



Data Glacier

Your Deep Learning Partner

G2M Case Study

Virtual Internship (LISUM02)

Aynur Cemre Aka

06-Aug-2021

Background

- 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.
- Aim of this project is to identify the right company to make the investment.
- There are 2 cab companies:
 1. Yellow Cab
 2. Pink Cab
- The analysis include:
 1. Data understanding
 2. Creating multiple hypothesis
 3. Visualization and checking hypotheses
 4. Finding the most suitable company for investment

Datasets

There are 4 datasets:

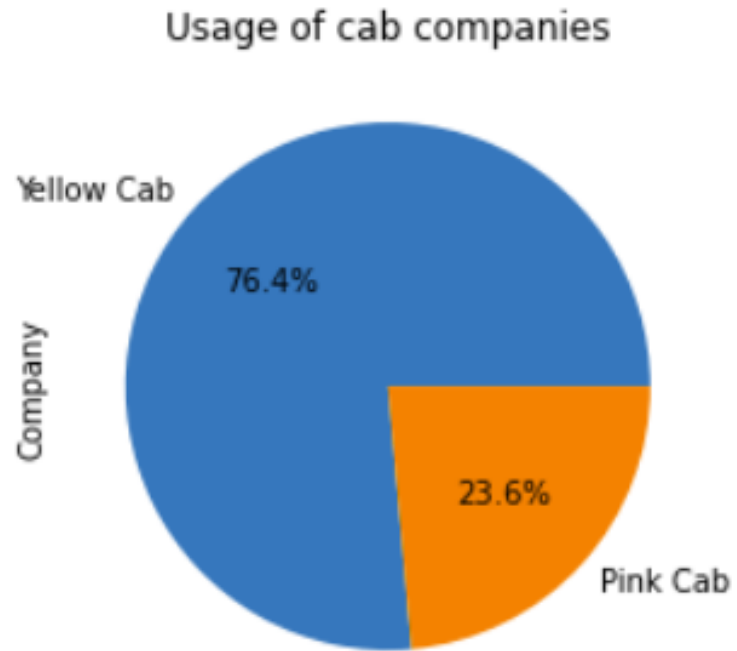
- Cab_Data.csv: includes information about Pink and Yellow Cab companies.
- City.csv: includes information about US cities (name of the city, population, cab users).
- Customer_ID.csv: includes information about customer profiles (ID of customer, gender, income, age).
- Transaction_ID.csv: includes customer's transaction information (ID of transaction, ID of customer, payment mode).

Changes:

- Cab_Data.csv:
 1. "Date of Travel" is changed from serial time to date time.
 2. Attribute names are changed.
 3. Checked for missing and unnecessary values and saved as Cab_Data.pkl.
- City.csv: checked for missing and unnecessary values and saved as City.pkl.
- Customer_ID.csv: checked for missing and unnecessary values and saved as Customer_ID.pkl.
- Transaction_ID.csv:
 1. Attribute names are changed.
 2. Checked for missing and unnecessary values and saved as Transaction_ID.pkl.

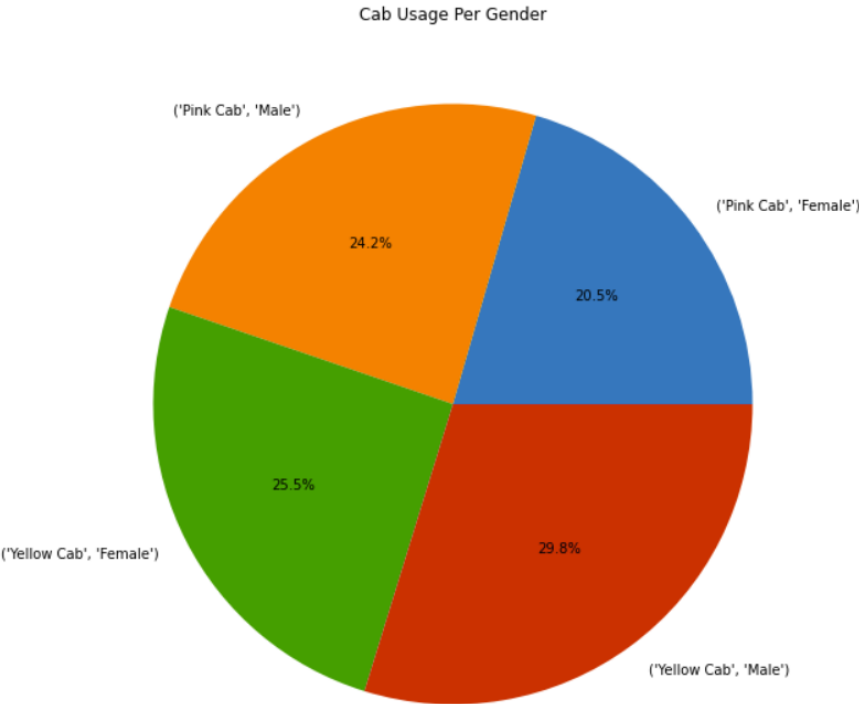
Exploratory Data Analysis (EDA)

Total Usage of Cab Companies



- Yellow Cab company is more preferred than pink cab company.

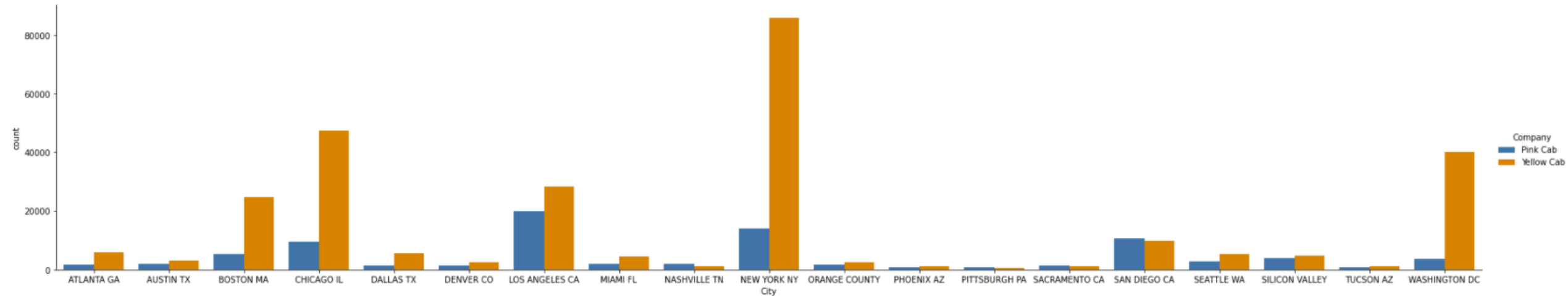
Cab Usage by Gender



Company	Gender	Payment_Mode		
Pink Cab	Female	Card	35.311169	15070.483064
		Cash	35.209879	15081.188938
	Male	Card	35.408755	15072.911708
		Cash	35.296201	15007.239208
Yellow Cab	Female	Card	35.251196	15070.910917
		Cash	35.355999	15024.989852
	Male	Card	35.409862	15033.178743
		Cash	35.325693	15051.852449

Company	Gender	Age	Payment_Mode	Income
Pink Cab	Female	35.270518	2	15074.779696
	Male	35.363596	2	15046.562618
Yellow Cab	Female	35.293129	2	15052.537345
	Male	35.376189	2	15040.649473

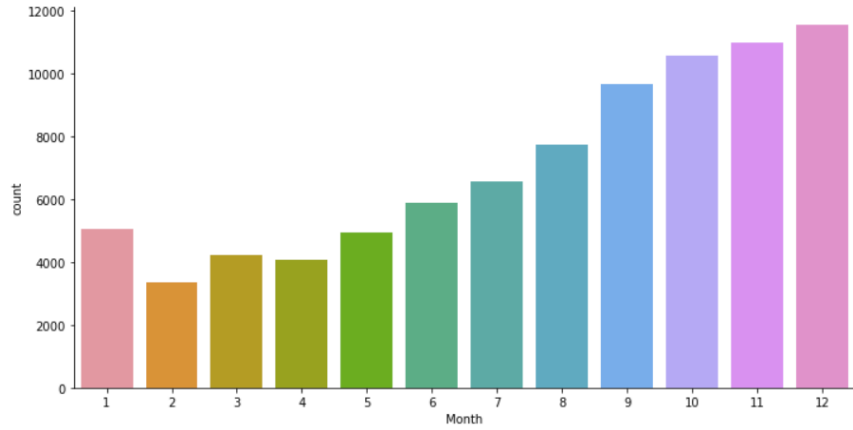
Usage of Cab Companies per City



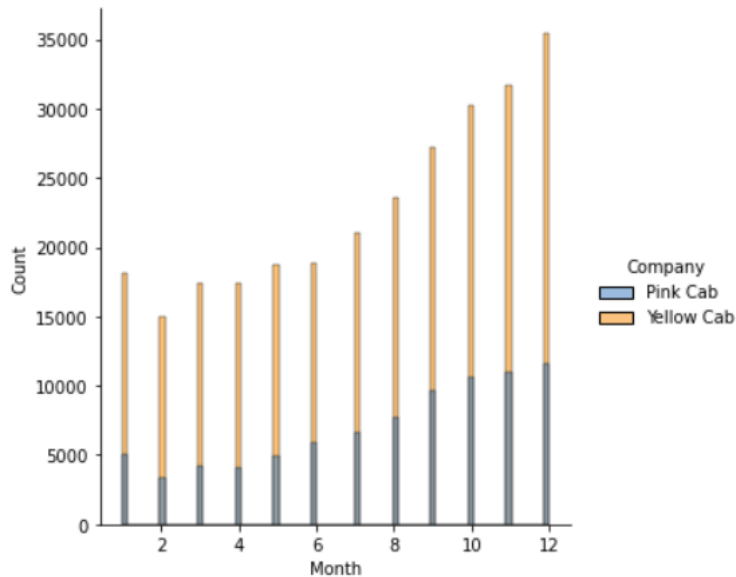
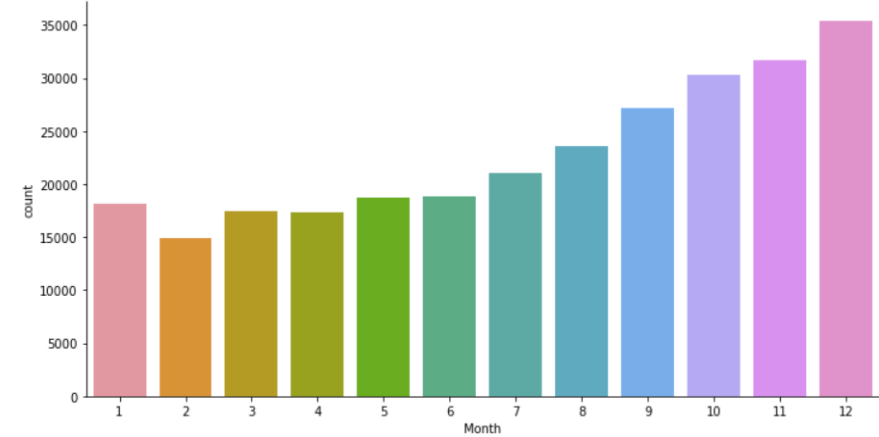
- Other than Nashville, Pittsburg, Sacramento and San Diego, Yellow Cab company is more preferred.

Monthly Usages of Cab Companies

Pink Cab:



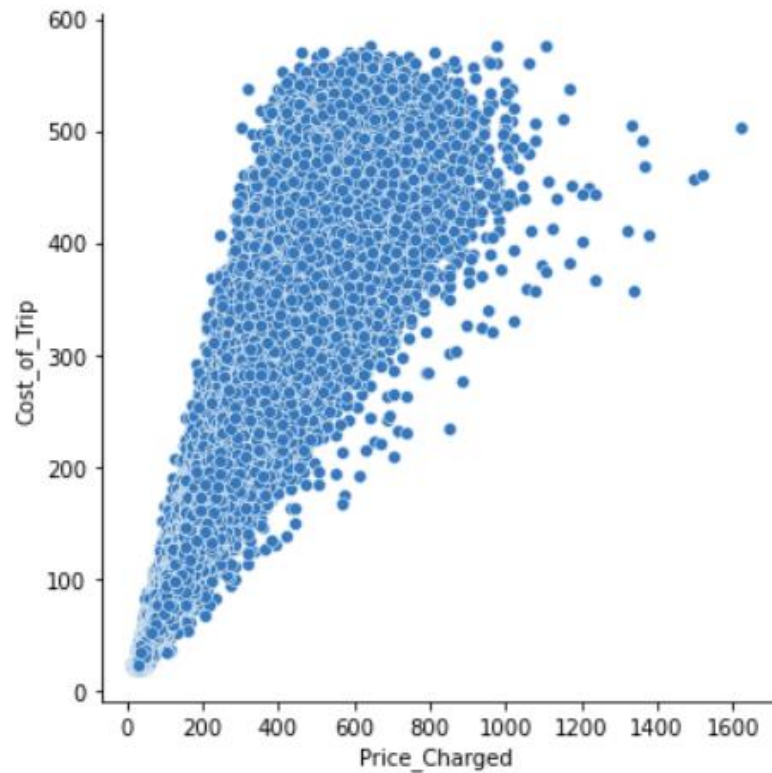
Yellow Cab:



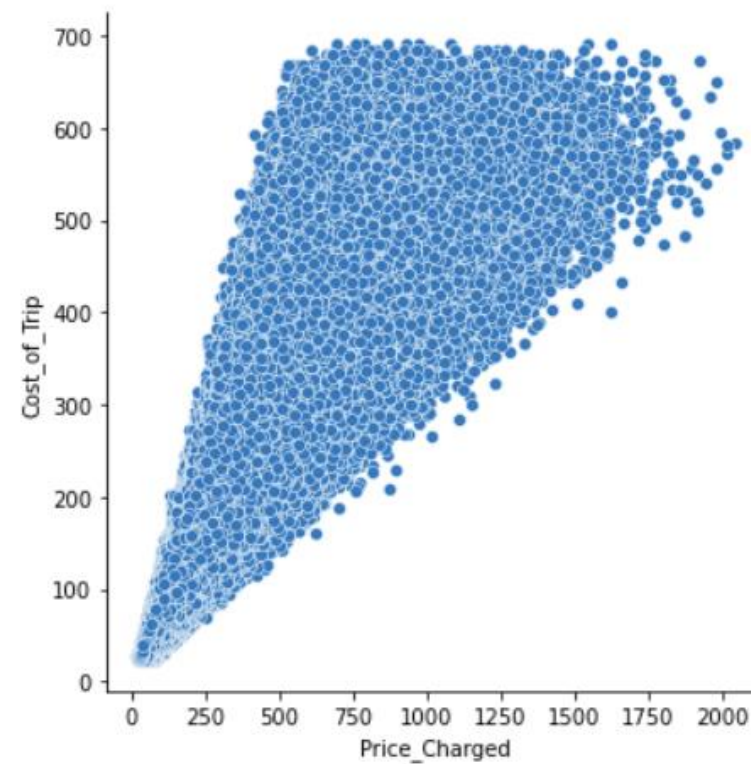
- For every month, Yellow Cab company has more travels than pink cab company.
- The changing trends by months are almost the same.

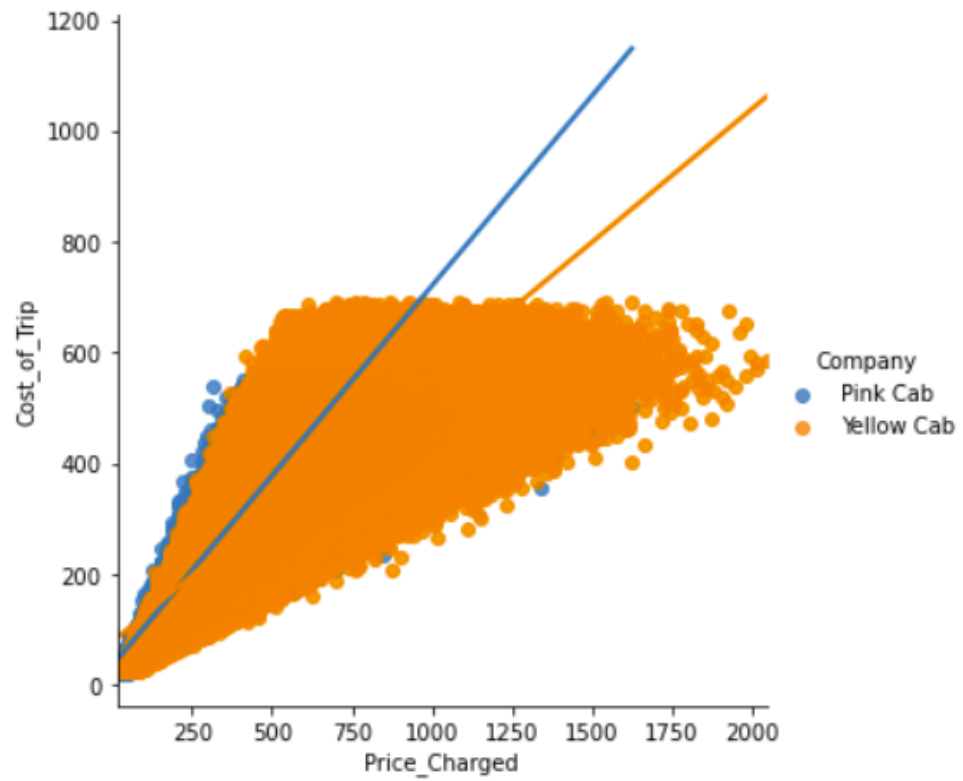
Distribution of Price Charged vs Cost of Trip

Pink Cab:



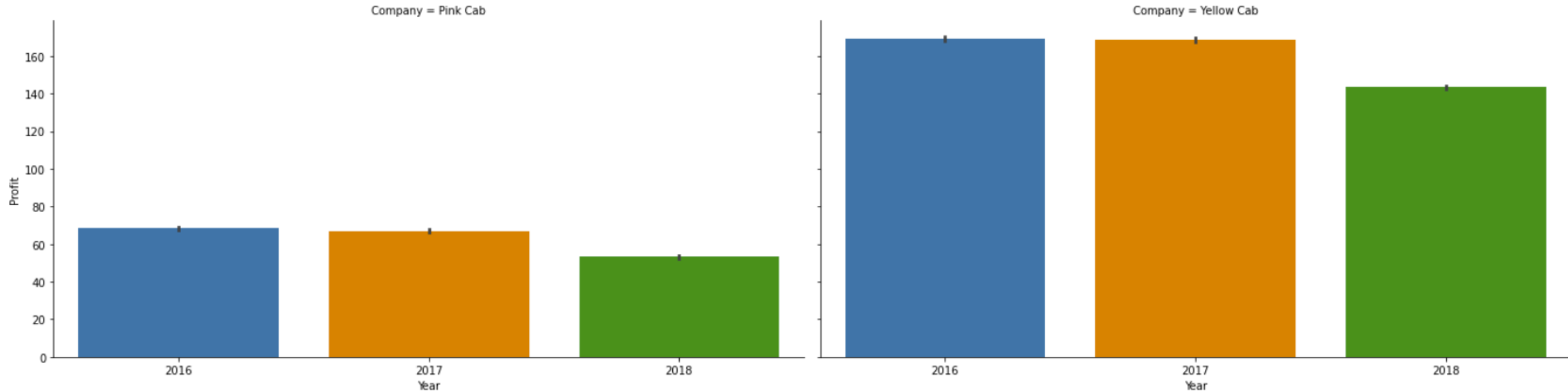
Yellow Cab:





- Both companies have positive revenue.
- Yellow Cab company earns more.

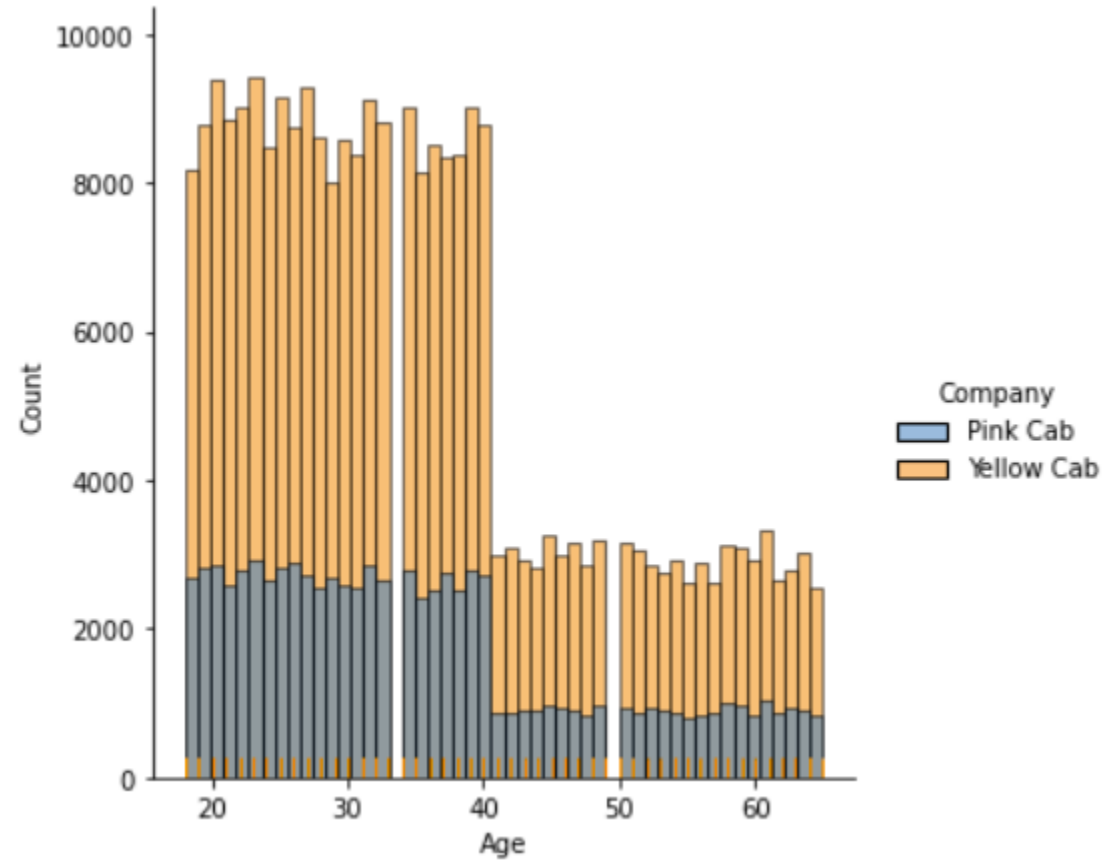
Profit Analysis



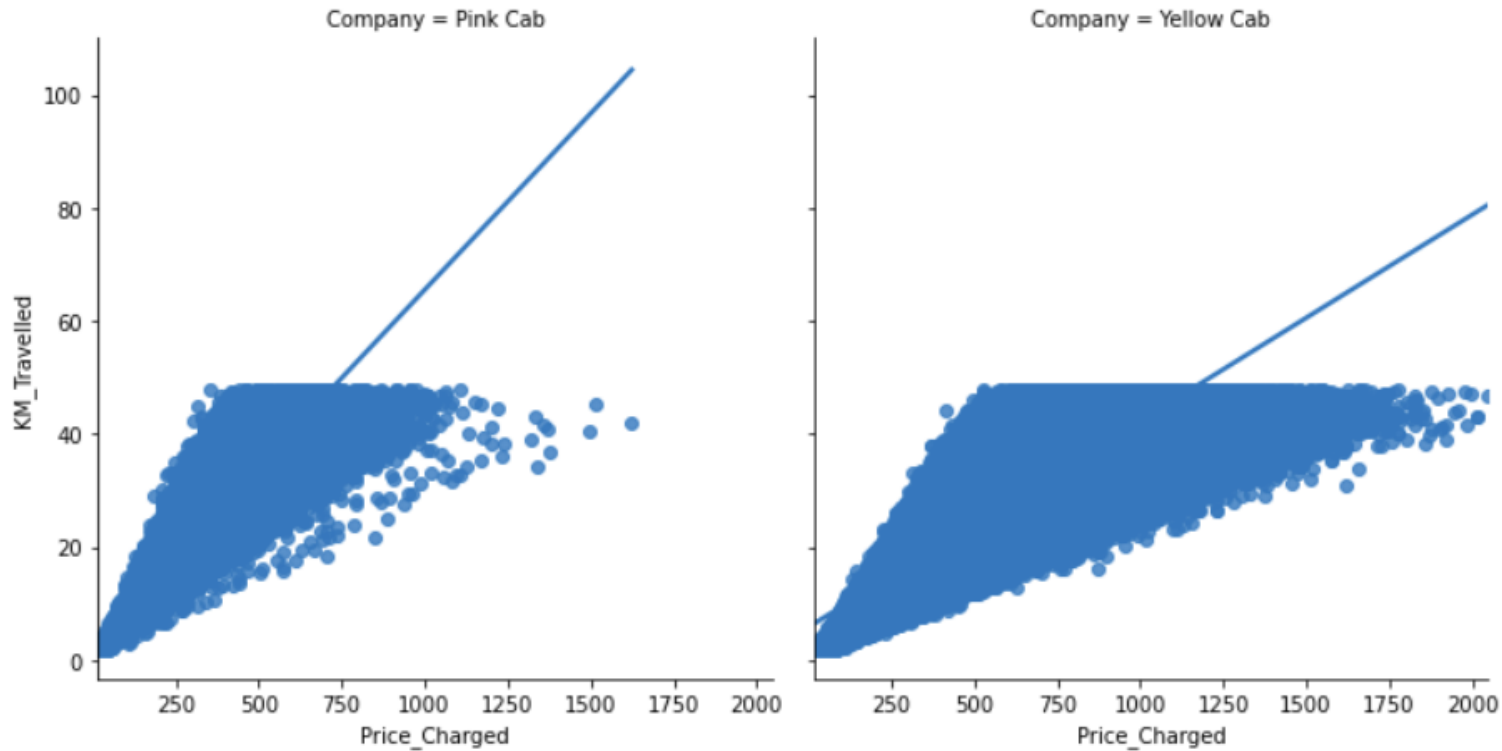
Company Name	Sum of Profits	Profit per Ride
Yellow Cab	\$44020373.17	\$160.26
Pink Cab	\$5307328.32	\$62.65

- Sum of profit and profit per ride values of Yellow Cab company is higher than Pink Cab company.
- Every year, profit of Yellow Cab is higher than Pink Cab company.

Distribution of Age by Company



Distribution of KM Travelled vs Price Charged



Information About Cab Companies

Pink Cab:

	count	mean	std	min	25%	50%	75%	max
Price_Charged	84711.0	310.800856	181.995661	15.6	159.970	298.06	441.505	1623.48
Cost_of_Trip	84711.0	248.148682	135.403345	19.0	131.868	246.33	360.180	576.00
KM_Travelled	84711.0	22.559917	12.231092	1.9	12.000	22.44	32.960	48.00

Yellow Cab:

	count	mean	std	min	25%	50%	75%	max
Price_Charged	274681.0	458.181990	288.386166	20.73	226.68	425.060	633.880	2048.03
Cost_of_Trip	274681.0	297.922004	162.548986	22.80	158.40	295.596	432.432	691.20
KM_Travelled	274681.0	22.569517	12.234298	1.90	11.99	22.440	32.960	48.00

Recommendations

From the evaluations of following points, it is found that Yellow Cab is a better choice :

- Usage of cab: 76.4% of customers prefer Yellow Cab. When evaluated according to cities, it is seen that Yellow Cab is preferred more in 15 out of 19 cities. It is preferred almost 5 times in New York, where taxis are used the most.
- Price Charged vs Cost of the Trip: Although Yellow Cab earns more and charges more, it's more preferred.
- Profit: The total earnings of Yellow Cab are more than Pink Cab, and the earnings from each drive are more than Pink Cab.
- KM Travelled vs Price Charged: When KM travelled and price charged is compared, it is seen that Yellow Cab charges more and from distribution of the graph, it is seen that yellow cab travels longer distances.
- Age of Customer: The age distribution and the usage rates of the companies are directly proportional. Considering this information, Yellow Cab is preferred more in age ranges since the usage rate is higher.

On the basis above, we recommend the Yellow Cab for investment.

Thank You