



**Data Glacier**

Your Deep Learning Partner

# Exploratory Data Analysis

## G2M Case Study

**1 March, 2021**

# Agenda

Executive Summary

Problem Statement

Approach

EDA

Recommendations

# Executive Summary

- 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.
- Objective : Produce unique insights to help XYZ firm in identifying the right company for making investment.
- Dataset: 4 individual files have been provided, these are:
  - Cab\_Data
  - Customer\_ID
  - Transaction\_ID
  - City

# Problem statement

- 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.
- Objective : Produce unique 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

# Approach

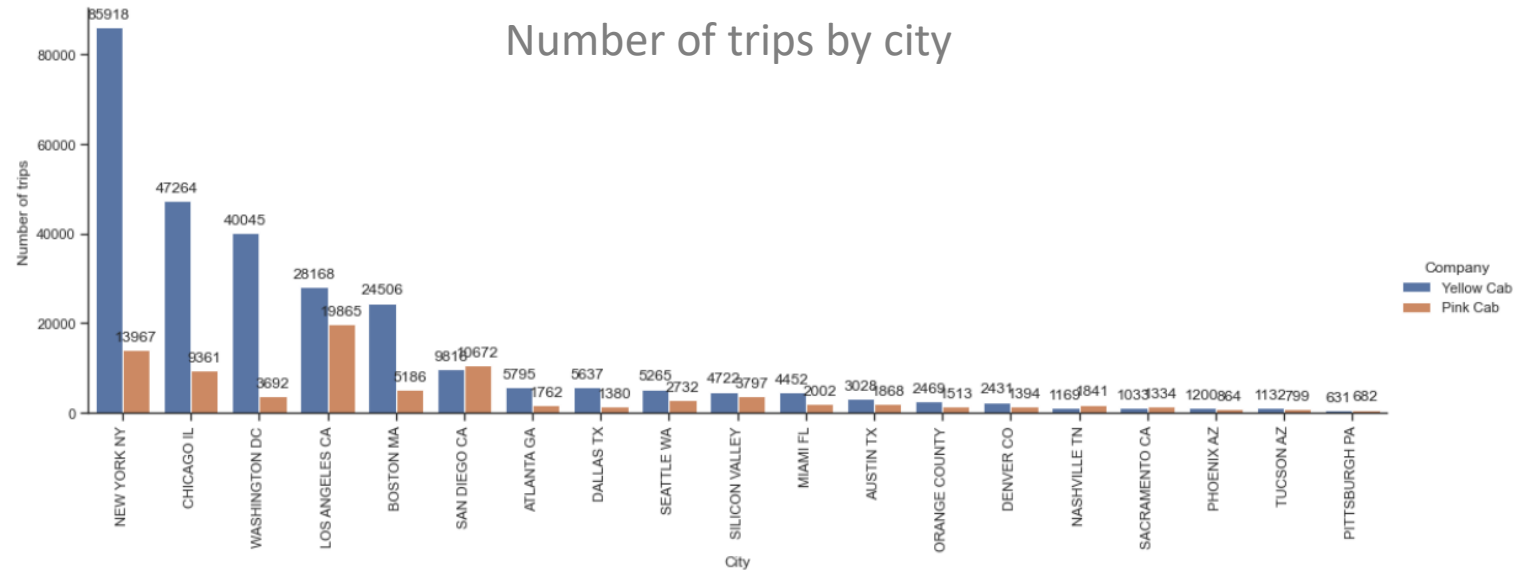
The analysis will have the following parts:

- Data Cleaning
- Data Understanding
- Communications of findings

Hypothesis:

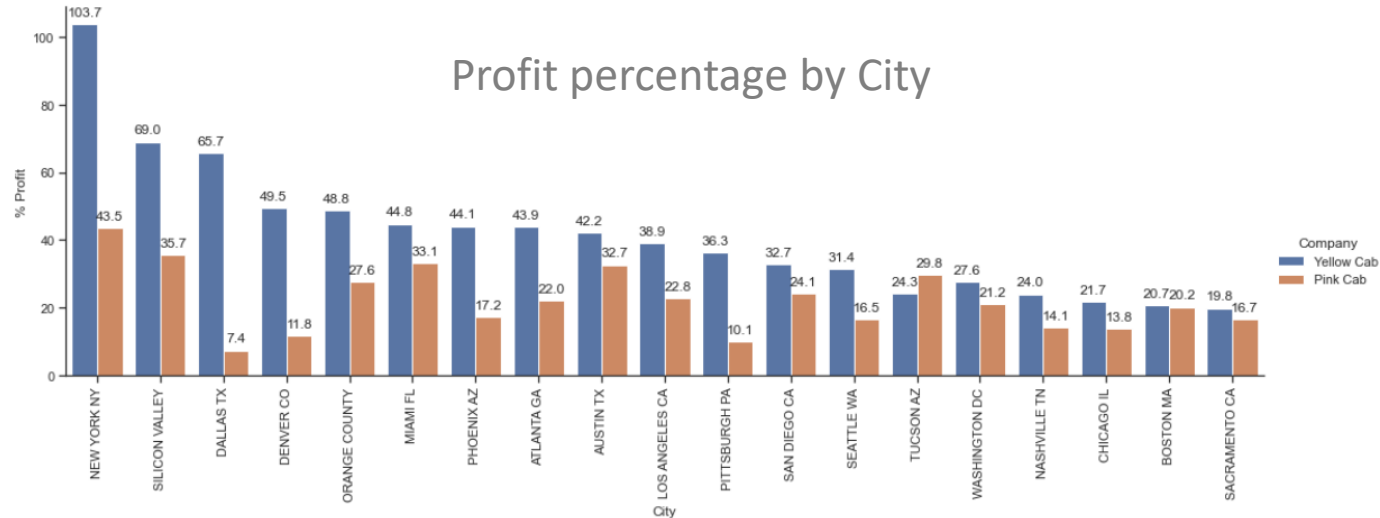
- What's is the city with more trips and the biggest percentage profit comparing the two companies using the cab service?
- What's the most frequent payment method and what more earnings gives using the cab service?
- Which are the seasons with the greatest movement in numbers of rides and more profit using the cab service?
- What's the relationship between customers' salaries and the amount of their trips using the cab service?
- Which is the company that better retains its customers?

# City Analysis

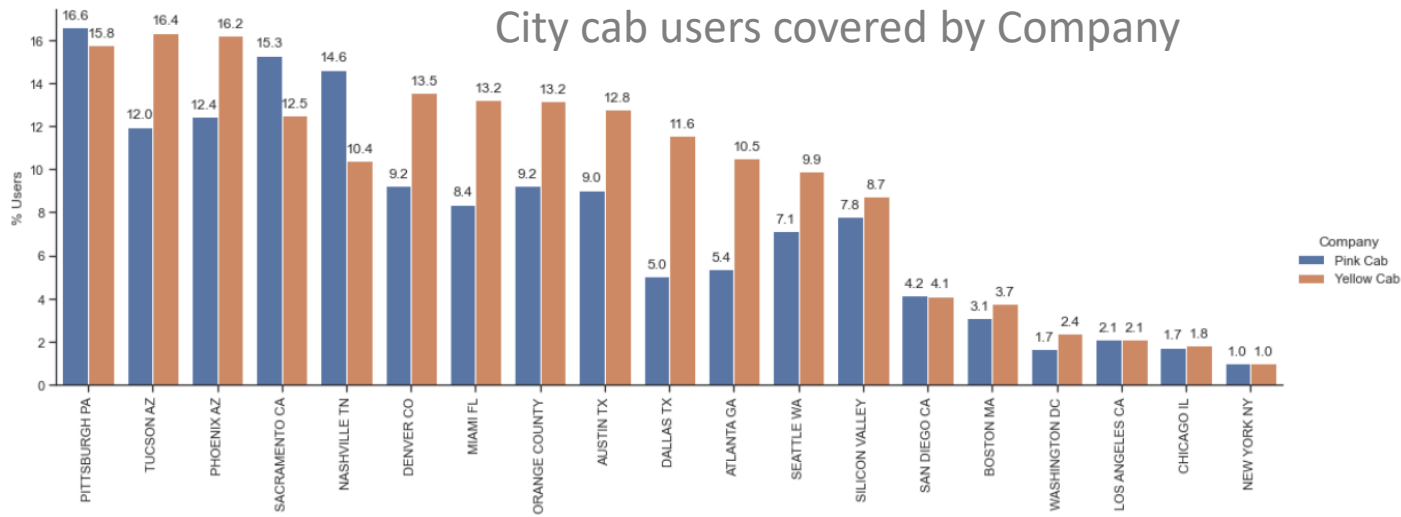


For example, Washington DC has a lot of rides, but the profit isn't very well.

The Yellow Cab Company has more rides in almost all cities, but the profit of the Pink Cab Company is not too far in many cities.

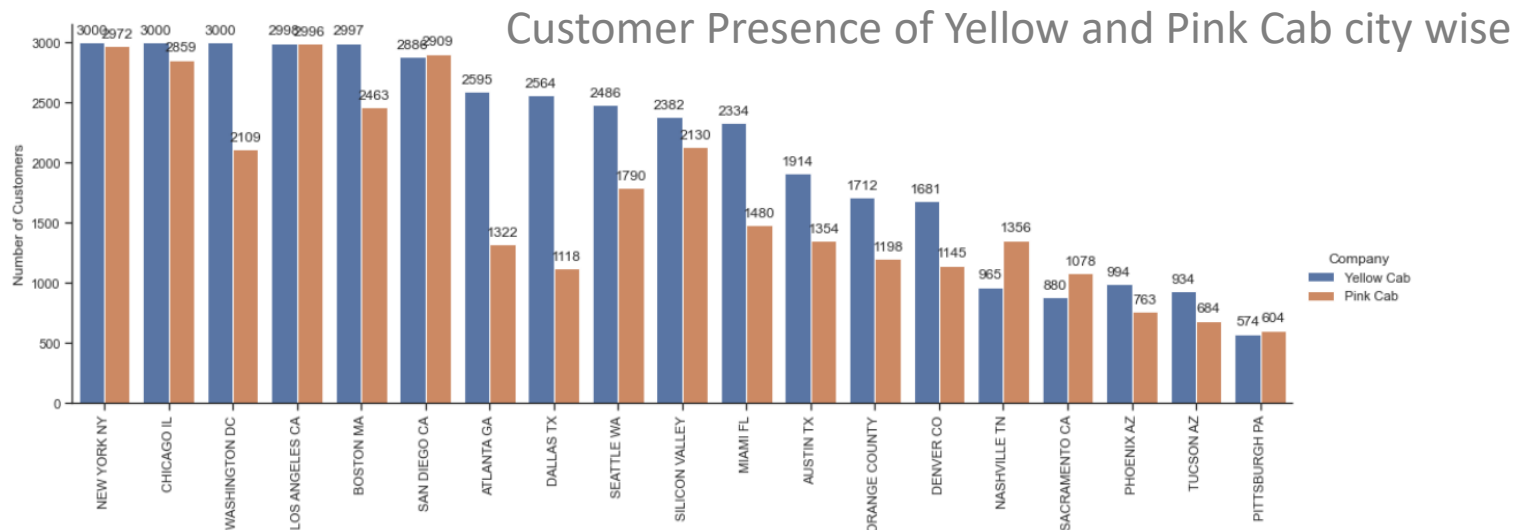


# City Analysis



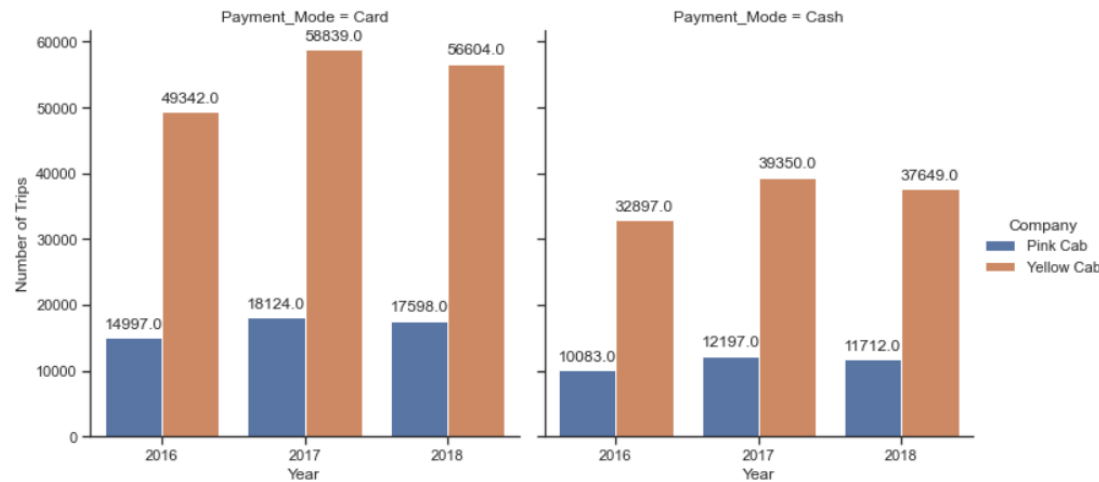
The distribution of customer in both companies is very similar, only the Pink Cap Company has more customer in the following cities:

- San Diego
- Nashville
- Sacramento
- Pittsburgh

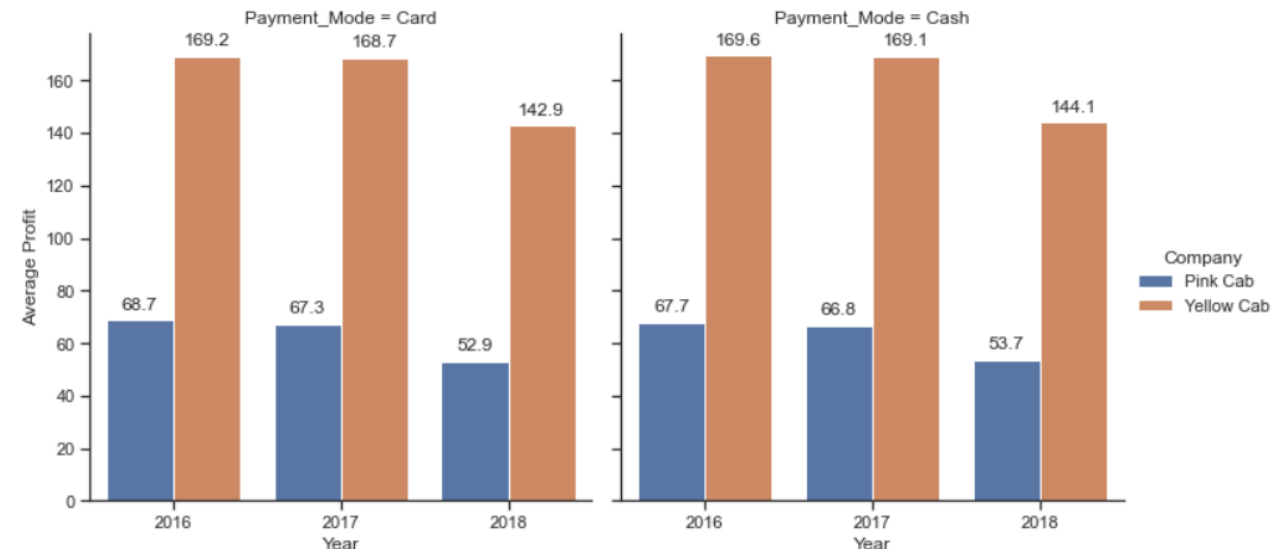


# Payment Method Analysis

## Number of Trips by payment method



## Average Profit by payment method

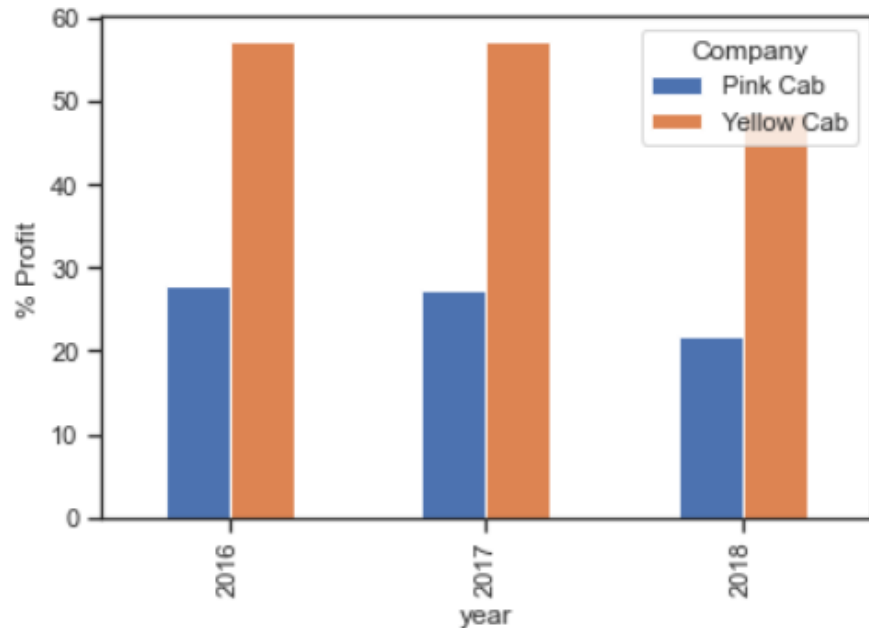


Both payment method have average profit seemed, but the payment method through card is more frequent in both companies.

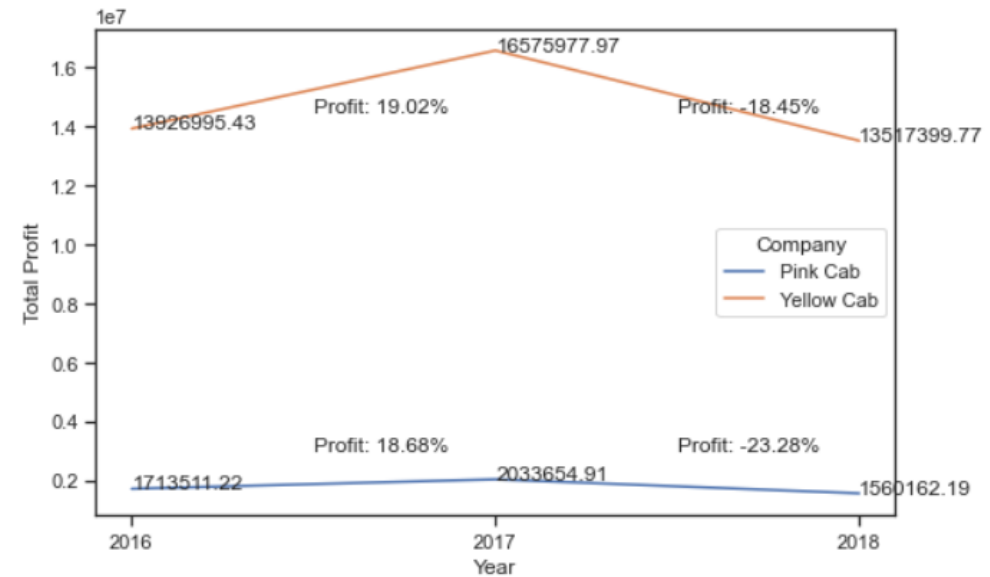


# Seasonality Analysis

Profit percentage average



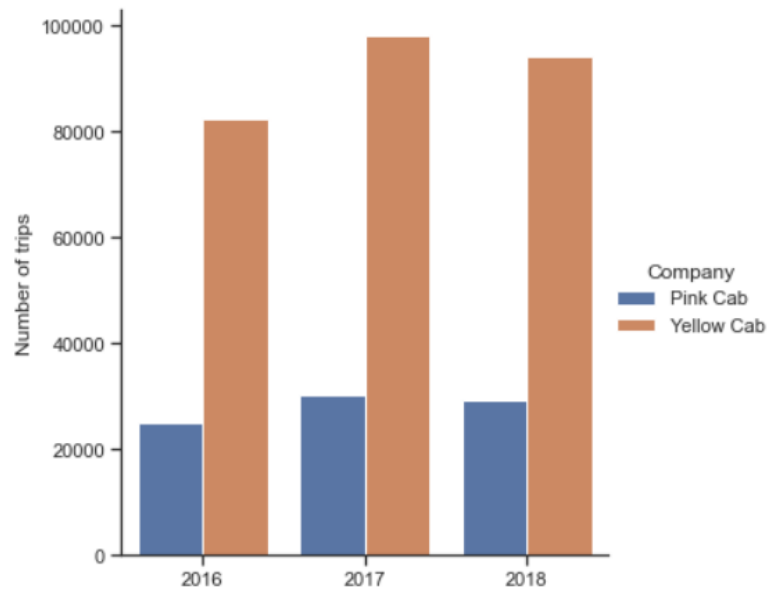
Yearly profit analysis



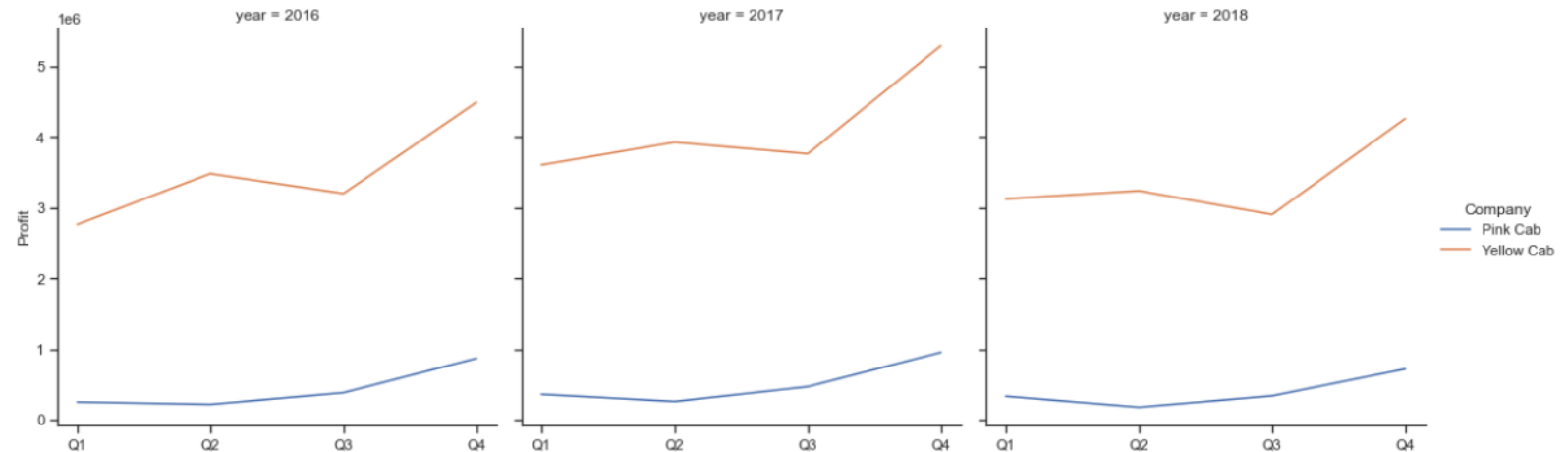
The profit on both companies is very similar, between the years 2016 and 2017 there is a improvement, and between the years 2017 and 2018, there are losses.

# Seasonality Analysis

Number of trips by year



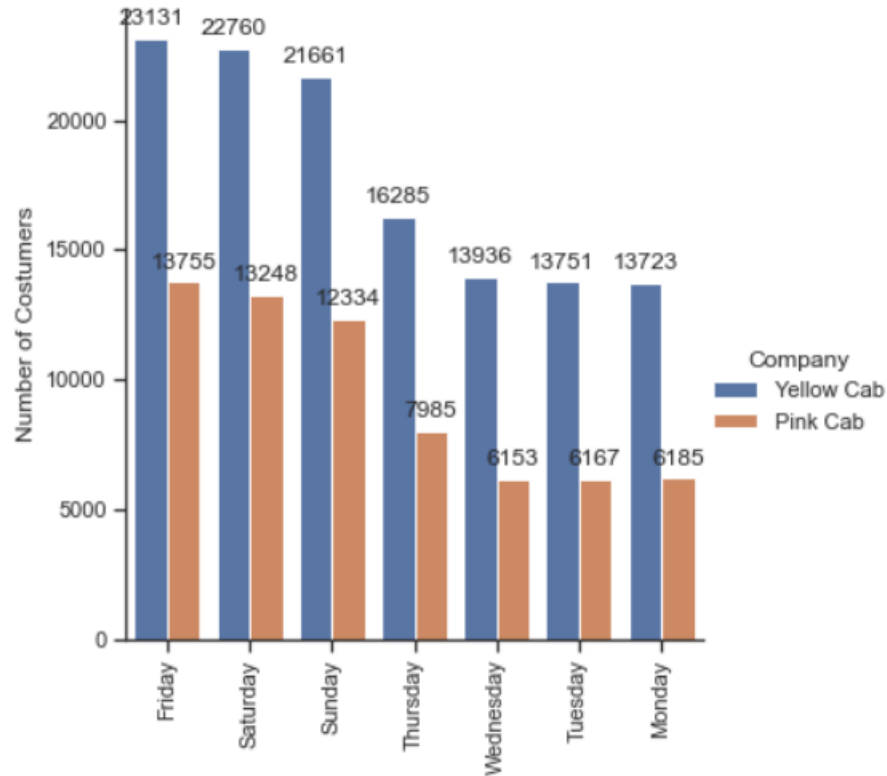
Quarterly profit analysis



There is a behavior different in both companies, for the second quartile the yellow cap profit increases, but the pink cap profit has a minimum fall. For the third quartile the yellow cap profit fall but the pink cap profit improves a little. By the end of the year, the yellow cap profit improves a lot and the pink cap profit also improves.

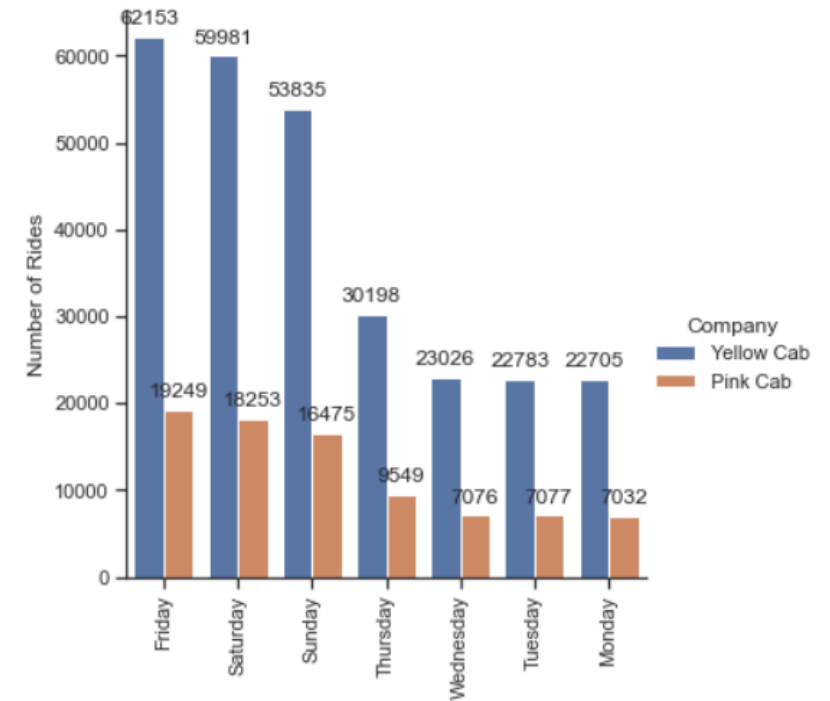
# Seasonality Analysis

Customer Analysis by day



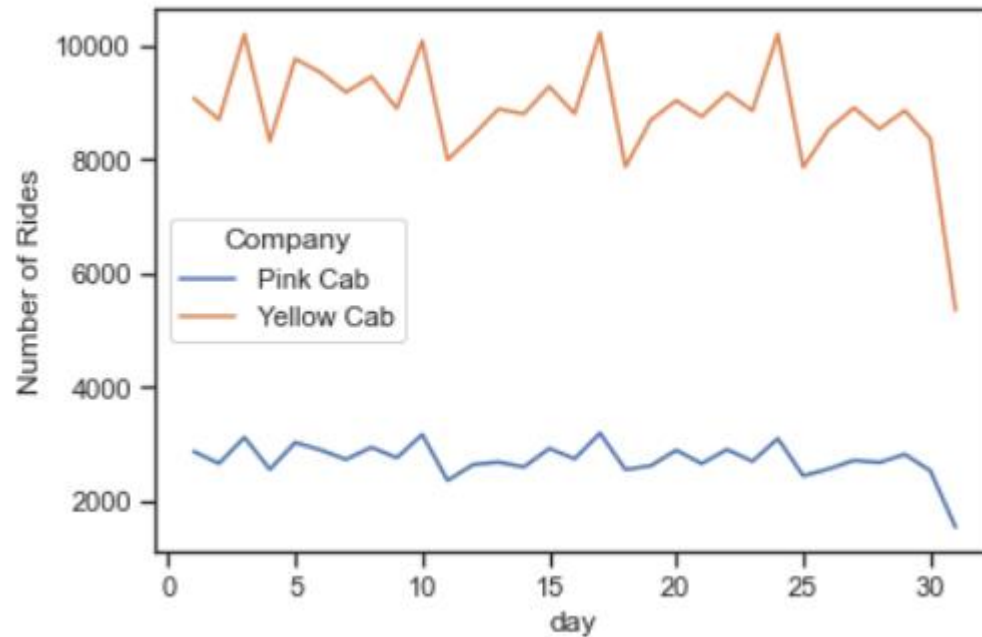
Analyzing day by day of the week, the Yellow Cab Company has more costumers and rides in each of the days.

Rides Analysis by day



# Seasonality Analysis

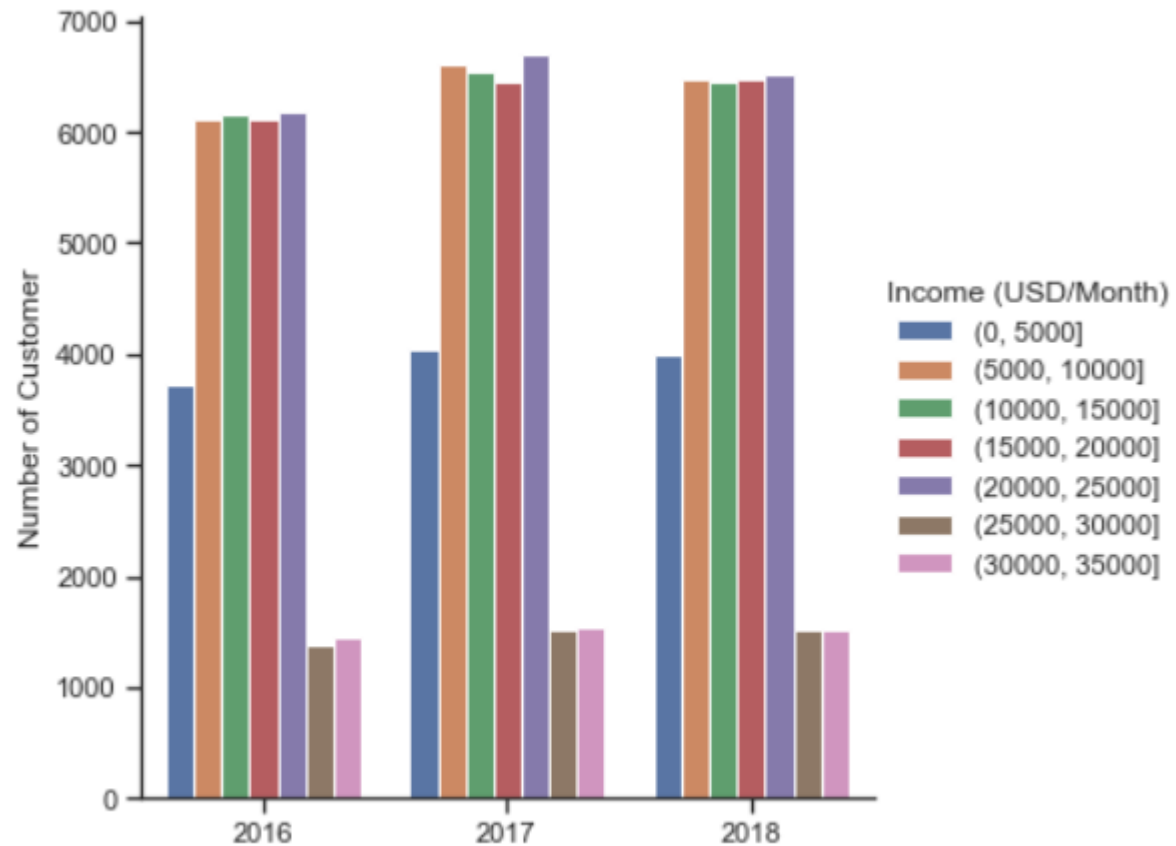
Demand by day



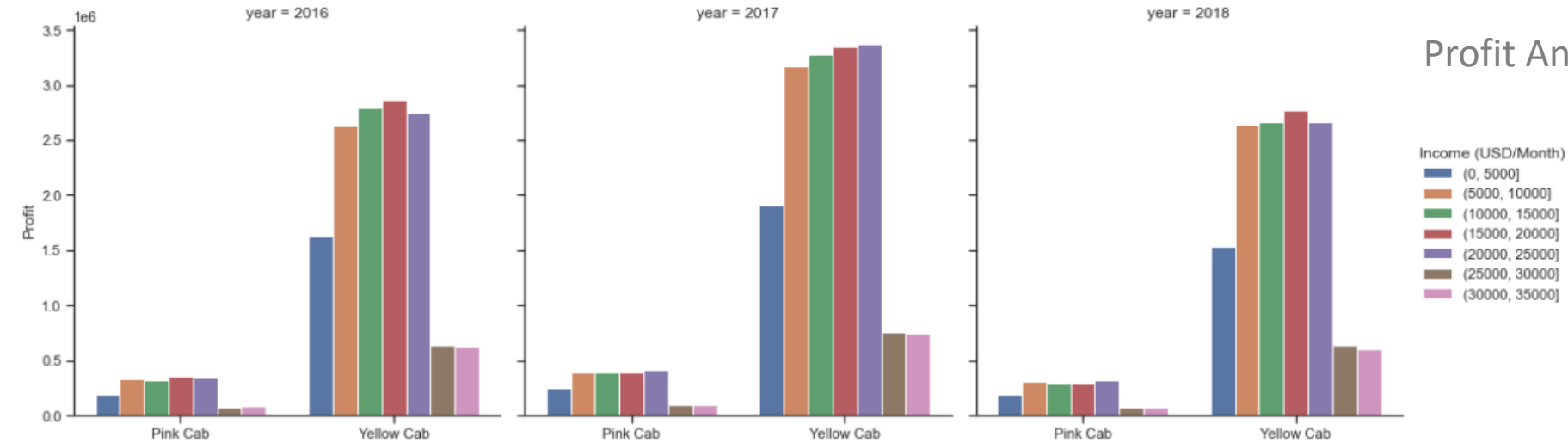
There is a similar pattern every 7 days, and this synchronizes with the Friday day in the previous analysis.

# Incomes Analysis

Number of customer by Income

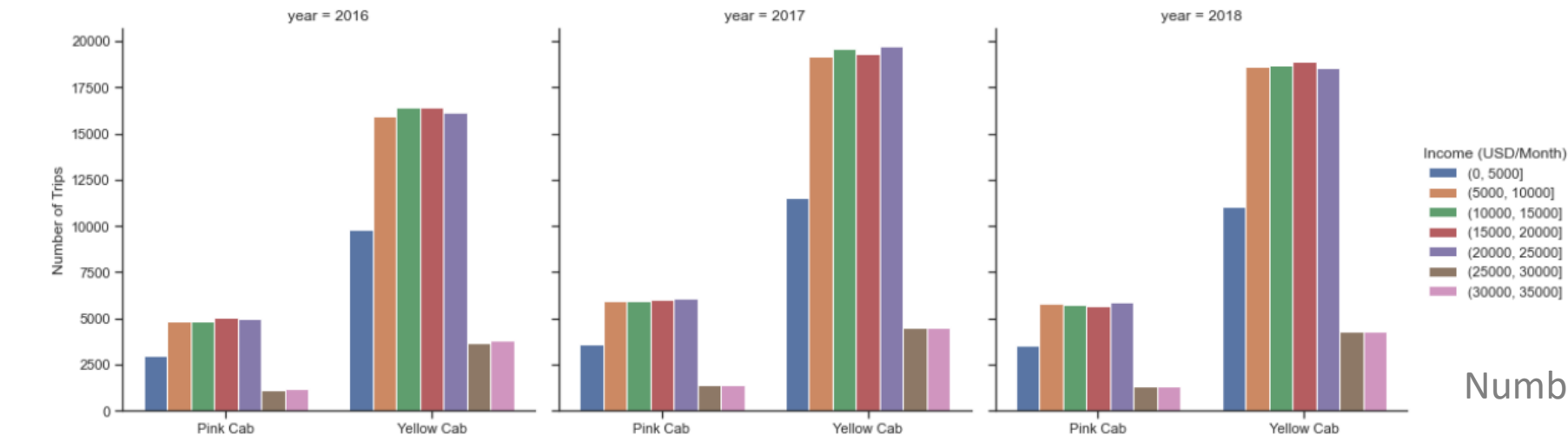


# Incomes Analysis



Profit Analysis by Income

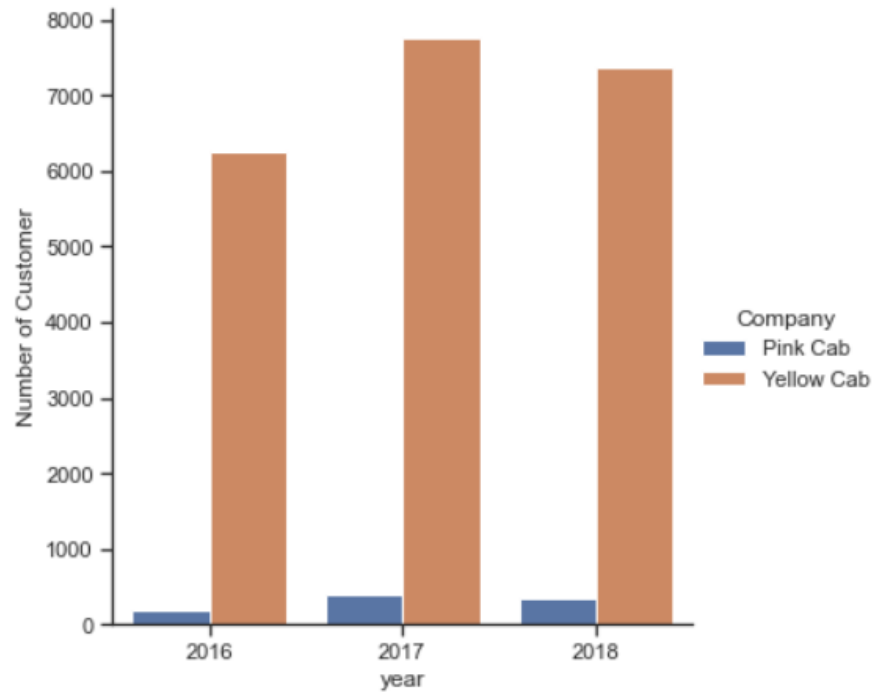
There is more numbers of trips and also the profit is better in the customer that have incomes between 5000 Dollars and 25000 Dollars, but also can related with the number of customer that have the same pattern seen in the previous graph



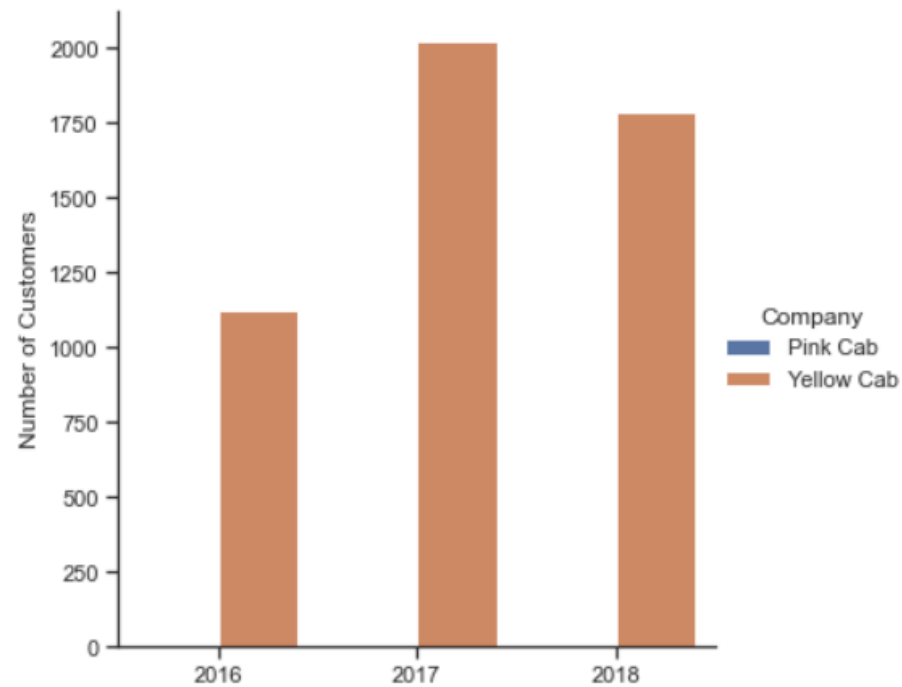
Number of trips by income

# Customer Retention Analysis

Customer retention (at least 5 rides)



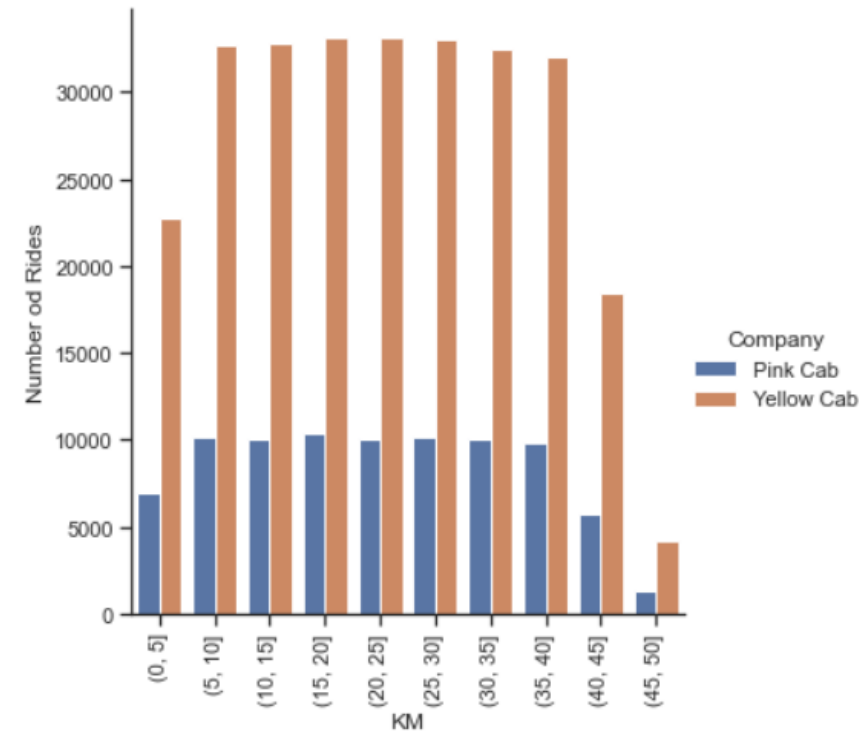
Customer retention (at least 10 rides)



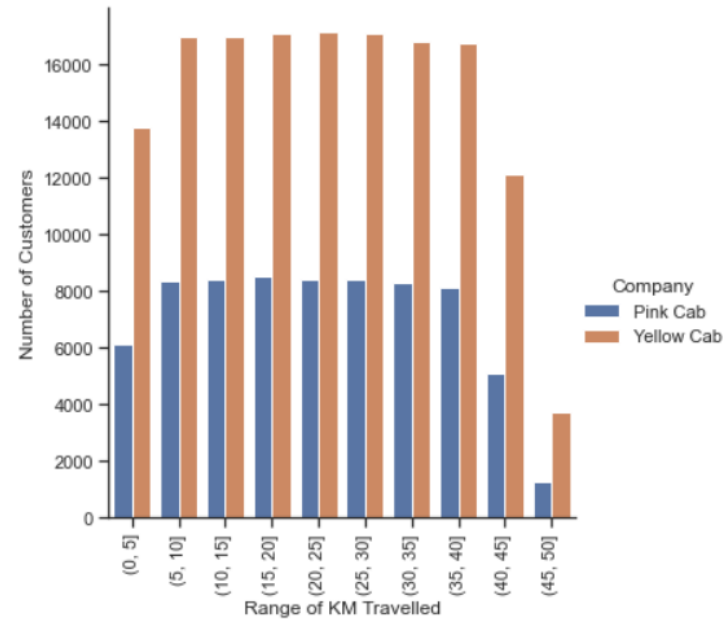
Clearly the Yellow Cap Company has a better retention of its customers.

# Ride distance Analysis

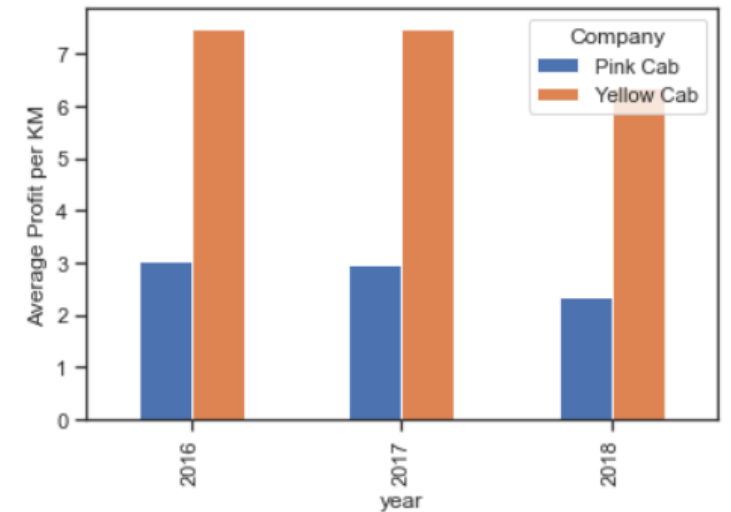
Number of Rides by Range of KM



Number of Customer by Range of KM

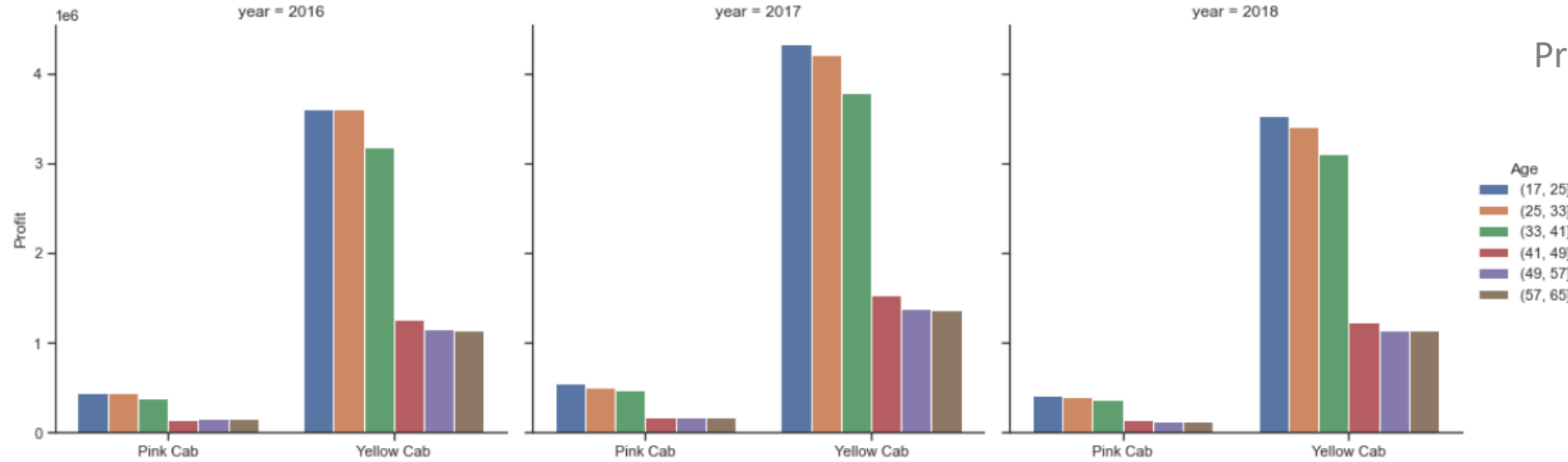


Average Profit per KM by year



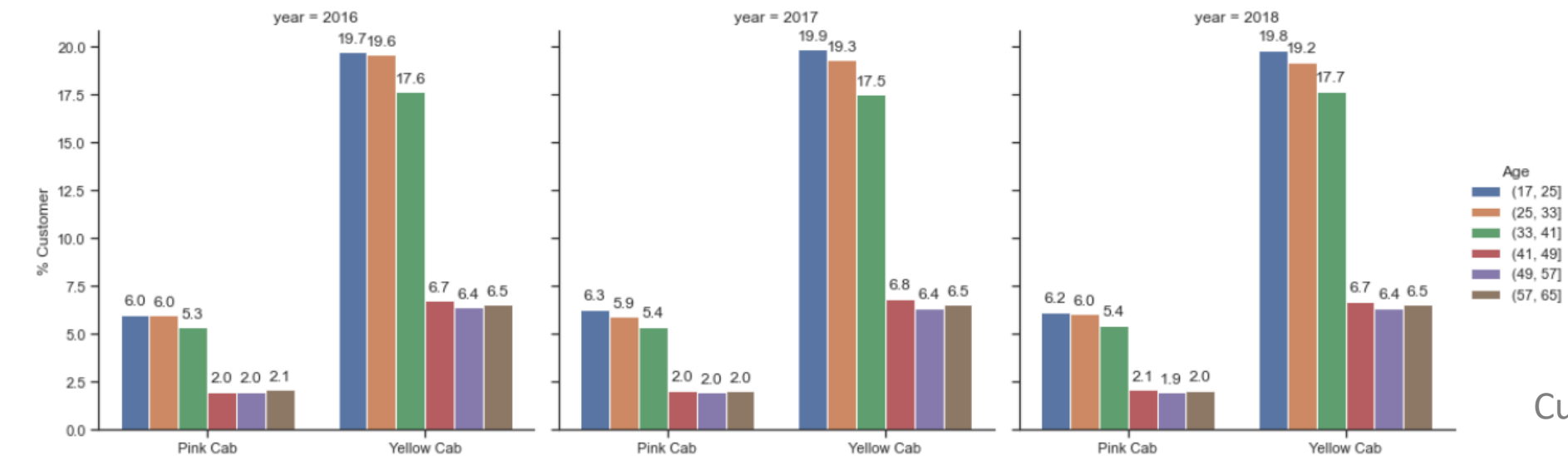


# Age Analysis



Profit by age

There are more customer who are between 17 and 41 years old.



Customer Percentage by Age

# Recommendations

The Cab Companies have been evaluated in the following points and found that:

**City reach:** Yellow cap has higher trips and customers reach in each city. In addition, Pink cap only has a city where they have a better profit and that city is Tucson.

**Payment method reach:** Clearly Yellow Cap has higher trips and average profit is better compared with Pink Cap, but both have the same pattern, improve in the numbers of trips and loss in the profit according the payment method.

**Season reach:** There is a pattern on the days of the month, It's that every 7 days there is an increase in the numbers of trips and that day is the Friday day. It is in both companies.

**Income reach:** There is a better contribution from customer with incomes between \$5000 and \$25000. But it's evident that the Yellow cap has higher contributions.

**Customer retention:** Two groups have been analyzed: at least 5 trips and at least 10 trips in each company. The conclusion is that the Yellow Cap has more costumers in these segments.

**Based on these conclusions, it is suggested to invest in the Yellow Cap Company.**

# Thank You