Exploratory Data Analysis

Data Analysis for Cab Investment

By: Ihab Ahmed
Date: 20.06.2022

Agenda

Introduction

Data Used

Hypothesis

EDA

EDA Summary

Recommendations

Data Glacier
Your Deep Learning Partner

Agenda

Introduction

Introduction

In this project, the company "XYZ" which is a private firm that is planning for an investment in Cab Industry and is taking G2M(Go-to-Market) strategy to understand the market before taking the final decision.

There are two rising cab companies to choose from:
- Yellow Cab Company, and,
- Pink Cab Company.

In this presentation, we are going to explore the two companies using the data provided and analyse their behaviour in the span of nearly three years to make sense of which company is worth investing in.

Agenda

Data Used

Data Used.

The following datasets is used in the analysis:

- "**Cab_Data"**: Which has the details of transaction of the two cab companies, cities, date of travel, KM travelled, Cost of each trip and price charged for each trip.

- **"Customer_ID"**: Mapping table that contains a unique identifier which links the customer's demographic details.

- **"Transaction_ID"**: Mapping table that contains transaction to customer mapping and payment mode.

- **"City"**: contains list of US cities, their population and number of cab users.

- **"US Bank holidays"**: Contain all official holidays.

Agenda

Hypothesis

Data Glacier
Your Deep Learning Partner

Hypothesis

Hypothesis and Questions:

1- Which company has maximum cab users at a particular time period ?

2- Does margin proportionally increase with increase in number of customers ?

3- What cities where people use cab the most? In these cities which cab company is most demanded ?

4- In public holidays which company is the most demanded ?

Business problem:

- Go to market strategy (G2M), investigate two cab companies activities before investment.

Data Glacier

Your Deep Learning Partner

Agenda

EDA

Exploratory Data Analysis: Cab Rides

Number of Pink Cab rides vs Yellow Cab rides (in the span of three years):

- Total Yellow Cab rides (approx): 280000.

- Total Pink Cab rides (approx): 85000

- As notices yellow cab rides is more than thrice the pink cab rides
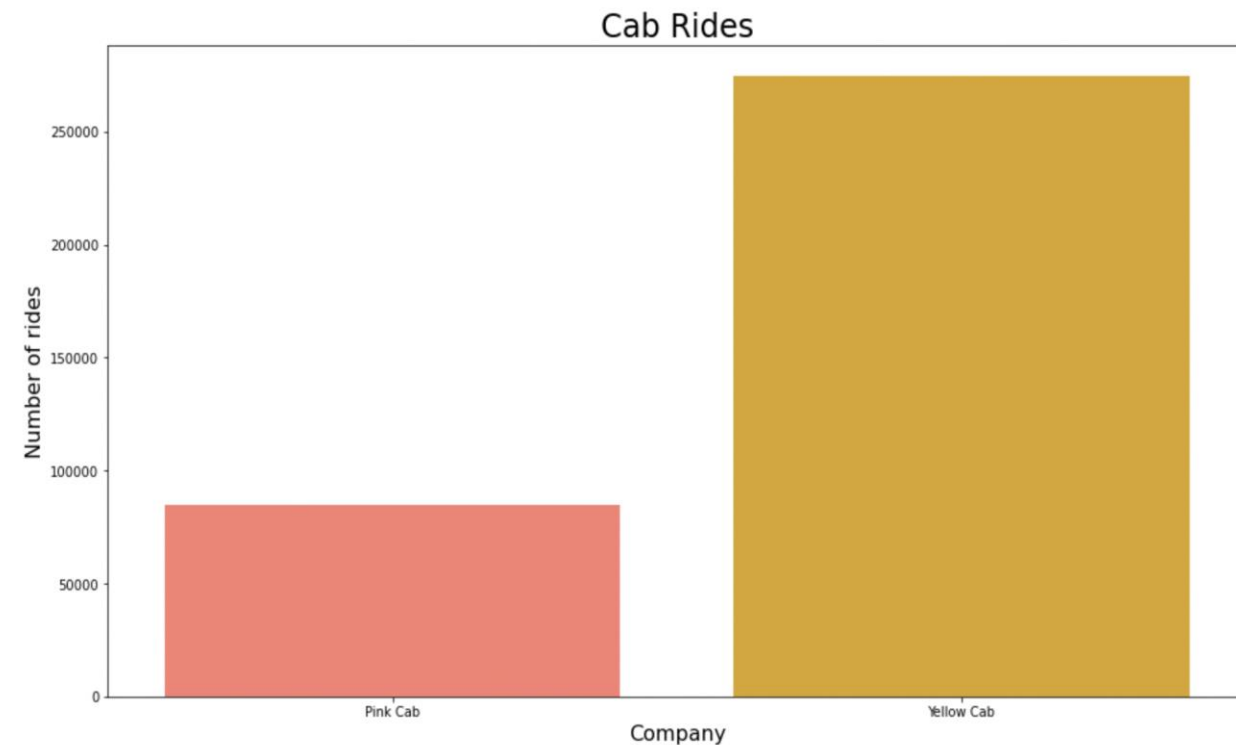
Exploratory Data Analysis: Revenue

Revenue of Pink Cab vs Yellow Cab (in the span of three years):

- Yellow Cab total revenue (approx): 125M $.

- Pink Cab total revenue (approx): 26M $.

- Yellow Cab has nearly 5x the revenue of Pink Cab.

Data Glacier
Your Deep Learning Partner
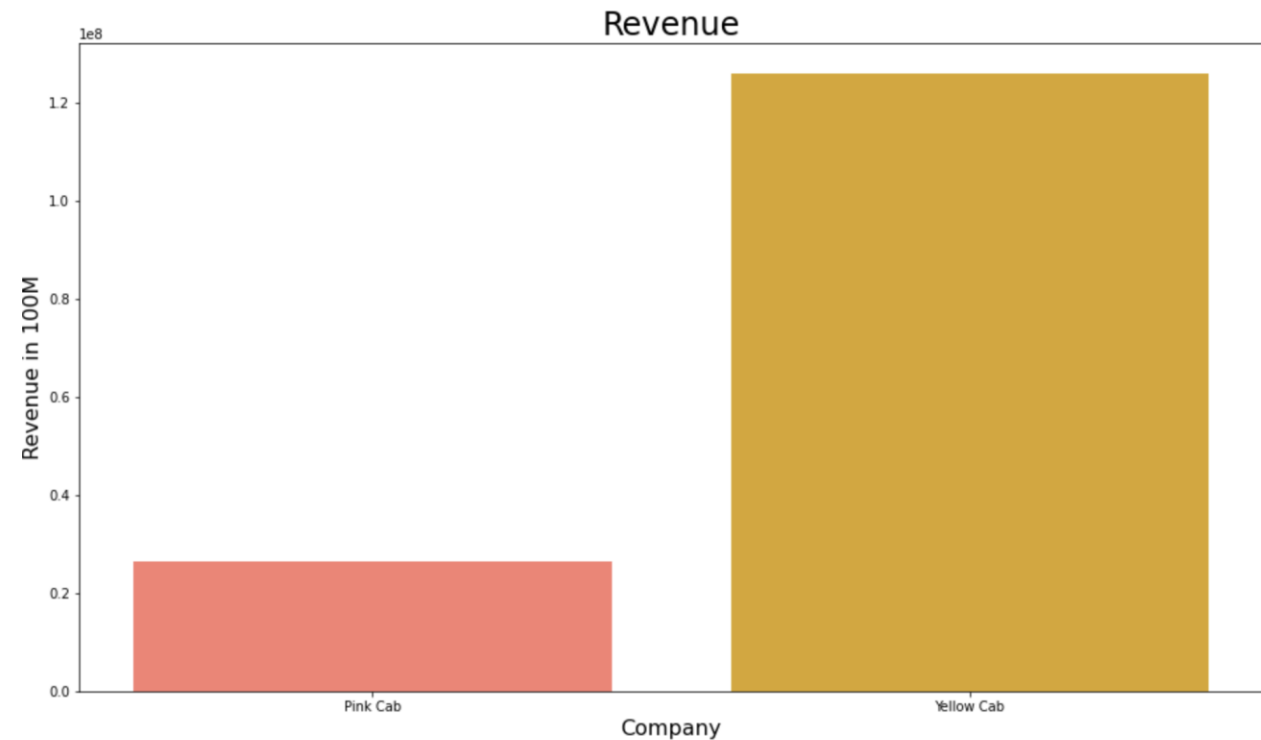
Exploratory Data Analysis: Income

Income of Pink Cab vs Yellow Cab (in the span of three years):

- Yellow Cab total Income (approx): 44M $.

- Pink Cab total Income (approx): 5.3M $.
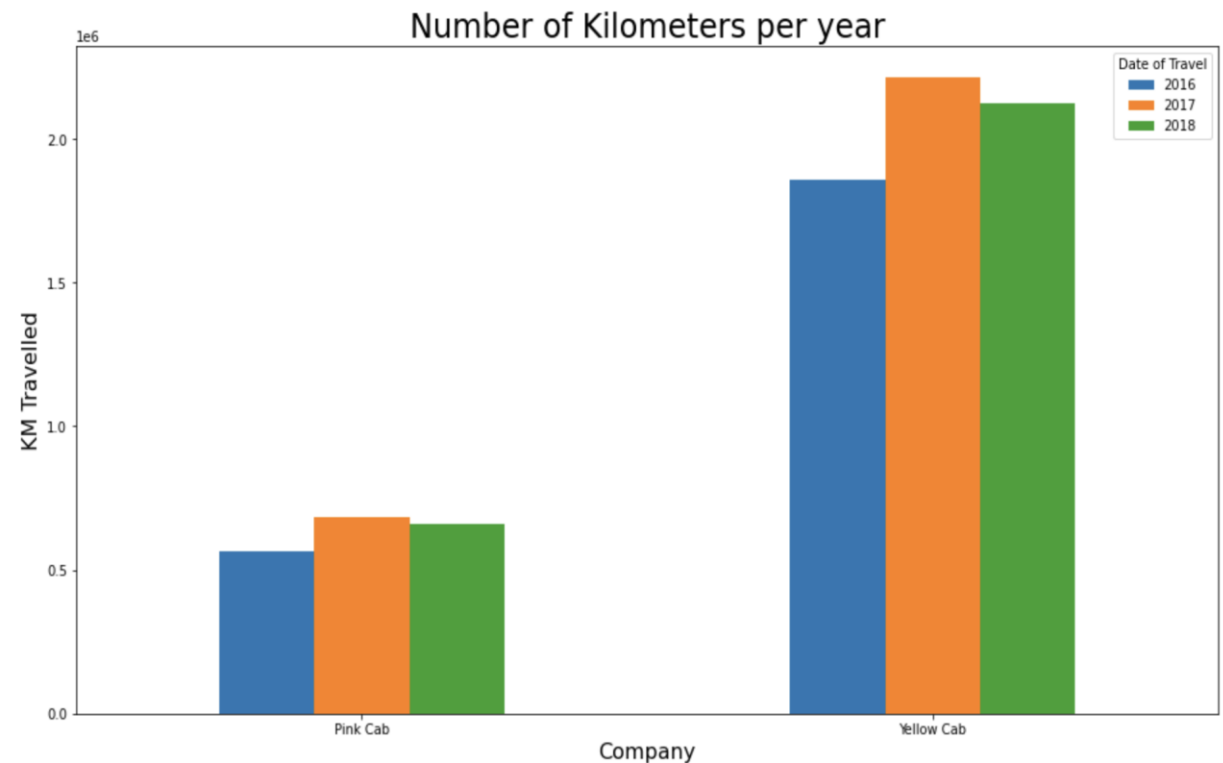
- Yellow Cab's income is more than 8x Pink Cab's Income.
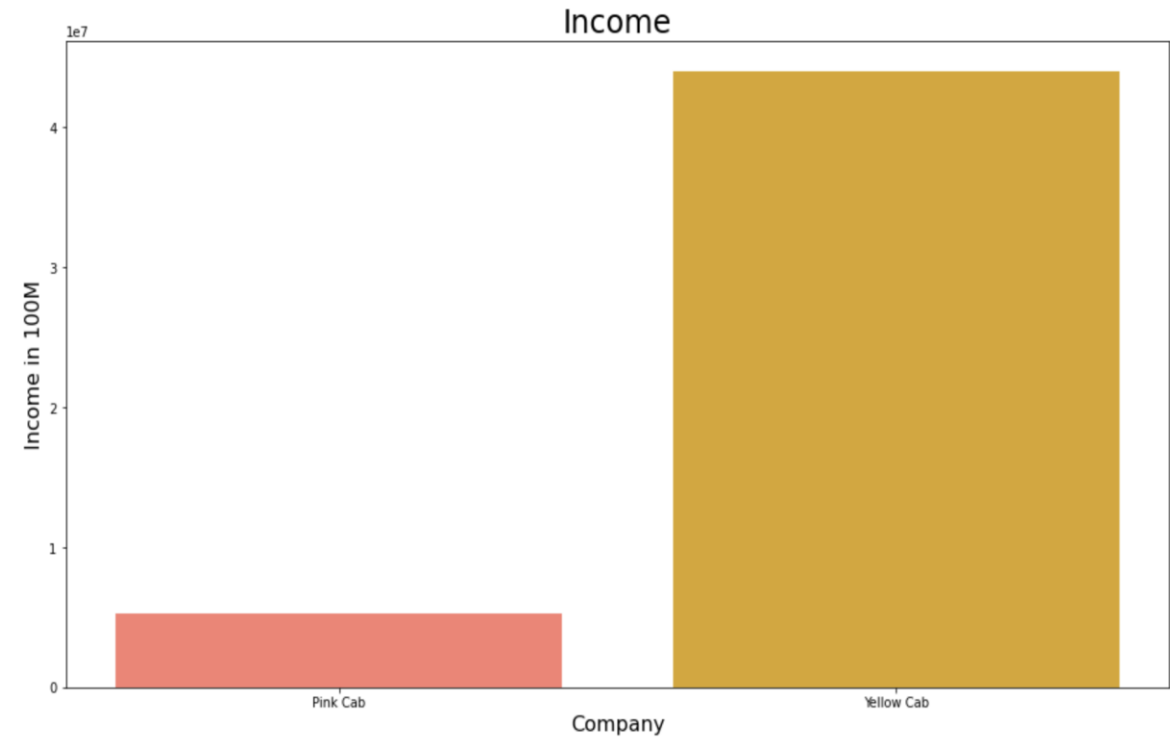
Exploratory Data Analysis: Yearwise

Total activity of each cab company in each year (2016, 2017, 2018)

<u>Kilometers Travelled:</u>

- Yellow Cab (2016): 1.8M Km.
- Pink Cab (2016): 560K km.

- Yellow Cab (2017): 2.2M Km.
- Pink Cab (2017): 680K Km.

- Yellow Cab (2018): 2.1M Km.
- Pink Cab (2018): 661K Km.

In every year Yellow Cab travel 3x more than Pink Cabs

Number of Kilometers per year

Data Glacier
Your Deep Learning Partner

Exploratory Data Analysis: Yearwise

Total Revenue of each cab company in each year (2016, 2017, 2018)

Kilometers Travelled:

- Yellow Cab (2016): 38M $.
- Pink Cab (2016): 7.9M $.

- Yellow Cab (2017): 45M $.
- Pink Cab (2017): 9.5M $.

- Yellow Cab (2018): 41M $.
- Pink Cab (2018): 8.8M $.

In every year Yellow Cab has approx. 5x more revenue than Pink Cab

Revenue per year

Data Glacier
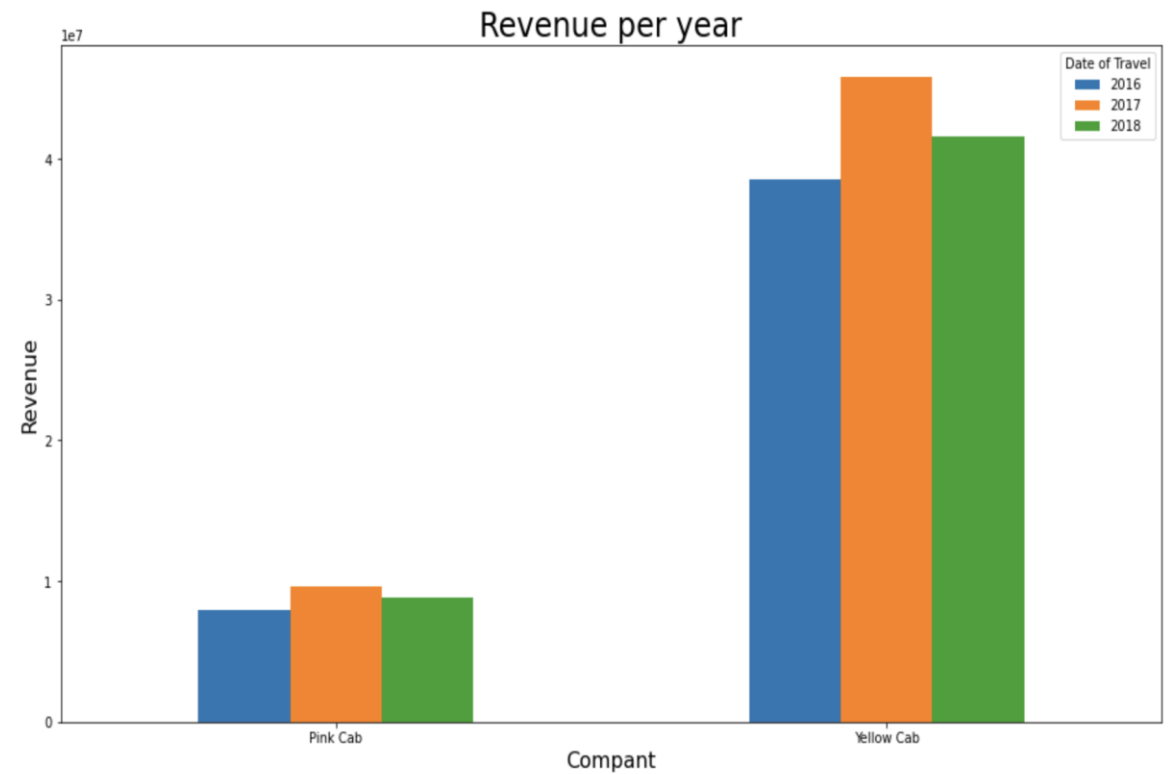
Exploratory Data Analysis: Yearwise

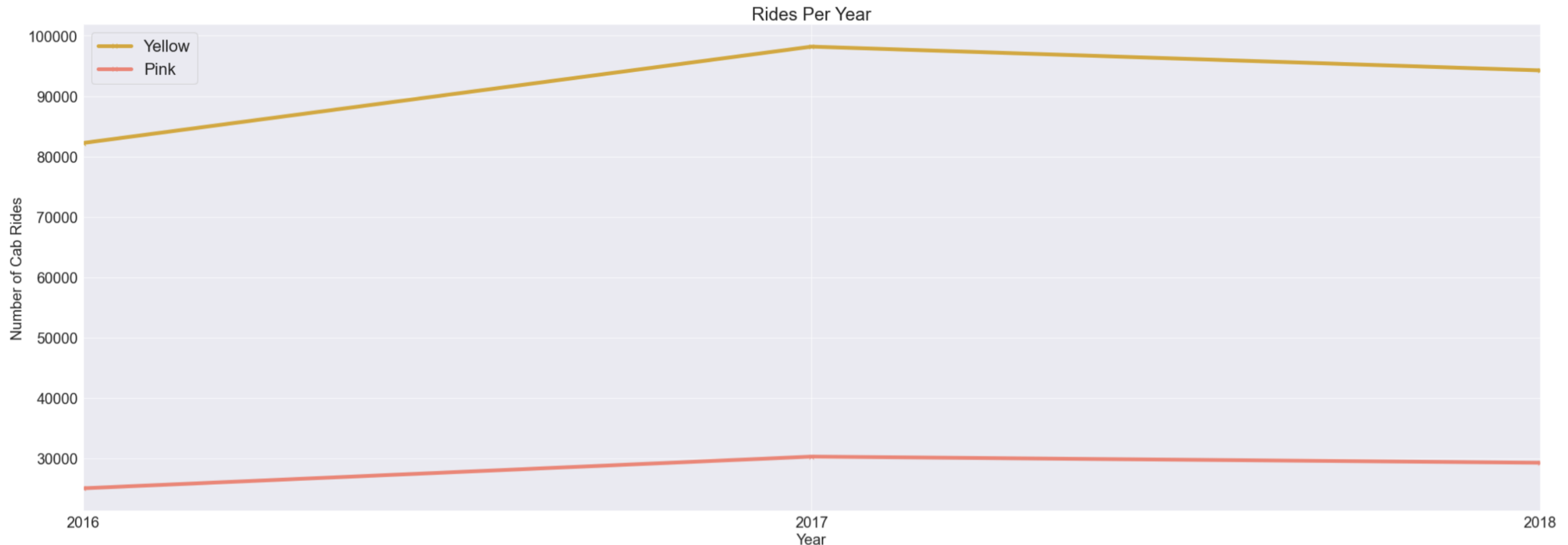Total Income of each cab company in each year (2016, 2017, 2018)

Kilometers Travelled:

- Yellow Cab (2016): 13M $.
- Pink Cab (2016): 1.7M $.

- Yellow Cab (2017): 16M $.
- Pink Cab (2017): 2M $.

- Yellow Cab (2018): 13M $.
- Pink Cab (2018): 1.5M $.

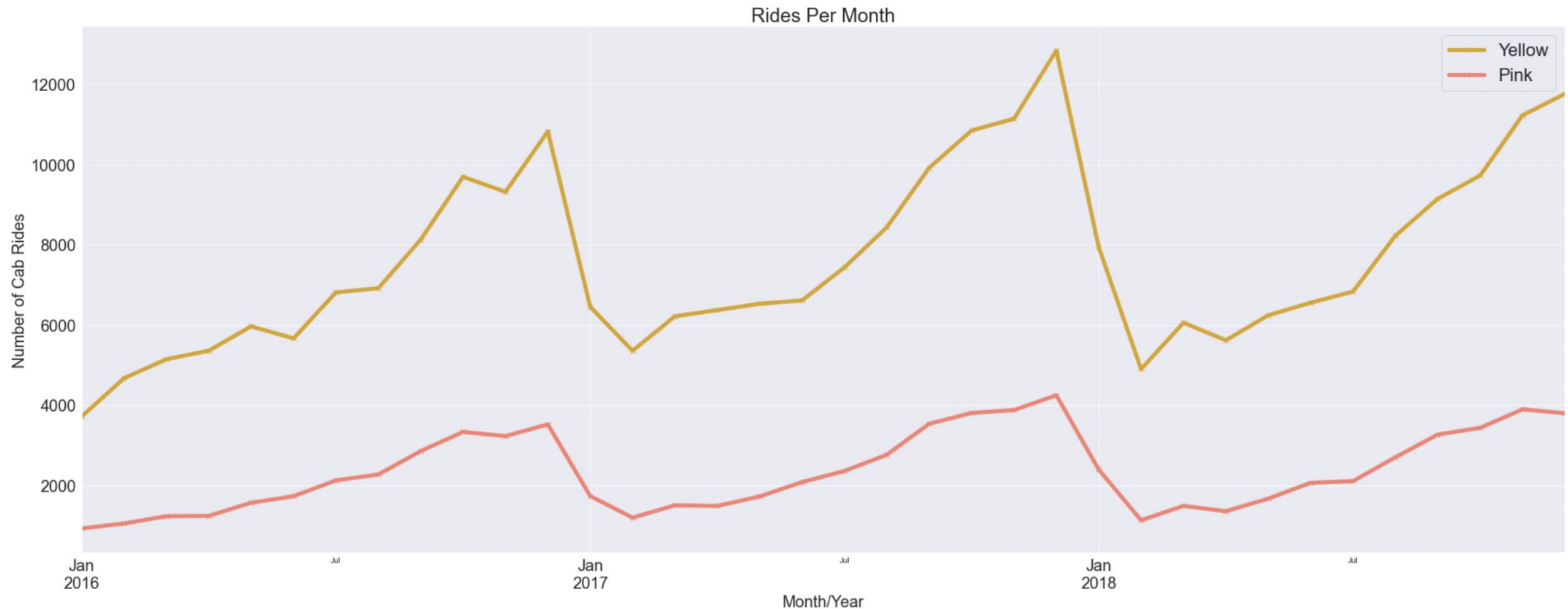In every year Yellow Cab has approx. 8x more profit than Pink Cab.
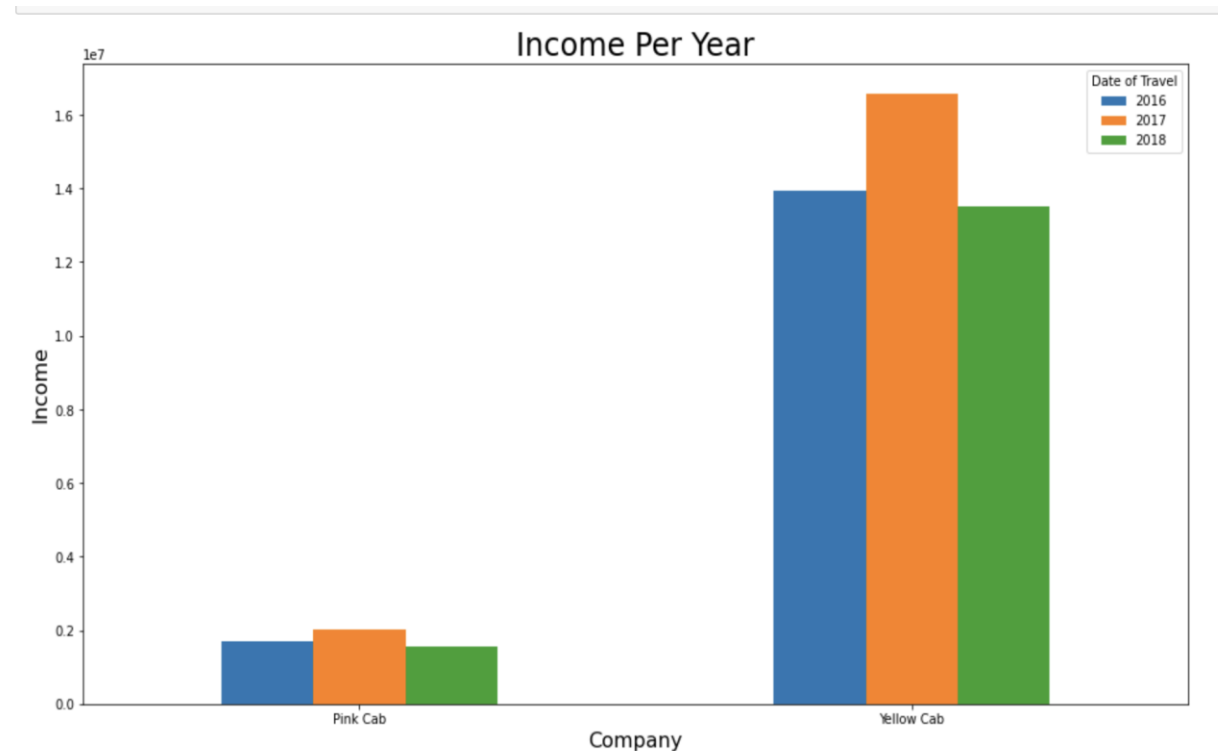
Exploratory Data Analysis: Yearwise

Q1- Which company has maximum cab users at a particular time period ?

Exploratory Data Analysis: Monthwise

Q1- Which company has maximum cab users at a particular time period ?

Hypothesis 1 answer

Q1- Which company has maximum cab users at a particular time period ?

Ans1- Yellow Cab company has much more cab users Yearly and monthly from 2016 to 2019

Data Glacier

Exploratory Data Analysis: Yearwise

Q2- Does margin increase with increase in number of customers ?

Exploratory Data Analysis: Monthwise

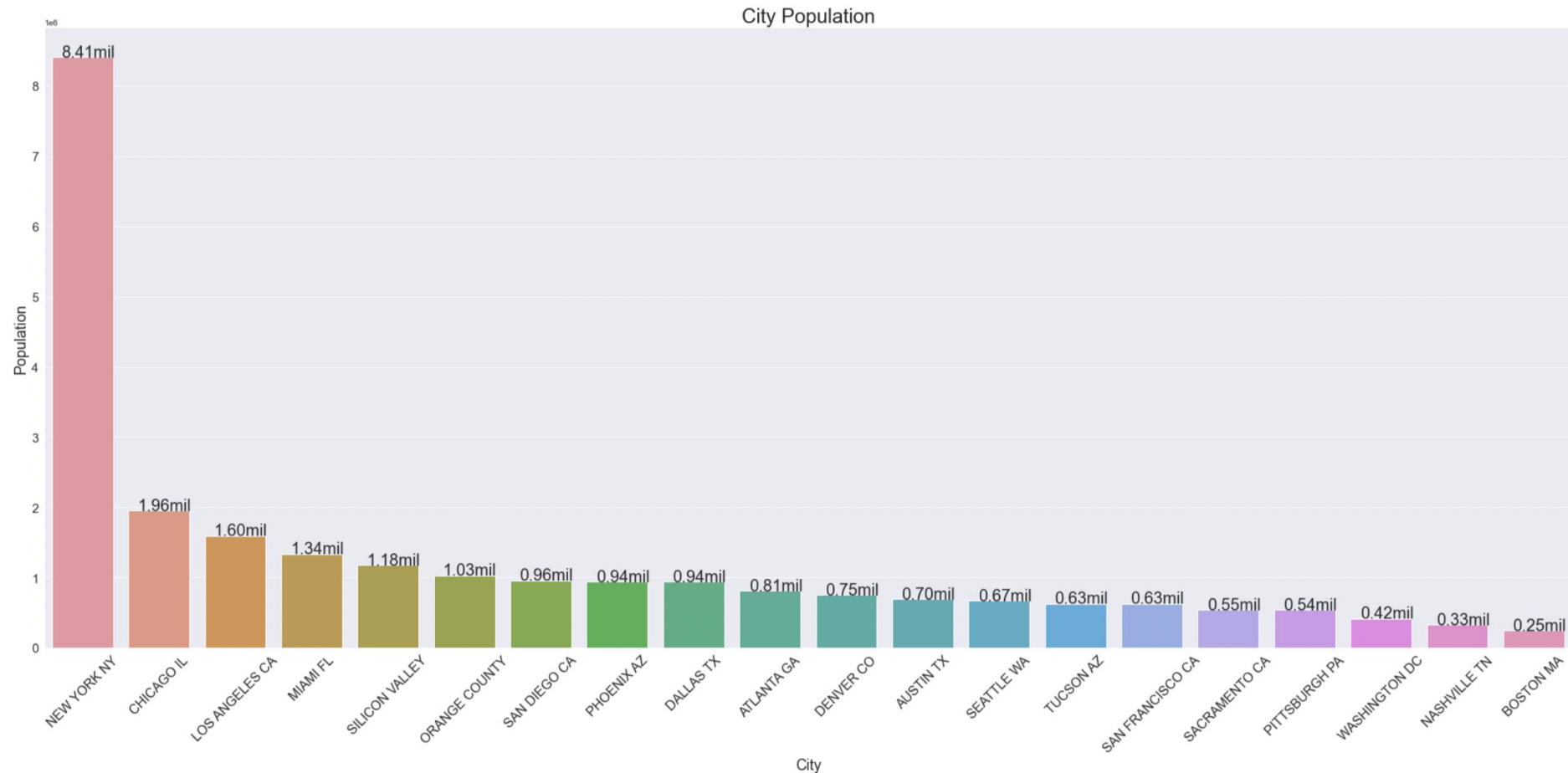Q2- Does margin increase with increase in number of customers ?

Data Glacier

Hypothesis 2 answer

Q2- Does margin proportionally increase with increase in number of customers ?

Ans2- Yes, Yellow cab company has much more income than pink cab company, since it has more customers
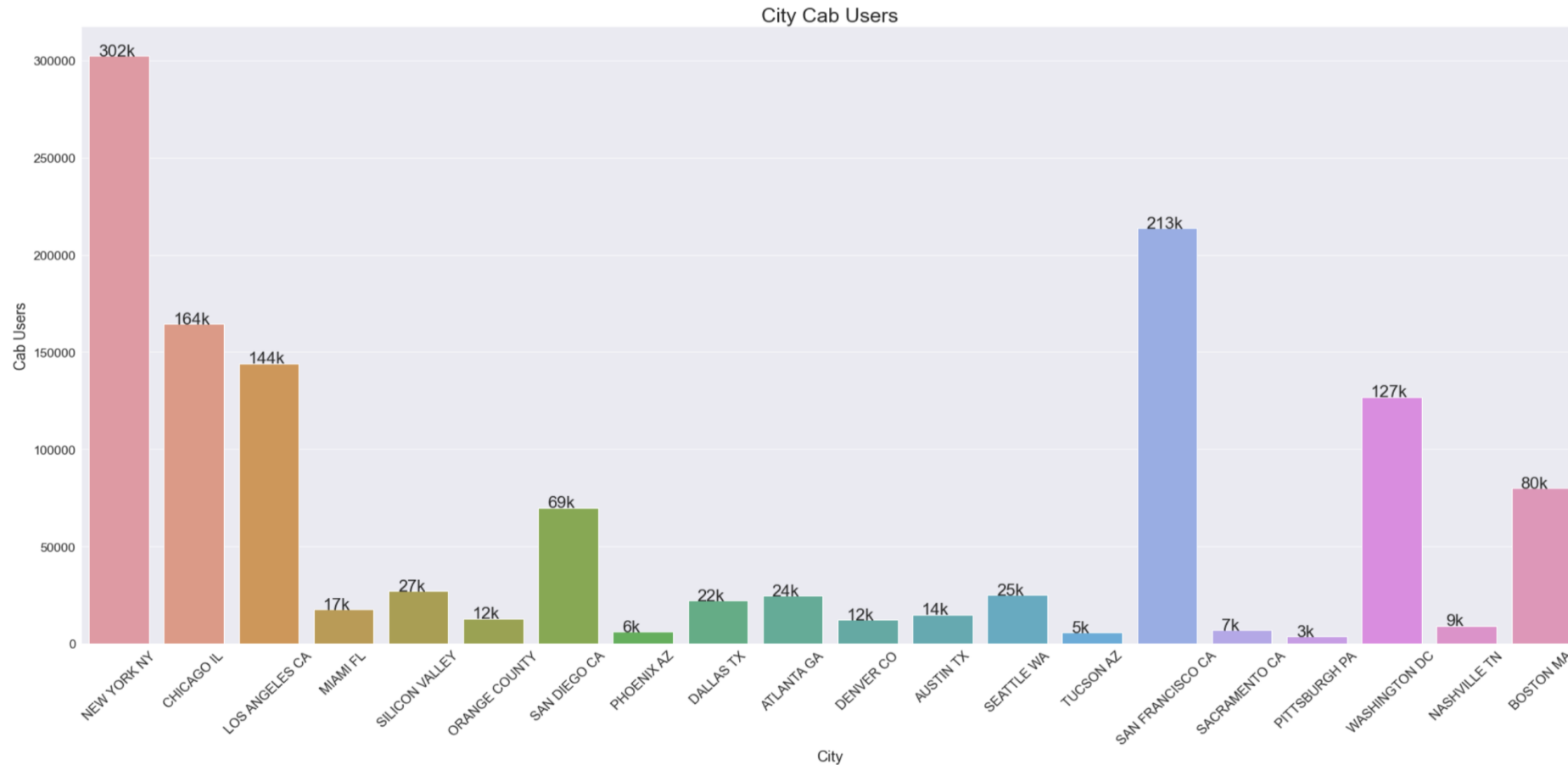
EDA: Cities, Cab Rides and Income
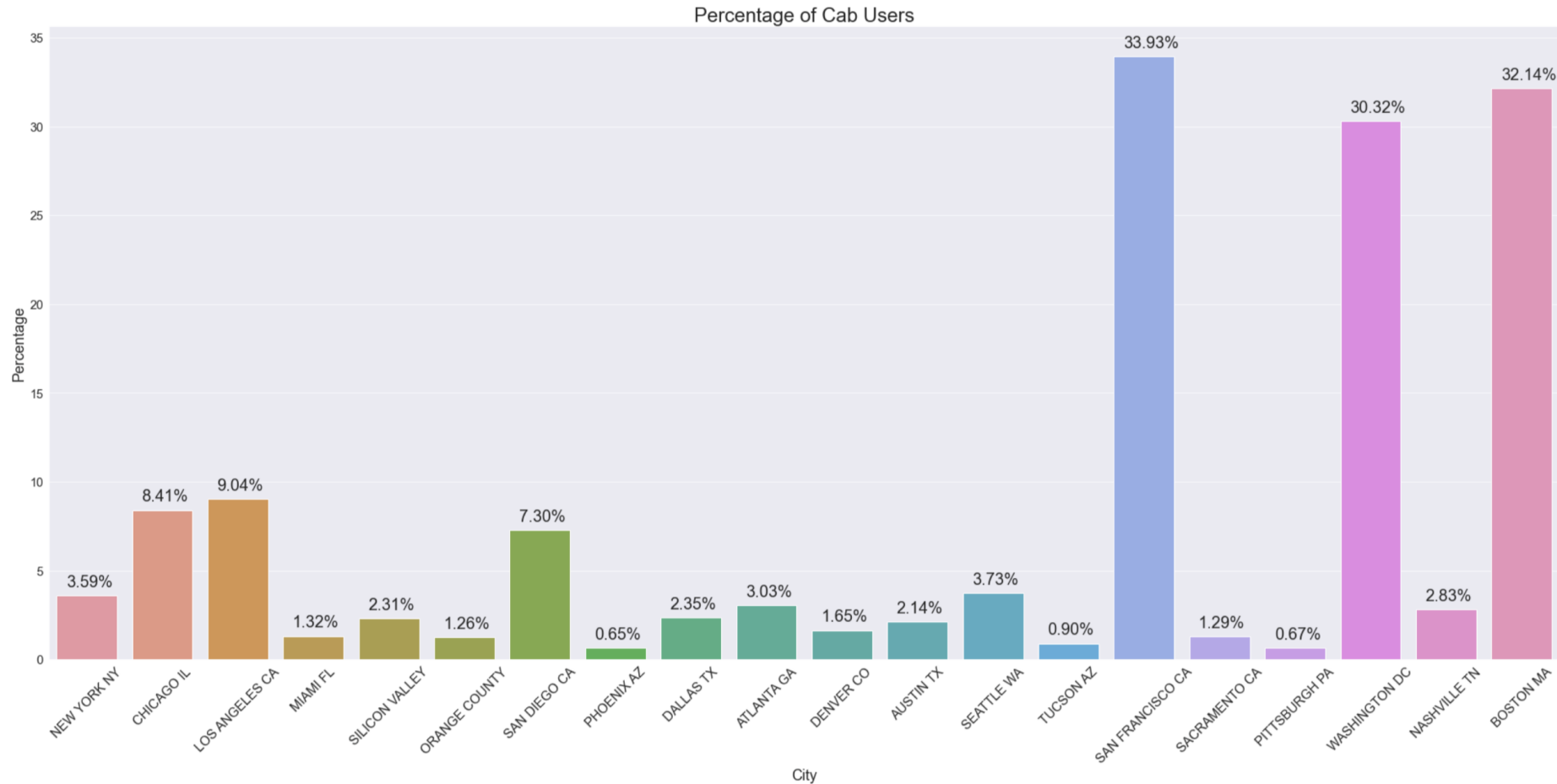
Population Distribution:

City Population

EDA: Cities, Cab Rides and Income

Cab User Distribution:
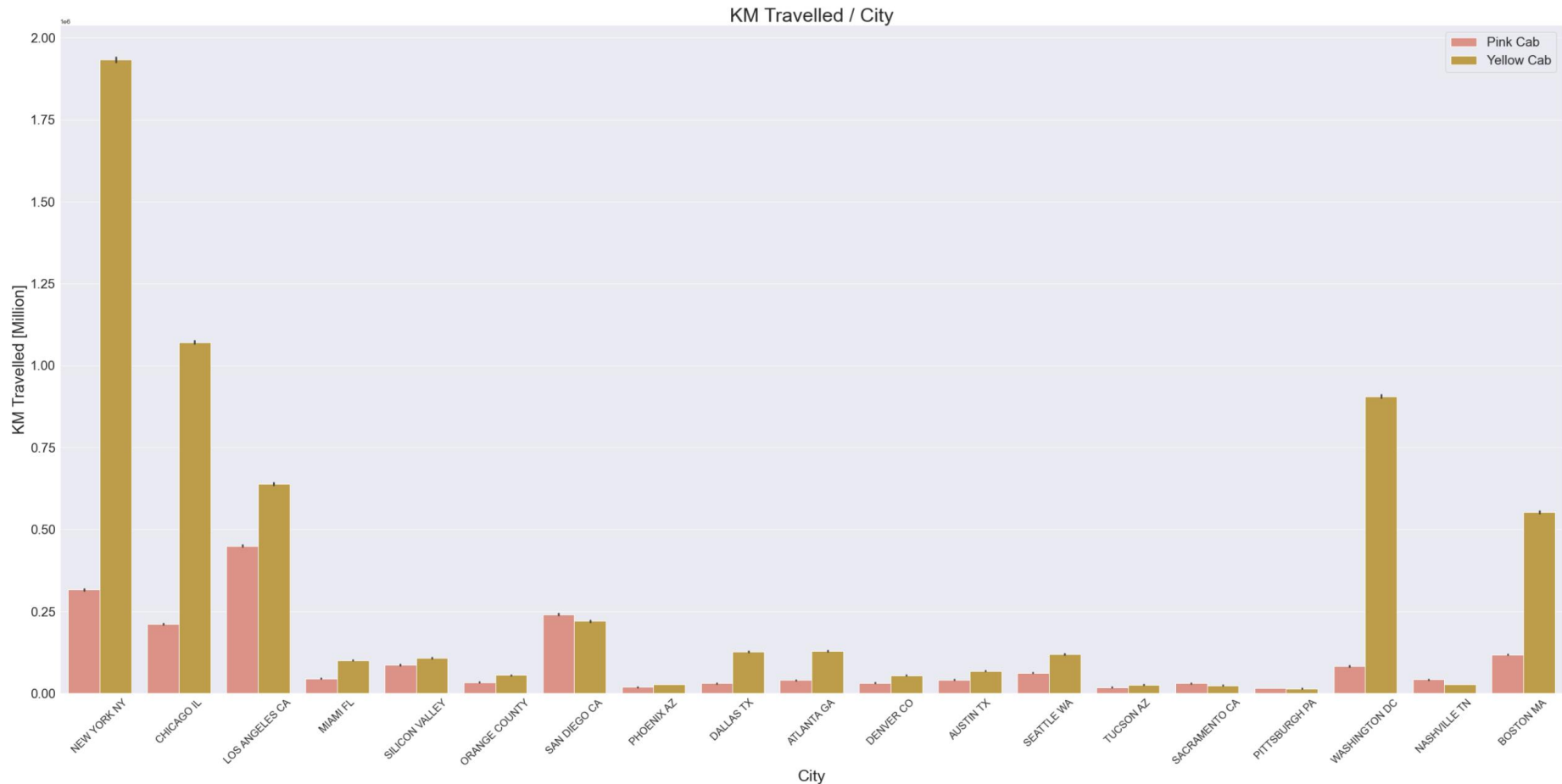
EDA: Cities, Cab Rides and Income

Percentage Distribution:

EDA: Cities, Cab Rides and Income

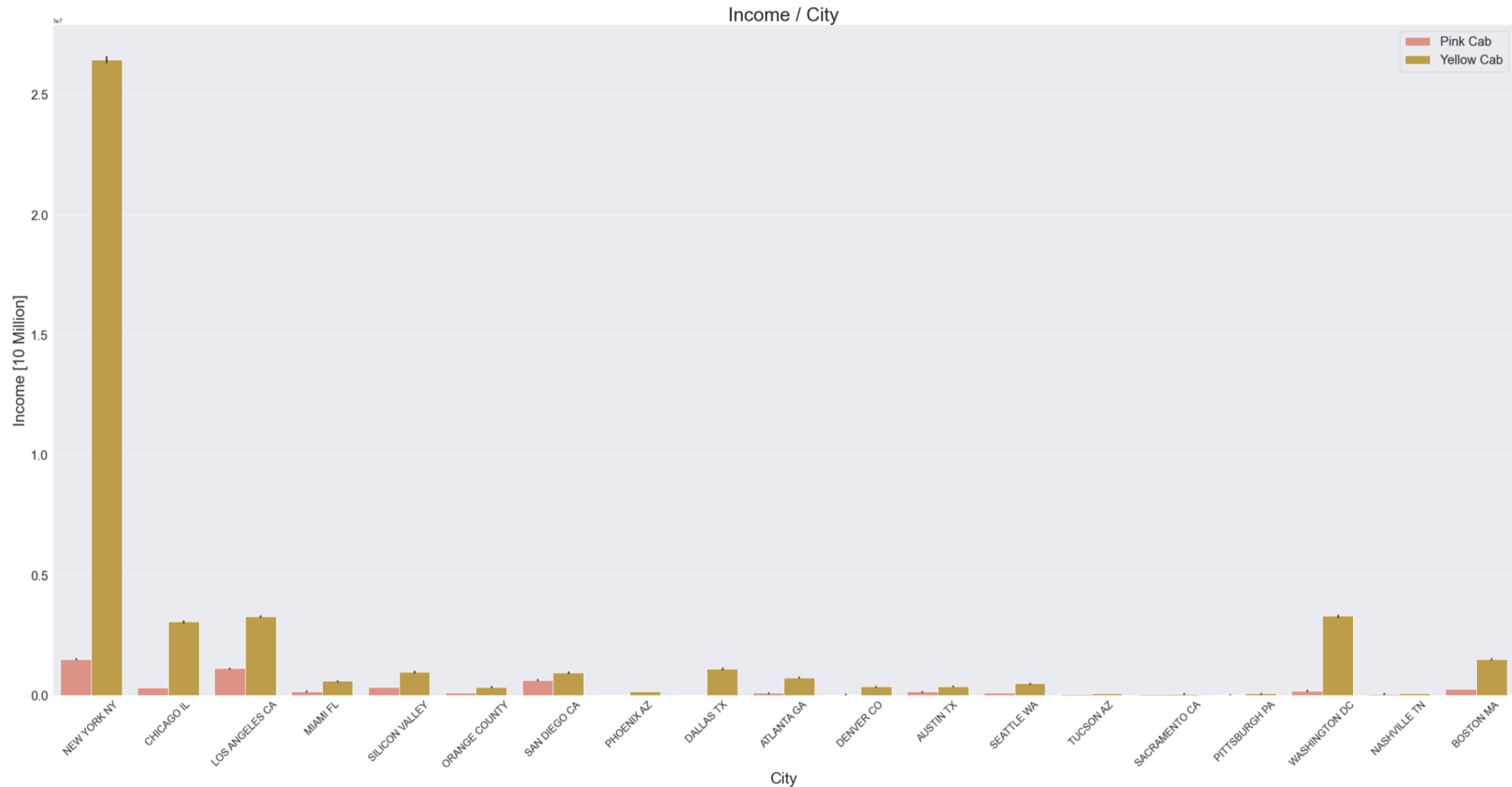Profit and Activity of each cab company in each city:

- Yellow/Pink Cabs KM Travelled per city:

EDA: Cities, Cab Rides and Income

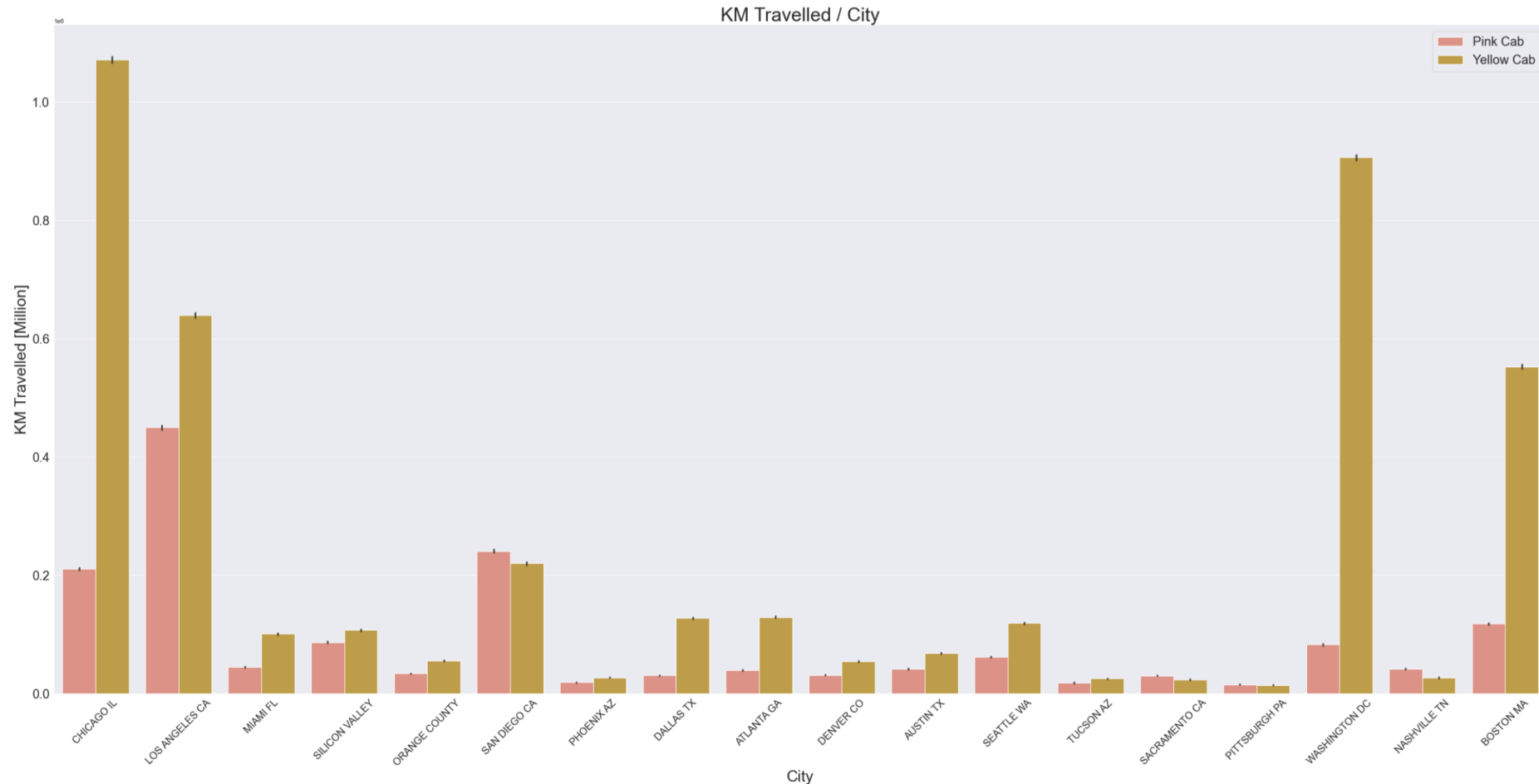Profit and Activity of each cab company in each city:

- Yellow/Pink Cabs Income per city:

EDA: Cities, Cab Rides and Income

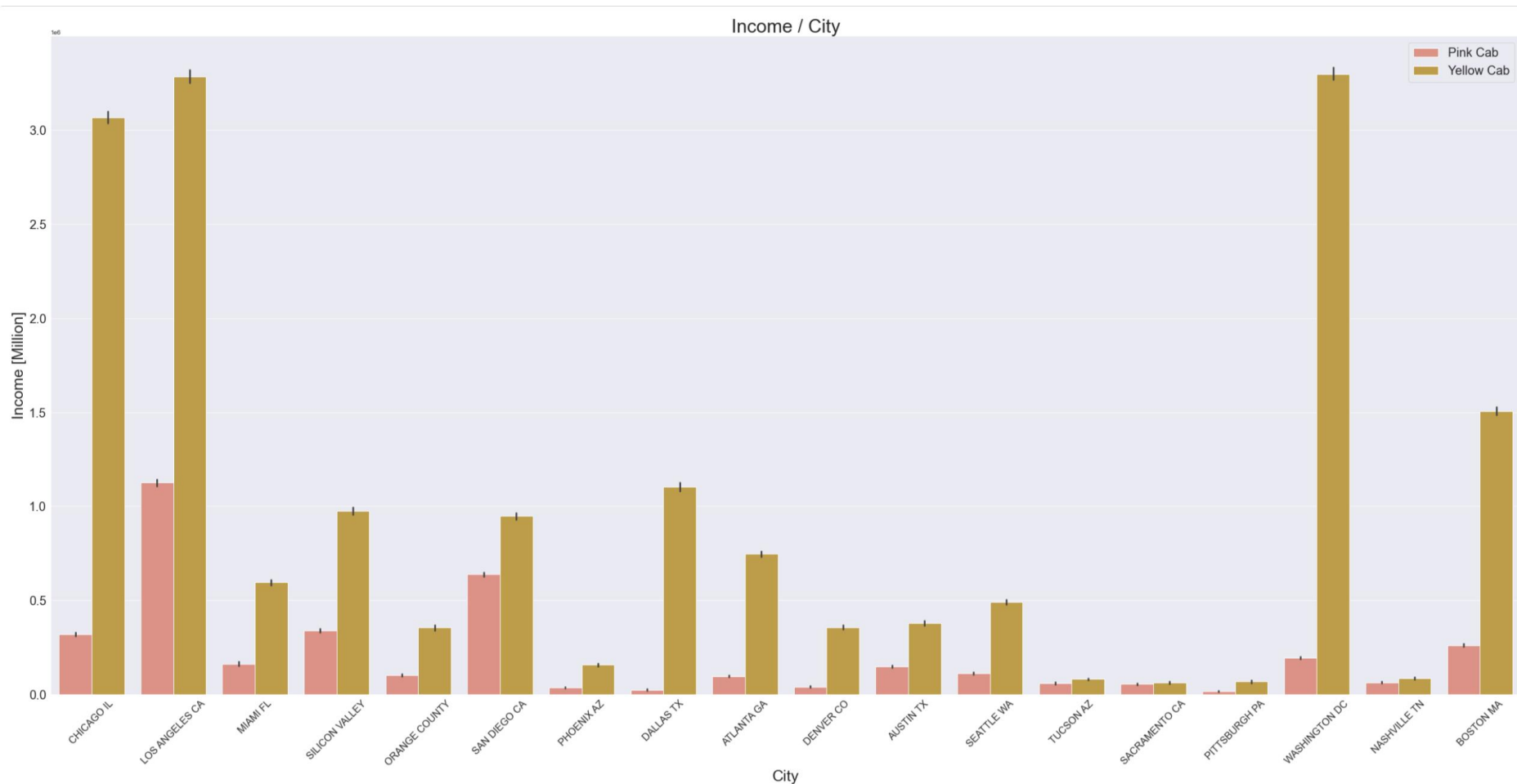Profit and Activity of each cab company in each city:

- **Yellow/Pink Cabs KM Travelled per city (without New York to make sense of other cities):**

EDA: Cities, Cab Rides and Income

Profit and Activity of each cab company in each city:

- **Yellow/Pink Cabs Income (without New York to make sense of other cities):**
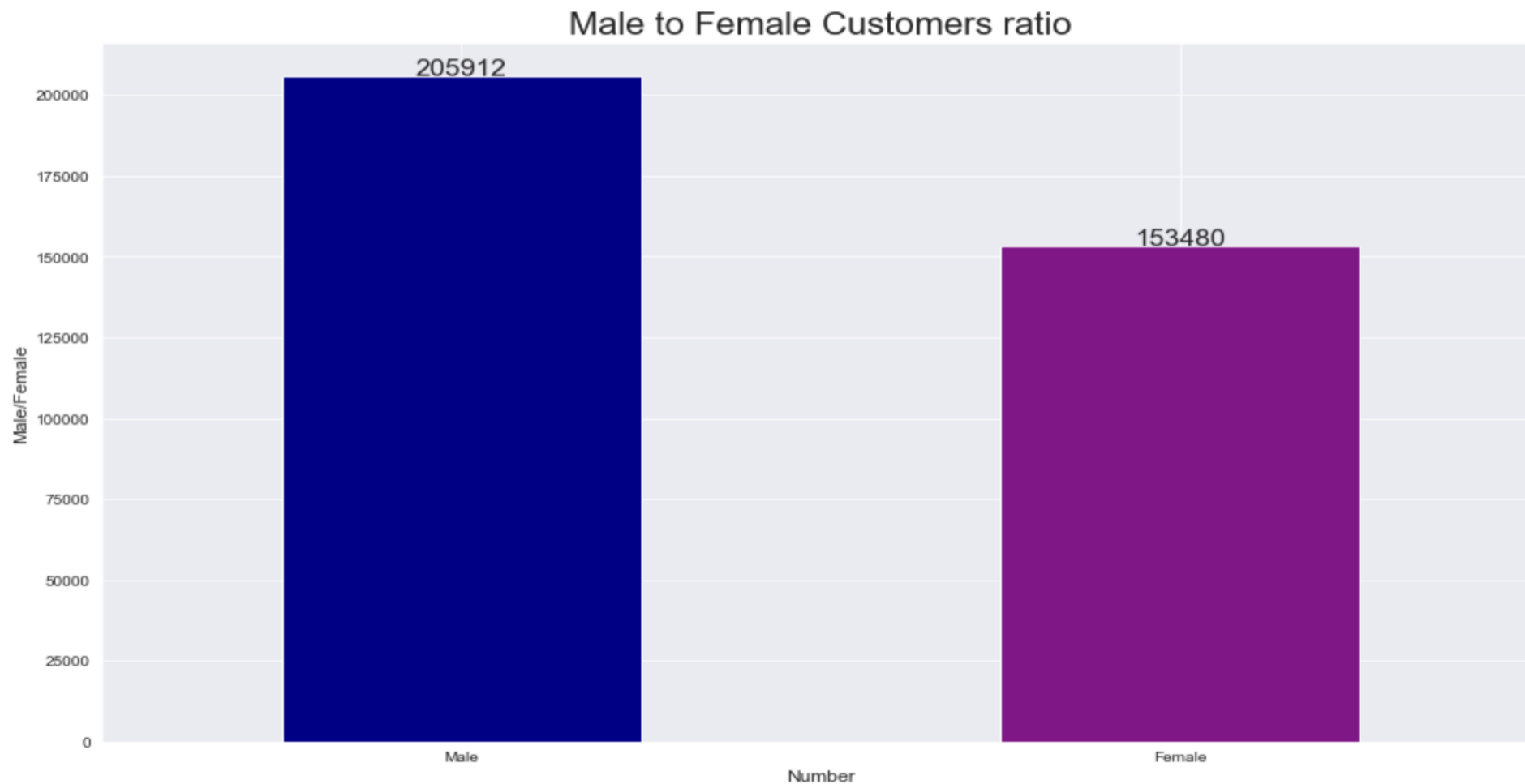
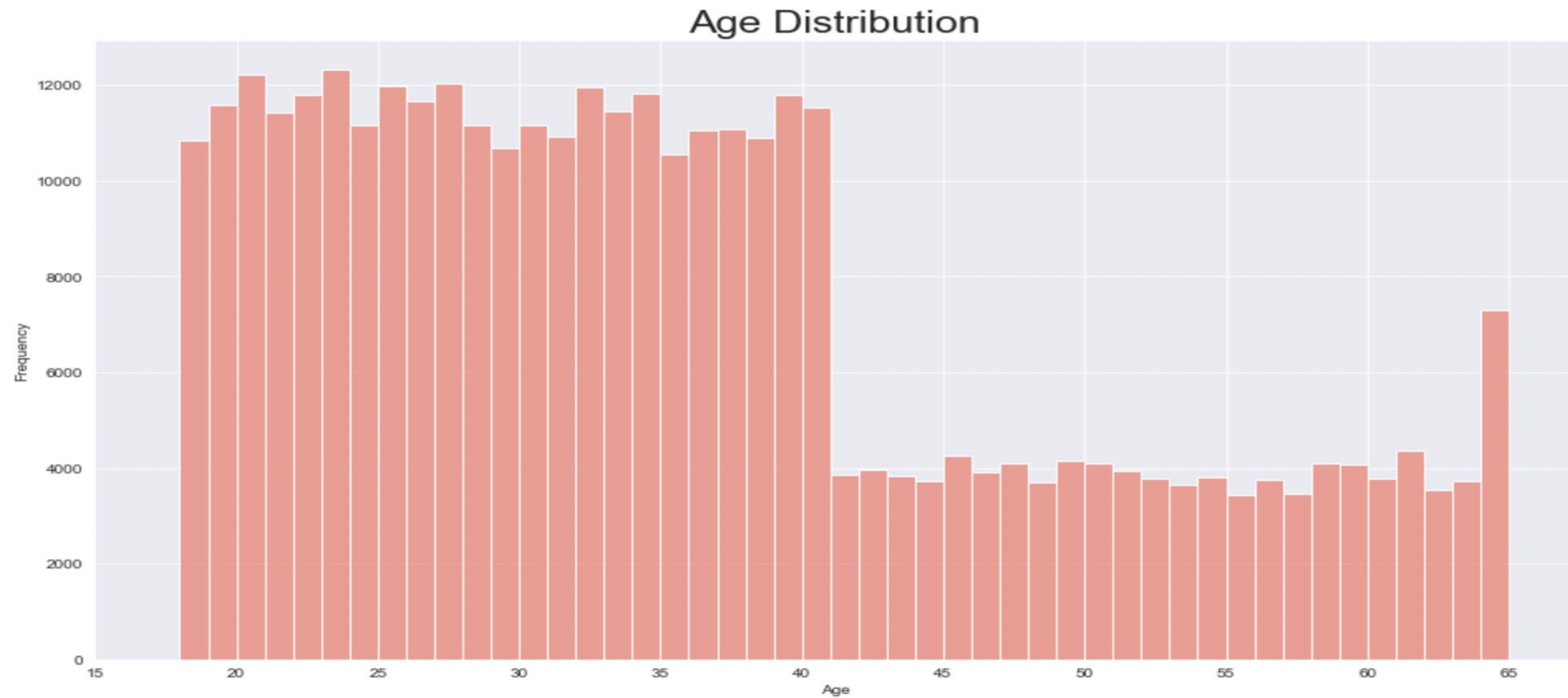Data Glacier

Hypothesis 3 answer

Q3- What cities where people use cab the most? In these cities which cab company is most demanded ?

Ans3- In almost every city, Yellow cab is much more demanded and profitable

Data Glacier

EDA: Male/Female Customers
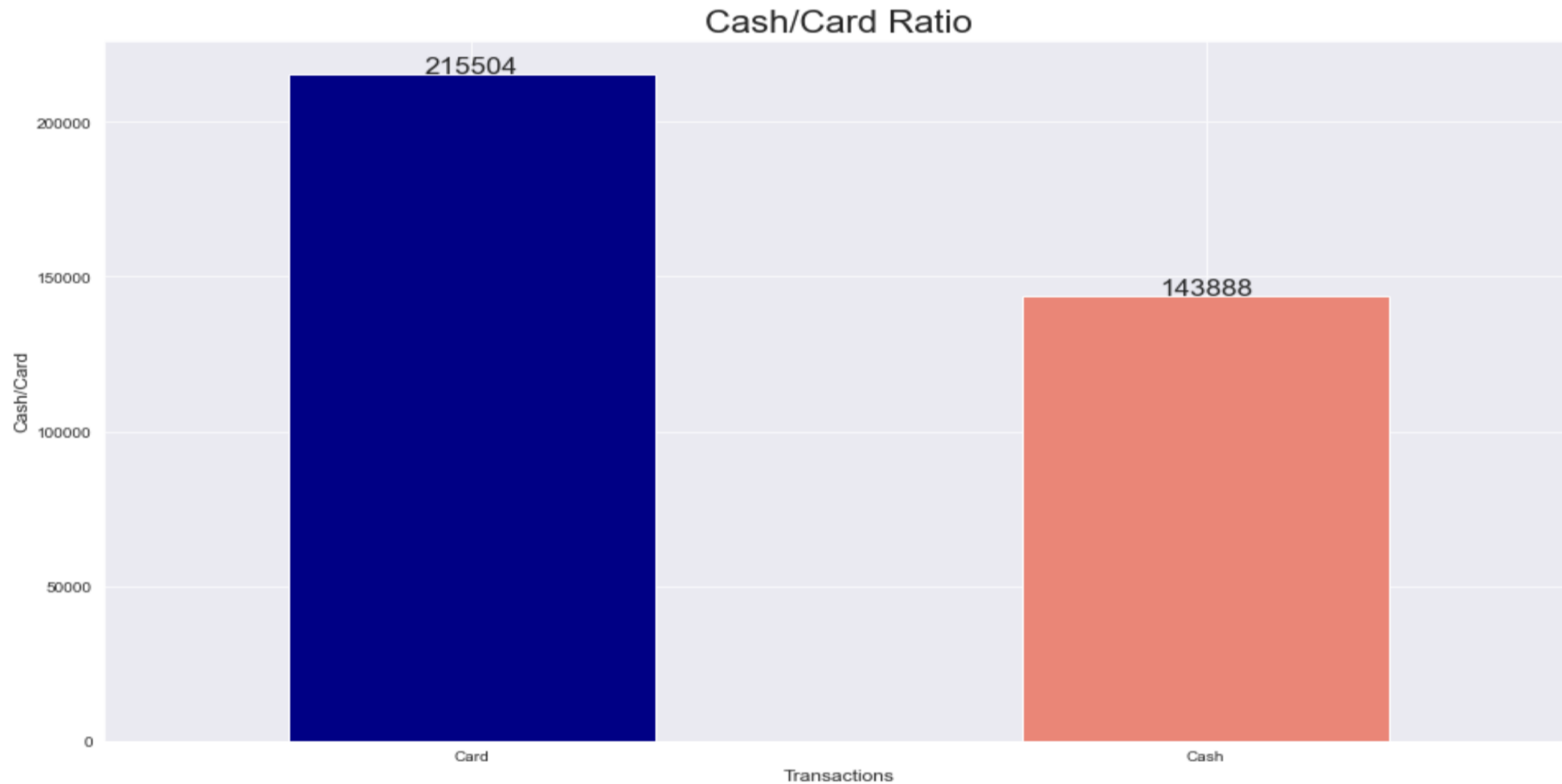
EDA: Age Distribution

Age Distribution

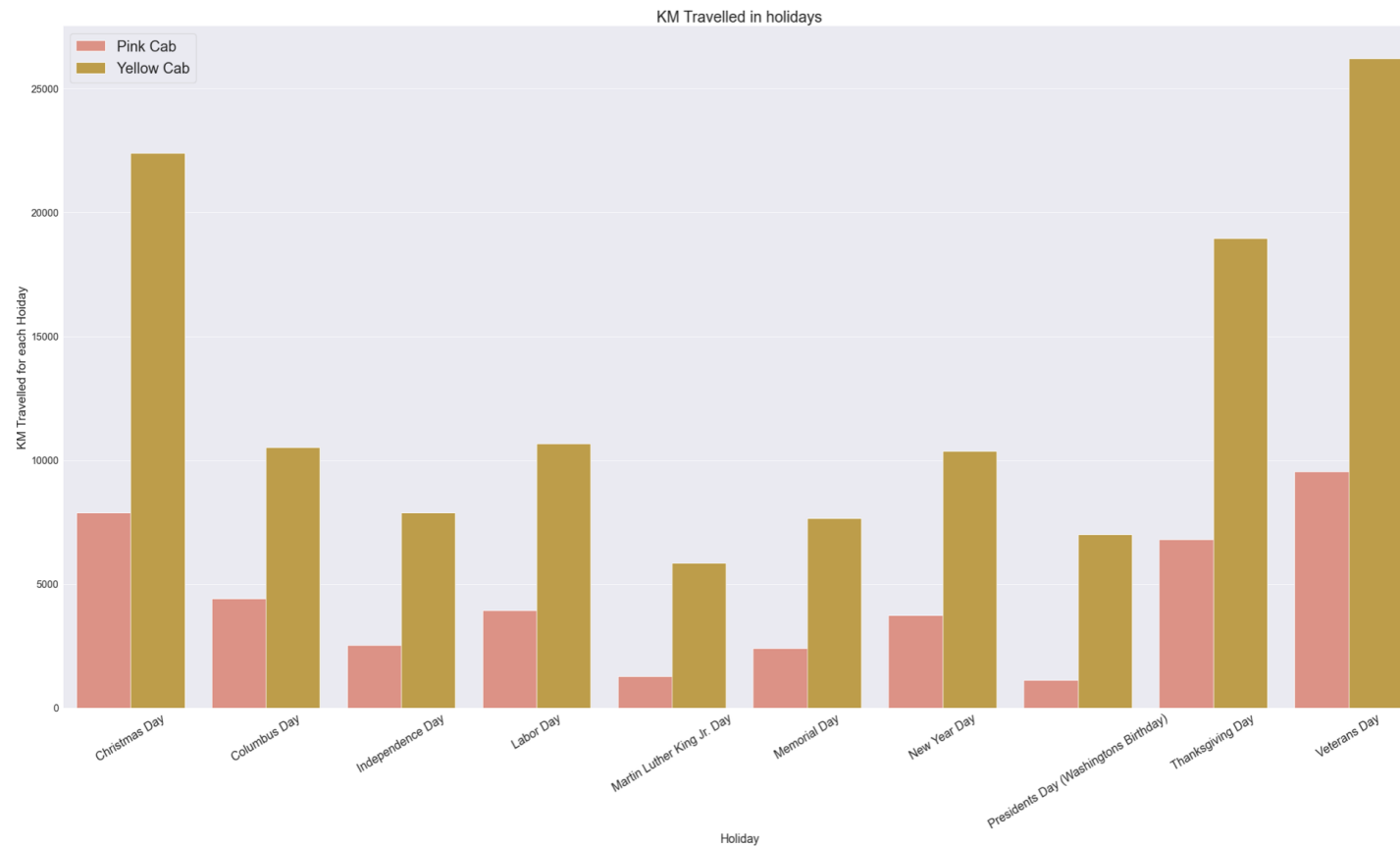**Most of the ages are between 16 and 40**

EDA: Cash/Card Payment

EDA: Holidays

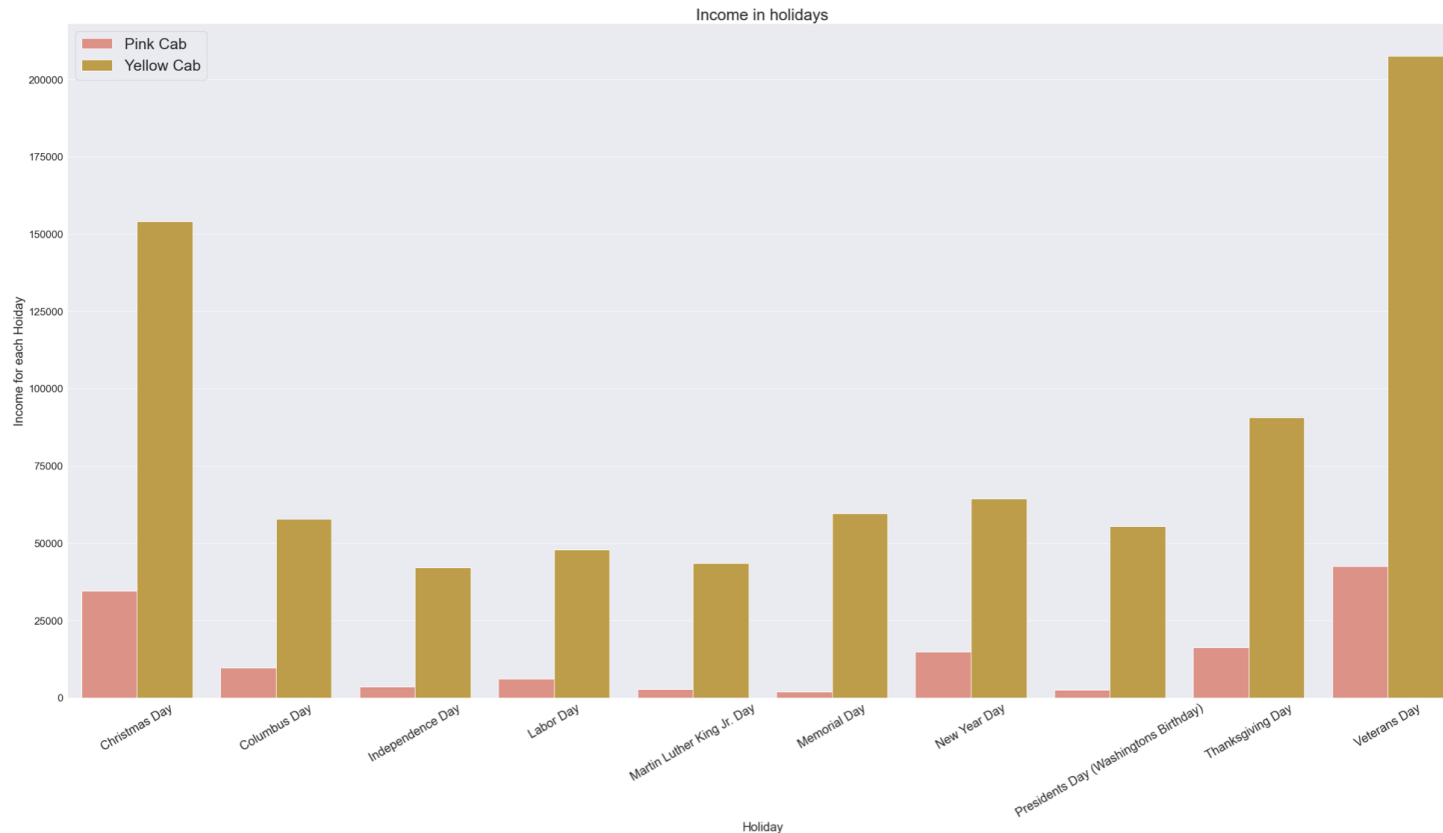## Q4: In holidays which company is the most demanded ?

- In every holiday the activity of Yellow Cab is minimum 4x the activity of Pink Cab.

EDA: Holidays

Q4: In holidays which company is the most demanded ?

Income in holidays

- In every holiday the Income of Yellow Cab is minimum 6x the Income of Pink Cab.

Hypothesis 4 answer

Q4: In holidays which company is the most demanded ?

Ans3- In every holiday, Yellow Cab is the most demanded and most Profitable

Agenda

EDA Summary.

EDA: Summary

Over the span of three years (2016, 2017, 2018):

- Total Yellow Cab rides are more than thrice Pink Cab rides.
- Total Yellow Cab revenue is 5x total Pink Cab revenue.
- Total Yellow Cab Income is 8x total Pink Cab Income.
- In every year Yellow Cab travel 3x more than Pink Cabs.
- In every year Yellow Cab has approx. 5x more revenue than Pink Cab.
- In every year Yellow Cab has approx. 8x more profit than Pink Cab.
- In almost every city, Yellow cab is much more demanded and profitable.
- In every public holiday, Yellow Cab is the most demanded and most Profitable.

Data Glacier
Your Deep Learning Partner

Agenda

Conclusion

Data Glacier
Your Deep Learning Partner

Conclusion

In Conclusion: Based on the investigation above, it is safe to say you should invest in the Company 'Yellow Cab'.

Data Glacier

Your Deep Learning Partner

Thank You

Data Glacier

Your Deep Learning Partner