

G2M Cab Investment Project

Baris Sonmez

AGENDA

- Problem statement
- Dataset information
- Exploratory data analysis
- Multiple hypothesis test

Problem Statement

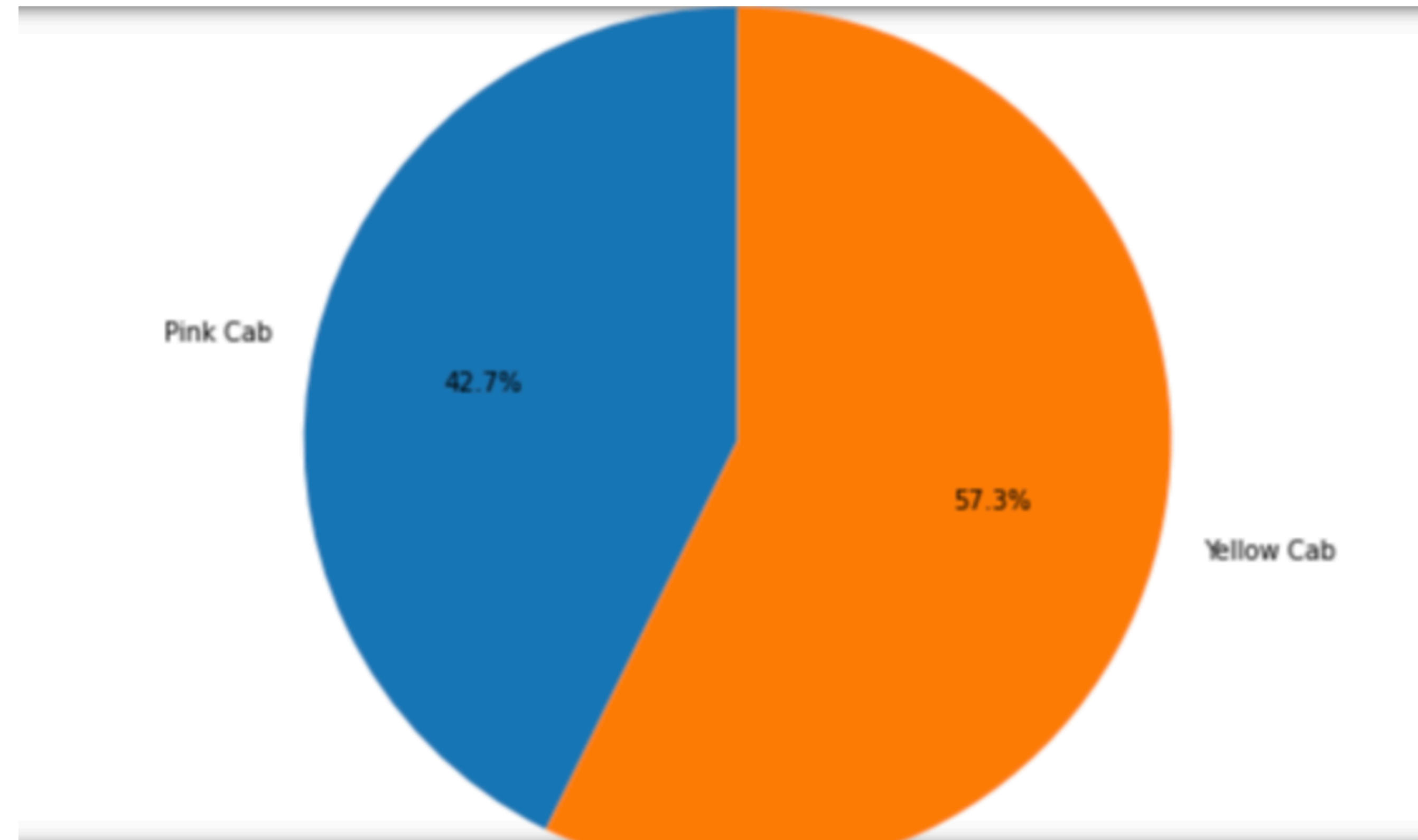
XYZ is a private equity firm in the US. Due to remarkable growth in the Cab Industry in the last few years and multiple key players in the market, it is planning for an investment in the Cab industry

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

Which Company has more Users?

There are two different taxi cab companies pink and yellow cab.

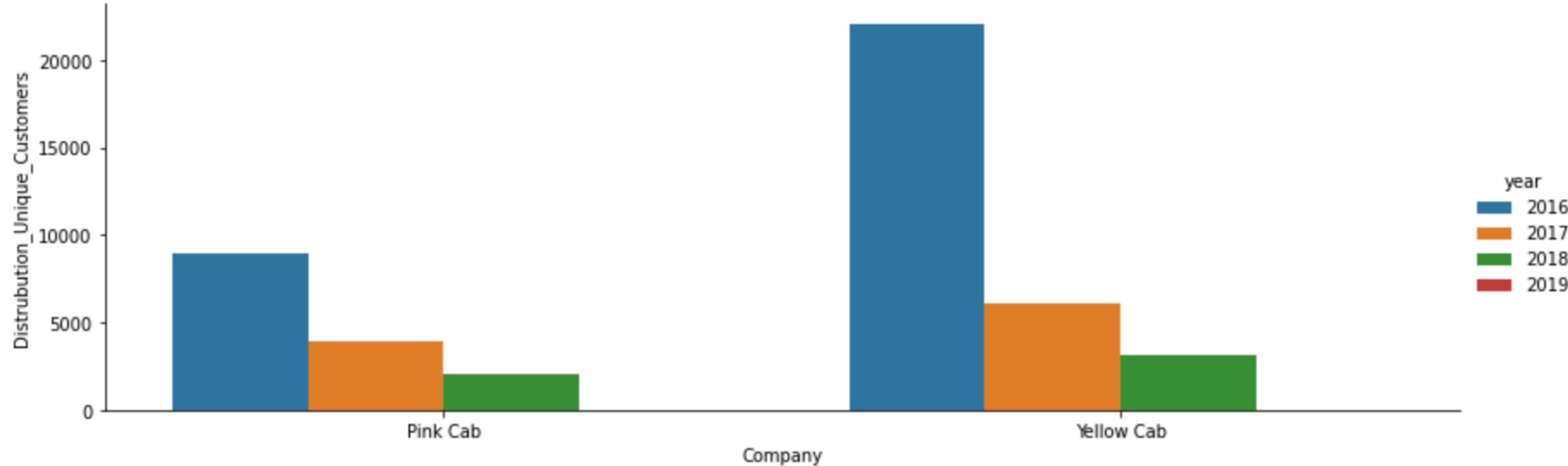
Pink cab services are having '42.7' percent and Yellow cab services are having '57.3' percent of customer shares respectively



Distribution of number of unique customers over years

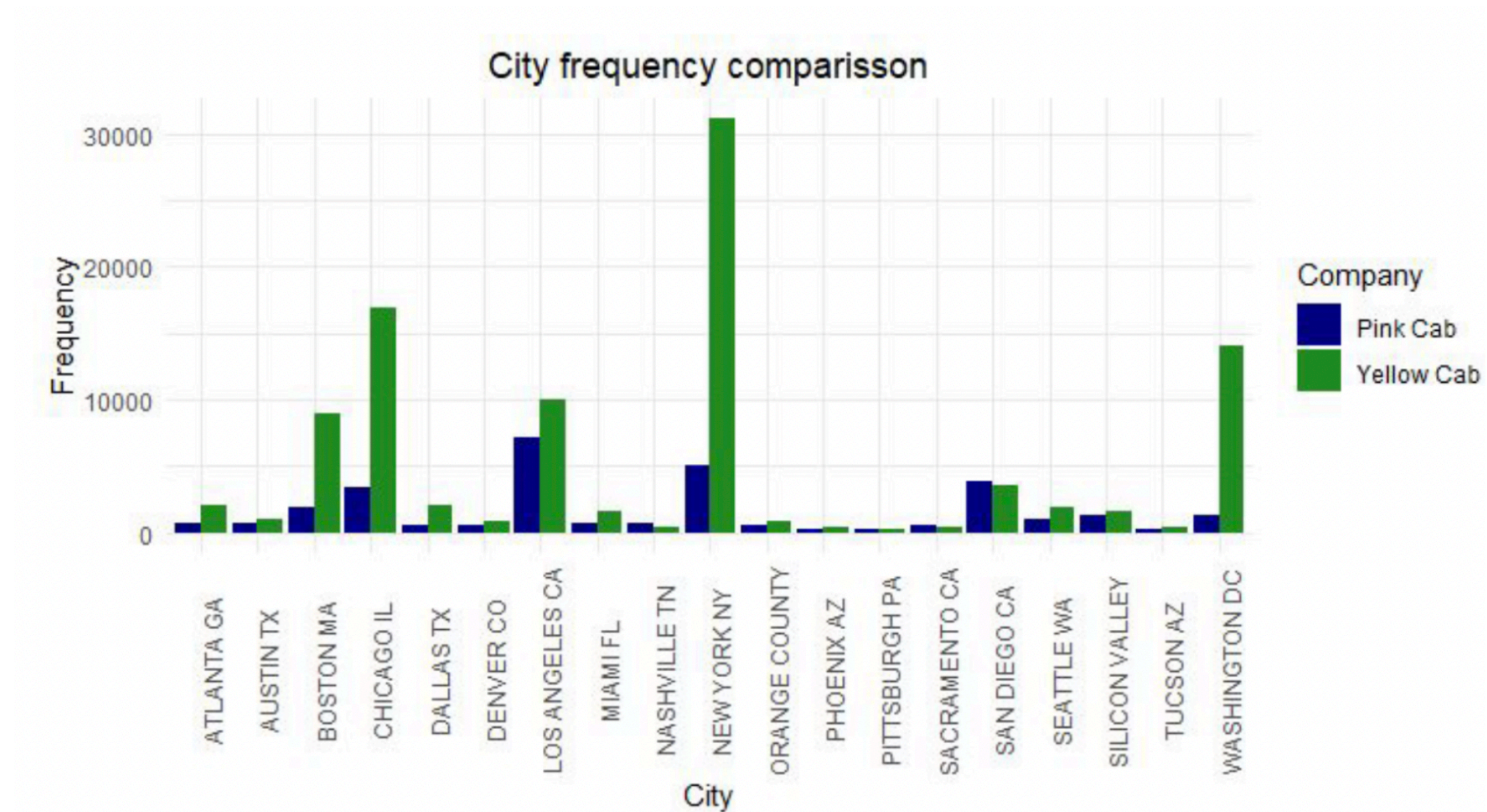
Dataframe Shape: (8, 3)

	Company	year	Distrubution_Unique_Customers
0	Pink Cab	2016	8914
1	Pink Cab	2017	3874
2	Pink Cab	2018	2037
3	Pink Cab	2019	4
4	Yellow Cab	2016	22095
5	Yellow Cab	2017	6117
6	Yellow Cab	2018	3096
7	Yellow Cab	2019	11



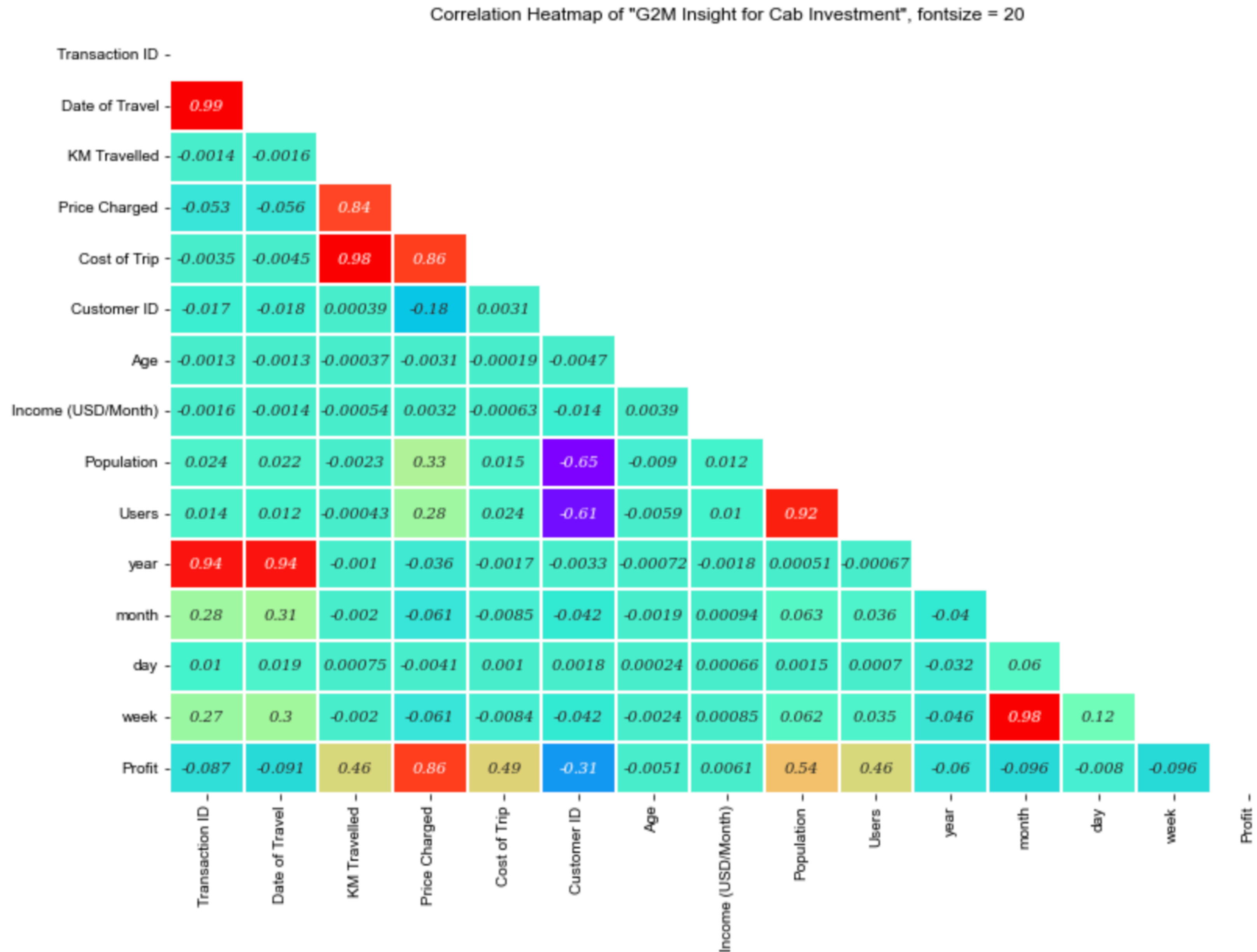
Which City has high Users?

- New York, Washington D.C. and Chicago are important cities for yellow cab company
- Los Angeles, New York and San Diego are important cities for pink cab company



Correlation Heat-map of G2M Insight For Cab Investment

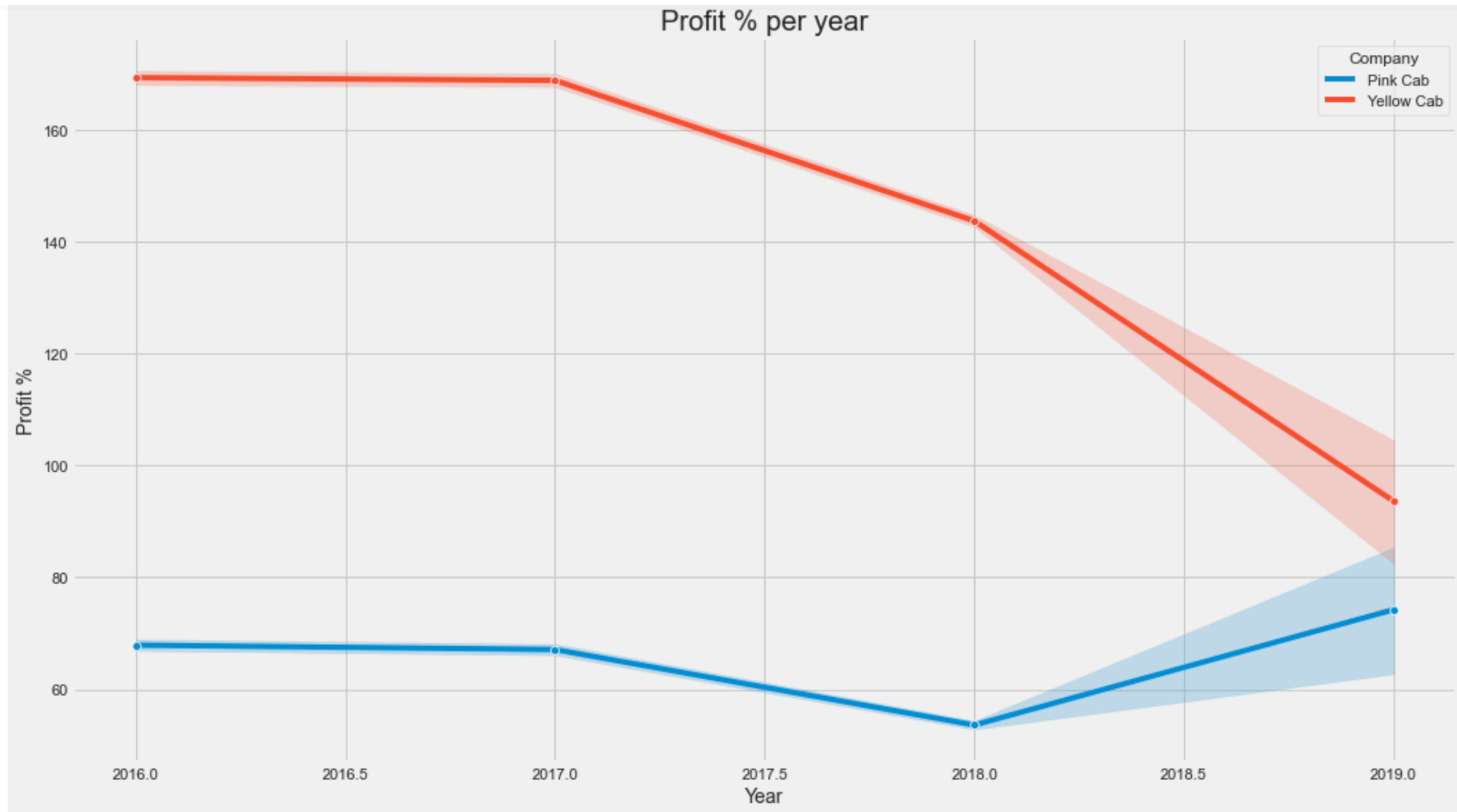
Price charged - KM
travelled , KM
travelled - cost of Trip and
Population -
users are having the
positive correlation



Which Company has a high price charged?

Pink cab taxi company's price charged average is 280 dollars while yellow cab's average is 380 dollars





Yellow cab taxi company's profit decreases after 2017, while pink cab taxi company's profit increases after 2018

Hypothesis 1: Cities which has population more than one million are more profitable

Pink Cab

50505 34206

P value is 3.645945885942243e-282

We accept the alternative hypothesis

(H1): There is a difference in profit by population for the Pink Cab.

Yellow Cab

172993 101688

P value is 0.0

We accept the alternative hypothesis

(H1): There is a difference in profit by population for the Yellow Cab.

Hypothesis 2: Which has km travelled more than 10 km are more profitable

Pink Cab

67552 17085

P value is 0.0

We accept alternative hypothesis (H1) that there is a difference regarding age for Pink Cab

Yellow Cab

219130 55249

P value is 0.0

We accept alternative hypothesis (H1) that there is a difference regarding km travelled for Yellow Cab

Hypothesis 3: In cities with high populations, the average profit from long-distance travel is higher than the average profit from short-distance travel in cities with low populations.

Pink Cab

40295 10162

P value is 0.0

We accept the alternative hypothesis (H1): There is a difference in profit by population for the Pink Cab.

Yellow Cab

34782 81111

P value is 0.0

We accept the alternative hypothesis (H1): There is a difference in profit by population for the Yellow Cab.