G2M Case Study

Virtual Internship

Jose Vicente Solorzano

11-October-2021

Agenda

1) Background
2) Data and Hypothesis
3) EDA (Exploratory Data Analysis)
4) Profit Analysis
5) Demand (travels) analysis
6) Profitability – non profitable travels
7) Customer retention
8) Conclusion

Data Glacier

Your Deep Learning Partner

1) Background

1) Background

Data Glacier
Your Deep Learning Partner

- XYZ is a private firm in US.
- Multiple key players in the market and growth of the industry.
- XYZ needs an outstanding understanding of the industry because they are analysing an investment between two companies: Yellow Cab Company and Pink Cab Company
- The main goal is to help XYZ with the needed information and knowledge to make the best decision posible.

Data Glacier

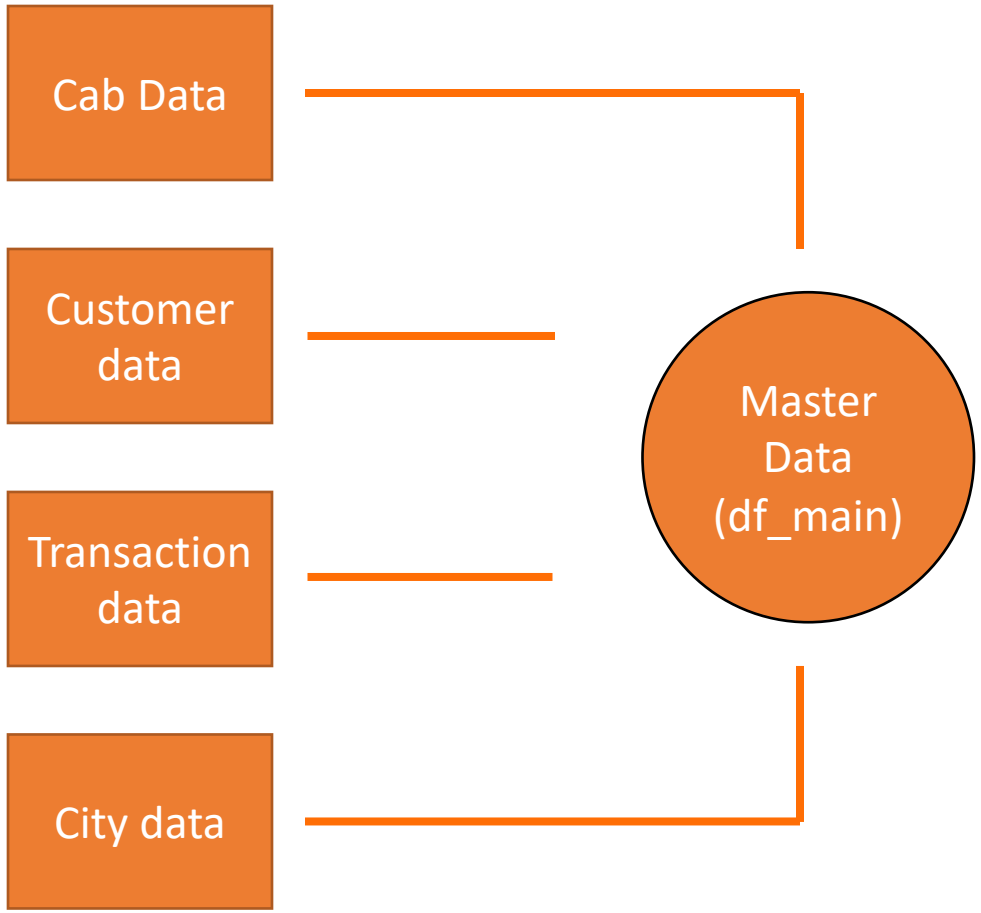
Your Deep Learning Partner

2) Data and Hypothesis

2) Data and Hypothesis

Data Glacier
Your Deep Learning Partner

Master data (df_main) formed by 4 datasets given…



Hypothesis:

1) Did the demand change over the years?

2) How is the demand by city?

3) How is the demand by age?

4) Is there any seasonability in the demand?

5) Is there any profitability diference among cities?

6) Is there any profitability diference among age range?

7) Are there customer retention in companies?

Data Glacier

Your Deep Learning Partner

3) EDA (Exploratory Data Analysis)

3) EDA

Data Glacier
Your Deep Learning Partner

Master data:
- 14 features (original)
- TimeFrame of data: 01-01-16 to 12-31-18
- 359.392 rows

Assumptions to take into account:

- “Profit per travel” was calculated by the difference of ”Price Charged” and “Cost of Trip”.

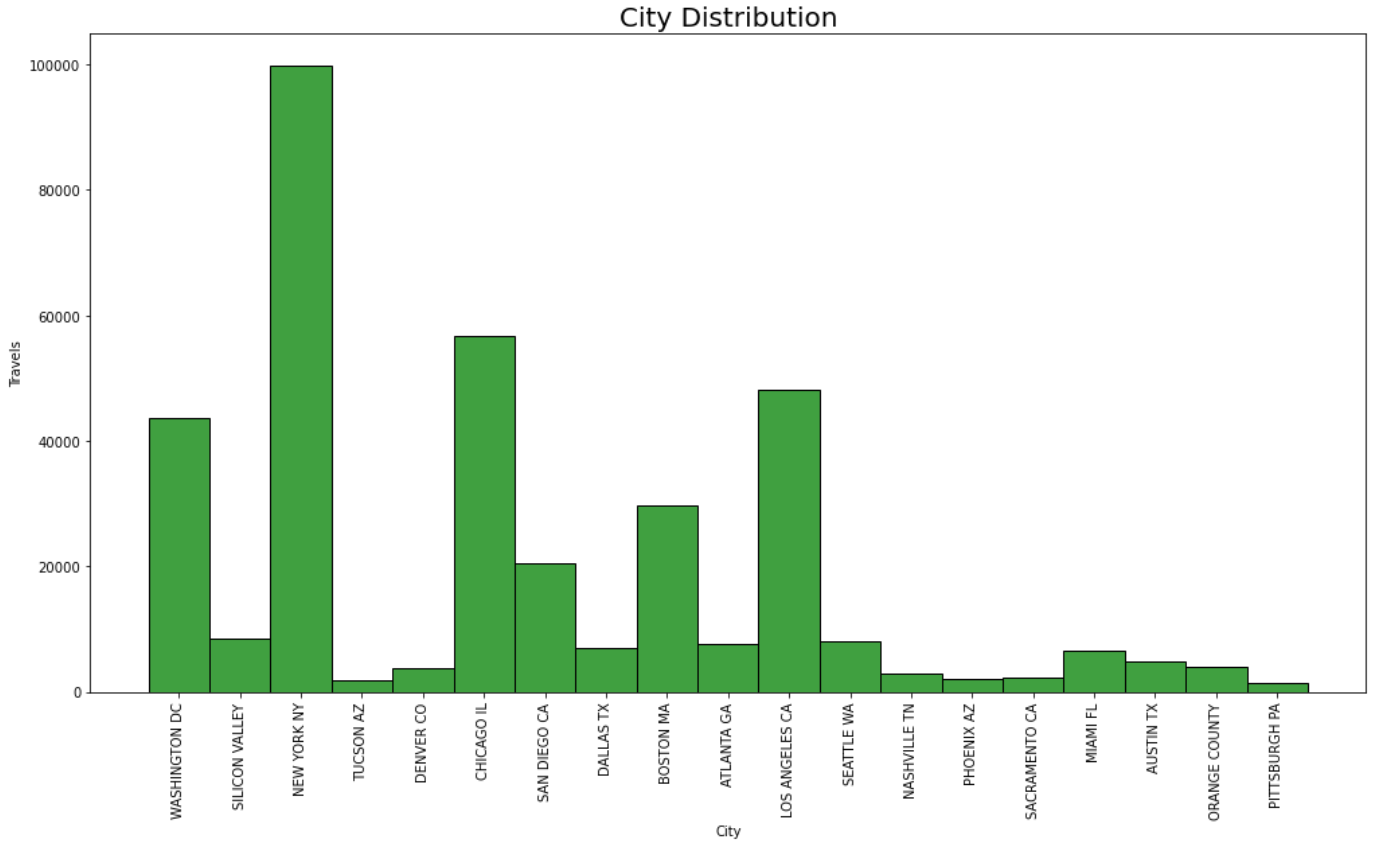
- Outliers were kept in this stage, because it is exploratory anlysis. For the model, I am going to evaluate the criterion for them.

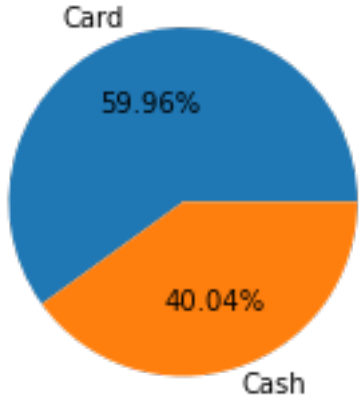
- Non profitable travel rate was calculated by considering “Profit per travel” < 0, grouped by the features I analysed

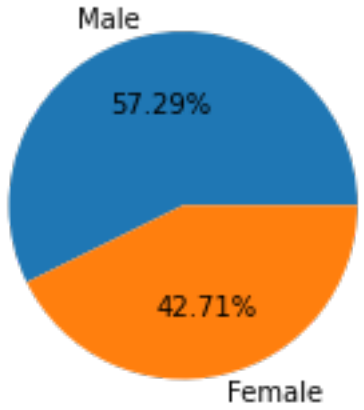
3) EDA

Data Glacier
Your Deep Learning Partner

Some information to get familiar with the data:



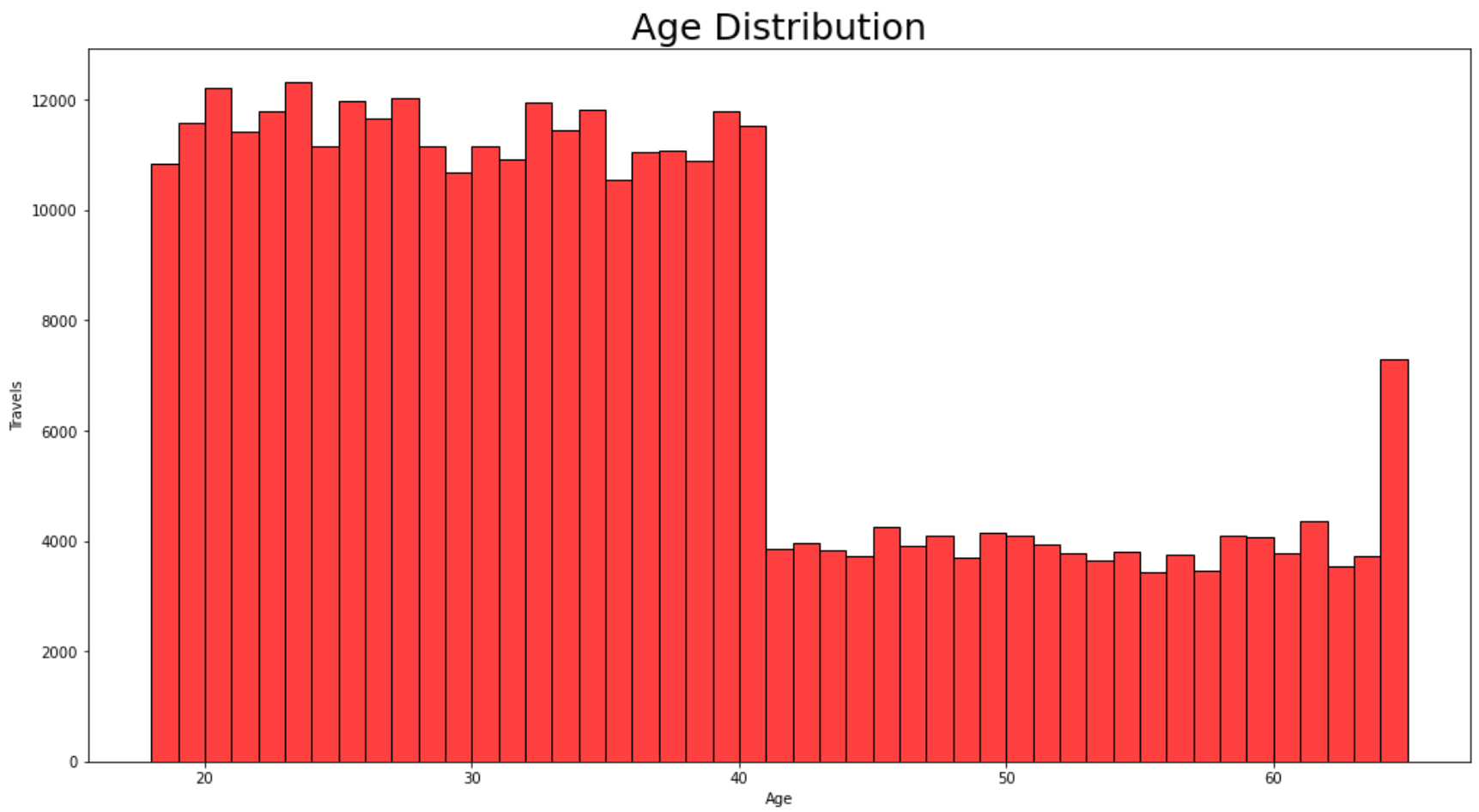




3) EDA

Data Glacier
Your Deep Learning Partner

Some information to get familiar with the data:



The demand of travels is around 10.000 and 12.000 till a huge drop at 41 y.o.

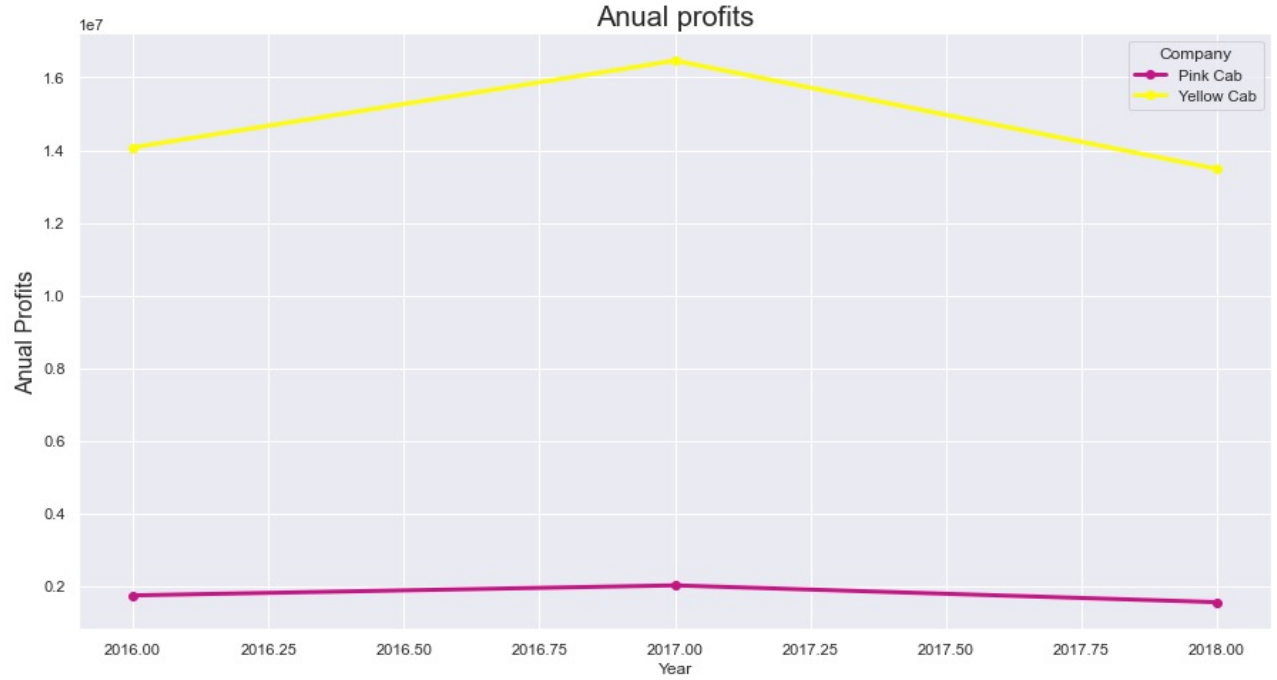
Data Glacier

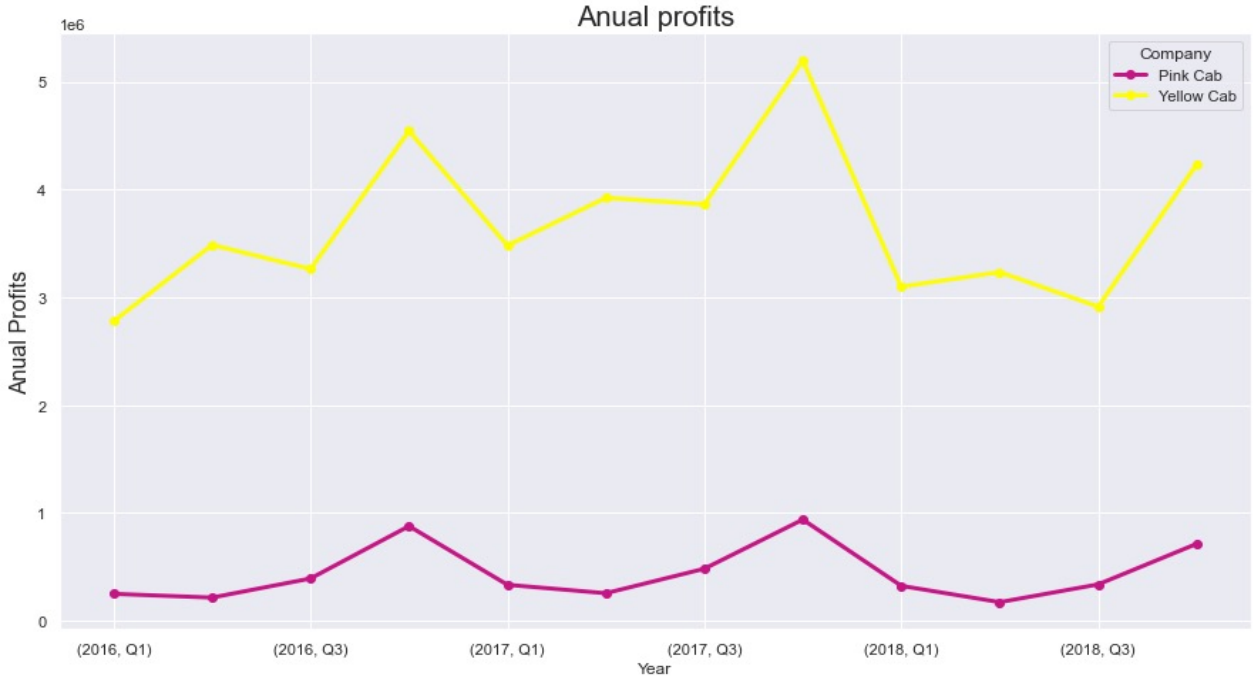
Your Deep Learning Partner

4) PROFIT ANALYSIS

4) Profit Analysis

Data Glacier
Your Deep Learning Partner



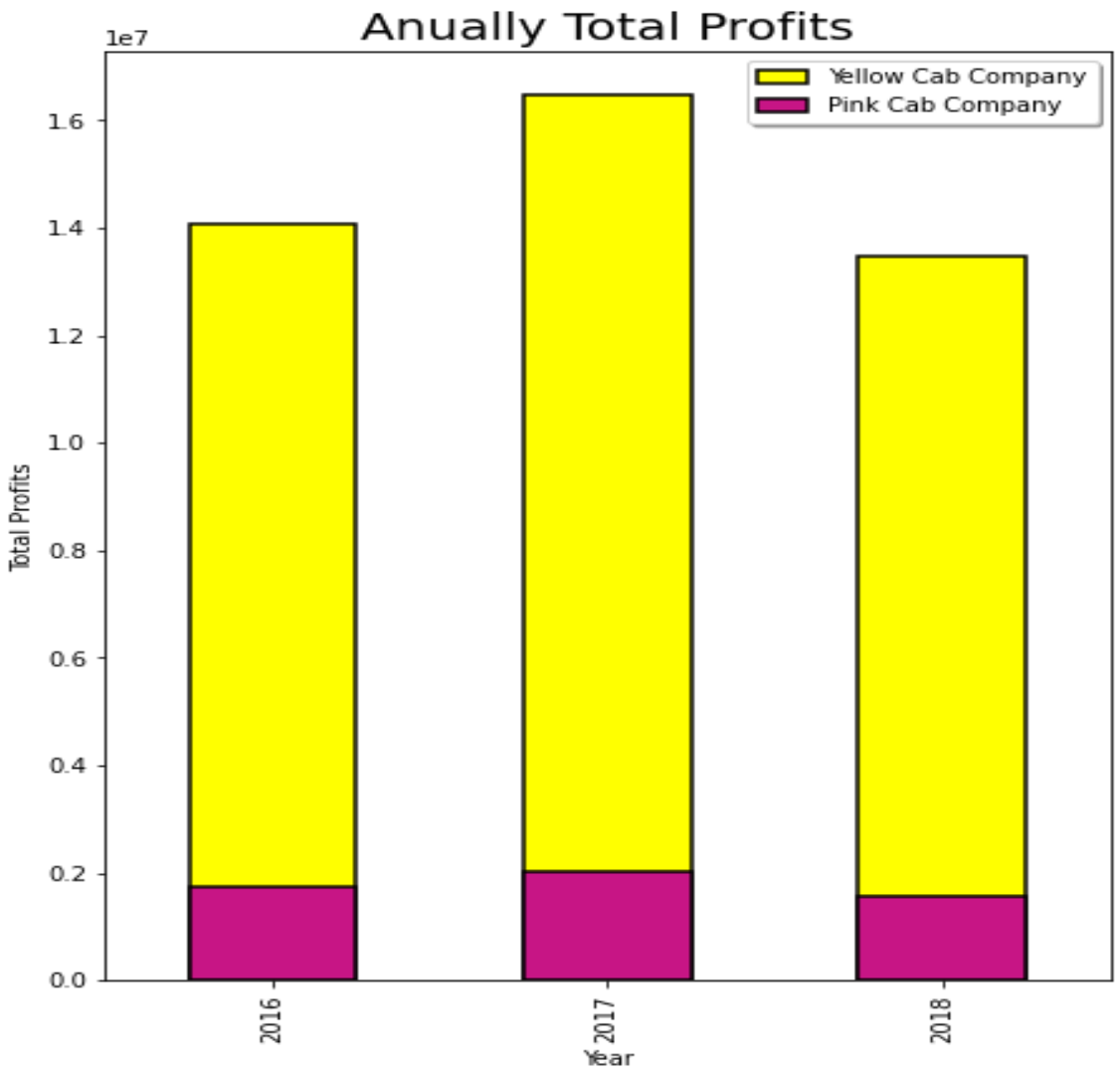


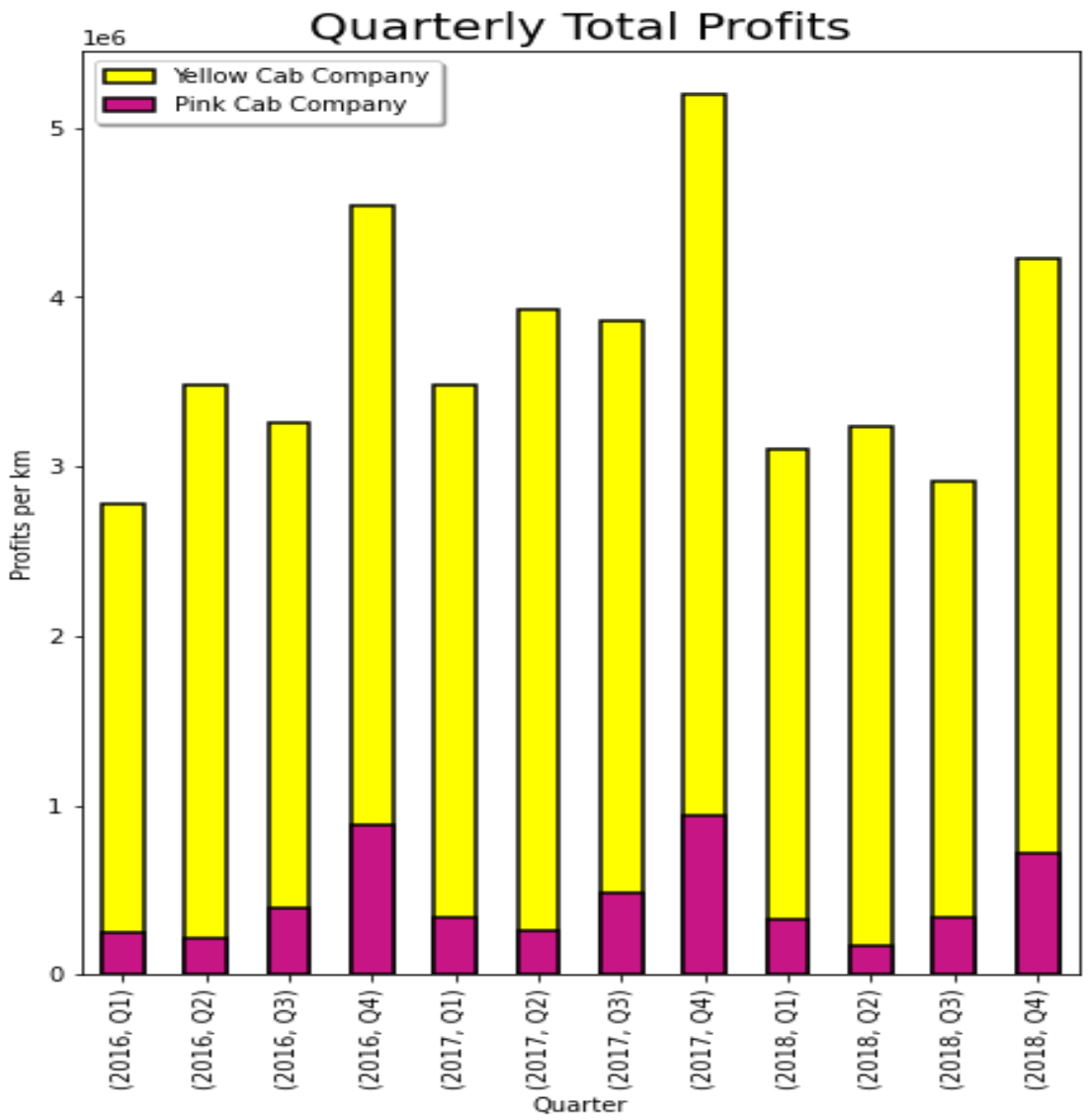
Yellow Cab profits are more tan 8 times greater than Pink Cab's.

Company	Year	
Pink Cab	**2016**	1739882.63
	2017	2015100.66
	2018	1552345.04
Yellow Cab	**2016**	14073886.48
	2017	16464267.14
	2018	13482219.55

4) Profit Analysis

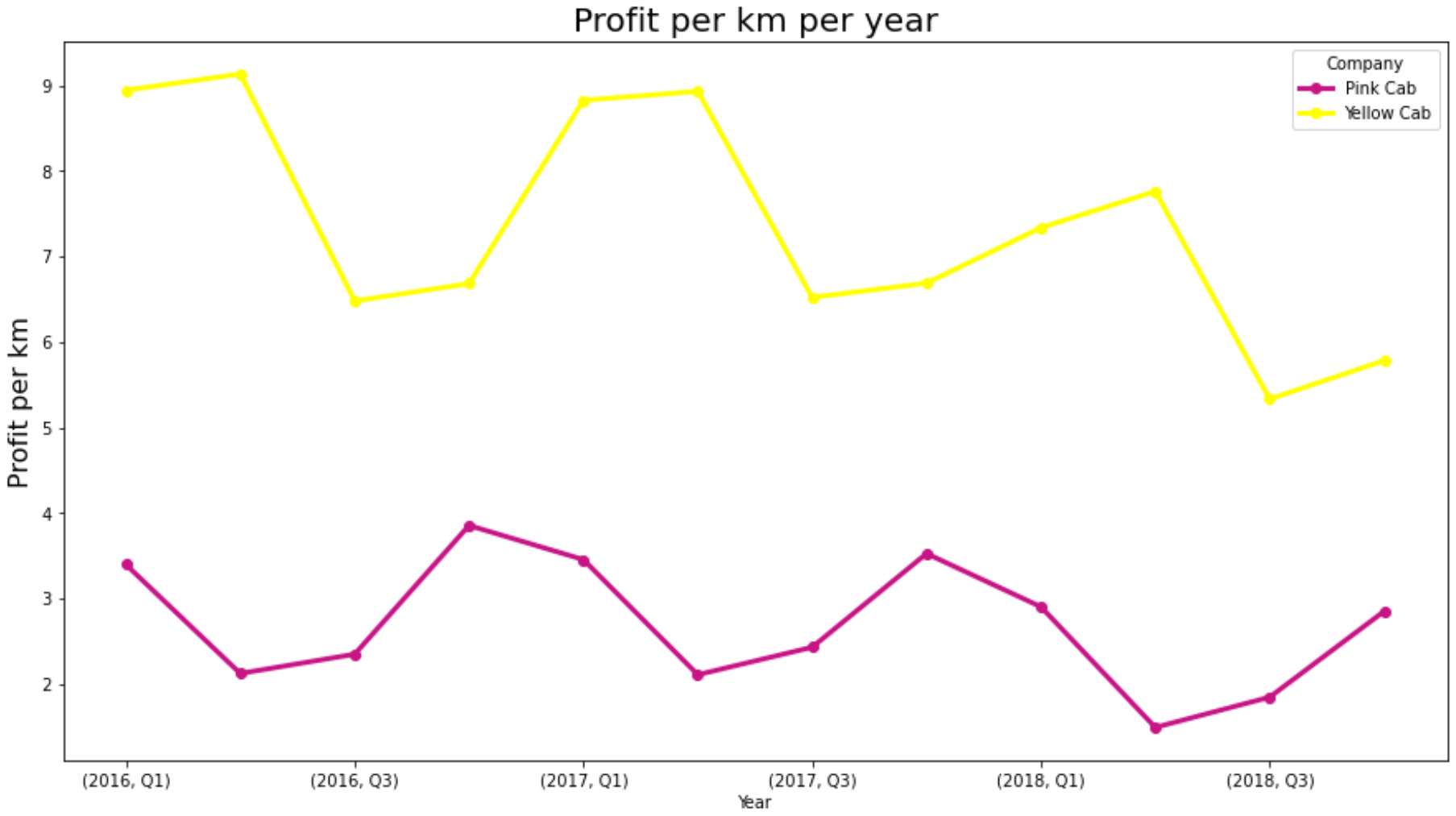
Data Glacier
Your Deep Learning Partner





4) Profit Analysis

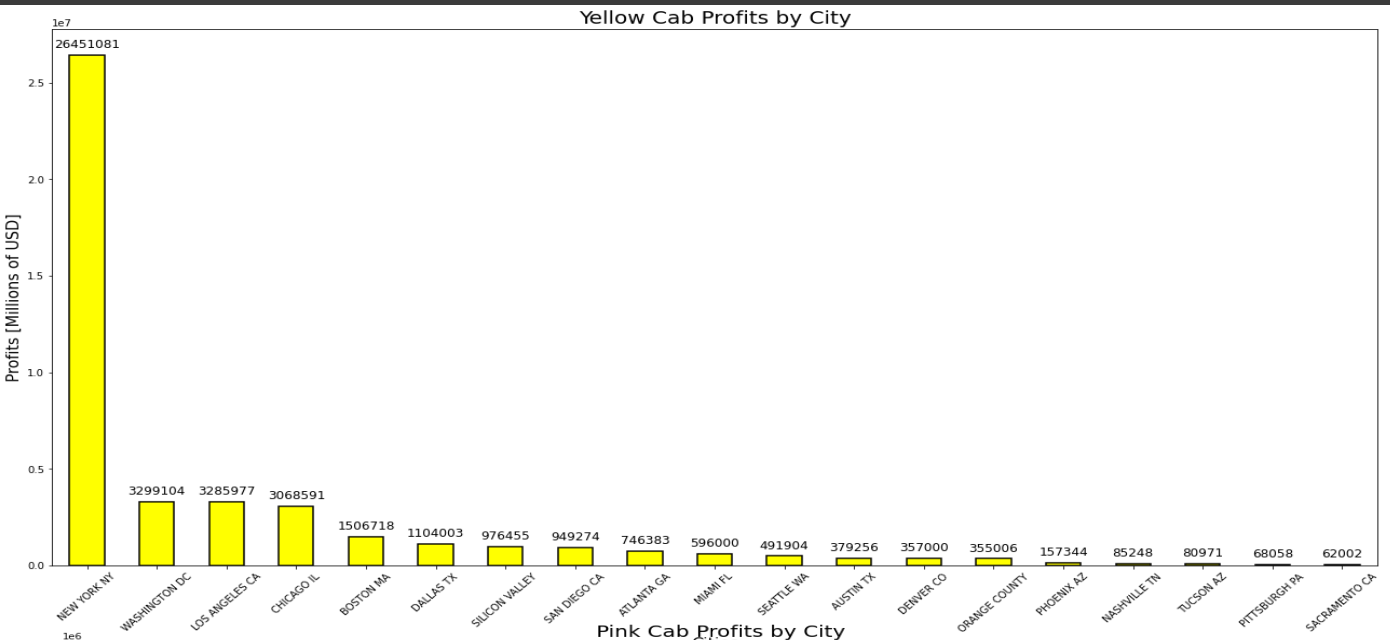
Data Glacier
Your Deep Learning Partner

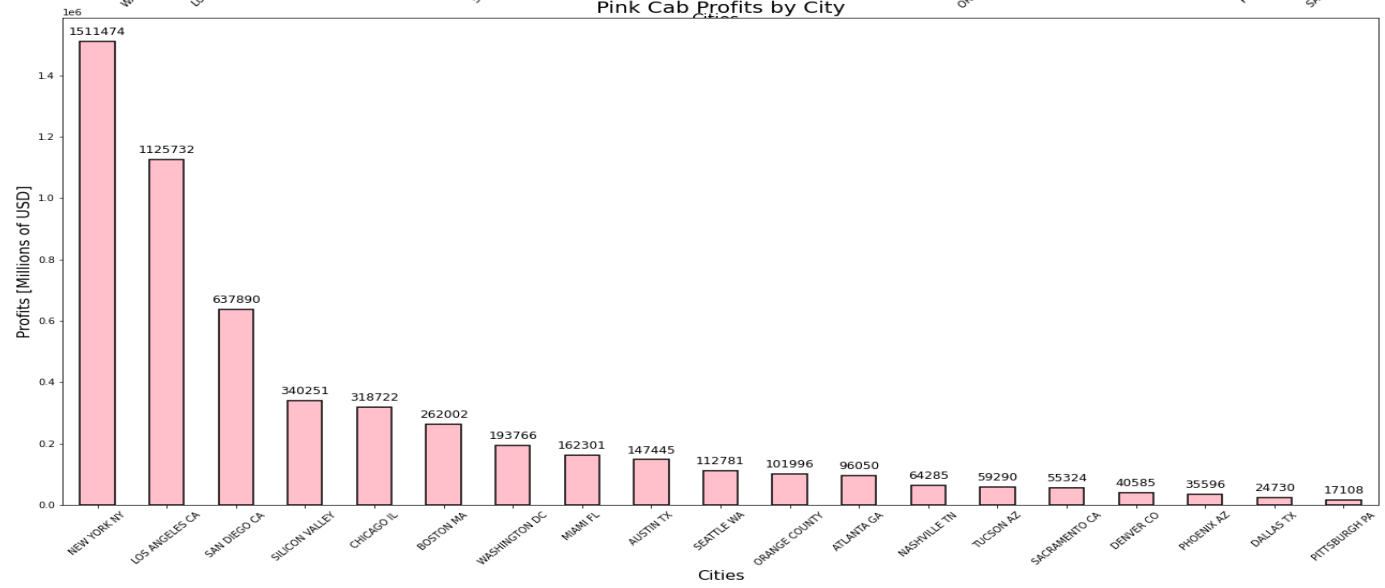


An interesting point to review is the profit per km.

The Profit per km of Yellow Cab Company is 3 times (average per quarter) greater tan Pink Cab.

4) Profit Analysis





The city producing the most profits for both companies is New York. Equally, this is logic because cities producing more profits are cities with bigger populations.

NEW YORK NY 839617028745
CHICAGO IL 110709236250
LOS ANGELES CA 76614412221
SAN DIEGO CA 19654281816
WASHINGTON DC 18319636083
SILICON VALLEY 10032051071
MIAMI FL 8642906370
BOSTON MA 7392357856
DALLAS TX 6616385436

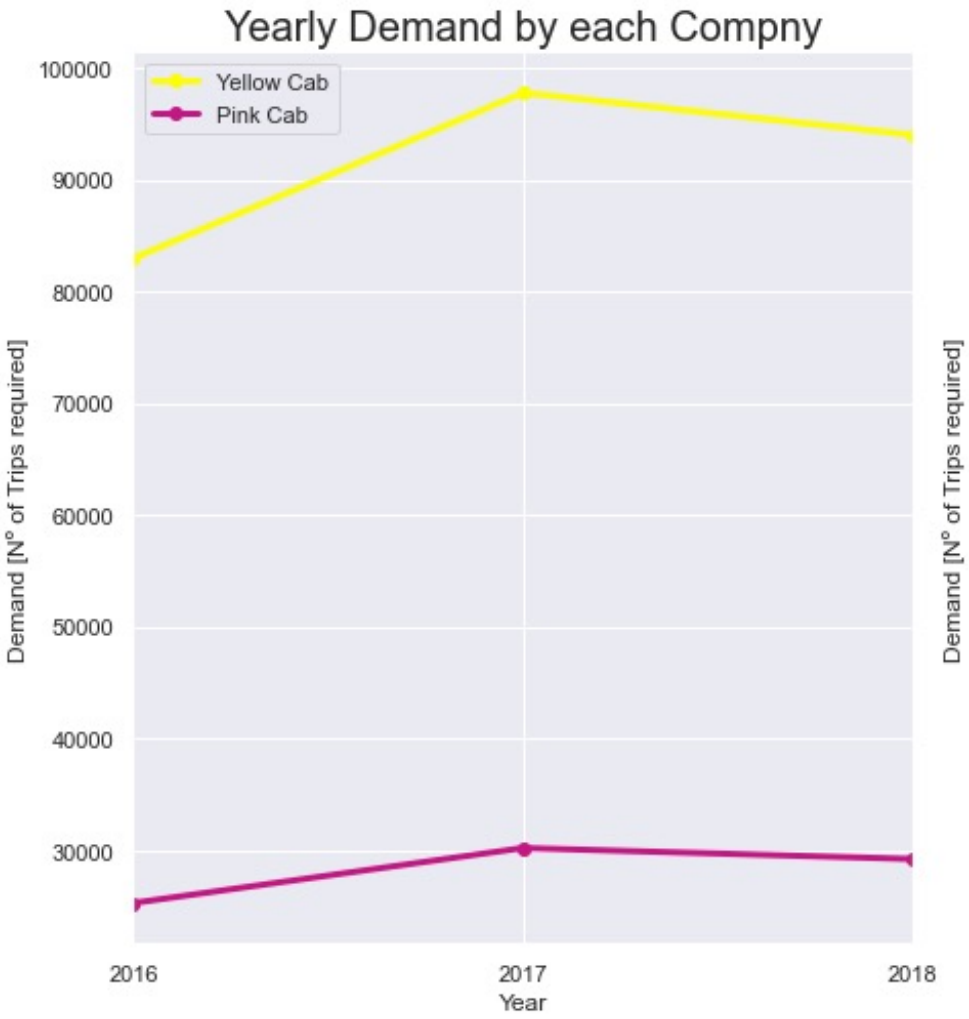
Data Glacier

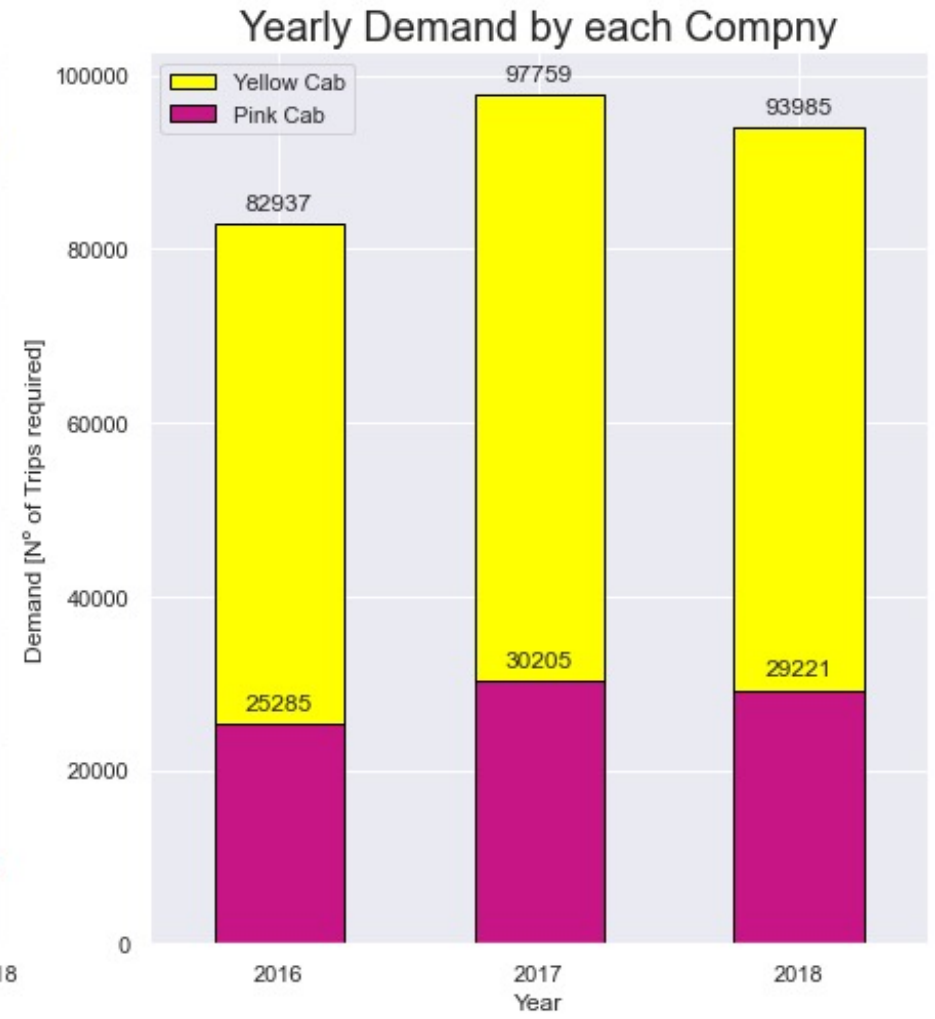
Your Deep Learning Partner

5) DEMAND ANALYSIS

5) Demand Analysis

Data Glacier
Your Deep Learning Partner



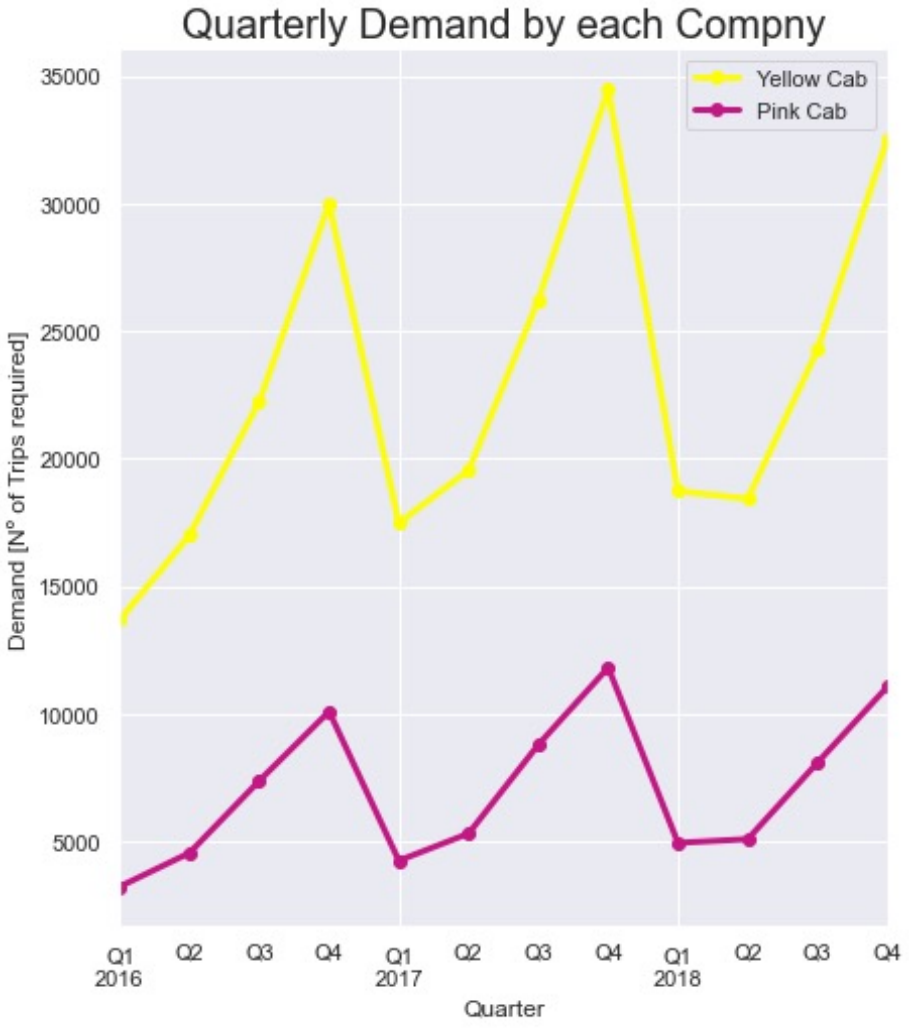


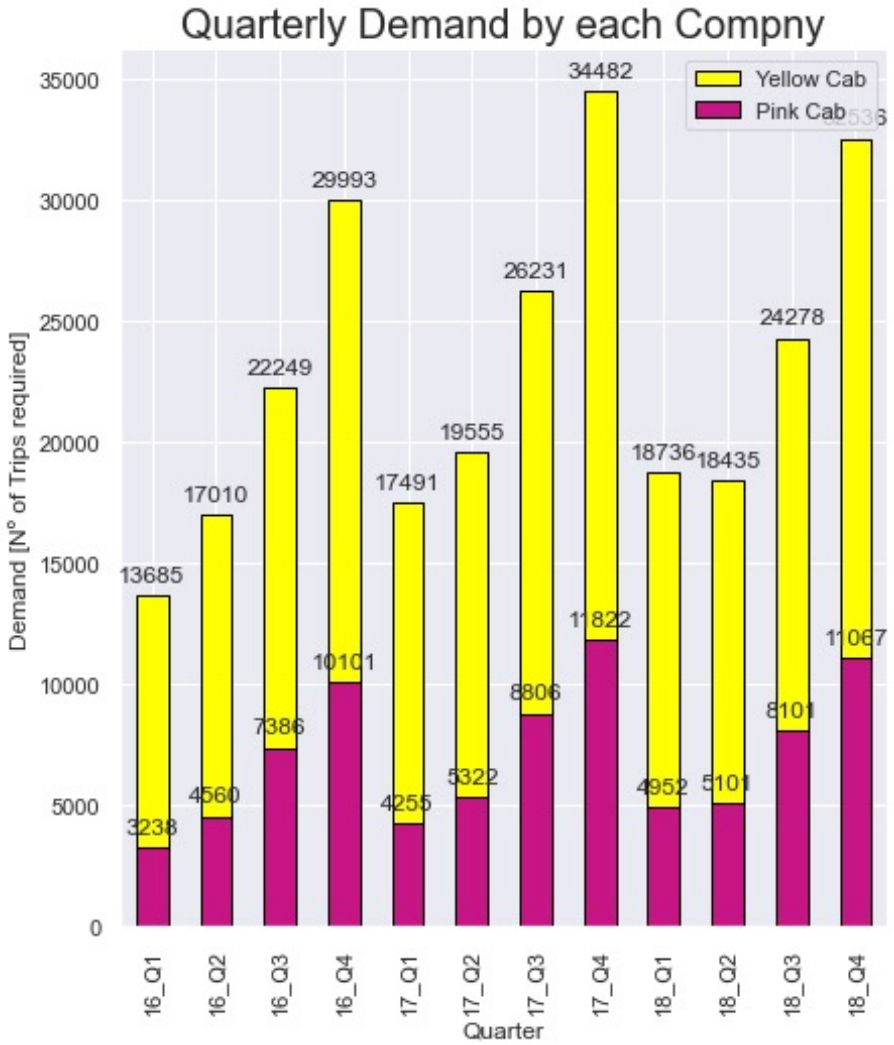
As well as profits, YC´s demand is way higher. PC represents just 30% of all YC's demand.

Demand cenceived as quantity of travels.

5) Demand Analysis

Data Glacier
Your Deep Learning Partner

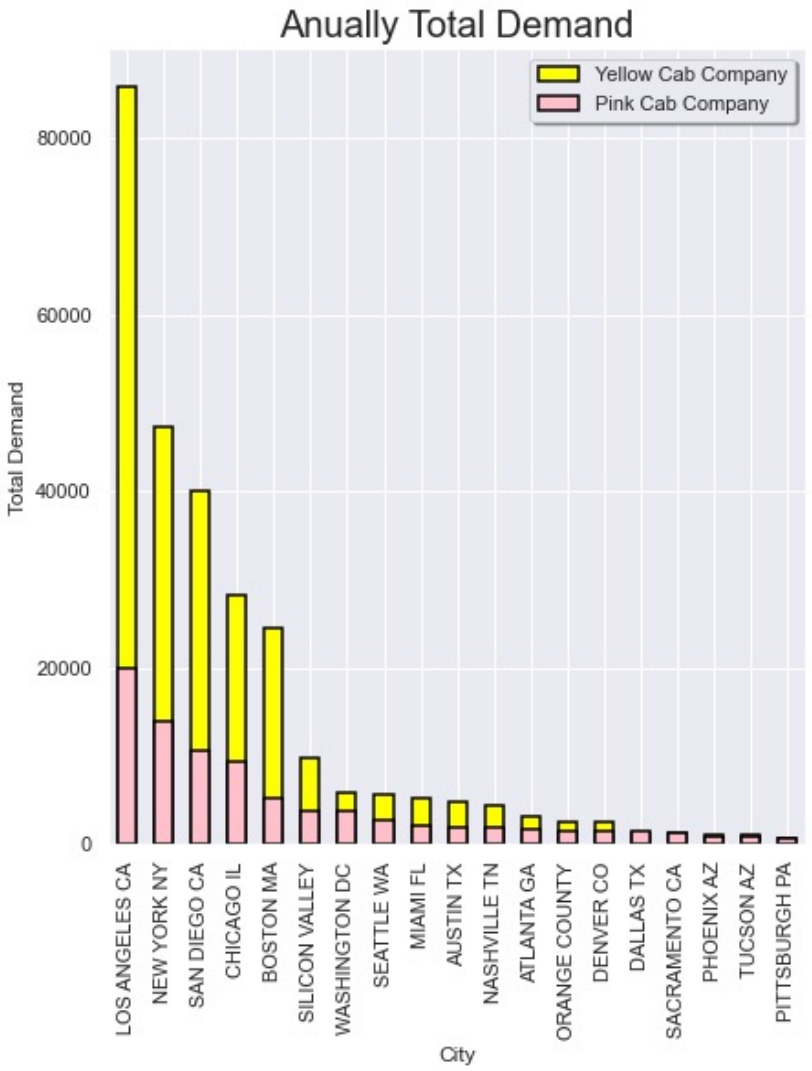


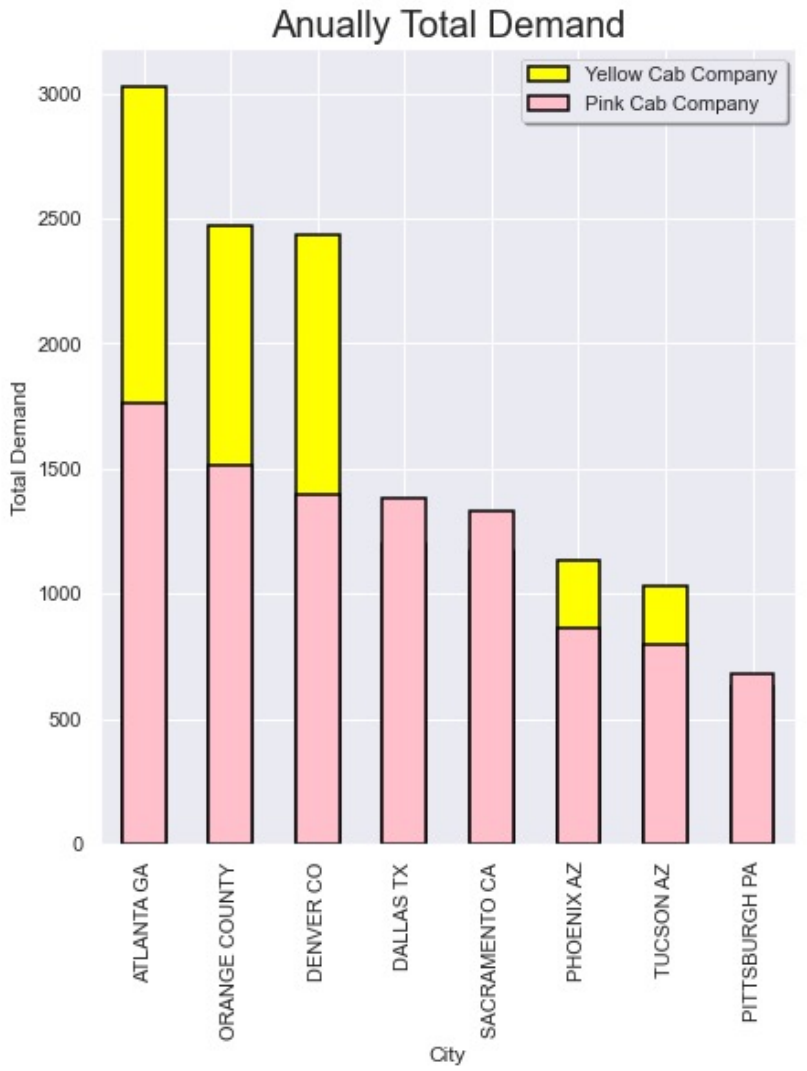


The demand for both companies follows a quite similar cycle. We can apreciate here more detailed.

5) Demand Analysis

Data Glacier
Your Deep Learning Partner





Demand distribution is similar to profit, but there is one point I want to highlight.

There are some cities where YC is almost not present. Keep this in mind because we will talk about it later.

Data Glacier

Your Deep Learning Partner

6) PROFITABILITY – NON PROFITABLE TRAVELS

6) Profitability Analysis

Data Glacier
Your Deep Learning Partner

As mentioned, I calculated a rate of non profitable travels, and it was contrasted with two features I considered could be interesting insights: Age and City.

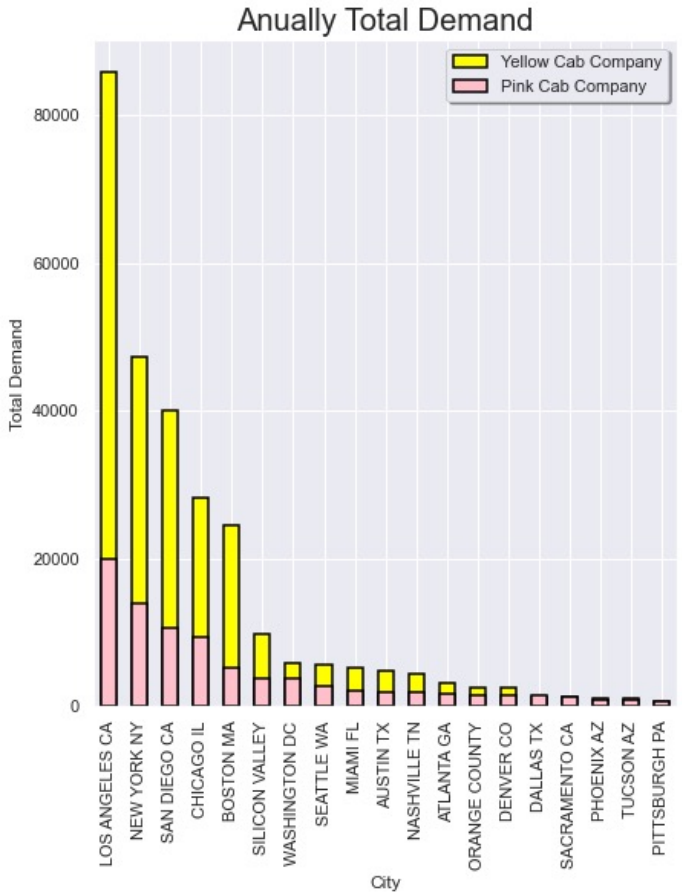
age_range	Transaction ID	total_travels	perc_non_prof
(17, 25]	6380	93344	0.07
(25, 40]	11760	169794	0.07
(40, 60]	5310	77343	0.07
(60, 66]	1369	18911	0.07

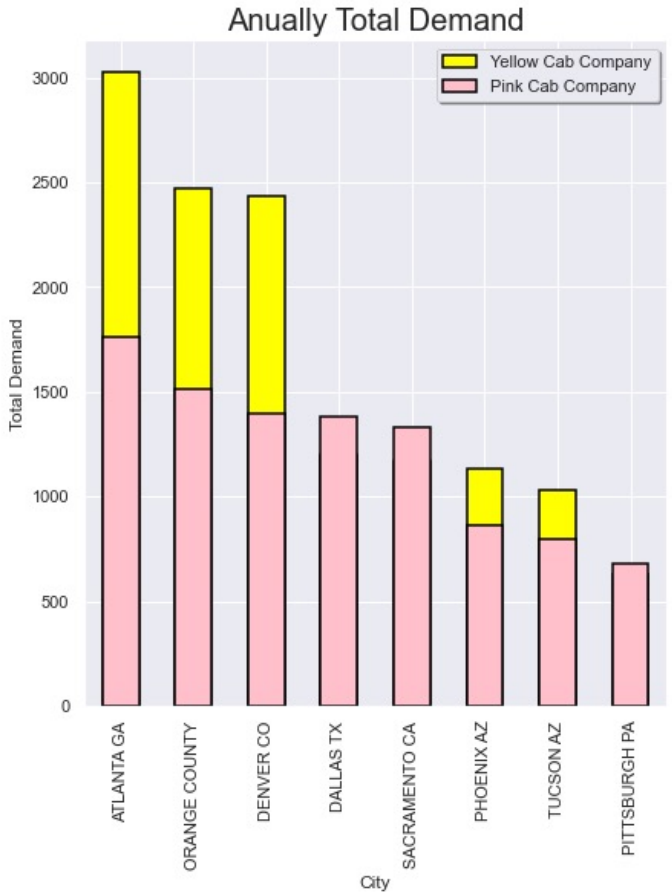
Age regarding, I divided population into the ranges shown above, but there were similar non profitable travel rate along the ranges, so the analysis was finished there.

6) Profitability Analysis

Data Glacier
Your Deep Learning Partner

Talking about non profitable travel rate among cities, 6 of the 8 cities we mentioned before where YC has the lowest marketshare, also are the least profitable cities (with highest non profitable travel rate).






```
The 8 cities with the highest non-profitable travel rates :
City
PITTSBURGH PA    0.26
NASHVILLE TN     0.23
SACRAMENTO CA    0.22
TUCSON AZ        0.14
CHICAGO IL       0.14
DENVER CO        0.13
BOSTON MA        0.13
PHOENIX AZ       0.12
Name: perc_non_prof, dtype: float64
The 8 cities with the lowest Yellow Cab market share :
AUSTIN TX         3028
ORANGE COUNTY     2469
DENVER CO         2431
PHOENIX AZ        1200
NASHVILLE TN      1169
TUCSON AZ         1132
SACRAMENTO CA     1033
PITTSBURGH PA      631
```

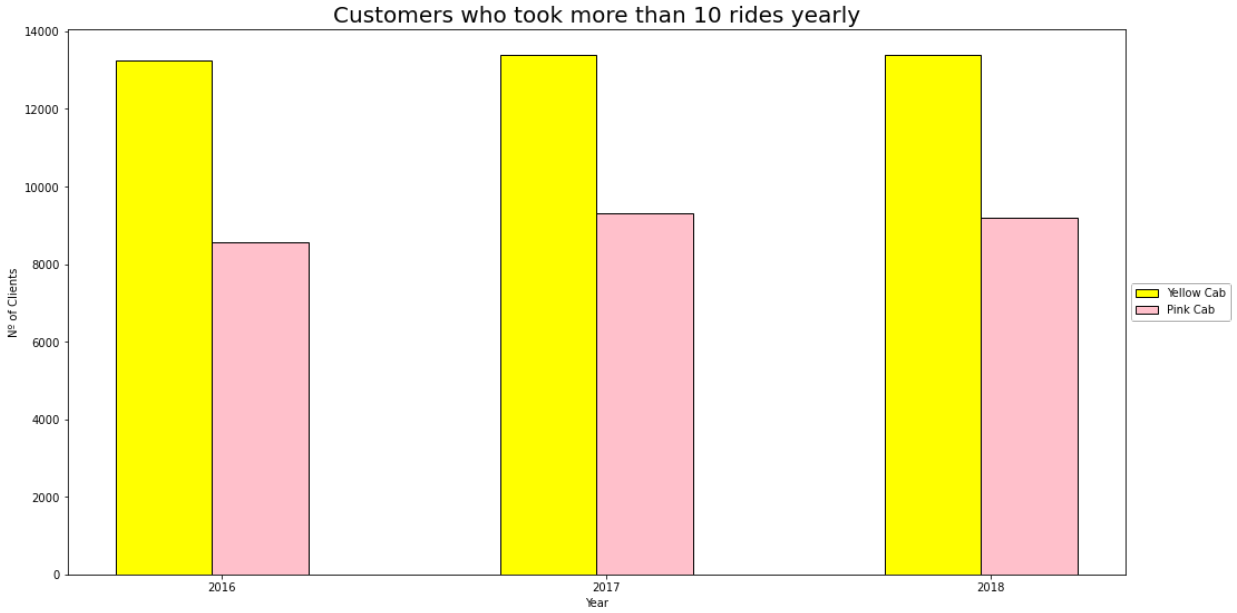
Data Glacier

Your Deep Learning Partner

7) CUSTOMER RETENTION

7) Customer Retention

Data Glacier
Your Deep Learning Partner



Customer retention analysis was in two dimensions. Customers taking more than 10 travels per year and customers taking more than 10 travels per month.

YC and PC customer retention yearly is similar, near 50% of all customers.

The main difference in monthly because YC is 15% (average) higher than PC.

Data Glacier

Your Deep Learning Partner

8) CONCLUSION

7) Conclusion

Data Glacier
Your Deep Learning Partner

Profits:
Yellow Cab profits are higher in every dimension analysed. Total profits, profits per travel and profits per km.

Demand:
Yellow Cab demand is higher in every dimension as well (consider the the few cities where Pink Cab has higher market share that we mentioned).

Seasonability and the variation along the quarters/years follows a similar cycle, but the variations and changes of demand in Yellow Cab are more abrupt.

Customer retention:
Both companies are similar at yearly levels, but opening the information in monthly dimension, YC retention is greater.

Final state:
Considering all the information we had, and the analysis developed over the weeks, I am allow to say that it would be recommended to invest on Yellow Cab Company.

Thank You