



# G2M Case study

Week two deliveries  
Zhihui Chen



# Background

The Cab Industry experienced remarkable growth in the last few years, our clients, **XYZ** is going to invest in this industry. They need to understand the two cab companies that domain the cab markets to **make their investment decisions**.

Our goal is to come out insights in different dimensions to figure out the best company to invest in.



# Data Understanding - basic information

There are four datasets, their time range is from **31/01/2016** to **31/12/2018**.

**Cab\_Data** – the transaction for 2 cab companies

**Customer\_ID** – contains the customer's demographic details

**Transaction\_ID** – transaction to customer mapping and payment mode

**City** – list of US cities, their population and number of cab users

There are no null values in these datasets



# Data Understanding - Data processing

The four datasets are combined into one final dataset based on the shared column contents, which is called **new**.

Some **new columns** are generated for analysis,

Priceperkm - the price charged with unit of Kilometers the journey covers

Additionalfees - the difference between price charged and costoftrip, this might contain the fees generated by company charges, tax, tips, etc..

State - extracted the state the journey happened from city column

Year - extracted year from exist column Dateoftravel

Month - extracted month from exist column Dateoftravel

Season - generated with respective month

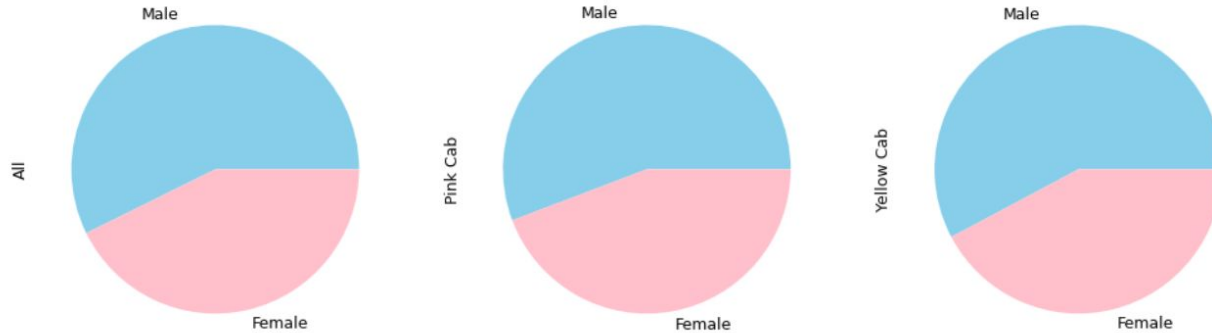
Some **new tables** for analysis,

Yellow/Pink - separate the table 'new' with respect to different companies

Yellow2016-2018/Pink2016-2018 - based on Yellow/Pink separate table into years

## Dive into datasets - User demographic info

Gender

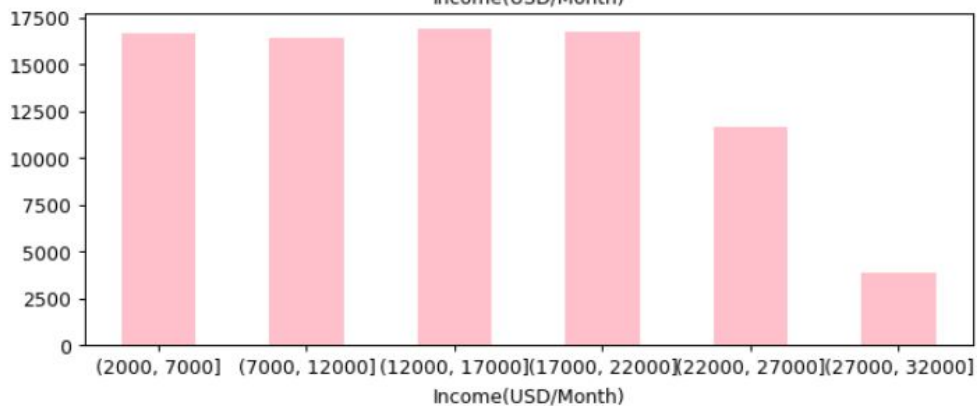
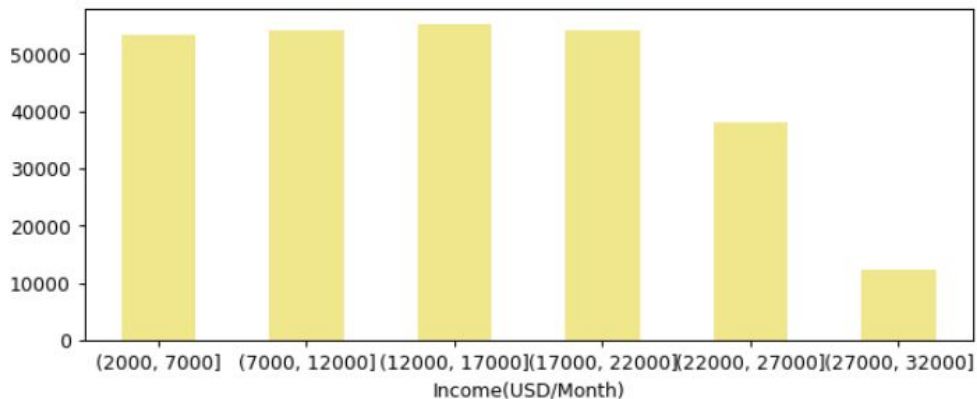


The result shows, there are more male users in these three dimensions. But in Yellow Cab, male users takes up more percentage than pink cab.

# Dive into datasets

Income in all  
companies and in  
pink/yellow cab

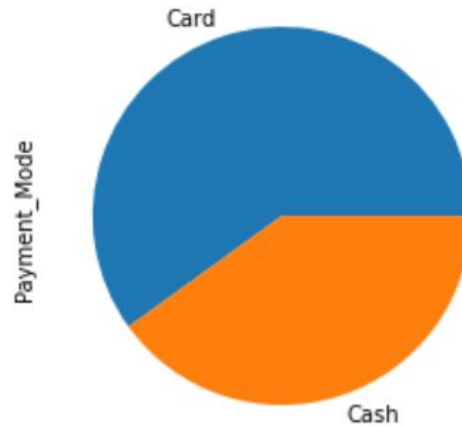
People with relatively less  
income are more likely to call a  
cab



# Dive into datasets - User demographic info

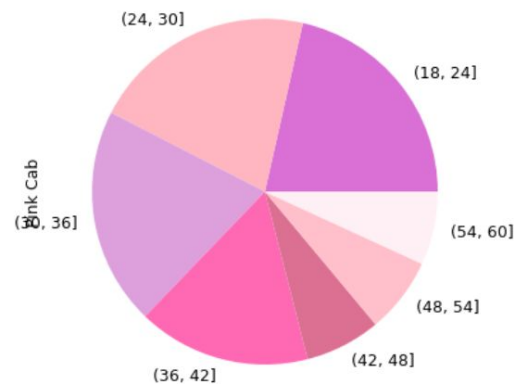
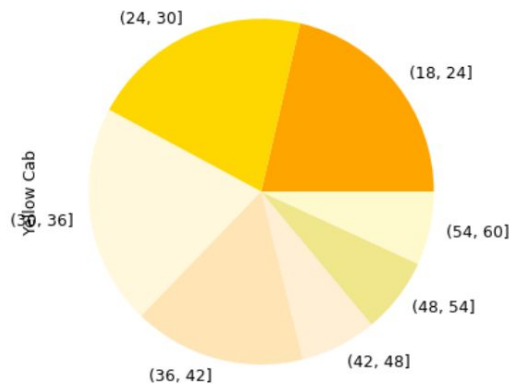
Payment methods

People generally prefer to pay by cards



## Dive into datasets - User demographic info

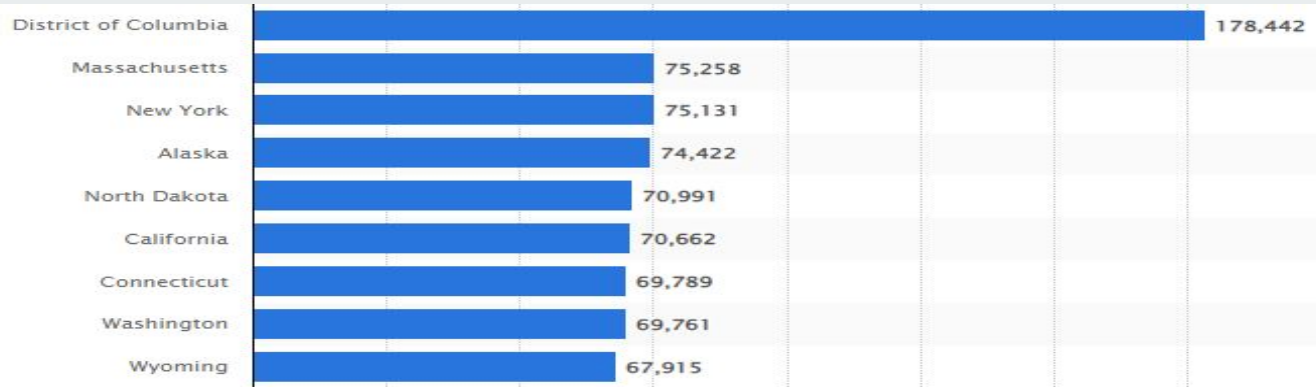
Age



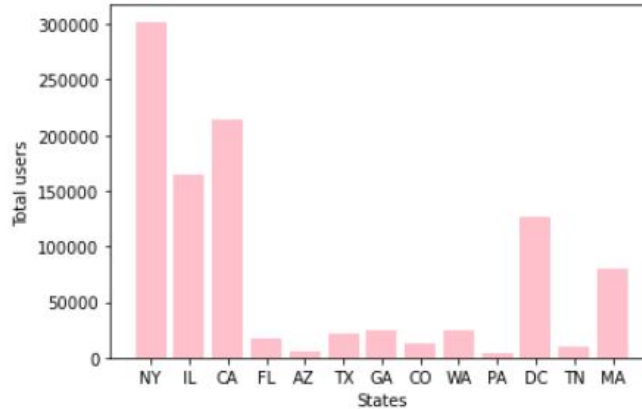
The composition of age looks similar in the two companies. And in both of the companies, users are **mainly under 36 years old**, cabs are more popular among the youngest



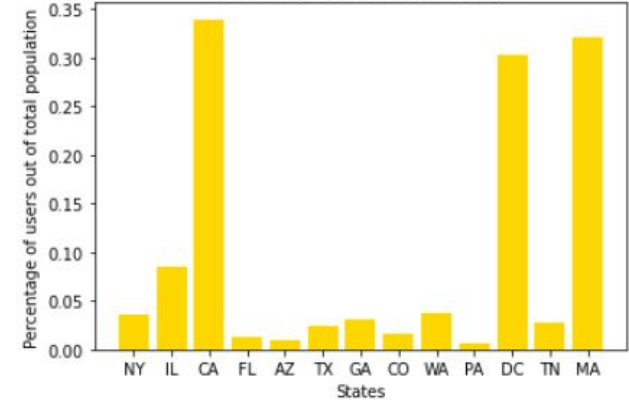
## Dive in



States and users



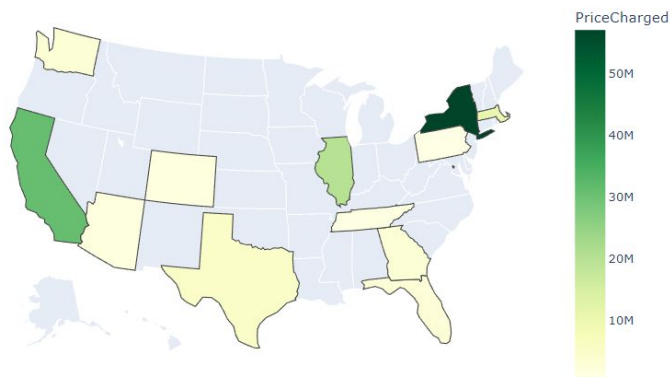
States and users (In percentage)



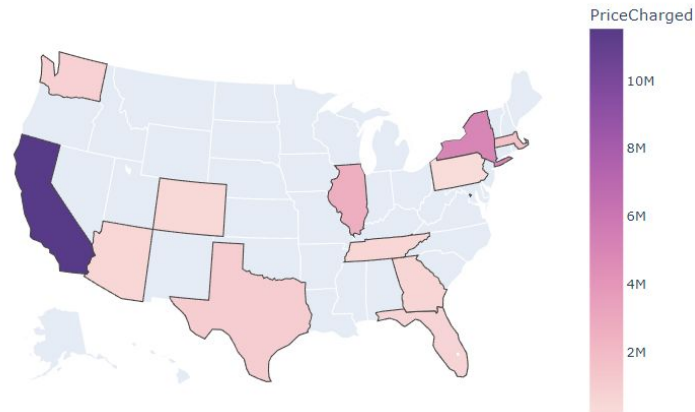
Most users:  
NY, CA, IL, DC, MA  
Most conversion  
rate:  
CA, MA, DC, IL, NY

States with the most highest population to users conversion rate are generally with higher education level and have higher GDPs.

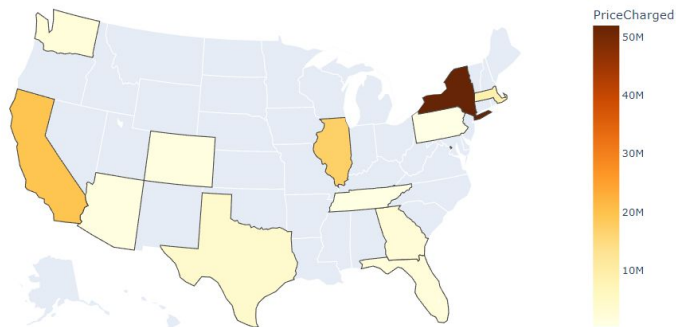
## Total priced charged by each company in different regions



All cab companies



Pink Cab

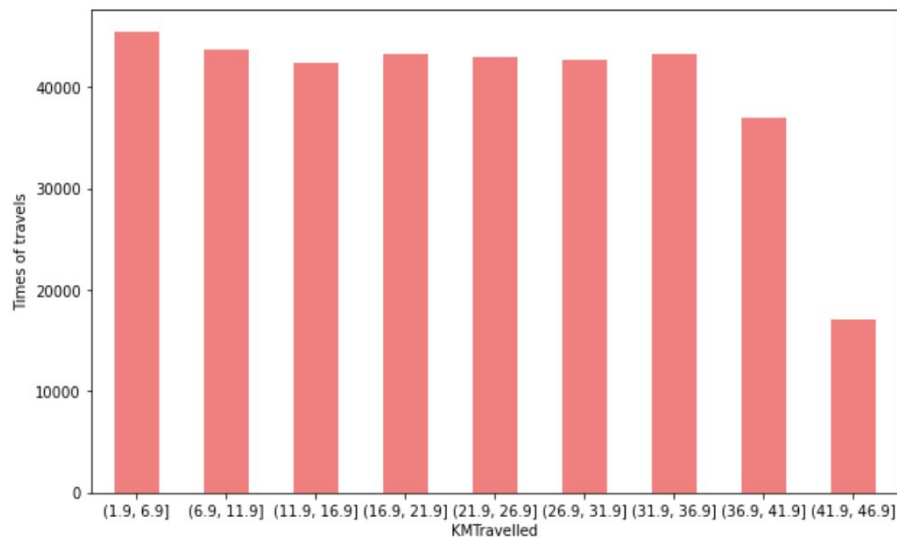


Yellow Cab

From the heat maps, in the east coast, Yellow cab is more welcomed, and charged more in total, while Pink cab is more welcomed in the west coast.

Combined with previous region analysis, NY, CA, IL, DC, MA have the most users and their conversion rate are also high, for more profits, the we need to mainly focus on the company are welcomed in more regions, that would be the Yellow Cab

## KM vs. trips

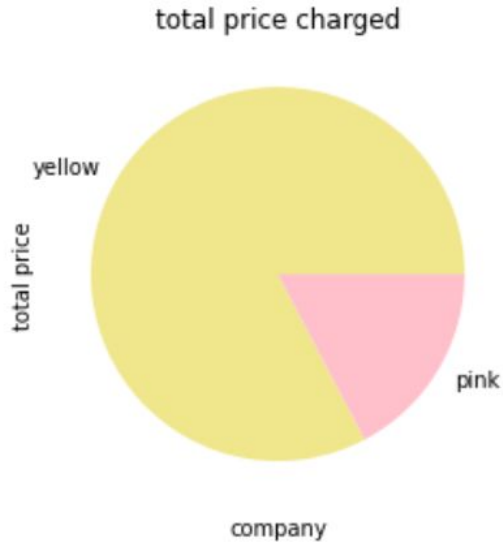


SAN - San Diego 3.1 miles  
BOS - Boston 3.1 miles  
PHX - Phoenix 3.2 miles  
SJC - San Jose 3.8 miles  
ABQ - Albuquerque 4.9 miles  
HNL - Honolulu 5.9 miles  
...

JFK - New York 15.4 miles  
ORD - Chicago 17.9 miles  
LAX - Los Angeles 18.2 miles  
PIT - Pittsburgh 19.1 miles  
IAH - Houston 20.1 miles  
DFW - Dallas 20.9 miles; Ft. Worth 24.5 miles  
MCI - Kansas City 21.0 miles  
DTW - Detroit 22.1 miles  
DEN - Denver 25.2 miles  
IAD - Washington DC 26.9 miles

As expected, the trips with respect to KM follows a uniform distribution. And the distance are all within the access to the airport from their downtowns.

# Total price charged by two companies

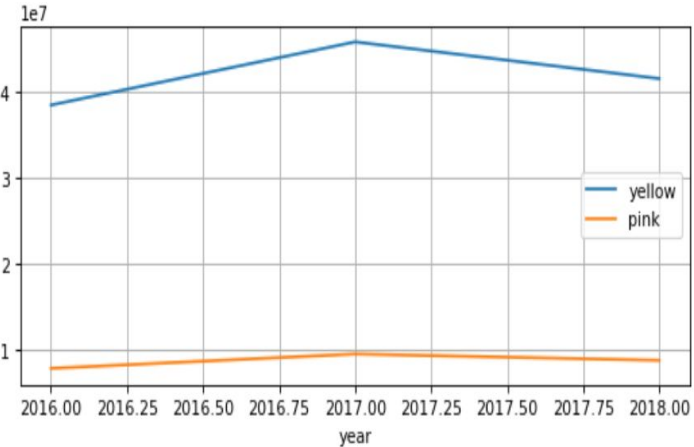
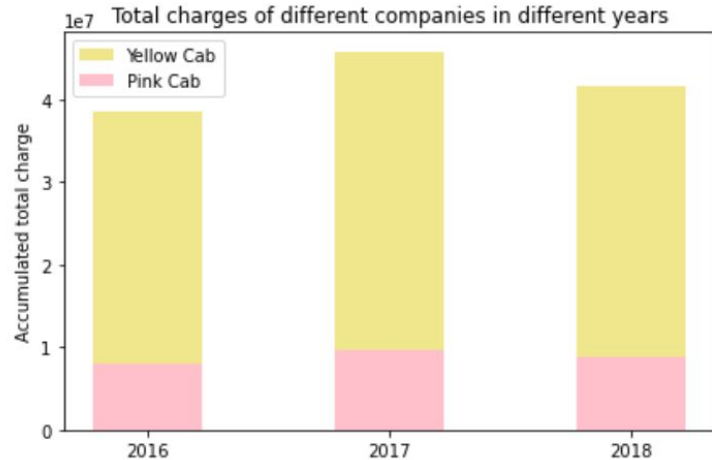


Accumulated total price charged by two companies in three years.

It is easy to find that **Yellow Cab** takes most parts of the company, approximately 80% of the cab markets' revenue.

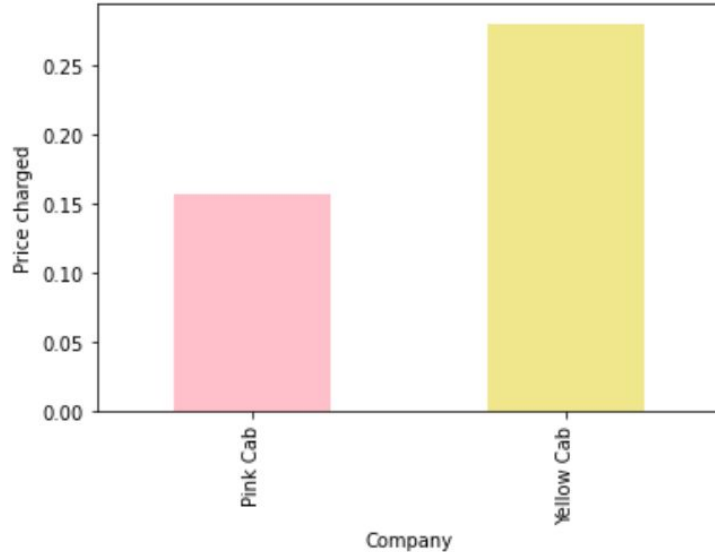
Growth rate	Yellow Cab	Pink Cab
2017	0.1743	0.1601
2018	-0.0834	-0.1026

## Total charges in different years in two companies



The conclusion is similar to previous part, Yellow Cab takes the most part of the cab market, and from the right chart, in 2018, the two companies all experienced a decline. The **growth rate** of **Yellow Cab** is also higher than Pink Cab.

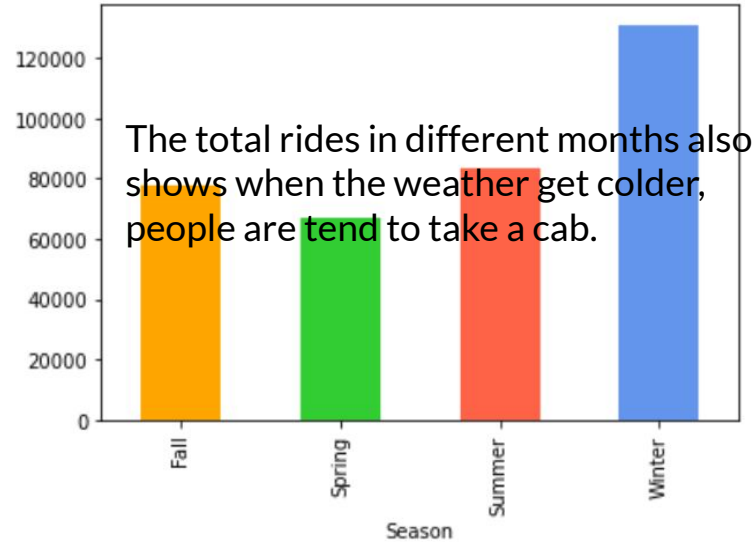
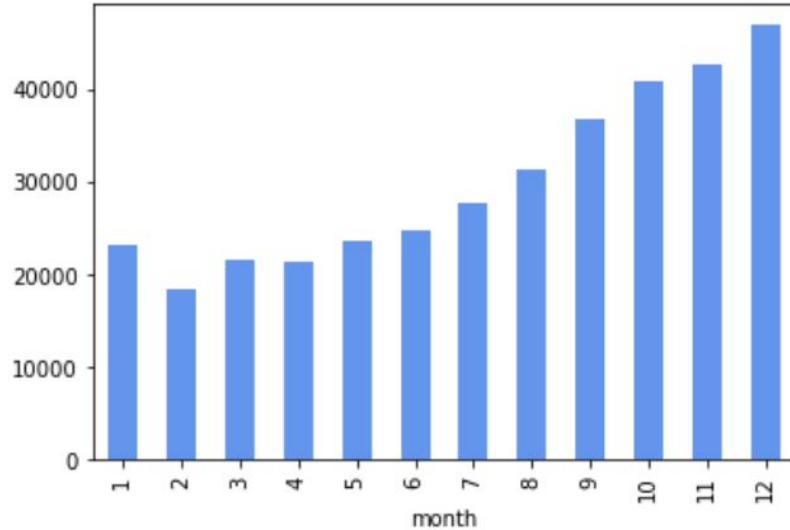
## Additional fee collected by the two companies



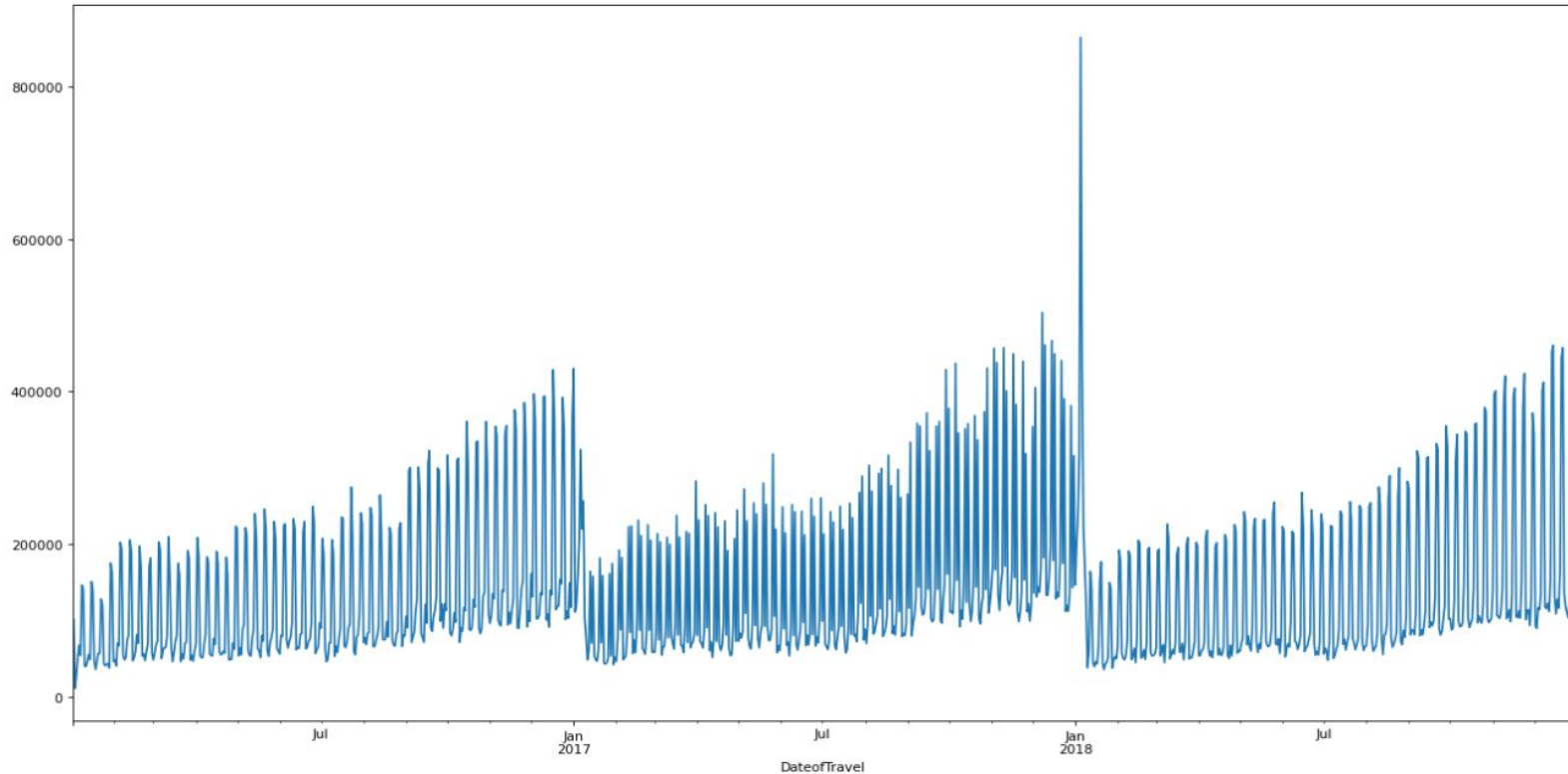
Excepted for the mileage fee and basic fees generated by calling a cab, cab companies might charge additional operating fees, tax, or tips. The difference between 'Priced charged' and 'Cost of trip' could cover these fees.

**Yellow Cab** collects more additional fees than Pink Cab. (The total accumulated price charged of yellow cab is much more than pink cab, to make sure the analysis is fair, the y axis indicates the **mean** value of additional fees divided by total price charged)

## Times of rides in different seasons

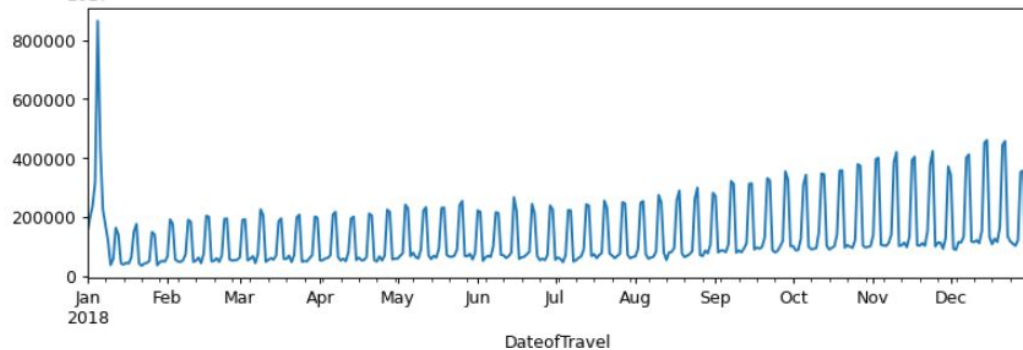
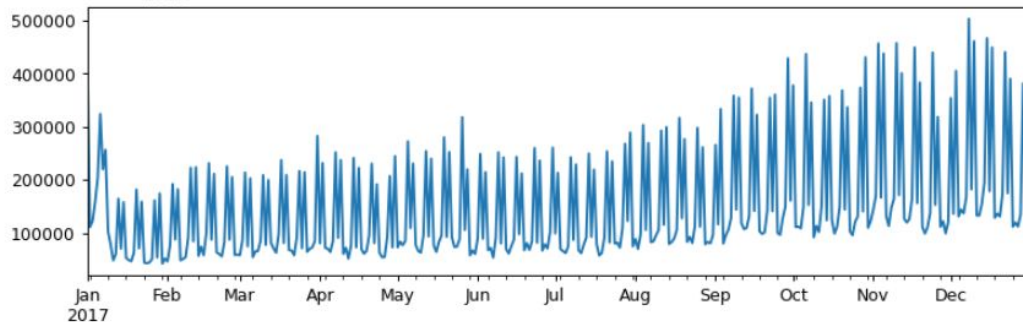
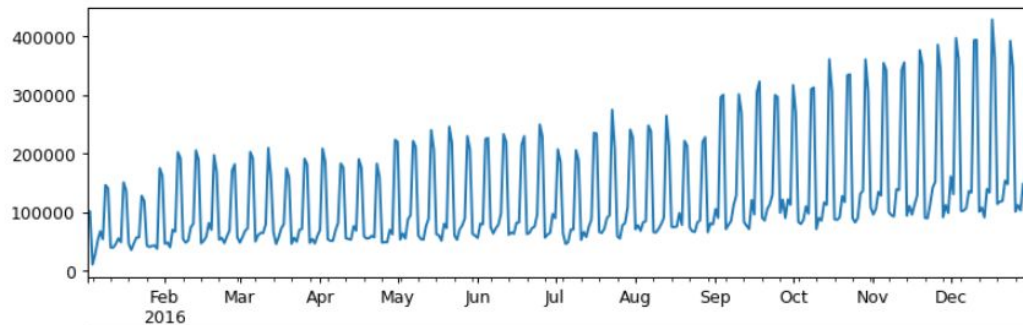


# Time series data

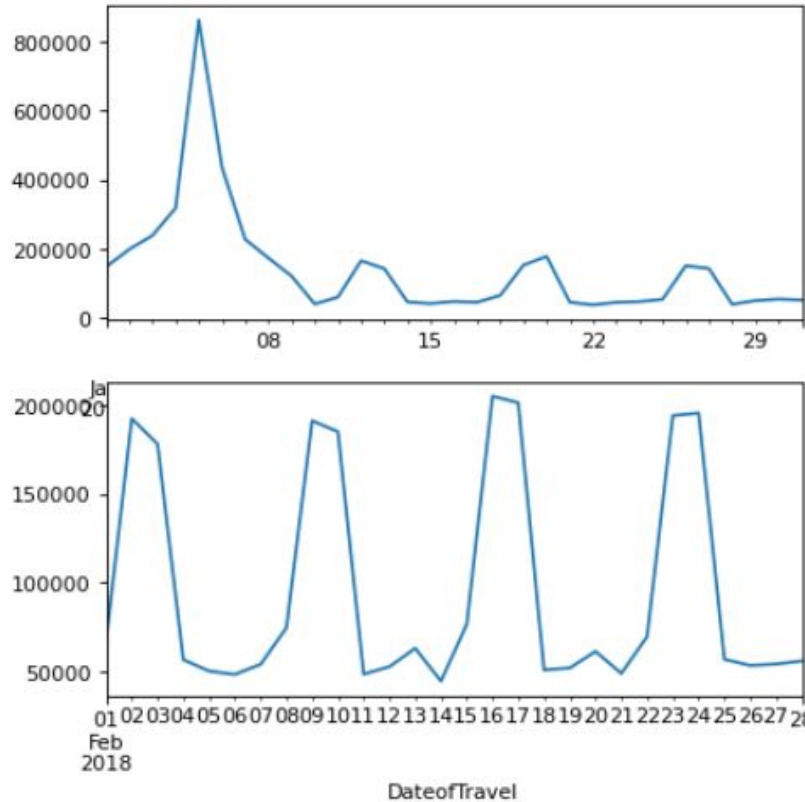


From the time series plot, there are obvious seasonalities and also an upward trend in each year when winter comes





When zoom in each year, the upward trend is more obvious. And there are also seasonalities in each month.



Take January and February in 2018 as examples. The pattern is repeated in each week, 5-7, 12-14 of Jan, 2-4 of Feb, are all weekends. The needs for cabs are much higher than work days.



# Recommendations

In general: **Yellow Cab** is a better choice for investment.

Growth rate: Based on the growth rate, **Yellow Cab** does a better job than the Pink Cab

Market occupation: Yellow Cab takes most part of the markets, while in the east coast, CA specifically, Pink Cab takes the lead. **Yellow Cab** still holds the overall leading position in the cab industry.

Profit/Revenue: In the three years, from 2016 to 2018, **Yellow Cab** made more profits than Pink Cab, and charged more on additional fees, this might indicate the special marketing and charging strategy of Yellow Cab.

Demographic: The demographic features of the two companies are similar, but Yellow Cab is more attractive to the small group of users, for example, people are older, and people with higher incomes, where their original percentage is smaller. This may also indicate that **Yellow Cab** is more popular in special groups, and can generally reach out to more potential customers.

Though Pink Cab are more popular in few dimensions, Yellow Cab is more worth to be invested.