Exploratory Data Analysis

Cab Investment

11/12/2021

Agenda

- Overview
- The Datasets
- The Hypothesis
- General Look
- Profit Analysis
- Demand Analysis
- Loyalty Rates
- Payment Mode Distribution
- Conclutions

Overview, The Problem & The Objective

XYZ is a private firm in US. 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. There are two companies that XYZ would invest in one of them.

The objective is to provide actionable insights to help XYZ firm in identifying the right company for making investment.

The Datasets

Four individual datasets are provided. (Time period of data is from 31/01/2016 to 31/12/2018):

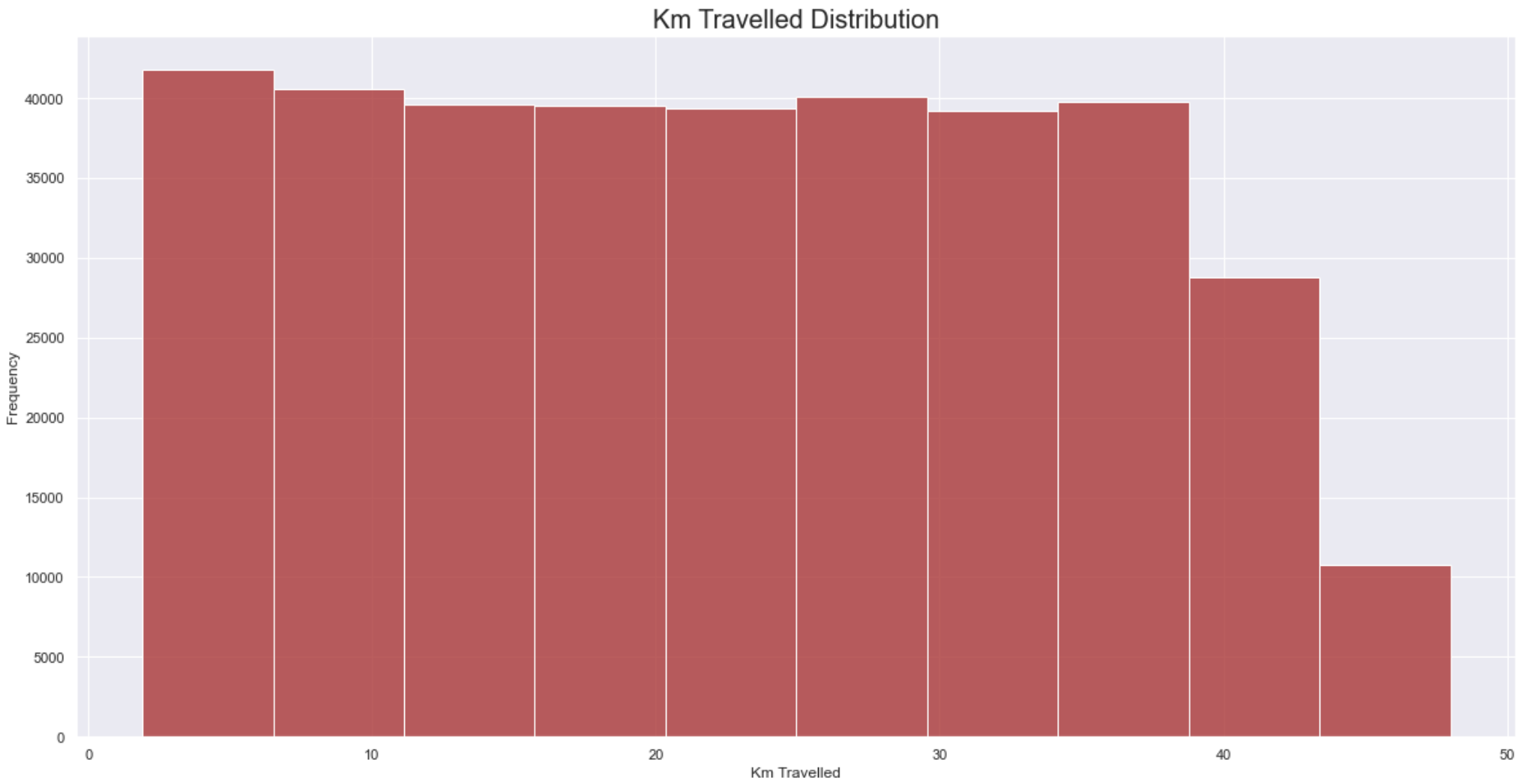
1. **Cab Dataset:** this dataset includes details of transaction for 2 cab companies.
2. **Customer Dataset:** this is a mapping table that contains a unique identifier which links the customer's demographic details.
3. **Transaction Dataset:** this is a mapping table that contains transaction to customer mapping and payment mode.
4. **City Dataset:** this file contains list of US cities, their population and number of cab users.

The Hypothesis

The hypothesis are:

1. The difference in customer's number gender-wise happens due to income inequality.
2. Does the demand of the cab industry tends to increase over time?
3. The percentage of profitable trips change by city.
4. How the demand varies according to age-wise?
5. Loyalty Rates.
6. Does the payment method fluctuate year-wise? Age-wise? City-wise?

General Look: The Frequency of Travelled km

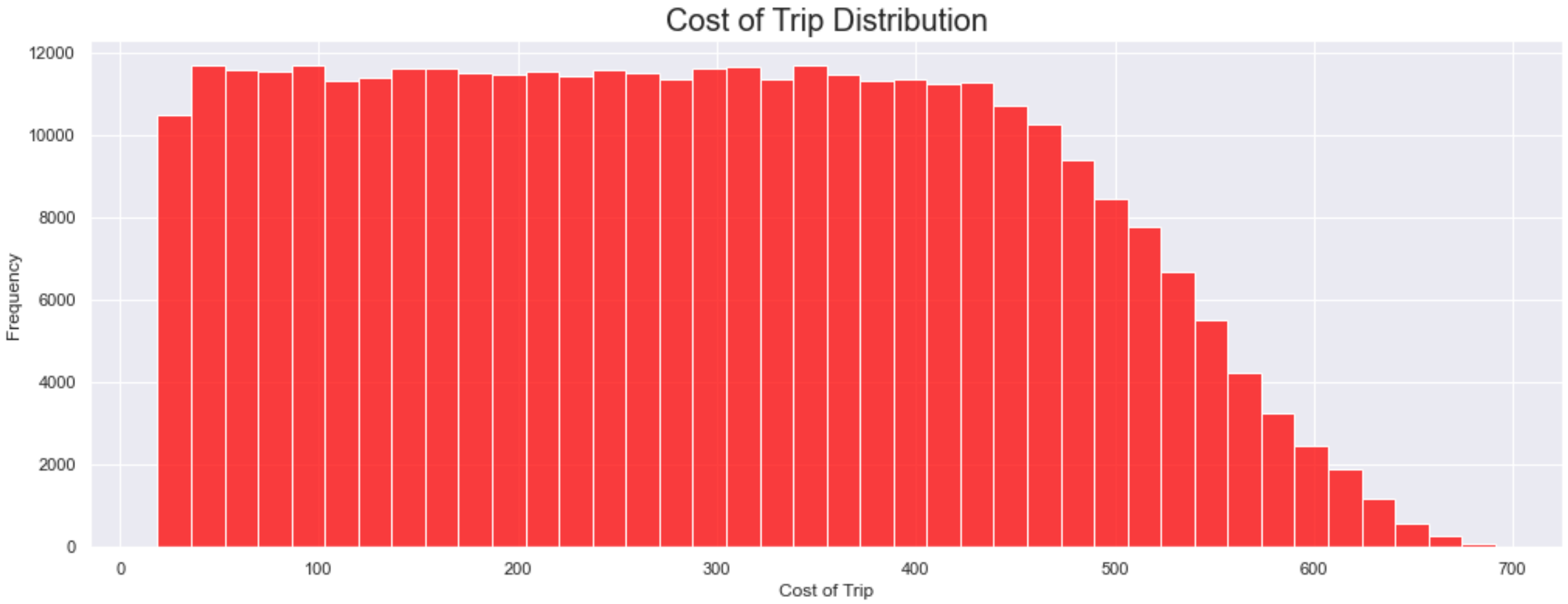


The number of trips decreases above 40 km.

General Look: The Frequency of Price Charged

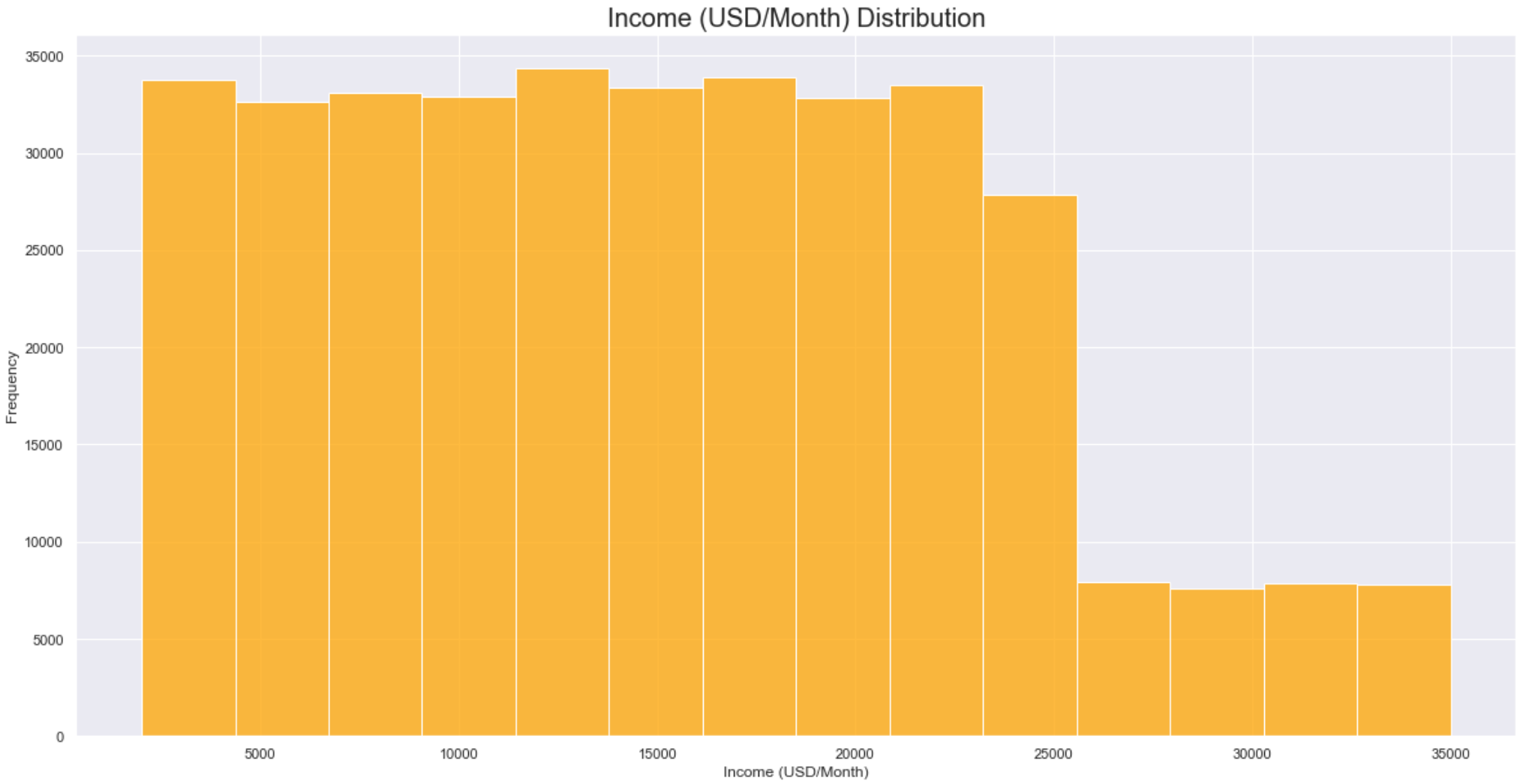
Most of customers paid between 0$ to 500$.

General Look: The Frequency of Cost of Trip



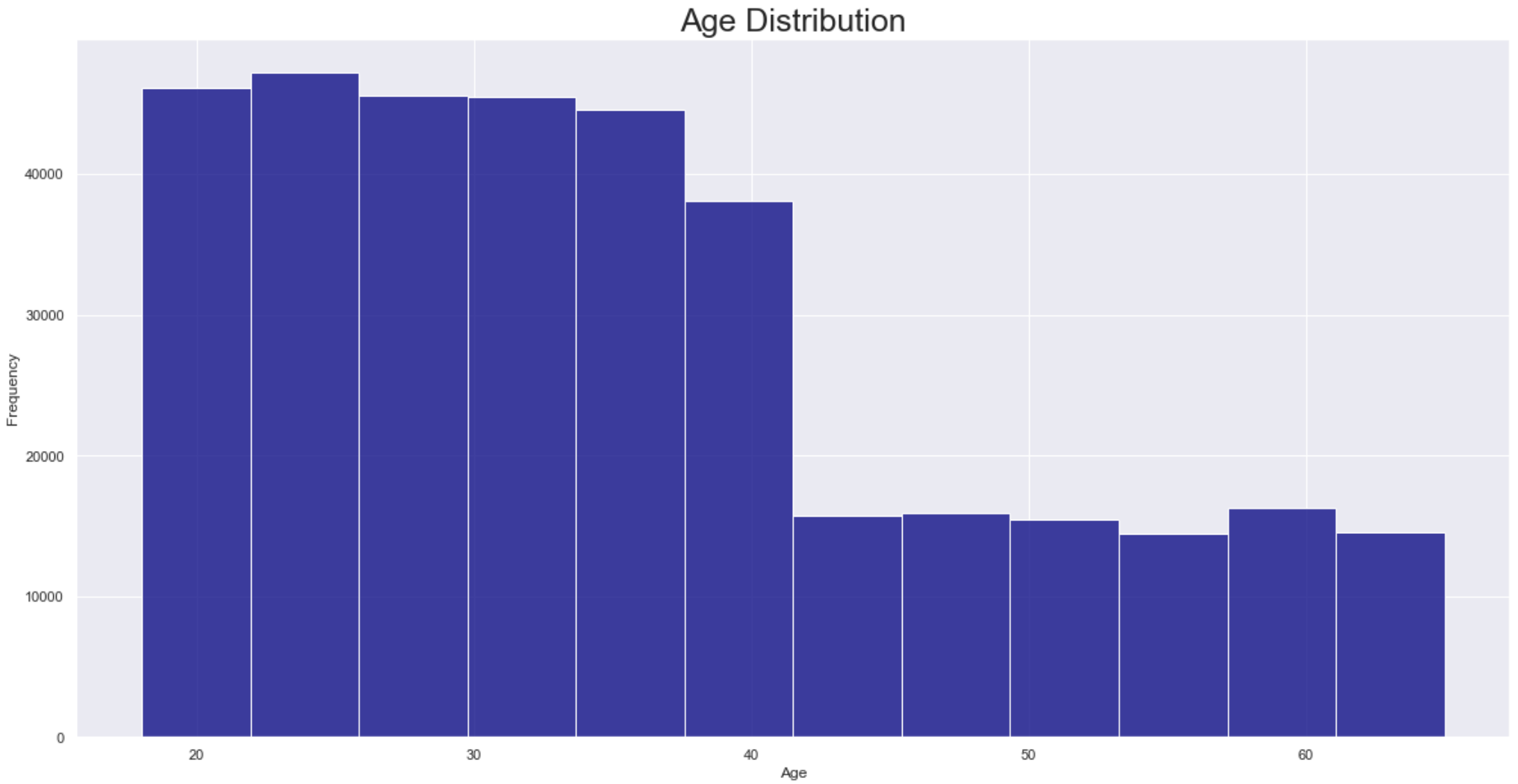
Most of trips cost between 0$ to 500$ then it start decreases.

General Look: The Frequency of Customers Income



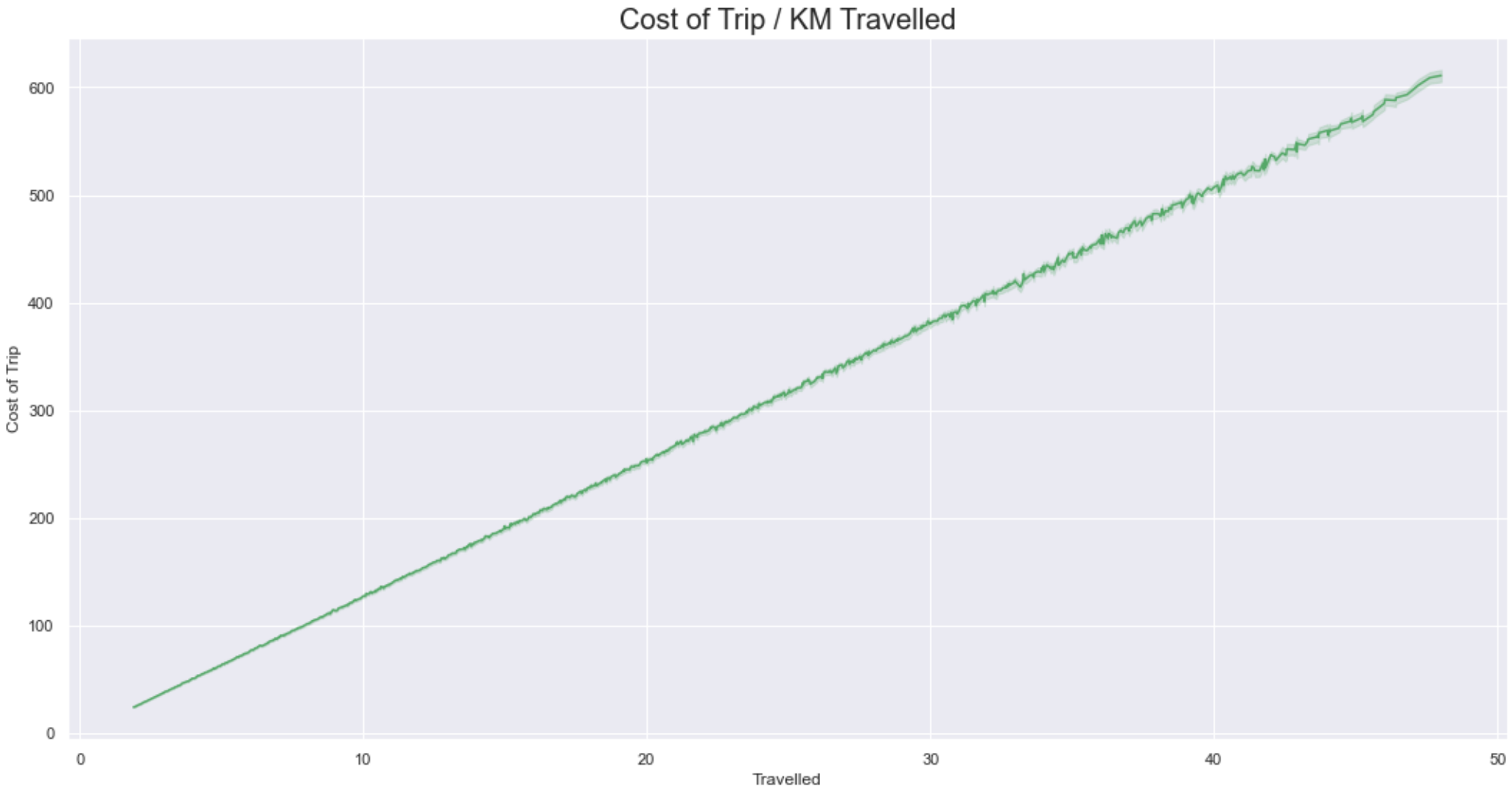
Most customers has monthly income between 0$ to 25000$.

General Look: The Frequency of Customers Income



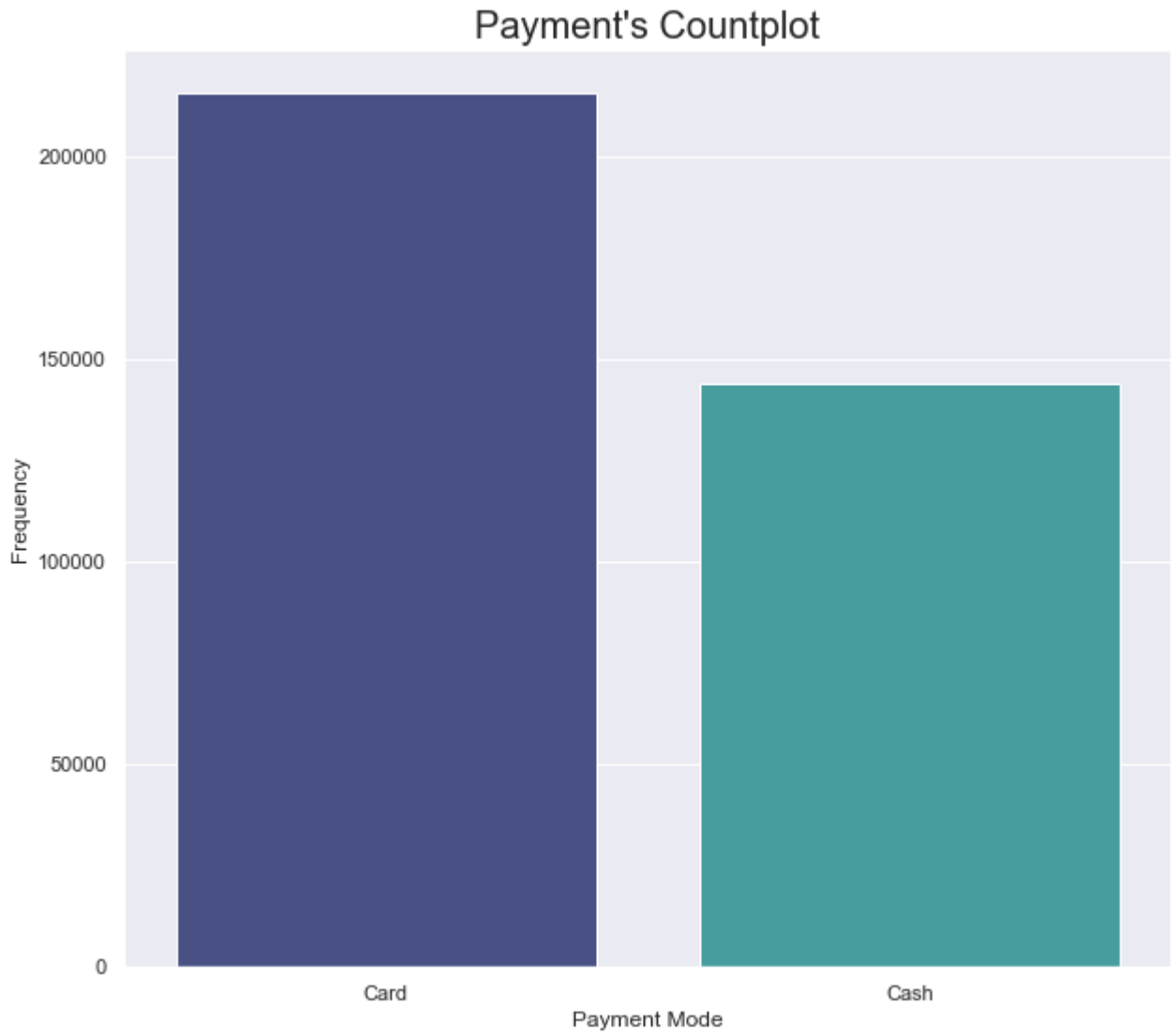
Most customer's age between 18 to 40.

Cost of Trip per KM Distribution



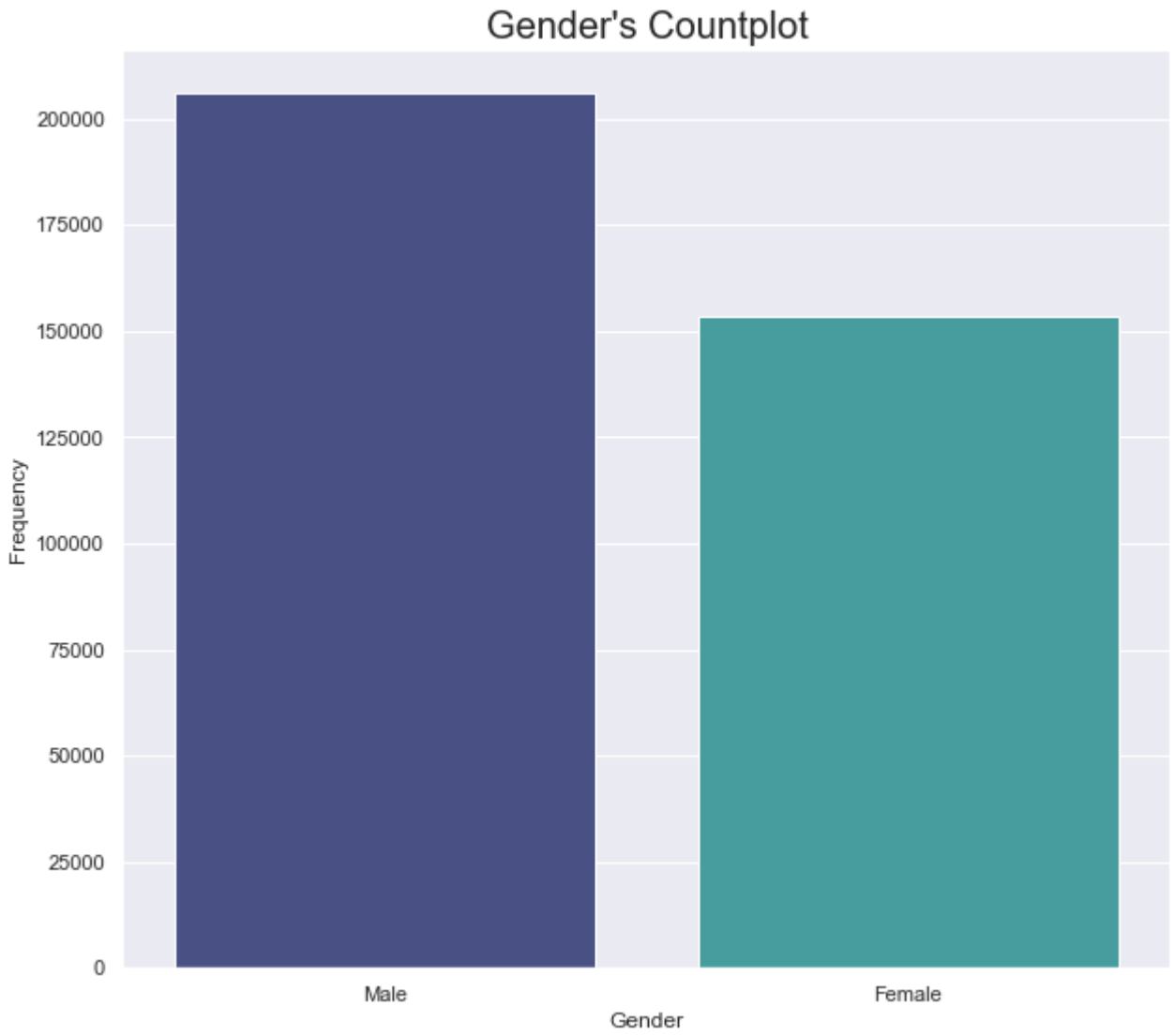
The Cost increases as distance increases.

Payment Mode Count



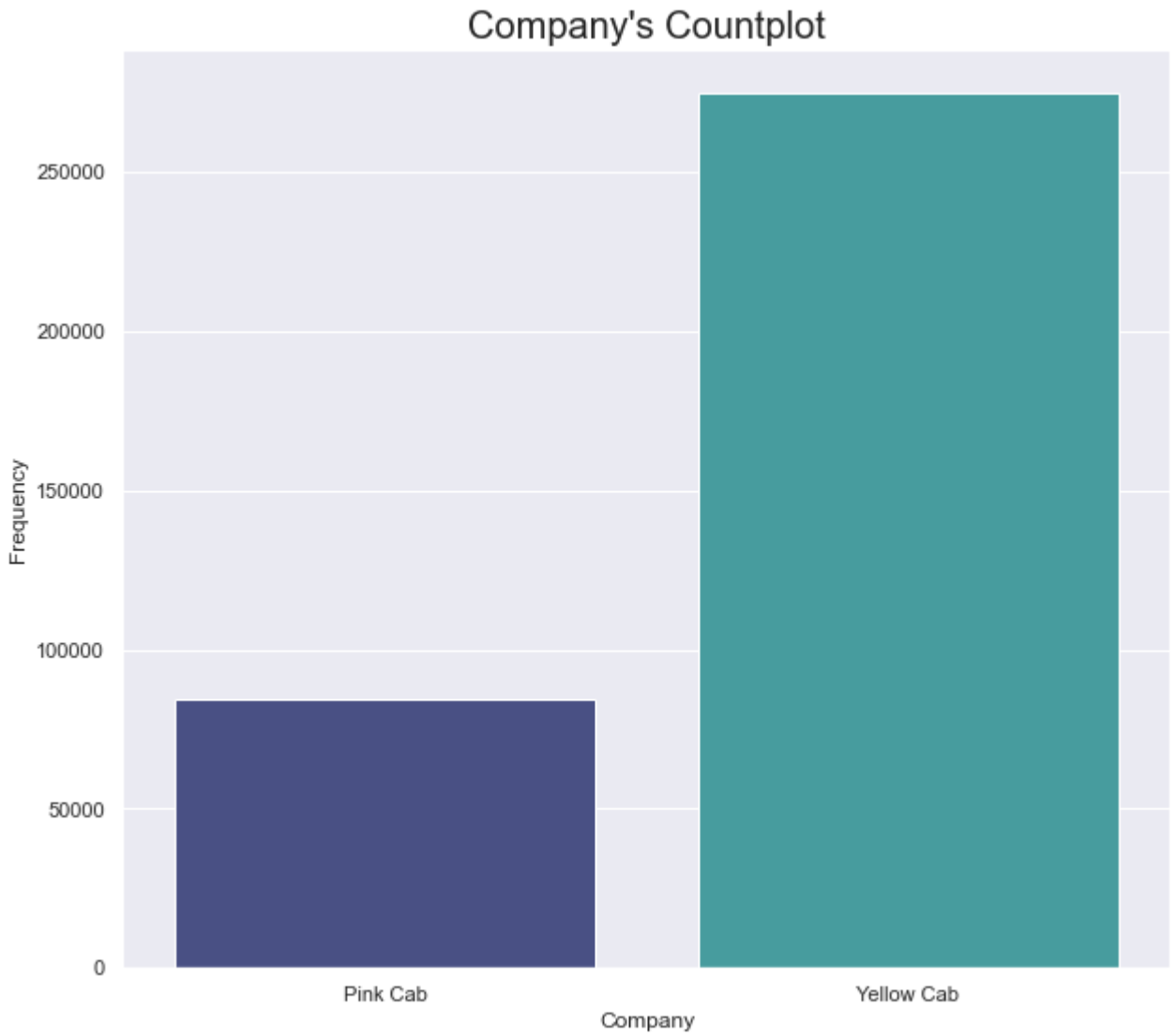
There are 71616 customers preferred to pay using card over cash.

Customer Gender Count



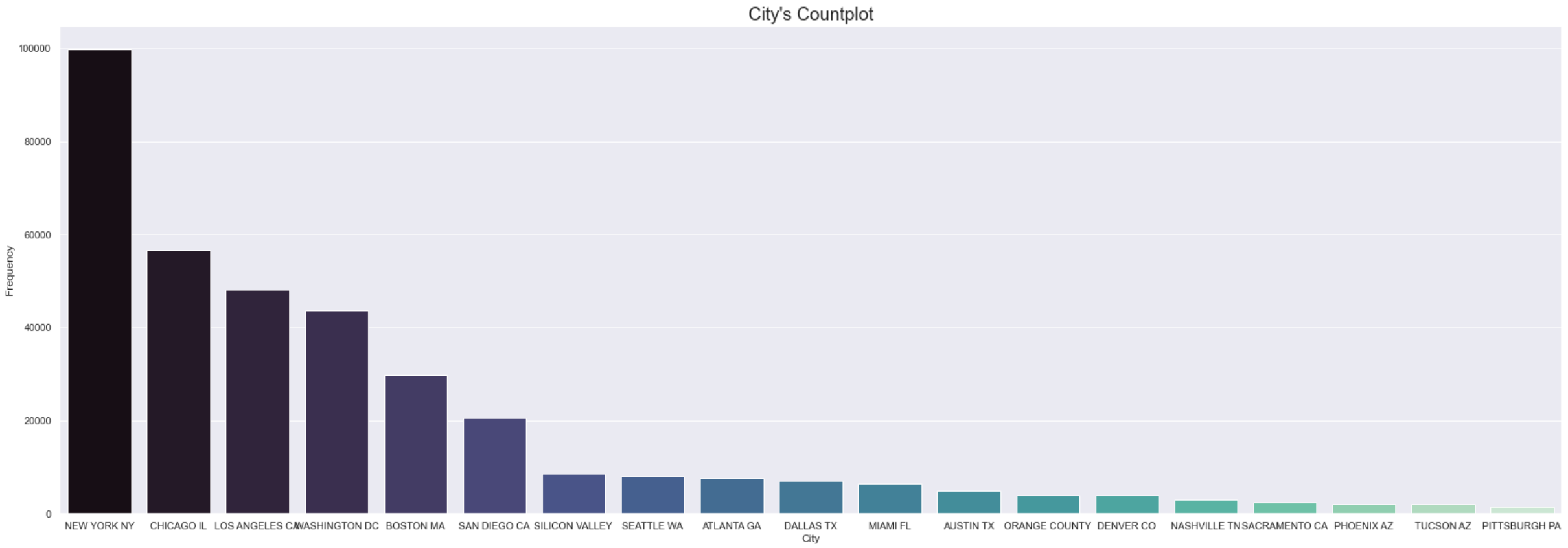
There are 52432 male customers more over female customers.

Yellow and Pink Companies Trips Count



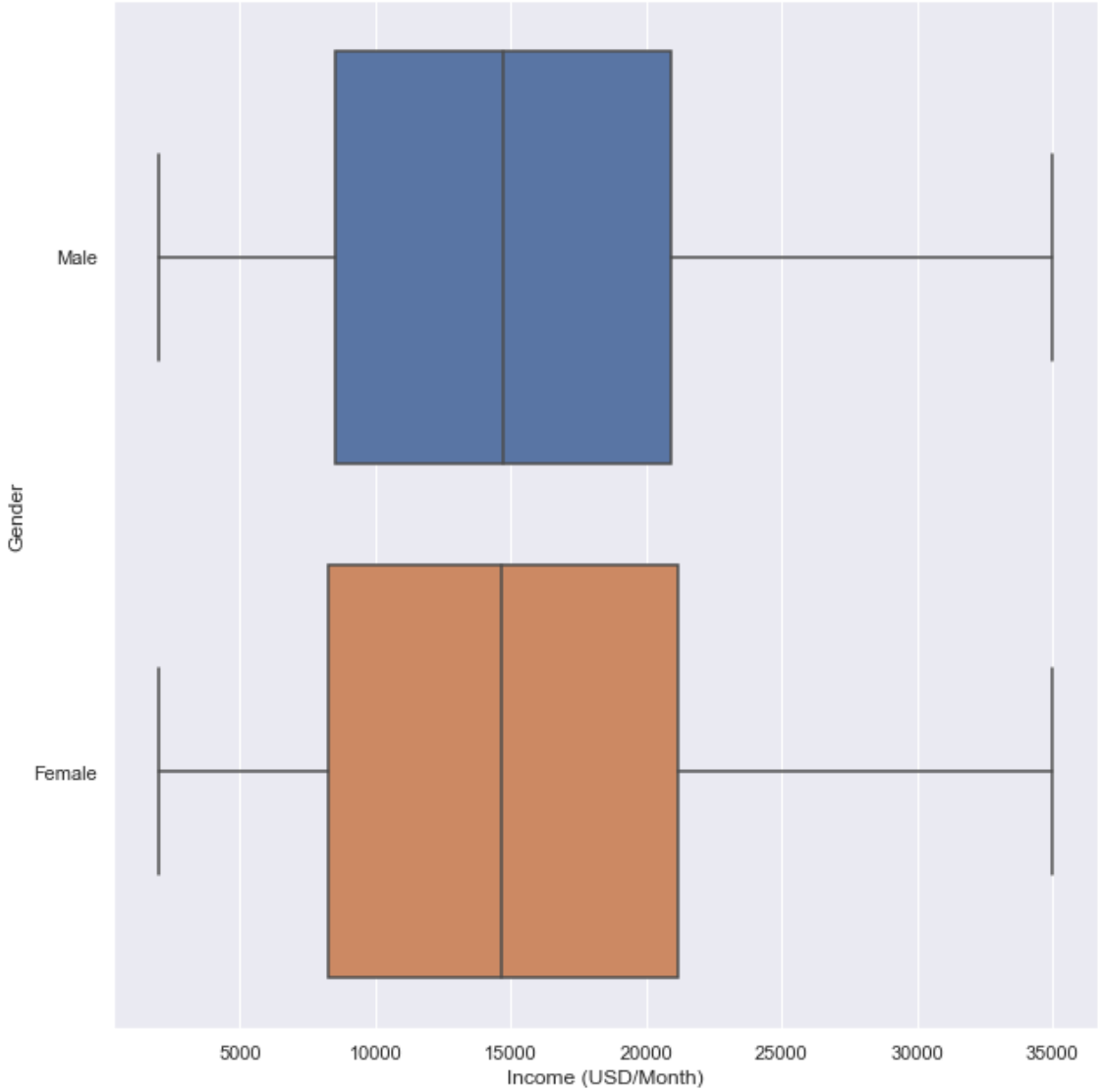
The customers made 189970 more trips provided by the yellow company over the pink company.

City-wise Trips Count



New York has the highest trips count then Chicago and Los Angeles after that.

Customer's Income Gender-wise

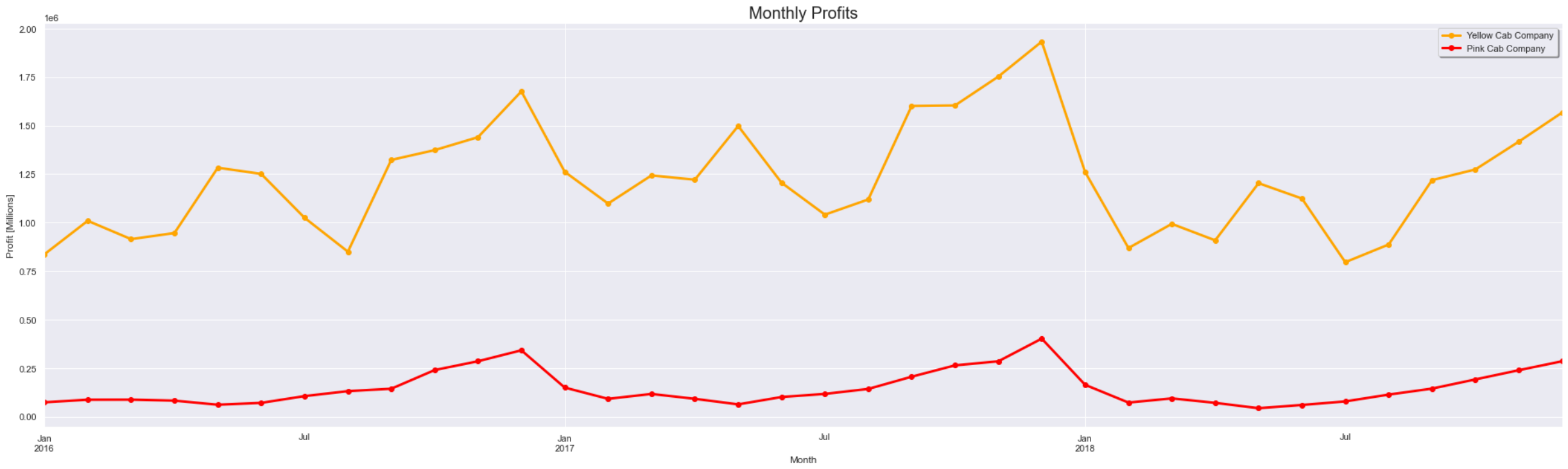


There is no difference between male & female customers' income. Which an answer to the 1st hypothesis.

Profit Analysis: Yearly Compare

The Yellow company profits over the last 3 years are eight times higher than the Pink company and the yellow company has 38,713,045$ more total profit than the pink company.

Profit Analysis: Monthly Compare

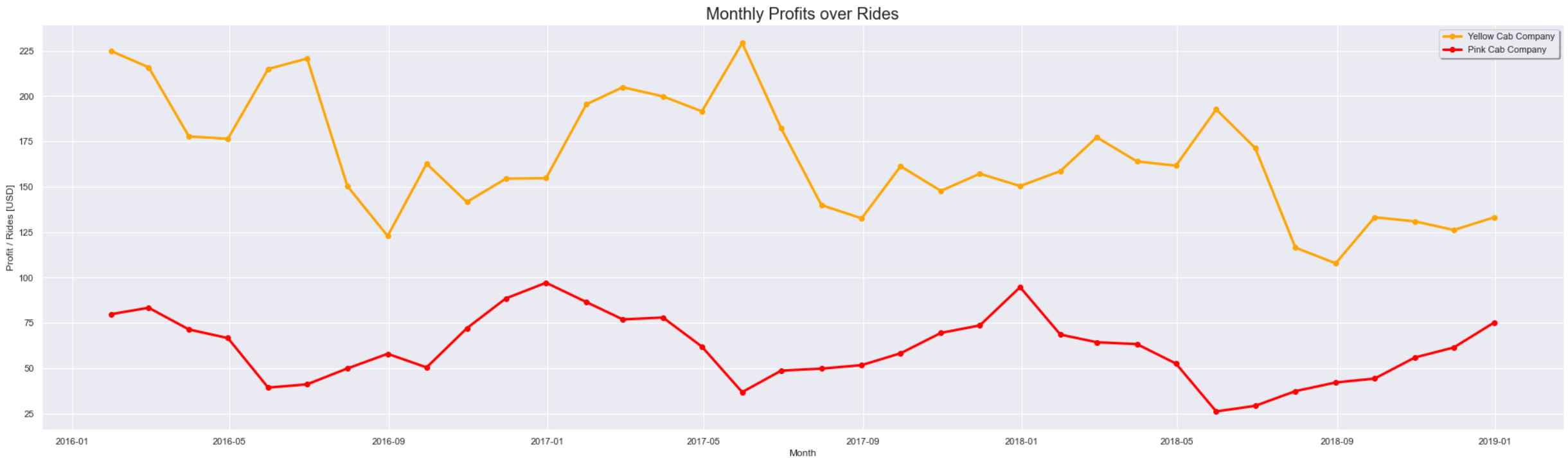


Over months, the yellow company's earnings are more stable, with fluctuations of 23%, while those of the pink company's vary in the order of 61%.

Annual Profits Over Rides

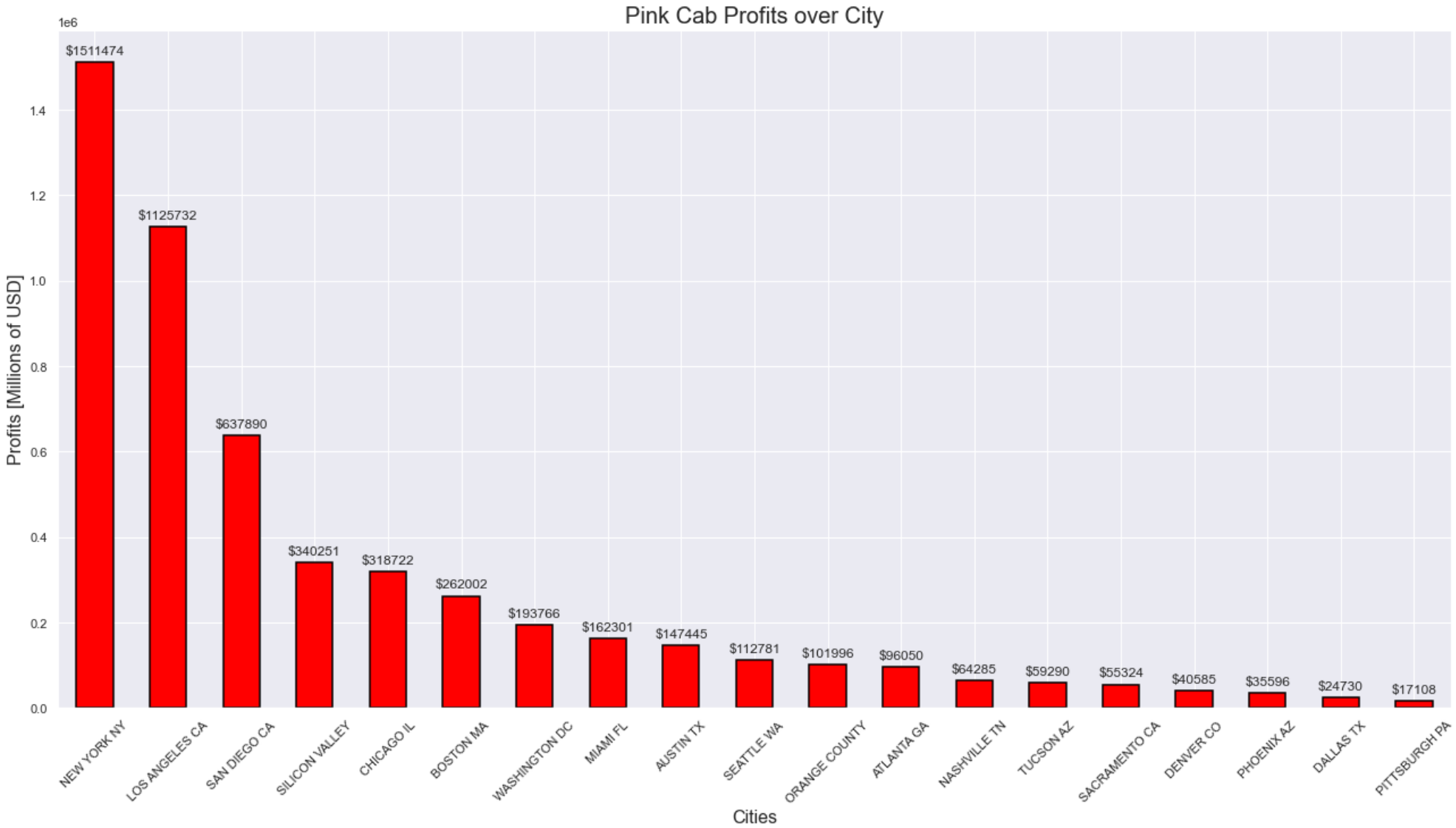
Both companies shows the same behavior, decrease over time.

Profits Over Rides: Monthly Compare



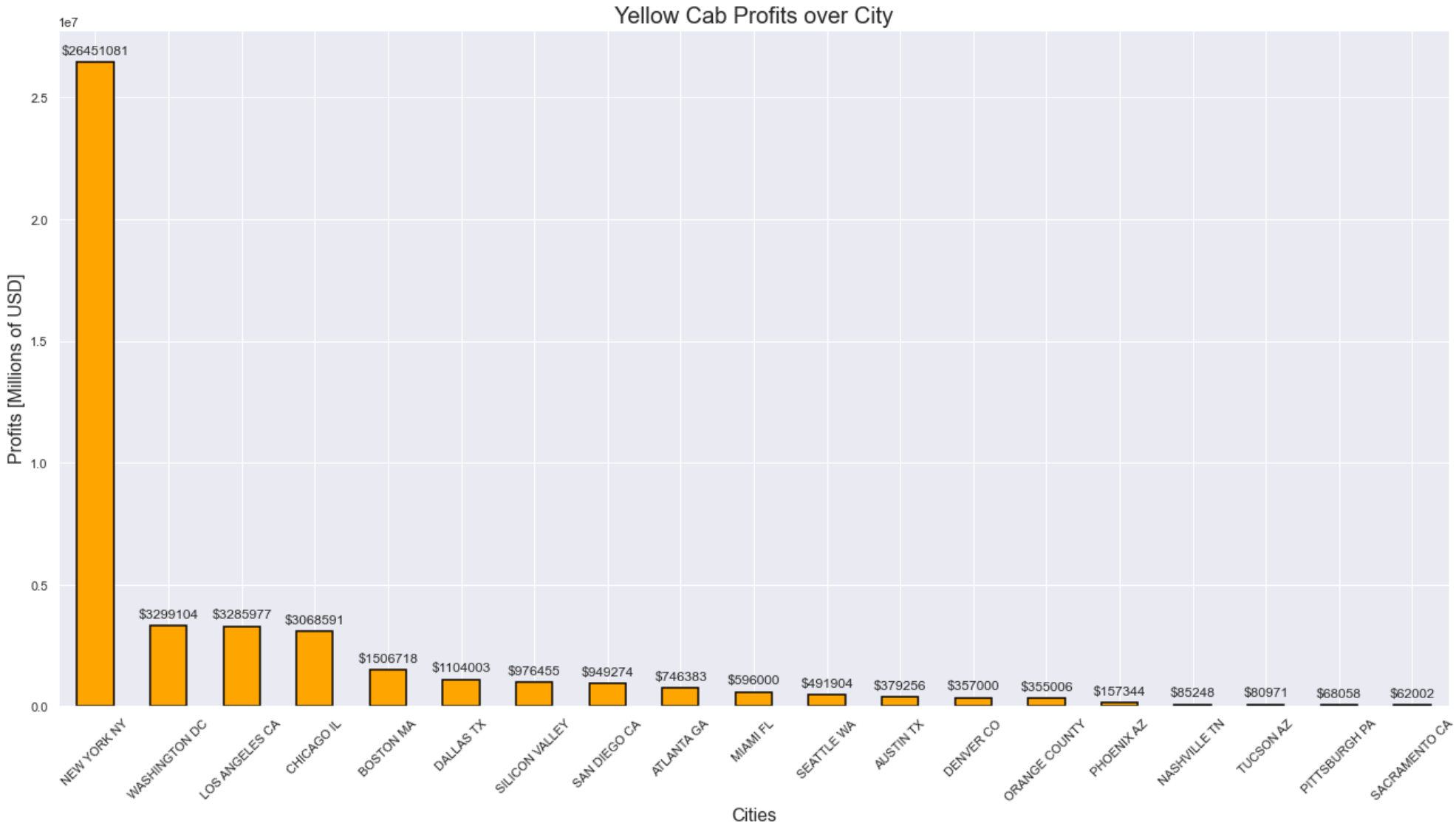
The yellow company is doing far better in profits per ride.

Average Profits per City: The Pink Company



The pink company's highest cities in terms of profits are New York has then Los Angeles and San Diego after that.

Average Profits per City: The Yellow Company



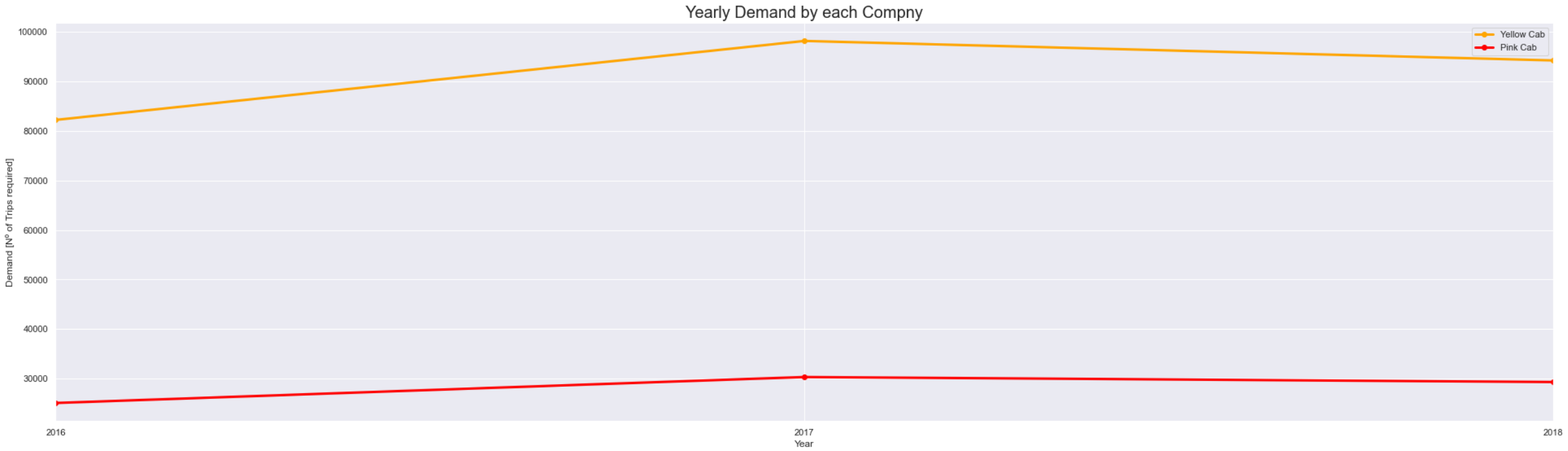
The yellow company's highest cities in terms of profits are New York has then Washington and Los Angeles after that. The yellow company has greater market share in every city.

Profit Analysis Summary

Profit per Ride: is an indicator that measures how efficient the company is, in terms of operational costs.

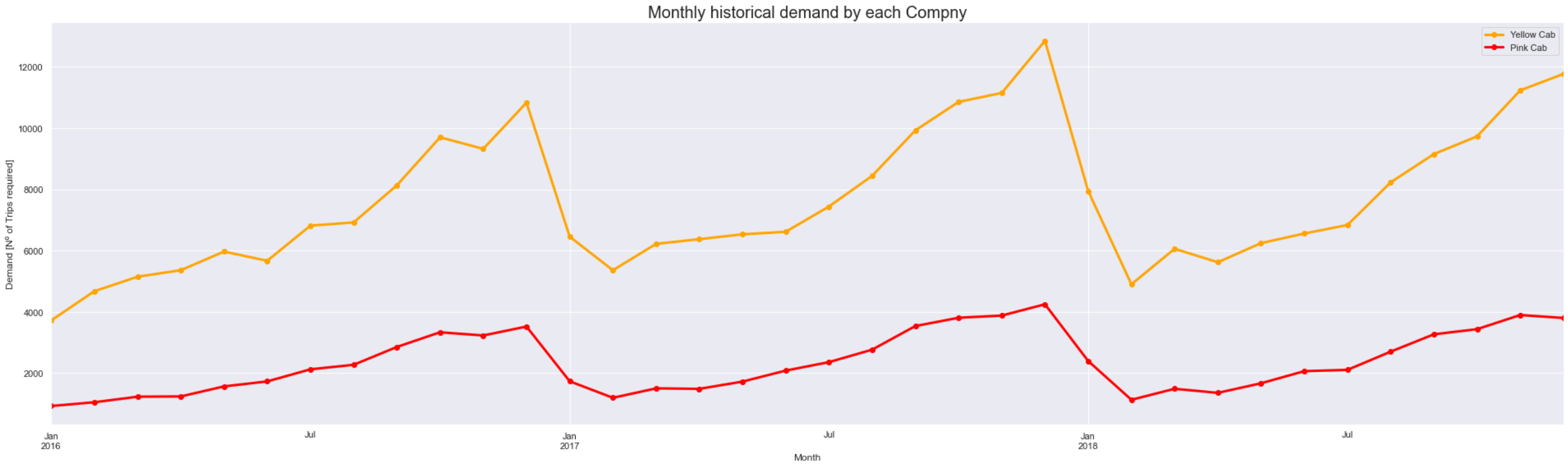
Now, we can answer the third hypothesis. Profitable rides change by cities, and of course - according to the Company- some differences assuming that at a profitable drive rate of 80% the operation performs well which is the case with the yellow company. It has a high performance according to it’s operations, maintaining a high level of profitable rides in every city. Pink Cab has a great too, except in Seattle to Pittsburgh. Moreover, Pittsburgh and Dallas performance are not Good at all, with a percentage of 61%. This means that in those cities, every 10 rides, almost 4 rides are not profitable.

Demand Analysis: Yearly Compare

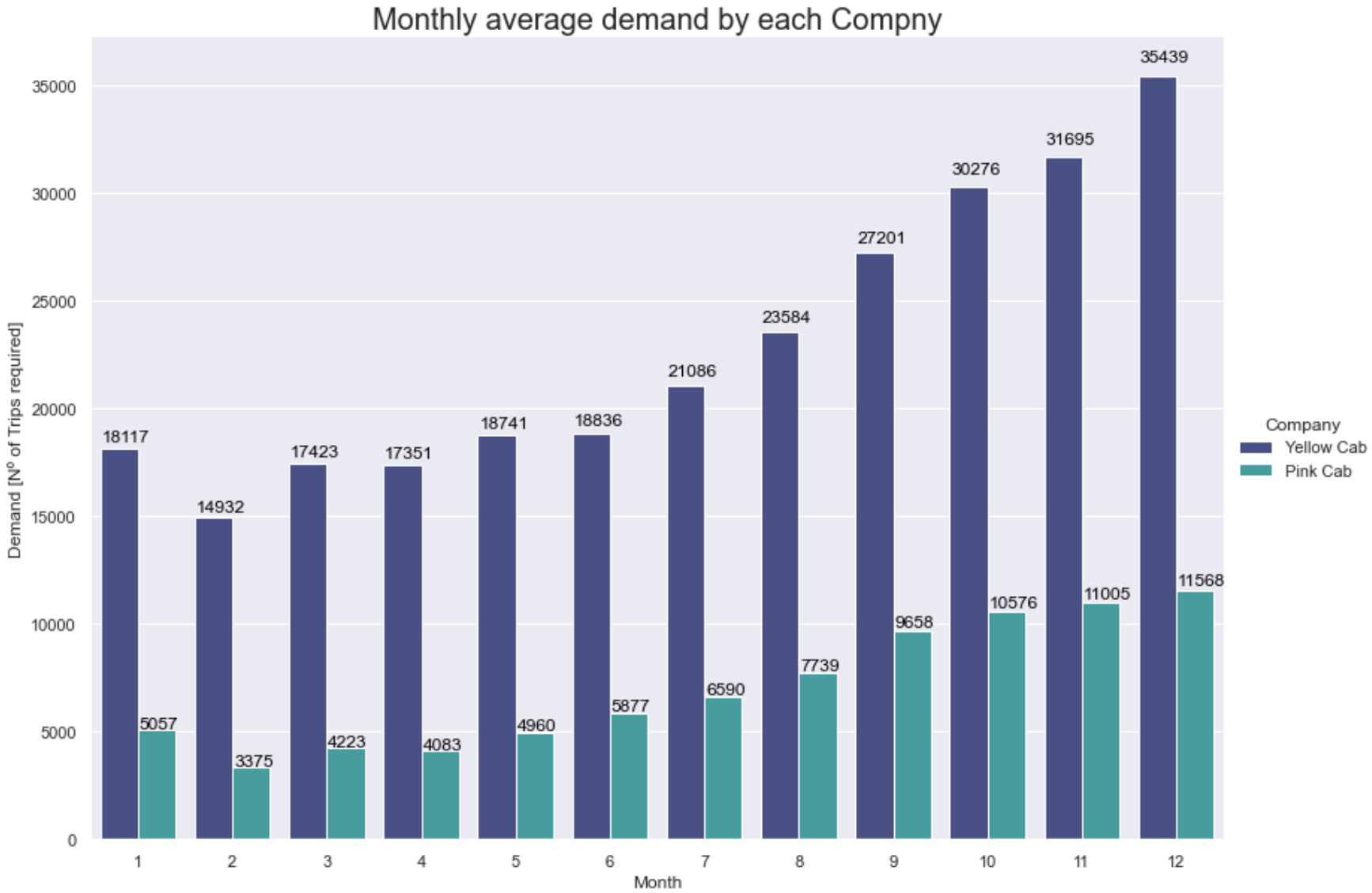


As expected, the yellow company has a little more than triple the demand of the pink company.

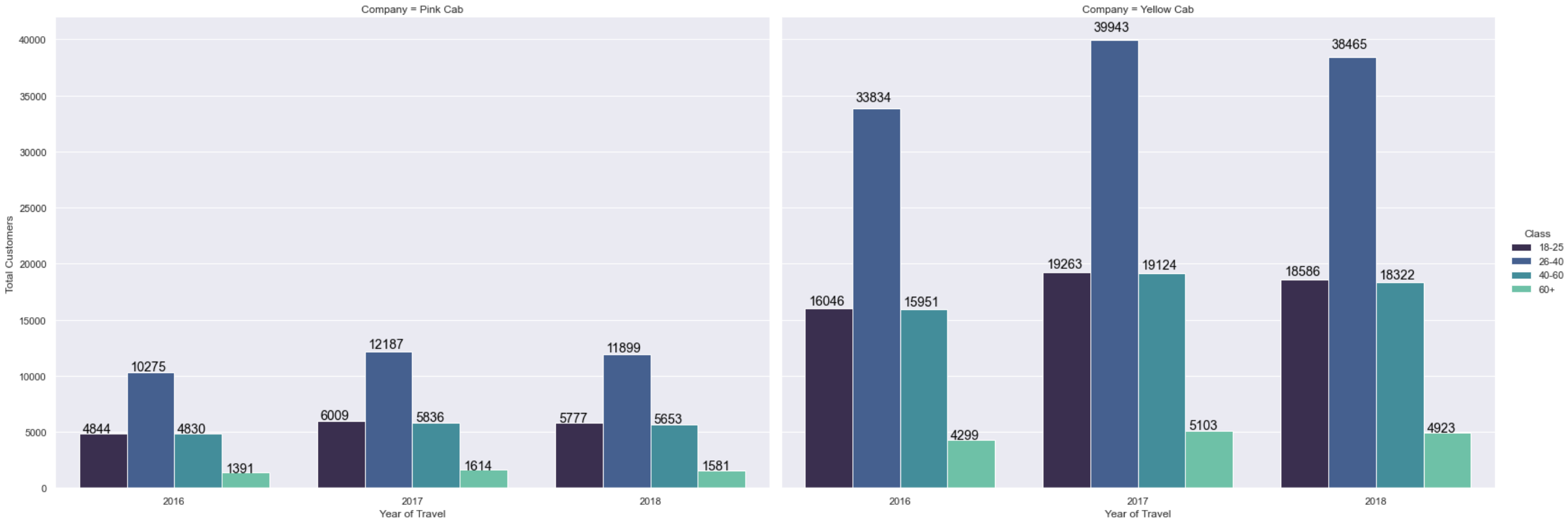
Demand Analysis: Historical Monthly Compare



Demand Analysis: Monthly Average Compare

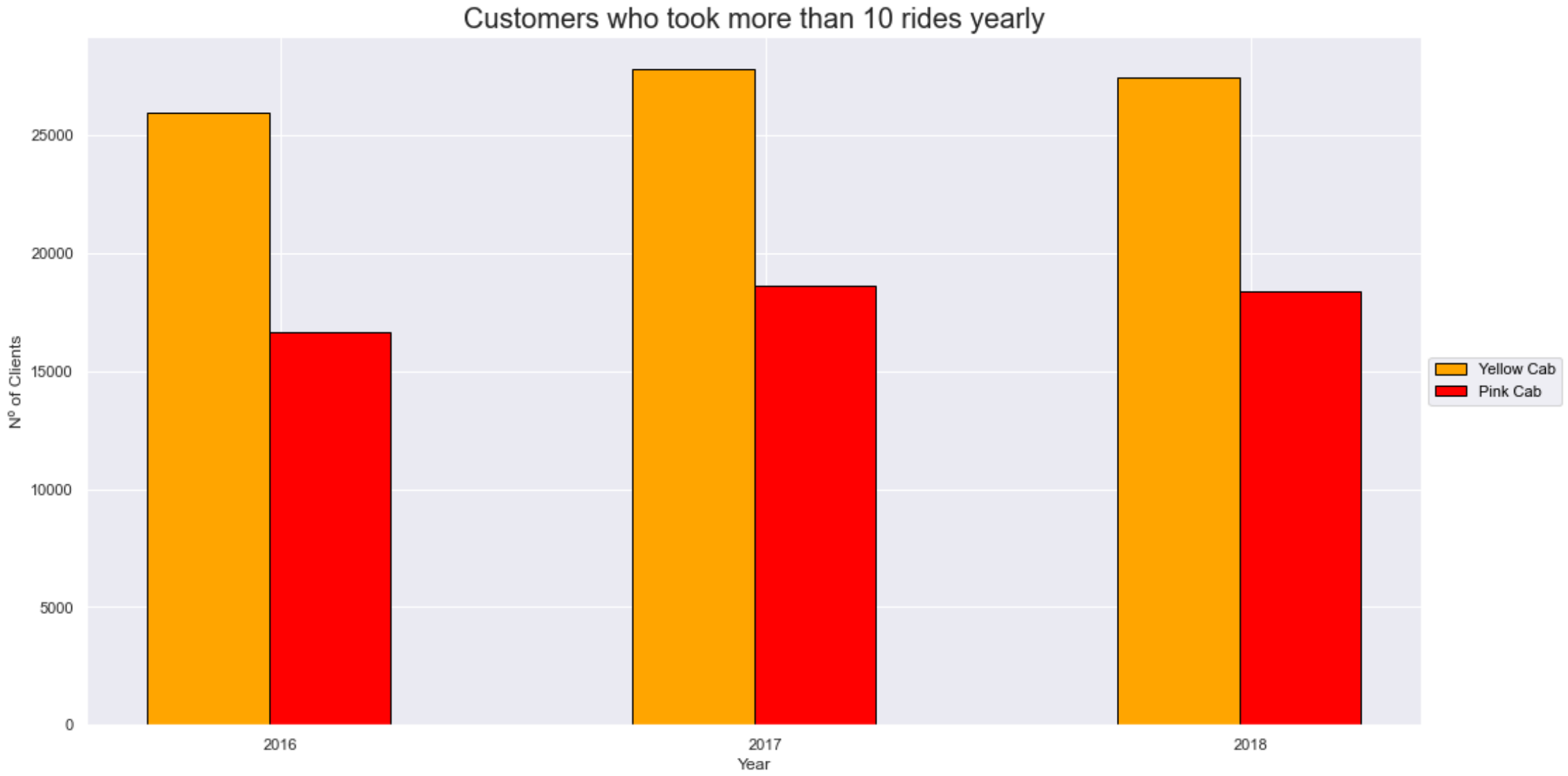


Demand Analysis: Age-wise

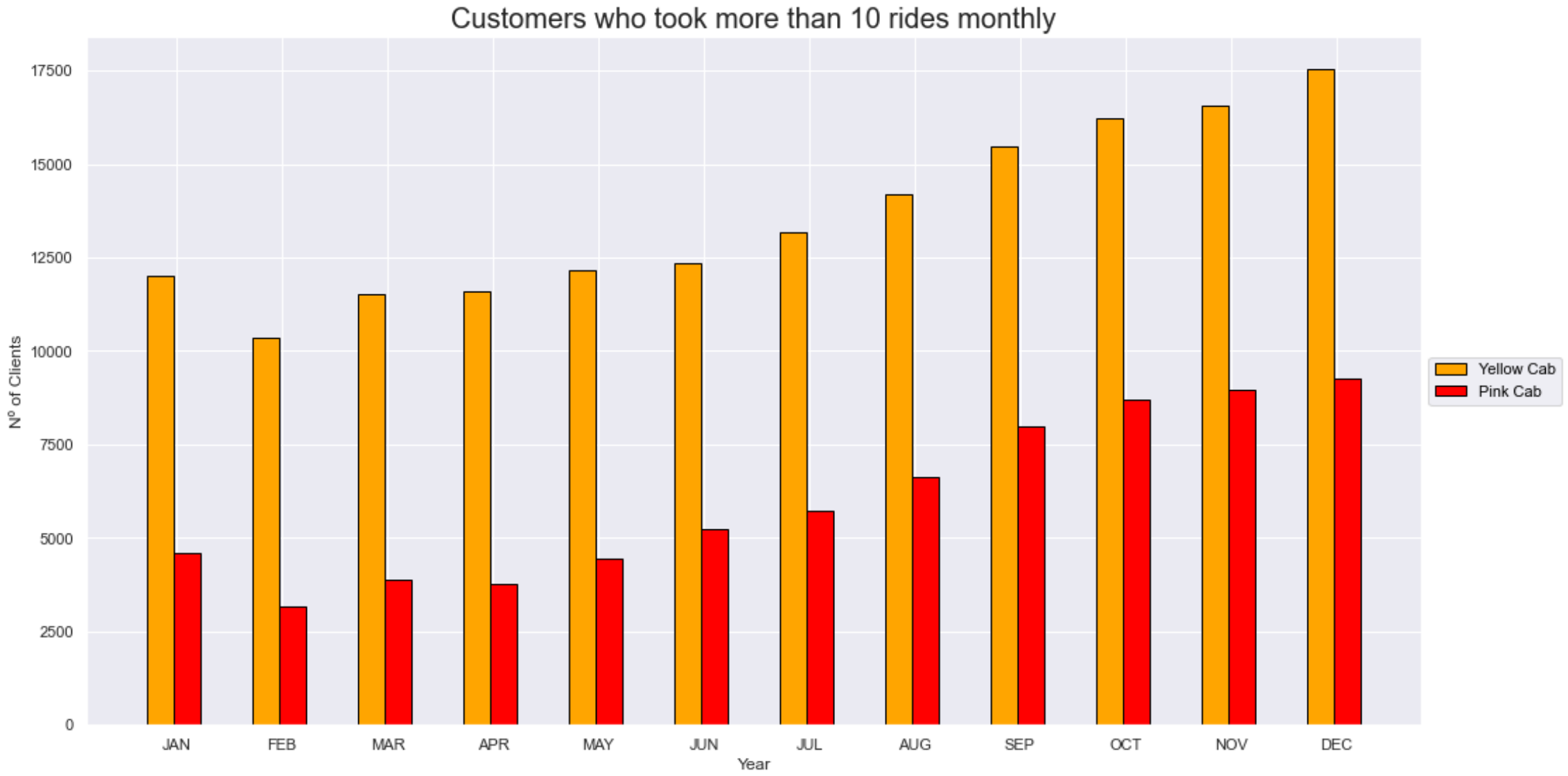


Both companies have the similar yearly demand ratios across the same age groups.

Loyalty Rates: 10 Yearly Rides Compare



Loyalty Rates: 10 Monthly Rides Compare



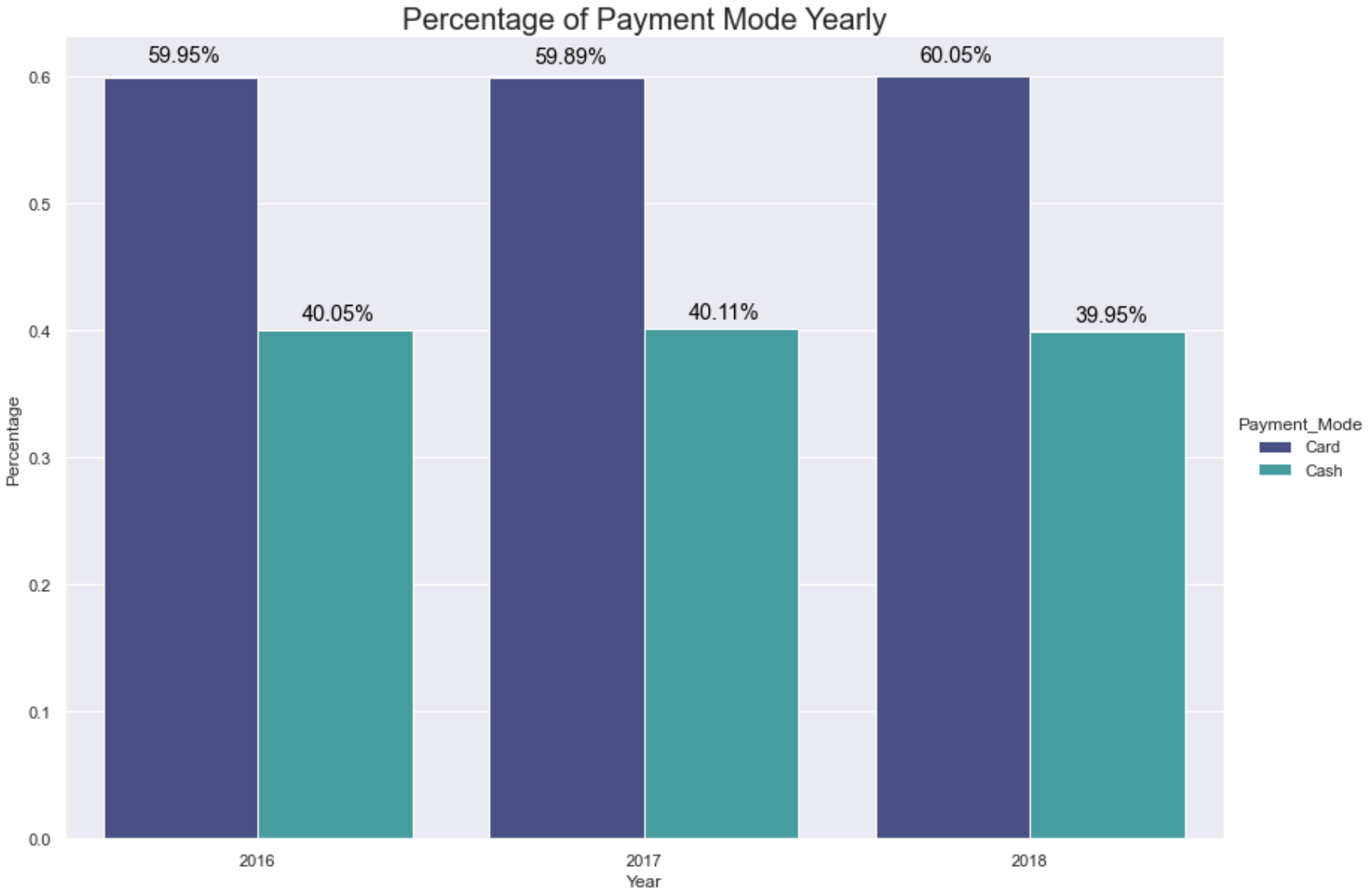
Loyalty Rates Summary

To analyze the Loyalty rates 2 classes has been defined:

1. **Medium Loyalty Customers:** Customers who took more than 10 rides yearly.
2. **High Loyalty Customers:** Customers who took more than 10 rides monthly.

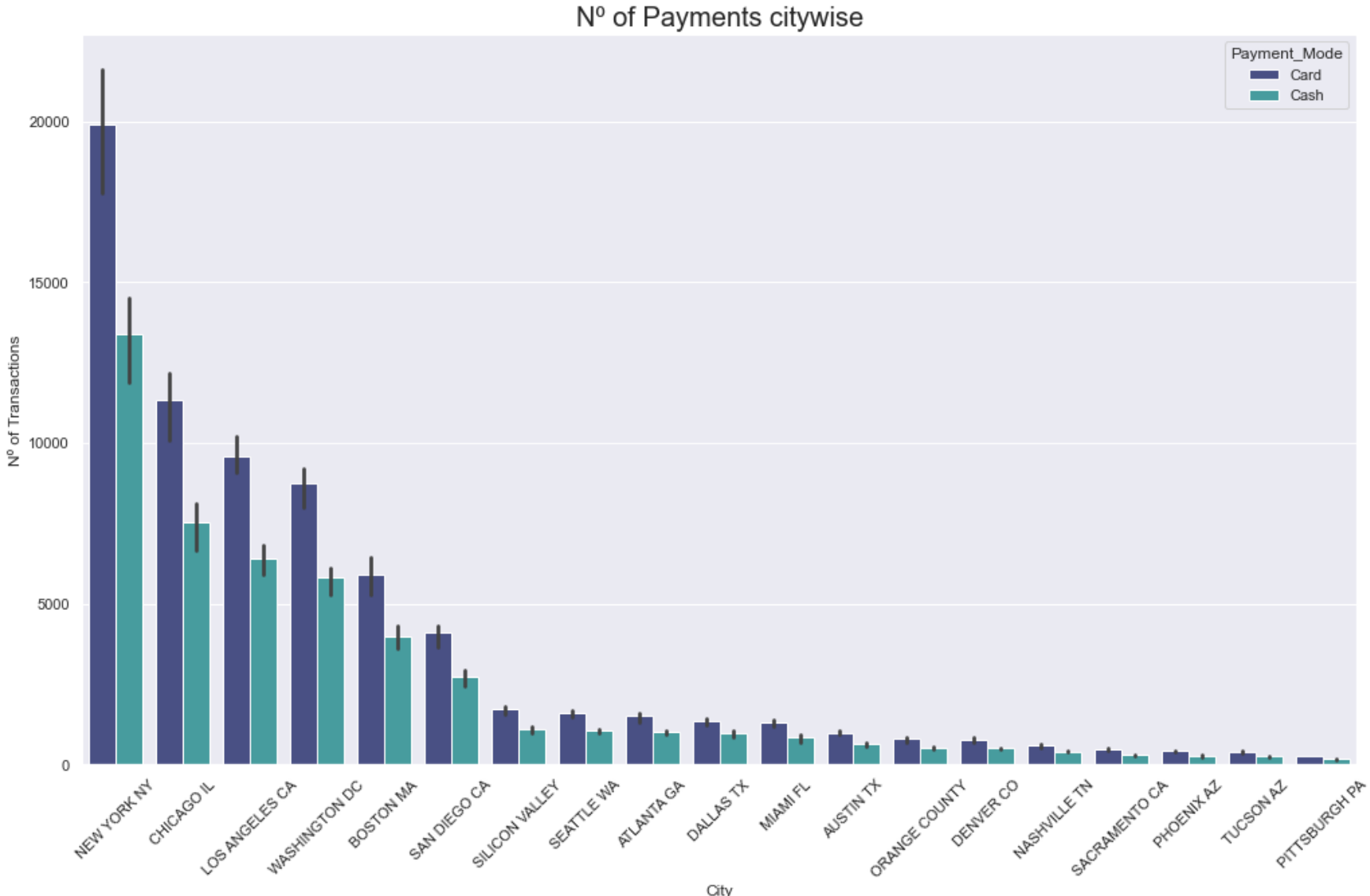
It is clearly that the yellow company is doing better in both classes of loyalty rates.

Payment Mode Distribution: Yearly Compare



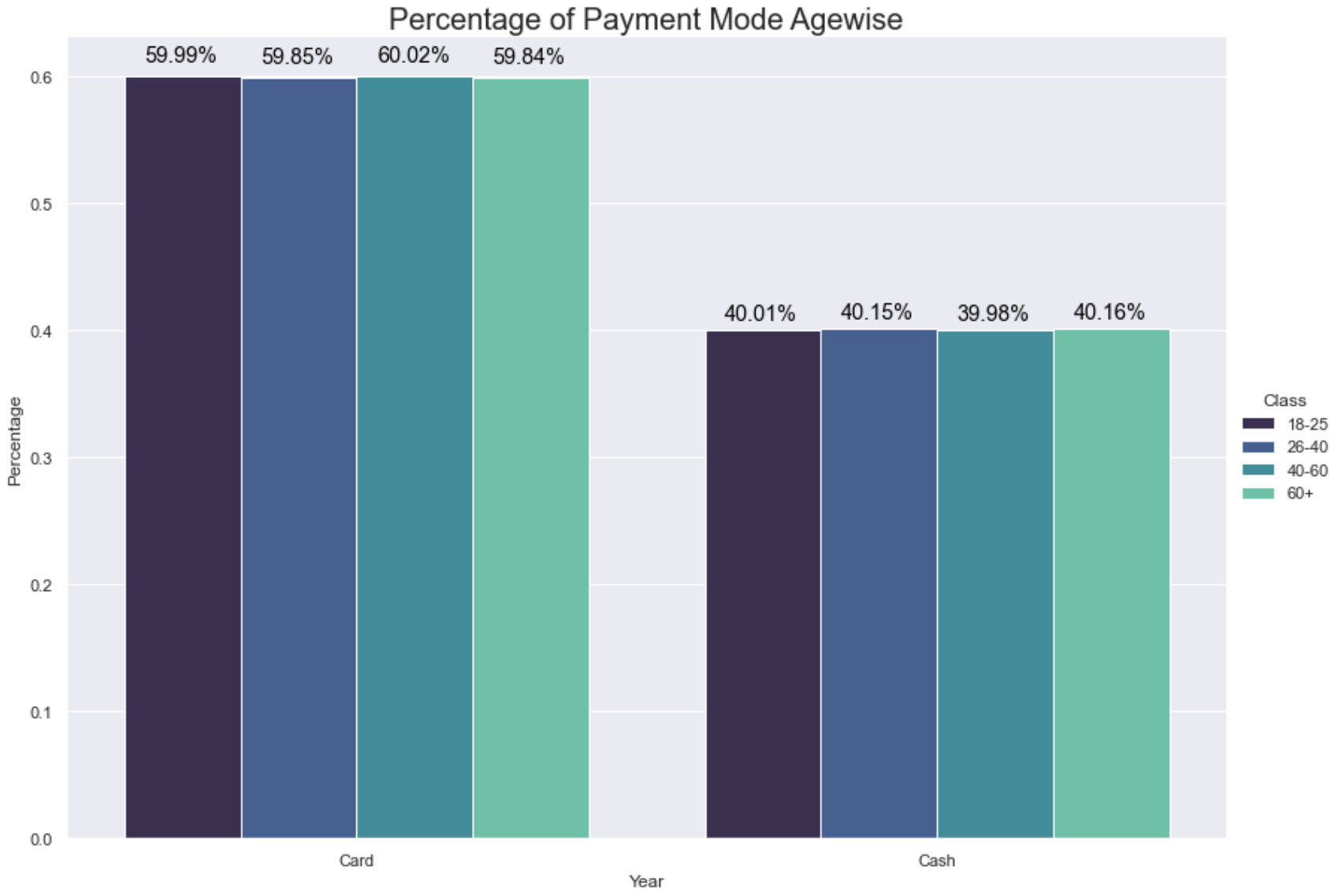
Minimal deviations in the payment method are observed over time.

Payment Mode Distribution: City-wise Compare



Minimal deviations of the payment method are observed in each city.

Payment Mode Distribution: Age-wise Compare



Minimal deviations of the payment method are observed with respect to the ages of the customers.

Conclutions

Both companies have been evaluated on following points and the Yellow company is better than the Pink company:

1. **Profit Analysis:**

- Higher profits over the time and less fluctuations monthly.
- Better profits over rides rate among time. The yellow company has 2.5 more profits over ride.
- Profits city-wise where the yellow company has greater market share in every city. In contrast, the pink company does not perform well in 8 cities.

2. **Demand Analysis:**

- The yellow company has more than triple the demand of the pink company.
- Both companies present the same distribution of Demand Age-wise.

3. **Client Analysis:**

- Assuming both classes, the yellow company has a higher High Loyalty Rate and Medium Loyalty Rate.
- Both companies present the same distribution of payment mode over time, city-wise and age-wise.

On the basis of above points , The yellow company is recommend for investment.

Thank You