



Data Glacier

Your Deep Learning Partner

Exploratory Data Analysis

G2M insight for Cab Investment firm

Jamila Hamdi

27-June-2021

Agenda

Executive Summary

Problem Statement

Approach

EDA

EDA Summary

Recommendations



- Provide actionable insights for XYZ to facilitate the identification of the best cab company for making investment.
- Find the most user-preferred cab service provider in each city in the United States by looking at every number of cab service users and making monthly, seasonal, and annual profits on weekdays and holidays.
- The analysis includes :
 - ✓ Data understanding
 - ✓ Checking data quality
 - ✓ Data visualization
 - ✓ Creating hypothesis
 - ✓ Recommendations for investment

Problem Statement



Data Glacier

Your Deep Learning Partner



OR



- ☐ **Description :** 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.
- ☐ **Objective :** help XYZ firm to identify which company, Yellow cab or Pink cab, is better for making investment .

Approach

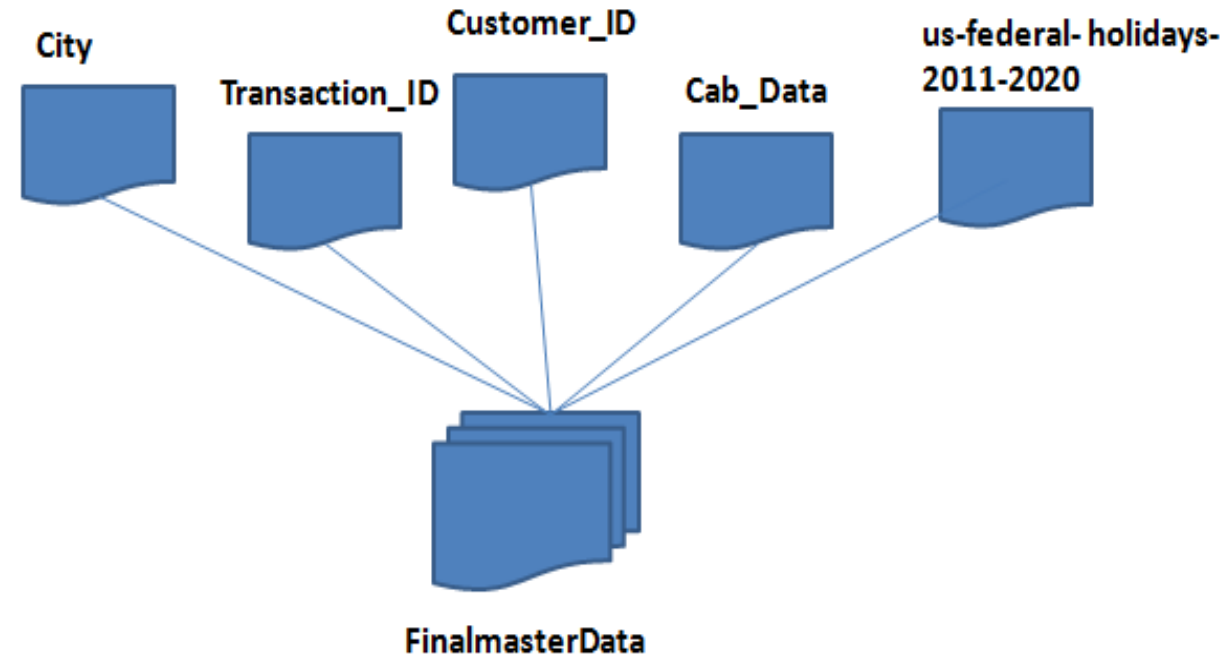


Data Glacier

Your Deep Learning Partner

Available datasets:

- **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 users.
- **Us_federal_holidays_2011_2020** : this file contains list of US holidays names and dates from 2011-2020.



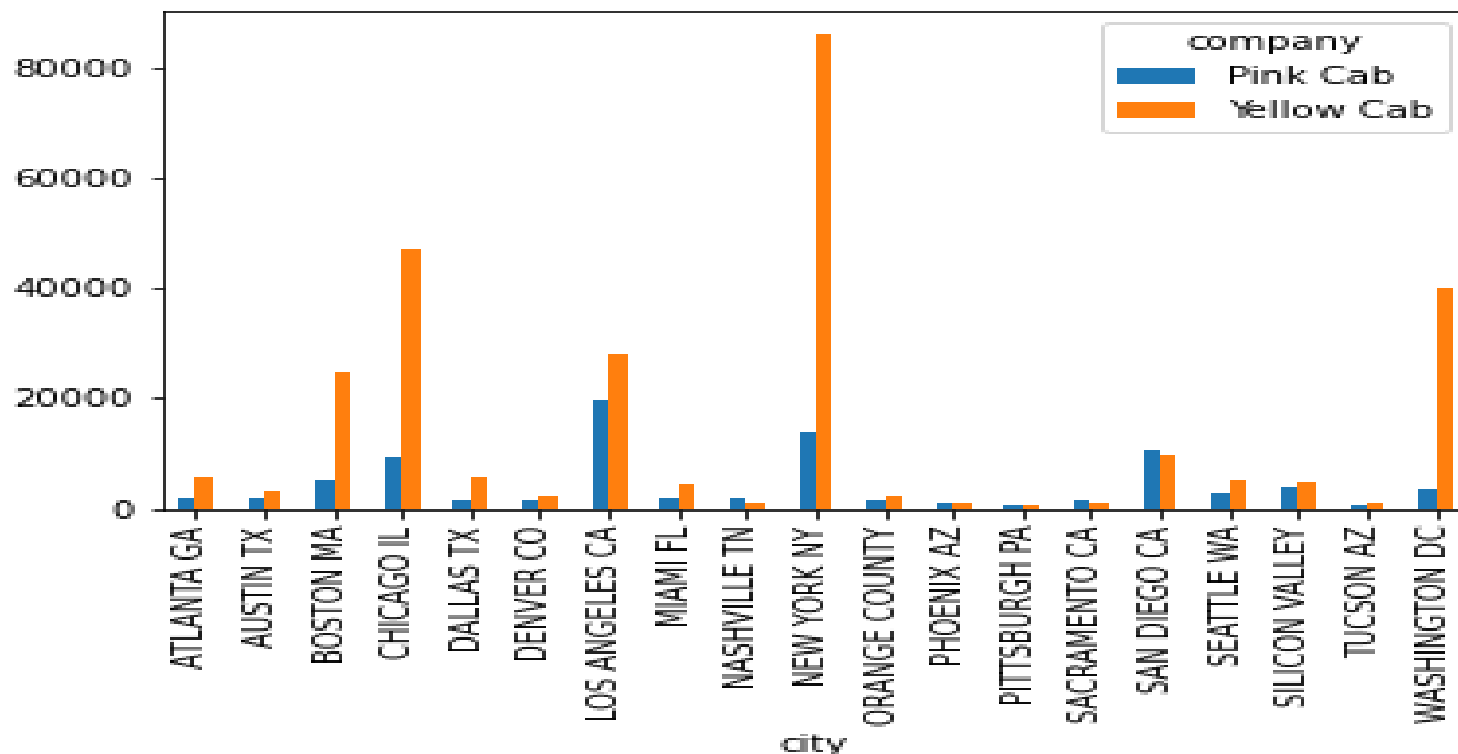


- Create a master dataset by merging all the five datasets.
- 8 new features were added and extracted from existing features which are:
 1. **Day, weekday, month and year** are extracted from the date of travel
 2. **Profit per one trip**, which is the difference between price charged and cost of trip
 3. **Profit per one km**, it's the profit per one trip value out of kilometres travelled
 4. **Price per km**, it's the price charged per one trip value out of kilometres travelled
 5. **Cost per km**, it's the cost of trip value out of kilometres travelled

➔ Exploratory Data Analysis approach was utilized to draw insights from the data.

- ✓ Which cab company is in high demand in different cities
- ✓ What's the preferred payment method used
- ✓ Which cab company is more preferred by users
- ✓ Which city has the highest number of users
- ✓ Which gender travels more t by taxi
- ✓ Age distribution of the users
- ✓ Profit distribution by year and company
- ✓ Customer distribution per company
- ✓ The profit margin of each company in each week, a season and year
- ✓ Analyse the distance travelled on weekdays against holidays
- ✓ Average kilometres travelled in holidays
- ✓ Cost of trip, price charged and profit per one trip during holidays

City analysis:

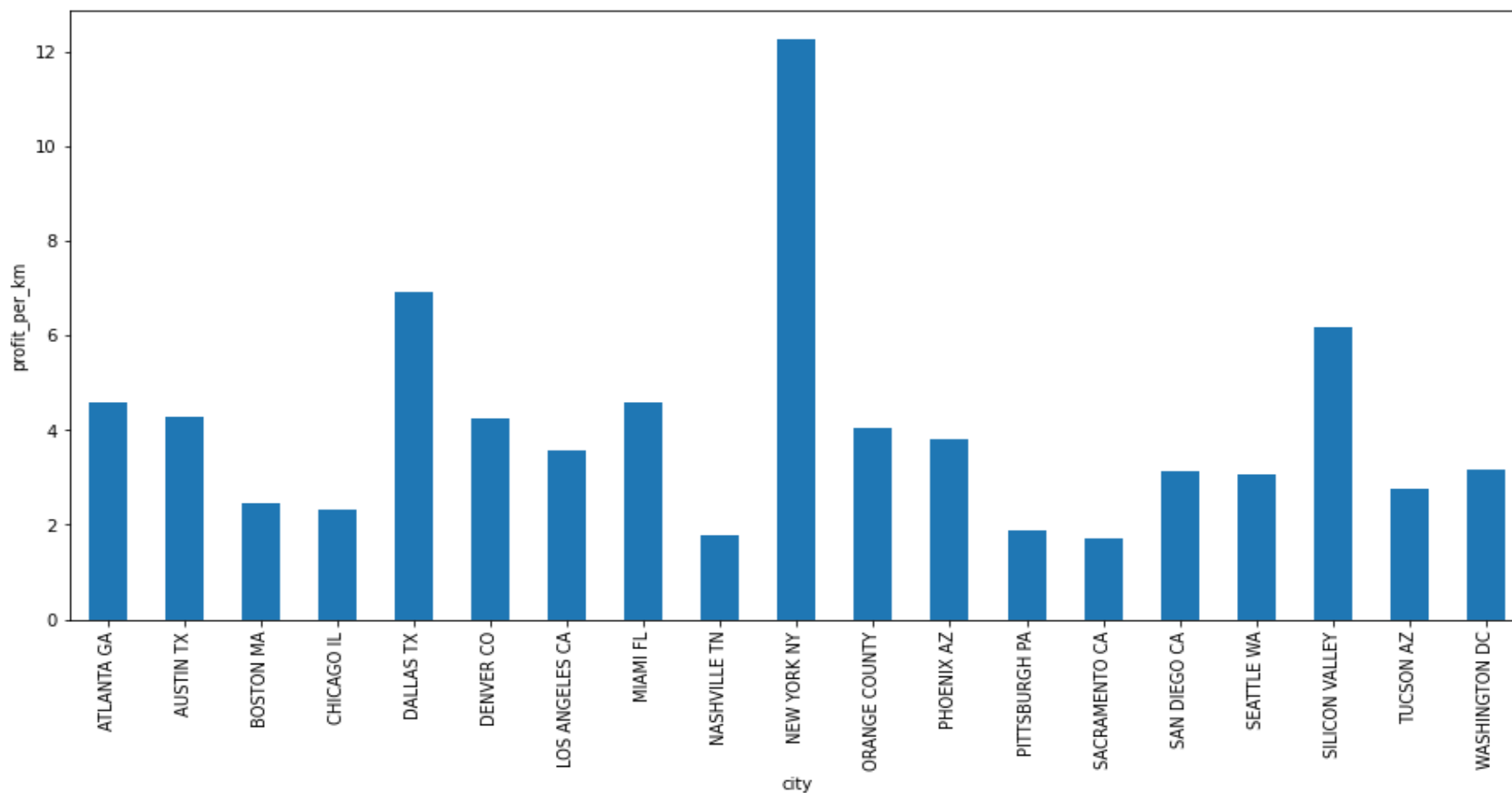


➔ In almost all cities, Yellow cab company it is more in demand than Pink cab company.

➔ New York City has the highest number of users . However, PITTSBURGH PA has the lowest number of users in the data set.

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

City analysis:

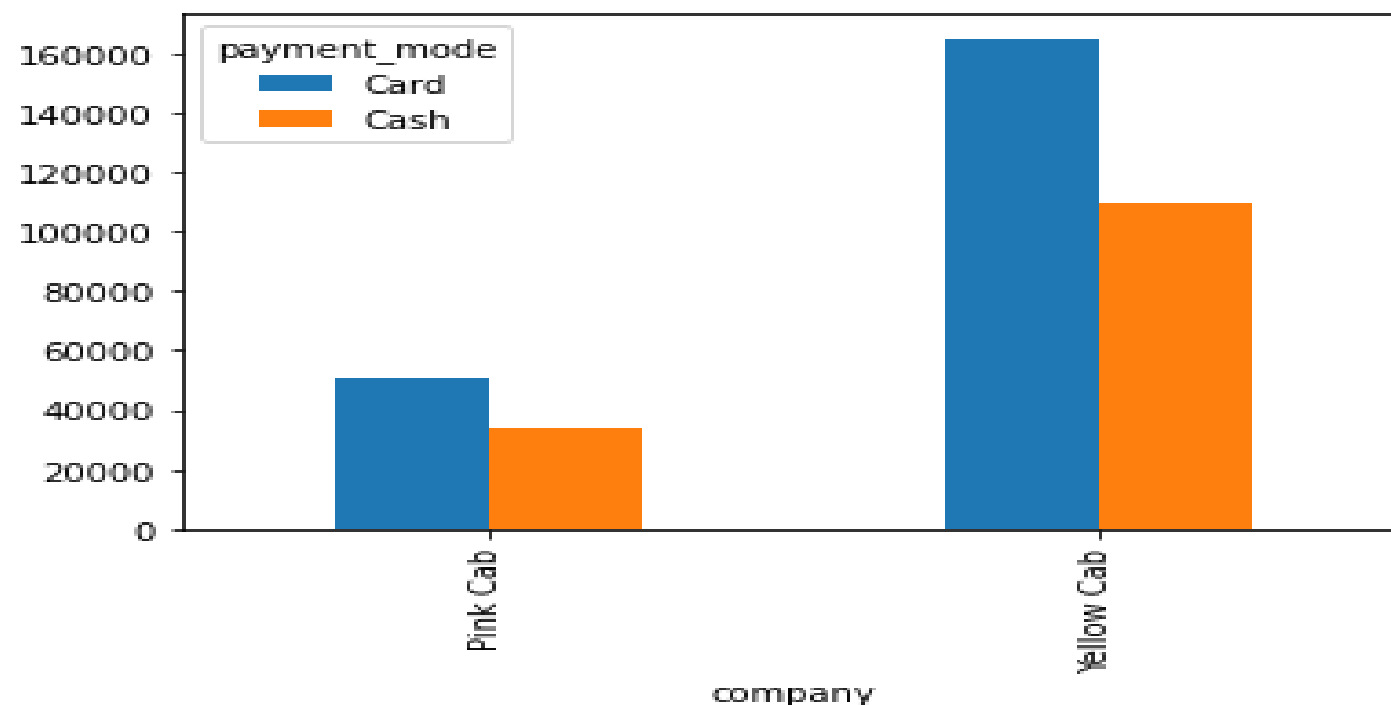


➔ The highest Profit per km is in favor of New York City.

➔ However the lowest Profit per KM is in favor of Sacramento CA.

- Payment mode analysis

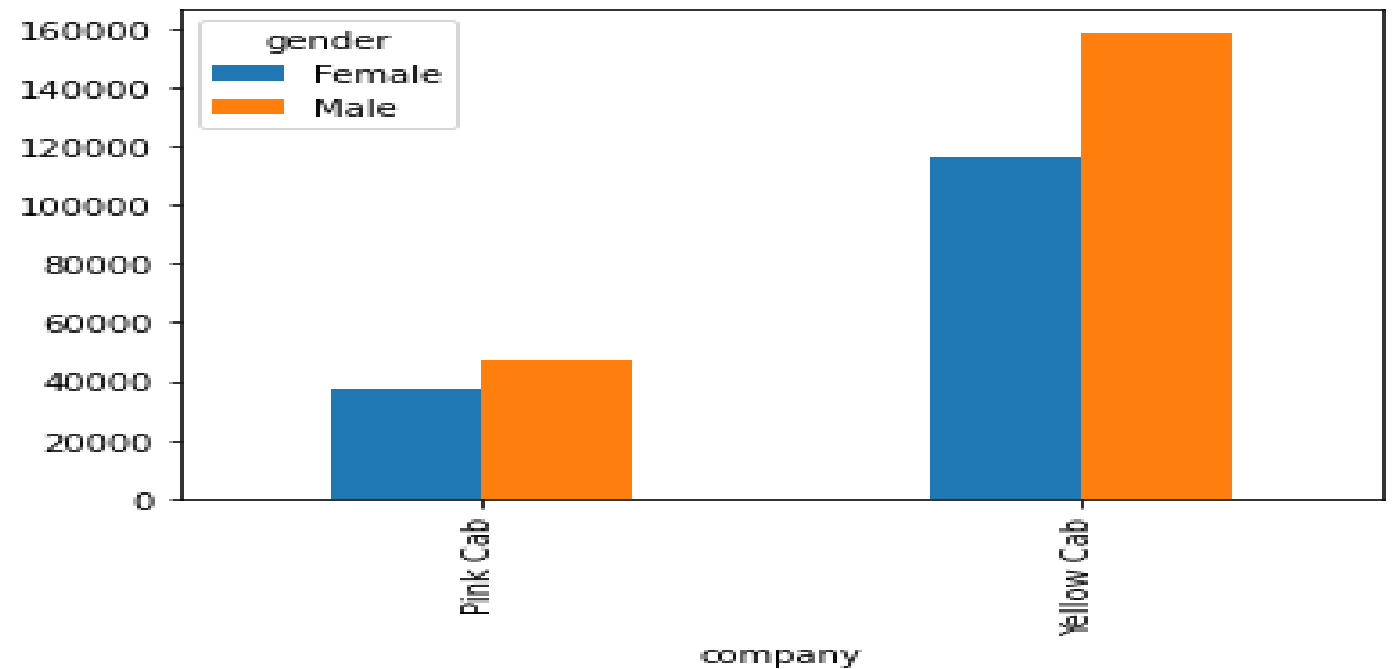
Company	Pink Cab	Yellow Cab	All
Payment mode			
Card	50719	164785	215504
Cash	33992	109896	143888
All	84711	274681	359392



➔ The Card payment method is more used than Cash payment in both, Pink cab and Yellow cab.

- Gender analysis

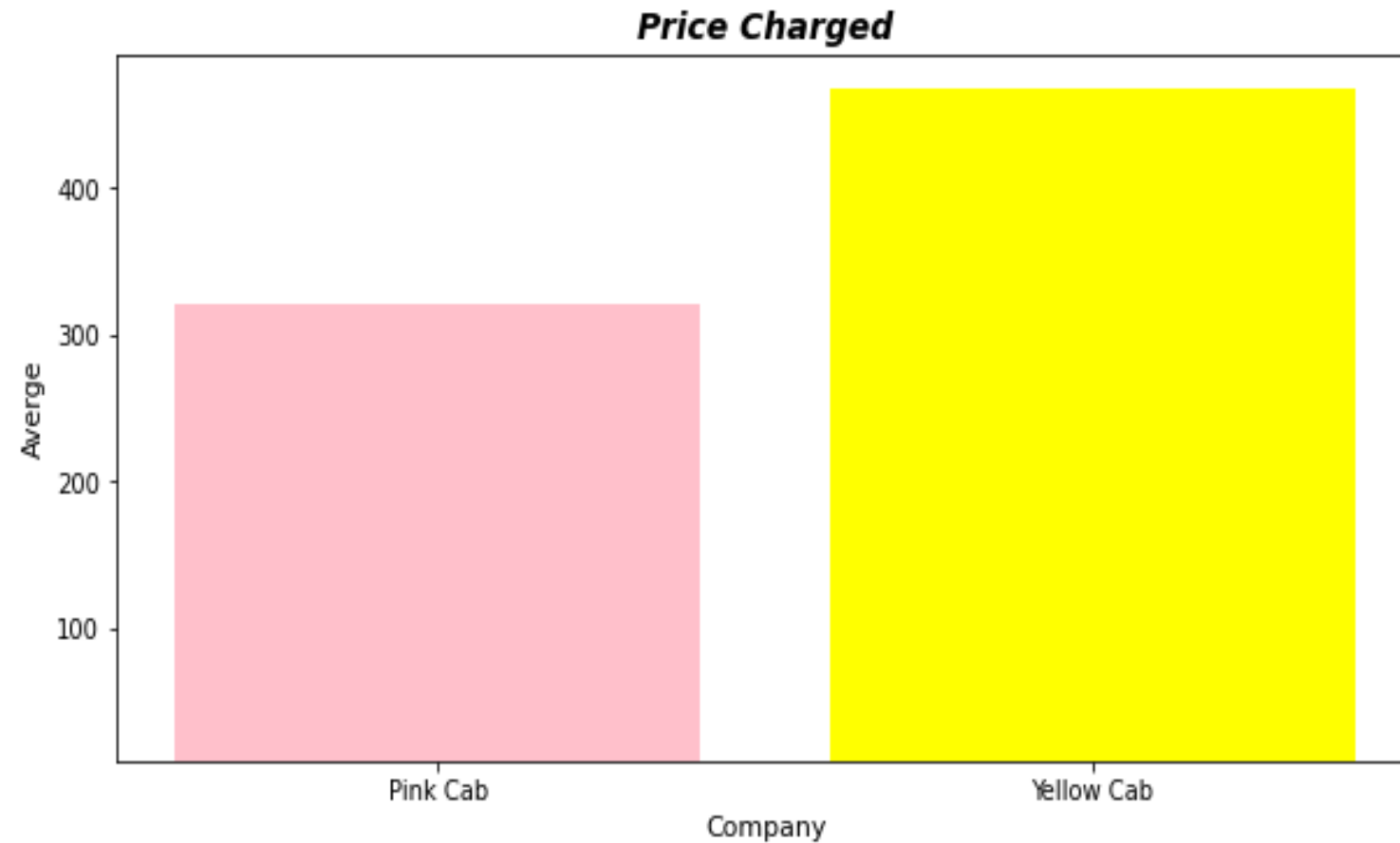
company	Pink Cab	Yellow Cab	All
gender			
Female	37480	116000	153480
Male	47231	158681	205912
All	84711	274681	359392



➔ The number of male users who travel by taxi is high compared to female users.

- Price charged analysis

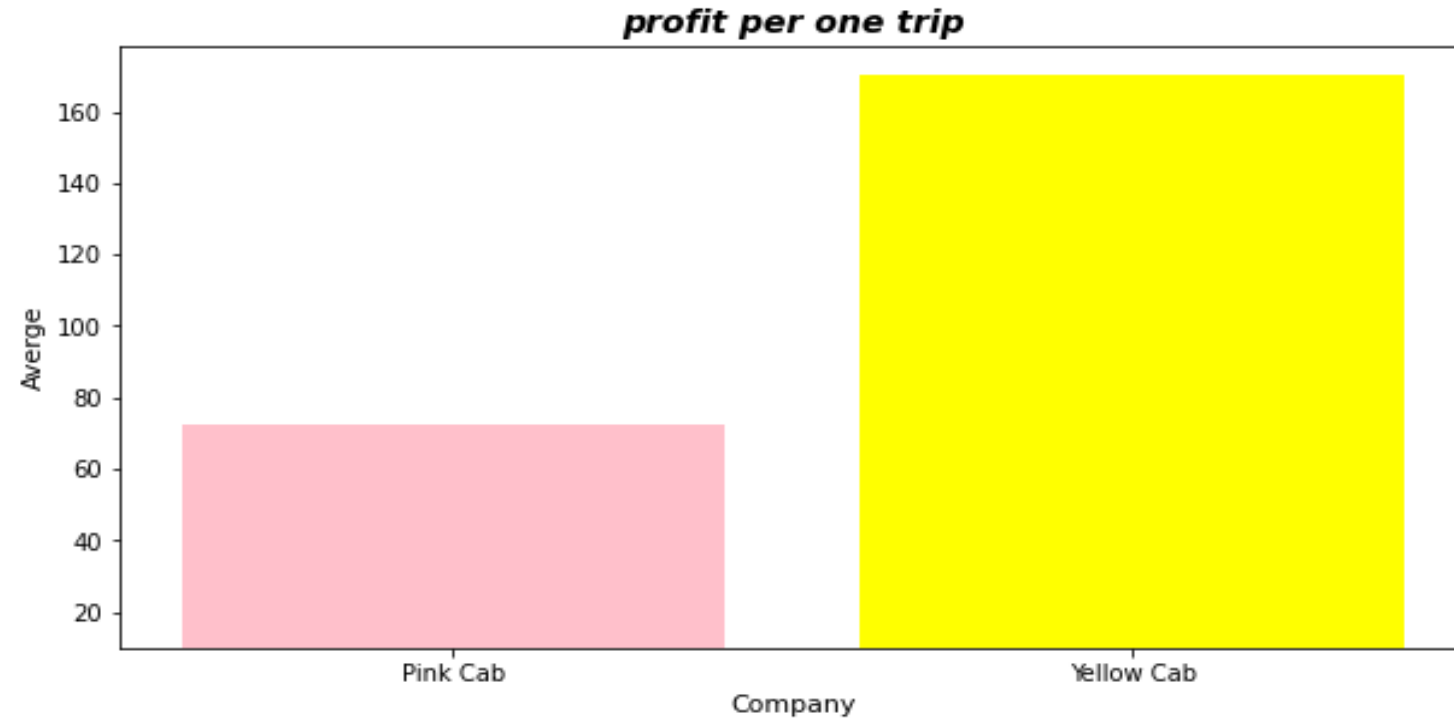
	Price charged
Company	
Pink Cab	310.800856
Yellow Cab	458.181990



➔ The charged price of the Yellow company is higher than that of the Pink company.

- Profit analysis

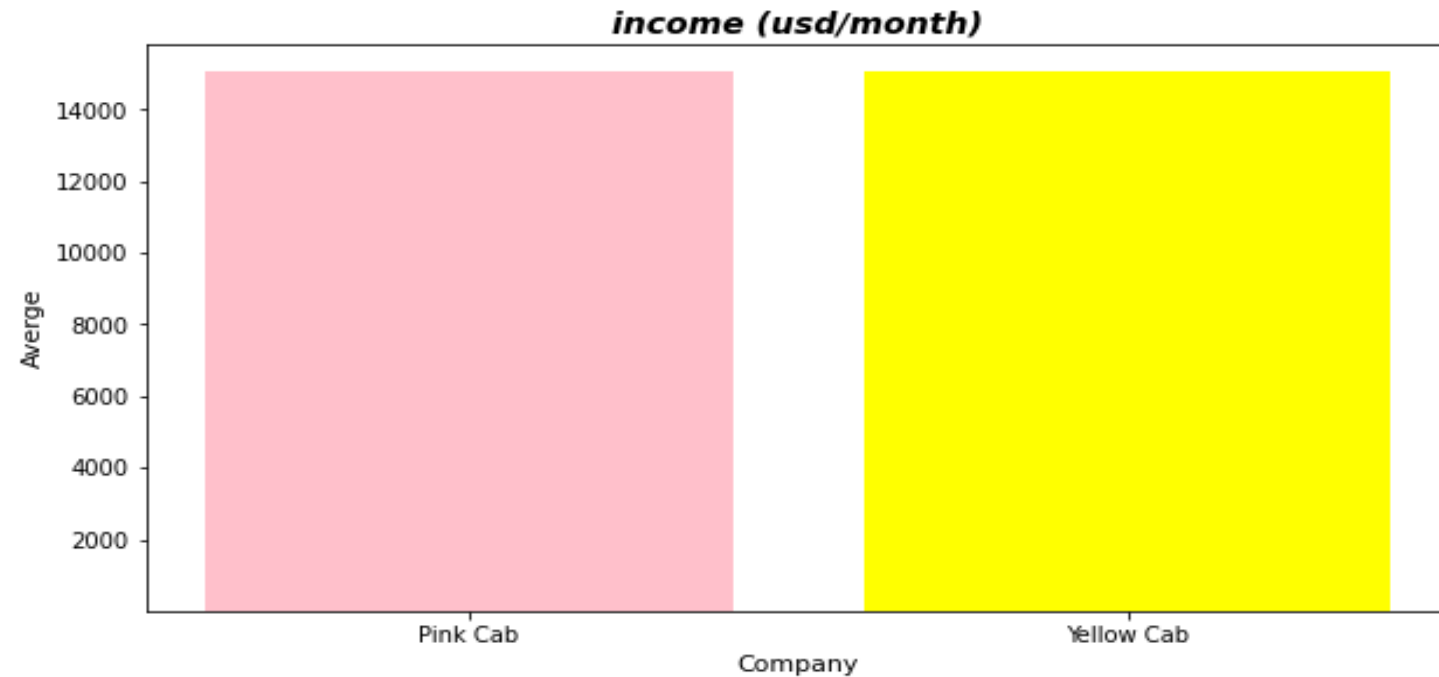
	Profit per one trip
company	
Pink Cab	62.652174
Yellow Cab	160.259986



➔ The Profit per one trip of the Yellow company is higher than that of the Pink company.

- Income analysis

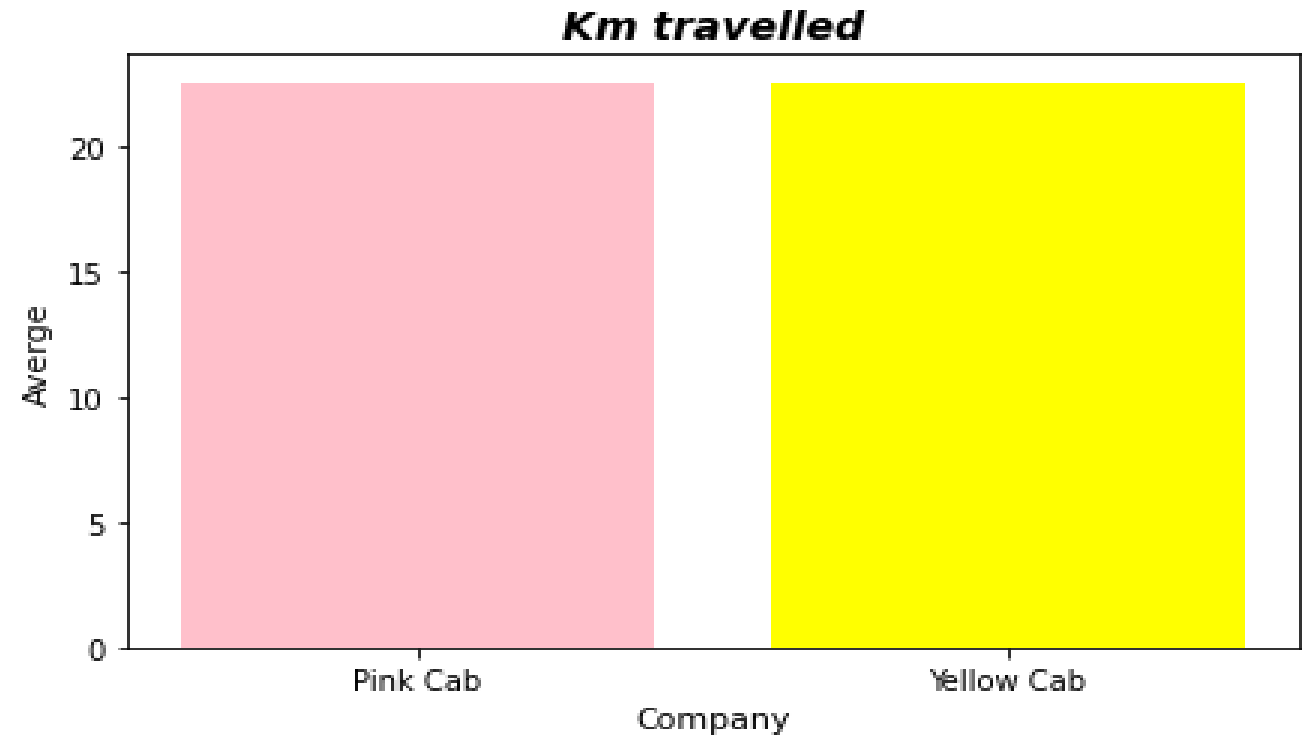
	Income (usd/month)
company	
Pink Cab	15059.047137
Yellow Cab	15045.669817



➔ There's no difference in the income (usd/month) by the two cab company.

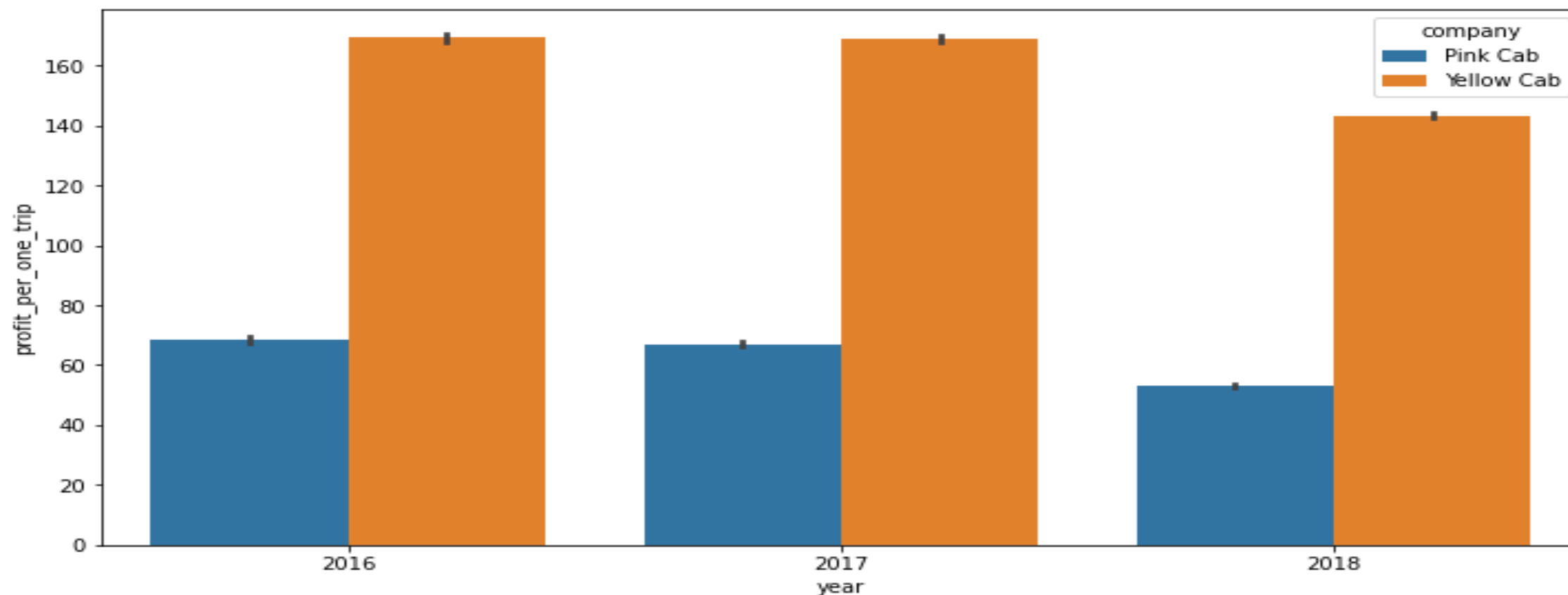
- Km travelled analysis

	Km travelled
company	
Pink Cab	22.559917
Yellow Cab	22.569517



➔ There's no difference in the KM travelled between Yellow and Pink Cab.

- Profit distribution by year and company



➔ Overall Yellow company profit per one trip is higher more than twice than Pink company in the three years.

- Customer per company analysis

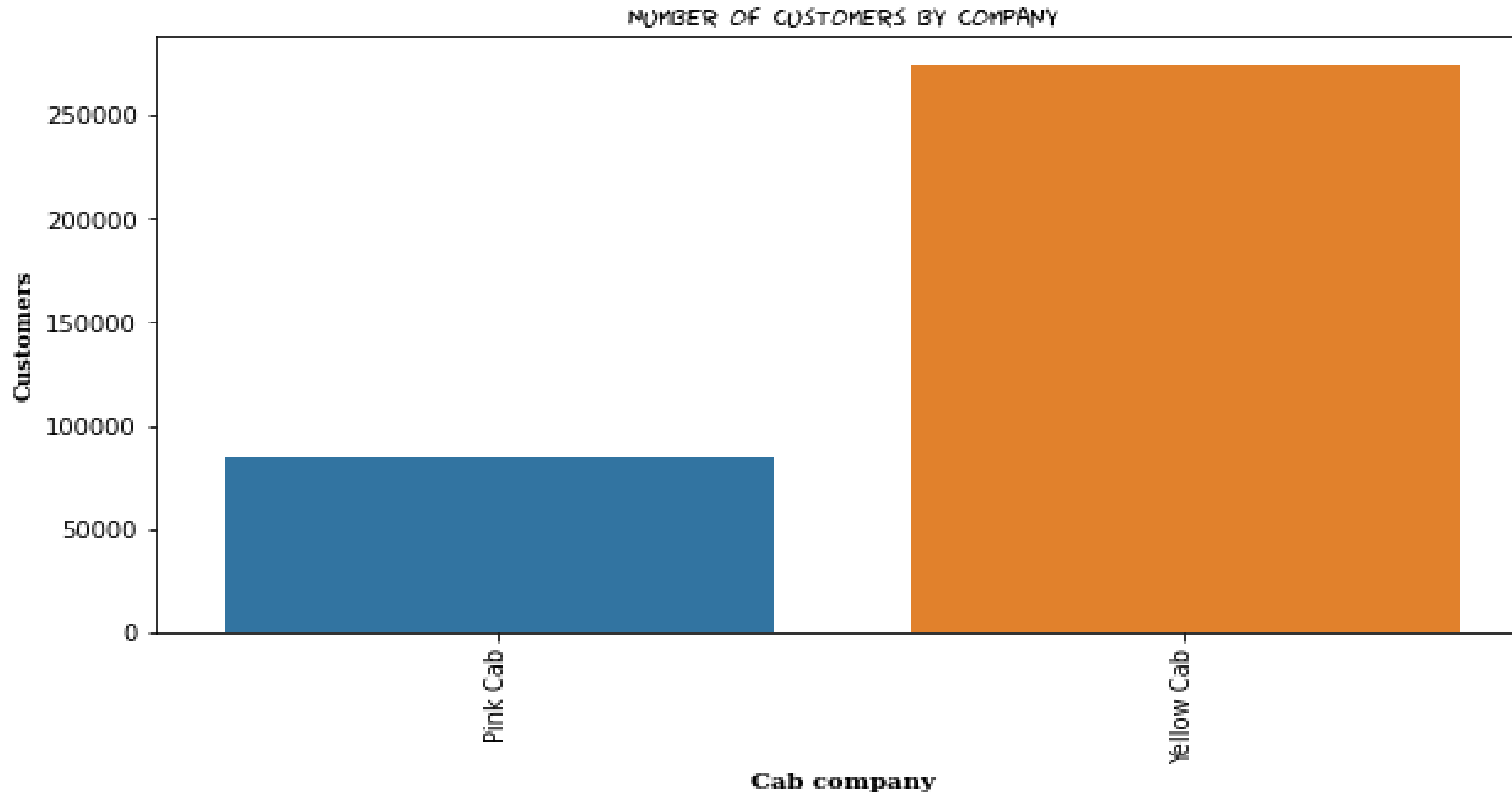
	customer_id	Size
company		
Pink Cab	831664558	84711
Yellow Cab	1049163053	274681

➔ Many customers choose to use Yellow cab company more to travel. the same customer sometimes travel with both companies.

➔ Yellow cab industry has the biggest number of customer.

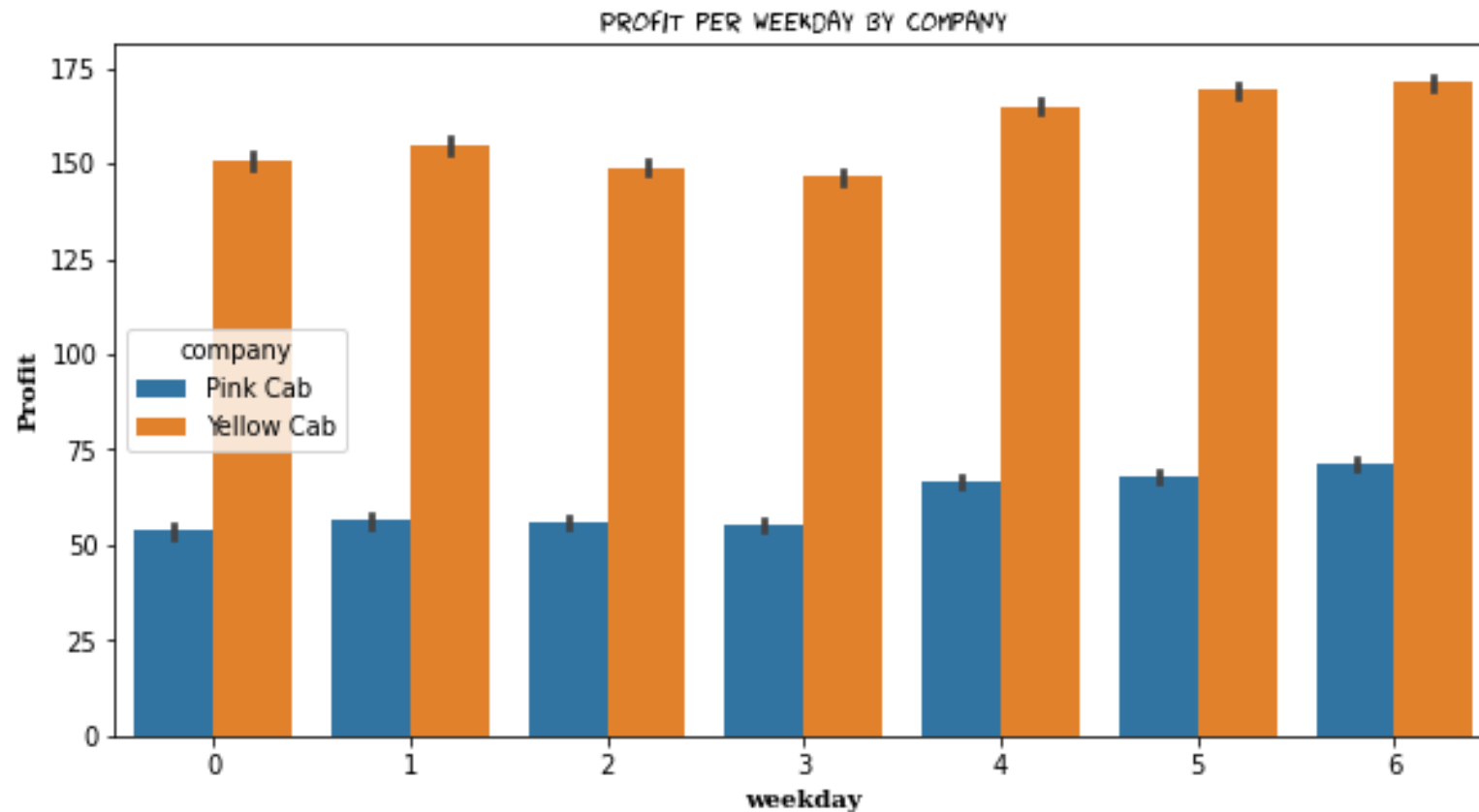
Customer_id	company	Size
1	Pink Cab	4
1	Yellow Cab	25
2	Pink Cab	4
2	Yellow Cab	36
3	Pink Cab	6
...
59998	Yellow Cab	6
59999	Pink Cab	2
59999	Yellow Cab	6
60000	Pink Cab	4
60000	Yellow Cab	14

- Customer per company analysis



➔ The highest number of customers in the cab industry is in the favor of Yellow cab

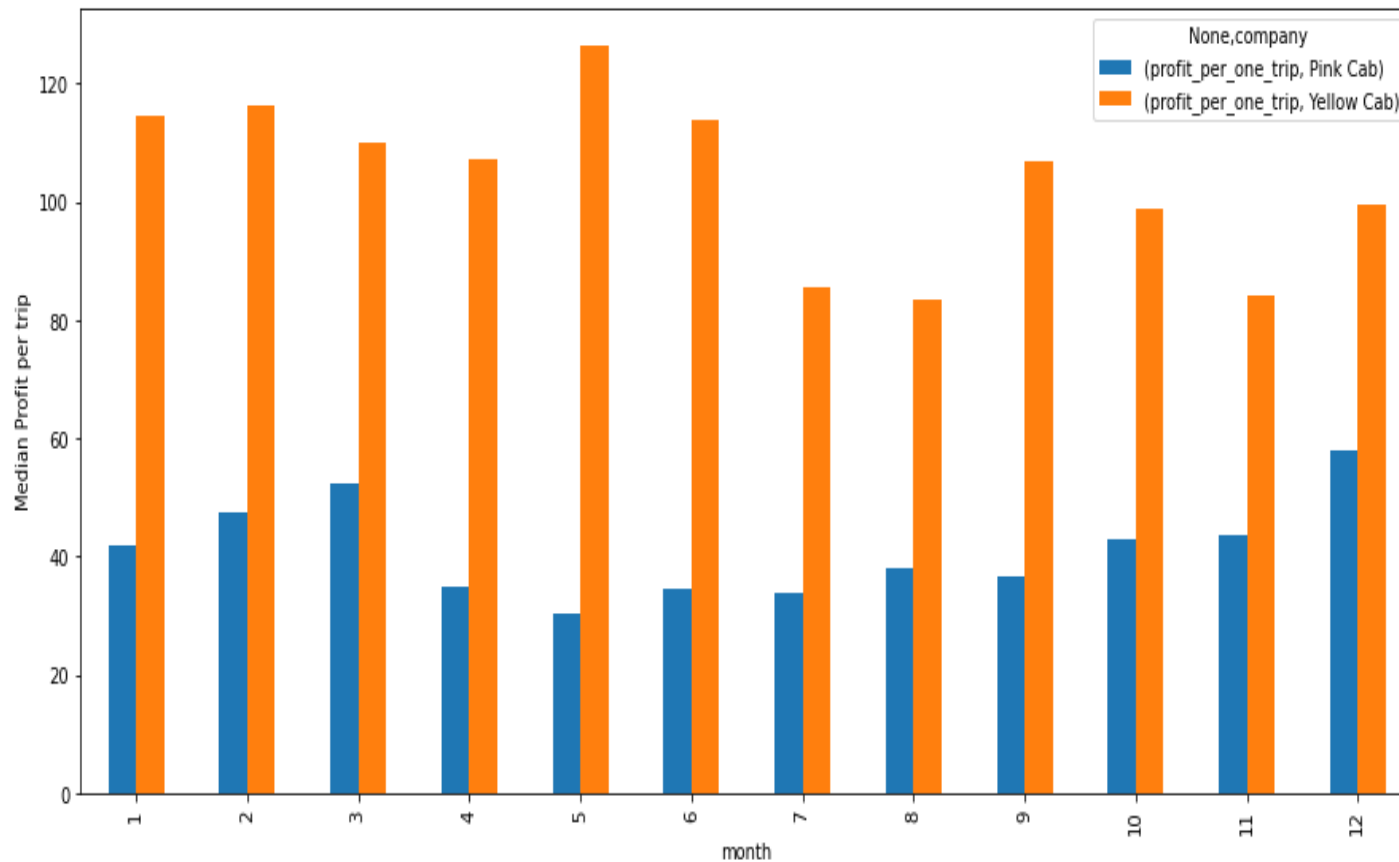
• Weekly analysis



➔ Profit per one trip increases in the weekends, especially, in Yellow cab. This increase is caused by the increase of kilometres travelled.

		km_travelled	profit_per_one_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

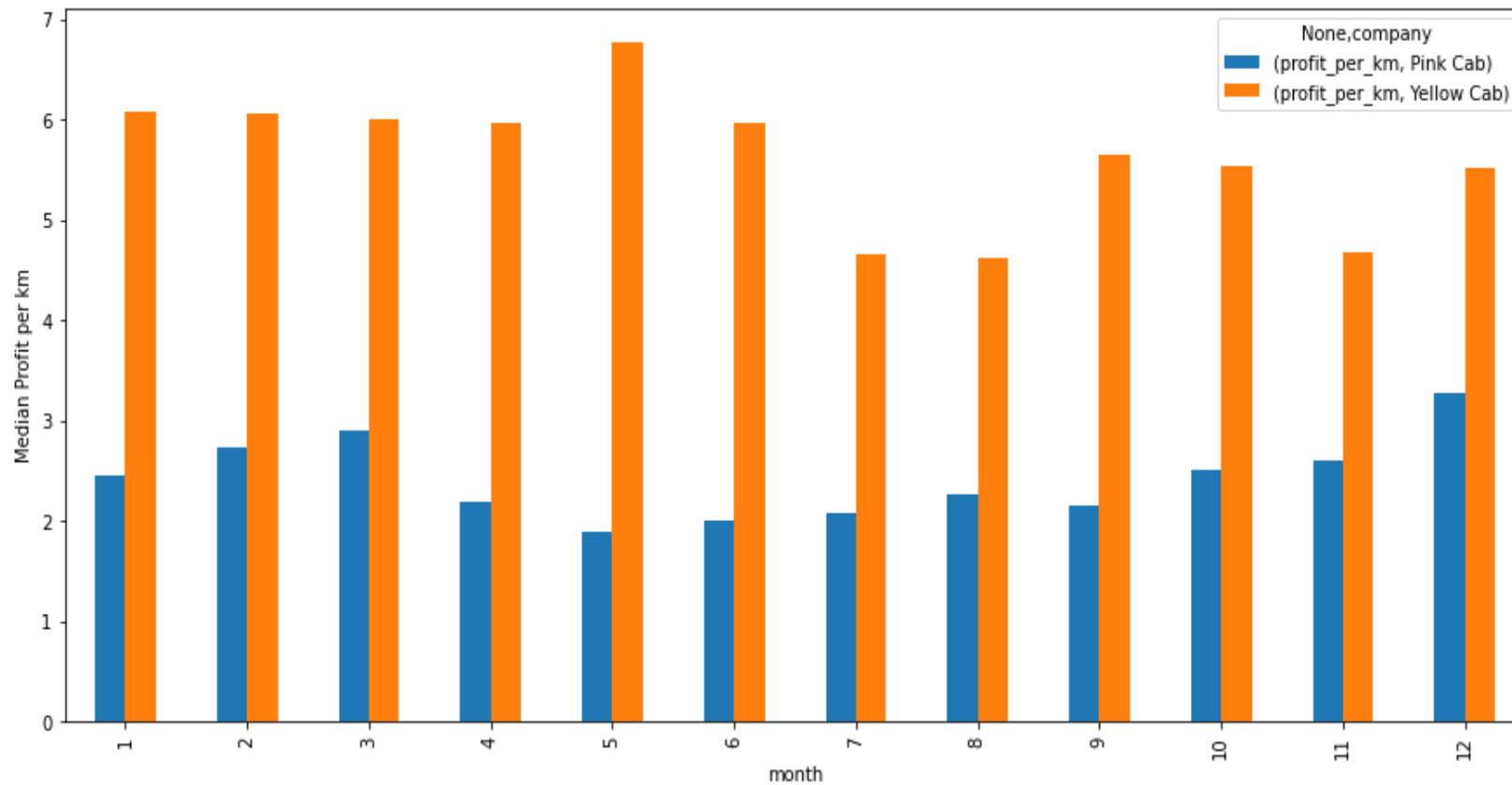
• Seasonality analysis



	km_travelled		profit_per_one_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

➔ The median profit per trip in each month for Yellow cab is always greater than Pink cab.

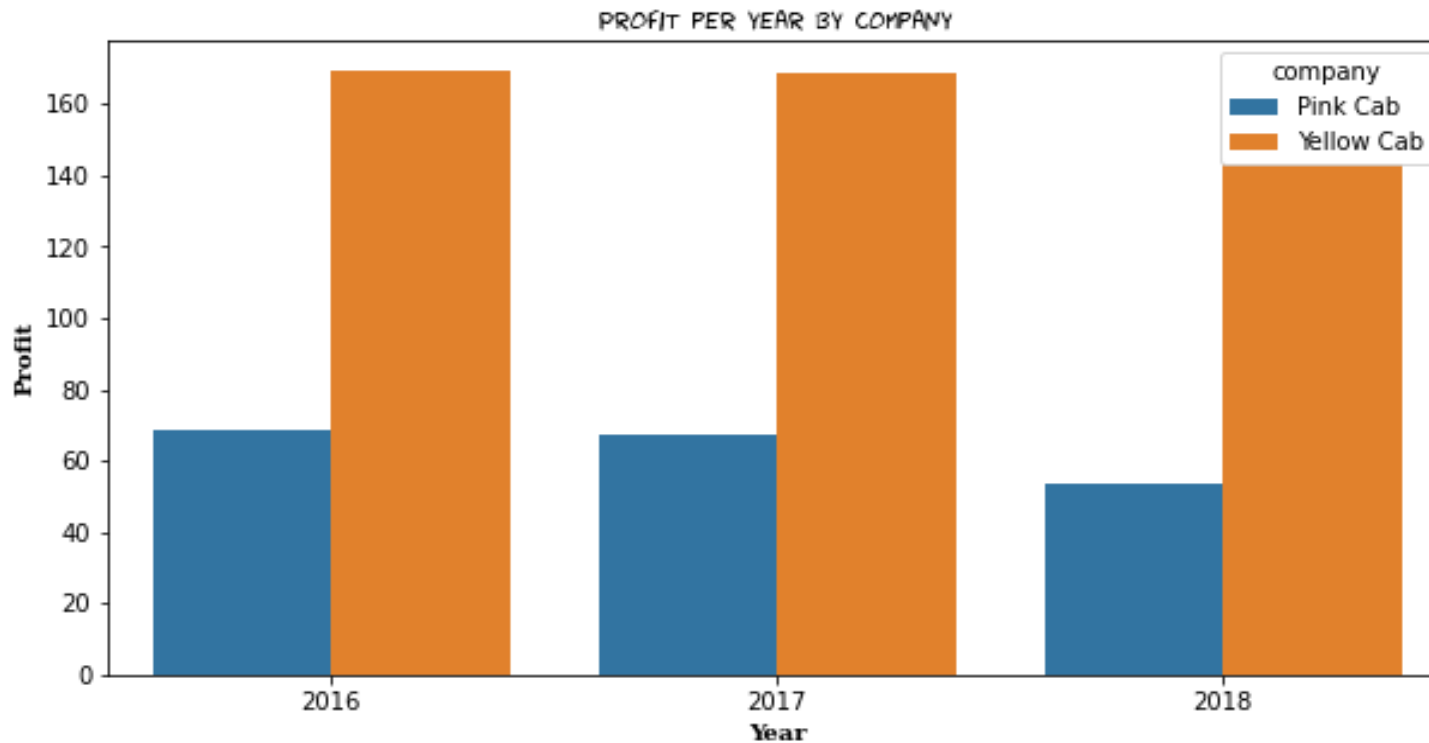
- Seasonality analysis



➔ The median profit per one trip and per km increase in the spring season especially in May.

➔ During July and August, both company's recorded a decrease in profit per km and profit per one trip. However, it is still higher than Pink Cab's profit. We can say that there's some seasonality

- Yearly analysis



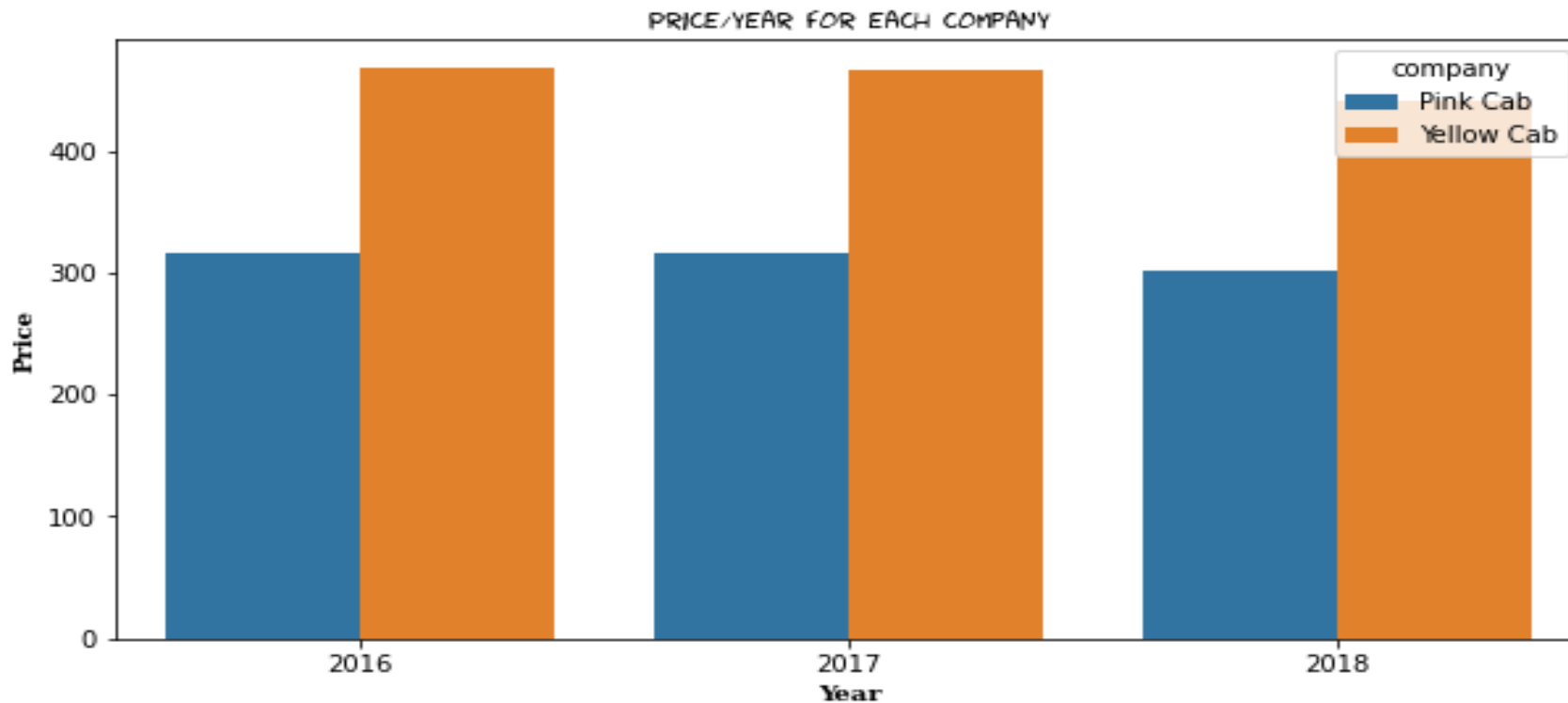
Mean profit per year by company

Year	Company	Profit per one trip
2016	Pink Cab	68.321819
2016	Yellow Cab	169.347821
2017	Pink Cab	67.070839
2017	Yellow Cab	168.817057
2018	Pink Cab	53.229689
2018	Yellow Cab	143.416122

➔ in 2018, profit margin has decreased in the Pink cab.

➔ Compared to Yellow cab profit, Pink cab profit is much less .

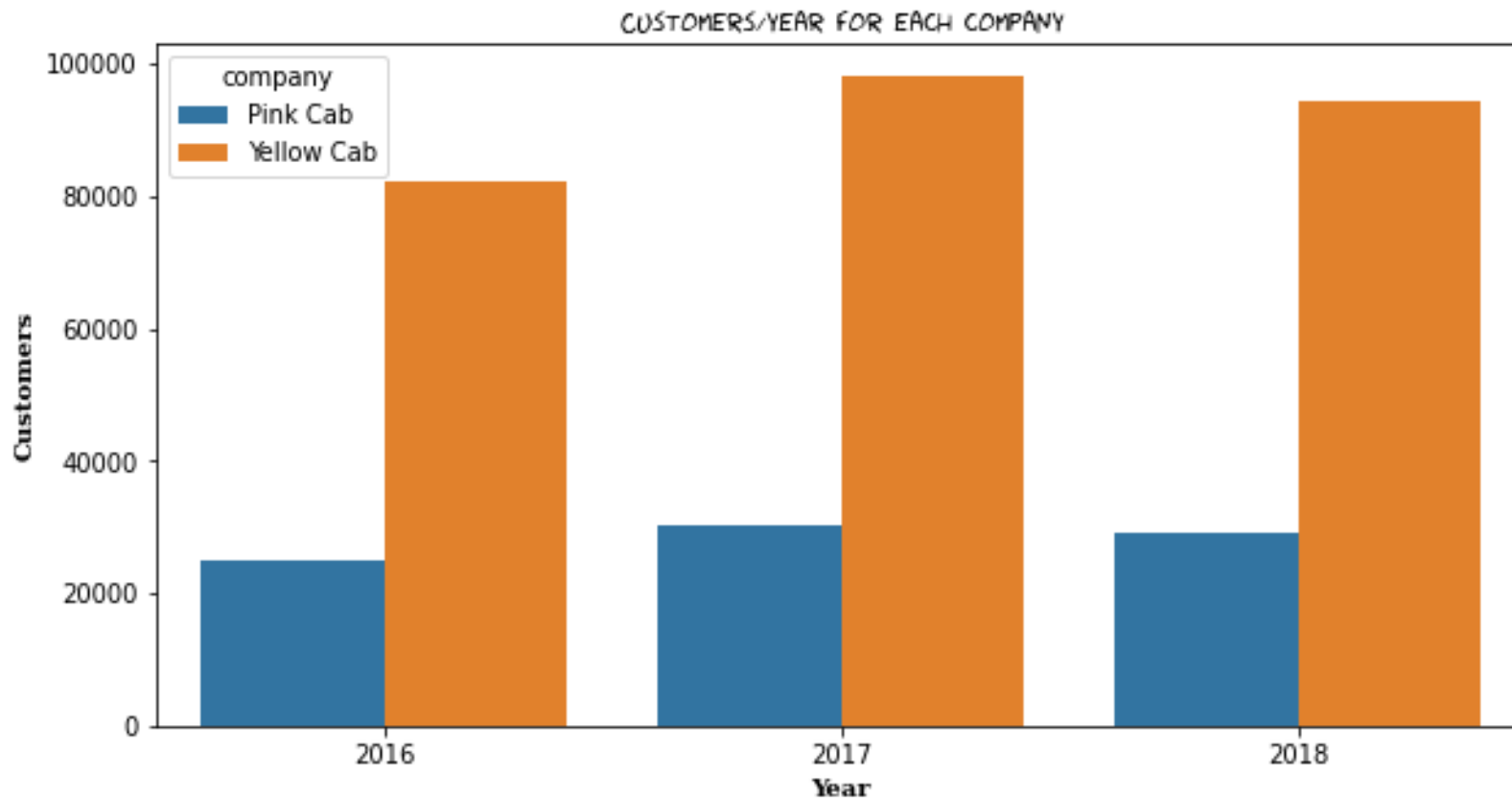
- Yearly analysis



Mean price per year by company

Year	Company	Price charged
2016	Pink Cab	315.330113
2016	Yellow Cab	467.918301
2017	Pink Cab	315.907442
2017	Yellow Cab	466.639950
2018	Pink Cab	301.642530
2018	Yellow Cab	440.875558

➔ Yellow Company has the highest price charged along three years.

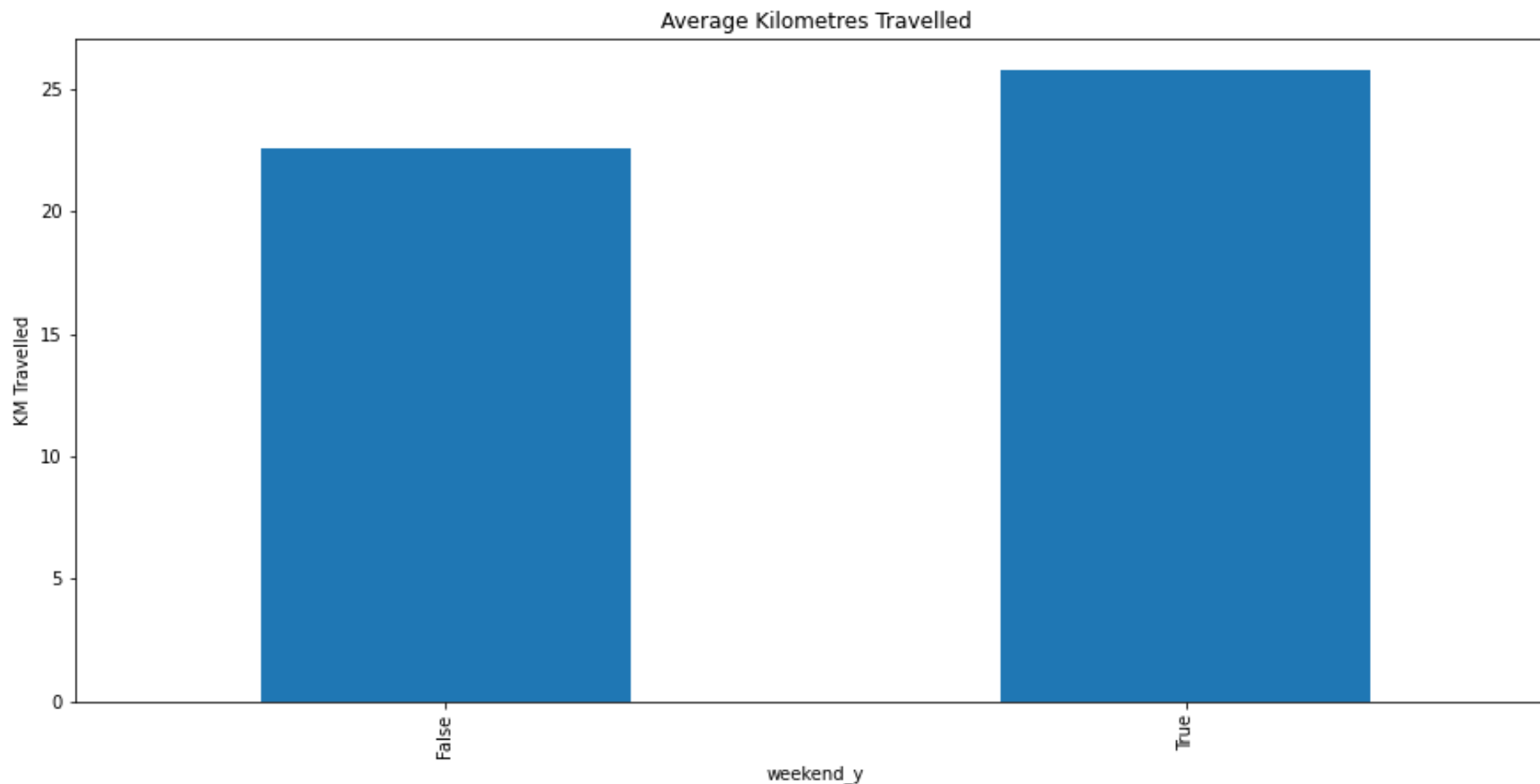


Cab user per year by company

Year	Company	Customer id
2016	Pink Cab	25080
2016	Yellow Cab	82239
2017	Pink Cab	30321
2017	Yellow Cab	98189
2018	Pink Cab	29310
2018	Yellow Cab	94253

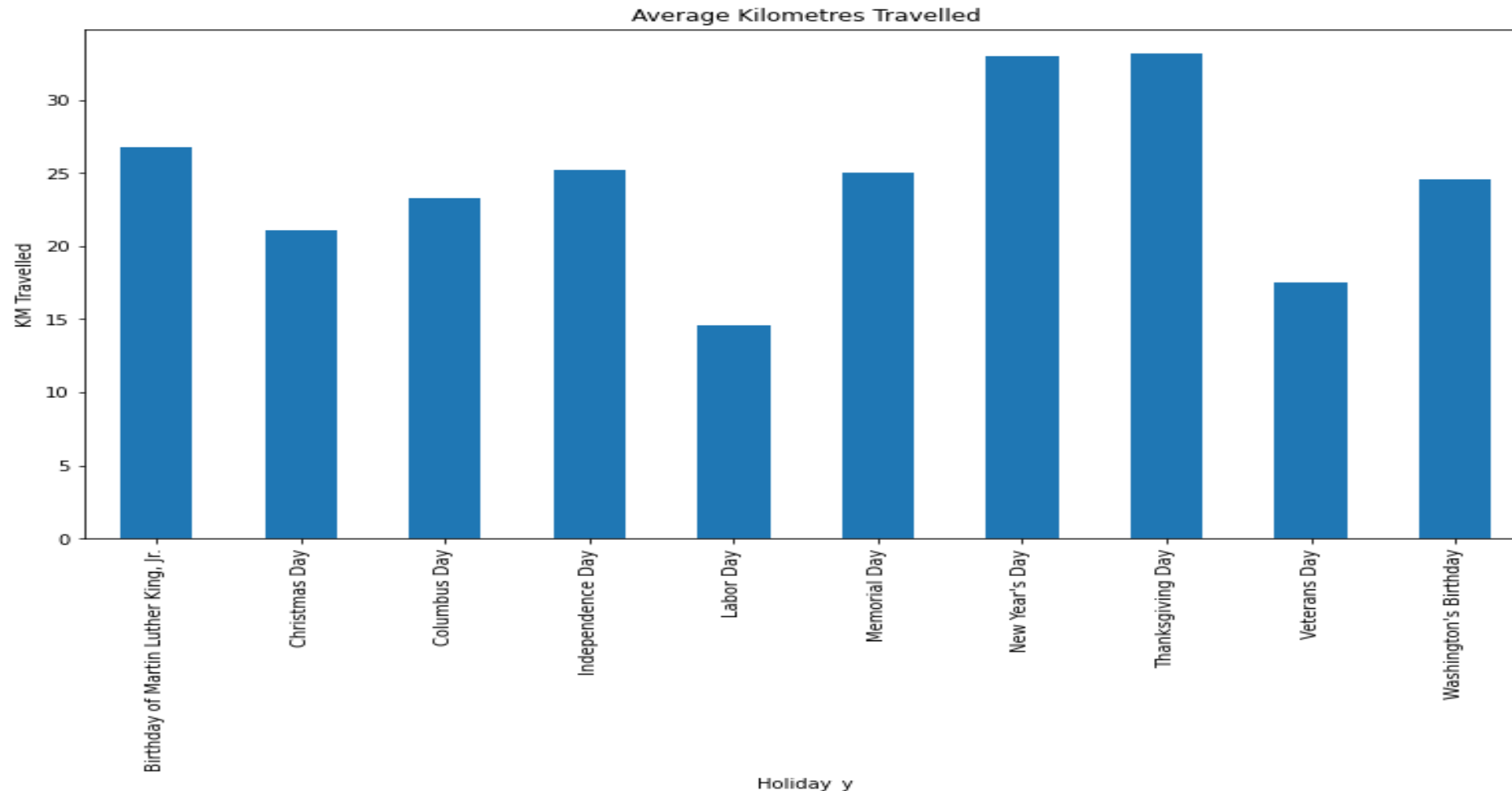
➔ Over years, Yellow company has the highest number of customers, especially in 2017.

- Average Kilometres Travelled in holidays, in weekends and weekdays



➔ In holidays, distance travelled for weekends is more big than weekdays.

- **Average Kilometres Travelled in holidays**



➔ The maximum average kilometres travelled is in Thanksgiving Day.

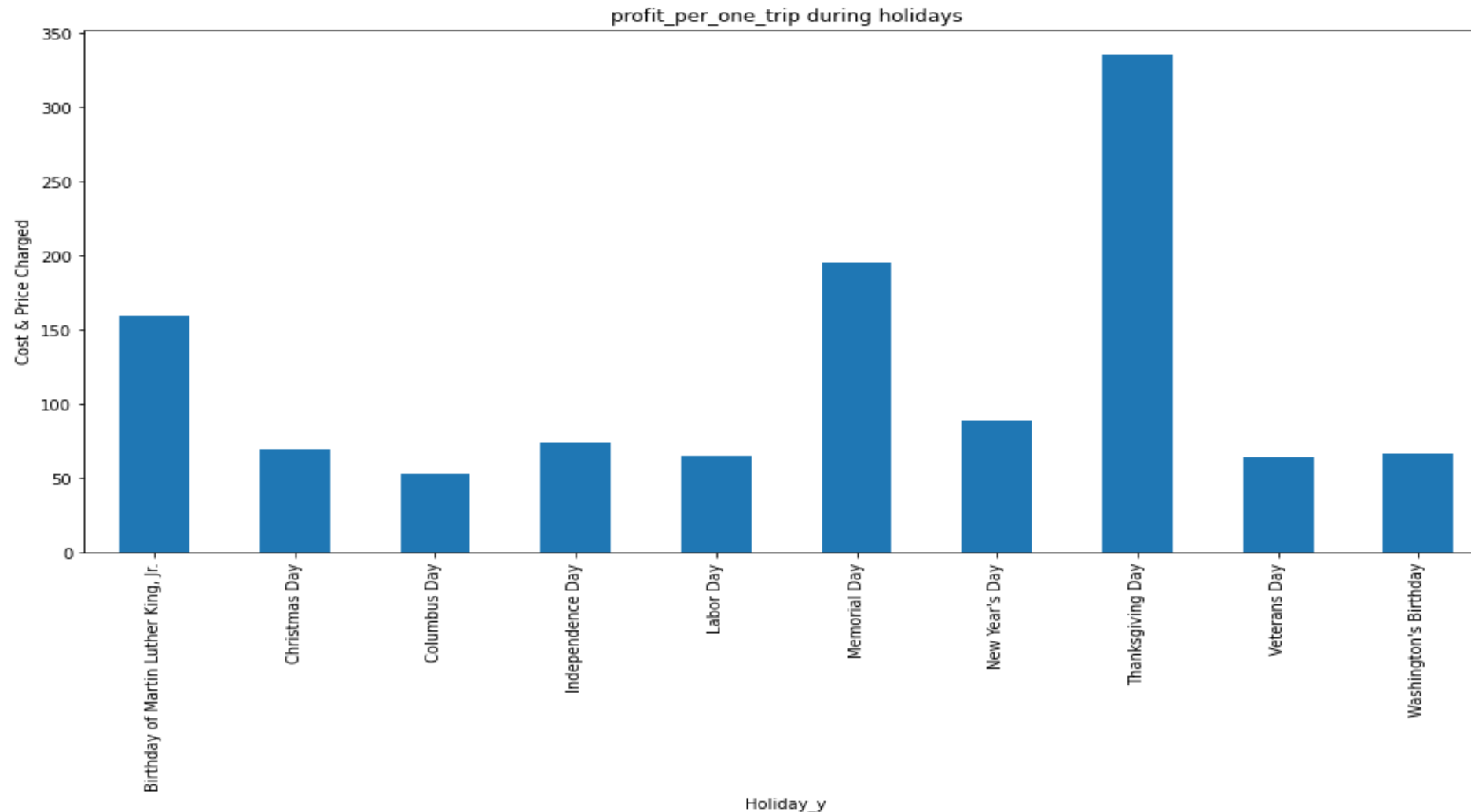
➔ The lowest km travelled is in the Labor day.

• Cost of trip and price Charged during holidays



➔ During Thanksgiving Day, price charged increase
➔ Cost of trip increase in new year's Day and it decrease in labor Day.

• Profit per one trip during holidays



- ➔ Over Thanksgiving Day, profit per one trip increase
- ➔ profit per one trip decrease in Columbus Day.



- After the follow-up of Exploring Data Analysis approach, we can say that Yellow Cab is preferred by most of the users in almost all the cities.
- Most of the users are Young and Middle-aged.
- Over years, Yellow company has the highest profit margin
- There's some seasonality in the profit.



Investing in Yellow Cab will be the best decision by following these points:

- **Yellow cab is the superior in terms of rides and profits. That's mean, that most rides are made with Yellow taxis.**
- **Yellow Cab has the biggest share of the profits.**
- **Over years, Yellow company has the highest number of customers in almost cities, especially in the populated cities . However, Pink cab has the highest number of customer in only four cities which are, San Diego, Nashville, Pittsburgh and Sacramento.**
- **Most of the customers for both Yellow cab and Pink cab are Young and Middle-aged.**
- **Weekly, monthly and annual profit of Yellow Cab is always much higher than Pink Cab.**
- **During July and August, both company's recorded a decrease in profit per km and profit per one trip. We can say that there's some seasonality.**
- **Payment mode of most of the customers of Yellow Cab is card. For Pink Cab, both card & cash payments are used.**

Thank You