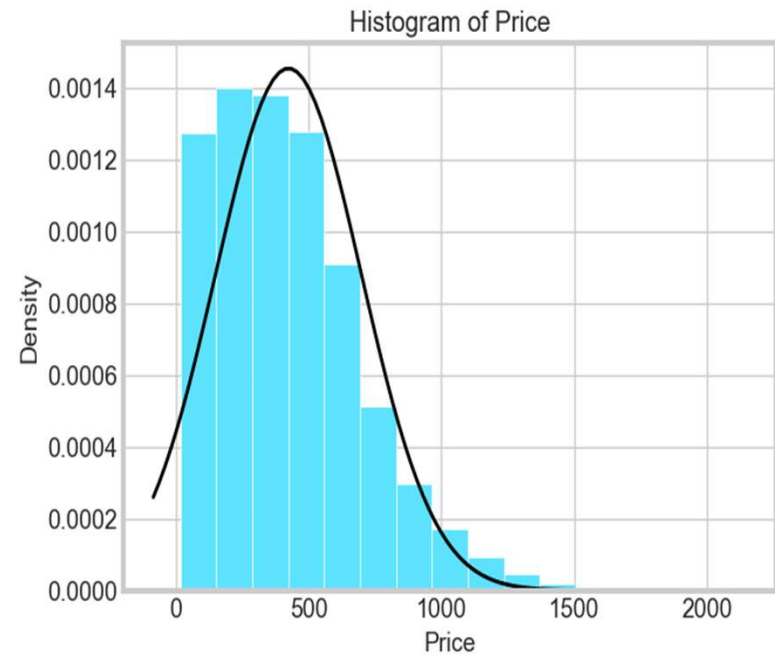
High correlation between Profit and Price Charged.

Users and Population are highly correlated.



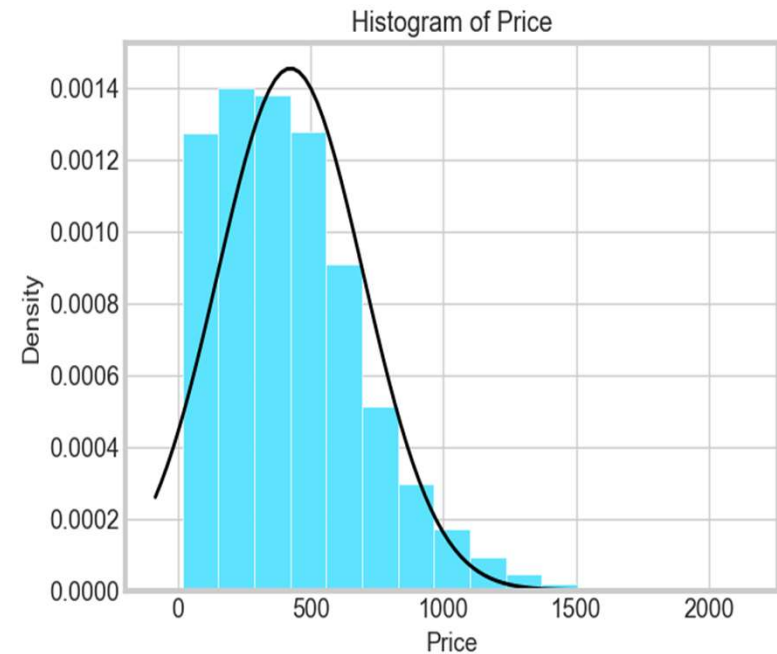
What are the prices charged for most of the rides?

Both cab companies have prices charged in the range of 125-375 dollars.



Distribution of age

Higher number of customers are in the age group of 20-22 years and lower in the range of 50-60 years.



Hypothesis Testing

Hypothesis 1: Is there any difference between the average prices of Yellow and Pink cab rides.

Ho: Avg price of Yellow cab rides = avg price of pink cabs
Ha: Avg price of Yellow cab rides \neq avg price of pink cabs

```
Mean price of Pink Cab: 311.0  
Mean price of Yellow Cab: 458.0  
Ttest_indResult(statistic=-140.37224404354248, pvalue=0.0)
```

The p-value $0 < 0.05$. As a result, we reject null hypothesis and accept alternative hypothesis.

Average prices of Yellow cab rides are not equal to average prices of Pink cab rides.

Hypothesis 2: Is there any difference between the average profit of Yellow and Pink cab rides.

Ho: Avg profit of Yellow cab rides = avg profit of pink cabs

Ha: Avg profit of Yellow cab rides \neq avg profit of pink cabs

Mean profit of Pink Cab: 63.0

Mean profit of Yellow Cab: 160.0

Ttest_indResult(statistic=-160.3715175947807, pvalue=0.0)

The p-value $0 < 0.05$. As a result, we reject null hypothesis and accept alternative hypothesis.

Average profit of Yellow cab rides are not equal to average profit of Pink cab rides.

Hypothesis 3: Is there any difference between the average profit of male and female customers.

Ho: Avg profit of male customers = avg profit of female customers

Ha: Avg profit of male customers \neq avg profit of female customers

Mean profit of male users of Pink Cab: 63.0

Mean profit of female users of Pink Cab: 62.0

Ttest_indResult(statistic=1.5754642478511207, pvalue=0.11515305900425798)

The p-value $0.1 > 0.05$. As a result, we cannot reject null hypothesis and gender has no effect on Profit in Pink Cab company.

Mean profit of male users of Yellow Cab: 163.0

Mean profit of female users of Yellow Cab: 156.0

Ttest_indResult(statistic=10.315494207195322, pvalue=6.060473042494056e-25)

The p-value $6.060473042494056e-25 < 0.05$. As a result, we reject null hypothesis and gender has significant effect on Profit in Yellow Cab company.

Recommendations

- We can provide discounts on Monday- Thursday to increase the rides on these days.
- Offer promotions in Winter when market is down.
- Give special discounts for long distance users to increase profits.
- Can offer sample products to attract female customers for cab rides.
- Invest more in New York city to get higher profits.

Conclusion

Yellow Cab is the right company to invest as per G2M strategy than Pink Cab due to higher number of rides, more users, high prices and high profits.

Thank You