



Data Glacier

Your Deep Learning Partner

G2M Case Study

Virtual Internship

27-Jun-2021

Agenda

Executive Summary

Problem Statement

Approach

EDA

EDA Summary

Recommendations

Background –G2M(cab industry) case study

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

The analysis has been divided into four parts:

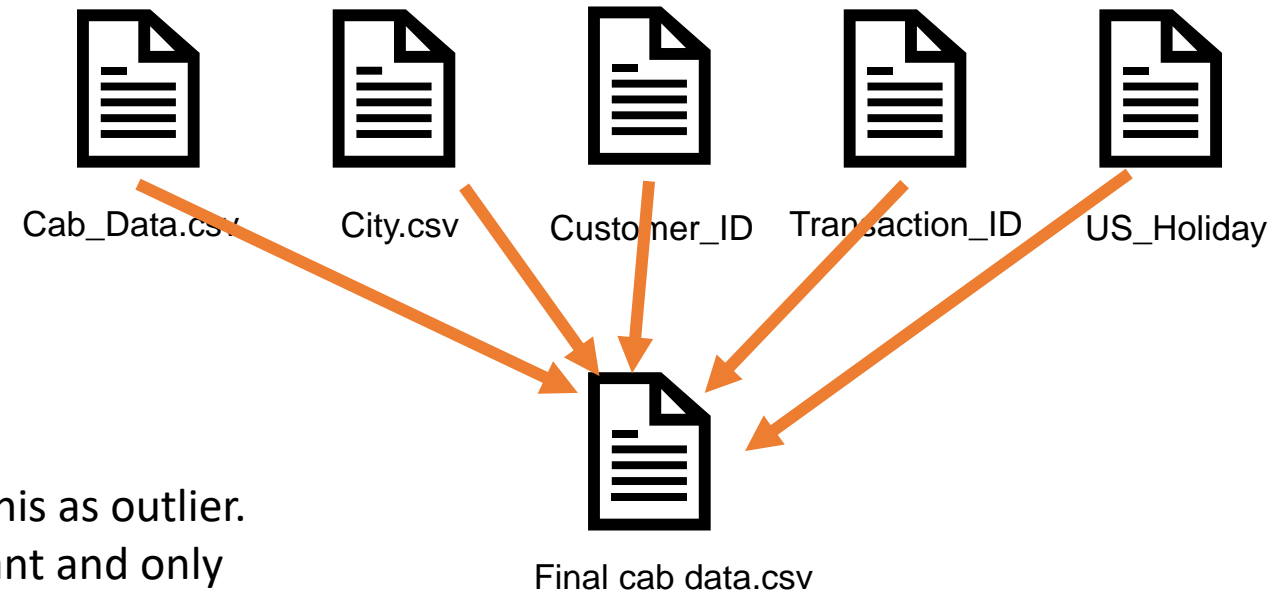
- Data Understanding
- Forecasting profit and number of rides for each cab type
- Finding the most profitable Cab company
- Recommendations for investment

Data Exploration

- 20 Features(including 5 derived features)
- Timeframe of the data: 2016-01-1 to 2018-12-31
- Total data points :359,392

Assumptions:

- Outliers are present in Price_Charged feature but due to unavailability of trip duration details ,we are not treating this as outlier.
- Profit of rides are calculated keeping other factors constant and only Price_Charged and Cost_of_Trip features used to calculate profit.
- Users feature of city dataset is treated as number of cab users in the city.



Tables

Company	Profit	Total Rides	Profit per Ride
Pink Cab	5,307,328	84,711	62.652
Yellow Cab	44,020,373	274,681	160.259

Profit Table

Age	18-25	26-40	41-60	60+
2016	0.257	0.474	0.213	0.055
2017	0.266	0.470	0.211	0.053
2018	0.259	0.473	0.214	0.054

Age Table
Pink Cab

Age	18-25	26-40	41-60	60+
2016	0.258	0.475	0.215	0.052
2017	0.260	0.471	0.217	0.052
2018	0.260	0.472	0.215	0.052

Age Table
Yellow Cab

Gender	Male	Female
2016	0.566	0.433
2017	0.556	0.444
2018	0.552	0.448

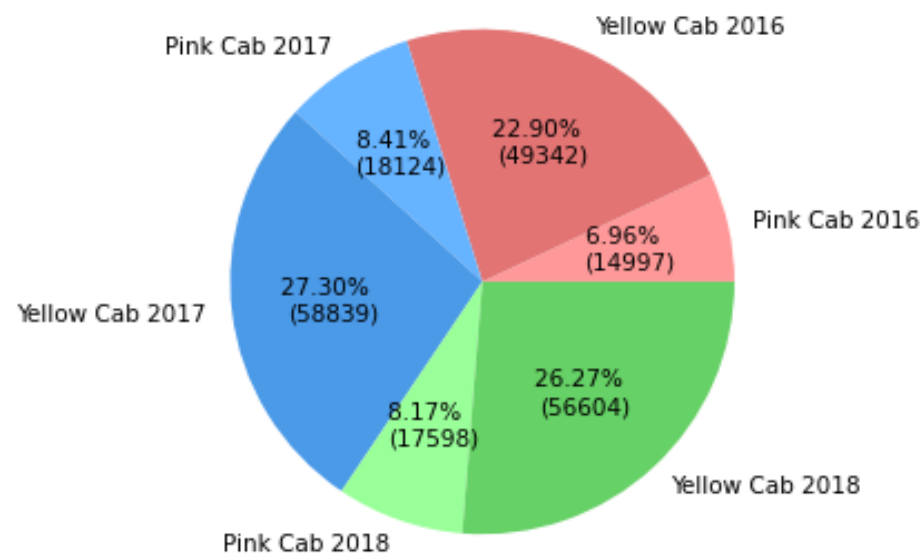
Gender Table
Pink Cab

Gender	Male	Female
2016	0.588	0.412
2017	0.574	0.426
2018	0.572	0.428

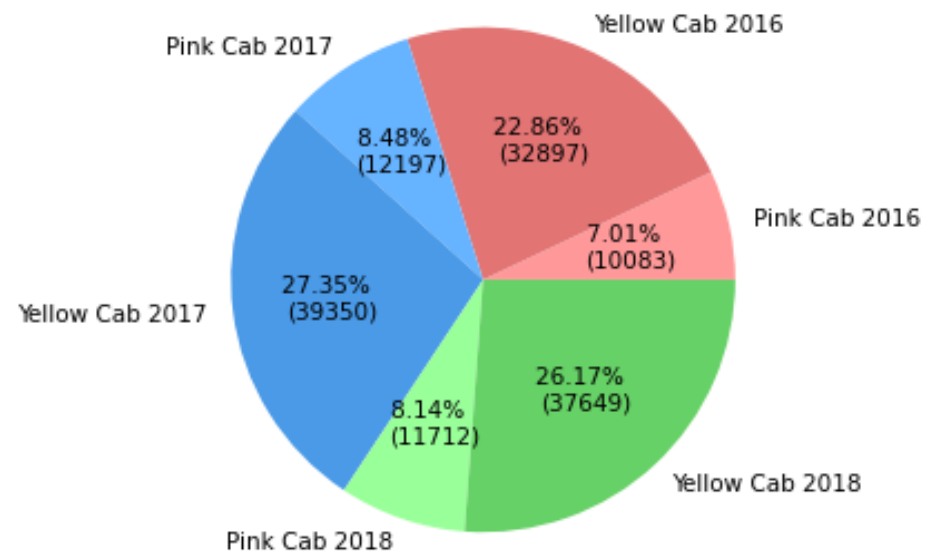
Gender Table
Yellow Cab

Payment

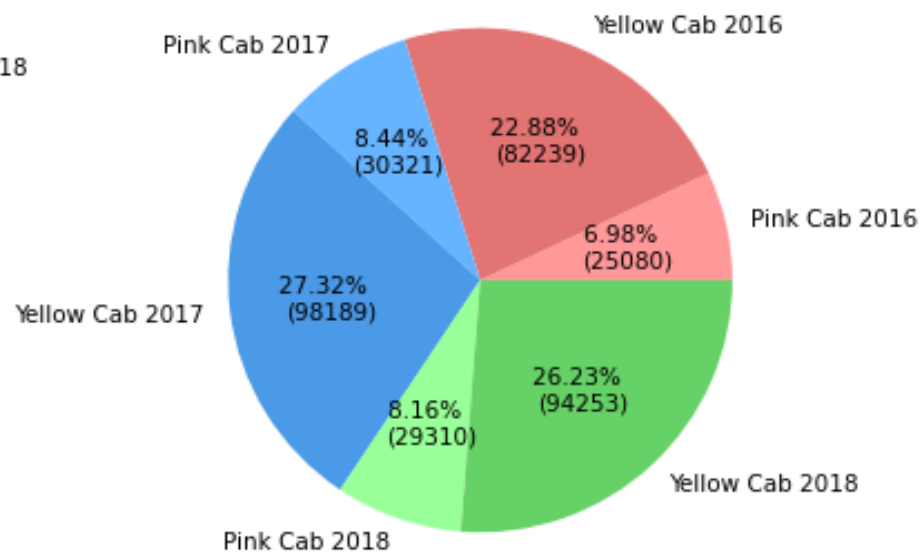
Payment State Percentages (Card)



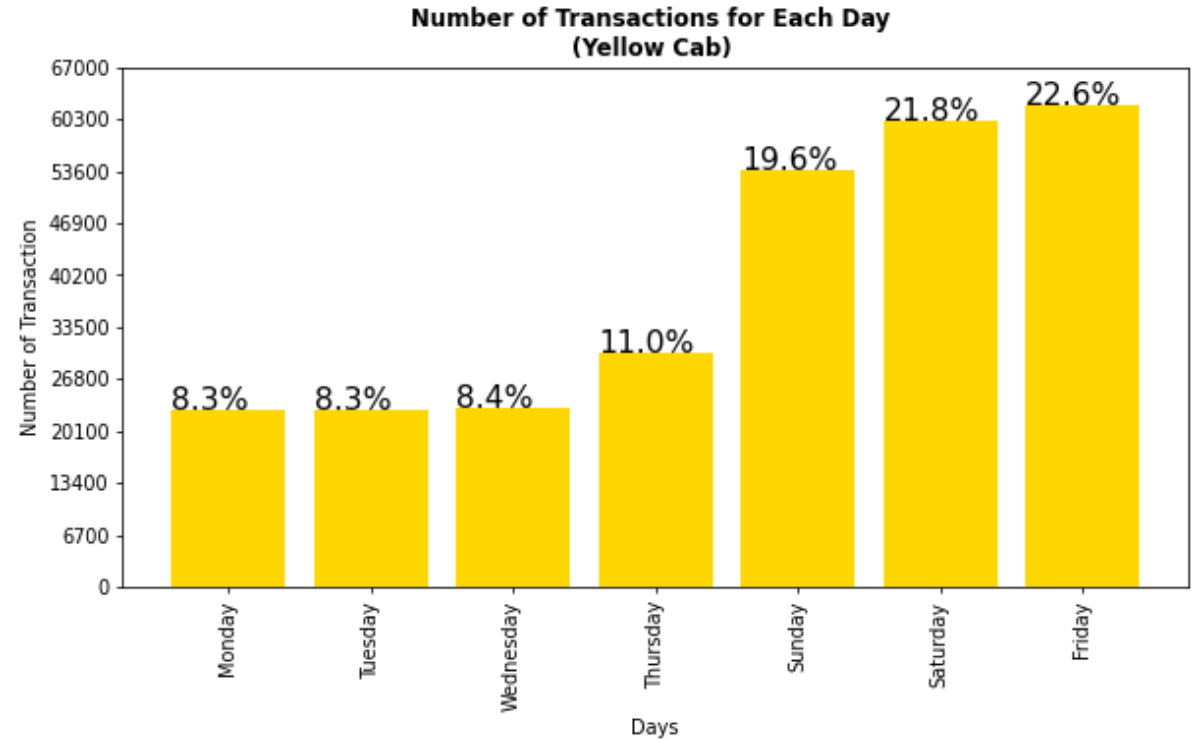
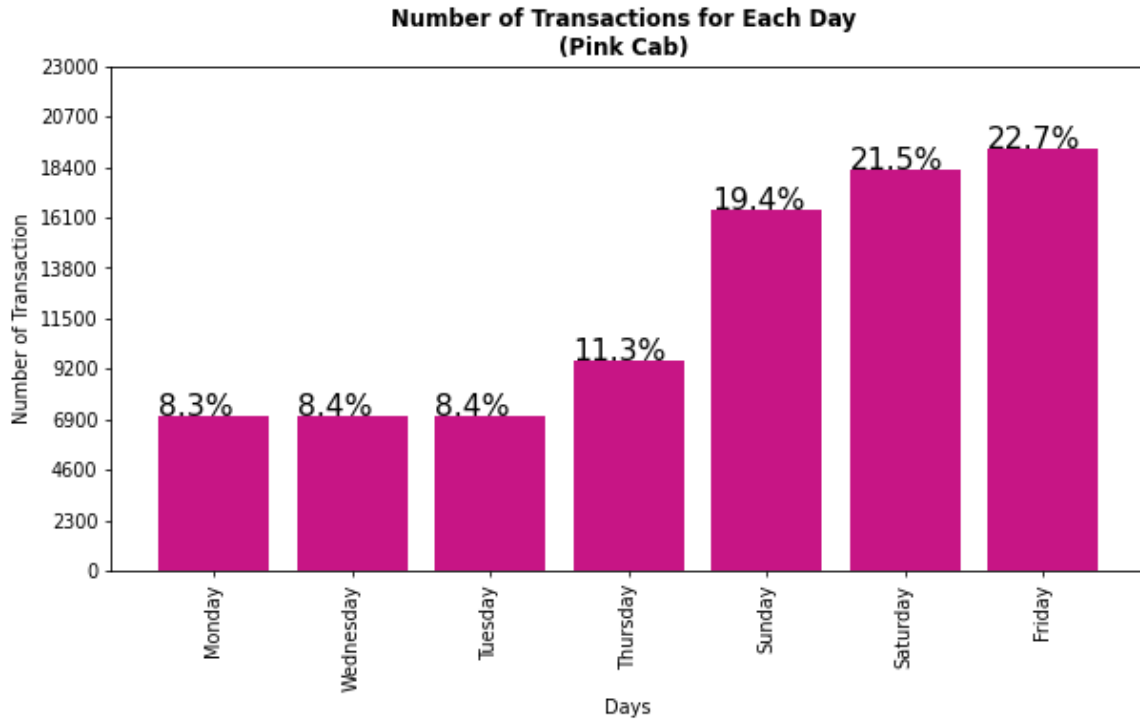
Payment State Percentages (Cash)



Payment State Percentages (All)



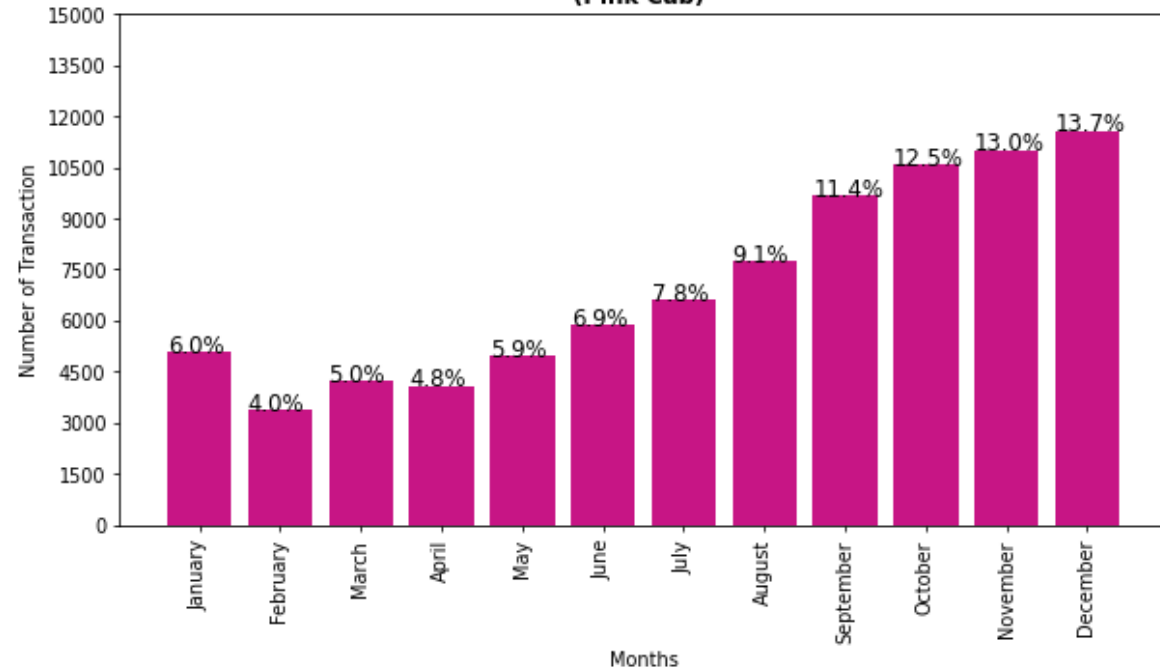
Transactions in Weekdays



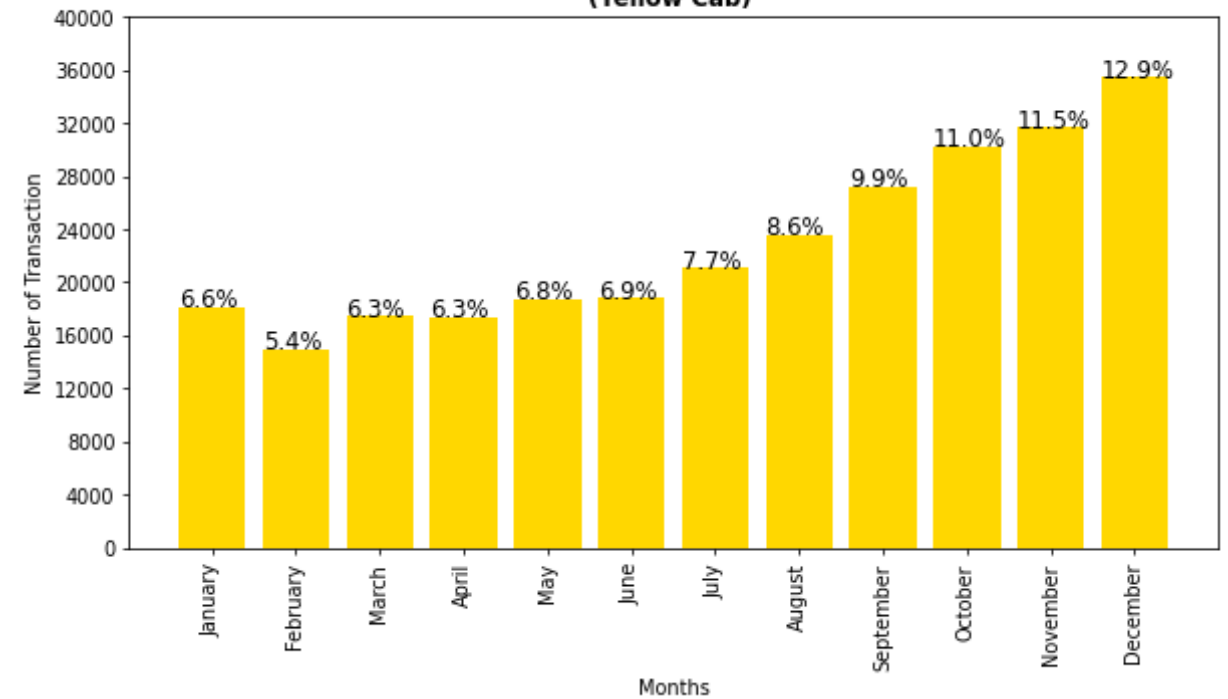
Both Cabs have the most transactions on Fridays and Weekends days.

Transactions in Each Month

Number of Transactions during each month
(Pink Cab)



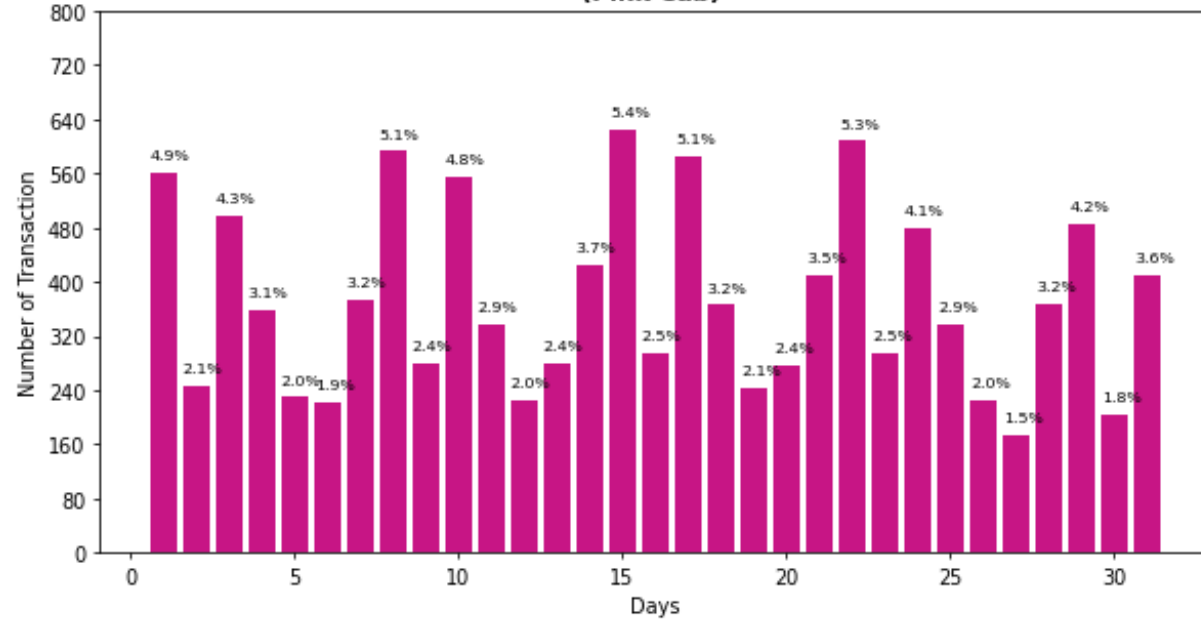
Number of Transactions during each month
(Yellow Cab)



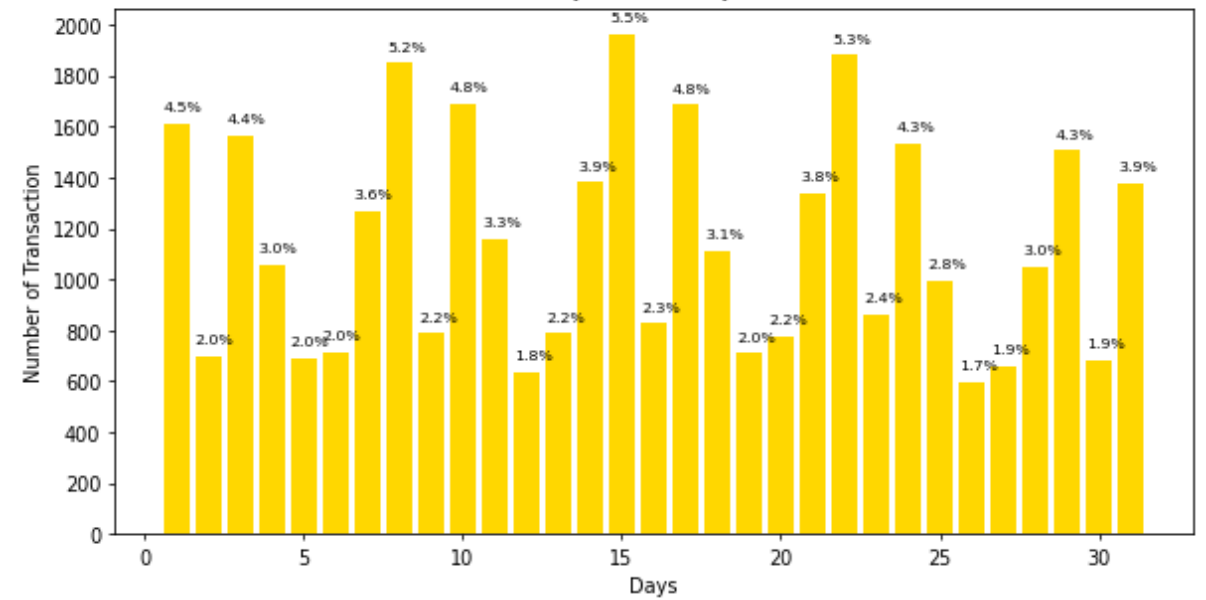
Both Cabs have the most transactions in December.

Transactions in December

Number of Transactions during December
(Pink Cab)



Number of Transactions during December
(Yellow Cab)

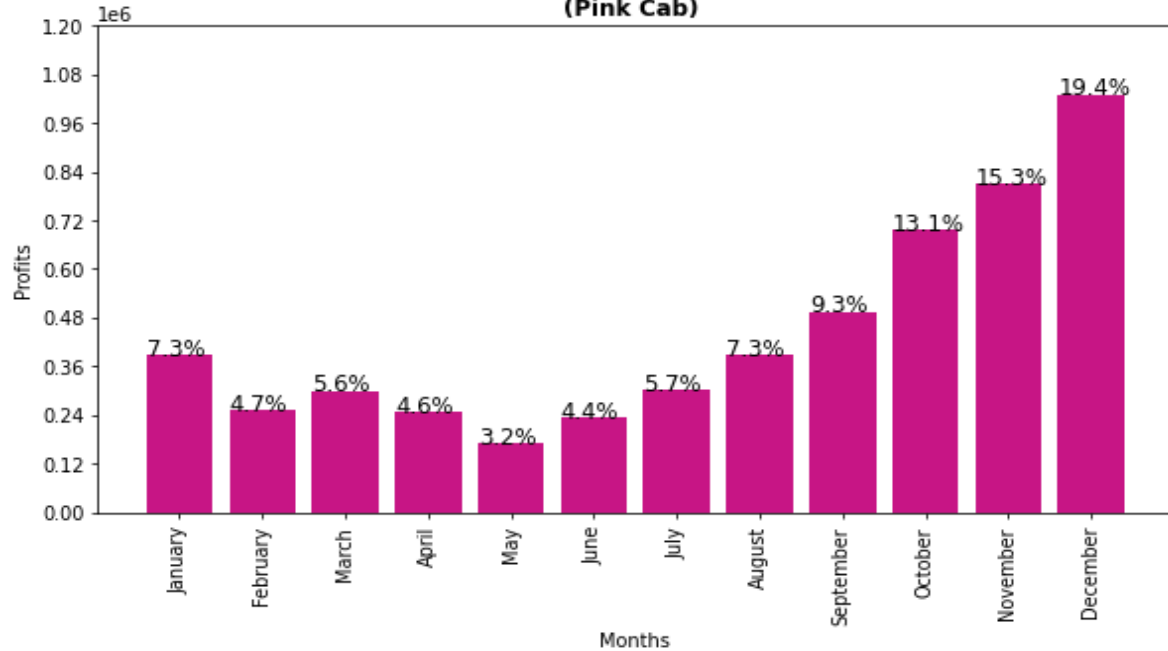


Both Cabs have the most transactions on the 15th day.

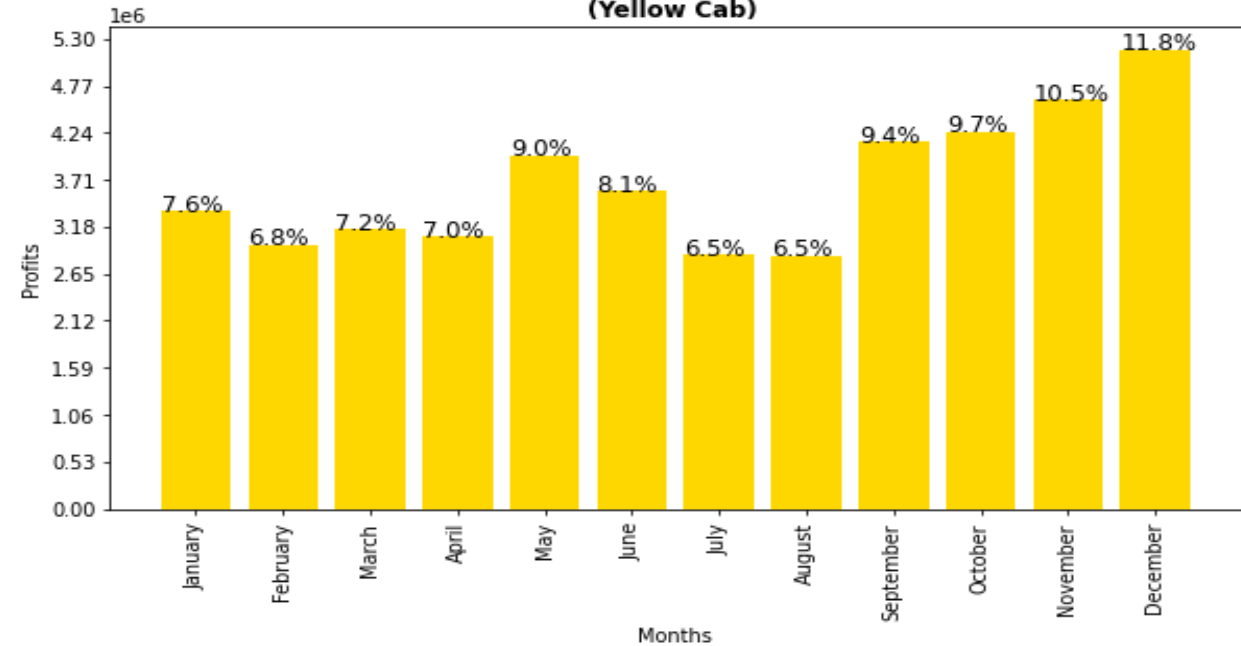
We can see that every 7 days there is increase in transaction numbers.

Profits for Each Month During The Year

Amount of Profits for each month (3 years)
(Pink Cab)



Number of Profits for each month (3 years)
(Yellow Cab)

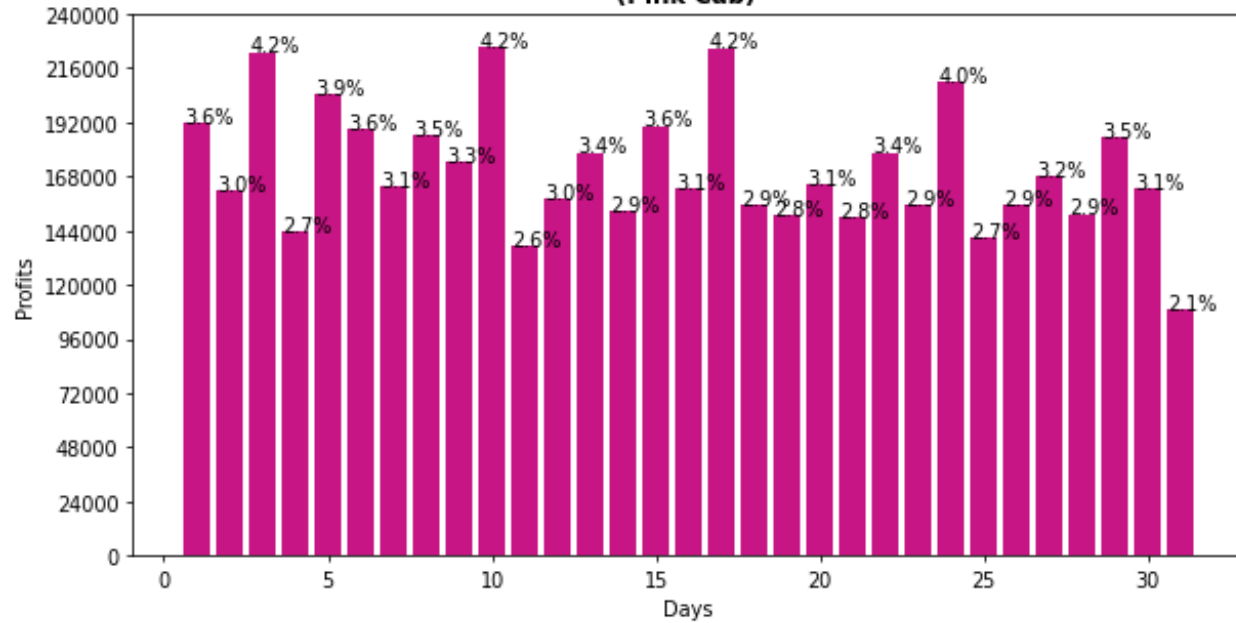


Both Cabs have the highest profits in December.

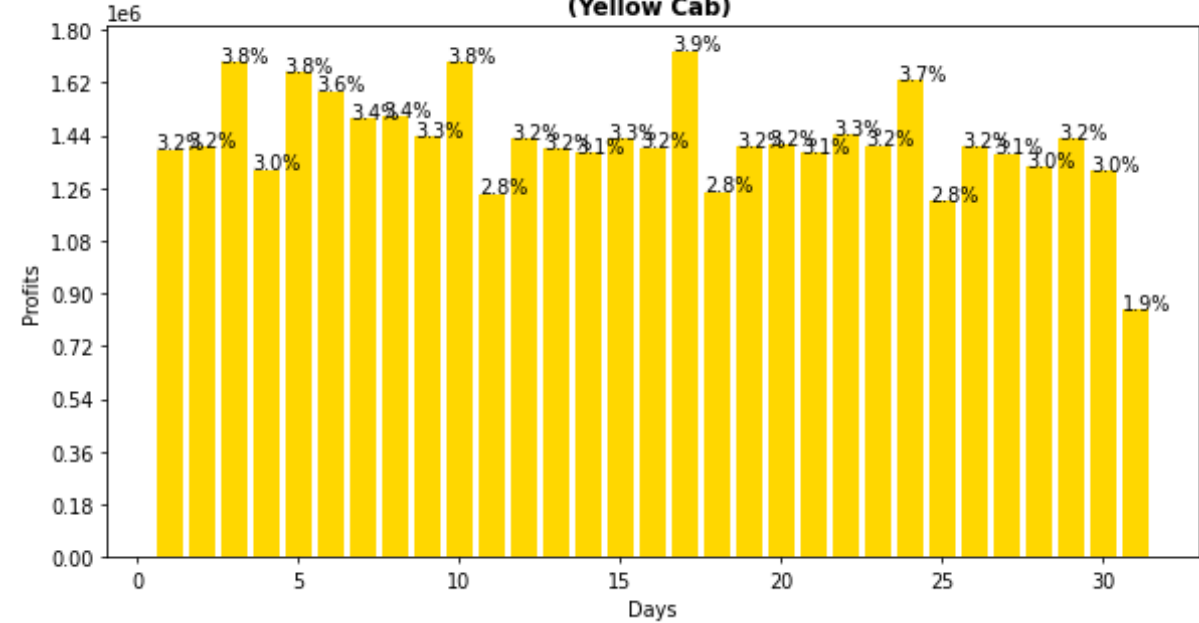
We learned: Most Transactions = Highest Profits.

Profits for Each Day During The Year

Amount of Profits for each day (3 years)
(Pink Cab)



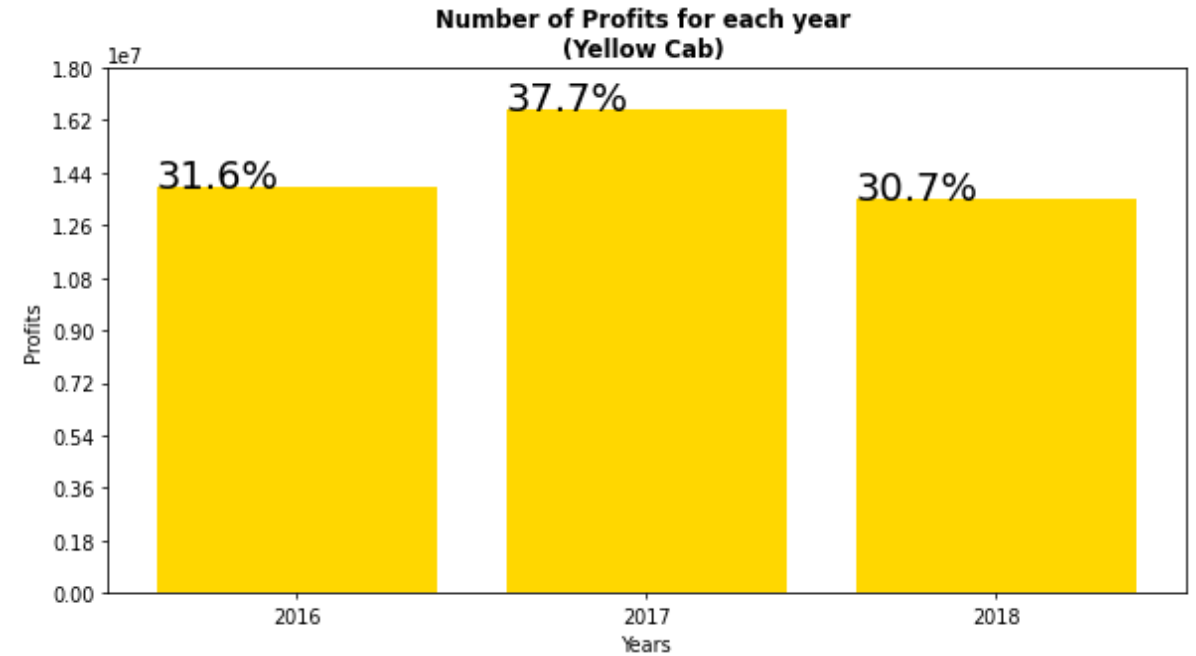
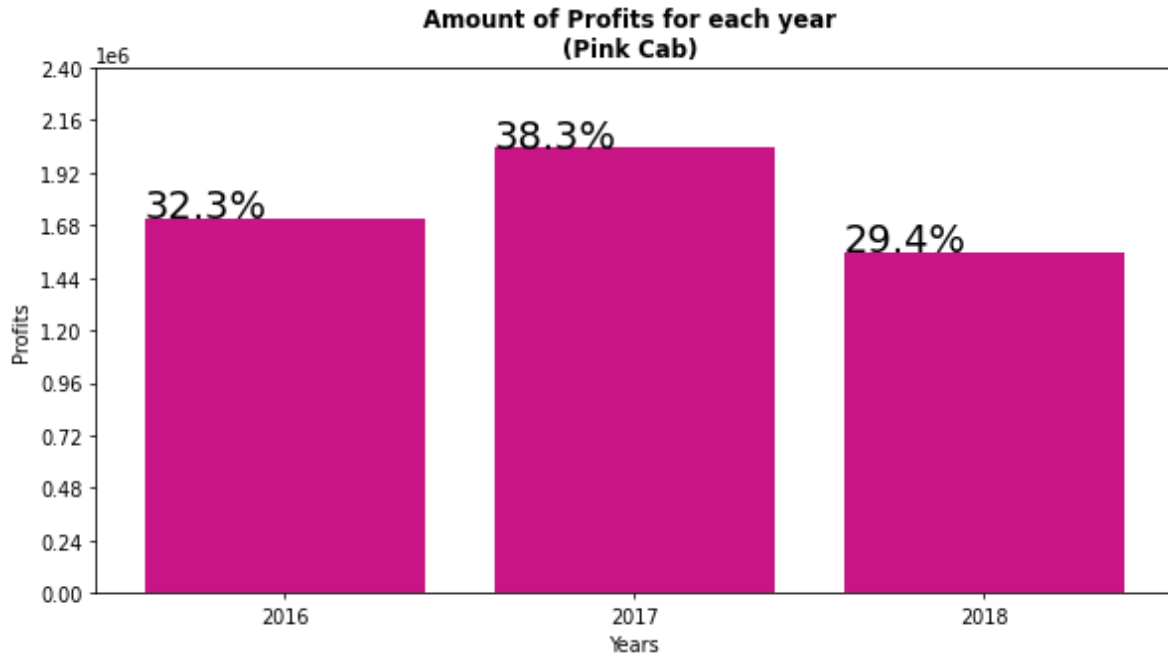
Number of Profits for each day (3 years)
(Yellow Cab)



Both Cabs have the most transactions on the 17th day.

Both Cabs have the highest profits in the days 3, 10, 17, and 24.

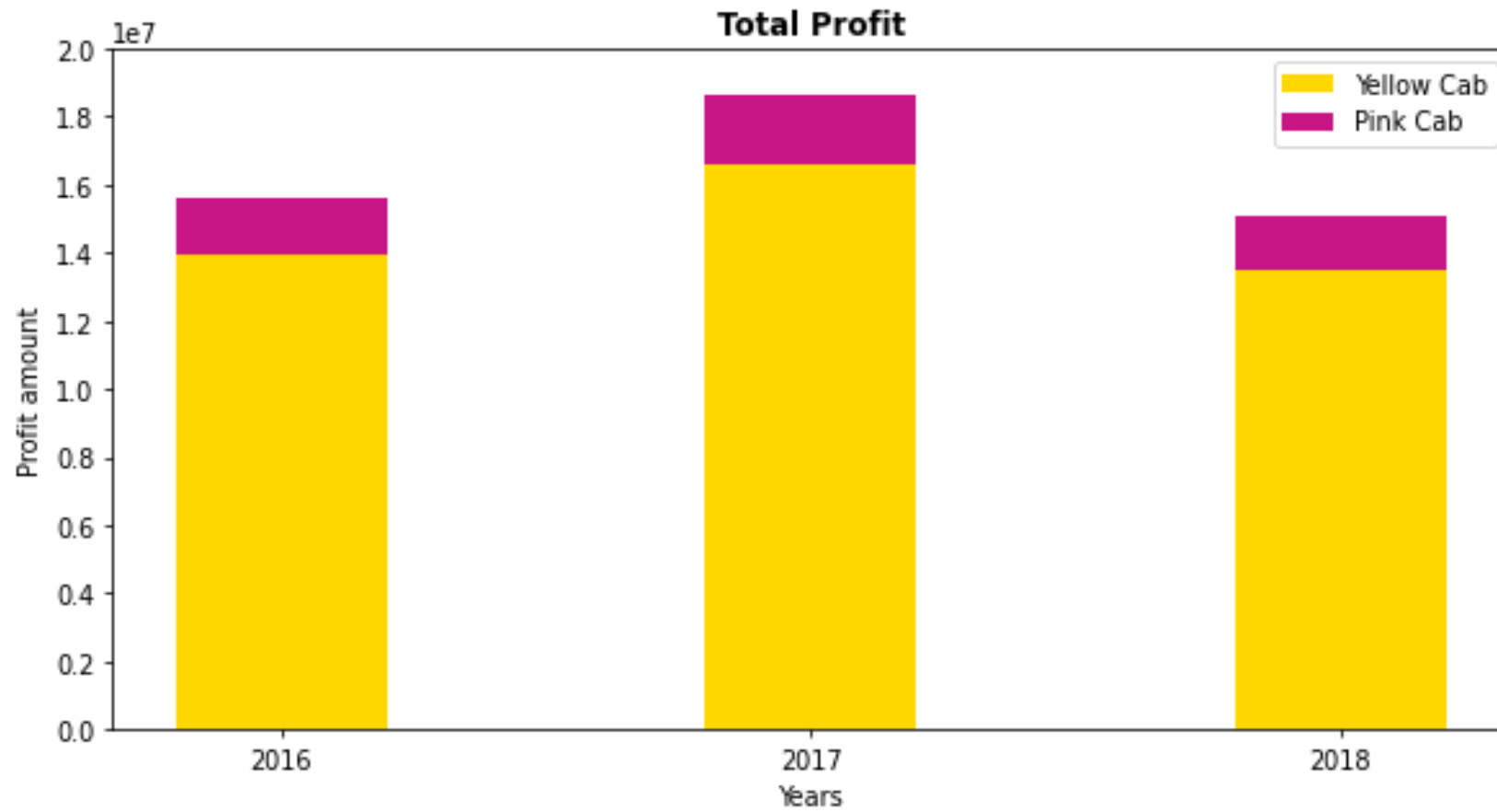
Profits in Each Year



Both Cabs have the highest profits in 2017.

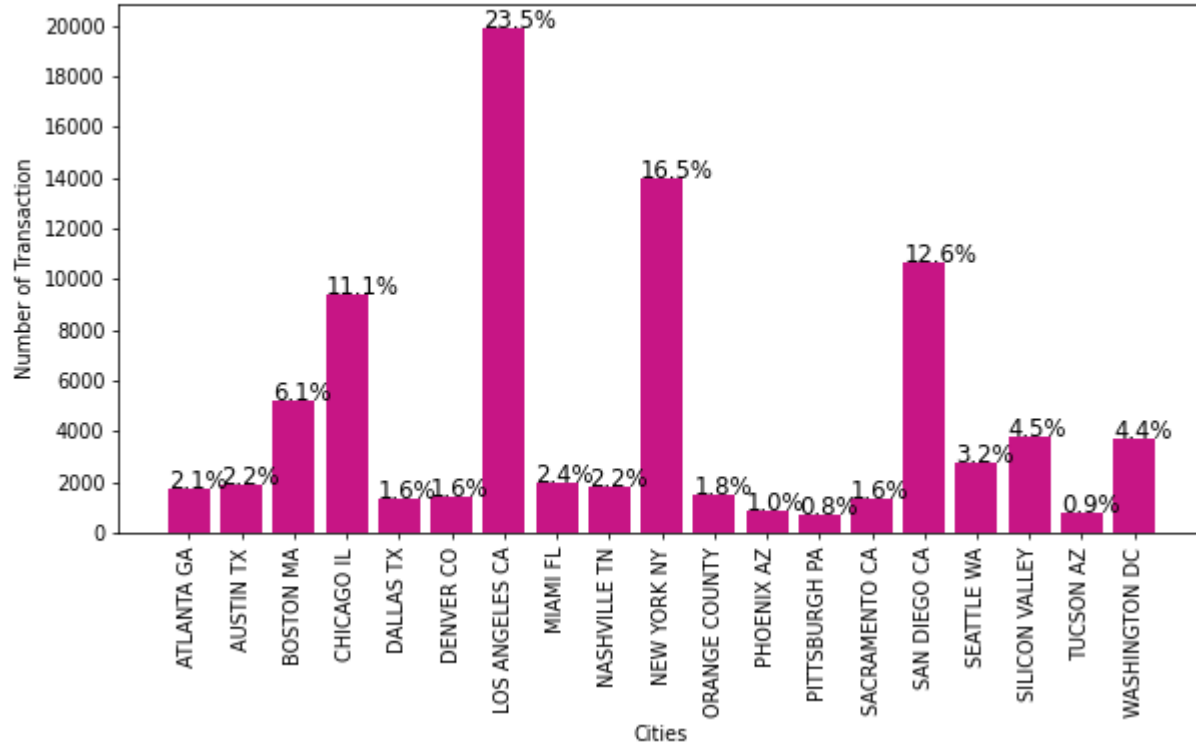
Both Cabs have the lowest profits in 2018.

Total Profit

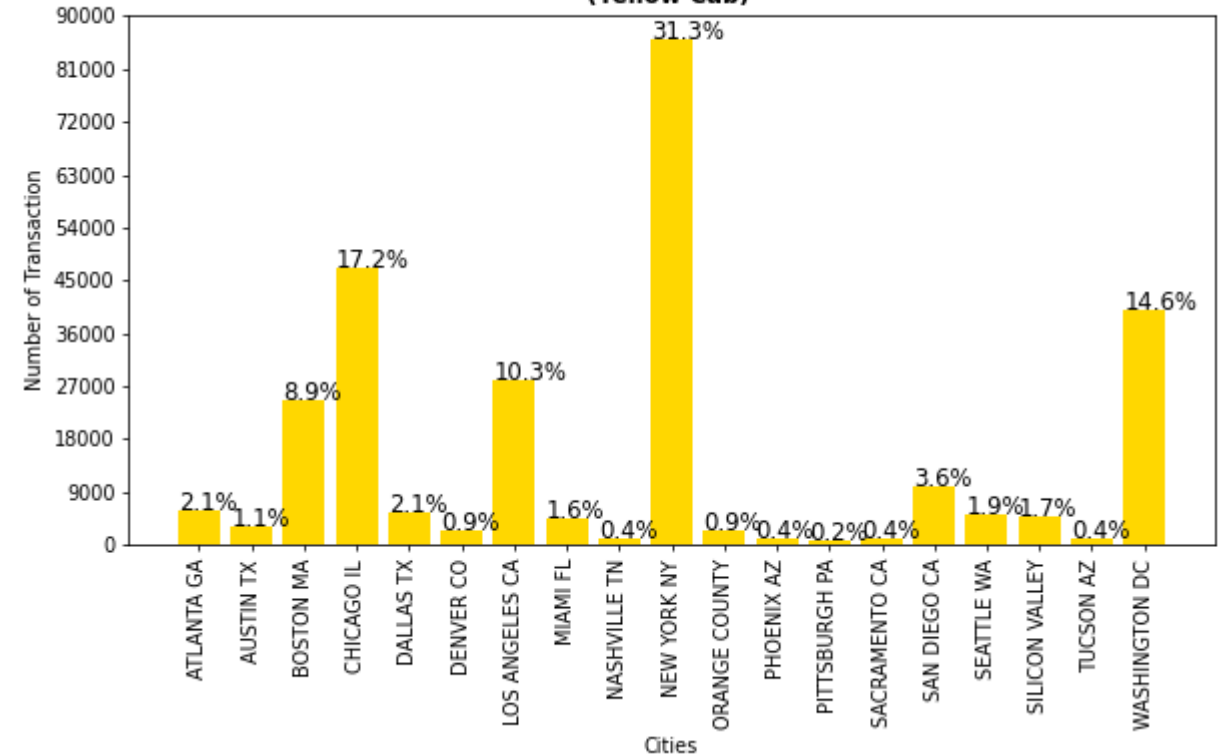


Transactions in Each City

Number of Transactions in Each city
(Pink Cab)



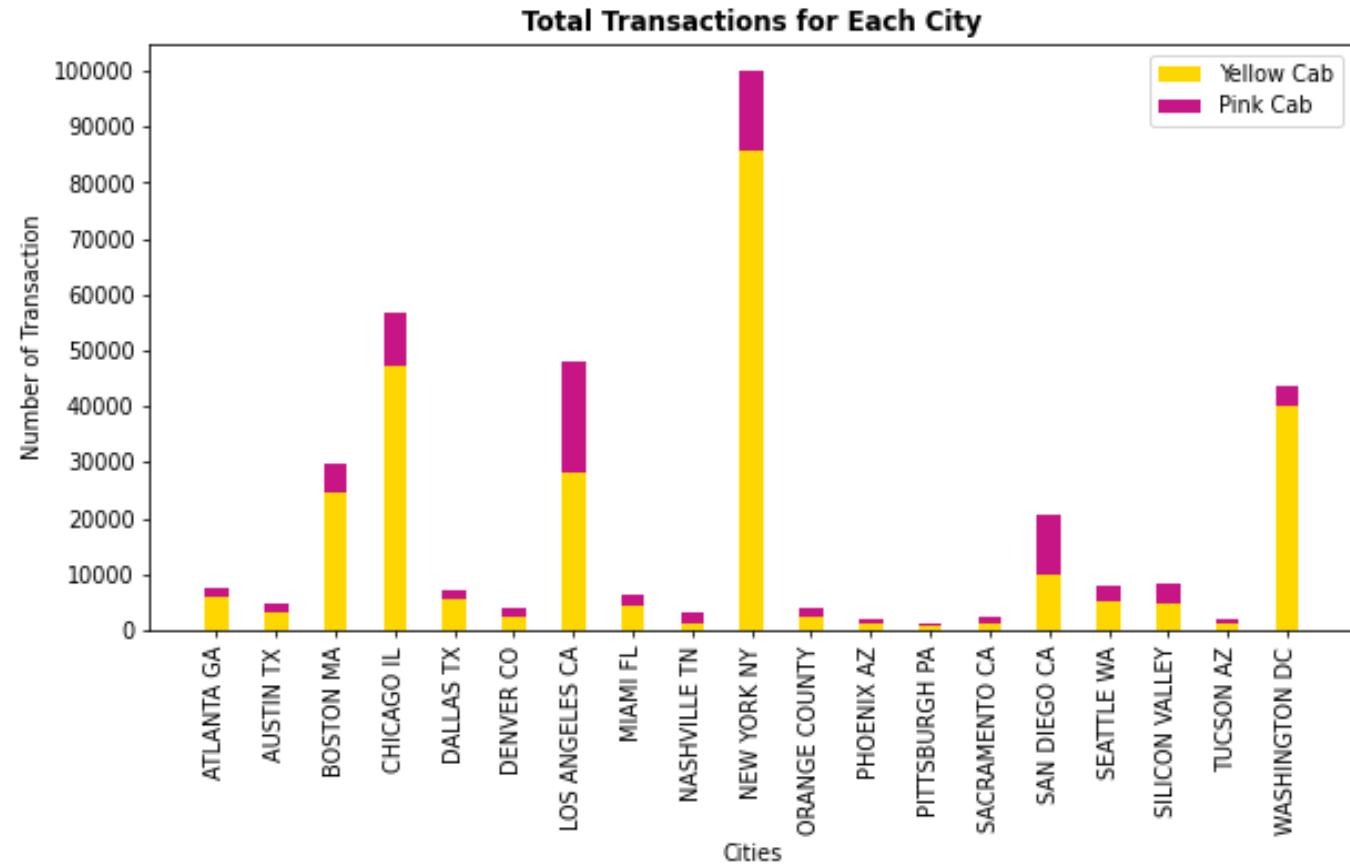
Number of Transactions in Each city
(Yellow Cab)



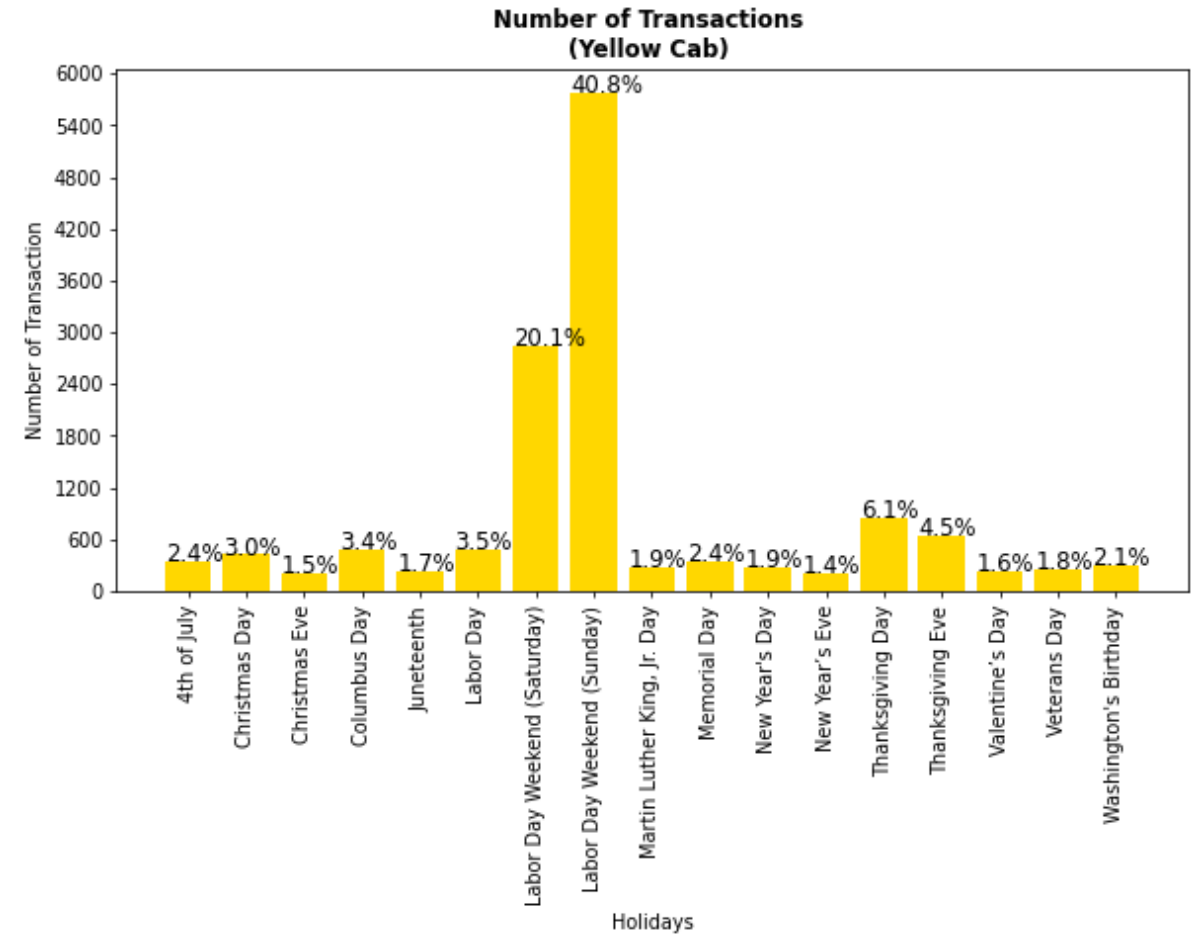
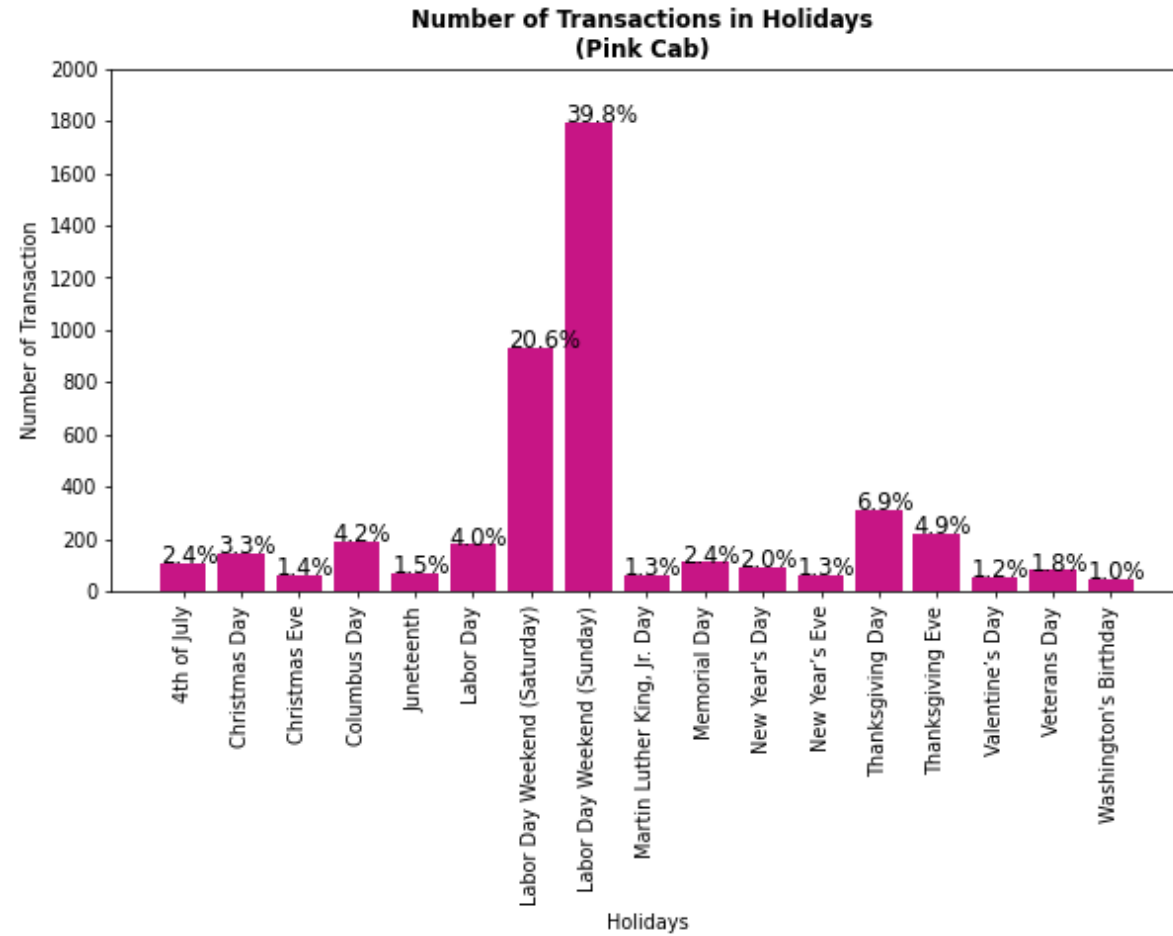
Yellow Cab has the most transactions in New York City.

Pink Cab has the most transactions in Los Angeles City.

Transactions in Each City (All)

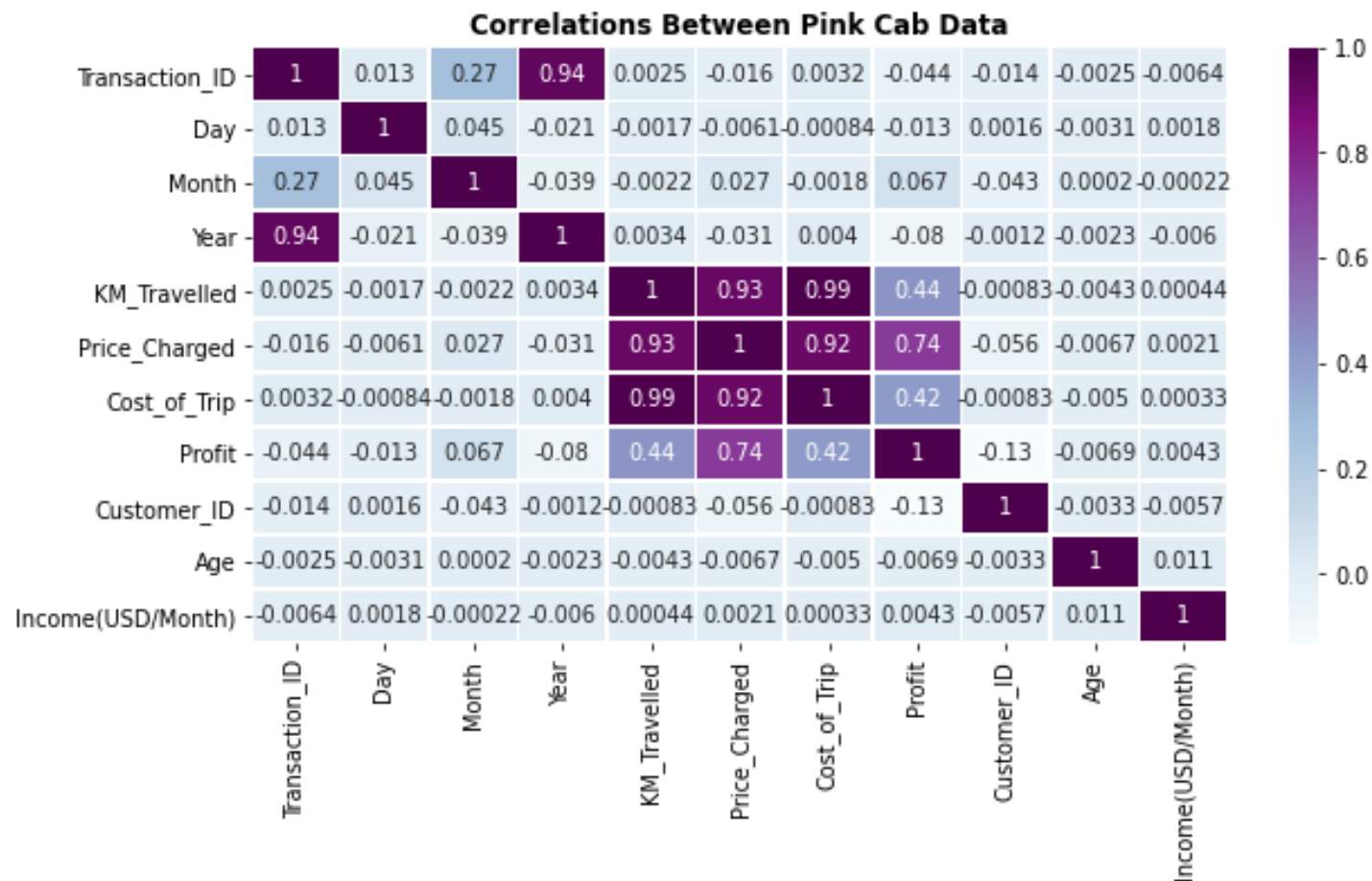


Transactions in Holidays



**There is no transactions in:
Western Easter Holiday and Eastern Easter Holiday.**

Correlations for Pink Cab

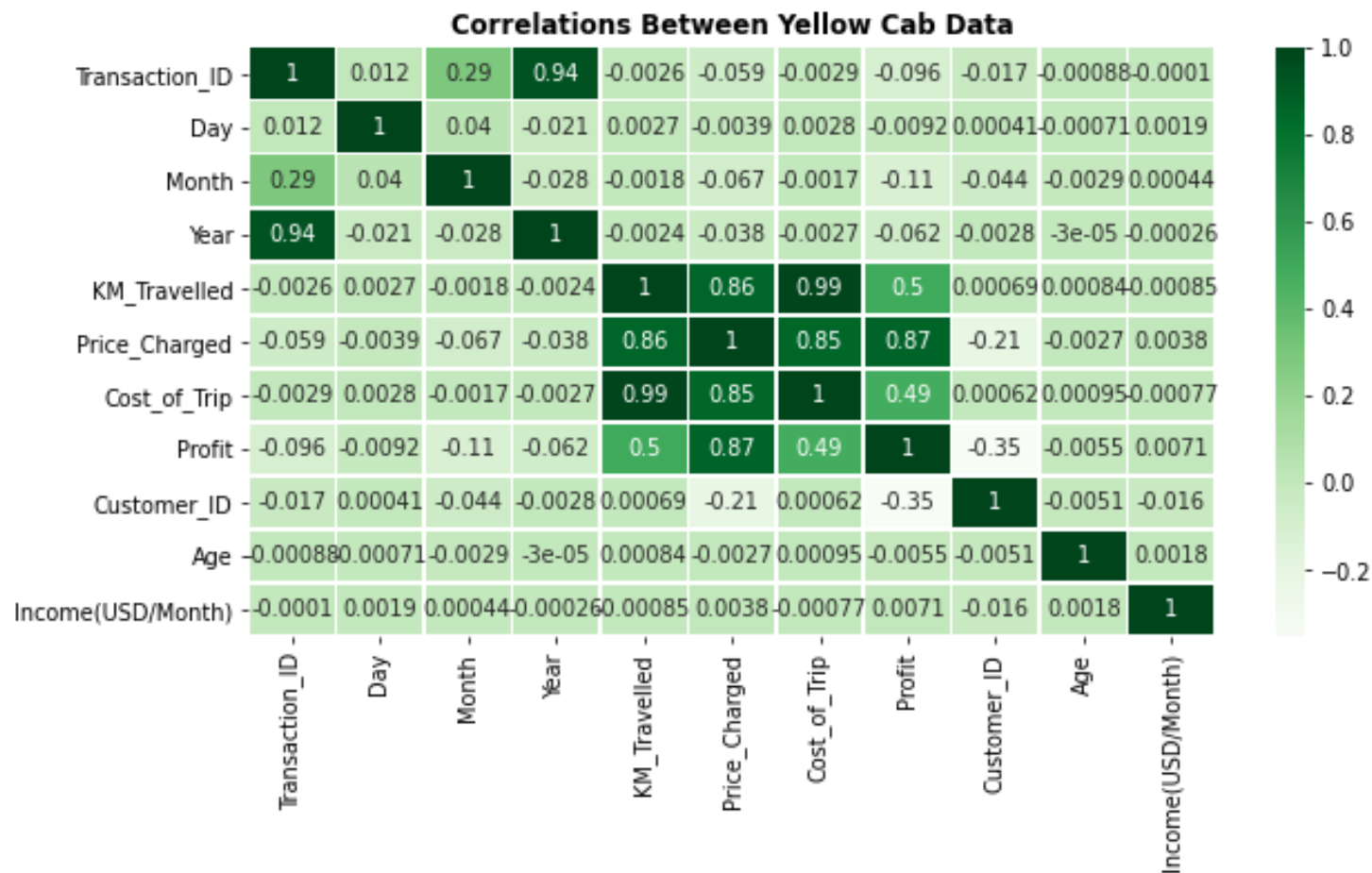


There are 4 correlations considered to be high here:

- KM_Traveled ~ Cost_of_Trip = 0.99
- KM_Traveled ~ Price_Charged = 0.93
- Price_Charged ~ Cost_of_Trip = 0.92
 - Price_Charged ~ Profit = 0.74

This correlation has no use
Transaction_ID ~ Year = 0.94

Correlations for Yellow Cab

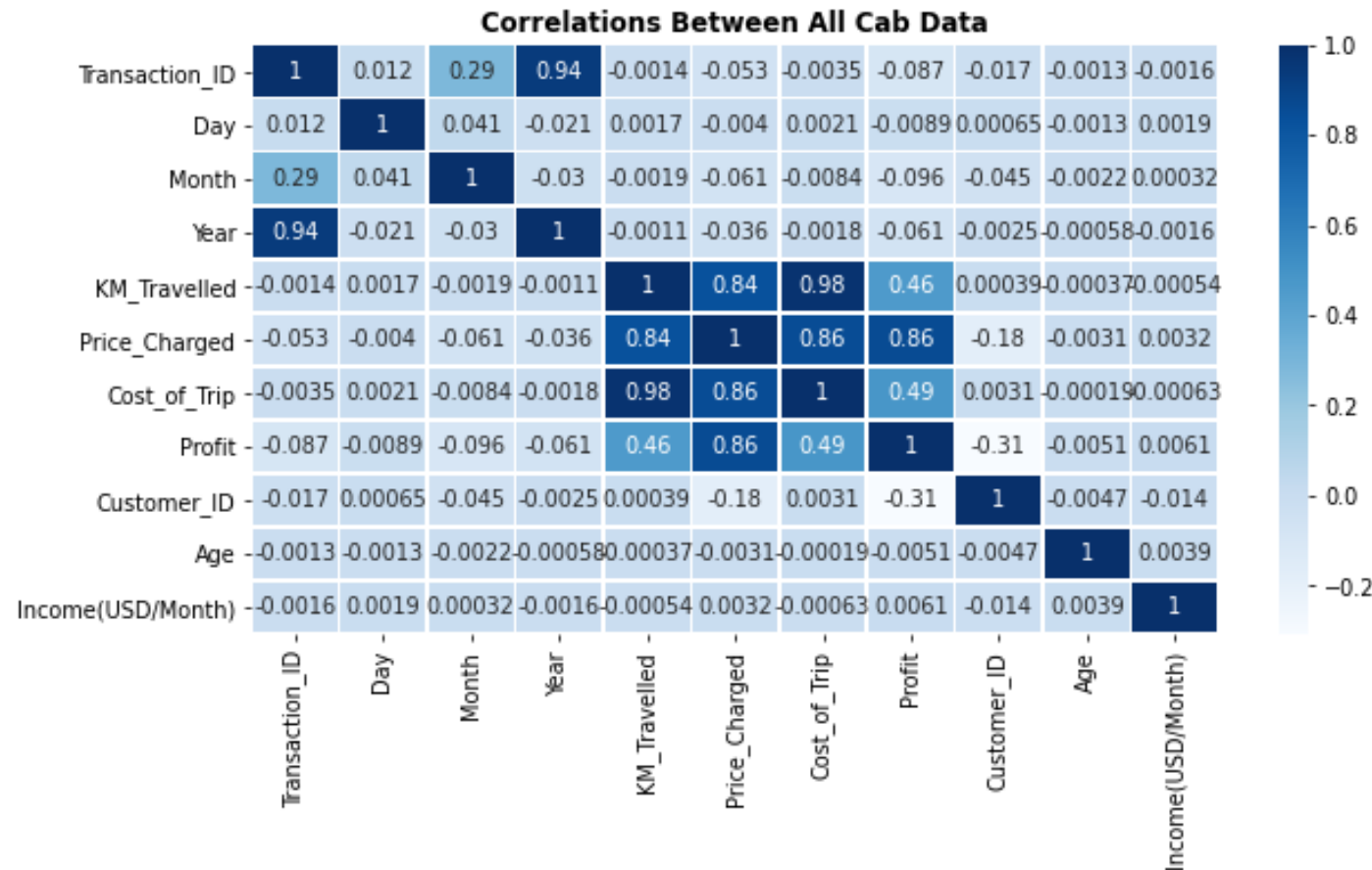


There are 4 correlations considered to be high here:

- KM_Traveled ~ Cost_of_Trip = 0.99
- KM_Traveled ~ Price_Charged = 0.86
- Price_Charged ~ Cost_of_Trip = 0.85
 - Price_Charged ~ Profit = 0.87

This correlation has no use
Transaction_ID ~ Year = 0.94

Correlations for All Cabs



There are 4 correlations considered to be high here:

- KM_Traveled ~ Cost_of_Trip = 0.98
- KM_Traveled ~ Price_Charged = 0.84
- Price_Charged ~ Cost_of_Trip = 0.86
 - Price_Charged ~ Profit = 0.86

This correlation has no use
Transaction_ID ~ Year = 0.94

This slide to show how the correlations will be if the two companies merged under one name.

Recommendations

We have evaluated both the cab companies on following points and found Yellow cab better than Pink cab:

According to Profits: The Yellow Cab company have shown difference here in terms of users, profits, and the average profit per ride. Literally Yellow Cab's average profit per ride is two and half times the average profit per ride of Pink Cab.

According to Age: Both companies have similar age ratios, mostly the focus is between 26-40 with percentage of 0.47 approximately, the ratios of 18-25 and 41-60 are close but the lead is to 18-25. Finally, the ratio of 60+ has the lowest ratio with percentage of 0.05 approximately.

According to Gender: Also, for gender, the two companies have a similar ratios, the ratio of male is always higher than females. One this I need to point to, in Pink Cab company, the ratio of females is slightly higher than the Yellow Cab company.

According to Success: The 2017 year was the best for both companies in terms of everything, such as profits and number of transactions. December was the best month for profits and number of transactions. Finally, the days 3,10, and 17 had the highest numbers of transactions and number of profits for both companies.

Recommendations (cont.)

According to Users: The Yellow Cab company have around 3.5 times the number of users that the Pink Cab company have.

According to Holidays: There is transactions everyday in the year (even holidays) except for two days, Western Easter Holiday and Eastern Easter Holiday. On the other hand, in the Labor Day Weekend day (Saturday and Sunday), both companies had transactions more than the transactions in the other holidays.

According to Payment: Both cabs' users prefer to pay via Card over Cash, percentages are very similar although.

According to Cities: Most of Yellow Cab company's transactions are in New York City. On the other hand, for Pink Cab company's transactions, mostly they are in Los Angeles City.

According to Correlations: The Yellow Cab company has shown higher correlation between **price_charged** and **Profit** than the Pink Cab Company. On the other hand, the Pink Cab company has shown higher correlations between **KM_Traveled** and **Price_Charged**, also between **Price_Charged** and **Cost_of_Trip**.

Based on these results, I will recommend the Yellow Cab Company for investment.

Thank You

Yousef Elbayoumi