Exploratory Data Analysis

G2M Insights for investing in cab company.

June 14, 2022

Agenda

- Executive Summary
- Problem Statement
- Approach
- EDA
- EDA Summary
- Recommendations

Data Glacier
Your Deep Learning Partner

Executive Summary

Data Glacier
Your Deep Learning Partner

- ✓ This report aimed to provide recommendations for the client, XYZ investment firm, so that they can invest in the best cab company, i.e., Yellow cab or Pink cab.
- ✓ The analysis includes EDA, Hypothesis tests, insights, and finally recommendations.

Recommendation

I recommend Yellow cab for investment, since it outperformed Pink cab in:

- ✓ Customer coverage
- ✓ Customer Loyalty
- ✓ Customer return
- ✓ Income-Wise Reach

Problem Statement

XYZ is a private equity 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.

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

Data sets

Below are the list of datasets which are provided for the analysis:

- **Cab_Data.csv –** this file includes details of transaction for 2 cab companies
- **Customer_ID.csv** – this is a mapping table that contains a unique identifier which links the customer's demographic details
- **Transaction_ID.csv –** this is a mapping table that contains transaction to customer mapping and payment mode
- **City.csv –** this file contains list of US cities, their population and number of cab user
- **USHoliday.csv –** this file contains list of US Holidays.

Time period of data is from 31/01/2016 to **31/12/2018.**

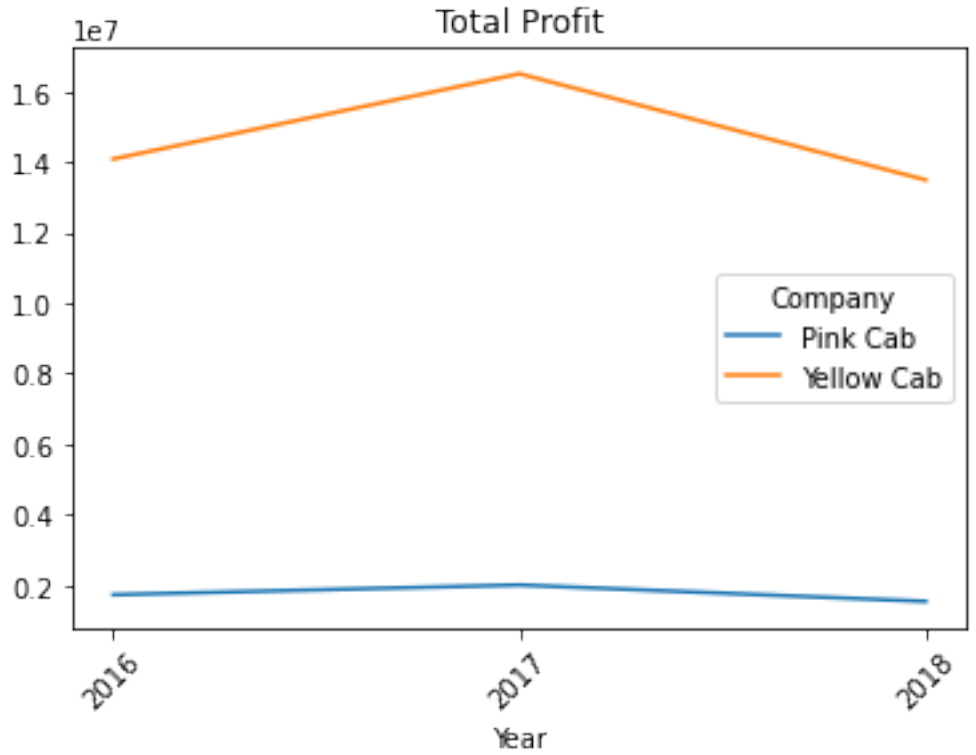
Data Exploration

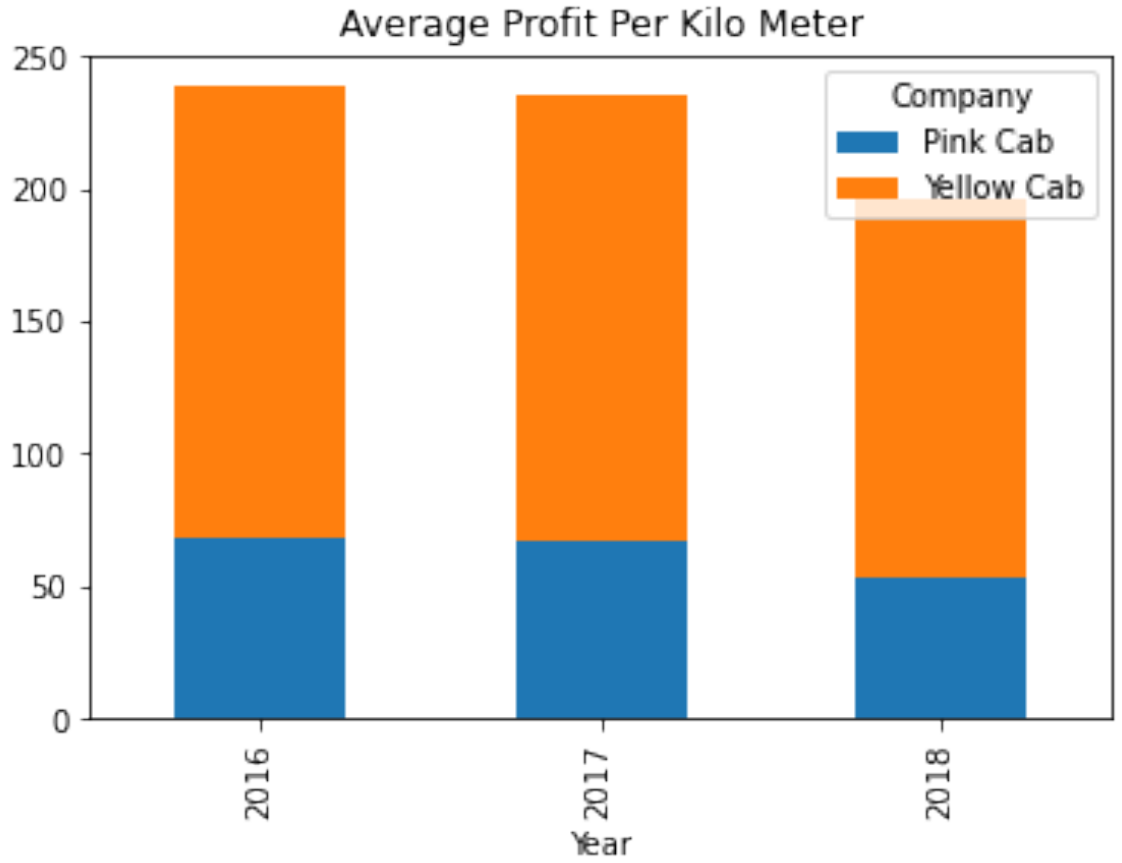
Assumption:

- Outliers are present in “Price Charged” feature, which belong to “Premium” trip, so I did not treat them.
- Profit of rides are calculated keeping other factors constant and only “Price Charged” and “Cost of Trip” features used to calculate profit.
- Negative profit are considered as loss.

Profit analysis

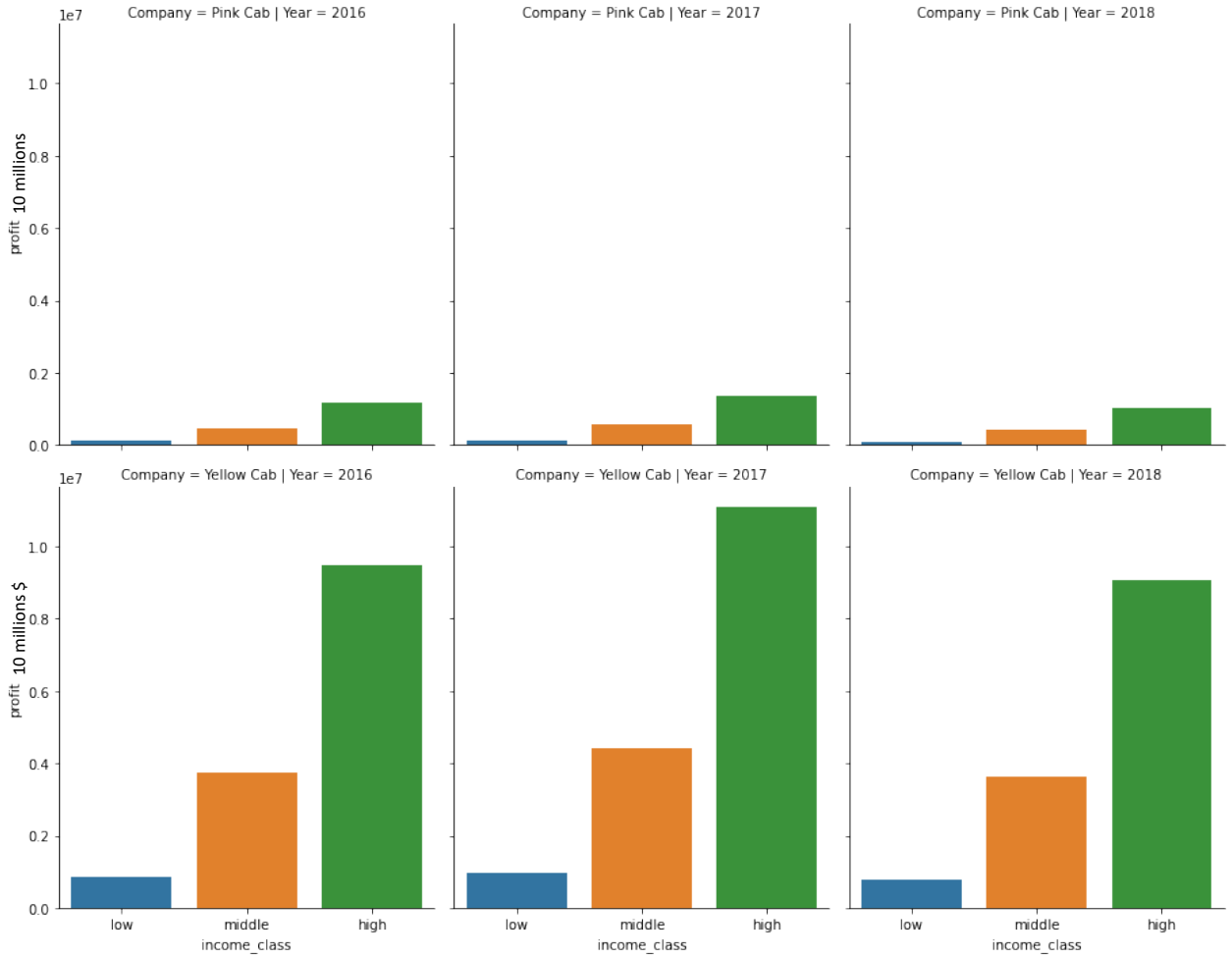
Data Glacier





Company	Profit	Total Rides	Profit per Rides
Pink Cab	5,308,147	84,739	62.64
Yellow Cab	44,042,849	274,826	160.26

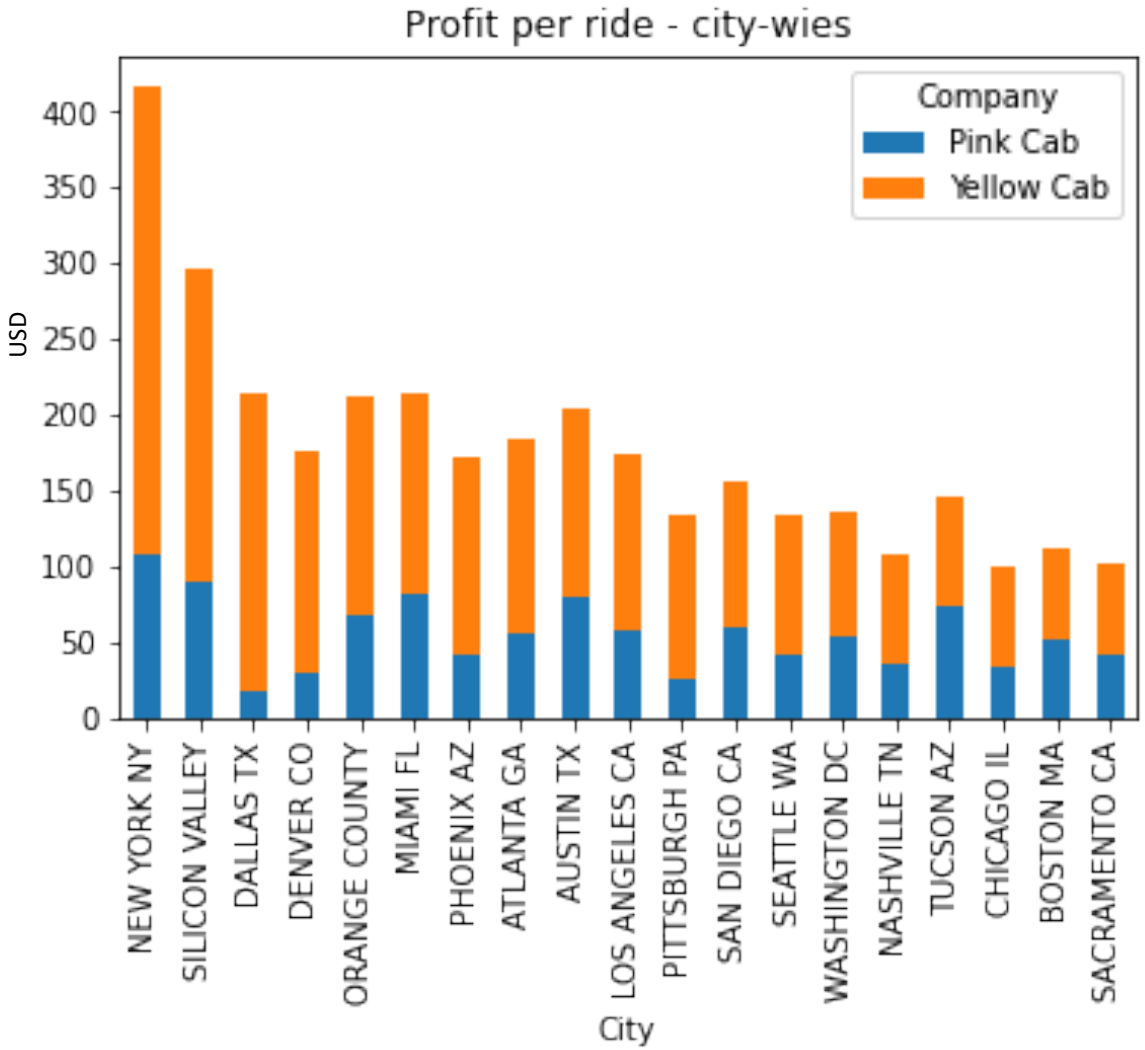
Profit and different income classes



The plot shows that high income family spend more money on cab company. Moreover, Yellow Cab benefits from high income customers more than its rival.

Pink Cab could not attract low income customers, although it has a number of them

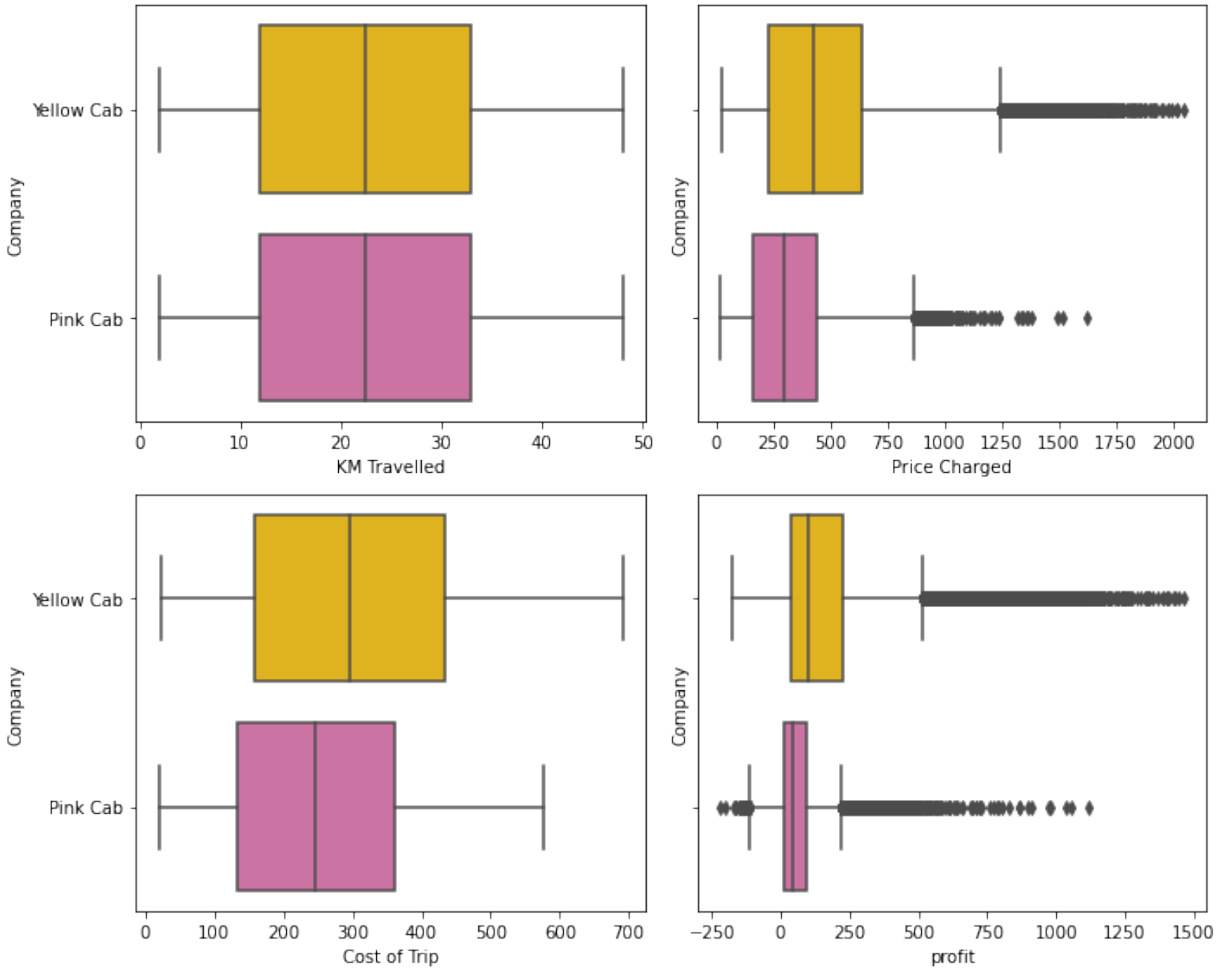
Profit per Ride base Analysis Cab wise



Due to the high number of "Premium" trips that Yellow Cab did, it was able to gain more profit per trip in any city.

Based on Hypothesis test: Premium and Profit are not independent (reject H0).

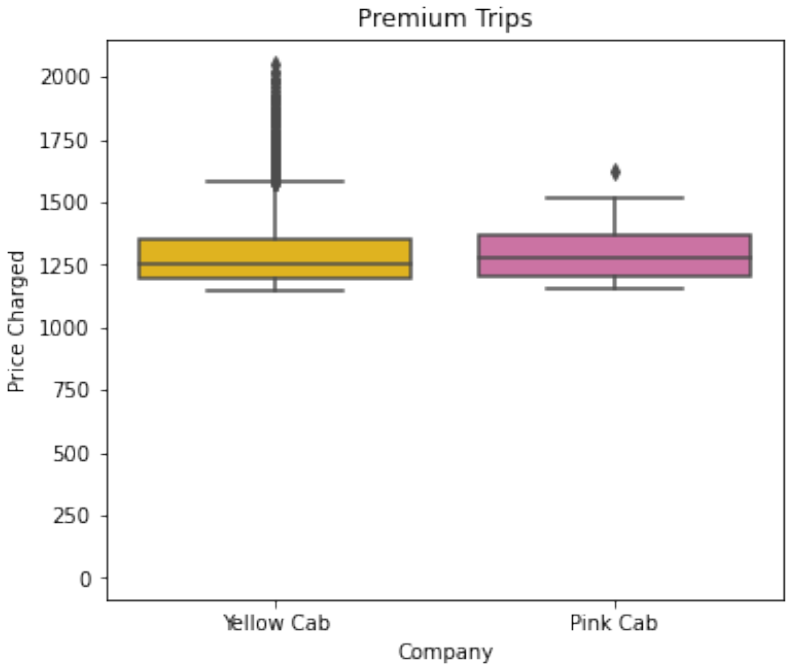
Profit’s factors

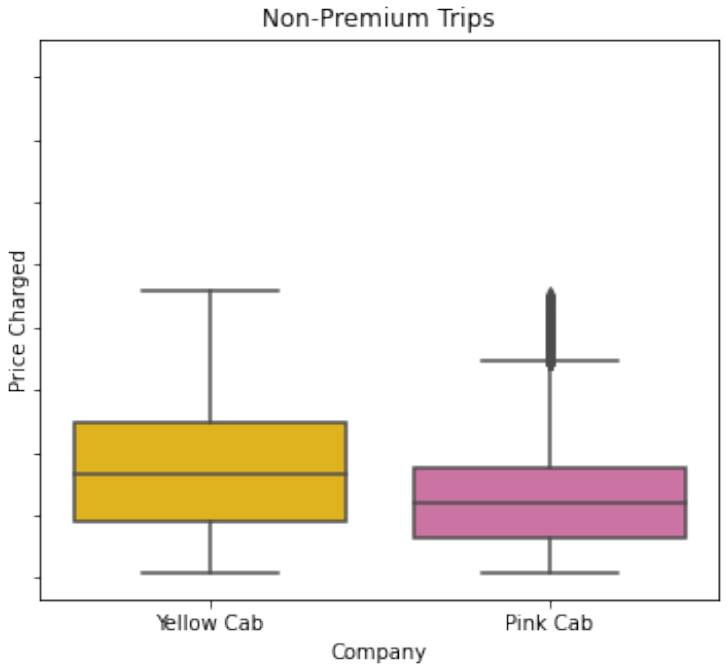


It has been concluded:

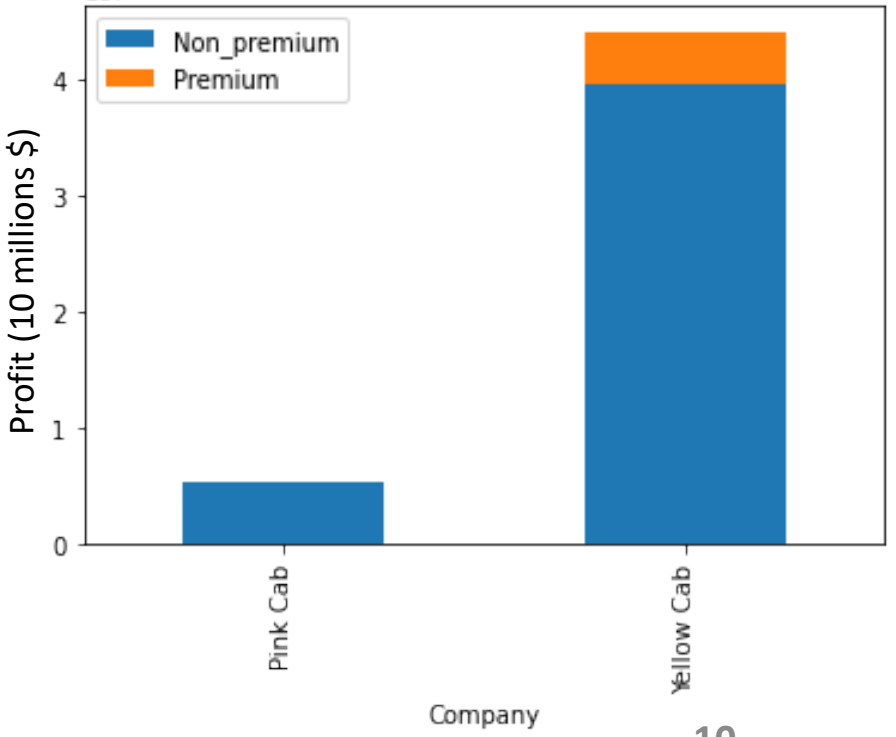
- Both companies have loss in their record, which should be investigated, according to the Q1 and outlier of profit boxplot.
- The boxplot of Charges (amount of payment) is similar to that of profit, which means that it is a key feature in profit. It can be assumed that they charge the customer with additional costs like a luxury cars or/and special services, including children car seats and so on, which can be named as "Premium".

Premium



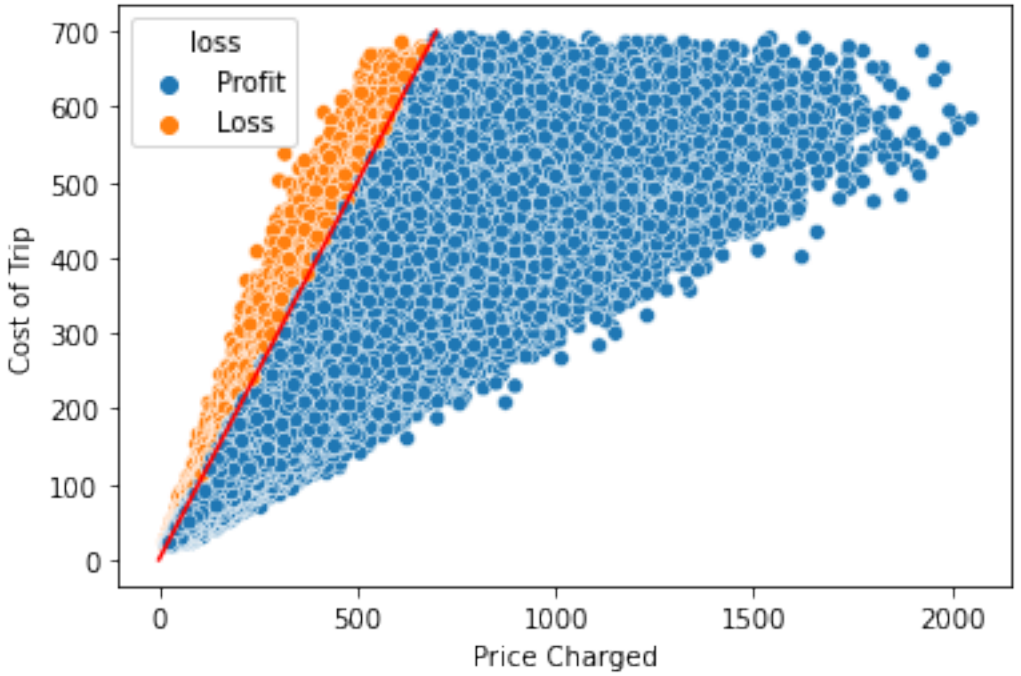


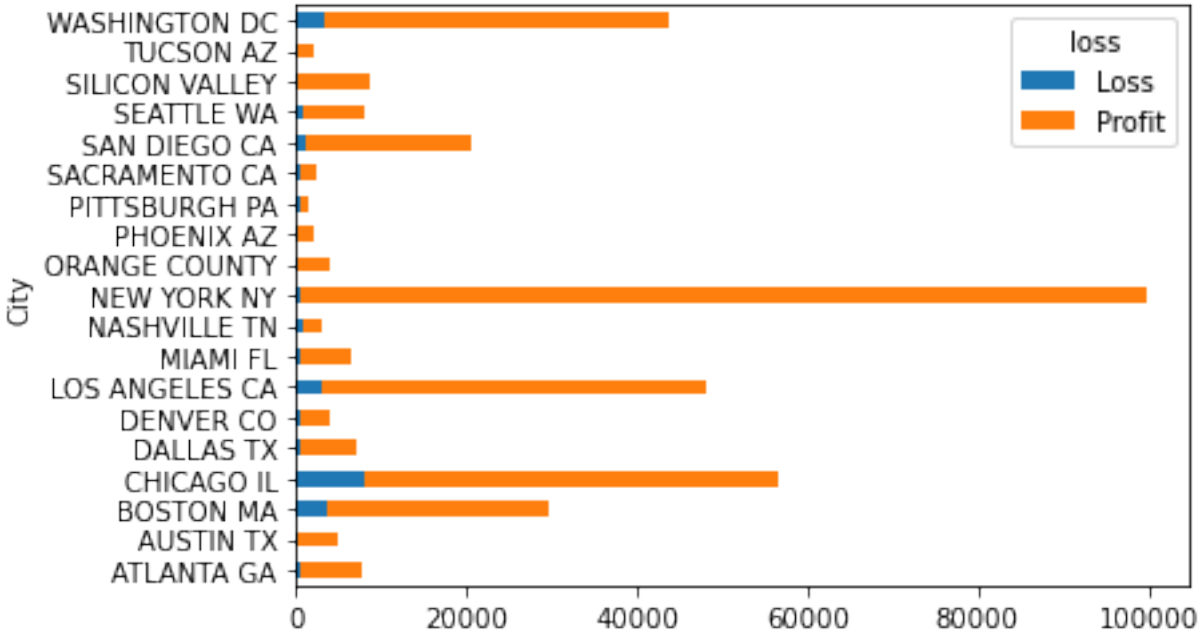
	Profit (USD)	
	Pink Cab	Yellow Cab
Non_Premium	5,292,701	39,547,550
Premium	15,446	4,495,303



"Premium" is calculated by Price Charged that are greater than Q3 + 1.5 * IQR, which is 1149.48.

Loss Analysis





- ✓ 70% of loss happens in three cities: PITTSBURGH PA, NASHVILLE TN, SACRAMENTO CA.
- ✓ New York has the lowest loss fraction.
- ✓ It can be drawn a conclusion that one of the factor to adjust the "Premium" is city.

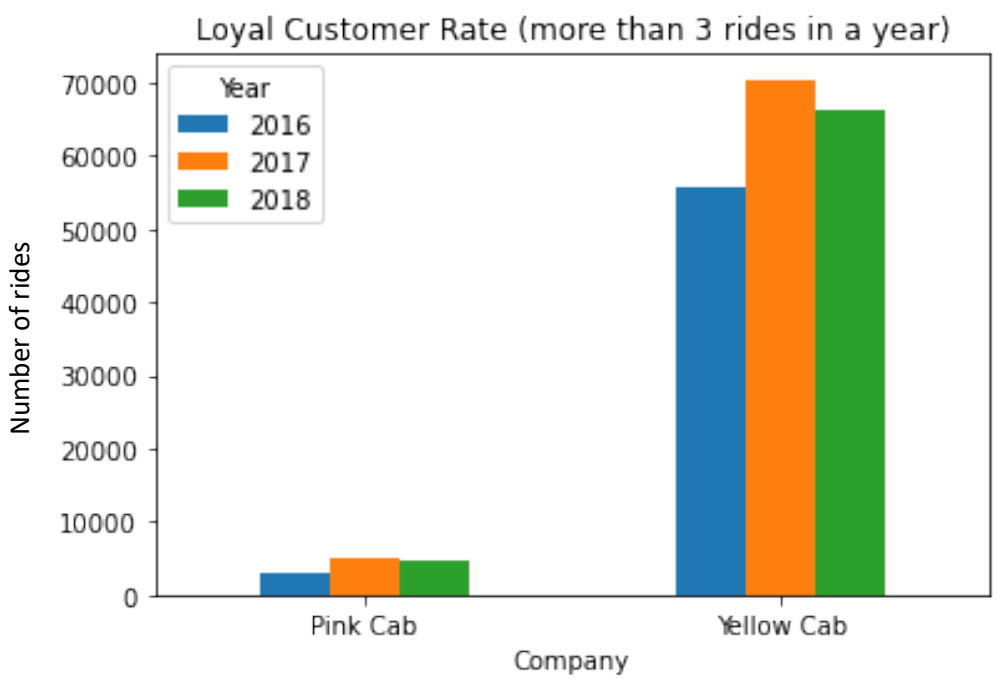
Loss VS seasonality and Holidays

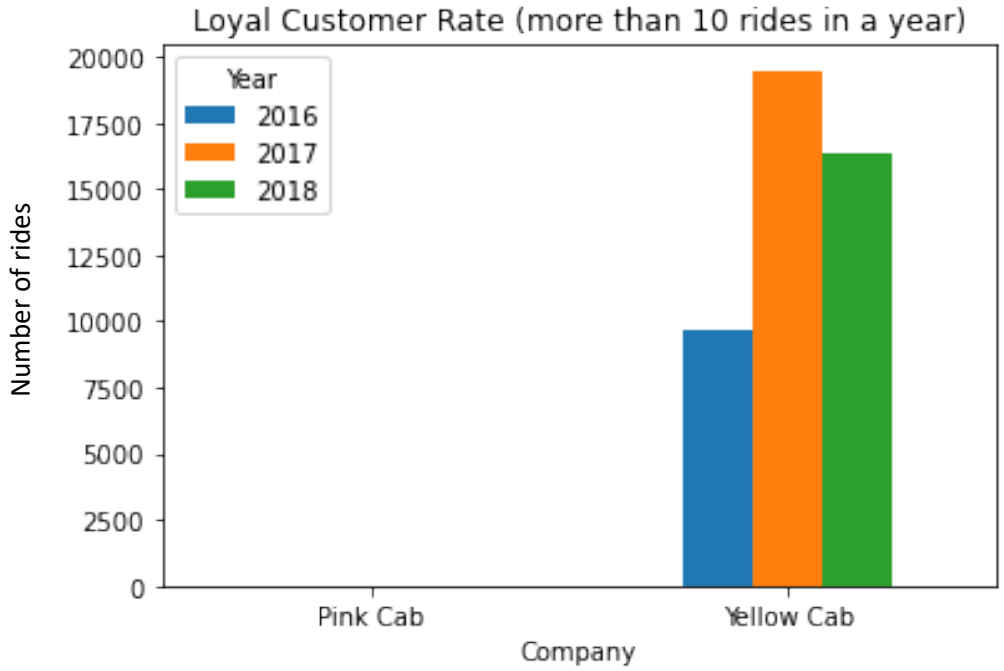
Profit/Loss	Non-Holidays	Holidays
Loss $	(23,607)	(1,237)
Profit $	318,884	15,837

The plot shows seasonality in loss. We have sharp seasonality on June, July, and October for Yellow cab Company. Also, for Pink Co., Loss increases during the June to October.

Holidays and Loss are independent (H0 holds true) by Pearson's Chi-Square Test.

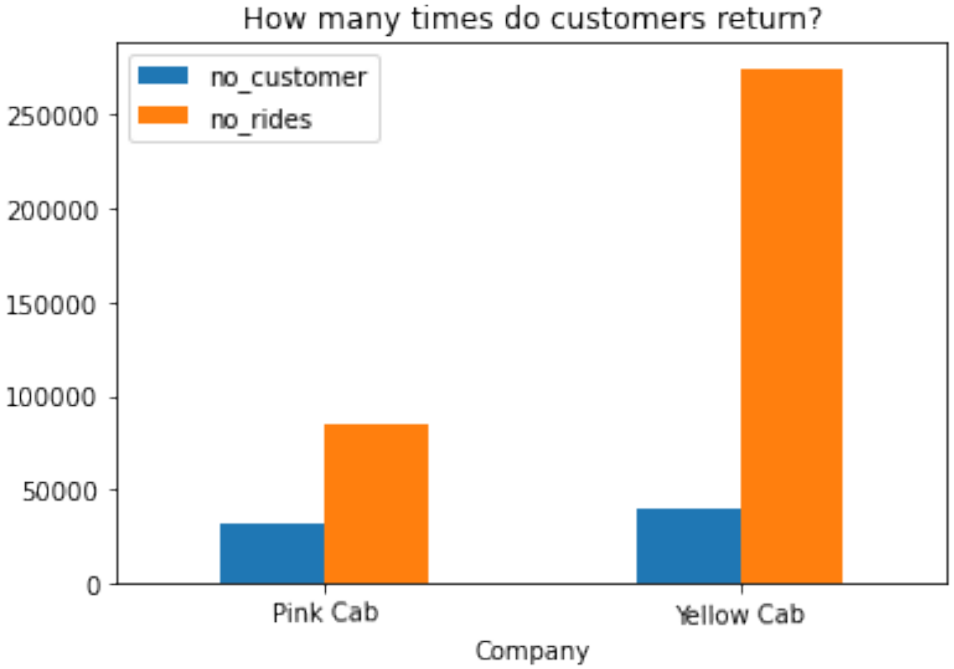
Customer Loyalty – year wise



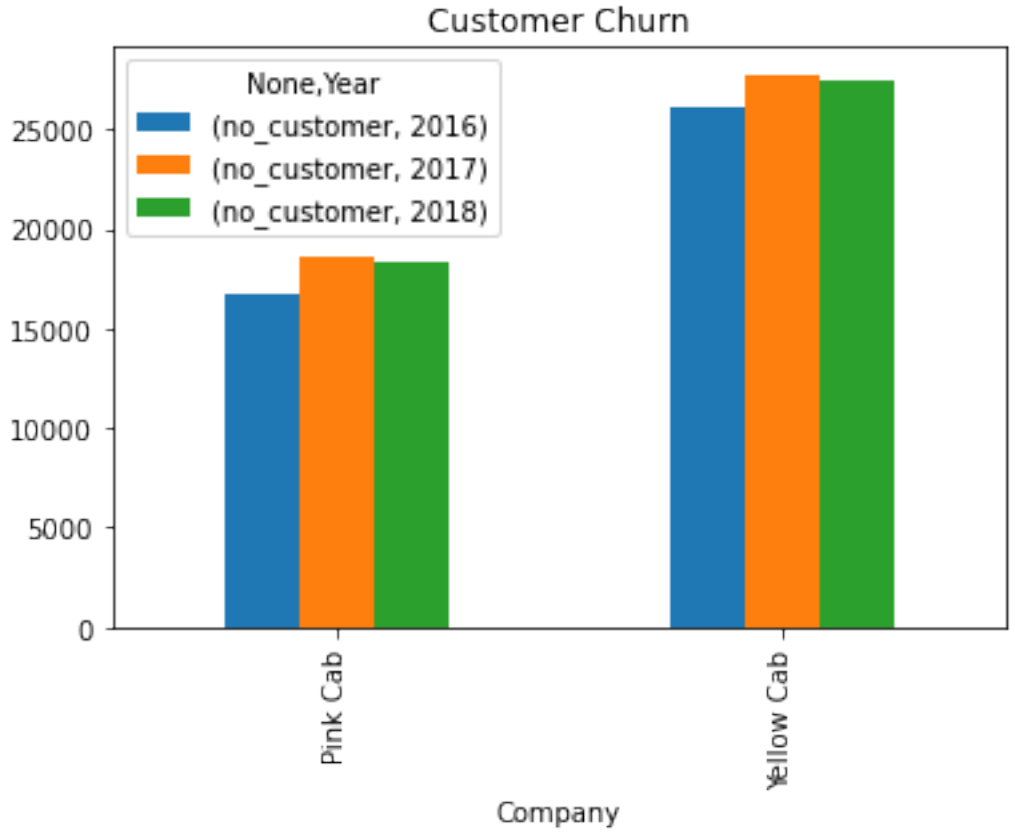


Customers of Pick cab company are not as loyal as that of Yellow cab.

Customer churn and return – year wise

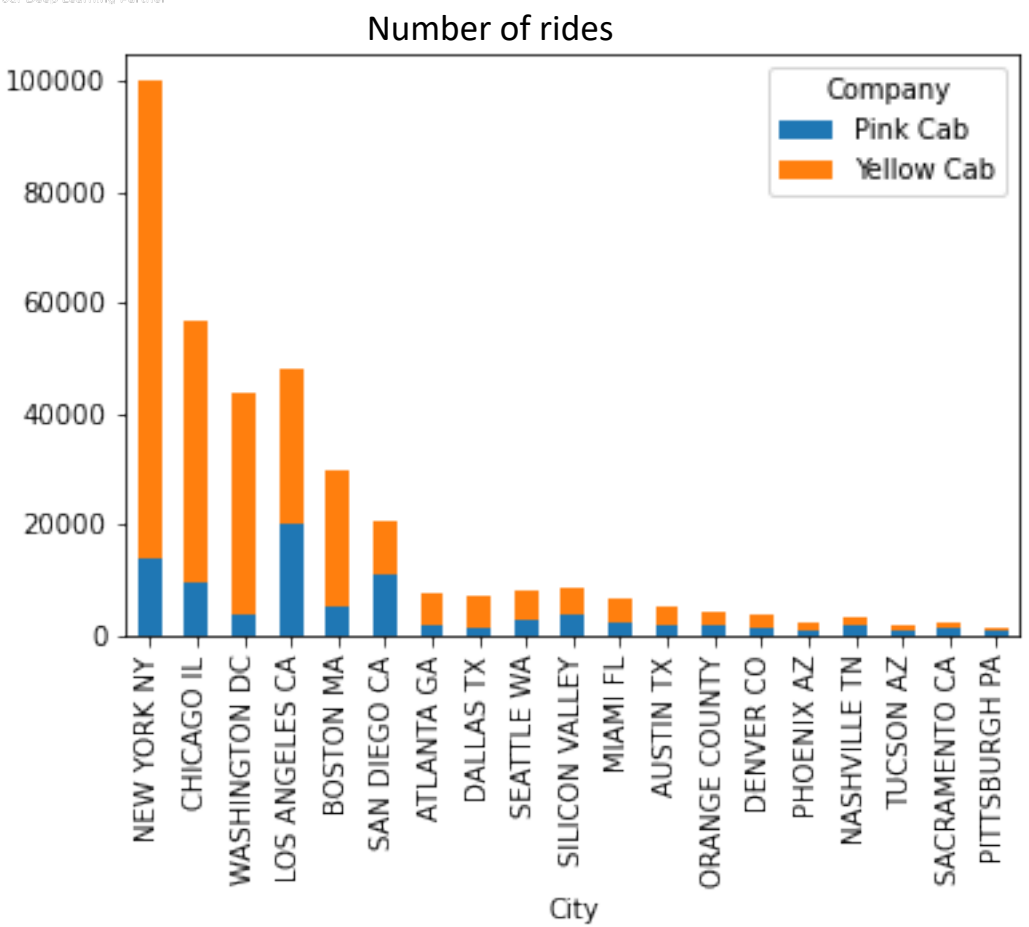


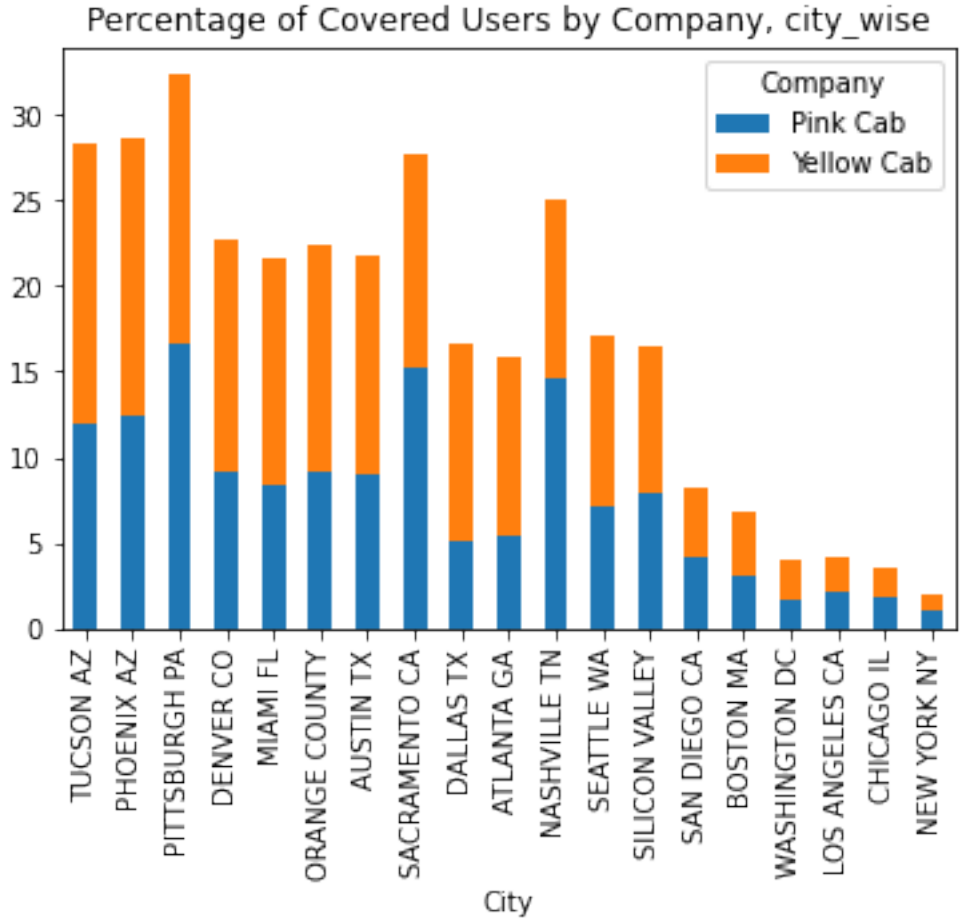
Yellow company ride per customer = 6.9
Pink company ride per customer = 2.6



Both companies, in a same way, increased customers in 2017, but could not retain some of them in 2018.

City Wise Cab Users Covered By Company





Not only does Pick cab fail to covered users, compared to Yellow cab, but it fall behind Yellow cab in benefit from it users covered.

Insights and Recommendation

Data Glacier
Your Deep Learning Partner

I have investigated in both the cab companies on following points; by doing so, I found Yellow cab better than Pink cab:

- ✓ Customers coverage : Yellow cab was able to covered more customers in 25 cities, compared to Pink can which could reach higher customers in 4 cities.
- ✓ Customer Loyalty: I evaluated customers' loyalty by repeat 3 rides and 10 rides with the same cab company, which shows that Yellow cab is doing way better than Pink cab in both these segments.
- ✓ Customer return: Customers prefer to reuse Yellow cab. Yellow cab rides per customer = 6.9, Pink cab rides per customer = 2.6
- ✓ Income wise Reach : Yellow cab offer services to 3 income classes, compared to Pink cab that is weak in absorbing low-income family.

All in all , I recommend Yellow cab for investment to XYZ company.

Thank You