Data Glacier

Your Deep Learning Partner

Exploratory Data Analysis

Cab Investment EDA

06 Sept. 2021

Table of Contents

Data Glacier
Your Deep Learning Partner

Problem Statement

Data Glacier
Your Deep Learning Partner

Problem Statement

Approach

EDA

EDA Summary

Recommendations

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 and as per their Go-to-Market(G2M) strategy they want to understand the market before taking final decision.

VS

Approach

Problem Statement

Approach

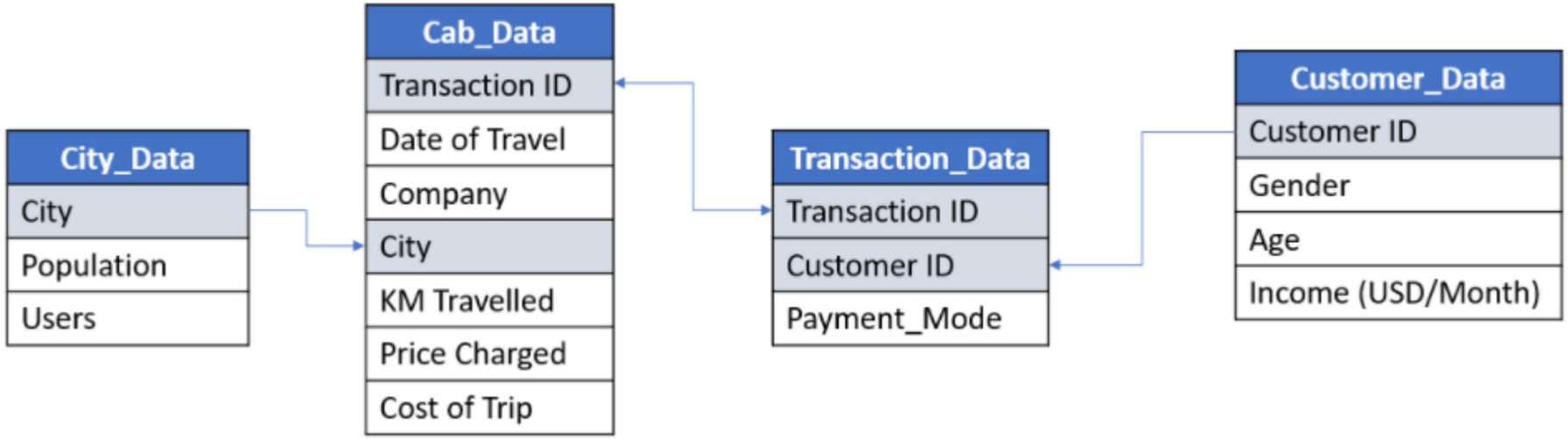
EDA

EDA Summary

Recommendations

Data Glacier
Your Deep Learning Partner

We were provided 4 datasets that we merge together like in the following picture to get a Master Dataset.



- 24 features (calculated ones included) :
  - 5 objects
  - 1 datetime
  - 18 numerical
- 359,392 points
- Timeframe of the data: 2016-01-02 to 2018-12-31

EDA

Problem Statement

Approach

EDA

EDA Summary

Recommendations

Data Glacier

Your Deep Learning Partner

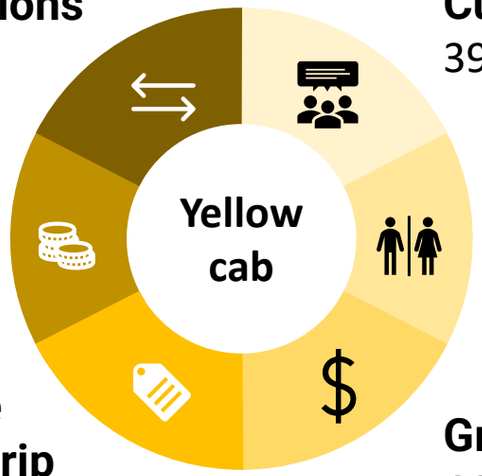
RESULTS FROM 2016 TO 2019

EDA : Global informations

Transactions
274,681

Customers
39,896

Margin
30.13 %



Genders
Male : 53.90%
Female : 46.10%

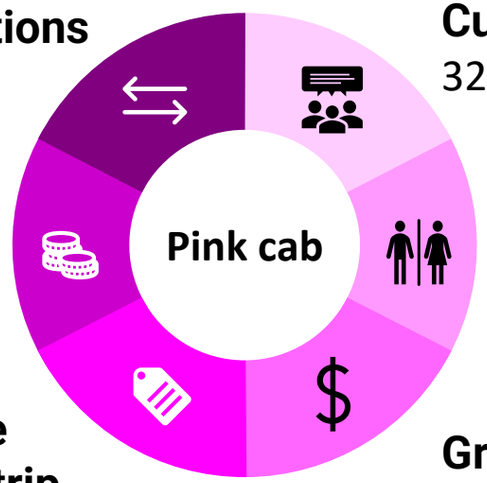
Average cost of trip
296 $

Gross profit
44,020,373 $

Transactions
84,711

Customers
32,230

Margin
17.38 %

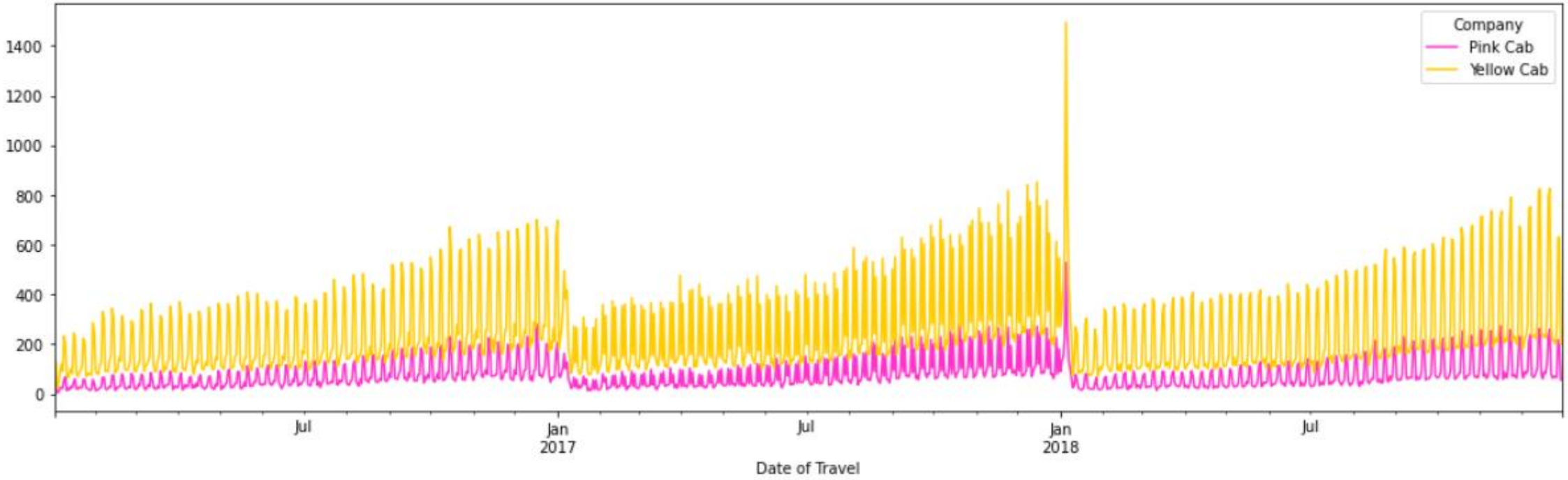


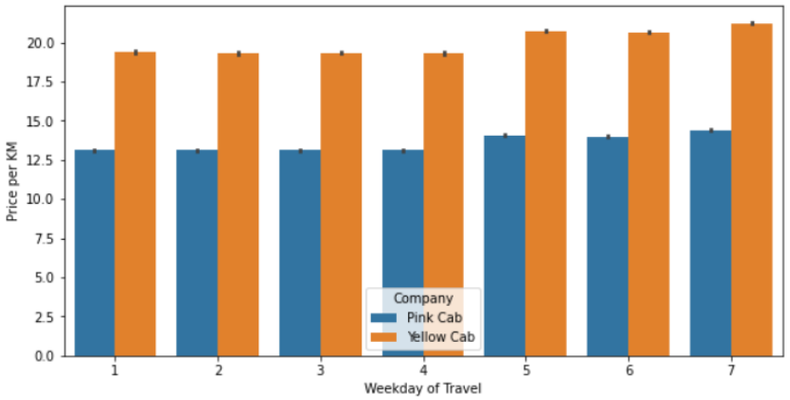
Genders
Male : 54.16%
Female : 45.84%

Average cost of trip
246 $

Gross profit
5,307,328 $

Rides seasonality





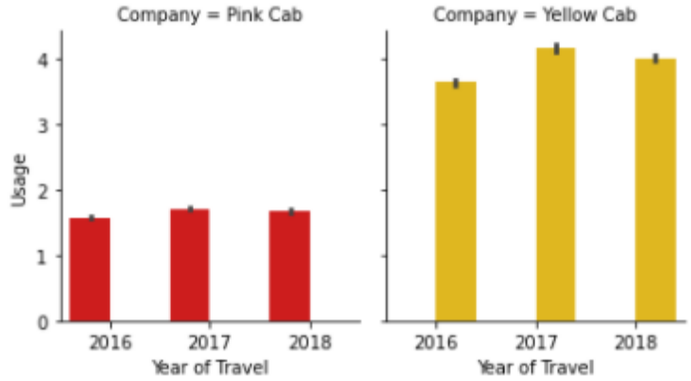
The distribution of the transactions according to the *date of travel* looks like a *time serie* with these properties :

- **Periodicity :** 1 year
- **Slope :** Positive
- **Noise :** High

There are more rides on the last 3 days of the week dans the other days. Pricing for these days are also higher than those in working days.

Year	Company	Customers	Common customers	% Common Customers
2016	Pink Cab	16,661	11,446	68.70
2016	Yellow Cab	25,937	11,446	44.13
2017	Pink Cab	18,643	13,014	69.81
2017	Yellow Cab	27,789	13,014	46.83
2018	Pink Cab	18,400	12,932	70.28
2018	Yellow Cab	27,470	12,932	47.08
All	Pink Cab	32,330	26,078	80.66
All	Yellow Cab	39,896	26,078	65.36

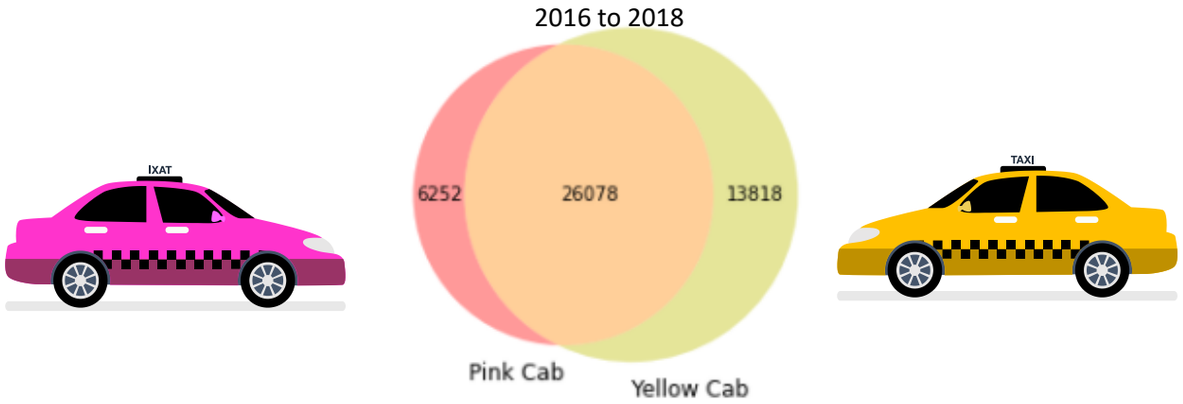
People use Yellow Cabs twice more than the Pink Cab ones

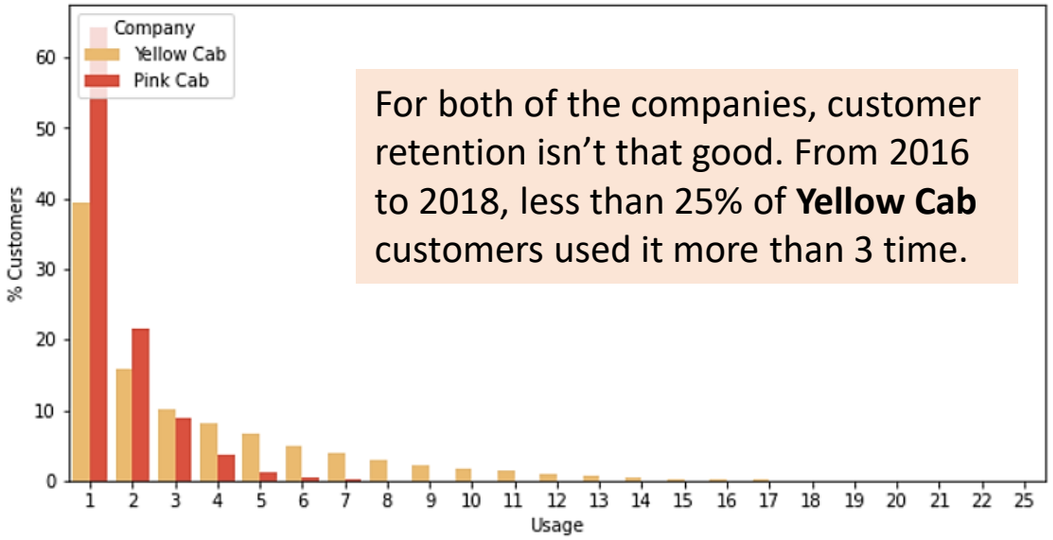


60% of Yellow Cab users took it at least twice over the last 3 years

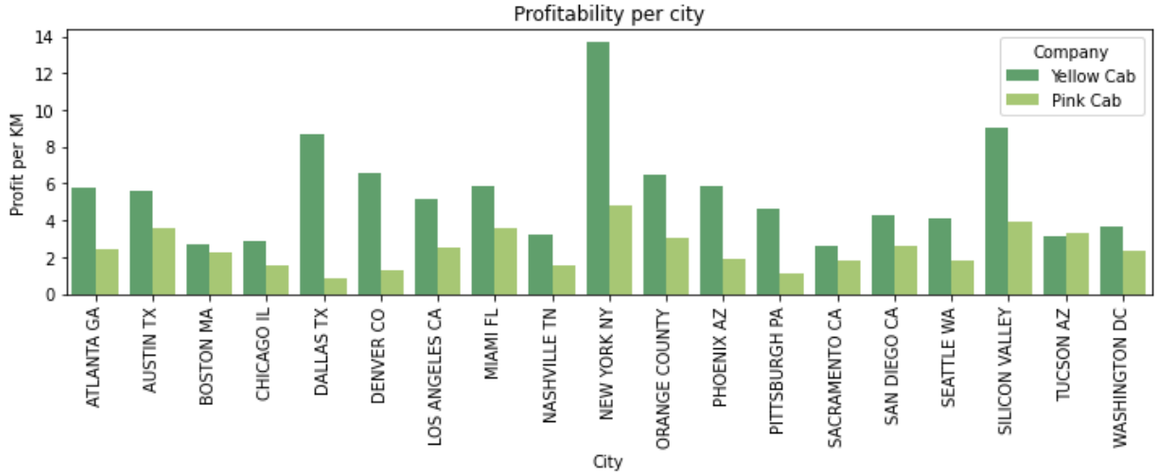
Only 36% of Pink Cab users took it at least twice

80% of Pink Cab users also took Yellow cab

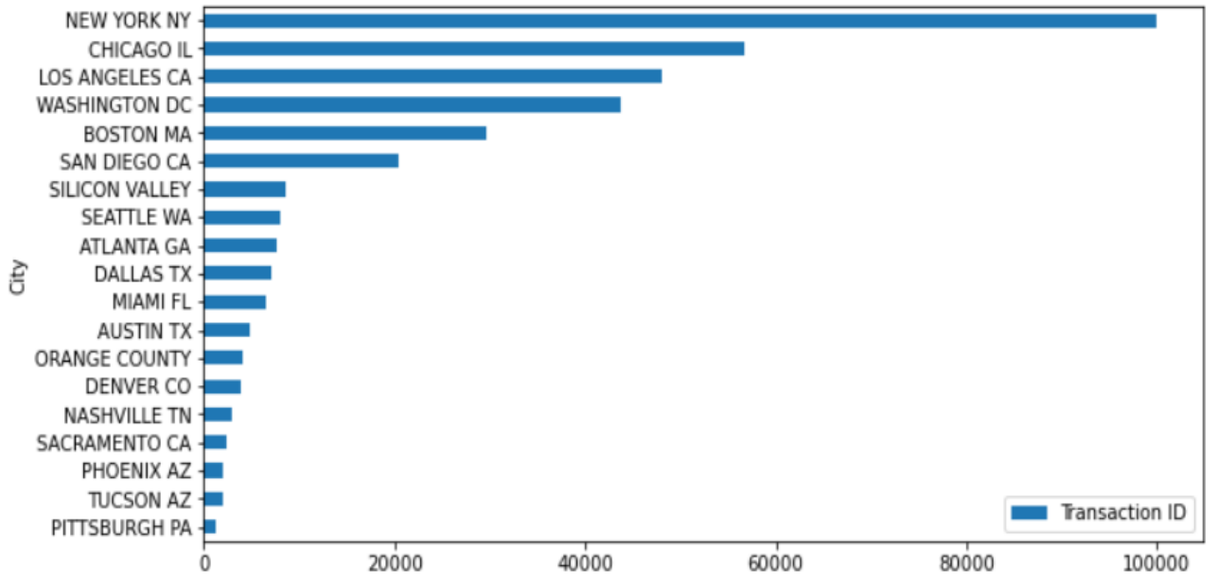




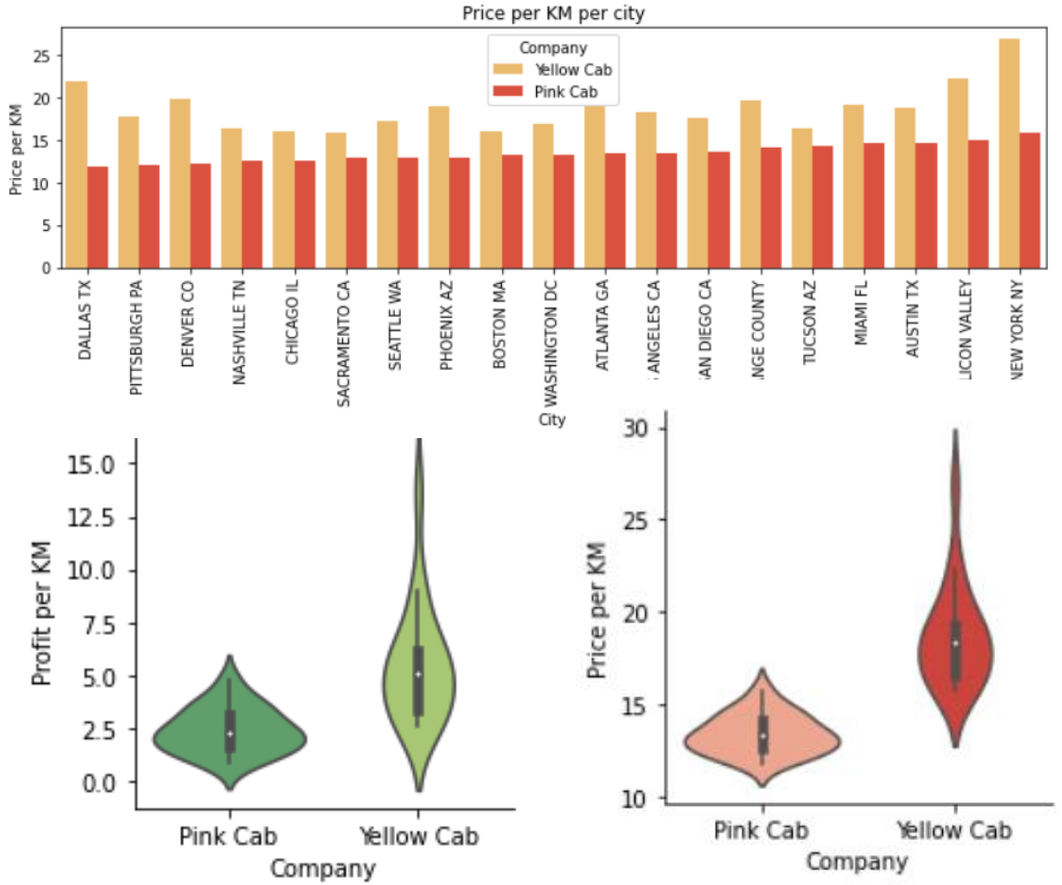
For both of the companies, customer retention isn't that good. From 2016 to 2018, less than 25% of **Yellow Cab** customers used it more than 3 time.



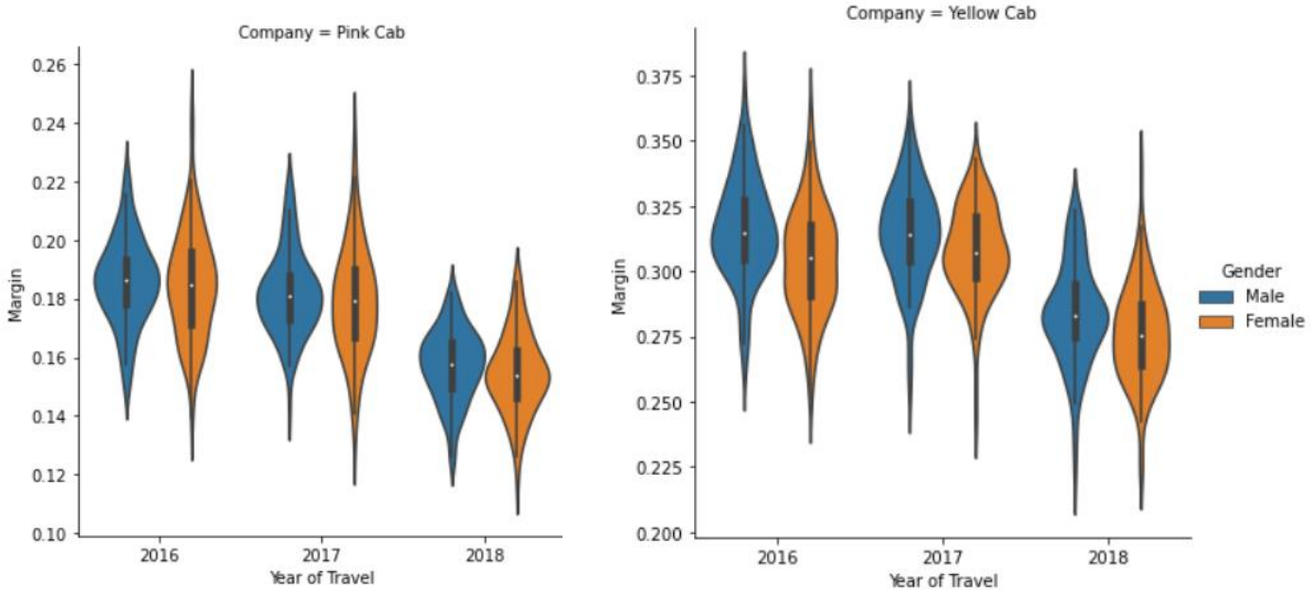
Cities like **Chicago**, **Los Angeles** and **Washington** tend to have high number of rides but they are not that profitable per KM



- It seems that Pink Cab (on the contrary of Yellow cab) doesn't have any particular pricing regarding of the city.
- Pricing for **Pink Cab** (on the contrary of **Yellow cab**) doesn't change a lot regarding the City.
- Price/KM is highly correlated with Profit/KM



The average Profit/KM is of **2.5$** and **5$** respectively for Pink then Yellow Cabs



- **Pink Cab** company doesn't have a difference in margin regarding the gender of its customer (Hypothesis 5 : Student Test on notebook)
- **Yellow Cab male** users generate more margin than **female** users.

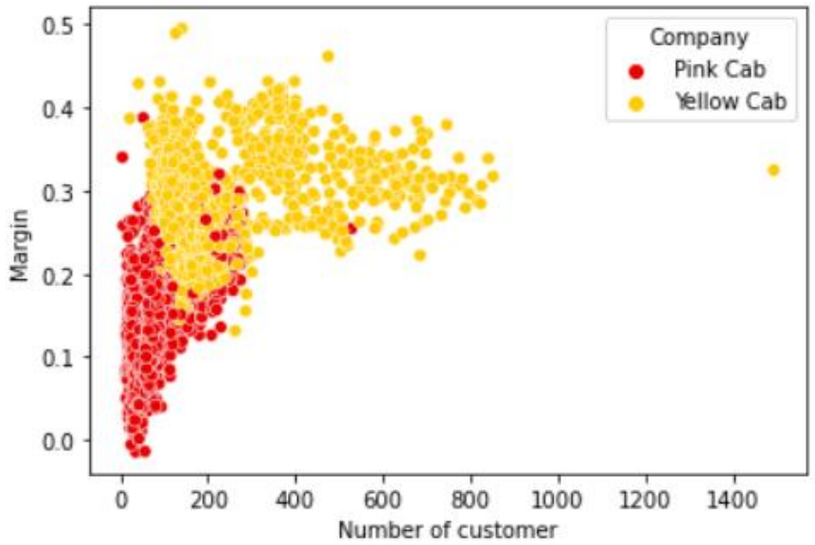
Globally, **Yellow Cab** margin decreases with the year : The margin is negatively correlated (**-0.51**) with the year (Same for **Pink Cab :**-0.6)

The correlation between the **Margin** and the **Number of customers** depends on the company :

- **Pink Cab :** moderated correlation ~48%
- **Yellow Cab :** low correlation ~18%

The more the **Pink Cab** has some customers, the more its margin increases by a half.

Margin decreases other the years.



EDA Summary

Problem Statement

Approach

EDA

EDA Summary

Recommendations

Data Glacier

Your Deep Learning Partner

After having explored the dataset, here are the declining conclusions we have :

1.Profitability : In term of profitability, Yellow Cab is 8 times more profitable than Pink Cab

- On the last 3 years, **Pink Cab** made *5.307.328* and **Yellow Cab**, *44.020.373*
- **NEW YORK NY** is the most profitable city (*27.962.555*) for both **Pink** and **Yellow Cabs**

2.Gender : There is nearly as many women as men | *~46%* of women and *~54%* of men . Company and Gender are independants.

3.Pricing : Yellow Cab proposes a very flexible **Pricing per KM** depending on lots of features like the City, the Gender; which is not the case of **Pink Cab** which doesn't.
From Friday to Sunday, the cab services are more expensive than the other days.
Pink Cab is the least expensive company among both companies (~3 times less on workdays and 2.5 times from Friday to Sunday)

4.Loyalty : Having high or low revenue doesn't inpact the frequency of usage of the cabs. **Yellow Cab** is the most used company.
But there is a low global customer retention :

5.Margin : The more **Pink Cab** has some users, the more it increases its margin. It's not the case for **Yellow Cab.**
Both of the companies lost some margin other the years.

6.Periodicity : The distribution of the transactions according to the *date of travel* looks like a *time serie* with these properties :

- **Periodicity :** 1 year
- **Slope :** Positive
- **Noise :** High

Recommendations

Problem Statement

Approach

EDA

EDA Summary

Recommendations

Data Glacier
Your Deep Learning Partner

The market is clearly in favor of the **Yellow Cab** company, even if during the last year, its *gross profit* has gotten a little decrease. *XYZ* should invest in **Yellow Cab** due to its market implantation.

If *XYZ*, wants to invest in a company with great potential, it look to **Pink Cab.**

Thank You