Exploratory Data Analysis

<G2M Insight for Cab Investment Firm>

<9/21/2022>

Agenda

Problem Statement

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 : Conduct EDA and 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 (EDA)
- Regression and forecast for profit of each cab type
- Recommendations for investment

Datasets

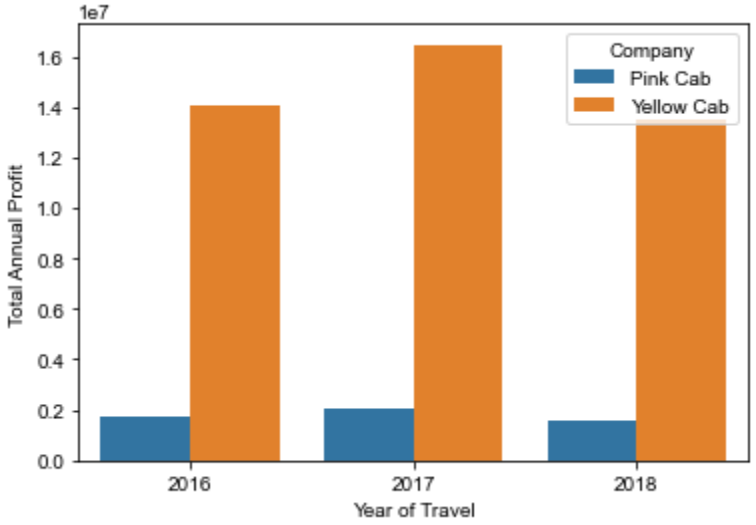
- Data files: (1) Cab_Data.csv; (2) City.csv; (3) Customer_ID.csv; (4) Transaction_ID.csv.
- Number of features: 21 features, include 6 derived features.
- Samples or data points: 359392
- Timeline: 2016-01-01 to 2018-12-31.

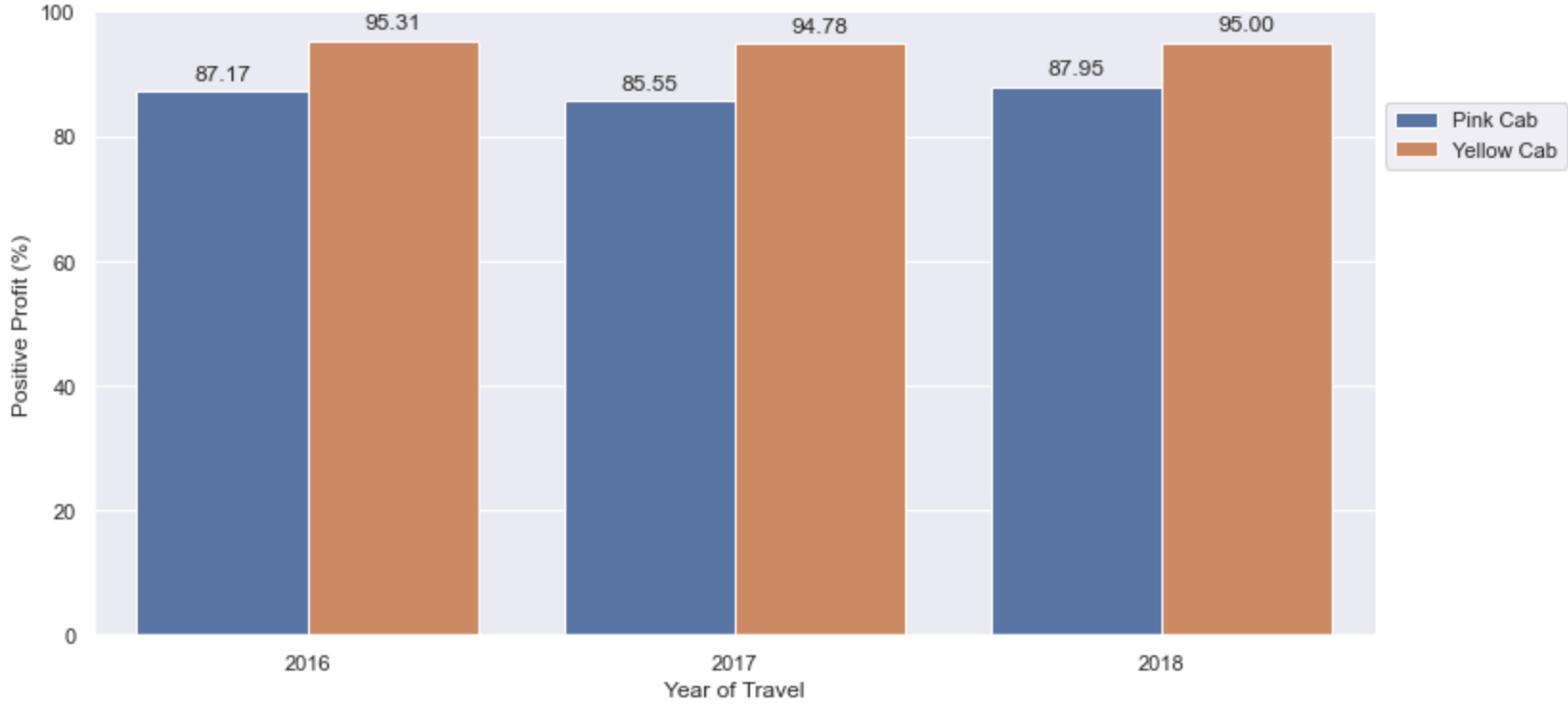
Assumptions:

- Profit per transaction is computed by the difference between “Price Charged” and “Cost of Trip”;
- Number of users in each city just need to count unique values of feature: “Customer ID”;
- Neglect other passengers in the same car as it was not recorded, only the one being charged has been recorded.

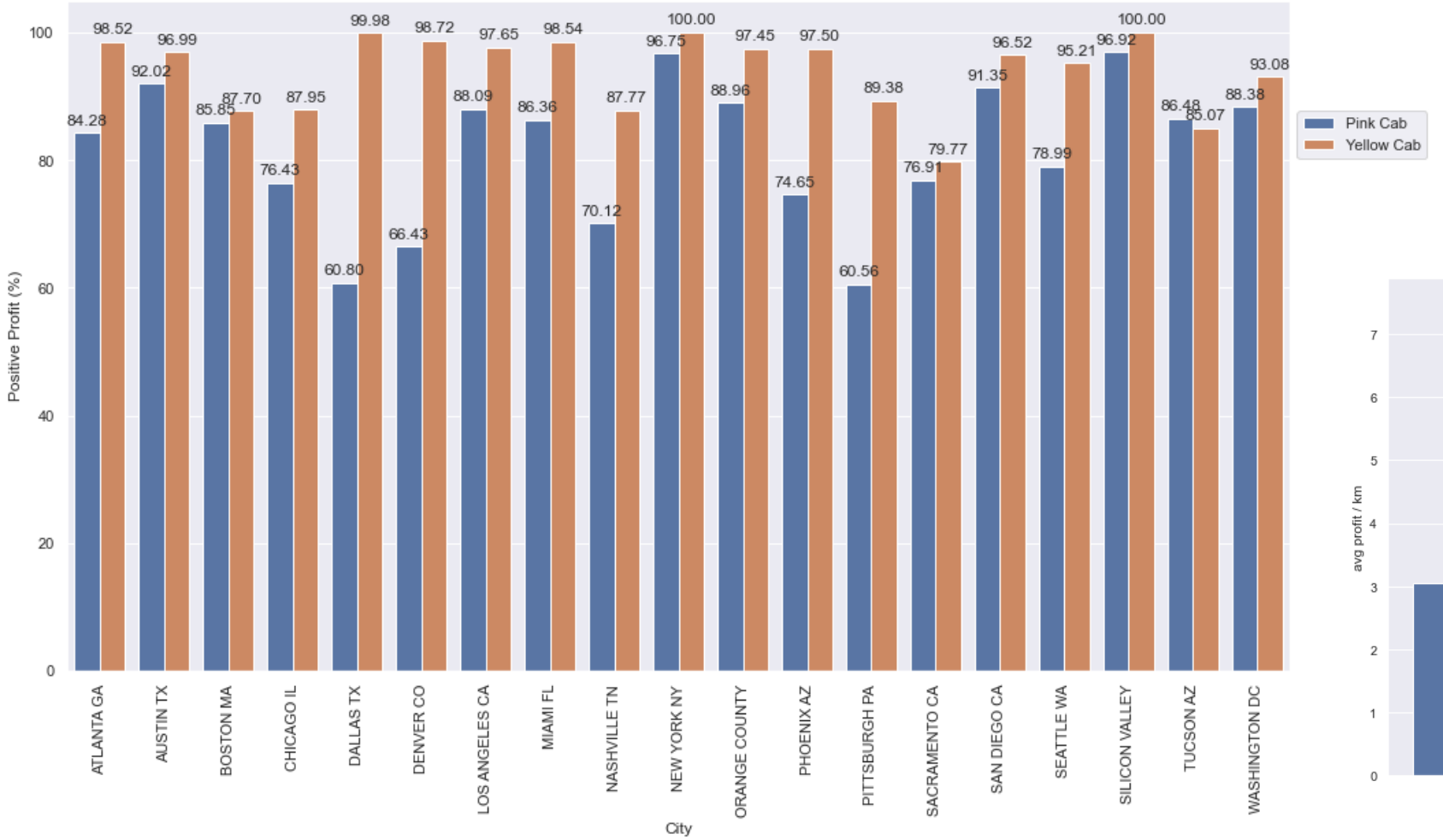
Annual Profit

2016-2018	Total Profit	Total rides	Average profit per ride
Pink Cab	5307328	84711	62.6222
Yellow Cab	44020370	274681	160.2600

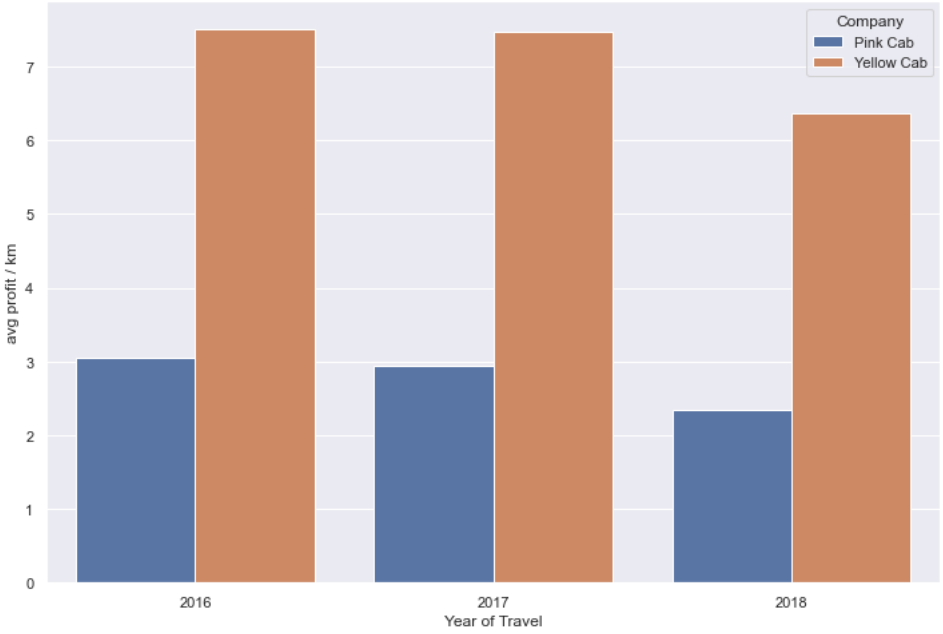




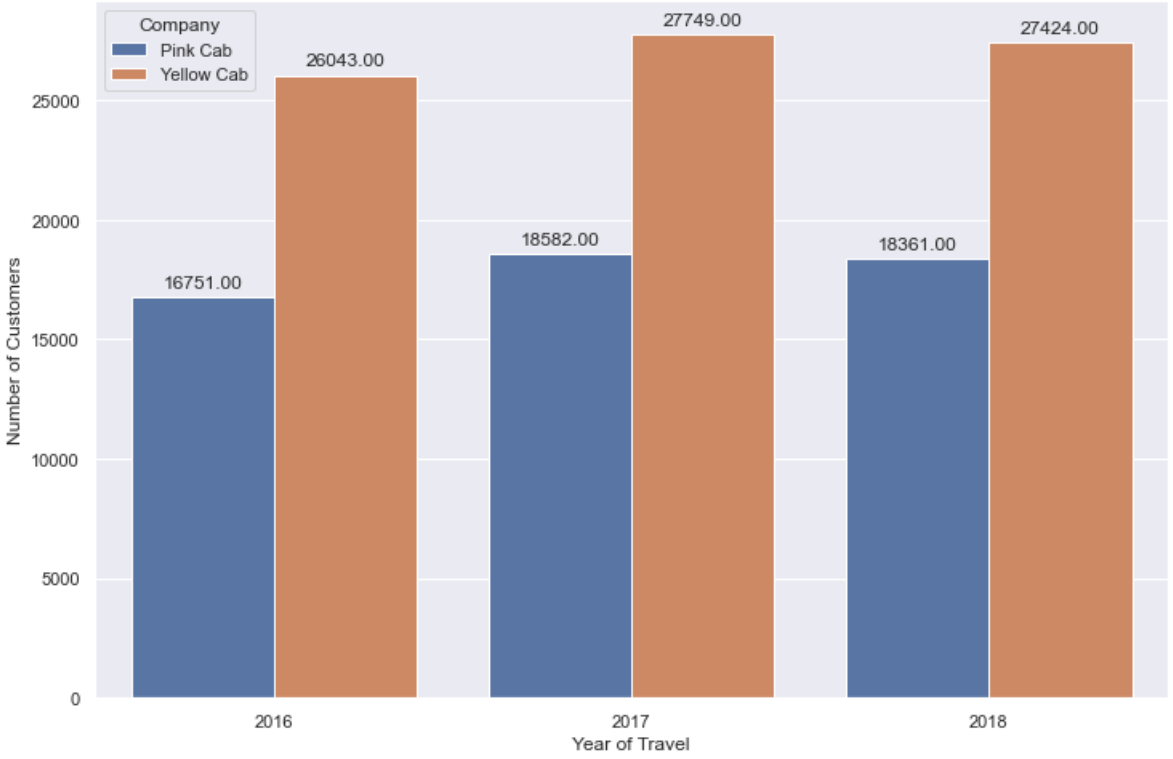
Positive Profit per City

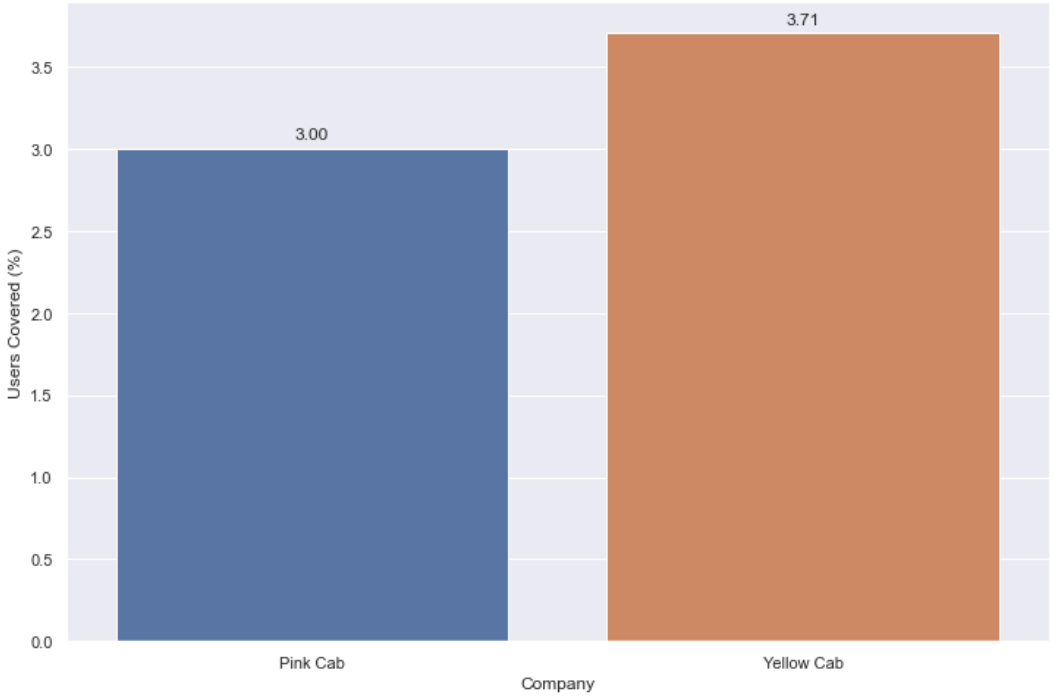


Mostly both company are profiting, but Yellow Cab has more profitable rides than Pink Cab, both in percentage or in average profit/km.



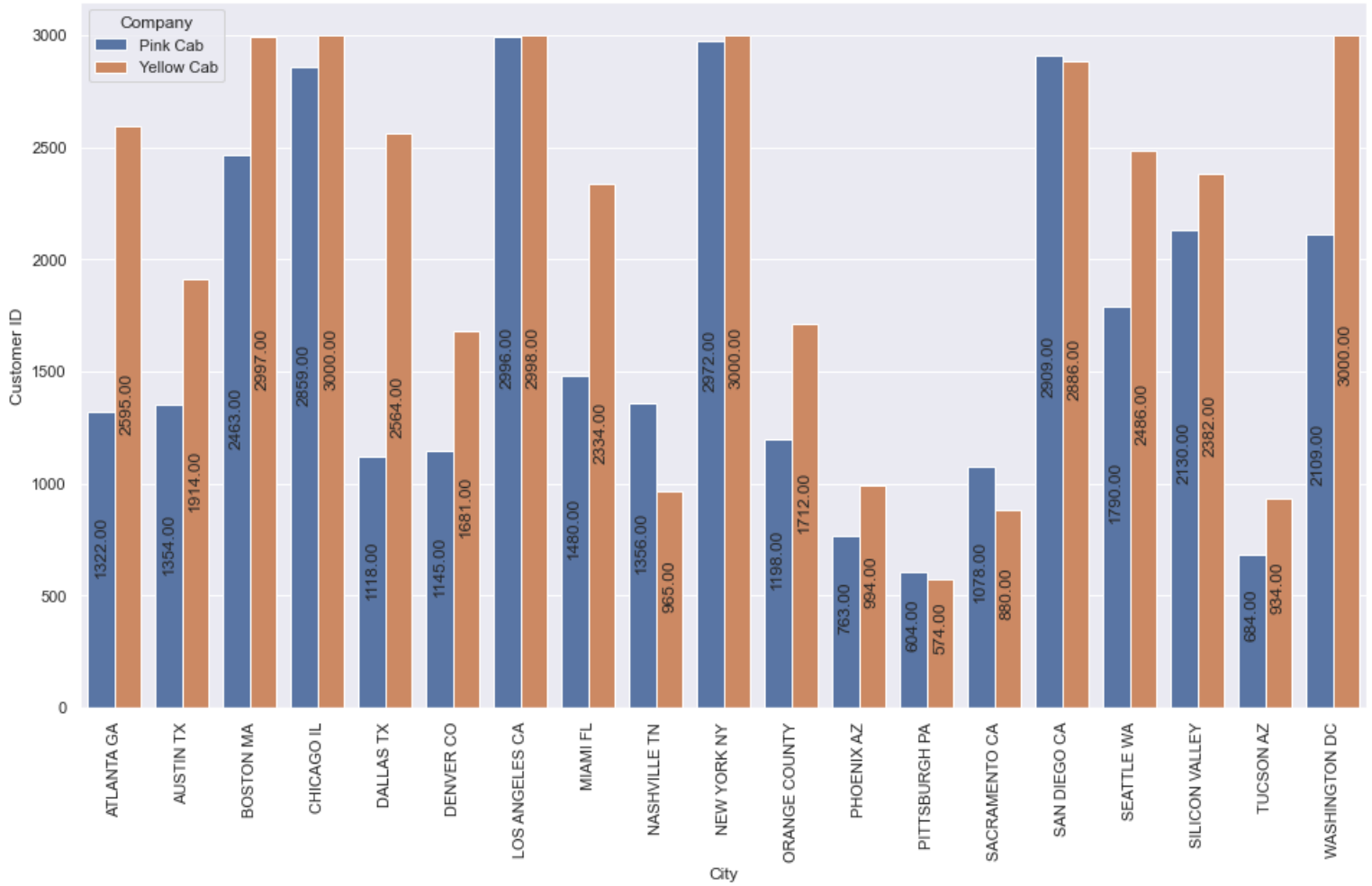
Customer Reach





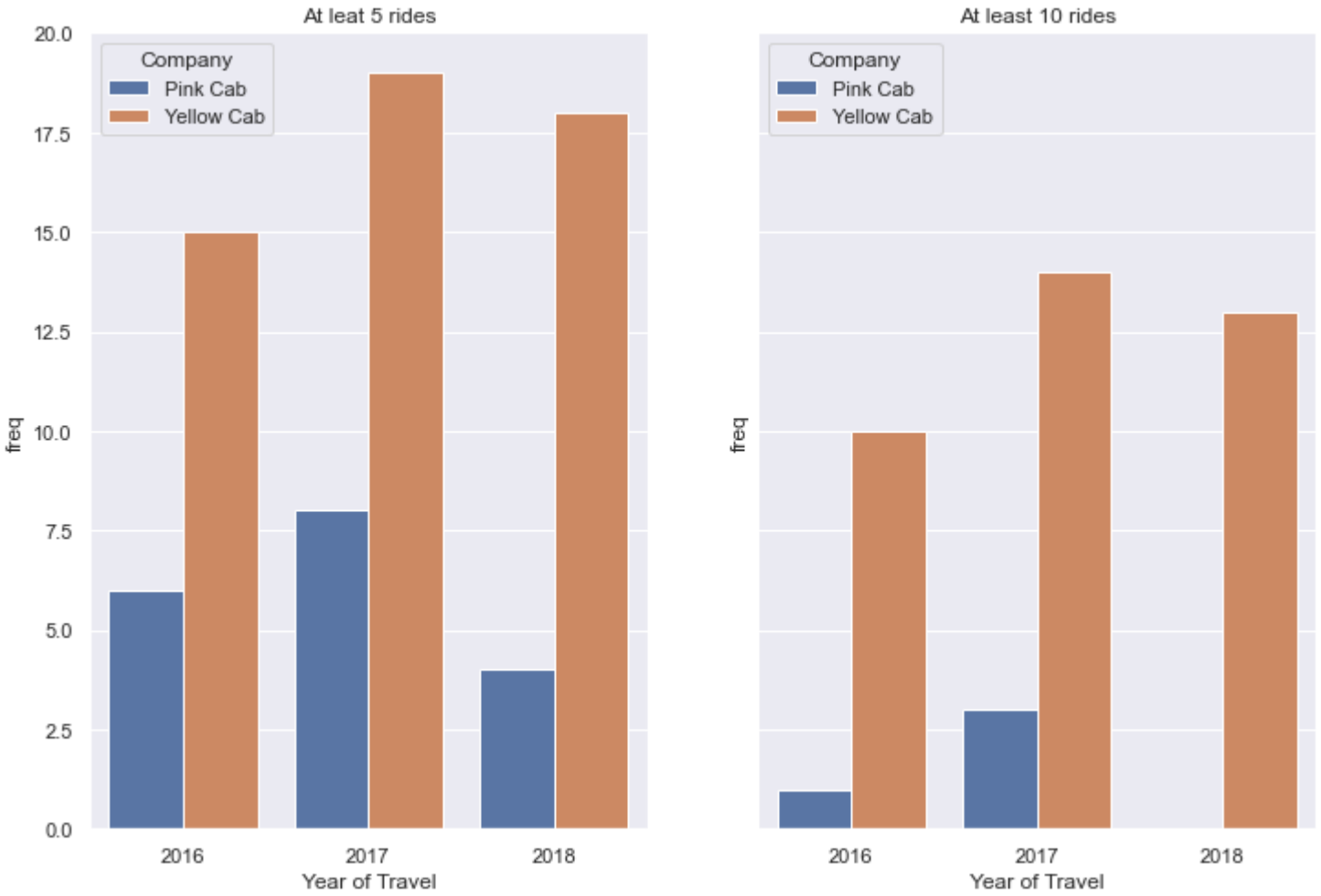
Note: there are still other company so user coverage is small for both company.

Customer Reach per City



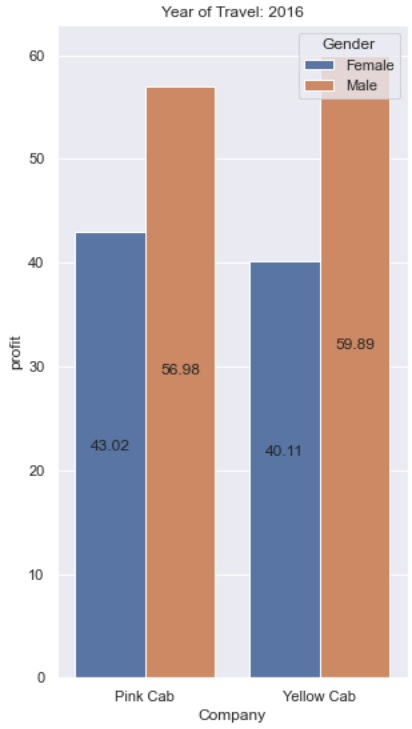
In most cities, Yellow Cab has more customer reach than Pink Cab, despite 4 cities: Nashville, TN; Pittsburgh, PA; Sacramento, CA; Sab Diego, CA.

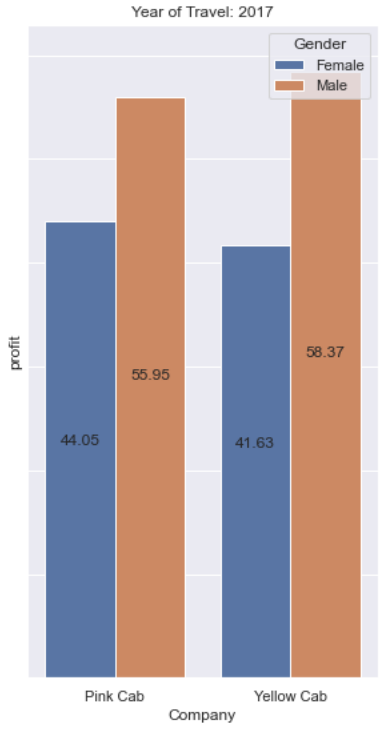
Customer Retention

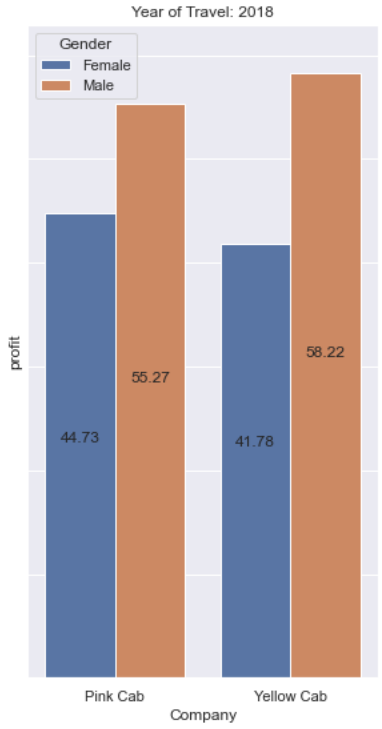


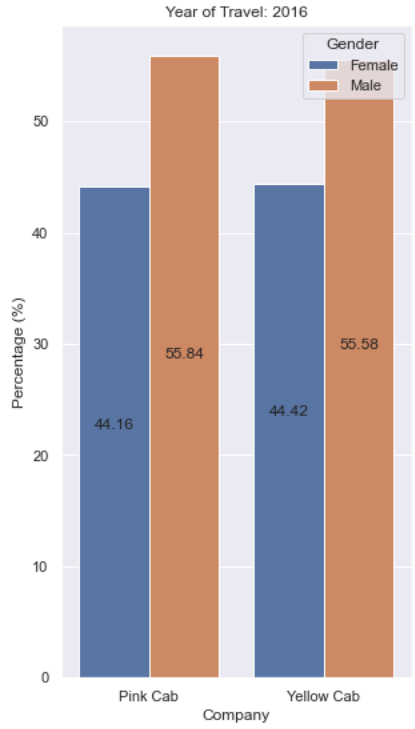
People using yellow cab with more than 5 rides annual is higher than pink cab; it is even more significant for people taking rides more than 10 times per year. Meaning that yellow cab has significant better customer retention.

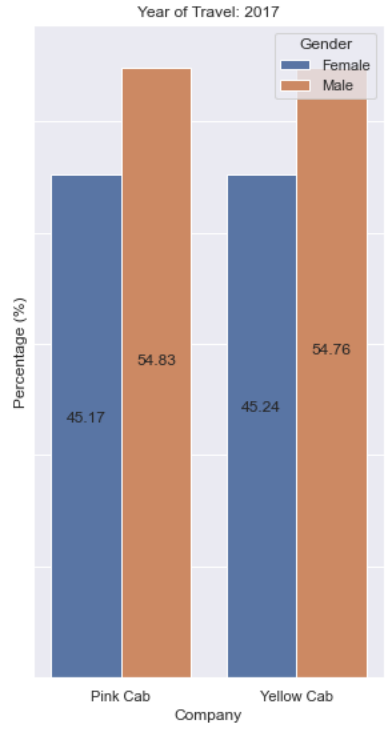
Gender Effect on Profit and Customer Base

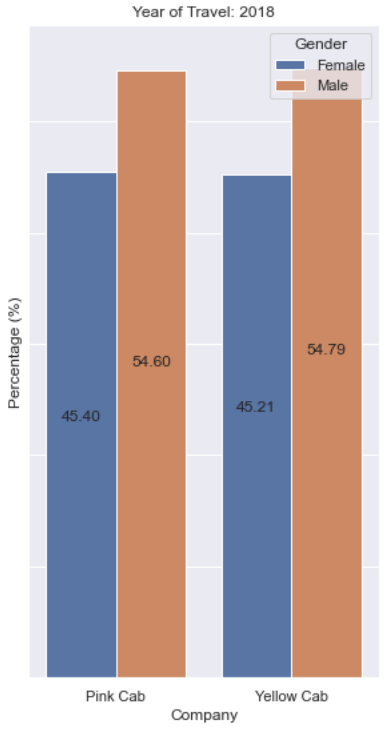




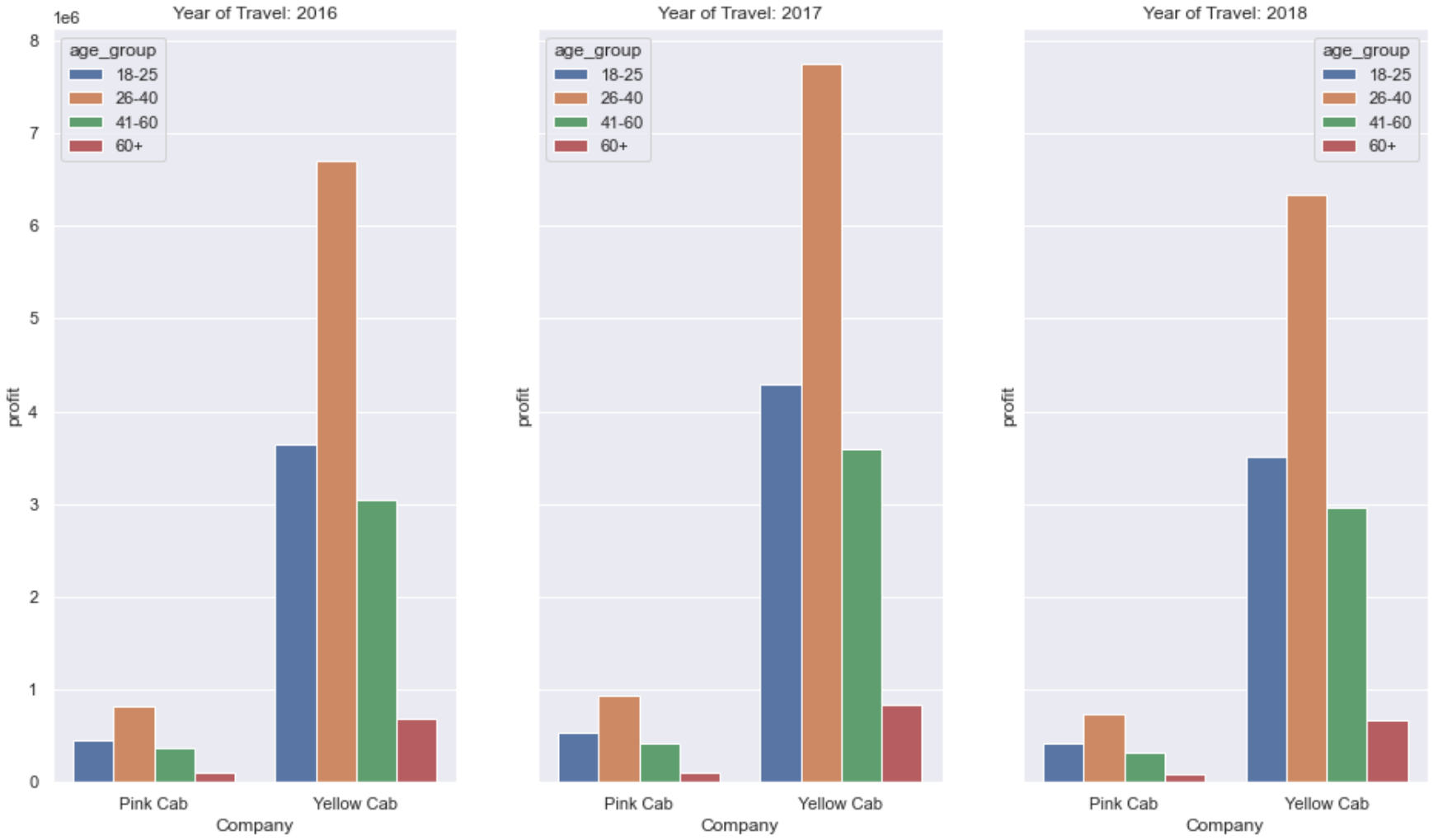


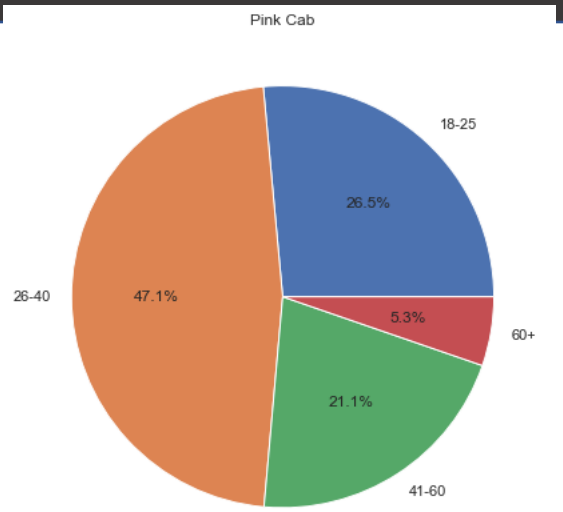


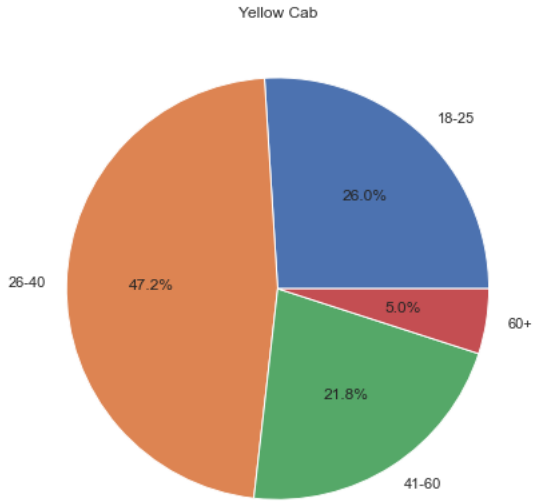




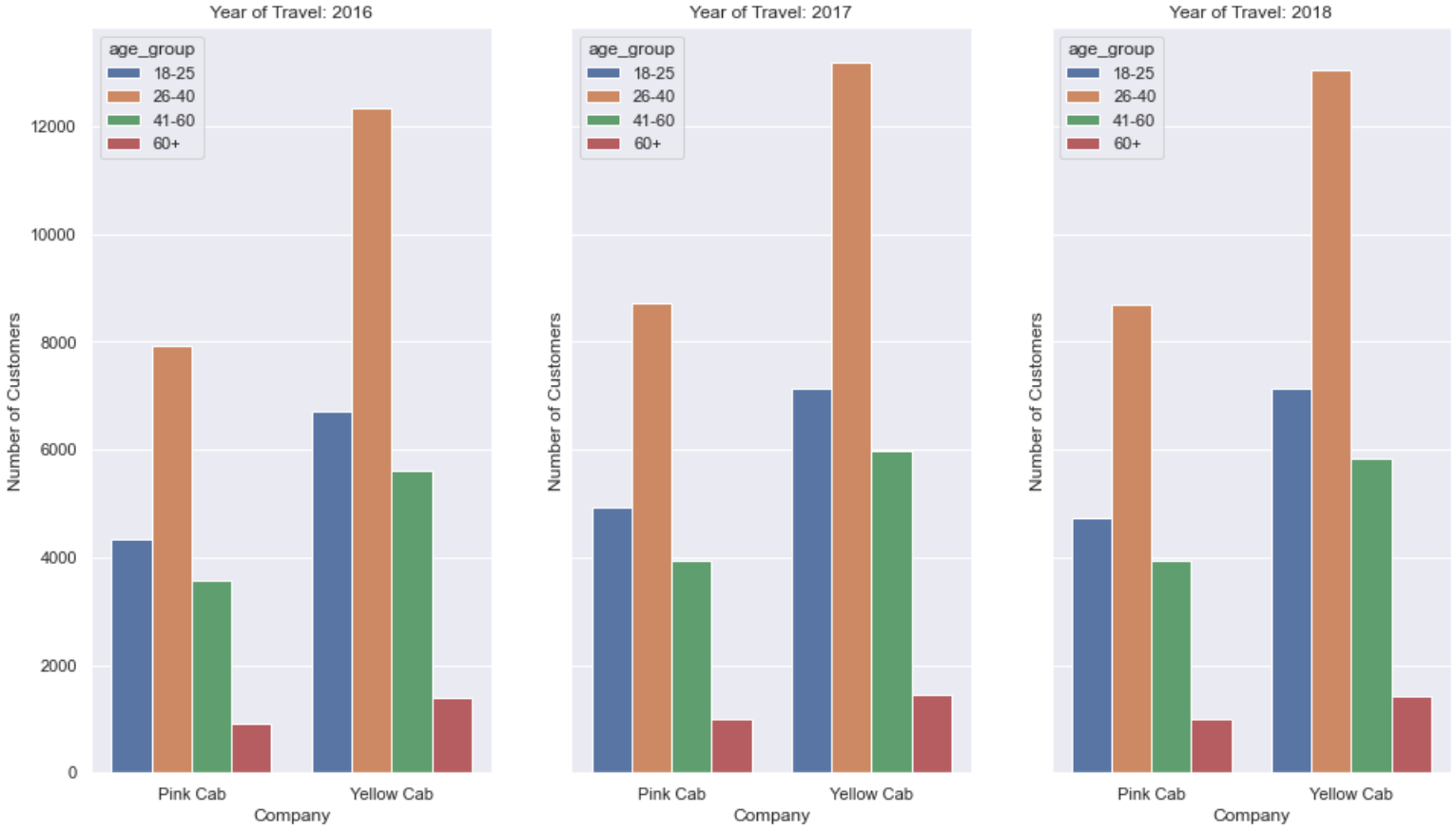
Age Effect on Profit

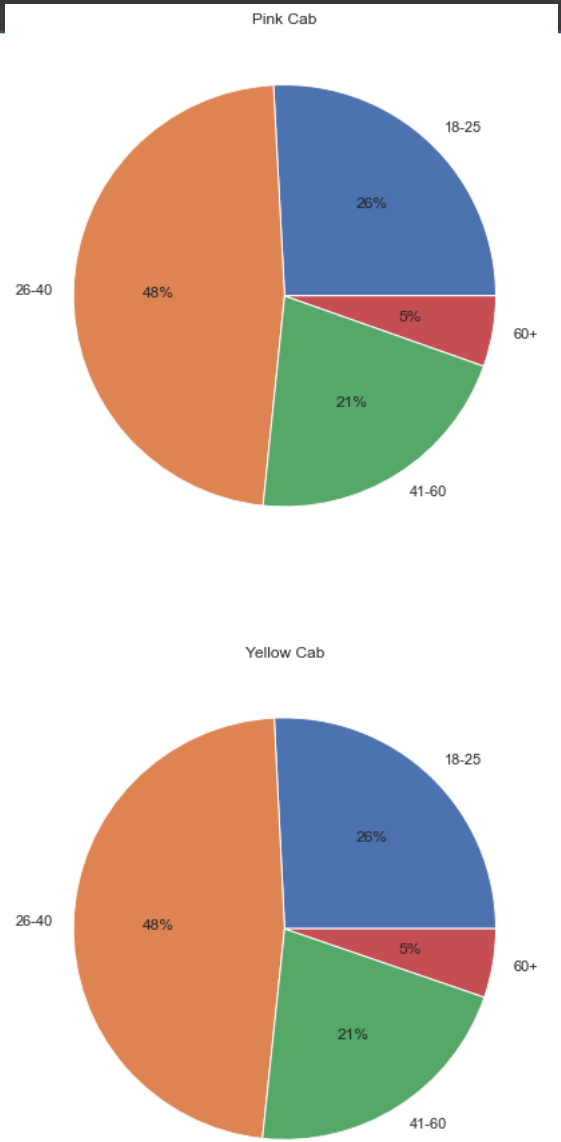




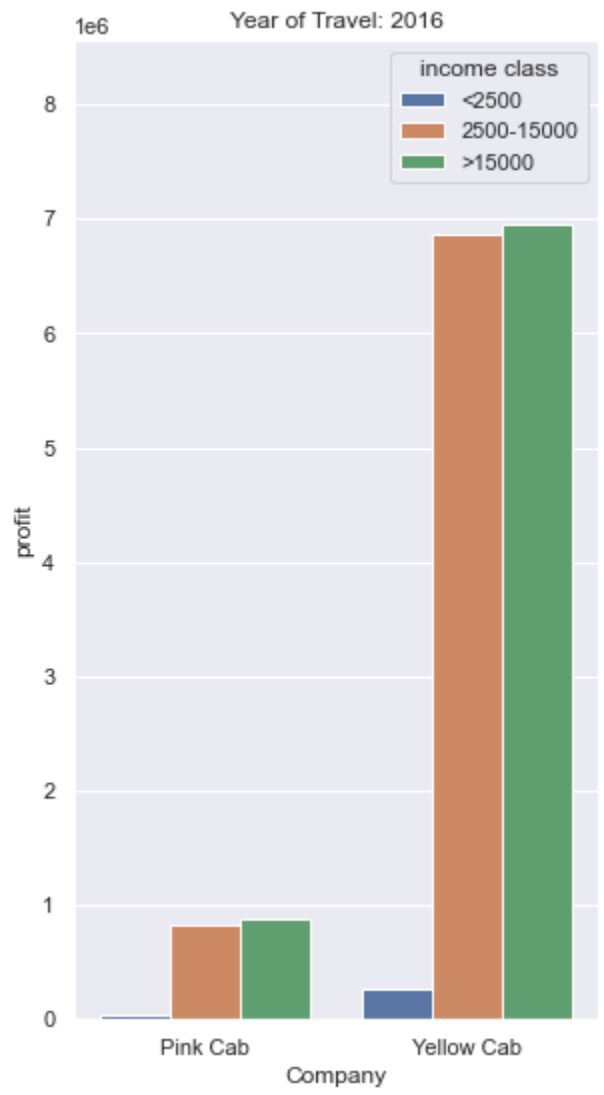


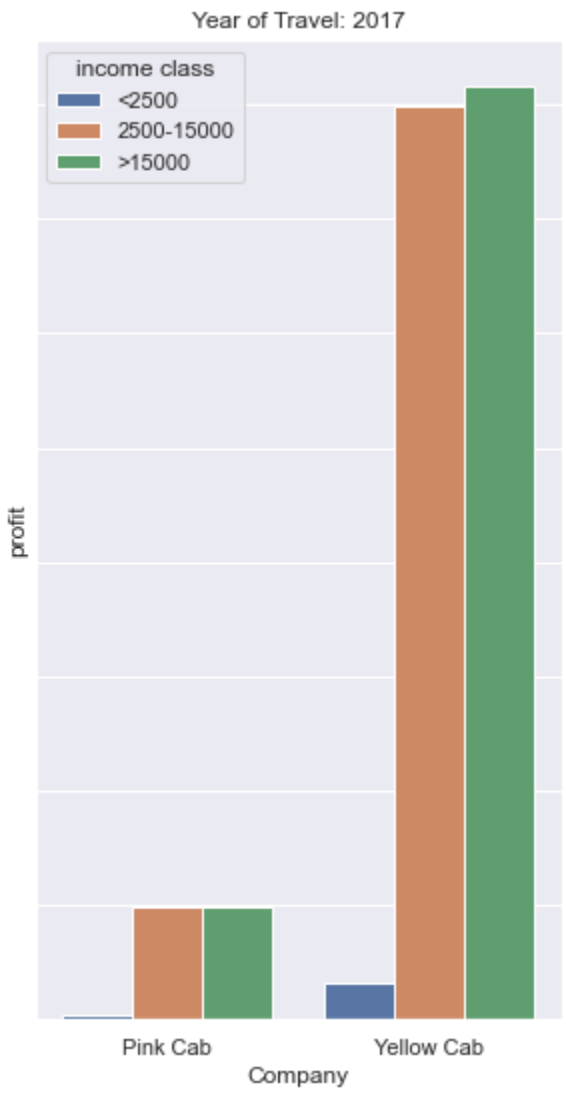
Age Effect on Customer Base

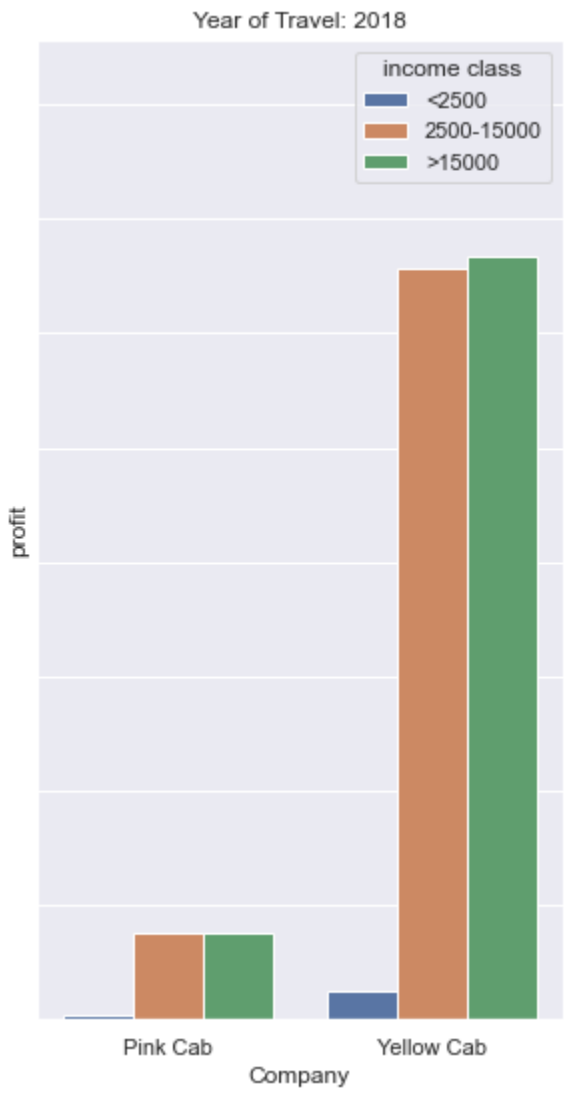


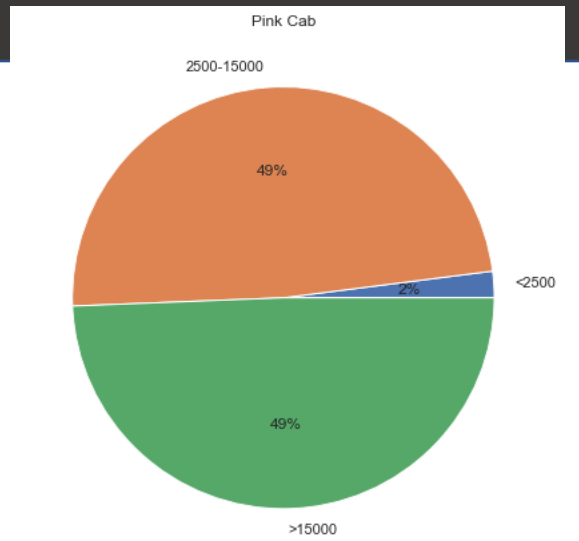


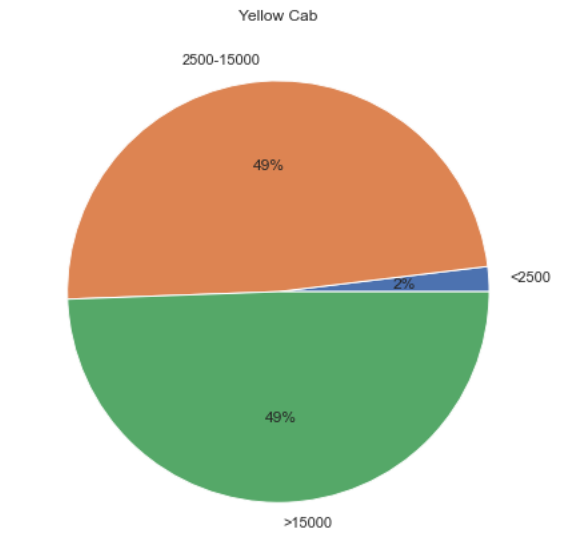
Income Class on Profit



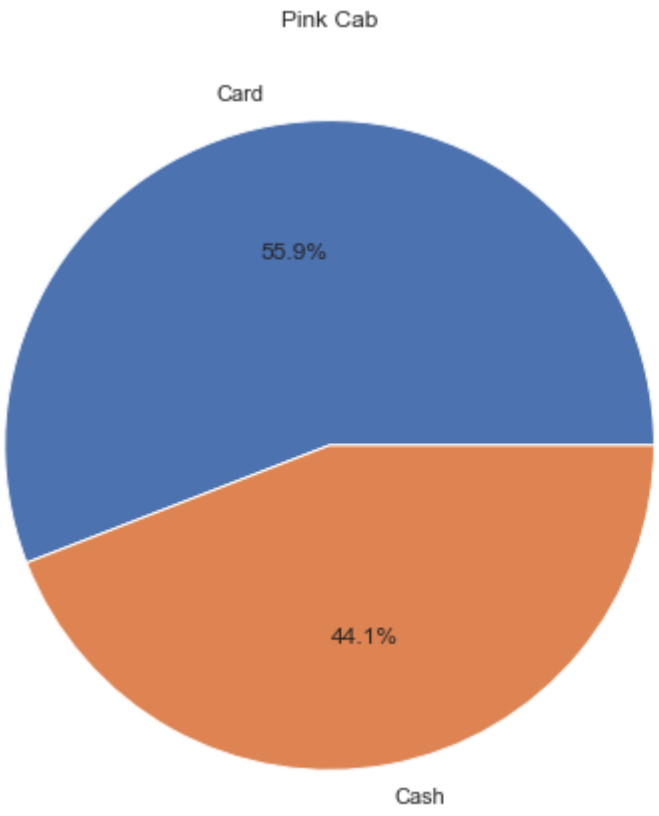


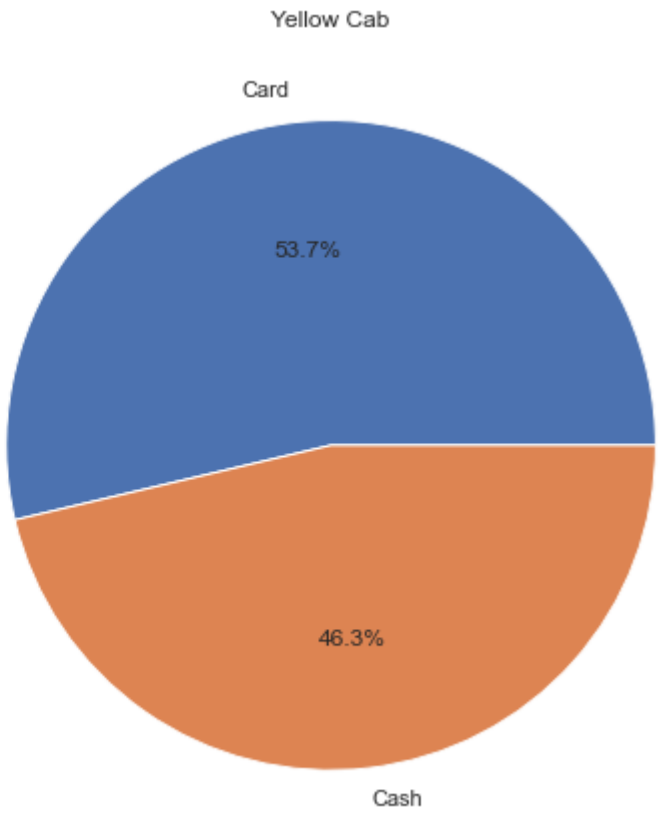






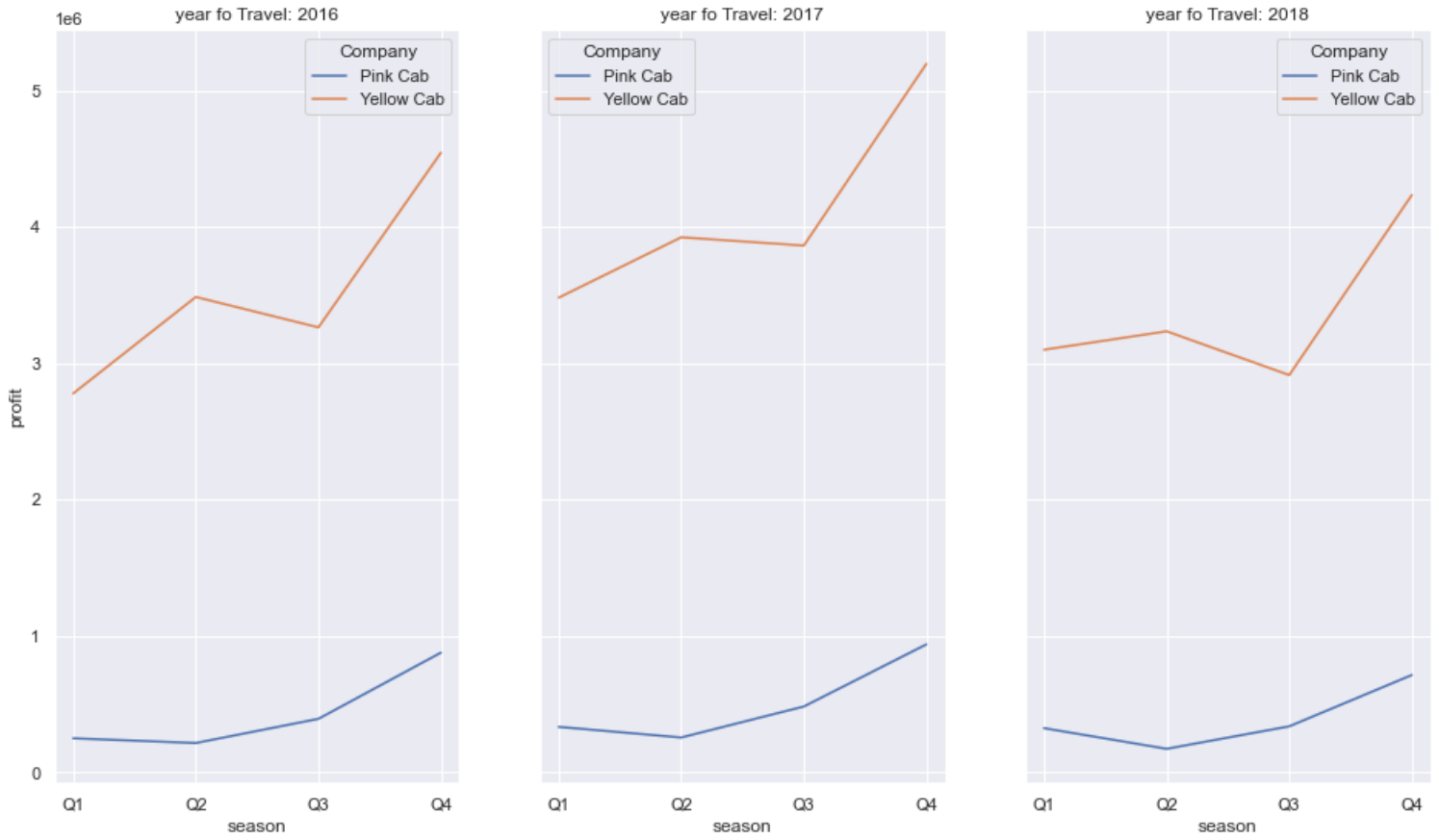
Payment Method on Customer Base





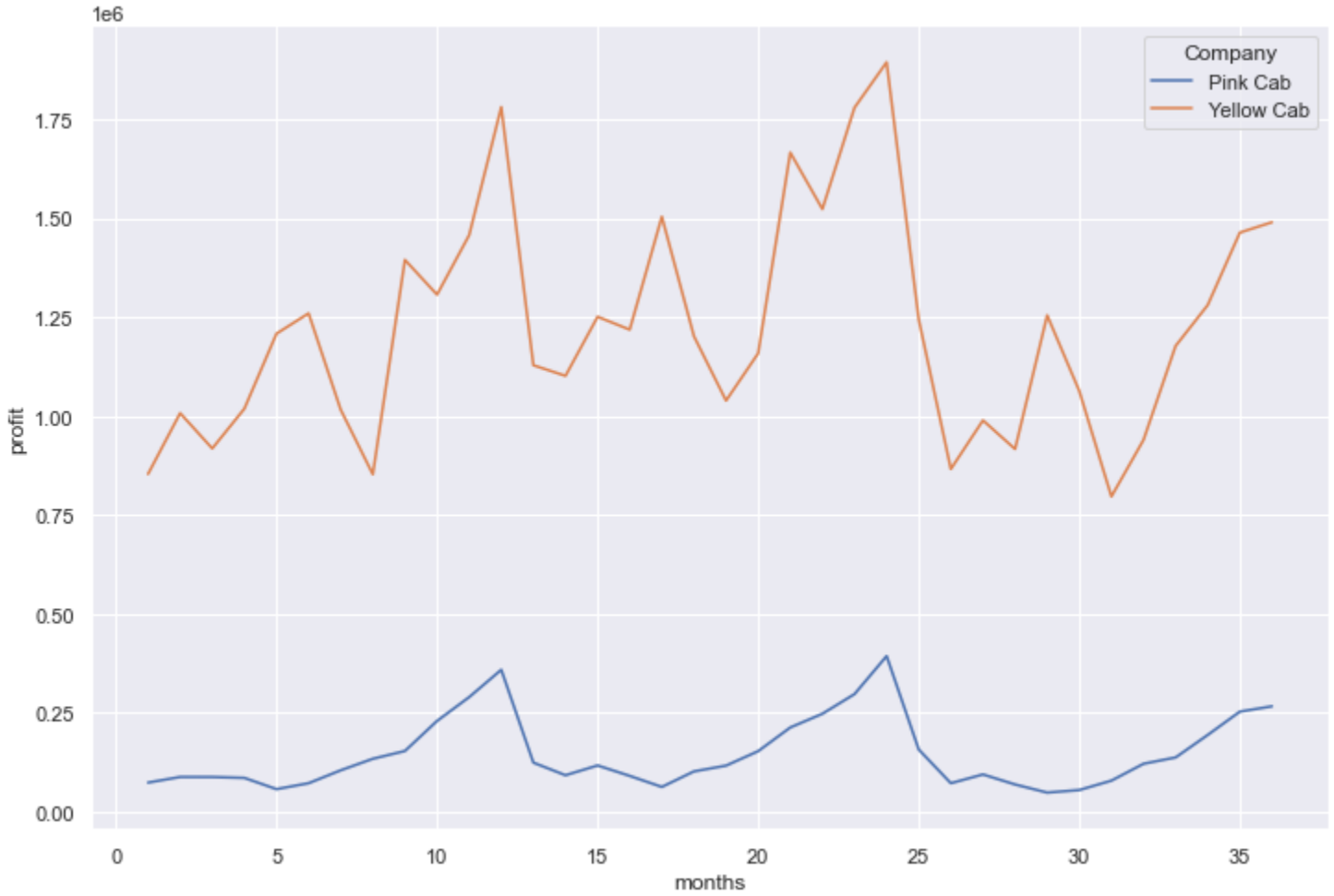
For both company, more people prefer card than cash.

Seasonal Change in Profits



Clearly, Yellow Cab has higher profits than Pink Cab; due to holidays at the end of the year, more profits/rides at Q4.

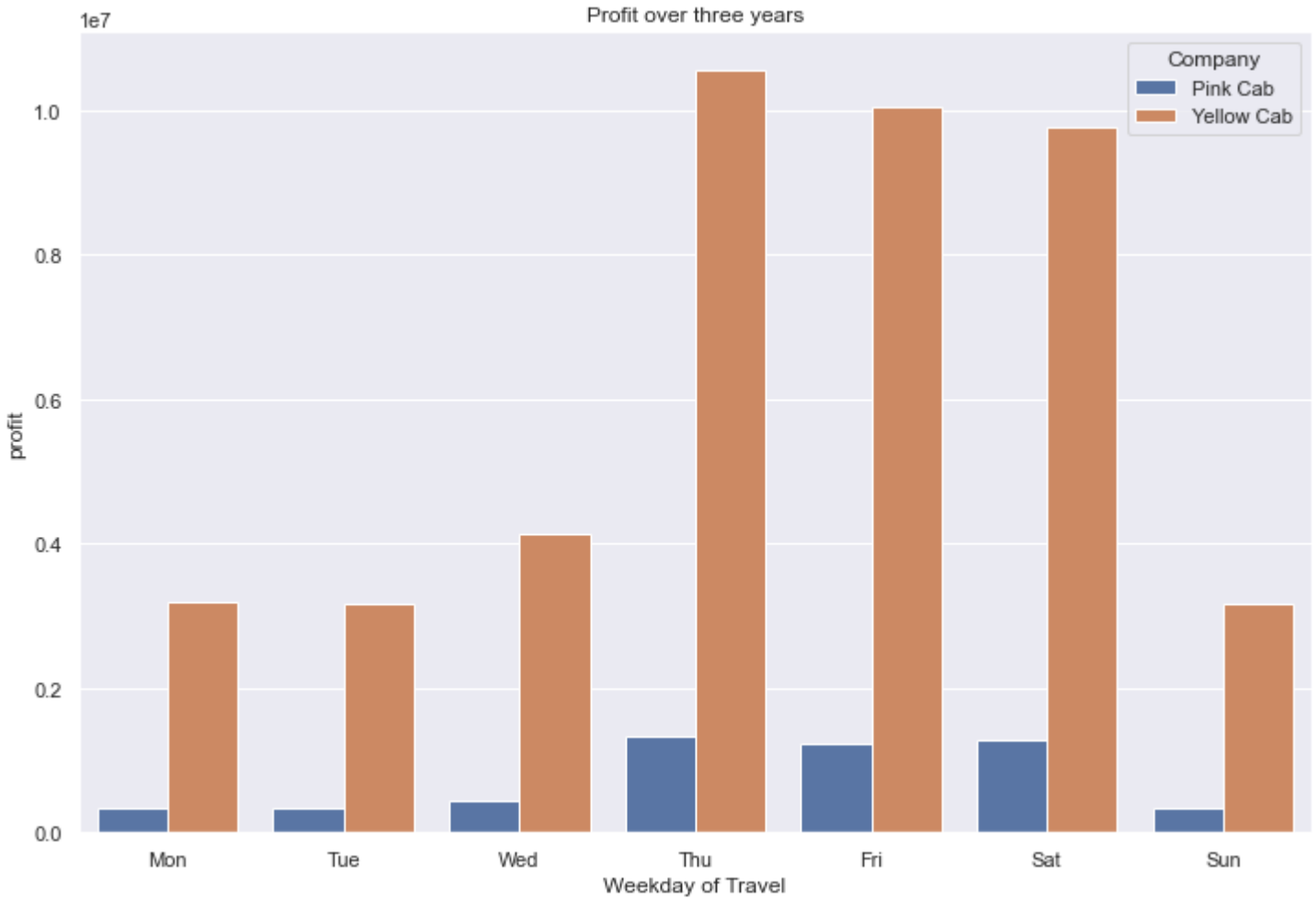
Monthly Change in Profits



Clearly, Yellow Cab has higher profits than Pink Cab; due to holidays at the end of the year, more profits/rides at Q4.

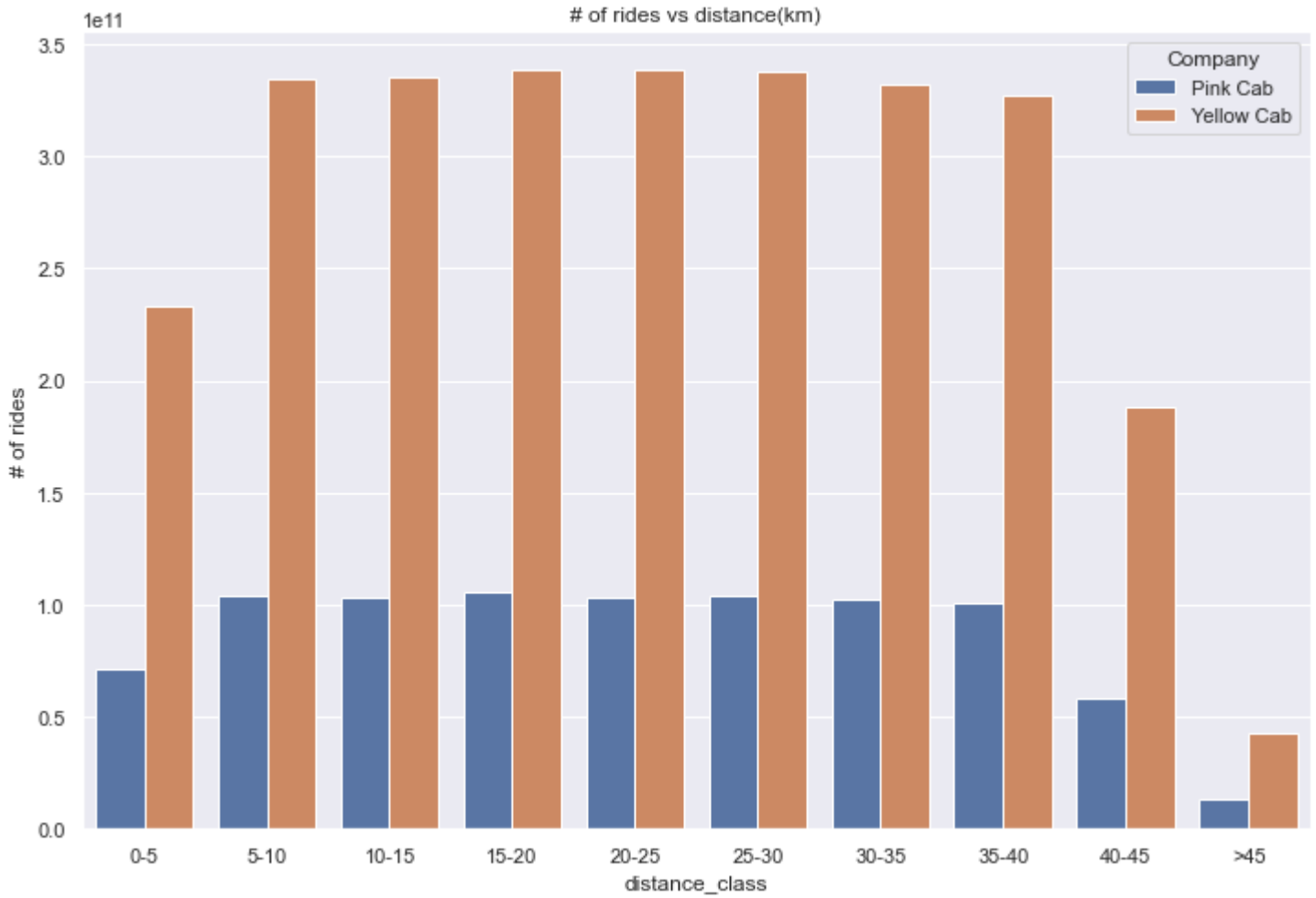
Consistent with last slides.

Profits in Weekdays



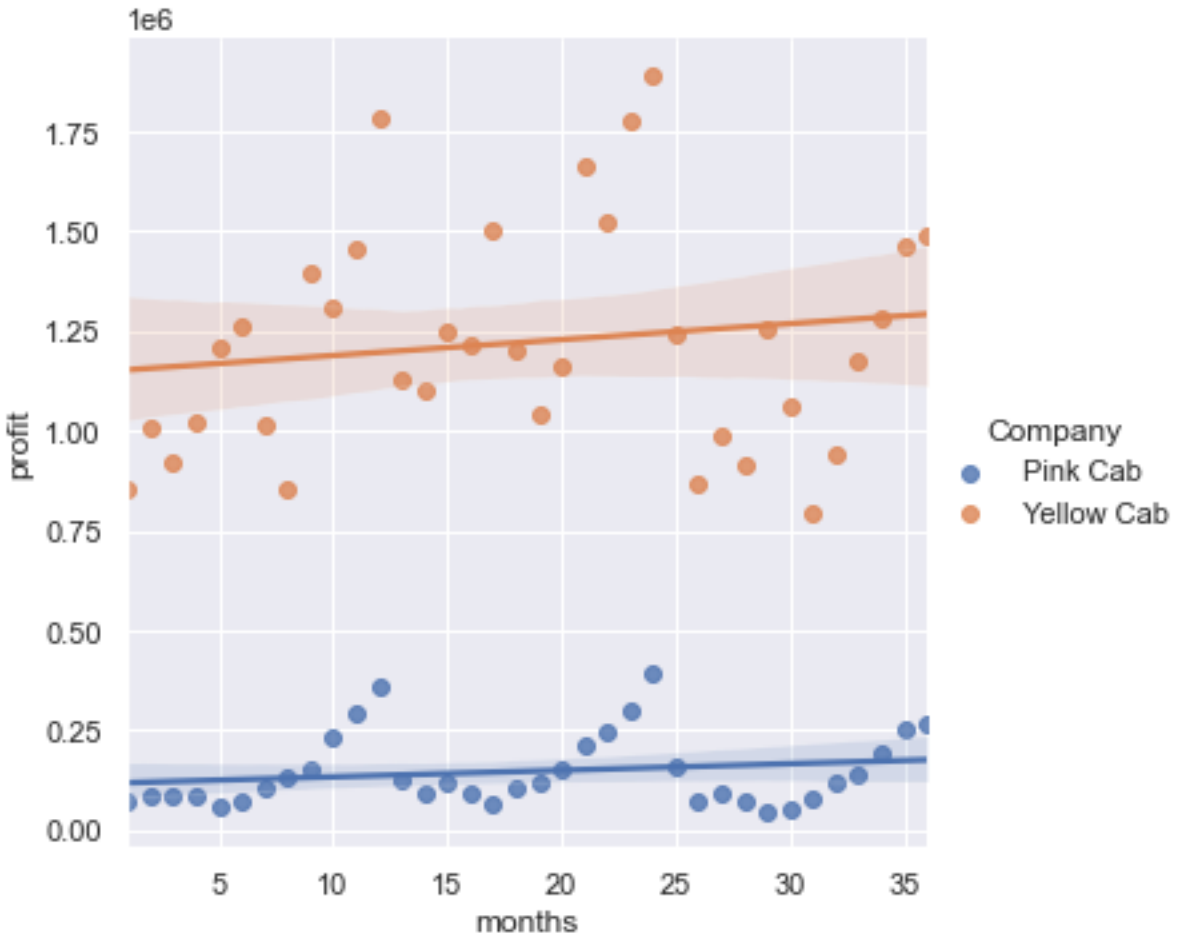
More profits/rides on Thu, Fri, Sat, the same for both company.

Number of Rides vs Travel Distance



People take rides mainly less than 45km; especially between 5 and 45.

Profit Regression and Prediction



Clearly next year, i.e., 2019, the annual profit of yellow cab will be around $1.35 million; while it is 0.24 million for pink cab.

Conclusion of Investment

Both company are evaluated with EDA in terms of profits, customer reach, customer retention, and effect of different factors on profit and customer base:

- During 2016,2017 and 2018, the annual profits of Yellow Cab is 7 to 8 times of annual profits of Pink Cab;
- In terms of proportion of profit rides, Yellow Cab is about 95% each year, while Pink Cab is 85% each year;
- The average profit/km of Yellow Cab is twice the value of Pink Cab;
- For a city wise comparison, Yellow Cab has more customer base than Pink Cab, hence the profits; Yellow Cab also has higher customer retention, i.e., it can keep more clients calling their rides, such as >5 rides per year, or >10 rides per year;
- Even 60+ people also use Yellow Cab which shows a full cover of all ranges of ages;
- Based on regression and predictions on the next year, we recommend to invest on Yellow Cab.

Thank You