



**Data Glacier**

Your Deep Learning Partner

# Exploratory Data Analysis

G2M Insight for cab Investment

**Somasundaram Palaniappan**

**21-July-2022**

# Agenda

Executive Summary

Problem Statement

Data Exploration

Approach

EDA

EDA Summary

Recommendations



- XYZ is the company give them analysis to find the ideal taxi company for an investment.
- Analysing the total number of cab service, customers and generating monthly, annual earnings on weekdays and weekends will help you identify the most popular cab service provider in each American city.

Analysis comprise:

- Data Understanding
- Testing the data quality
- Visualisation of data
- Forming a hypothesis
- Investment Recommendations.

# Problem Statement



**Data Glacier**

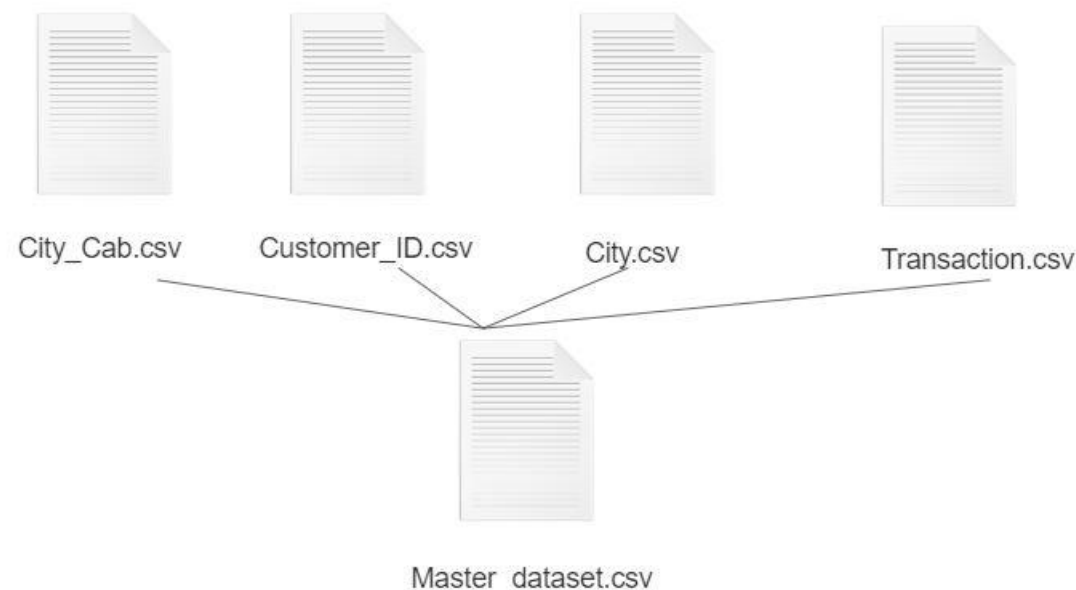
Your Deep Learning Partner

- Statement: XYZ is a private firm in US. Due to remarkable growth in 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: Help XYZ company to find which company is better for making investment.



### Datasets:

- 22 Features
- Timeformat of data: 2016-01-31 to 2018-12-31
- Total datapoints: 359392
- Cab\_Data.csv : this file contains details of transaction for 2 cab companies
- Customer\_ID.csv : this is a mapping table that contains a unique identifier which links that 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 users





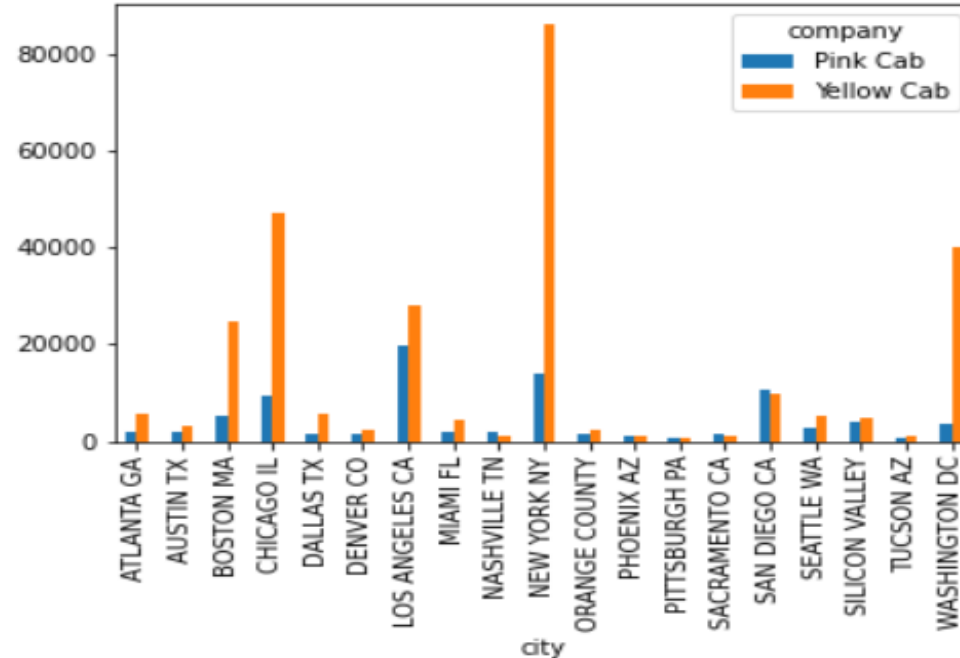
- Master dataset has been created by merging 4 datasets.
- New Features are added using the existing features in dataset.
  1. Day, Month, Weekday, Year – It was extracted from the date of travel
  2. Profit per trip – Shows the difference between the price charged and cost of trip
  3. Profit per KM – It's the profit per trip out of KM travelled
  4. Price per KM – It shows the price charged out of KM travelled
  5. Cost per KM – Shows the cost of trip out of KM travelled



- In order to derive the insights from the data, an exploratory data analysis approach was used.
- Which taxi service is demand in various cities in America ?
- Which mode of payment is most popular ?
- Which taxi business is more favoured by customers ?
- Identify the city with the most users
- Which gender uses taxis more frequently
- Distribution of user's age
- Distribution of profits by company and year
- Distribution of customers by firm
- The profit margin for each business throughout a week, season and year.

Which taxi service is demand in various cities in America ?

- Newyork city in Yellow cab company has high users and next high users is in Chicago IL. Whereas the pink cab company has low number of users compared to yellow cab.
- It shows the results of which city uses the more customers towards the yellow cab
- Whereas the Pittsburgh has the lowest customers users in the city
- Highest profit is in favour of Newyork , whereas the lower profit is on Sacramento CA.



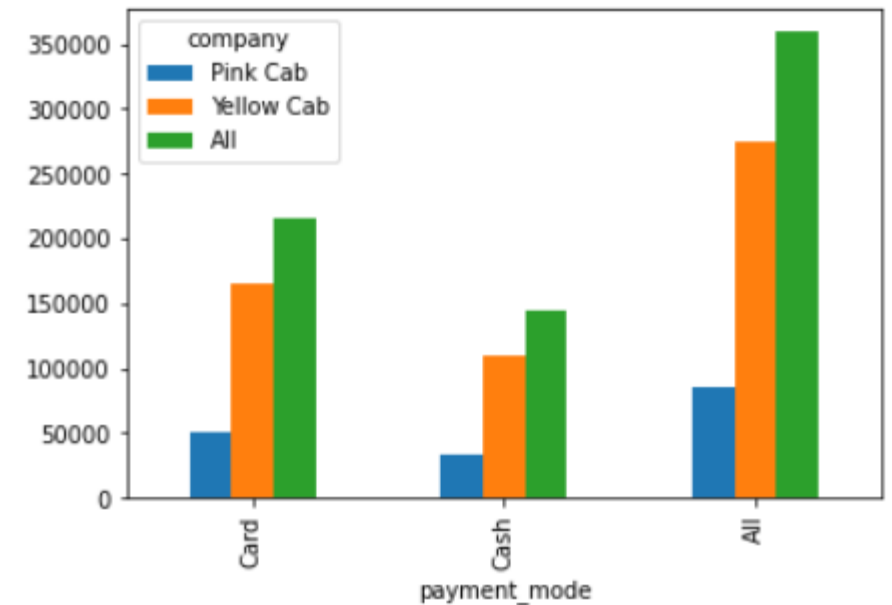
company	Pink Cab	Yellow Cab
city		
ATLANTA GA	1762	5795
AUSTIN TX	1868	3028
BOSTON MA	5186	24506
CHICAGO IL	9361	47264
DALLAS TX	1380	5637
DENVER CO	1394	2431
LOS ANGELES CA	19865	28168
MIAMI FL	2002	4452
NASHVILLE TN	1841	1169
NEW YORK NY	13967	85918
ORANGE COUNTY	1513	2469
PHOENIX AZ	864	1200
PITTSBURGH PA	682	631
SACRAMENTO CA	1334	1033
SAN DIEGO CA	10672	9816
SEATTLE WA	2732	5265
SILICON VALLEY	3797	4722
TUCSON AZ	799	1132
WASHINGTON DC	3692	40045



Which mode of payment is most popular ?

- Comparatively Card payment is much high in yellow cab than in pink cab. yellow cab Cash payment little high compared to pink cab.

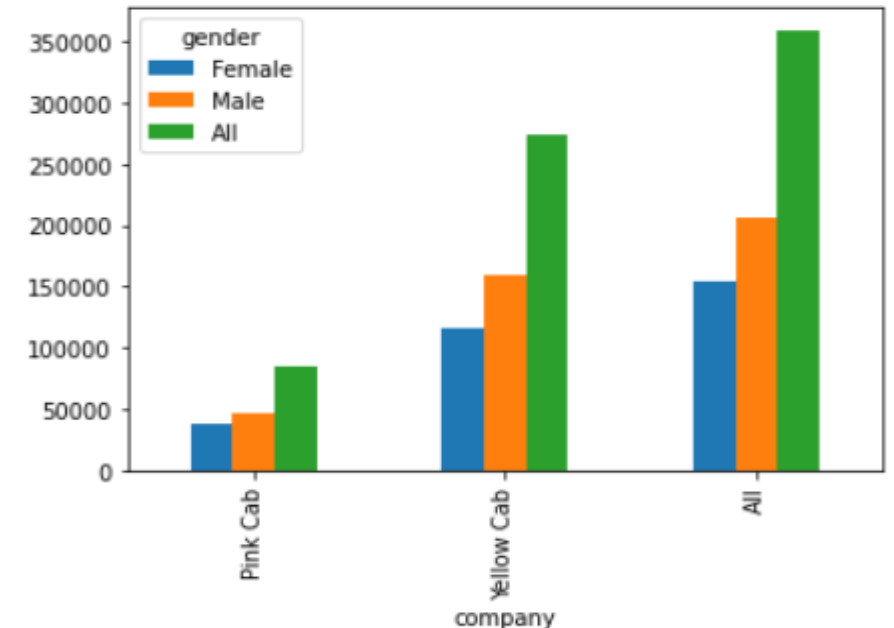
company	Pink Cab	Yellow Cab	All
payment_mode			
Card	50719	164785	215504
Cash	33992	109896	143888
All	84711	274681	359392



## Gender Analysis

- With this gender variation users yellow cab has been used by more male and female compared to pink cab users of male and female. Male Gender uses more cab compared to female gender.

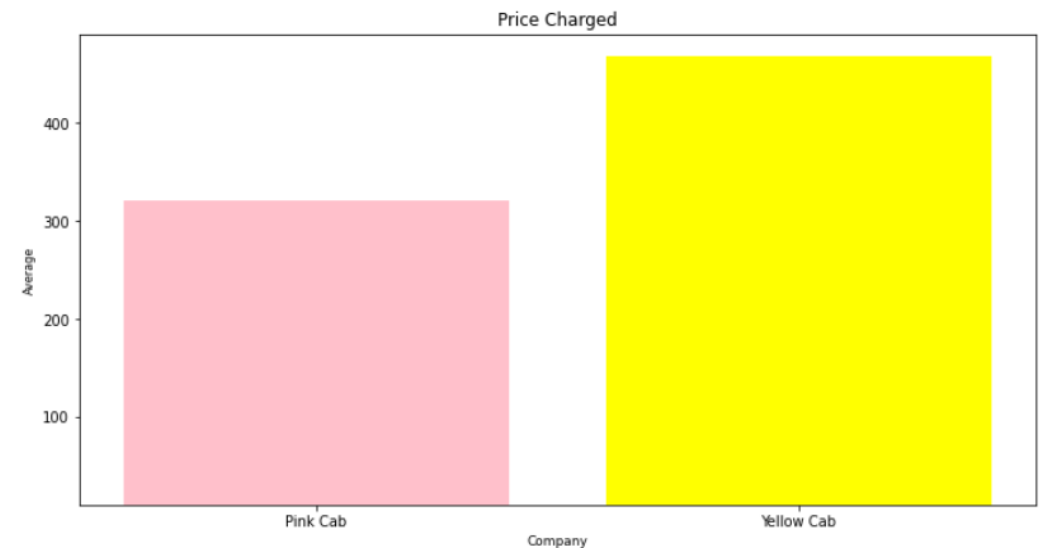
gender	Female	Male	All
company			
Pink Cab	37480	47231	84711
Yellow Cab	116000	158681	274681
All	153480	205912	359392



## Price Charged Analysis

- Yellow cab charges the high amount than pink cab

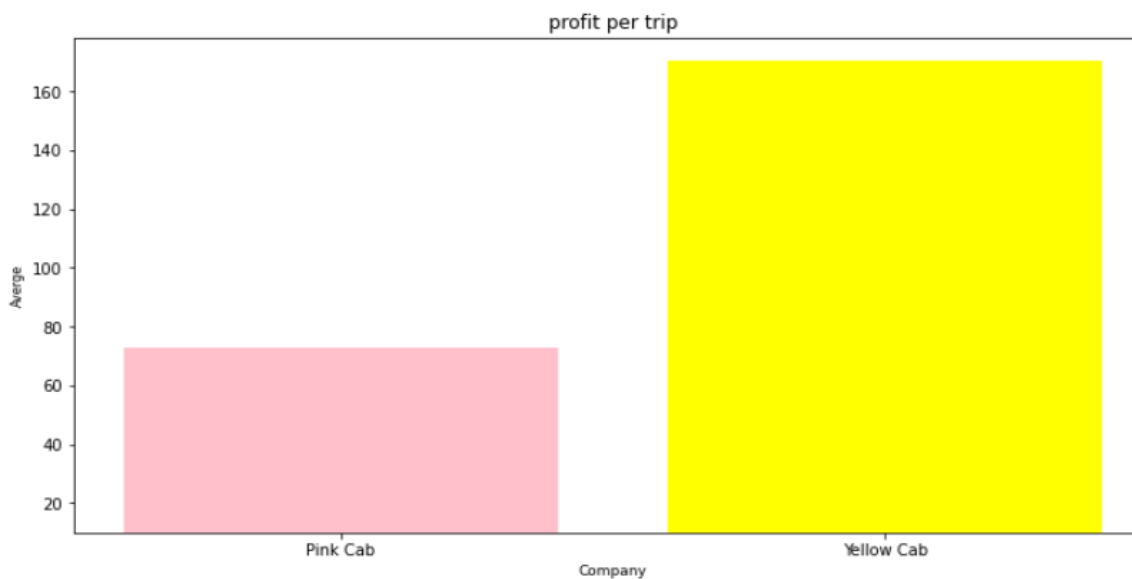
price_charged	
company	
Pink Cab	310.800856
Yellow Cab	458.181990



## Profit Analysis

- Yellow cab profits is high amount than pink cab

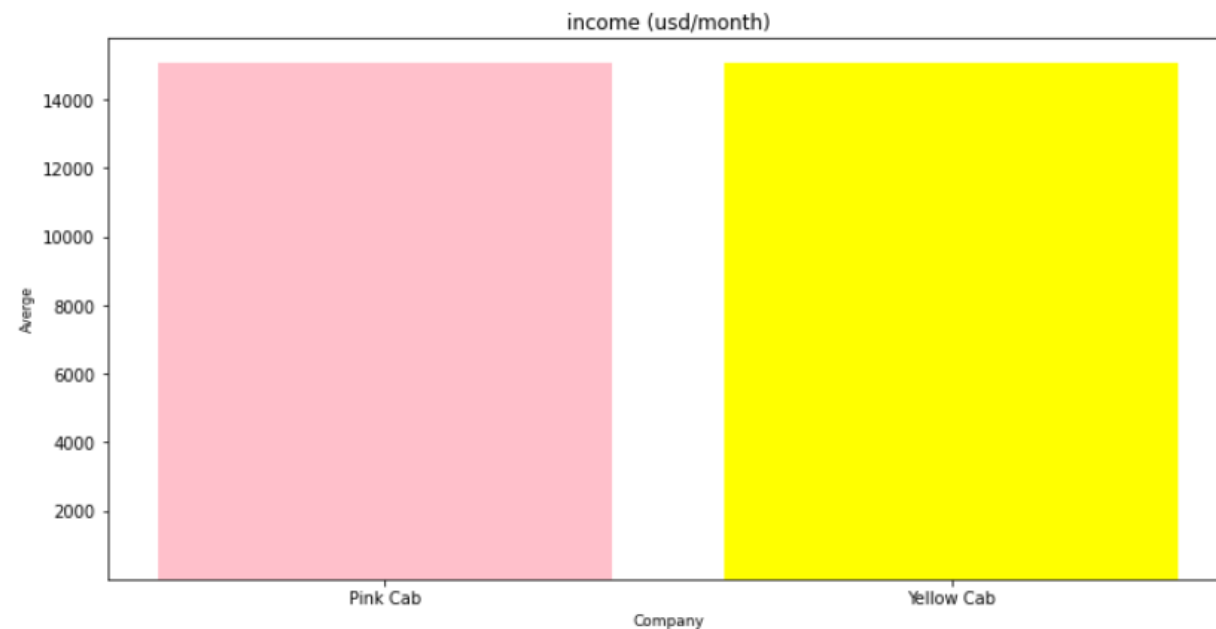
profit_per_trip	
company	
Pink Cab	62.652174
Yellow Cab	160.259986



## Income Analysis

- There is no difference in pink and yellow cab in income analysis

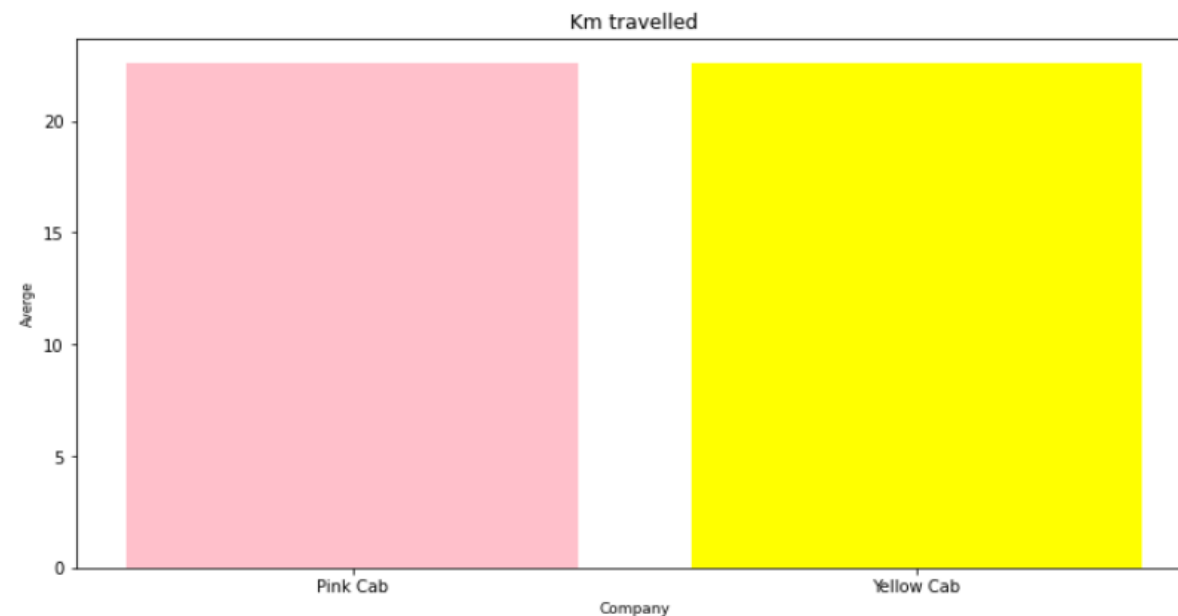
income_(usd/month)	
company	
Pink Cab	15059.047137
Yellow Cab	15045.669817



## KM Travelled Analysis

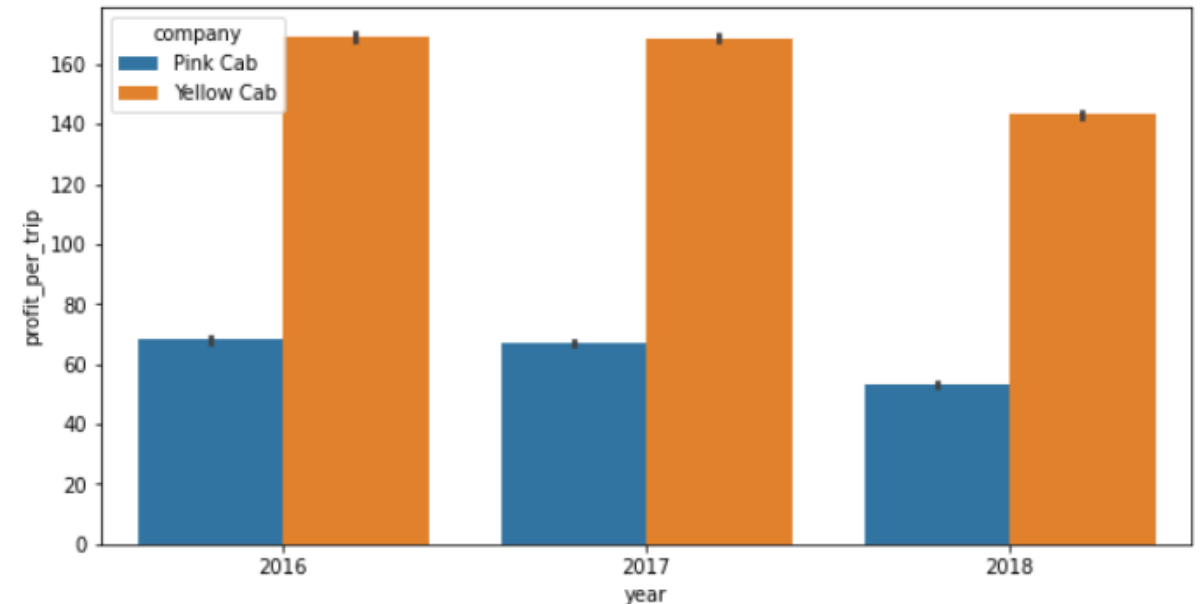
- There is no difference in pink and yellow cab in KM travelled

km_travelled	
company	
Pink Cab	22.559917
Yellow Cab	22.569517



## Distribution of profits by company and year

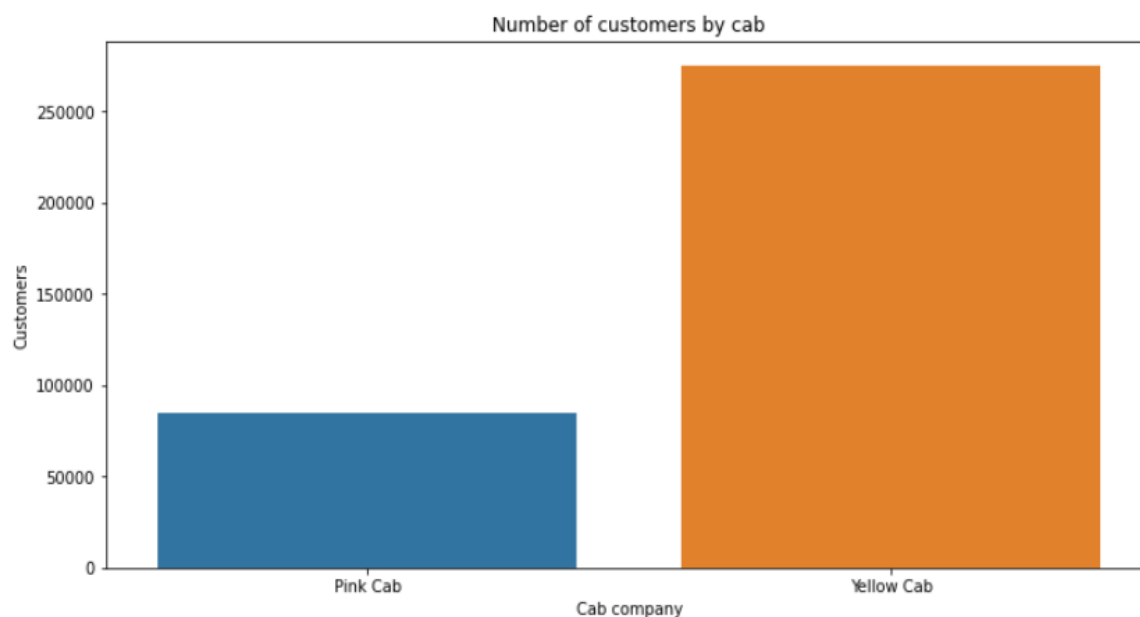
- Thus, profit per trip visualization shows that the 2016 and 2017 Yellow cab has higher profit whereas the 2018 it has fall down. In Pink cab Profit was stable in 2016 and 2017 and 2018 it has fall down



## Distribution of customers by firm

- More customer uses the yellow cab, but the same customers uses the both yellow cab and pink cab

company		customer_id	size
Pink Cab		831664558	84711
Yellow Cab		1049163053	274681

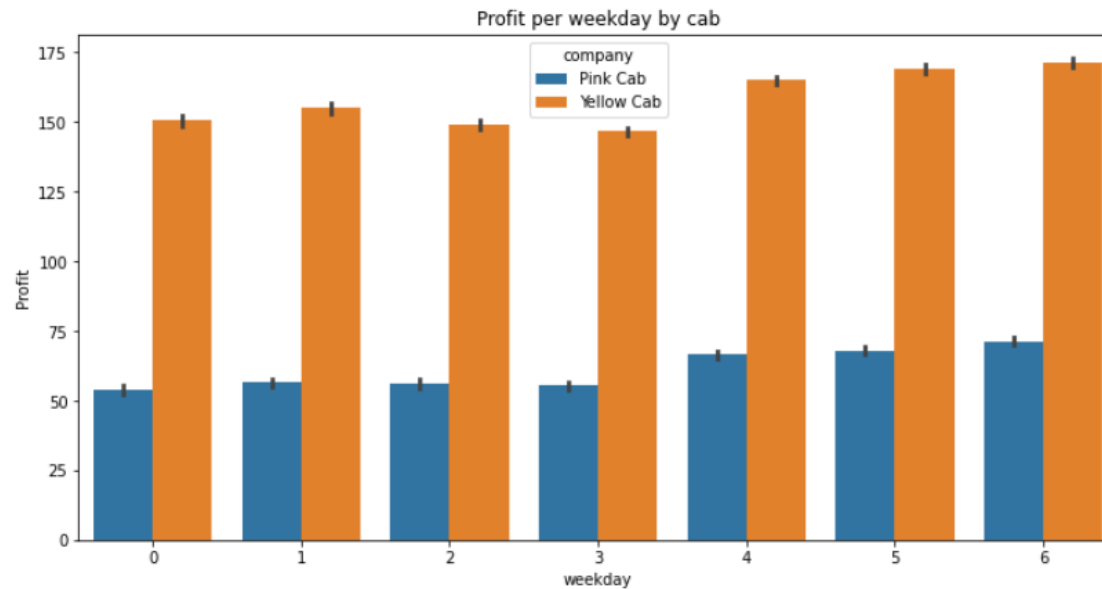


	customer_id	company	size
0	1	Pink Cab	4
1	1	Yellow Cab	25
2	2	Pink Cab	4
3	2	Yellow Cab	36
4	3	Pink Cab	6
...	...	...	...
72221	59998	Yellow Cab	6
72222	59999	Pink Cab	2
72223	59999	Yellow Cab	6
72224	60000	Pink Cab	4
72225	60000	Yellow Cab	14



## Weekly Analysis

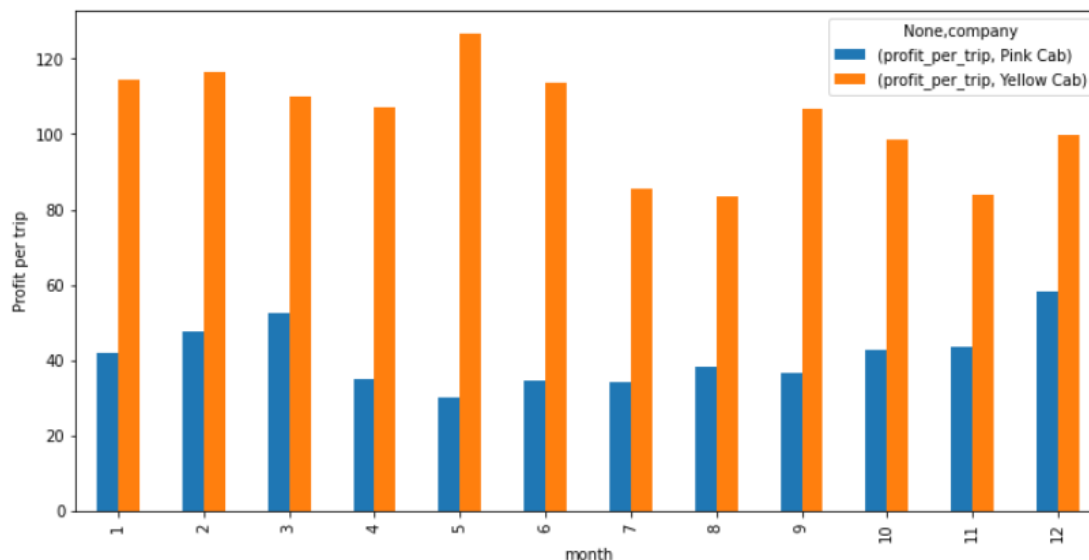
- In weekends particularly, Profit on per trip getting Increased in yellow cab on the same side increase in km travelled



		km_travelled	profit_per_trip	profit_per_km
weekday	company			
0	Pink Cab	22.05	34.3105	2.058884
	Yellow Cab	22.42	93.1312	5.103434
1	Pink Cab	22.61	35.6000	2.134995
	Yellow Cab	22.41	96.6506	5.391836
2	Pink Cab	22.26	36.4670	2.106509
	Yellow Cab	22.42	93.1888	5.149244
3	Pink Cab	22.40	36.0680	2.139563
	Yellow Cab	22.26	89.9904	5.025000
4	Pink Cab	22.66	45.9650	2.612900
	Yellow Cab	22.47	106.1260	5.792595
5	Pink Cab	22.31	46.1200	2.706624
	Yellow Cab	22.60	109.9228	5.876486
6	Pink Cab	22.54	47.7005	2.751732
	Yellow Cab	22.60	112.7820	6.159612

## Seasonality Analysis

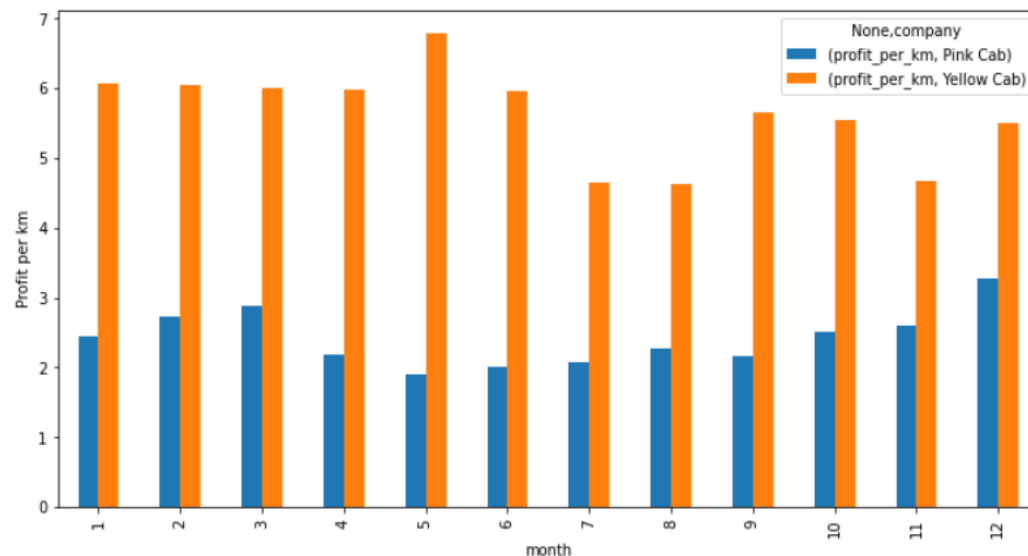
- Yellow cab has the greater margin than the pink cab



	km_travelled		profit_per_trip		profit_per_km	
company	Pink Cab	Yellow Cab	Pink Cab	Yellow Cab	Pink Cab	Yellow Cab
month						
1	22.355	22.54	41.935	114.6160	2.454020	6.074740
2	22.310	22.57	47.520	116.3060	2.736672	6.051205
3	23.520	22.22	52.398	109.8300	2.890560	6.010245
4	22.080	22.47	34.840	107.1488	2.184878	5.971364
5	22.000	22.23	30.260	126.4988	1.898216	6.777619
6	22.600	22.31	34.558	113.7192	1.999074	5.966703
7	22.260	22.66	34.030	85.4900	2.071250	4.656637
8	22.230	22.31	38.230	83.3244	2.269637	4.618318
9	22.440	22.42	36.803	106.8900	2.157562	5.650641
10	22.420	22.54	42.980	98.7100	2.514739	5.537637
11	22.420	22.61	43.768	83.9762	2.598498	4.671760
12	22.610	22.60	58.124	99.6280	3.268564	5.509328

## Seasonality Analysis

- Profit is higher in may month But it got lower in july and august month on both profit per Km and trip

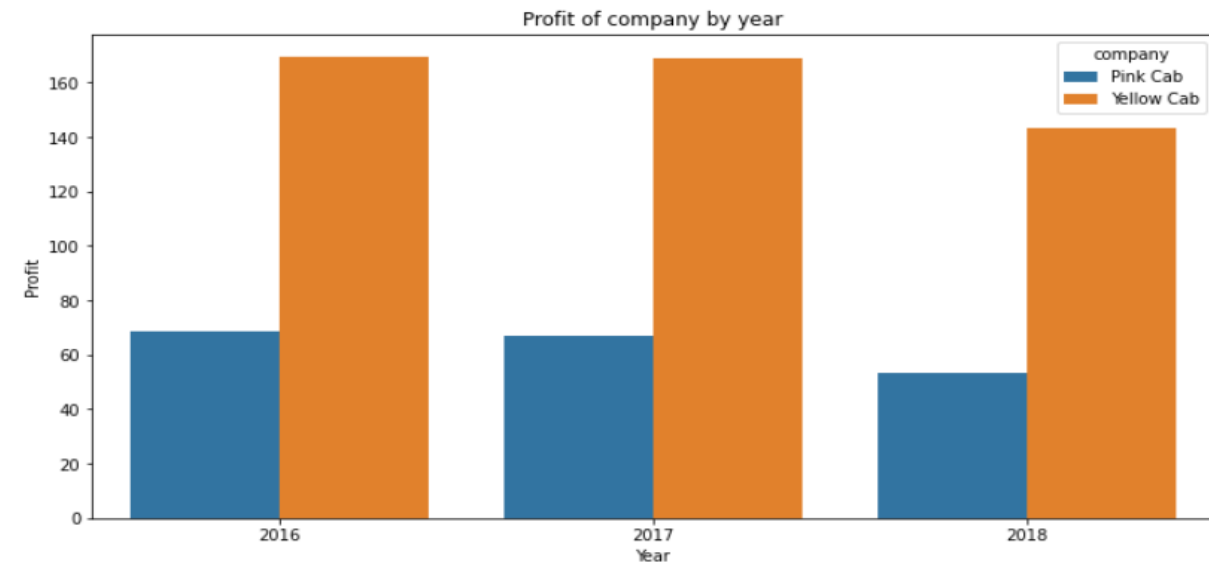


	km_travelled		profit_per_trip		profit_per_km	
company	Pink Cab	Yellow Cab	Pink Cab	Yellow Cab	Pink Cab	Yellow Cab
month						
1	22.355	22.54	41.935	114.6160	2.454020	6.074740
2	22.310	22.57	47.520	116.3060	2.736672	6.051205
3	23.520	22.22	52.398	109.8300	2.890560	6.010245
4	22.080	22.47	34.840	107.1488	2.184878	5.971364
5	22.000	22.23	30.260	126.4988	1.898216	6.777619
6	22.600	22.31	34.558	113.7192	1.999074	5.966703
7	22.260	22.66	34.030	85.4900	2.071250	4.656637
8	22.230	22.31	38.230	83.3244	2.269637	4.618318
9	22.440	22.42	36.803	106.8900	2.157562	5.650641
10	22.420	22.54	42.980	98.7100	2.514739	5.537637
11	22.420	22.61	43.768	83.9762	2.598498	4.671760
12	22.610	22.60	58.124	99.6280	3.268564	5.509328

## Yearly Analysis

- In 2016 and 2017, the yellow cab has higher profit and pink cab remains same on both year. Whereas in 2018 there is decrease in yellow cab and pink cab

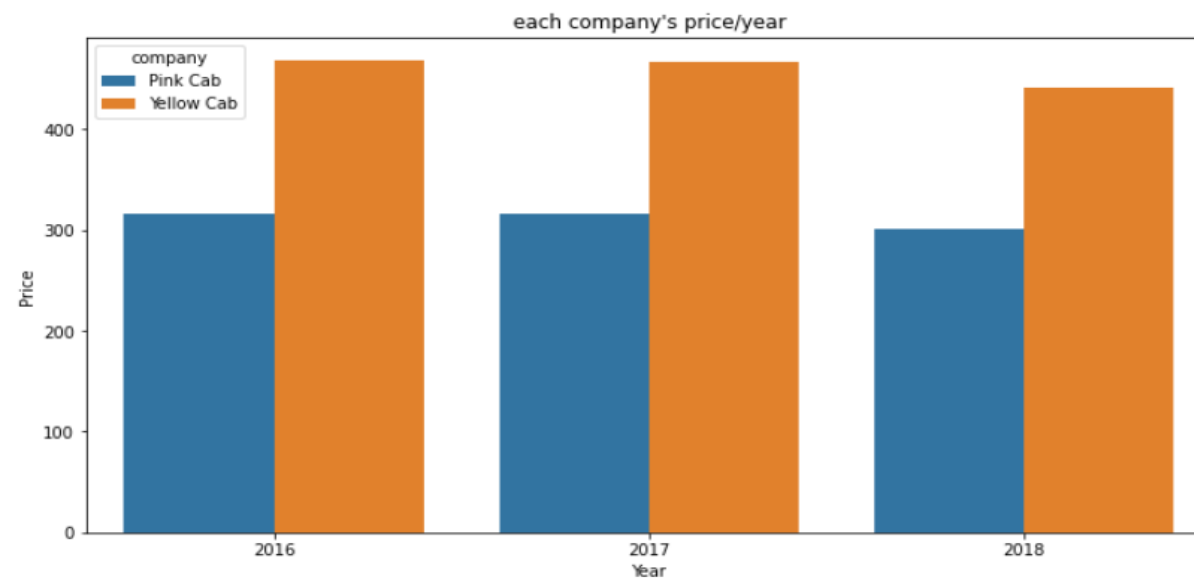
	year	company	profit_per_trip
0	2016	Pink Cab	68.321819
1	2016	Yellow Cab	169.347821
2	2017	Pink Cab	67.070839
3	2017	Yellow Cab	168.817057
4	2018	Pink Cab	53.229689
5	2018	Yellow Cab	143.416122



## Yearly Analysis

- In each year yellow cab is high and increased the price rates

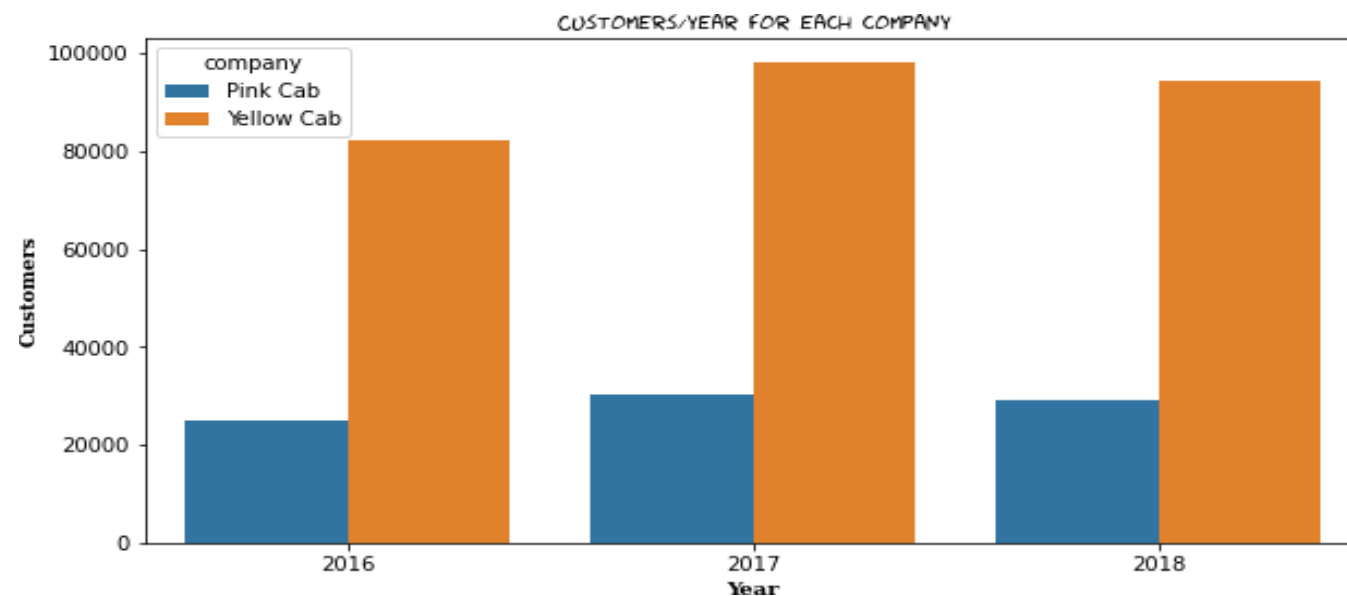
	year	company	price_charged
0	2016	Pink Cab	315.330113
1	2016	Yellow Cab	467.918301
2	2017	Pink Cab	315.907442
3	2017	Yellow Cab	466.639950
4	2018	Pink Cab	301.642530
5	2018	Yellow Cab	440.875558



## Cab user per year by company

- In 2017 Yellow cab has highest users compared to pink cab

	year	company	customer_id
0	2016	Pink Cab	25080
1	2016	Yellow Cab	82239
2	2017	Pink Cab	30321
3	2017	Yellow Cab	98189
4	2018	Pink Cab	29310
5	2018	Yellow Cab	94253





- In Practically all cities, Yellow cab is preferred by most users, according to the results of the Exploratory data analysis technique.
- Young and Middle-aged users make up the majority.
- Yellow company has historically had the highest profit margin
- The profit exhibits some seasonality.



- Yellow cab is the best decision to invest:
- Yellow cab has more profit and big share in last over years and also in terms of rides.
- It has high number of customer particularly in 2017.
- Customer uses the yellow and pink cab mostly are aged between 20-40
- Big cities with huge population are prefer yellow cab
- There are some seasonality because both cabs recorded deflation in July and august on profit per Km and trip.
- Yellow cab has always higher profit in all these days compared to pink cab.
- Most customers used Both card and cash payment for pink cabs, Whereas most card payment has been done by yellow cab.



# Thank You