G2M Case Study

Virtual Internship

30-May-2022

Background –G2M(cab industry) case study

- XYZ is a private equity firm based in the United States. It is going to invest in the Taxi Industry due to the spectacular expansion in the last few years and the presence of many significant competitors in the sector.

- Objective : Provide actionable insights to assist the XYZ firm in picking the best company to invest in.

The analysis has been divided into four parts:

1. Hypothesis or Business Questions
2. Description of the Datasets
3. Exploratory Data Analysis
4. Results of the Business Questions/Hypothesis
5. Recommendations for investment

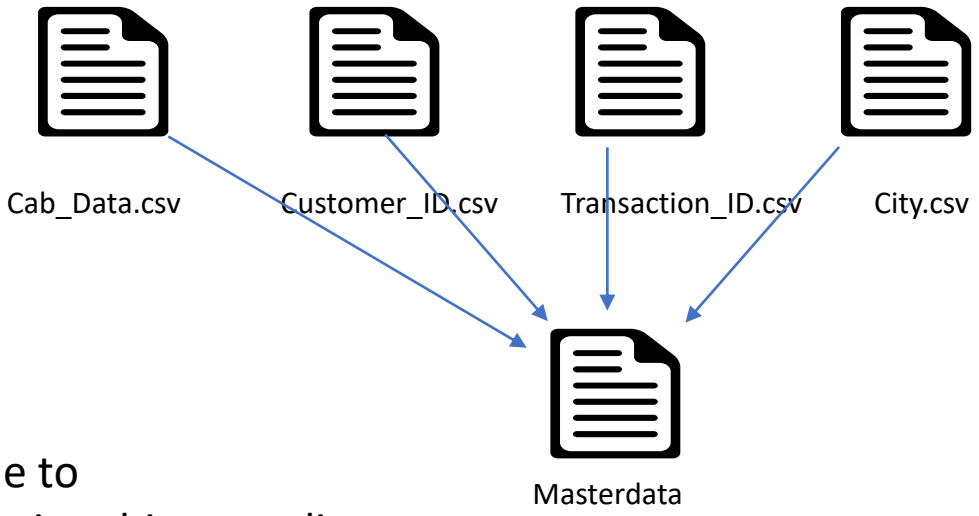
HYPOTHESIS OR BUSINESS QUESTIONS

BUSINESS QUESTIONS

- **What is the market size of the Cab company between 2016 to 2018.**
- **What is the profit rate of the cab company between 2016 and 2018**
- **What is the Market share of the Cab Company Citywise?**
- **What are the differences in demand based on age?**

Description of the Datasets

- 20 Features(including 3 derived features)
- Timeframe of the data: 2016-01-31 to 2018-12-31
- Total data points :355,032



Assumptions:

- Outliers are present in Price_Charged feature but due to unavailability of trip duration details ,we are not treating this as outlier.

- Profit of rides are calculated keeping other factors constant and only Price_Charged and Cost_of_Trip features used to calculate profit.

- Users feature of city dataset is treated as number of cab users in the city. we have assumed that this can be other cab users as well(including Yellow and Pink cab)

Description of the Datasets

Company	City	KM Travelled	Price Charged	Cost of Trip	Year	Month	Day	Customer ID	Payment_Mode	Gender	Age	Income (USD/Month)	Population	Users	Profit
Yellow Cab	WASHINGTON DC	10.20	202.54	135.8640	2016	1	1	52536	Card	Male	24	10606.0	418,859	127,001	66.6760
Yellow Cab	WASHINGTON DC	5.94	88.50	74.1312	2016	3	25	52536	Cash	Male	24	10606.0	418,859	127,001	14.3688
Pink Cab	WASHINGTON DC	15.21	179.33	153.6210	2016	11	8	52536	Cash	Male	24	10606.0	418,859	127,001	25.7090
Yellow Cab	WASHINGTON DC	8.08	123.21	104.7168	2017	4	16	52536	Card	Male	24	10606.0	418,859	127,001	18.4932
Yellow Cab	WASHINGTON DC	13.80	283.74	168.9120	2017	6	23	52536	Cash	Male	24	10606.0	418,859	127,001	114.8280
Pink Cab	WASHINGTON DC	28.00	424.20	280.0000	2017	10	30	52536	Card	Male	24	10606.0	418,859	127,001	144.2000

359392

Description of the Datasets

```
Transaction ID                int64
Date of Travel       datetime64[ns]
Company                      object
City                         object
KM Travelled                float64
Price Charged               float64
Cost of Trip                float64
Year of Travel                int64
Month of Travel               int64
Day of Travel                 int64
Customer ID                   int64
Payment_Mode                 object
Gender                       object
Age                           int64
Income (USD/Month)          float64
Population                    int32
Users                         int32
dtype: object
```

```
# Let's check if we have missing values
masterdata.isnull().sum()
```

```
Transaction ID         0
Date of Travel         0
Company                0
City                   0
KM Travelled           0
Price Charged          0
Cost of Trip           0
Year of Travel         0
Month of Travel        0
Day of Travel          0
Customer ID            0
Payment_Mode           0
Gender                 0
Age                    0
Income (USD/Month)     0
Population             0
Users                  0
dtype: int64
```

EDA(Statistical Analysis)

- **Multivariate Correlation**: We try to find out if there is a linear relationship between the variables in the masterdata since we have more than one variable to relate with.

	Transaction ID	KM Travelled	Price Charged	Cost of Trip	Year	Month	Day	Year of Travel	Month of Travel	Day of Travel	Customer ID	Age	Income (USD/Month)
Transaction ID	1.000000	-0.001429	-0.052902	-0.003462	0.941475	0.284724	0.014824	0.941475	0.284724	0.014824	-0.016912	-0.001267	-0.001570
KM Travelled	-0.001429	1.000000	0.835753	0.981848	-0.001094	-0.001773	0.000767	-0.001094	-0.001773	0.000767	0.000389	-0.000369	-0.000544
Price Charged	-0.052902	0.835753	1.000000	0.859812	-0.036903	-0.059639	-0.006820	-0.036903	-0.059639	-0.006820	-0.177324	-0.003084	0.003228
Cost of Trip	-0.003462	0.981848	0.859812	1.000000	-0.001766	-0.008309	0.000726	-0.001766	-0.008309	0.000726	0.003077	-0.000189	-0.000633
Year	0.941475	-0.001094	-0.036903	-0.001766	1.000000	-0.033169	-0.022449	1.000000	-0.033169	-0.022449	-0.002480	-0.000497	-0.001679
Month	0.284724	-0.001773	-0.059639	-0.008309	-0.033169	1.000000	0.055164	-0.033169	1.000000	0.055164	-0.045030	-0.002376	0.000585
Day	0.014824	0.000767	-0.006820	0.000726	-0.022449	0.055164	1.000000	-0.022449	0.055164	1.000000	0.000763	-0.002291	0.000927
Year of Travel	0.941475	-0.001094	-0.036903	-0.001766	1.000000	-0.033169	-0.022449	1.000000	-0.033169	-0.022449	-0.002480	-0.000497	-0.001679
Month of Travel	0.284724	-0.001773	-0.059639	-0.008309	-0.033169	1.000000	0.055164	-0.033169	1.000000	0.055164	-0.045030	-0.002376	0.000585
Day of Travel	0.014824	0.000767	-0.006820	0.000726	-0.022449	0.055164	1.000000	-0.022449	0.055164	1.000000	0.000763	-0.002291	0.000927

EDA(Statistical Analysis)

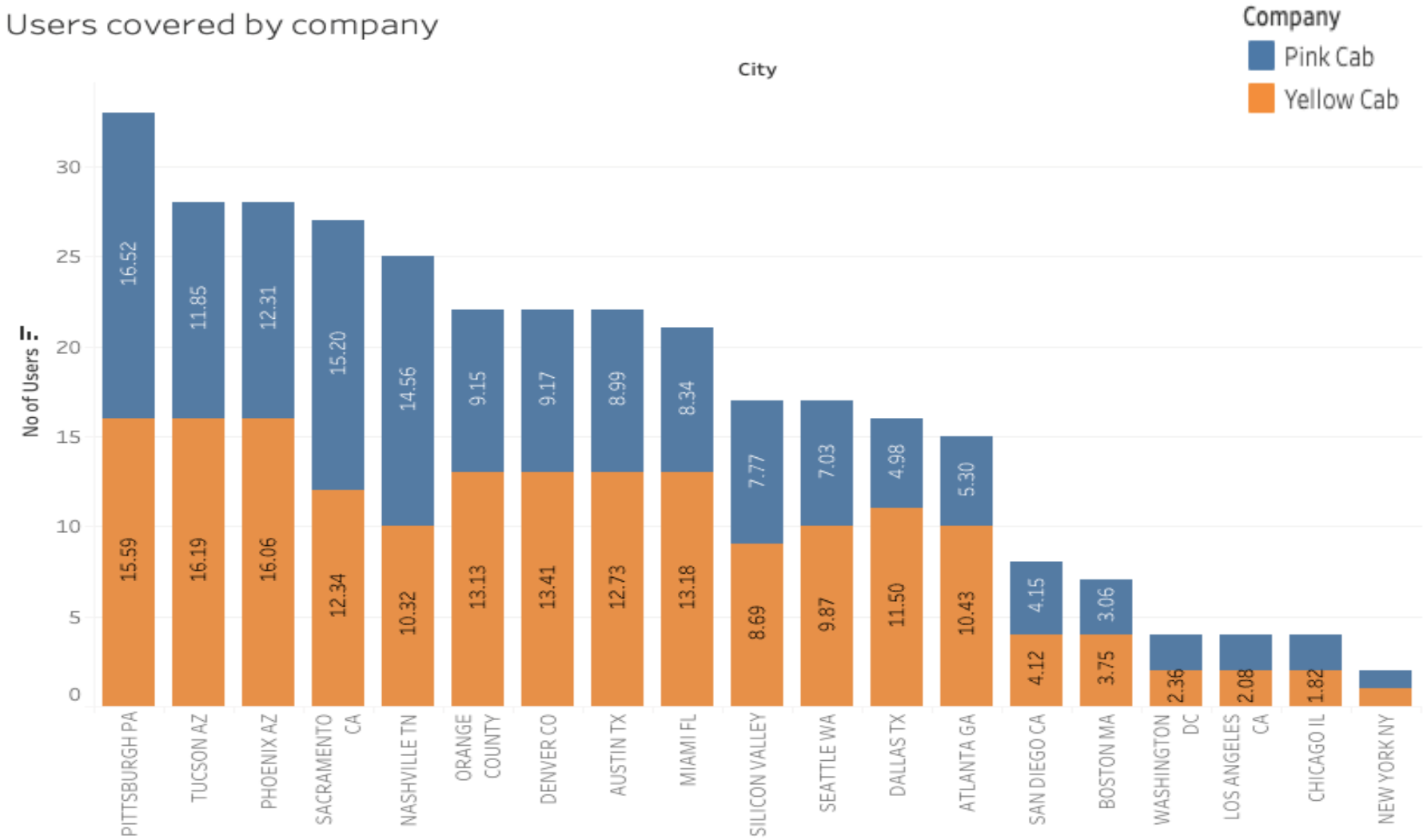
- **Timewise Analysis**: the dataset has a feature 'Date of Travel' which measures the events over a period of time.
- Here we will have a descriptive statistics per time interval.

Using the Date of Travel as index of the dataset,we are able to relate other features with respect to the period of time of events.

Date of Travel	Transaction ID	Company	City	KM Travelled	Price Charged	Cost of Trip	Year	Month	Day	Customer ID	Payment_Mode	Gender	Age	Income (USD/Month)	Populati
2016-01-01	10001340	Yellow Cab	WASHINGTON DC	10.20	202.54	135.8640	2016	1	1	52536	Card	Male	24	10606.0	418,8
2016-01-01	10001254	Yellow Cab	SILICON VALLEY	42.12	1342.05	520.6032	2016	1	1	13894	Card	Male	32	21898.0	1,177,6
2016-01-01	10001242	Yellow Cab	SILICON VALLEY	18.02	658.65	240.0264	2016	1	1	14513	Card	Male	62	3892.0	1,177,6
2016-01-01	10000885	Yellow Cab	NEW YORK NY	32.20	940.73	452.0880	2016	1	1	1698	Card	Male	23	22120.0	8,405,8
2016-01-01	10000961	Yellow Cab	NEW YORK NY	19.04	634.46	253.6128	2016	1	1	85	Card	Male	19	19765.0	8,405,8

RESULTS(Market size of the Taxi Company City wise)

City Wise Cab Users Covered By Company

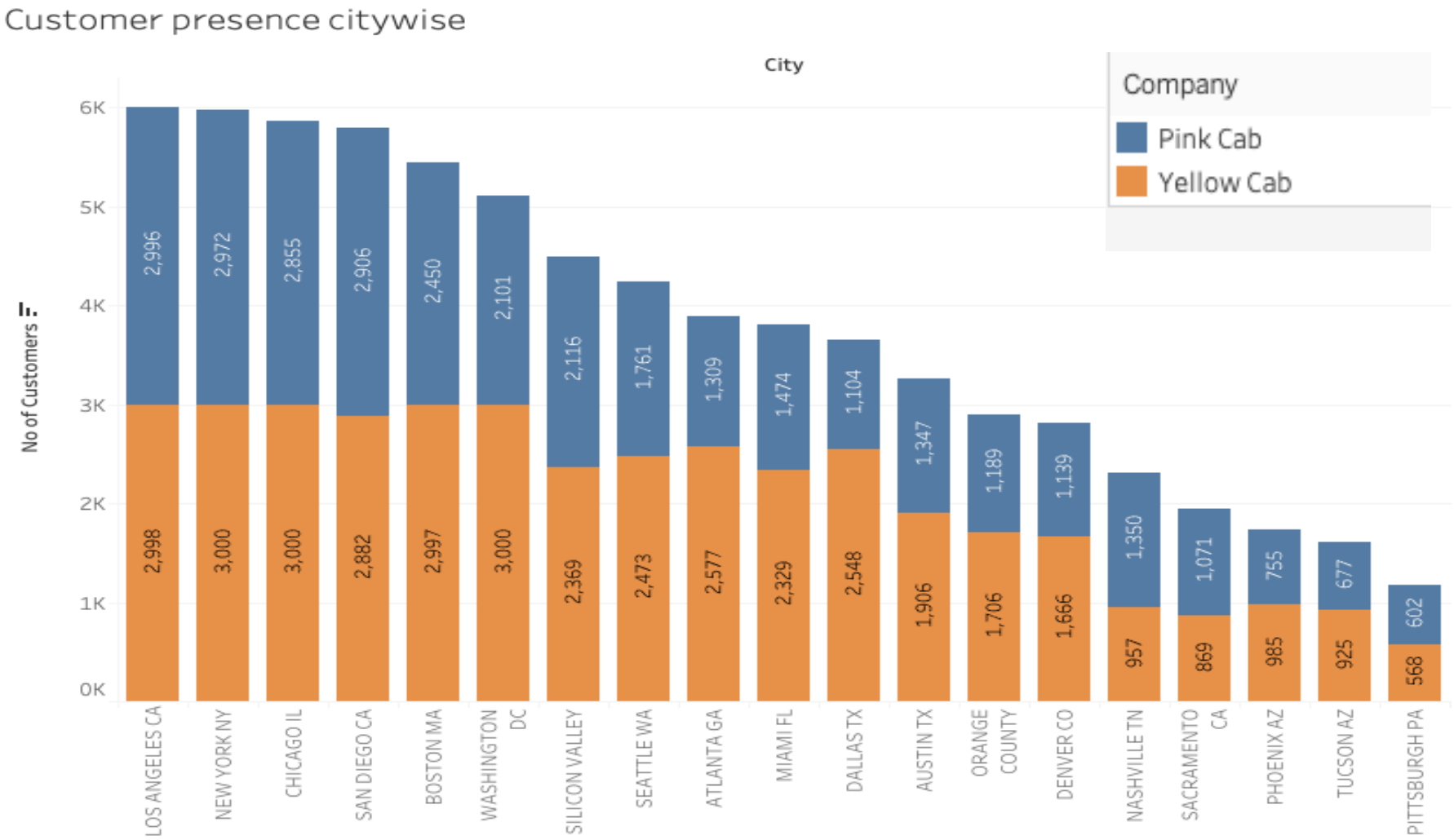


- This is the number of users covered by Yellow and Pink cab In the city against **all cab users** present In the city

Out of 19 Cities Pink cab have higher customer reach as compared to Yellow cab ,in following 4 citie

- SanDiego
- Nashville
- Sacramento
- Pittsburgh

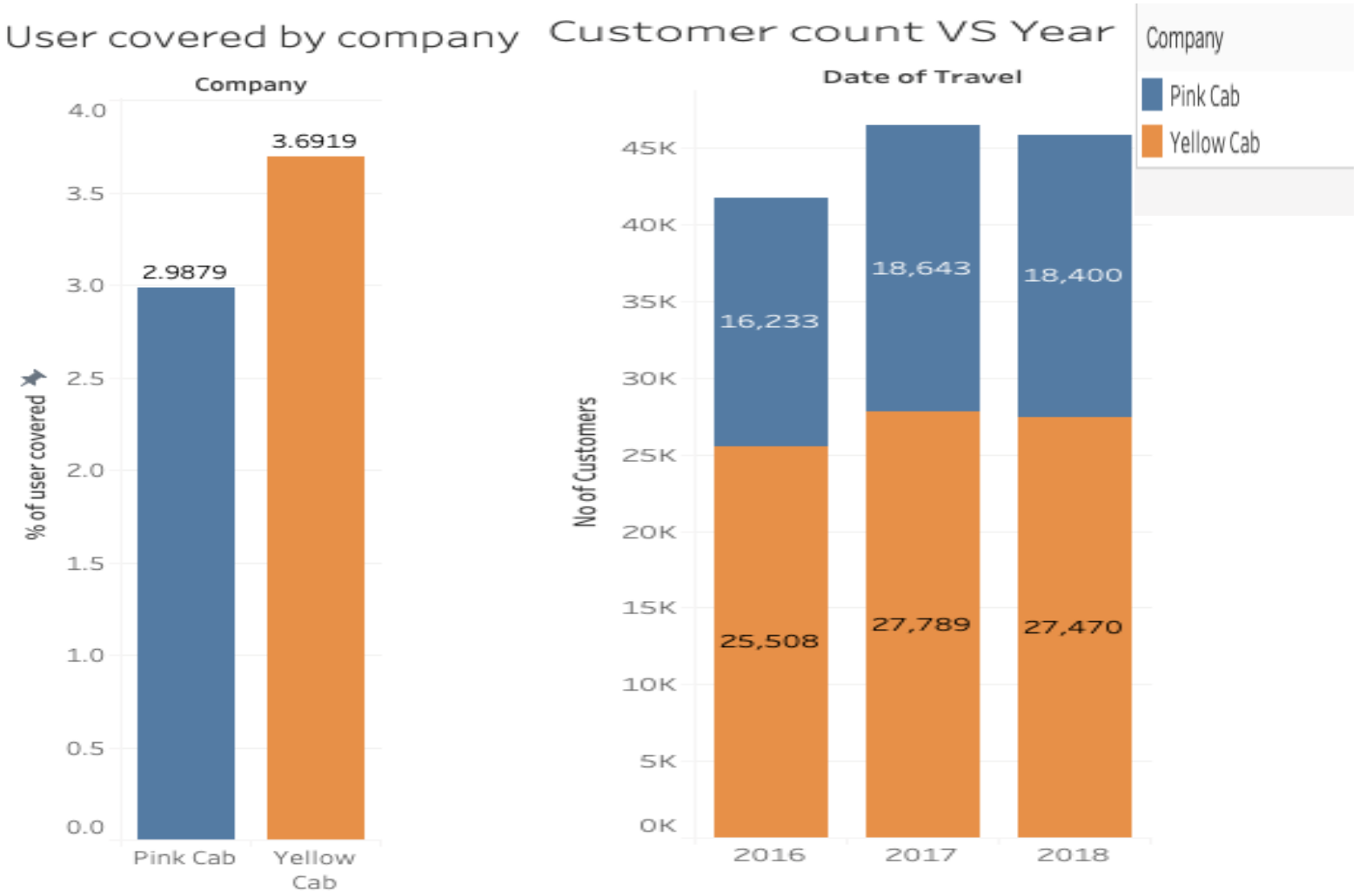
Customer Presence of Yellow and Pink cab city wise



Out of 19 Cities Pink cab have higher customer reach as compared to Yellow cab ,in following 4 cities:

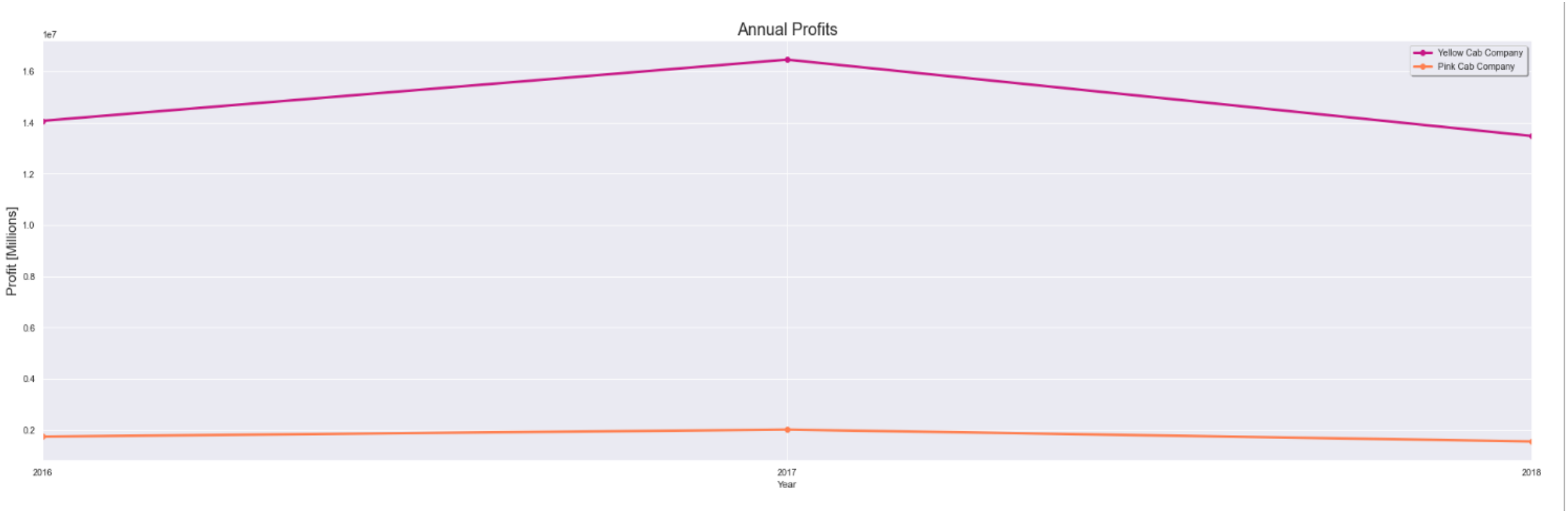
- SanDiego
- Nashville
- Sacramento
- Pittsburgh

User Covered by Company and Customer base Year wise



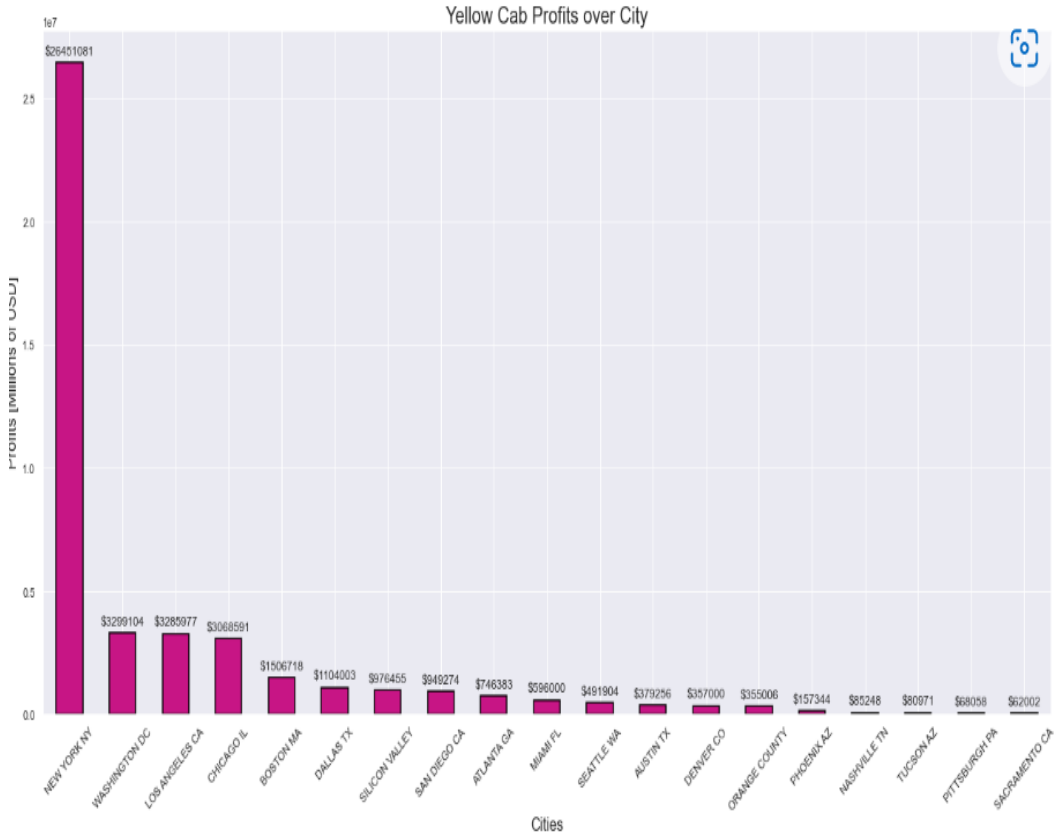
- Yellow cab has reach of around 3.7% cab users while Pink cab has 2.9% reach.
- We have considered all cab users present in 19 cities to calculate Yellow and Pink cabs coverage.
- Yellow cab higher customer base as compared to Pink cab.
- There is 1.3% loss in customer count of Pink cab from 2017-2018 and for the same period Yellow cab lost 1.14%

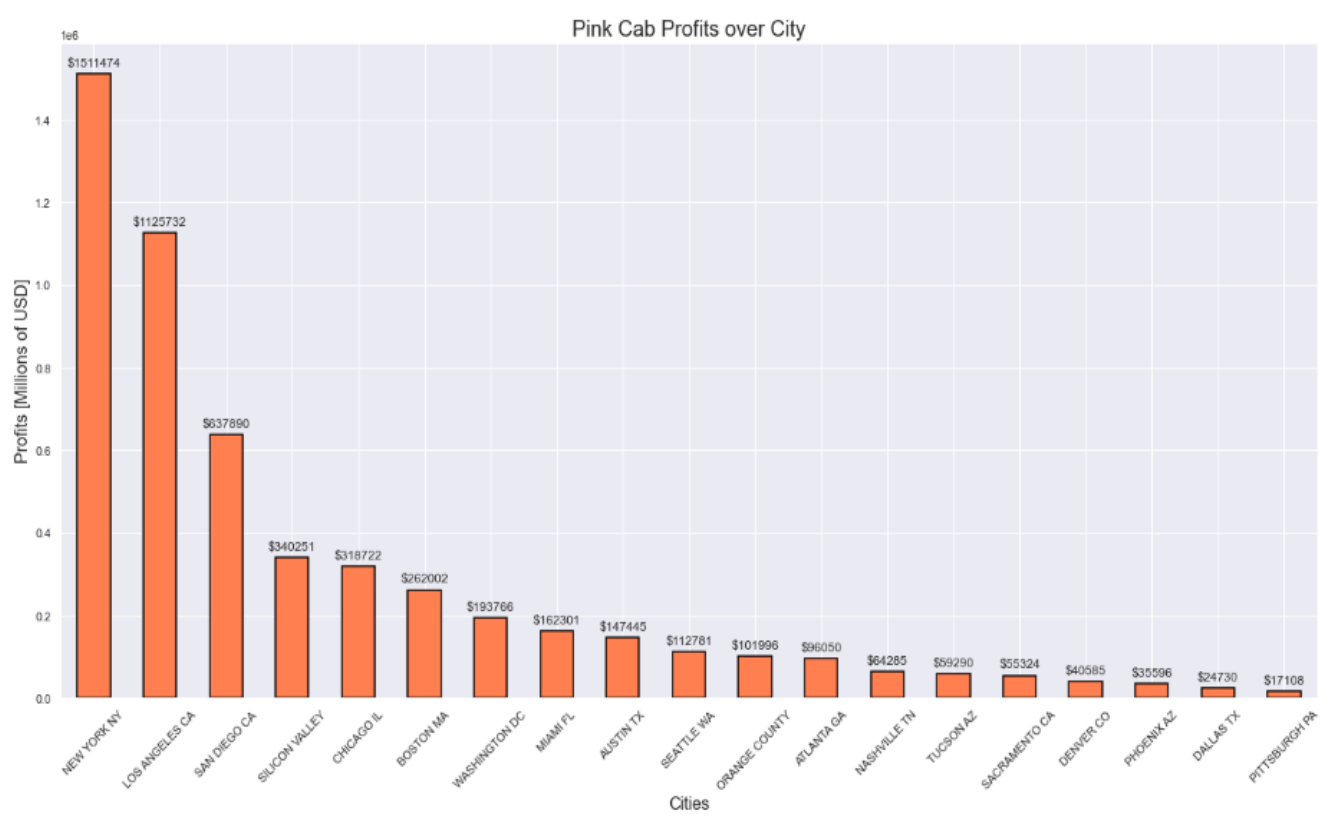
RESULT(Yearly Profit Analysis Of the Taxi Company)



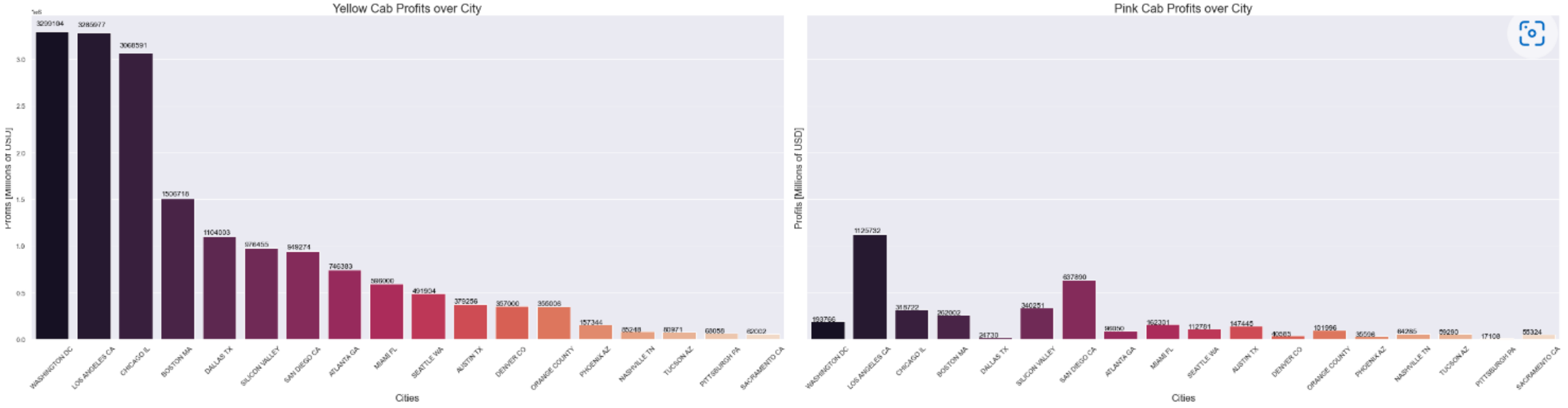
YellowCab's profits over the last 3 years are eight times higher than PinkCab's Company

RESULT(Market Share of the Taxi Company Citywide)





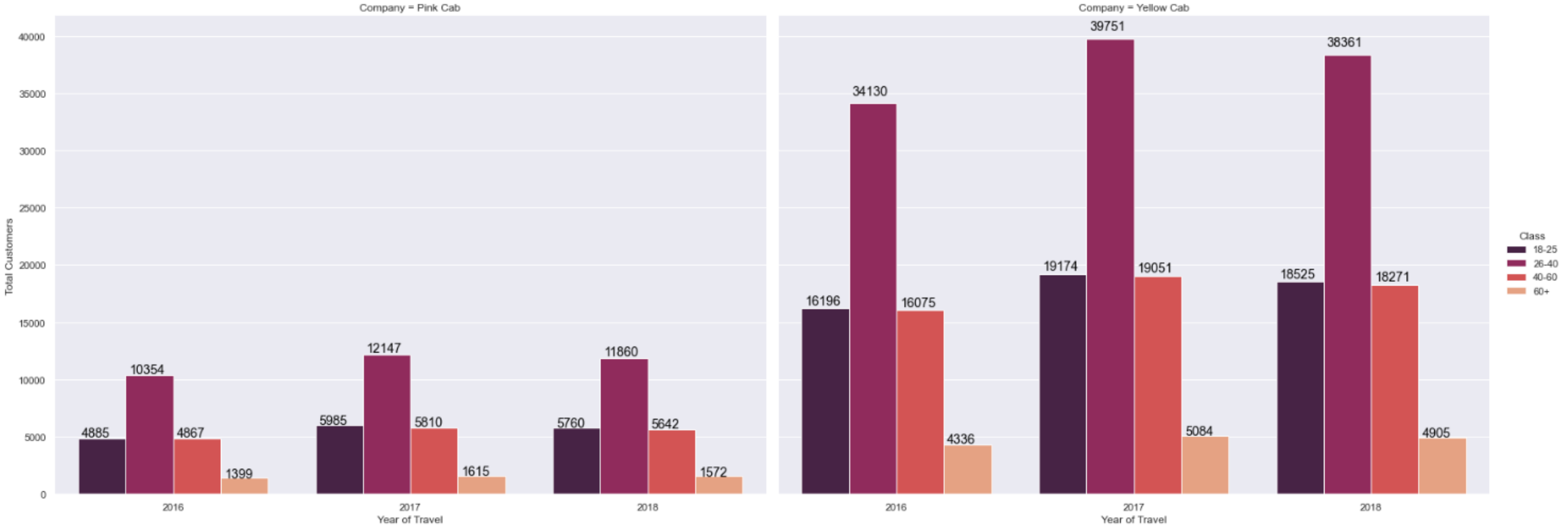
RESULT(Market Share of the Taxi Company Citywide)



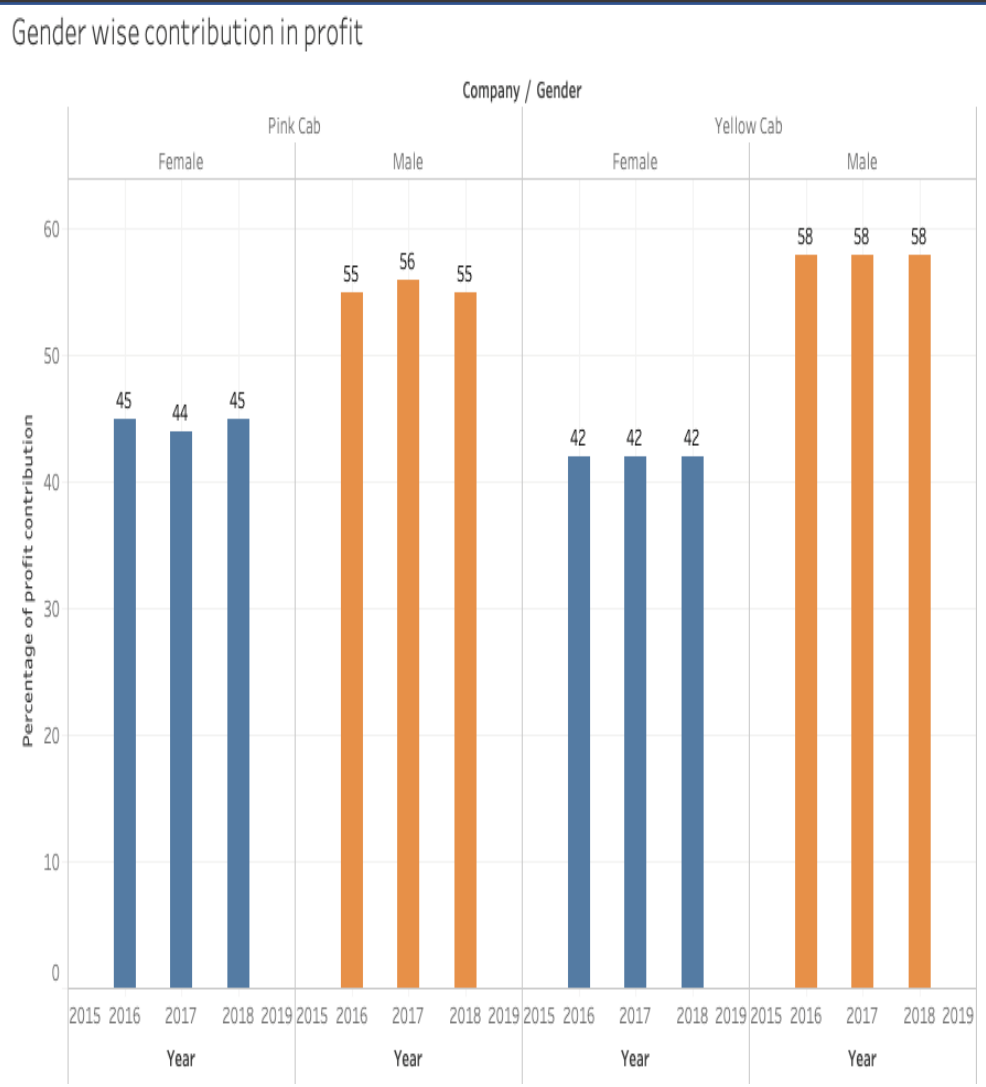
In this picture New York has been removed from both companies to get a better sight of the Profits over the other cities.

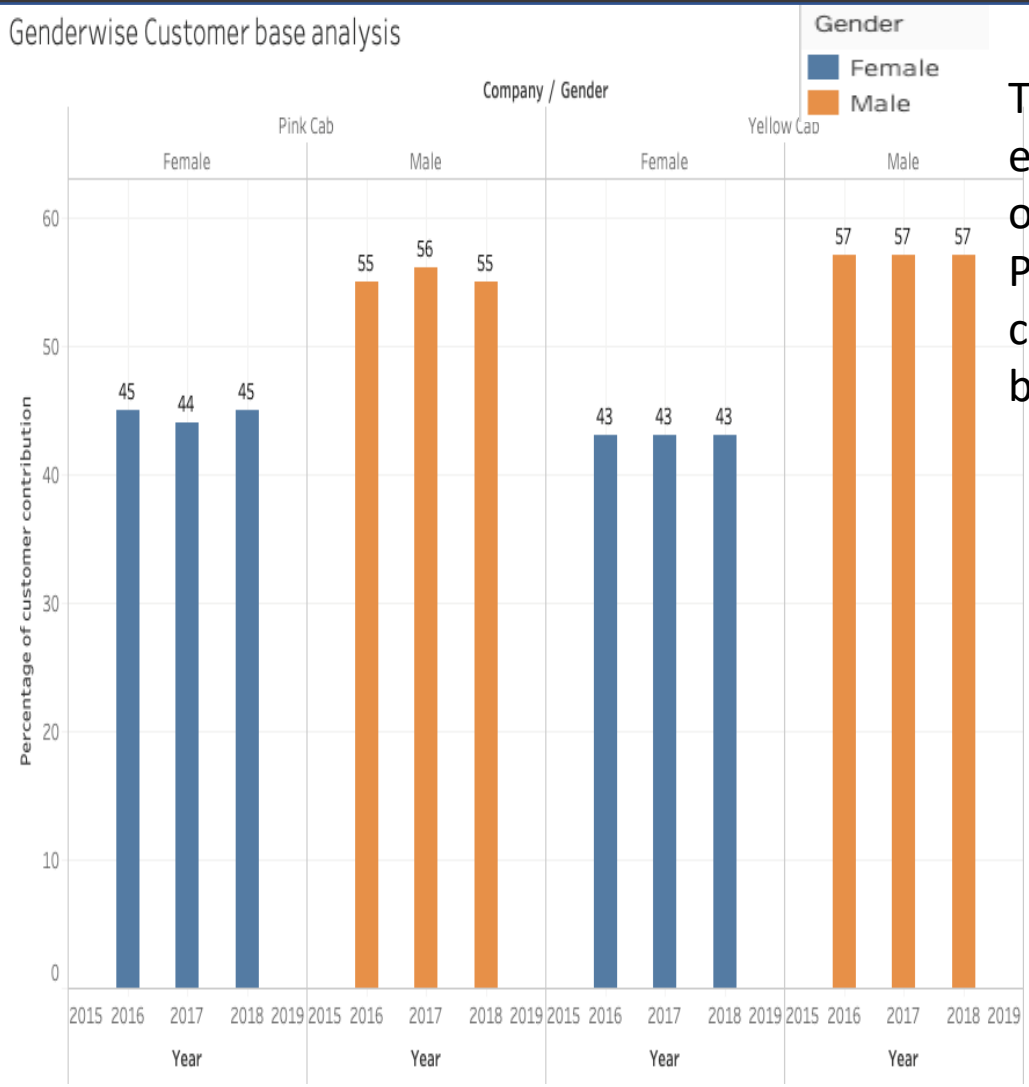
We can conclude that YellowCab has larger marketshare in every City

RESULT 4(customer base Analysis Gender wise



RESULT(Profit and customer base Analysis Gender wise





There is almost equal distribution of gender in the Profit and customer base of both the cabs.

Recommendations

I compared the two cab companies on the following criteria and found Yellow cab to be superior to Pink cab:

1. Pink have more customers than Yellow Cab but yet Yellow Cab has a larger market share in all cities than Pink Cab, Hence we can conclude that Pink Cab has not made more profit despite its larger customer presence but yellow Cab was able to make more profit with its limited users.

2.Profit Estimation : Higher profits over time and fewer monthly volatility

- City Profits: Yellow Cab has a larger market share in each city. Yellow Cab has a strong presence in every city. Pink Cab, on the other hand, performs poorly in eight cities.

3.. Demand Assessment: Yellow Cab Company's demand is more than treble that of Pink Cab Company’s.

- Demand Agewise: Both companies show the same Demand Agewise distribution.

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

Thank You