



Data Glacier

Your Deep Learning Partner

G2M INSIGHT FOR CAB INVESTMENT FIRM

INTERN TRAINEE : VIDYA GANESAN

BATCH NUMBER: LISU06

PROJECT REVIEWER: DATA GLACIER

LOCATION: UNITED ARAB EMIRATES

DATE: 09-02-2022

Background –G2M(cab industry) case study

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.

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

Companies:

✓ Pink Cab

✓ Yellow Cab

The analysis has been analyzed,

☐ Data Understanding

☐ Data Visualization (EDA)

☐ Obtained Statistical Info

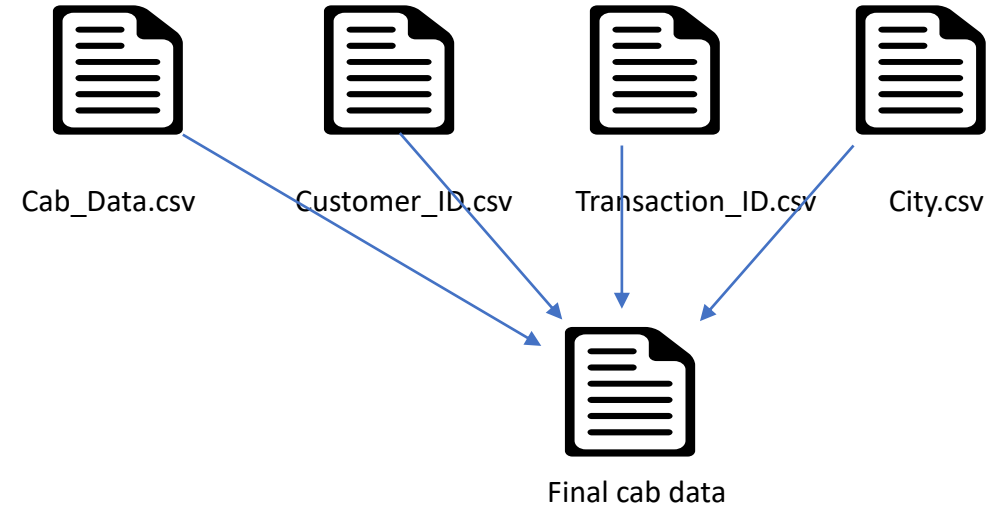
☐ Conducted data inferential analysis for the given data set.

Data Exploration

There are 4 datasets given in the format as .csv

- City_Data
- Customer_ID
- Transaction_ID
- City

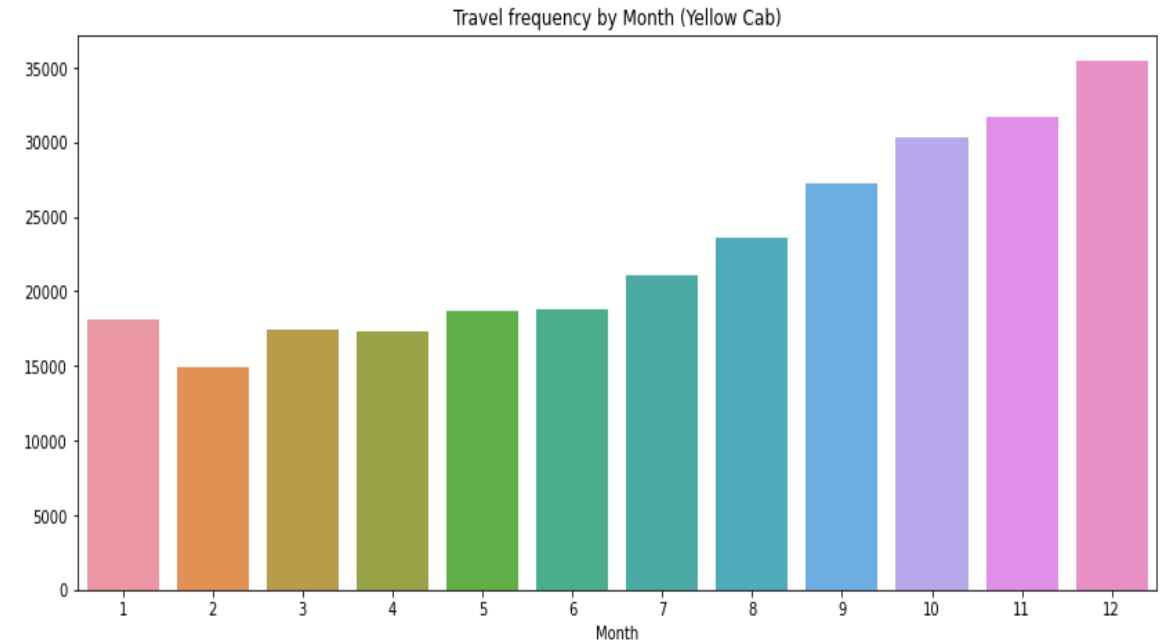
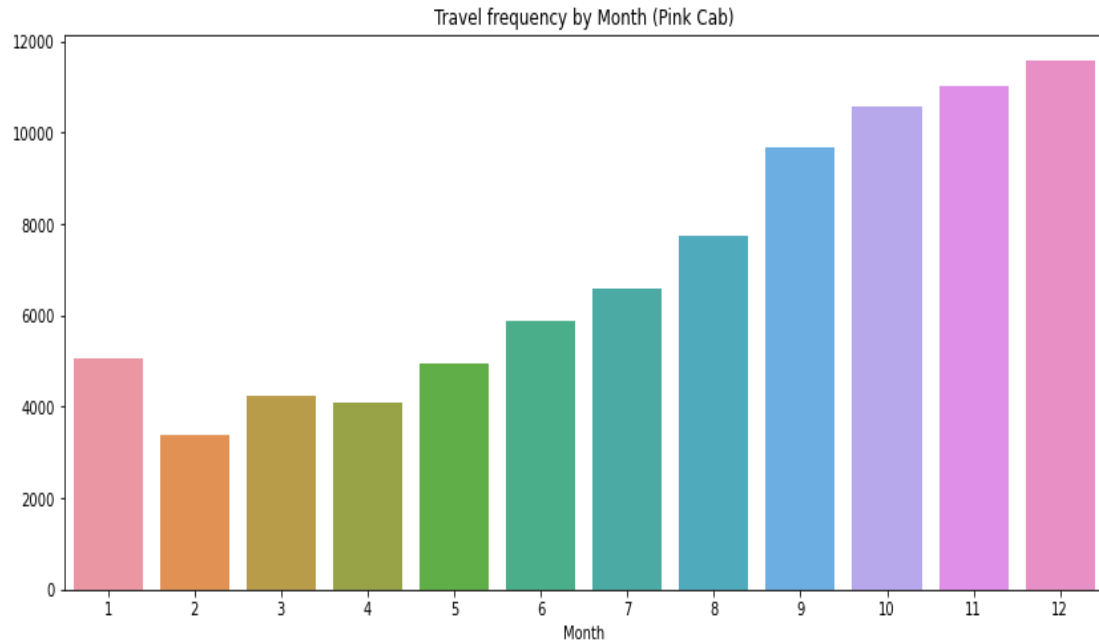
Merged these 4 datasets into a master dataset.



DATA VISUALISATION

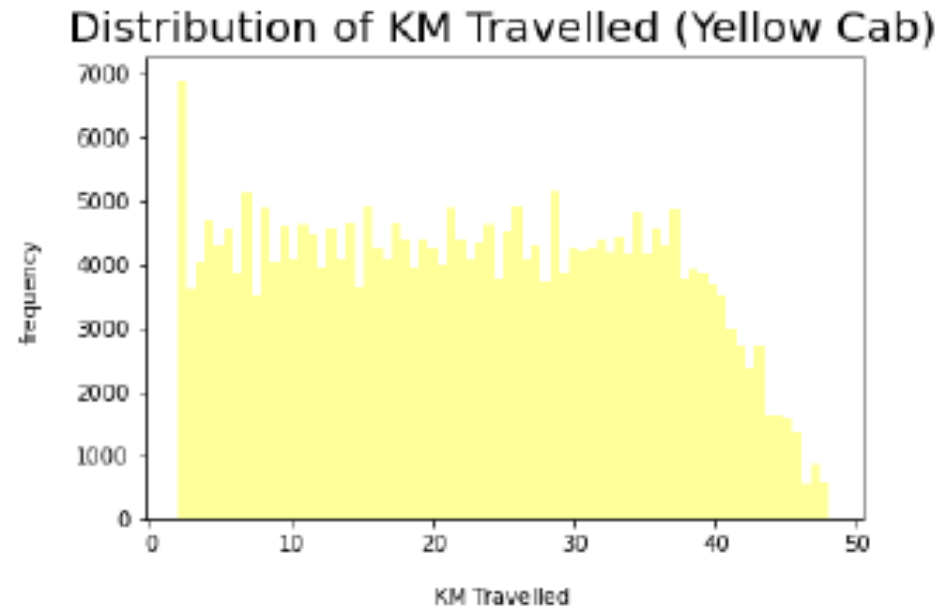
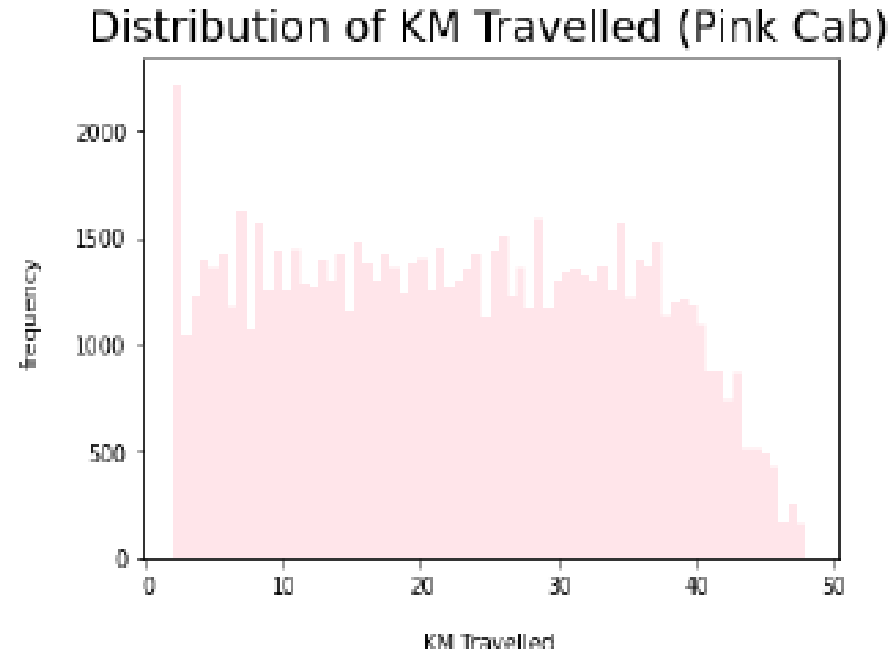
EXPLORATORY DATA ANALYSIS

Comparison of travel frequency between two companies



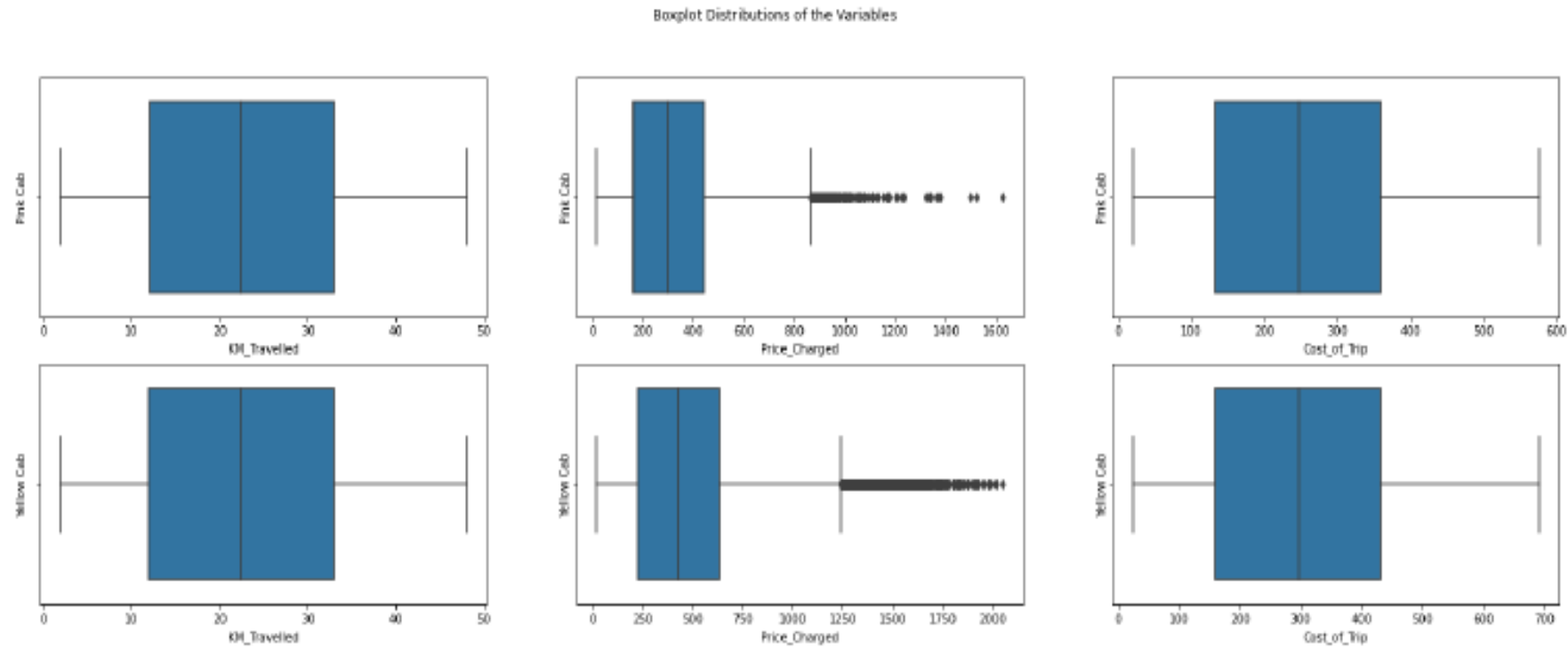
- Yellow cab has the high travel frequency compared to pink cab in the month of December.

Distribution of KM Travelled for both companies



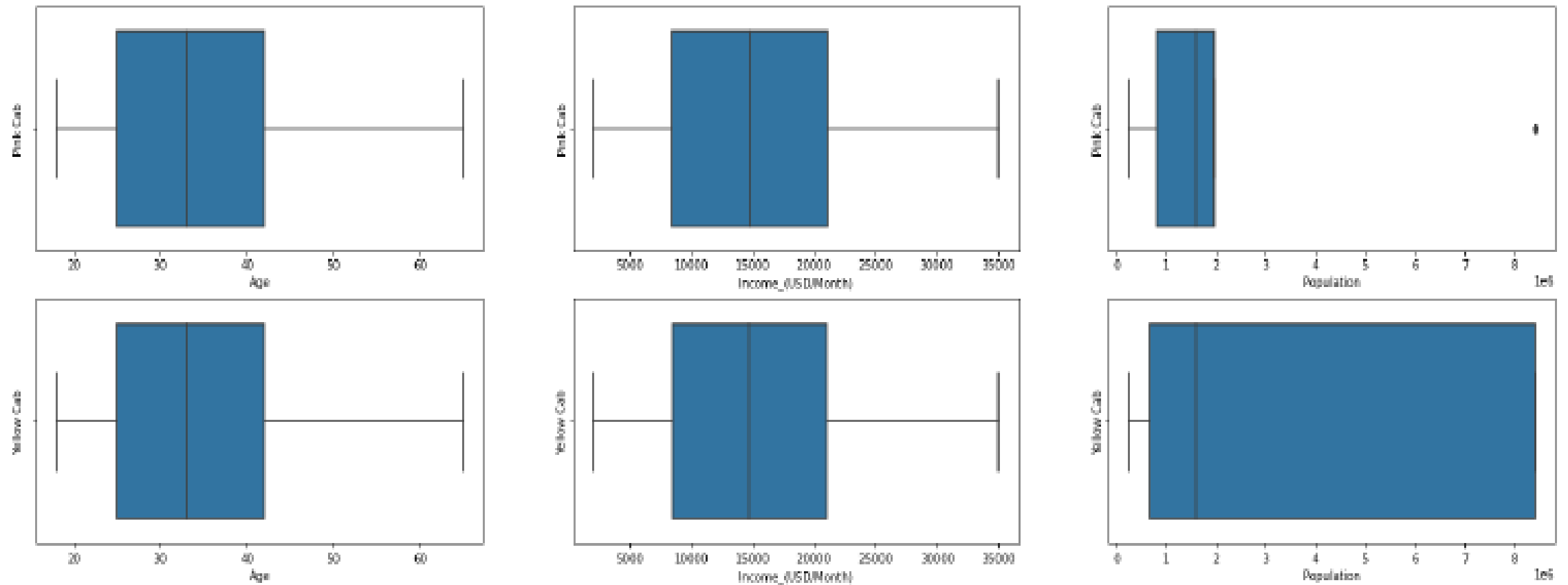
- Both pink and yellow cabs have approximately the same range of rides i.e. 2 to 48 kms

Box Plot Distribution for different variable for two companies

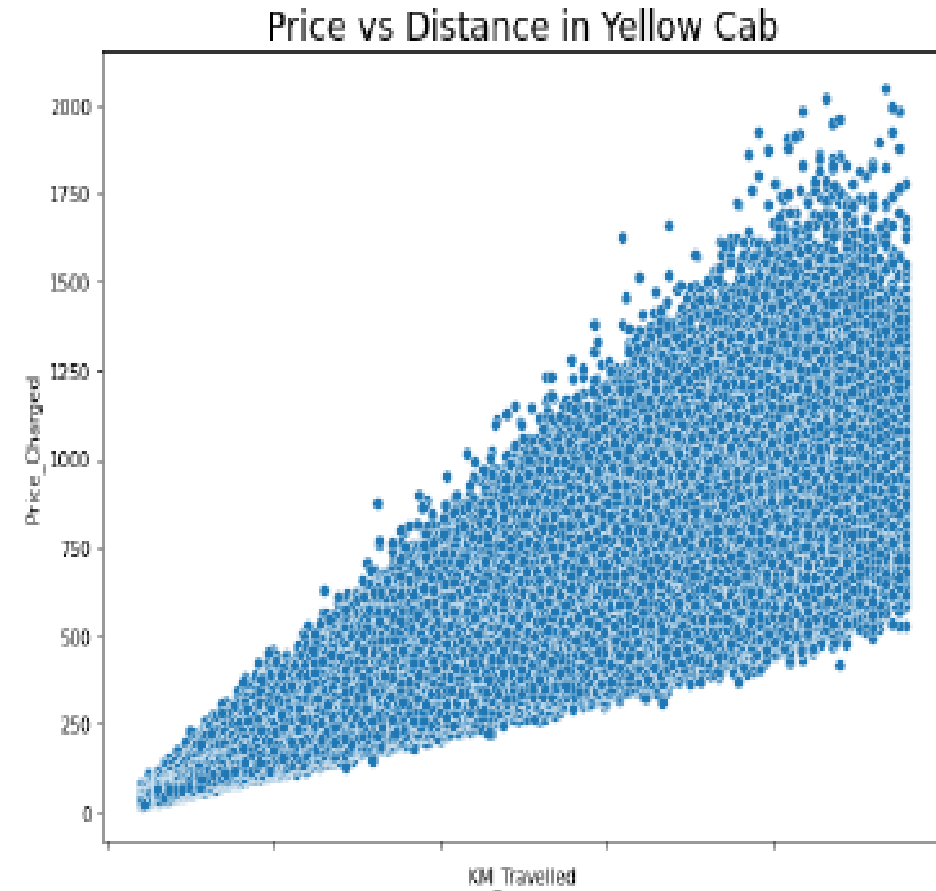
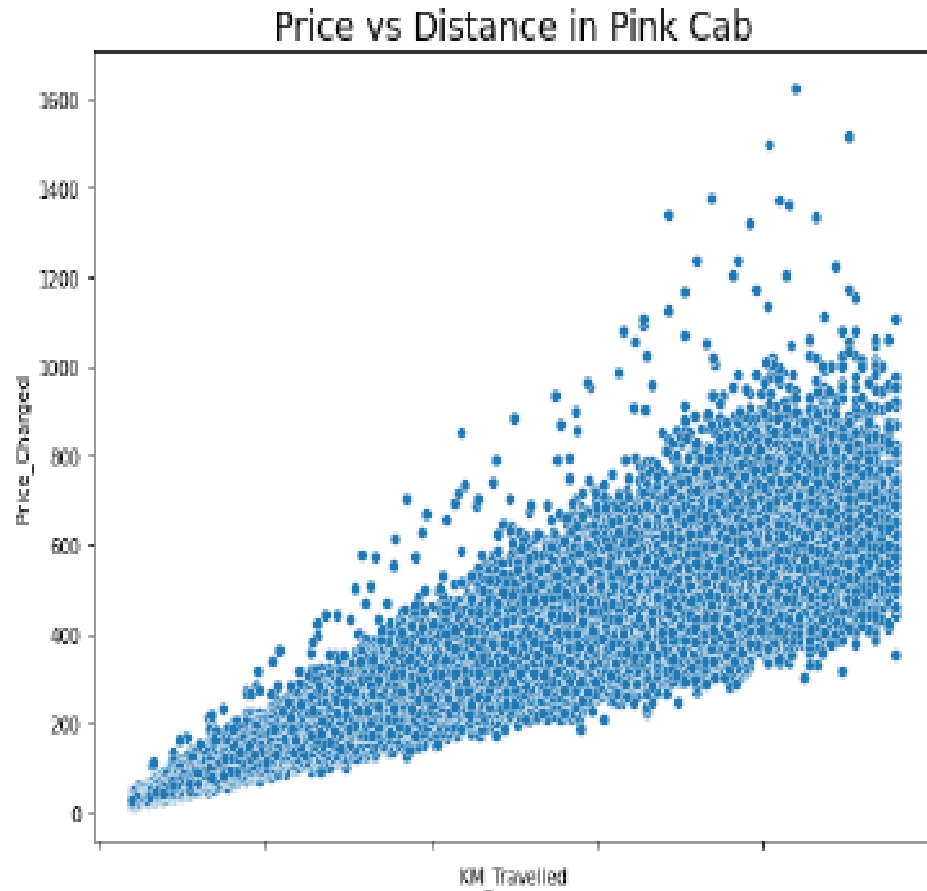


Boxplot Distributions of the Variables

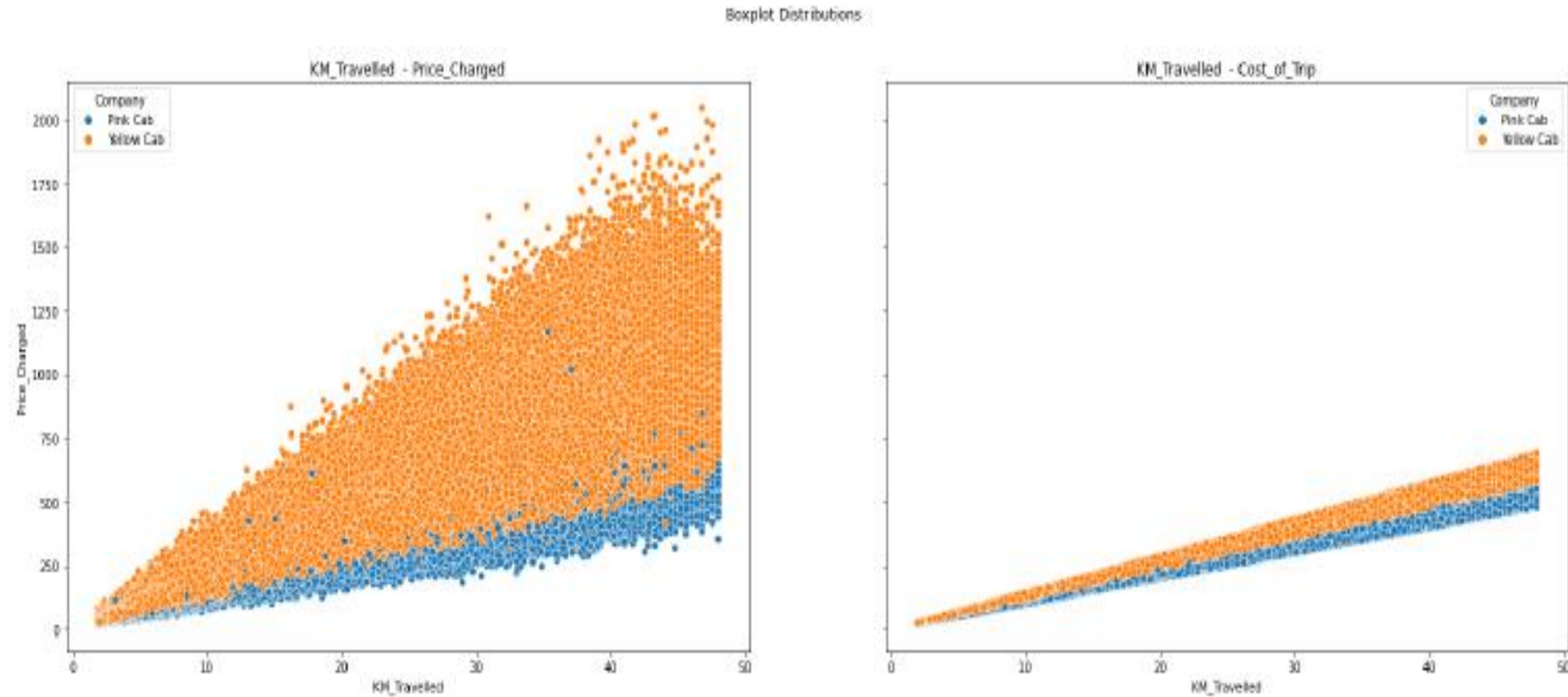
Boxplot Distributions of the Variables



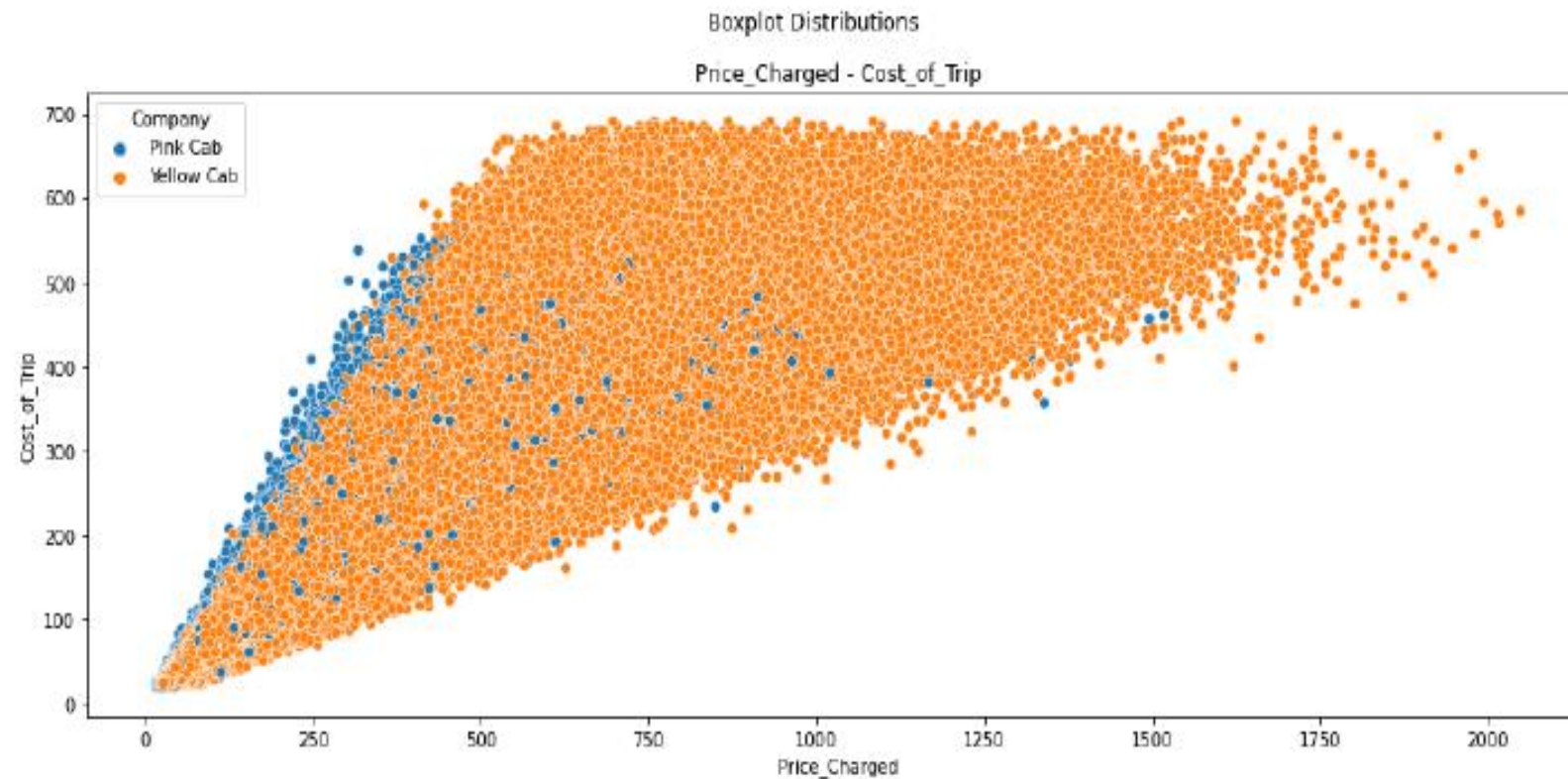
Scatter Plot for price vs distance for two companies



Scatter Plot between KM travelled and price charged ,cost of trip



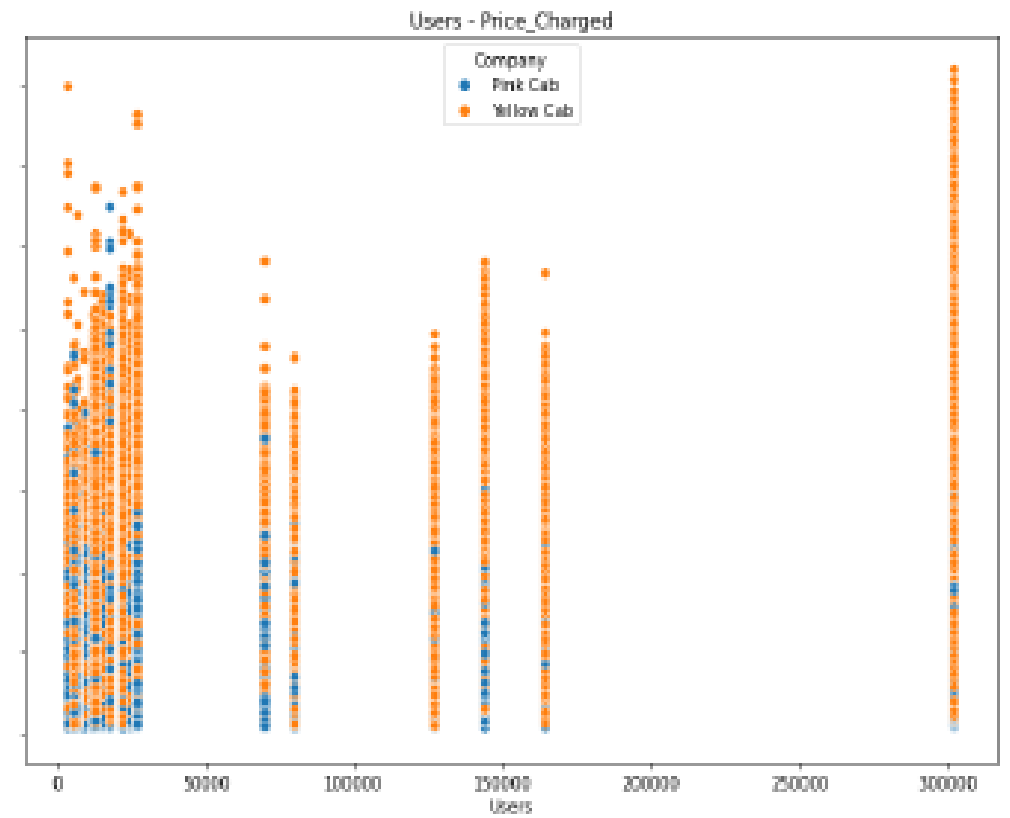
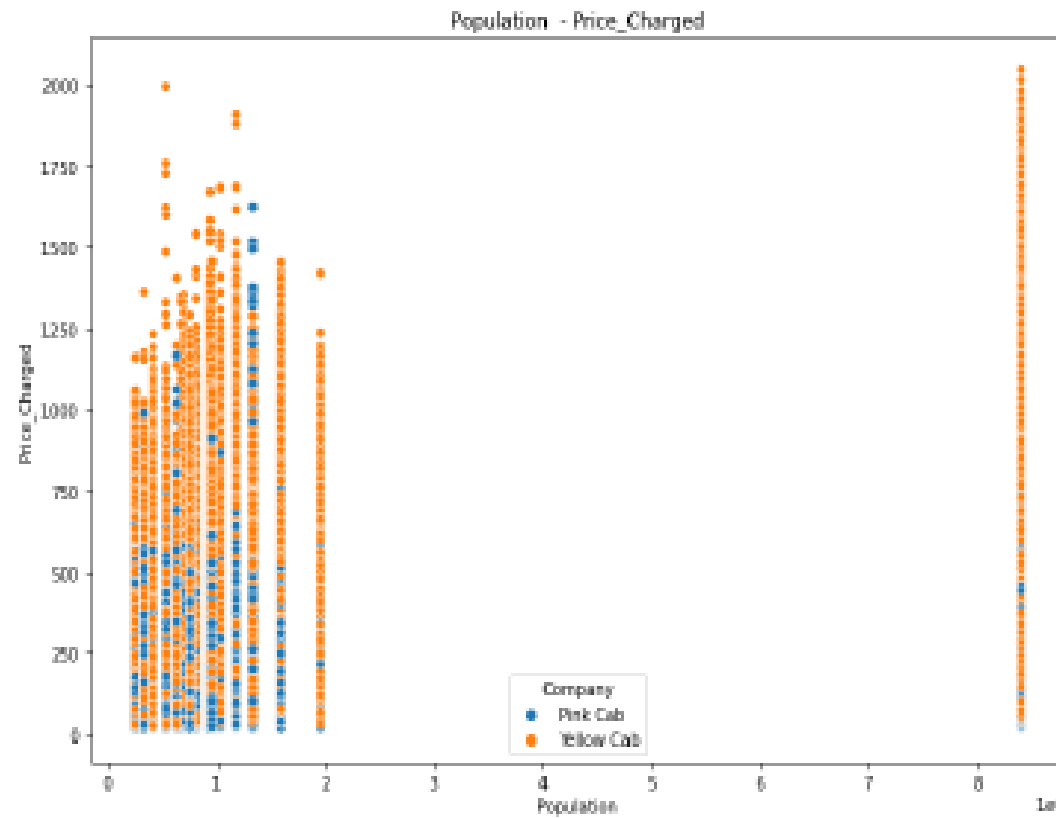
City Wise Cab Users Covered By Company



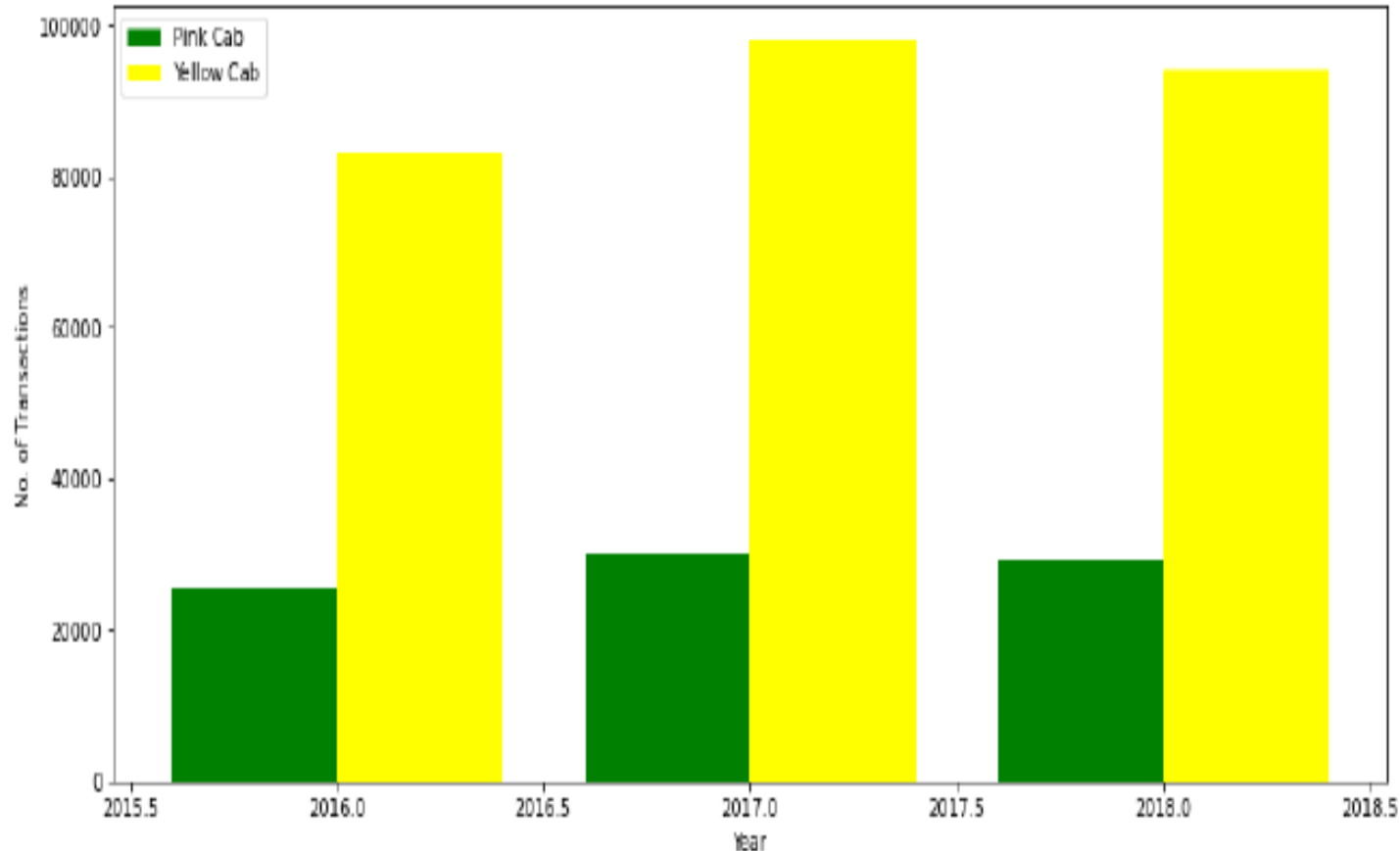
- This is the number of users covered by Yellow and Pink cab In the city against **all cab users** present In the city

Scatter Plot for Pink cab population vs price charged

Pink Cab Firm Boxplot Distributions

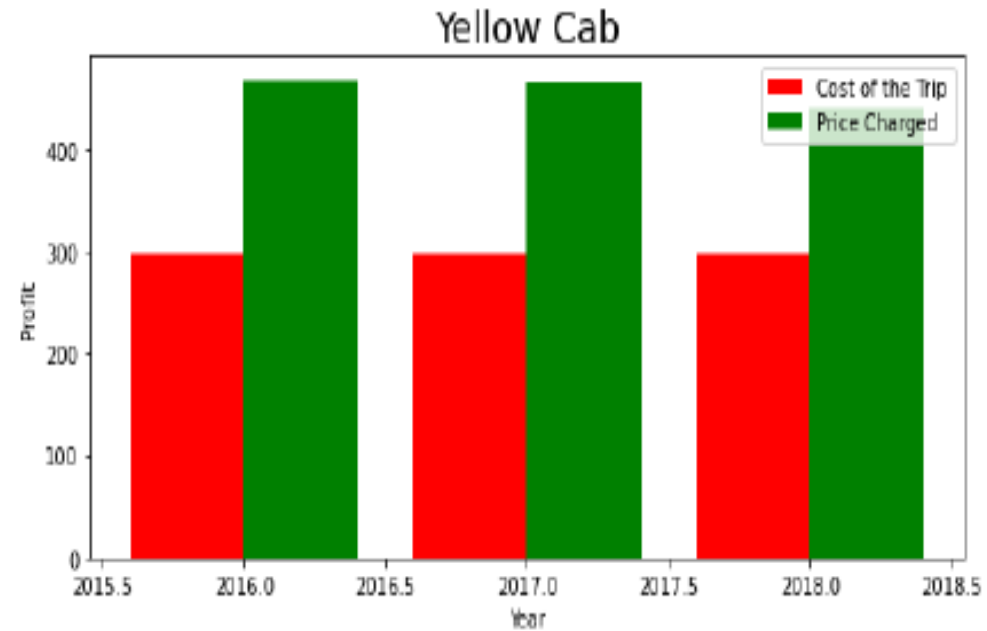
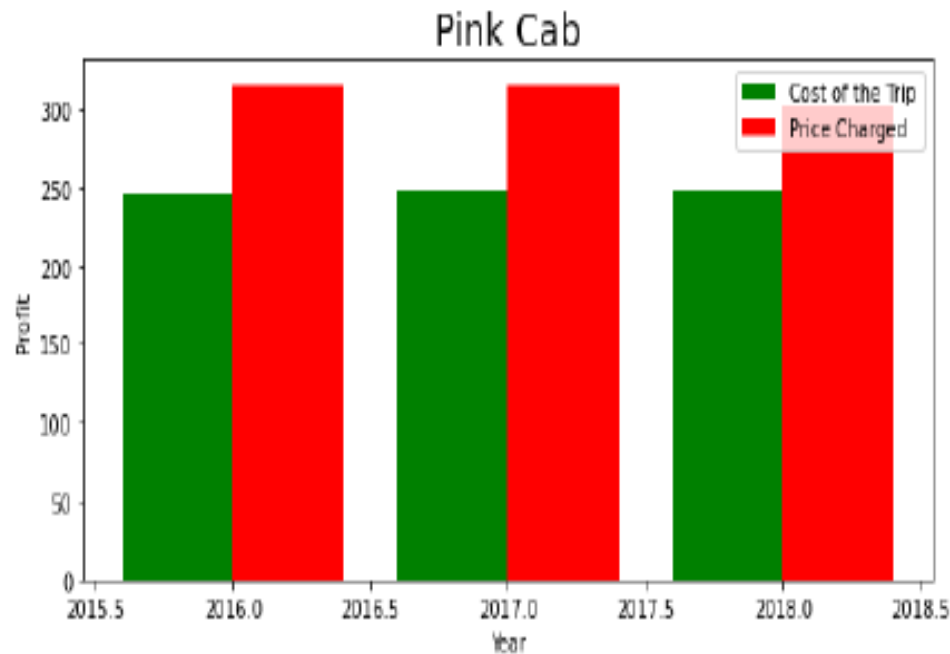


Yearly transaction comparison between two cab companies



➤ Yellow cab transaction is high compared to pink cab on yearly basis.

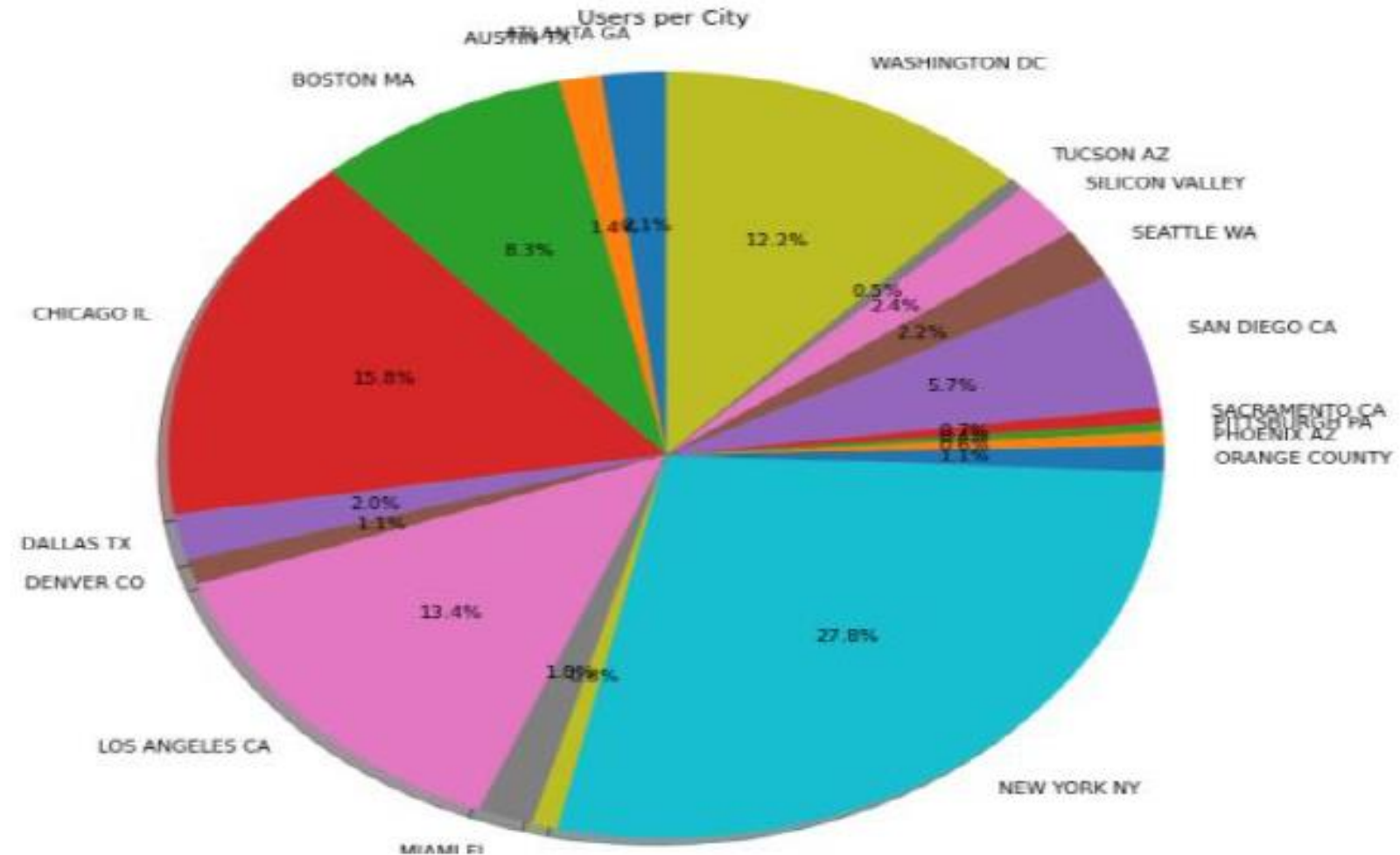
Profit Margin for two cab companies



- Profit Margin = Price charged - cost of trip
- Yellow cab has high profit margin compared to pink cab.

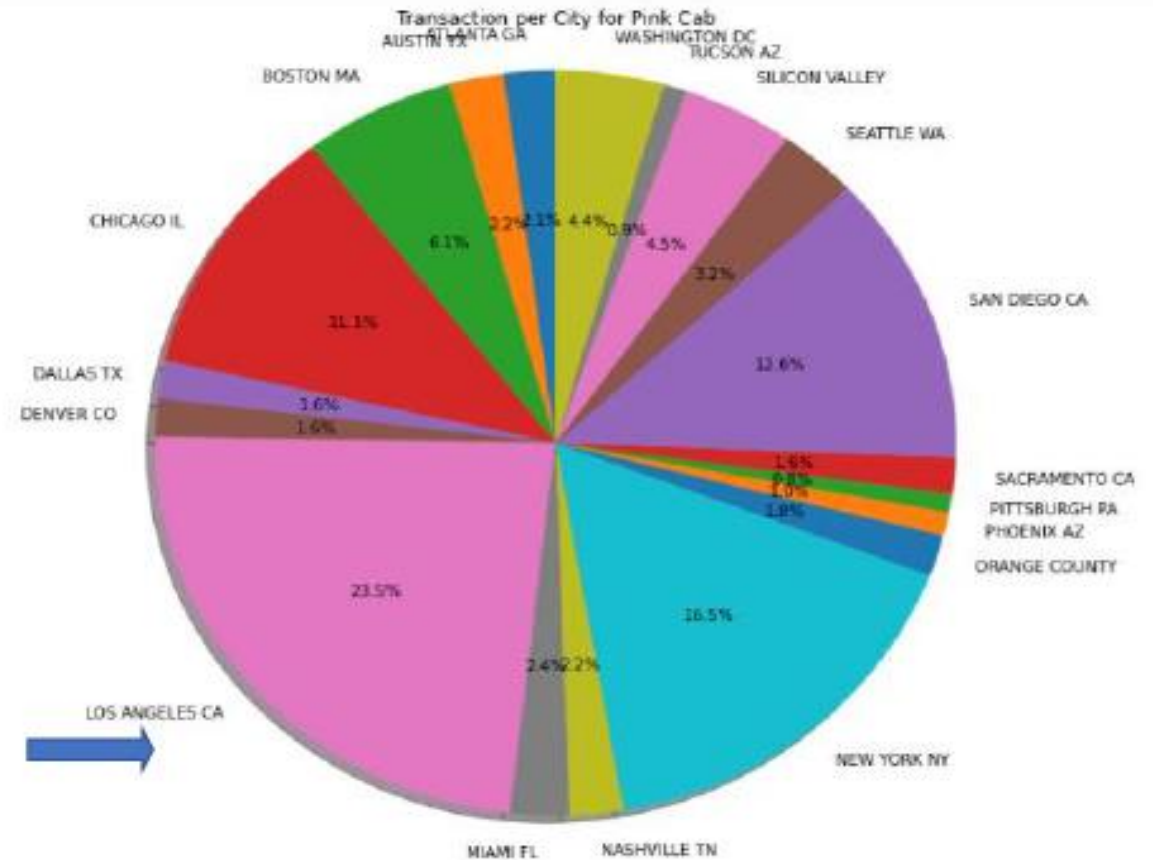
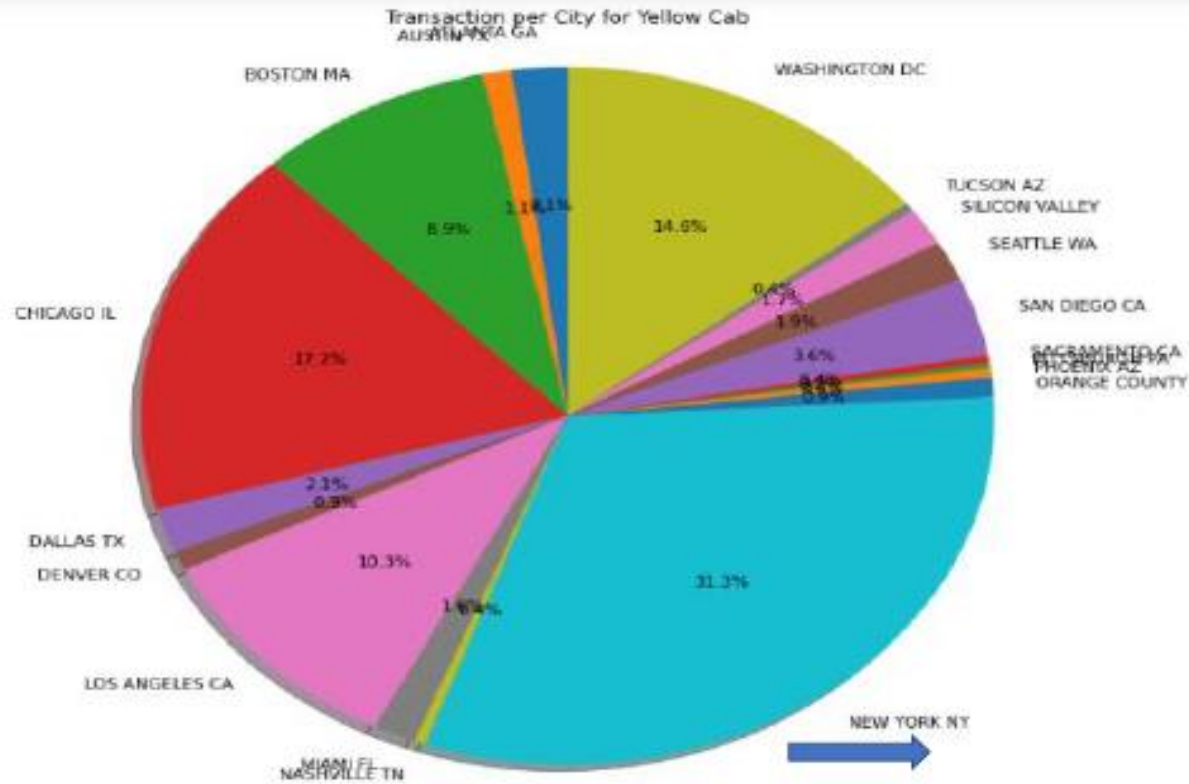
DATA INTERFENETIAL ANALYSIS

Cab user as per city



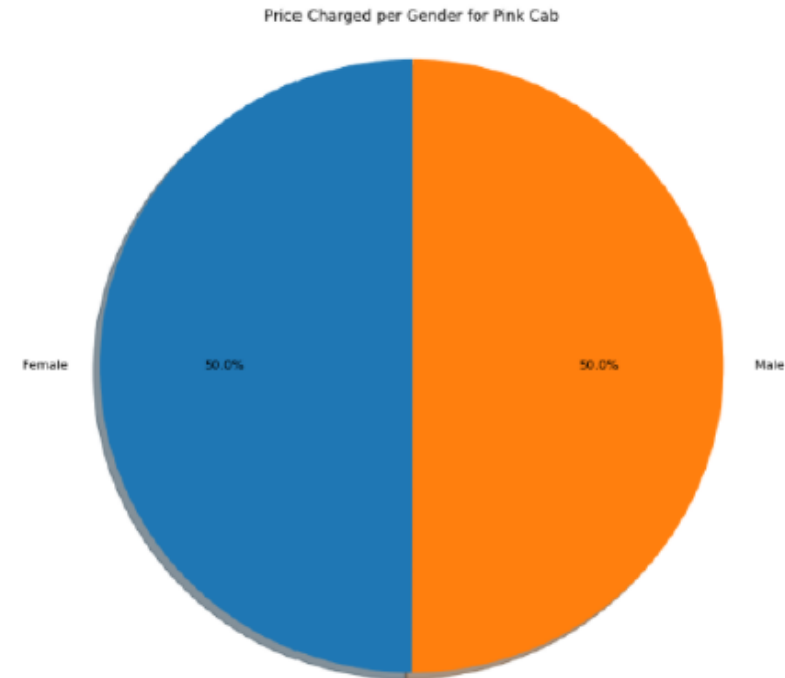
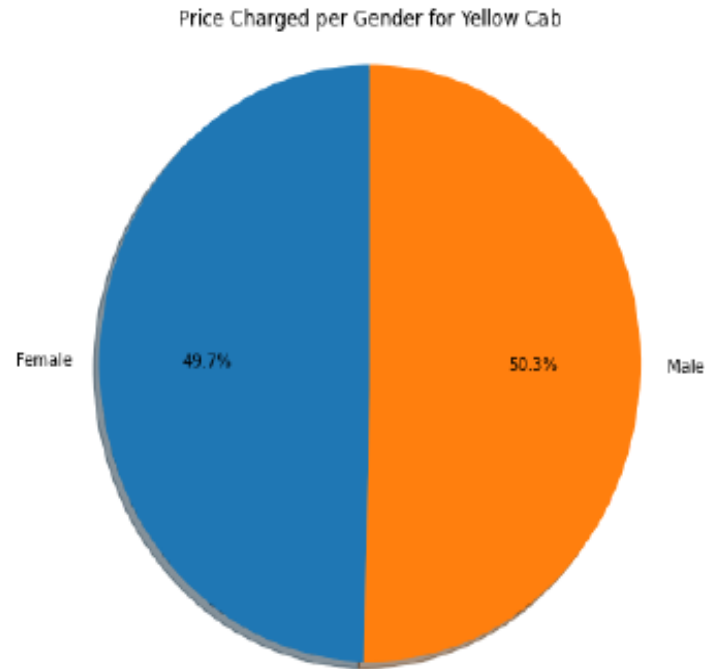
- New York city has the highest cab users with 28% followed by Chicago with 16% and Los Angeles with 13%

Transaction as per city for both companies



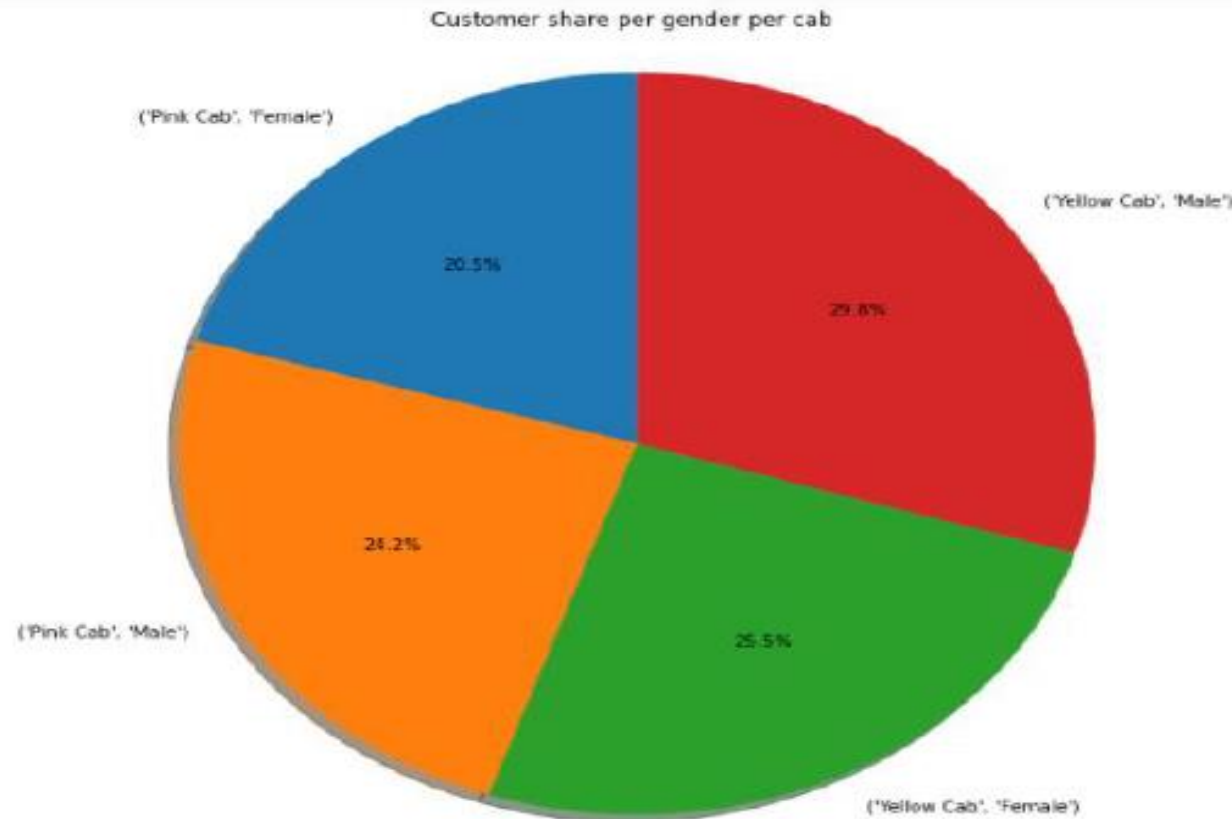
- Transaction for yellow cabs is high with 31% in New York and the highest cab users city is New York with 28%.
- Transaction for pink cab is higher in Los Angeles City with 24%.

Charges of price comparison between two cab companies



- Yellow cab charges less for female customers compared to pink cabs
- Pink cab charges the same for all customers.

Customer share per gender per cab



➤ Female Customer in yellow cabs (25.5%) is higher than to pink cabs (20.5%).

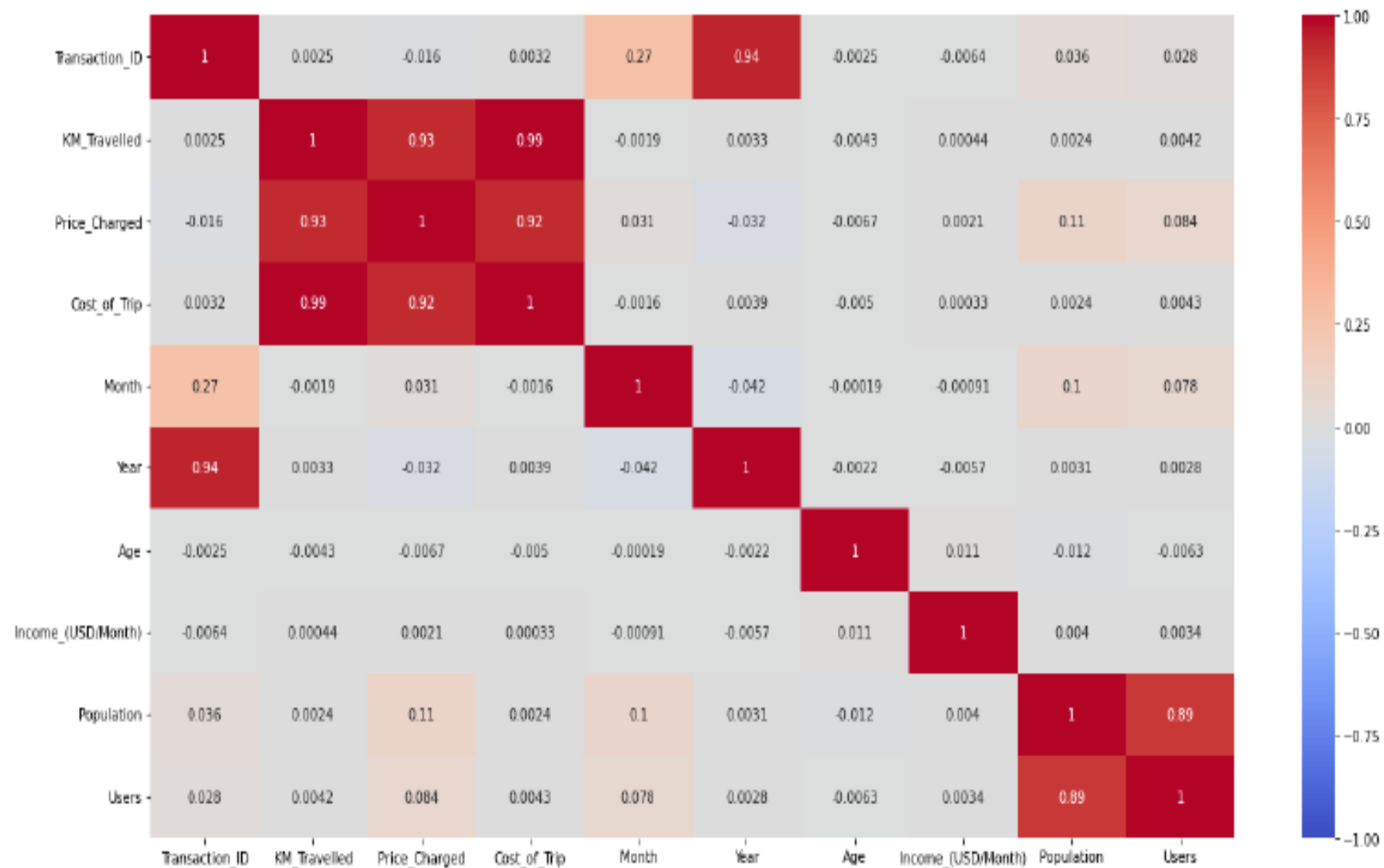
STATISTICAL ANALYSIS FOR THE DATASET

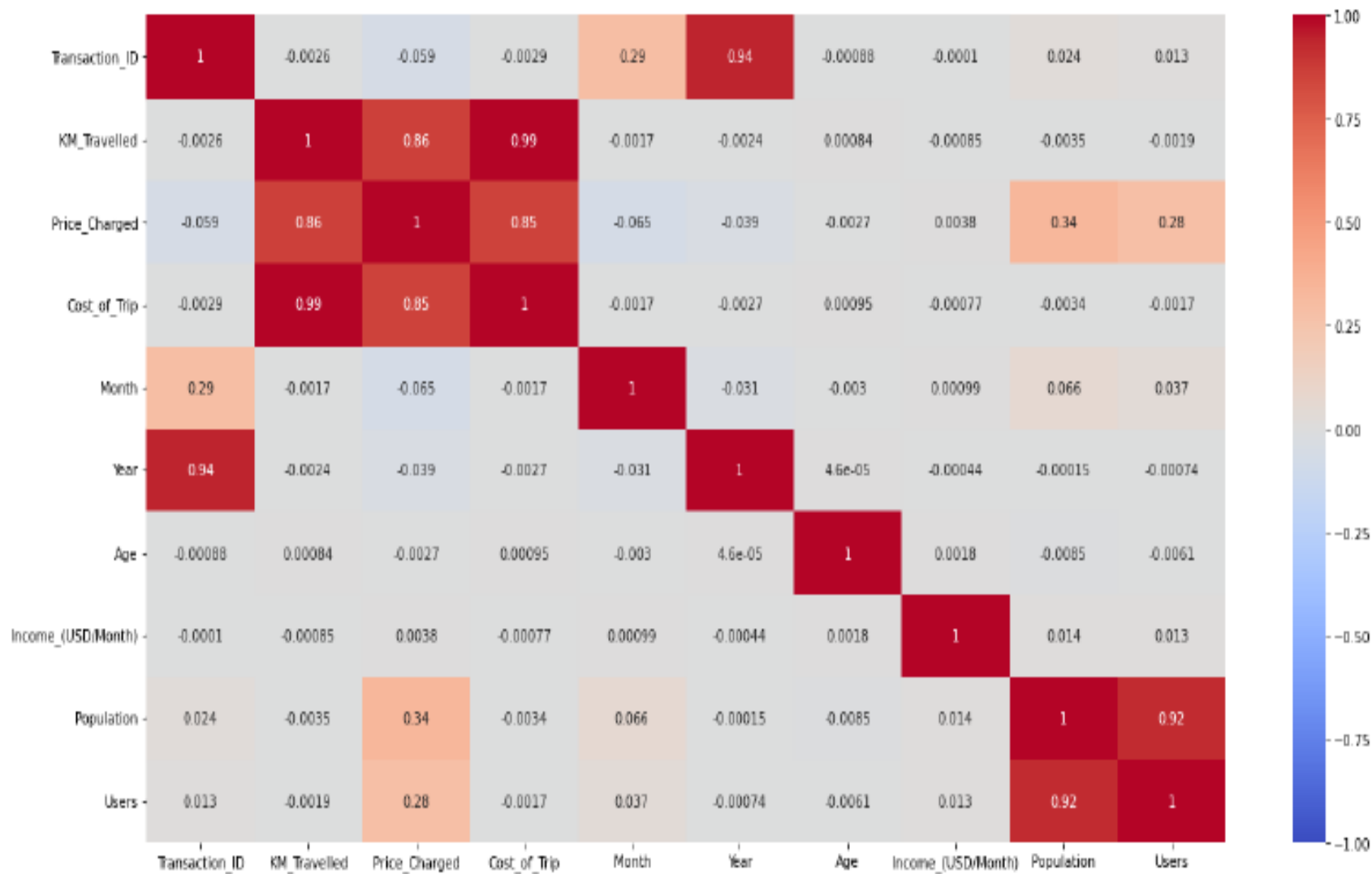
Statistical Info for Pink Cab

		Mean	Median	Median/Mean ratio of	Q1 Value	Q2 Value	IQR value	Outliers
	Transaction_ID	10223938.2	10225904	1.000192	10110143.5	10336416.5	226273	0
	KM_travelled	22.5599	22.44	0.99468	12	32.96	20.96	0
	price charged	310.8	298.06	0.959	159.97	441.505	281.534	237
	cost of trip	248.148	246.33	0.99267	131.868	360.18	228.312	0
	month	7.8161	9	1.15145	5	11	6	0
	year	2017	2017	0.9999	2016	2018	1	0
	Age	35.3224	33	0.93425	25	42	17	0
	income	15059.0417	14713	0.97702	8371	21055	12684	0
	population	2350641.508	1595037	0.67855	814855	1955130	1140245	13967
	users	125590.8133	144132	1.147	27247	164468	137221	0

STATISTICAL INFO FOR YELLOW CAB

	Mean	Median	Median/Mean ratio of	Q1 Value	Q2 Value	IQR value	Outliers
Transaction_ID	102239.22	102259	1	1011.5	1033.5	22627	0
KM_travelled	22.5599	22.44	0.99412	12	32.96	20.96	0
price charged	310.8	298.06	0.959	159.97	441.505	281.53	25866
cost of trip	248.148	246.33	0.9922	131.868	360.18	228.312	0
month	7.8161	9	1.151	5	11	6	0
year	2017	2017	0.99992	2016	2018	2	0
Age	35.322	33	0.934	25	42	17	0
income	15059.0471	14713	0.97702	8371	21055	12684	0
population	2350641.5	1595037	0.67855	814885	1955130	1140245	85918
users	125590.81	144132	1.147272	27247	164468	137221	0





➤ KM_travelled is correlated with price charged followed by cost of trip.

Recommendations

We have evaluated both the cab companies on following points and found Yellow cab better than Pink cab:

- **Customer Reach** : Yellow cab has higher customer reach in 25 cities while Pink cab has higher customer reach in 4 cities. We have also observed that Yellow cab is doing good in covering other cab users as compared to Pink cab.
- **Customer Retention:** We have analyzed this in two segments : at least 5 drive and at least 10 drive with the same cab company. And we found that Yellow cab is doing far better than Pink cab in both these segments.
- **Age wise Reach** : Yellow cab has customer in all age group and it's been observed that it's even popular in 60+ age group as equally as its in 18-25 age group.
- **Average Profit per KM:** Yellow cab's average profit per KM is almost three times the average profit per KM of the Pink cab.
- **Income wise Reach** :Both the cabs are very popular in high and medium income class but here also Yellow cab is performing better than Pink cab in offering their services to all the three income class group (low, medium and high)
- **Ride count and Profit Forecasting** : Both the companies are facing loss in the profit and no. of ride. Yellow cab's forecasted profit loss is around 1.83% while Pink cab's loss in 3.1%.Pink cab is facing more loss even when its forecasted no of ride loss is lesser than Yellow cab.

On the basis of above point , we will recommend Yellow cab for investment.

Thank You



Data Glacier

Your Deep Learning Partner