Exploratory Data Analysis

G2M Insight For Cab Investment Firm

By Yanjun Lin
2022-07-07

Agenda

Executive Summary

Data Exploration

Approach

EDA

EDA Summary

Recommendations

Summary

XYZ is a private equity firm in the US. Due to remarkable growth in the Cab Industry in the last few years and multiple key players in the market, it is planning for an investment in the Cab industry.

As the intern whom the project was assigned to, I will conduct full quantitative analyses on existing datasets and provide market trends on both Yellow and Pink cab companies to assist XYZ making investment decisions.

The analysis has been divided into four parts:
• Data Understanding and Visualization
• Finding the most users Cab company
• Finding the cheapest Cab company for users
• Finding the most profitable Cab company
• Multiple Hypothesis and Investigate

Data Exploration

- Cab_Data.csv – this file includes 7 features and details of transactions for 2 cab companies

- City.csv – this file contains 3 features includes a list of US cities, their population, and the number of cab users

- Customer_ID.csv – this is a mapping table with 4 features that links the customer's demographic details

- Transaction_ID.csv – this table contains 3 features of customer transactions and payment mode

- Geo.csv – this file includes 3 features: city name and its coordinates

Data Exploration
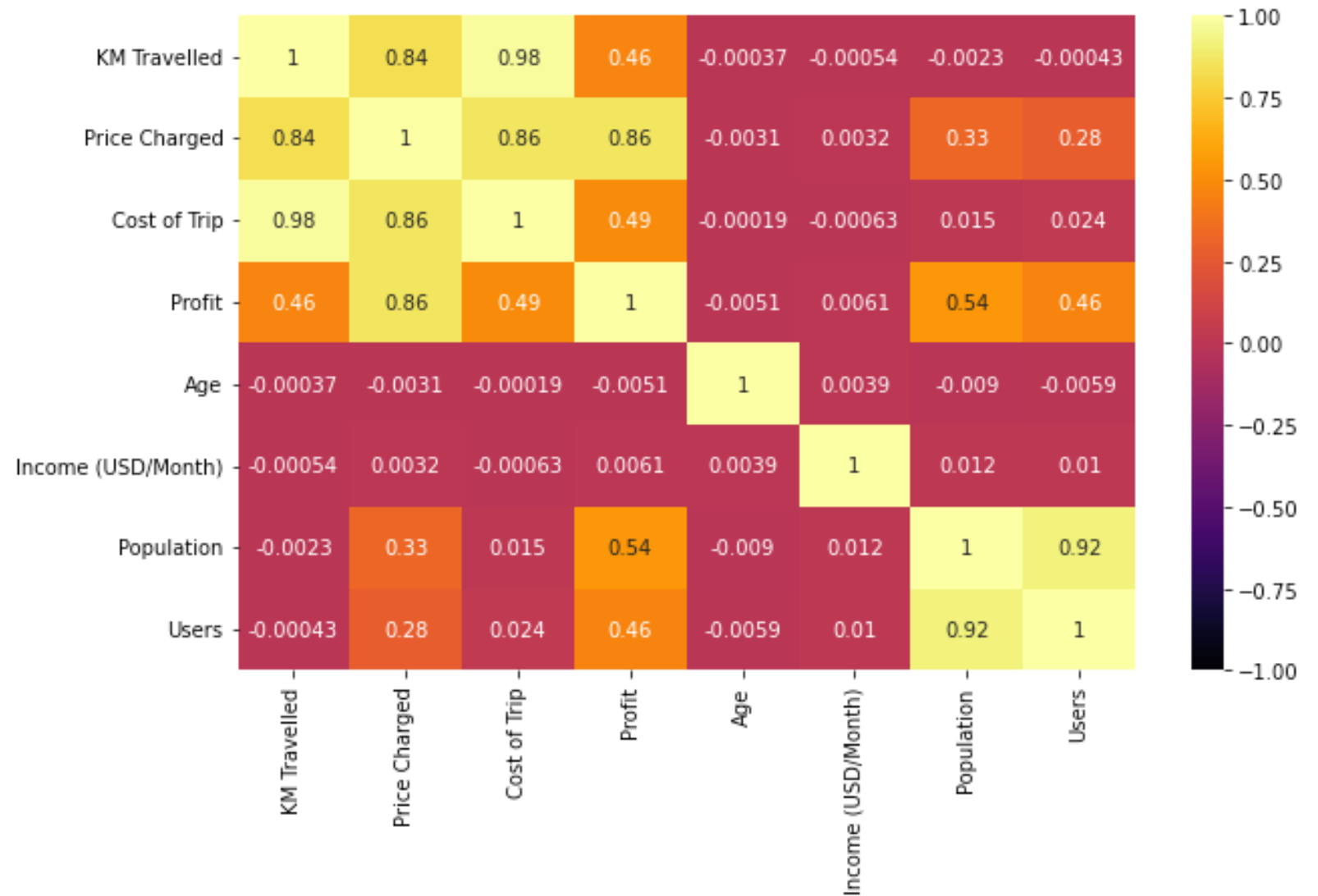
We can see from the above image that KM Travelled, Price Charged, Cost of Trip, Population, and Users have positive correlation with each other

Assumptions

Yellow vs. Pink
- How many rides were taken by customers in each company?
- Which city cost more for each company for every KM travelled? (Ave cost per KM by company by city)
- Which company in which city has the highest profit margin?
- Dominate cities by company

Users
- Does age group in each gender and each company matter?
- Does gender matter for each company?
- Are there differences between Age group in each city?
- Is travel distance influenced by Age?
- Relationship of age and income
- Dominate income group in genders by company

Cities
- User out of population by city
- Income by city

Dates
- How were cab demands influenced over the years for both companies overall and in each city?
- Are there certain time of the year and month affects cab demands? (Seasonal effects year and month)
- Is cab demand affected by day of the week?
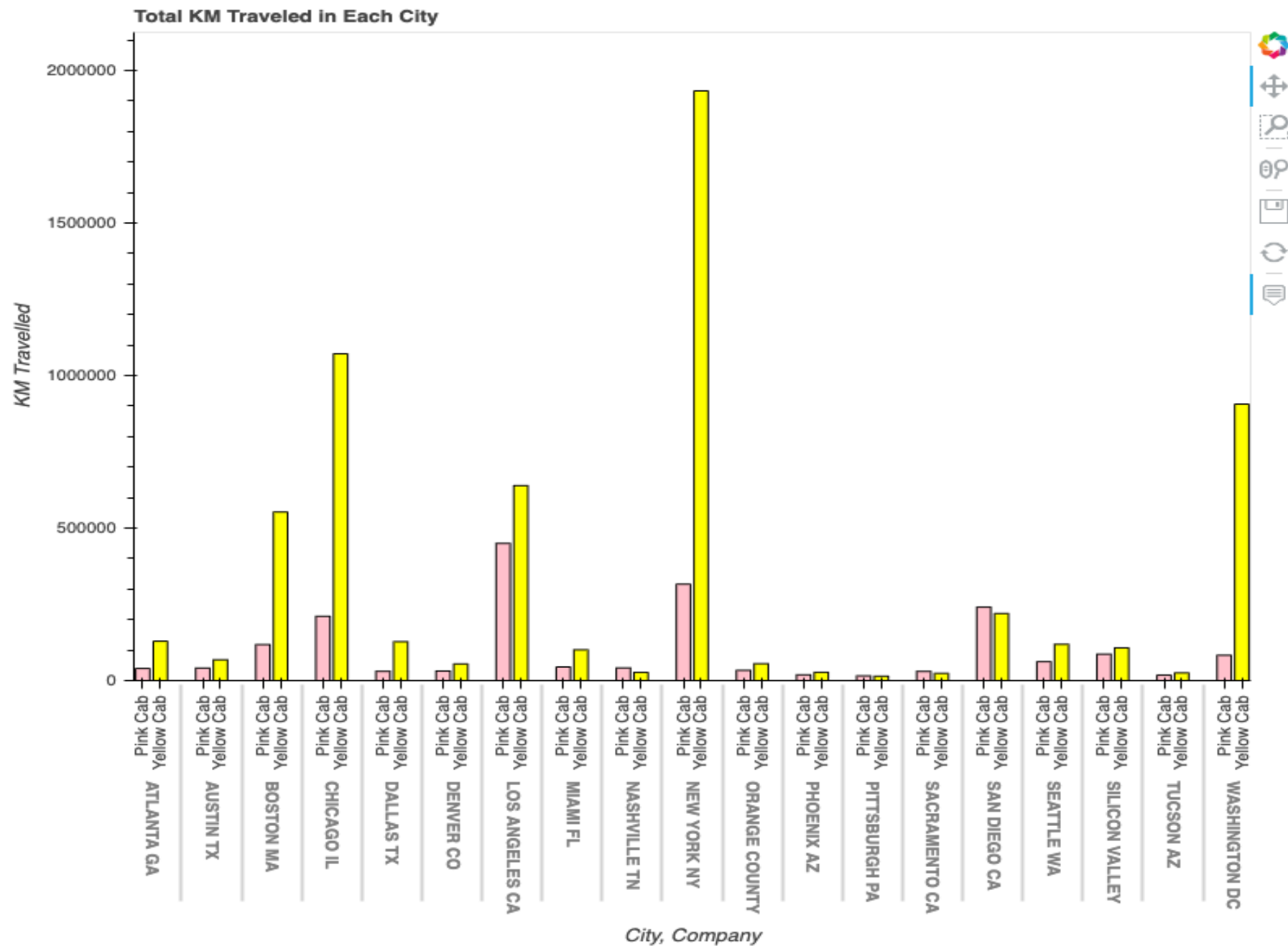- How was profit margin affected by time?

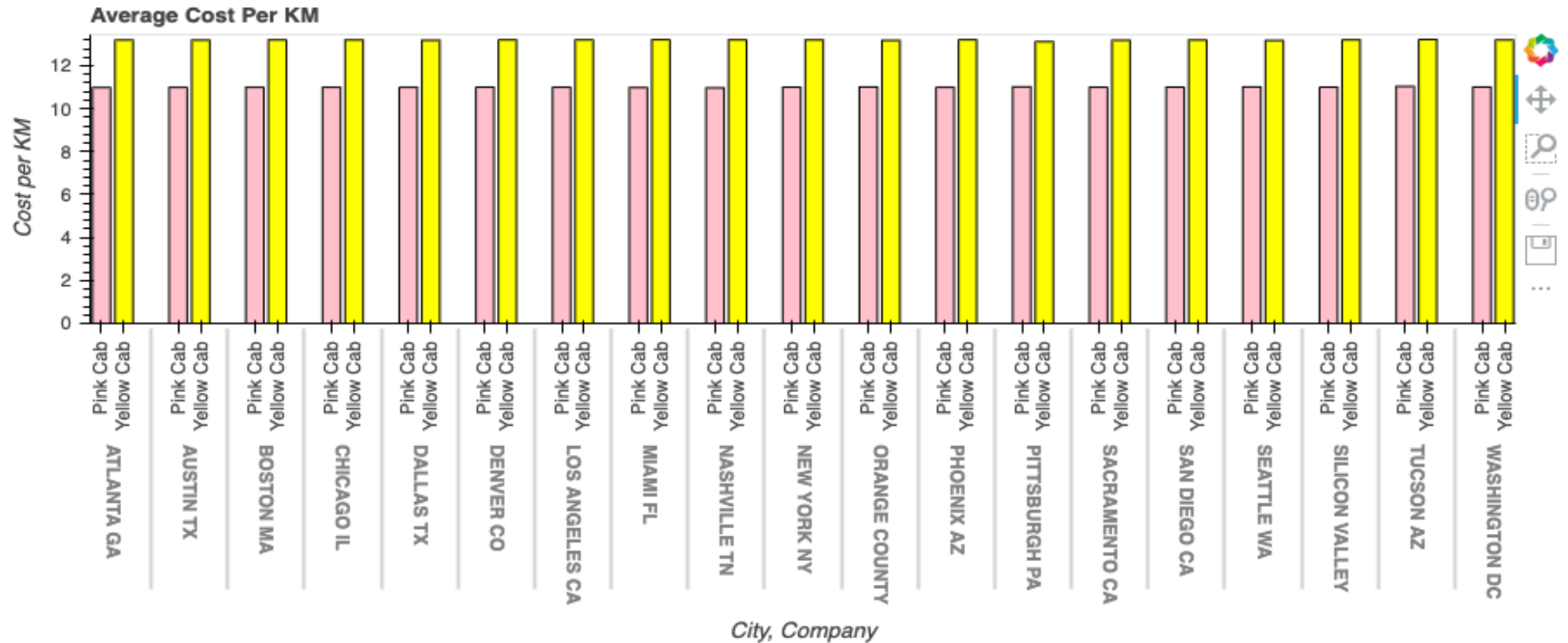# EDA

Total Rides Distribution 2016-2018

Pink Cab — 23.6%
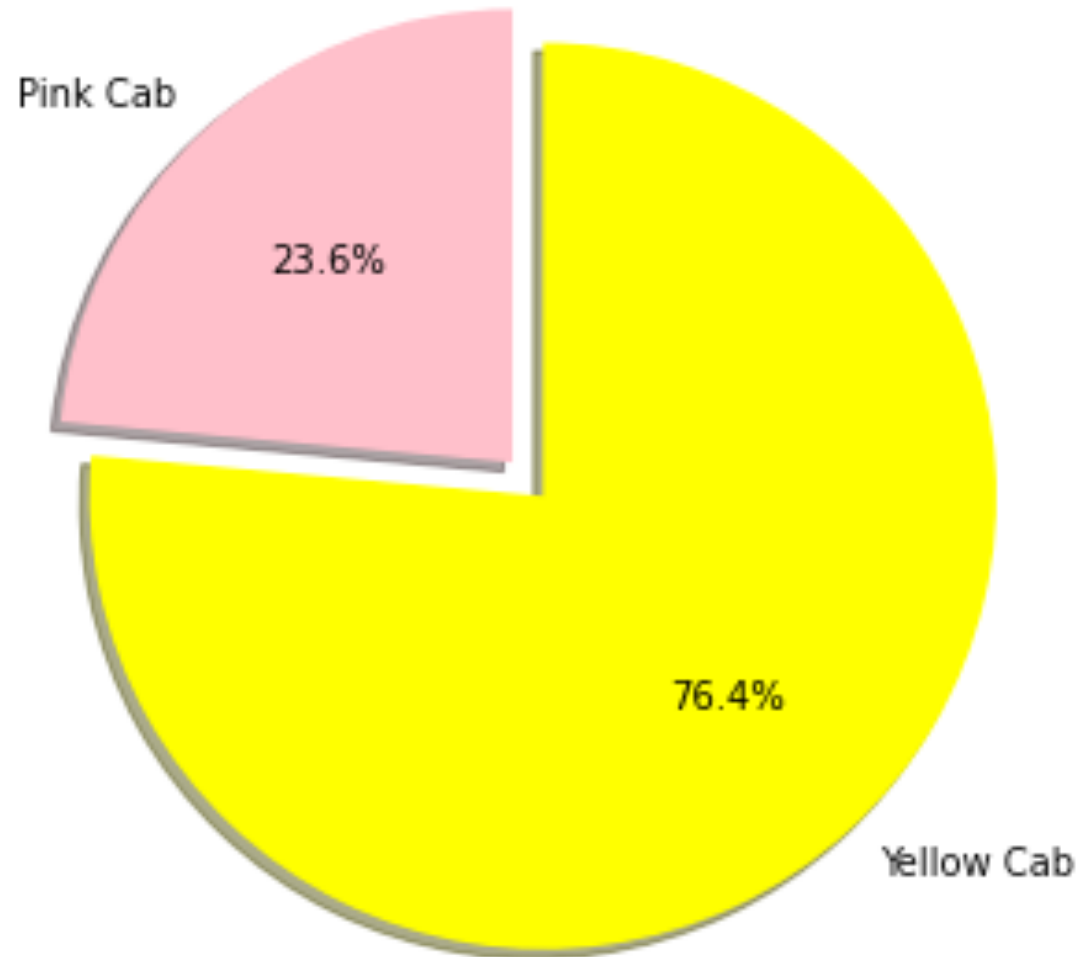
Yellow Cab — 76.4%

During the analyzing period, more than 4 times of the rides were taken in Yellow cab than in Pink cab.

EDA

Total KM Traveled in Each City

EDA

Average Cost Per KM

The bar chart above shows it generally cost $2 more every trip in all cities for Yellow cab than Pink cab

As we found out earlier that there are way more Yellow cabs than Pink cabs, which means it cost way more to drive and ride a Yellow cab, as well as to run its' company

The average cost per KM in all cities are generally the same

EDA

Average Profit Margin
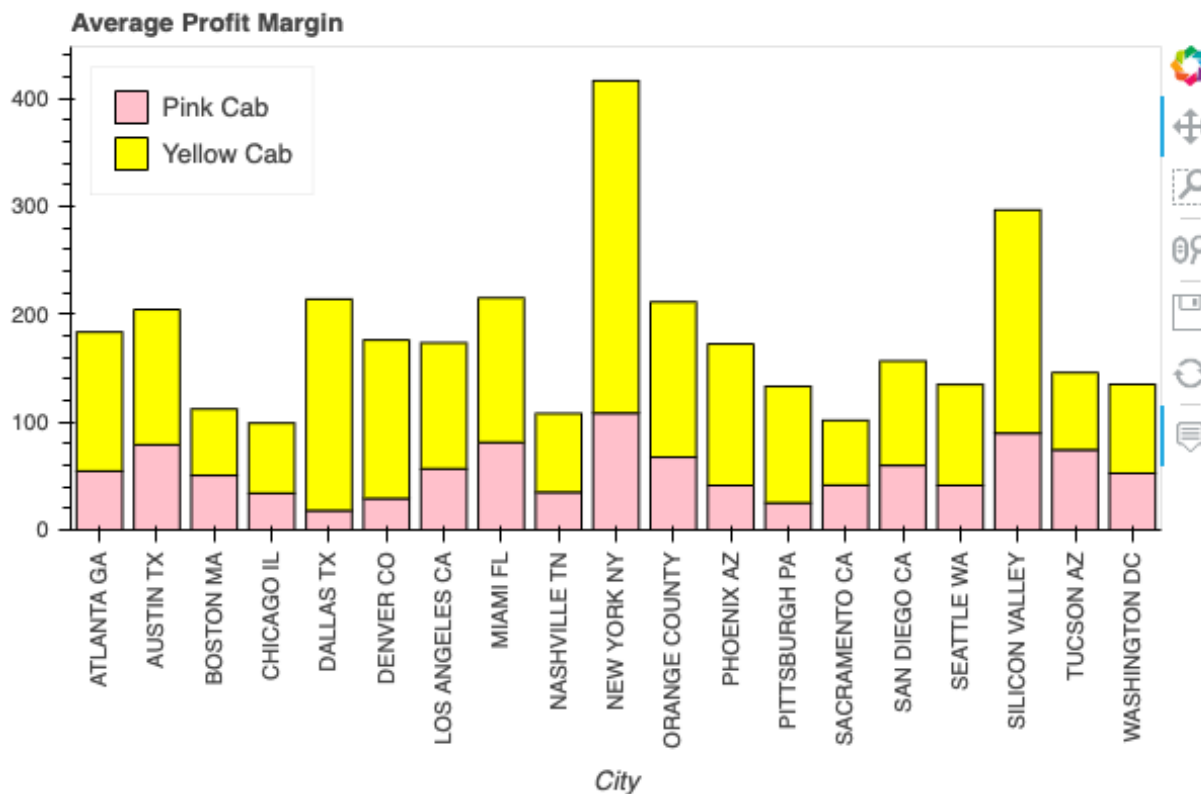
If the average cost per KM is generally the same among all cities provided (Previous section), the the higher the average profit margin means customers were charged higher

With a slightly lower average cost per KM, Pink cab shows low profit margin, they are charging customers way lower than Yellow cabs

Generally, Pink cab has lots of rooms for profit margin in all cities, but especially in Dallas, Denver, New York, Phoenix, Pittsburgh, and Silicon Valley

EDA

Total Pink Cab Rides 2016-2018

Total Yellow Cab Rides 2016-2018
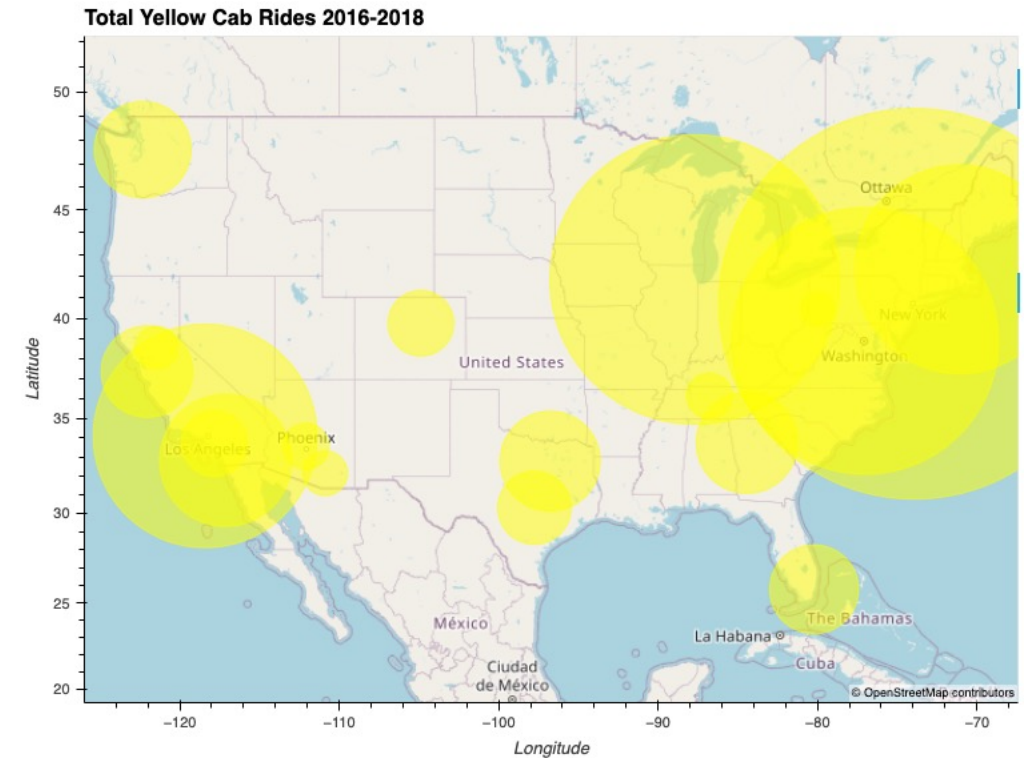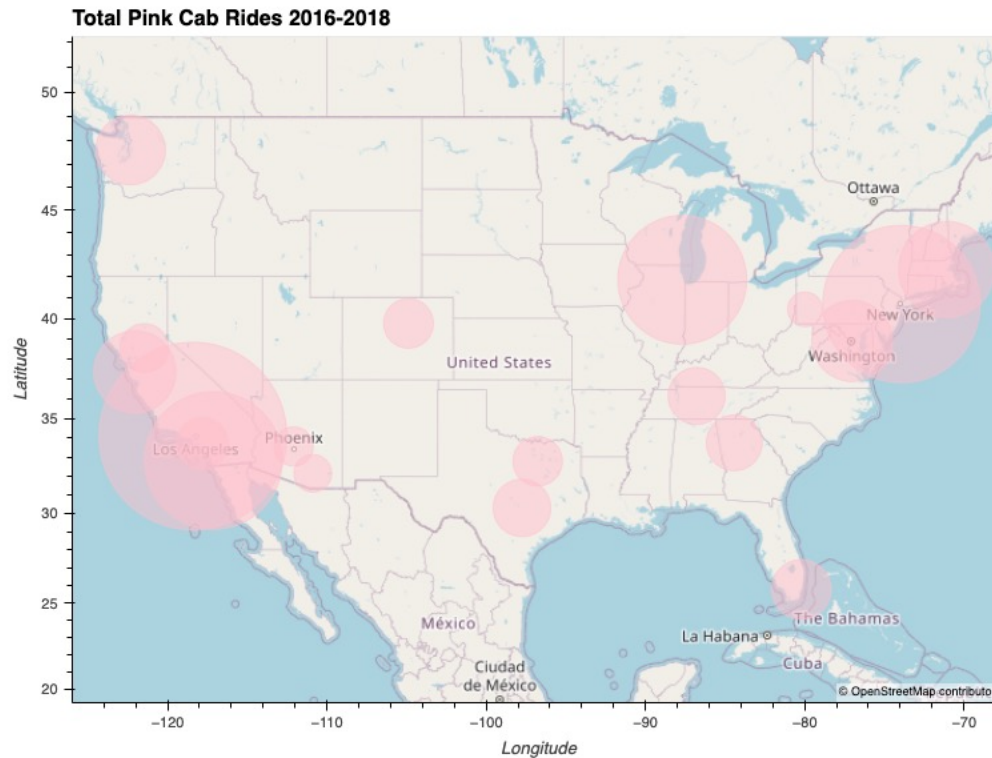
Yellow cabs outrun Pink cabs by a huge amount especially in east coast cities such as, New York, Boston, Chicago, and DC

Pink cabs has more advantage on the west coast cities such as Los Angeles, San Diego, Silicon Valley while Yellow cabs dominate the east coast

We can't directly see how many Yellow or Pink cabs are running in each city, but it is obvious that there are way more Yellow cabs than Pink cabs in most of the cities

There are high demands for cabs in Boston, Chicago, Los Angeles, New York, and DC; and moderate demands for San Diego, Seattle, Silicon Valley, and Atlanta than some of the Midwest cities

EDA
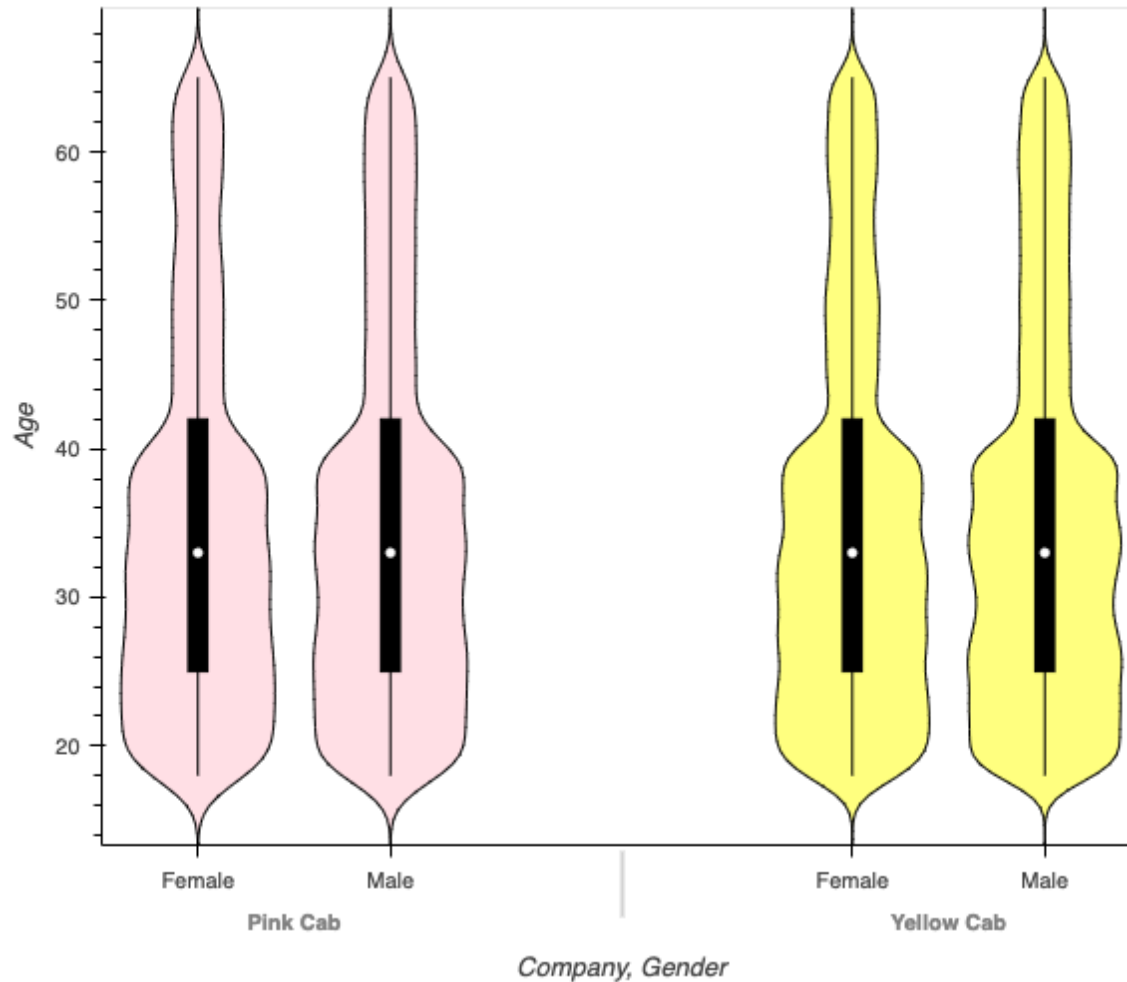
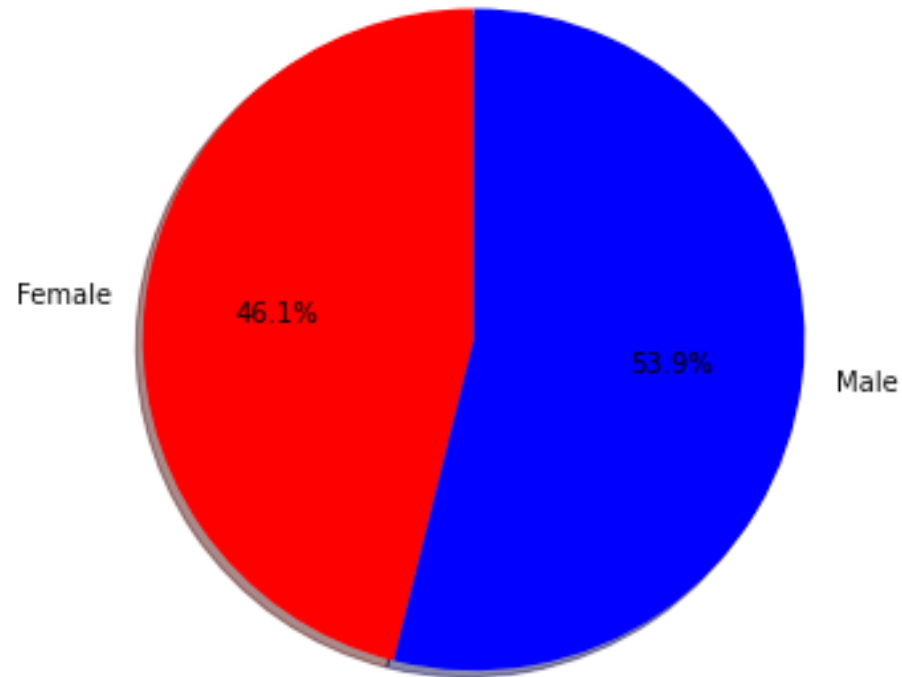Both companies dominate user age group is 20-40 years old

Gender distribution for both companies are quite similar

Gender distribution within each company are generally even between Female and Male users
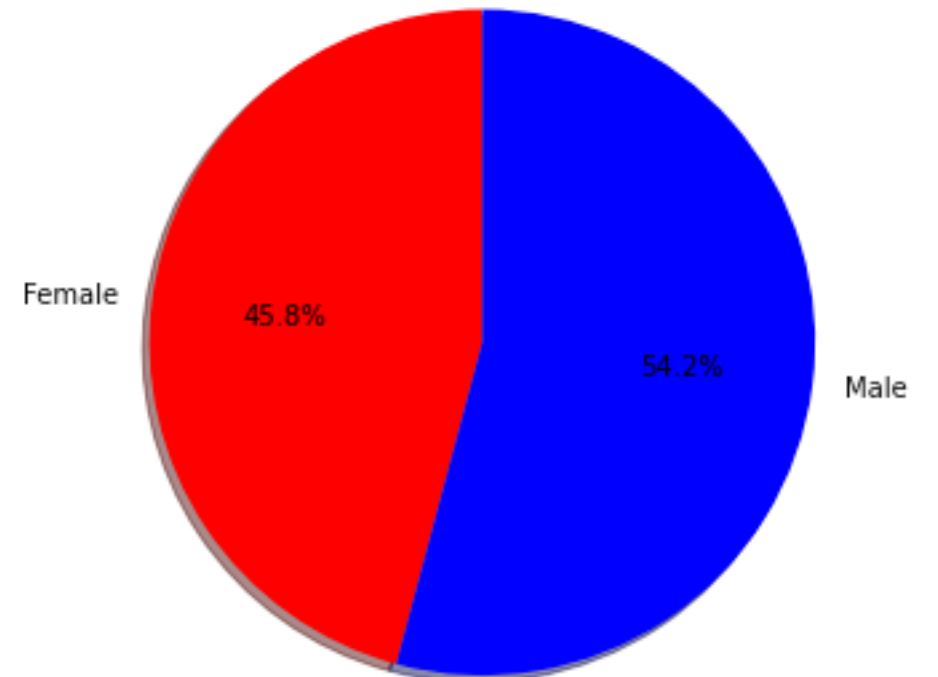
The average user age for both companies and both genders is about 33 years old
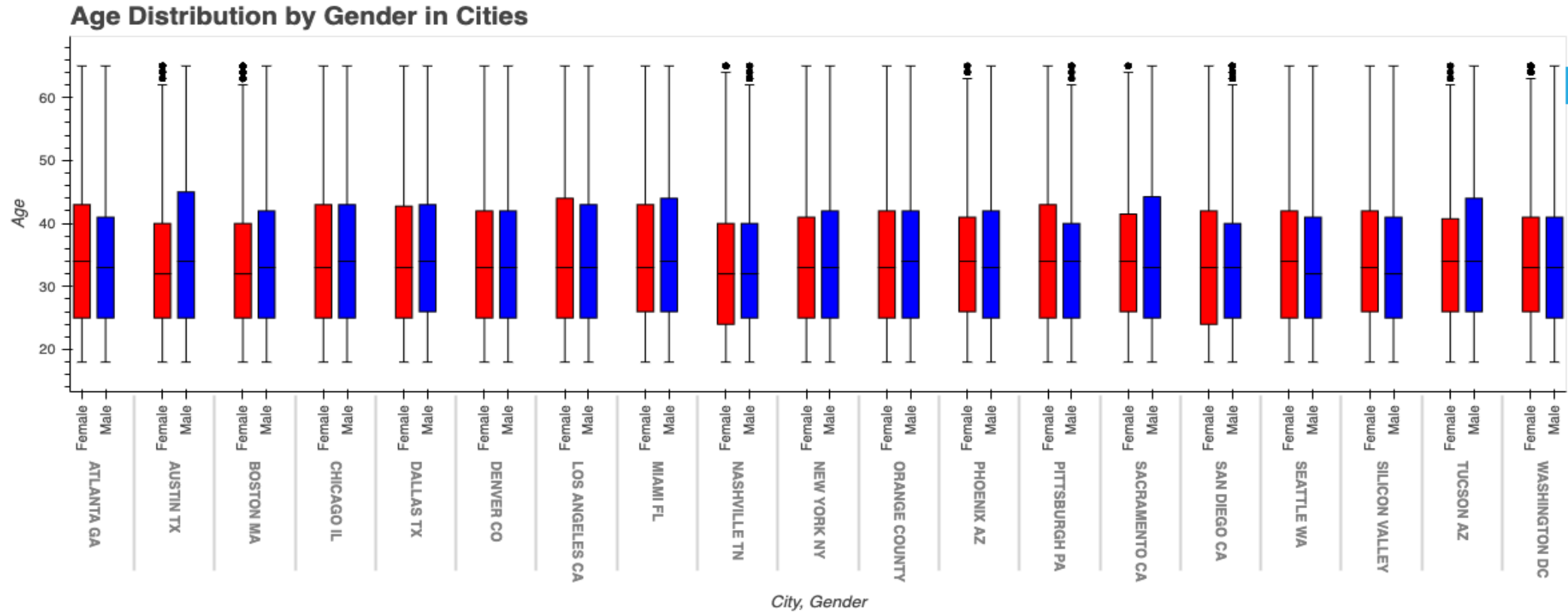
EDA

Yellow Cab Gender Distribution

Female 46.1%

Male 53.9%

Pink Cab Gender Distribution

Female 45.8%

Male 54.2%

Color Pink is not an influencer in gender distribution for Pink Cab

Both companies has more male than female customers

EDA

Age Distribution by Gender in Cities
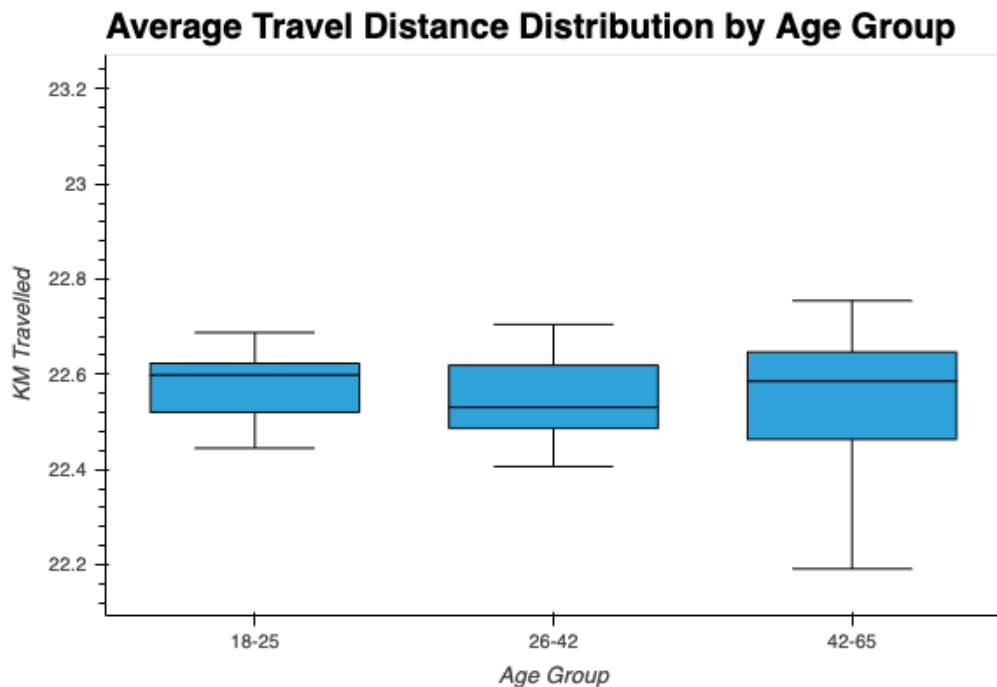
Age distributions by gender are quite even in each city. There is no big age difference between higher and lower cab demand cities

Pink cabs has wider female customer age ranges than males in cities such as Los Angeles, San Diego, Silicon Valley Nashville, Pittsburgh, Seattle where Pink Cab performs better than other cities

EDA
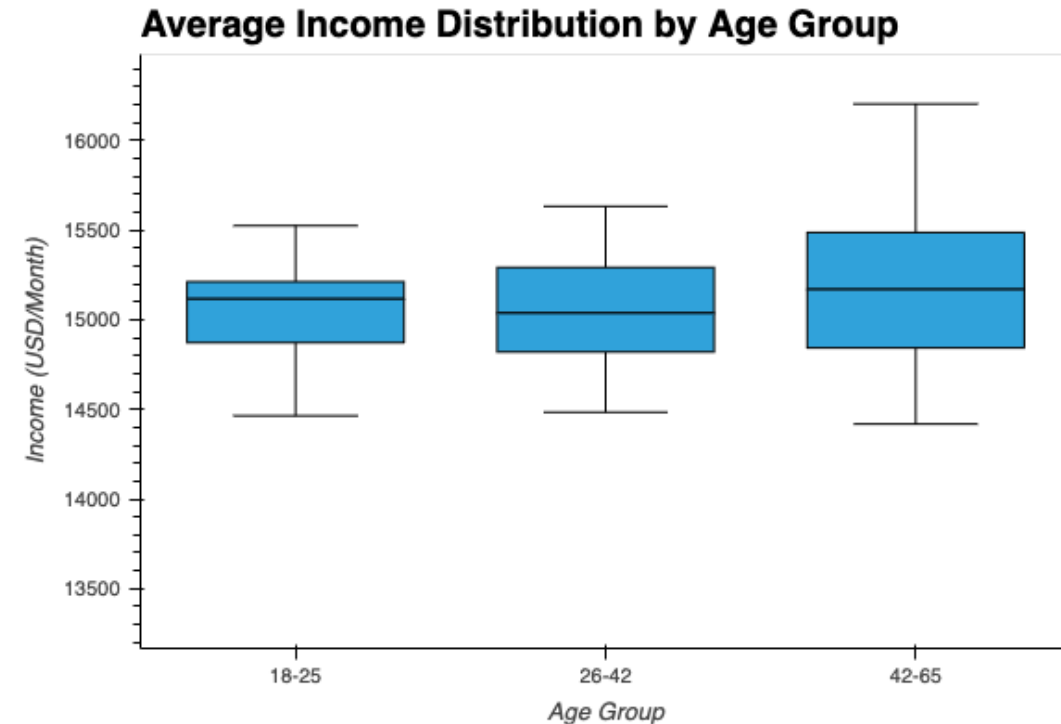
As age increases, cab travel distance generally increases, so does the variation of travel distance

Customers from younger age groups generally travel within 23 KMs while customers from 42-65 age group have higher needs taking cabs to closer destinations

As age increases, income generally increases, so does the income variation

Customers from younger age groups have incomes between 14500 and 15700, while customers from 42-65 age group have income as high as 16300 and as low as 14400

EDA

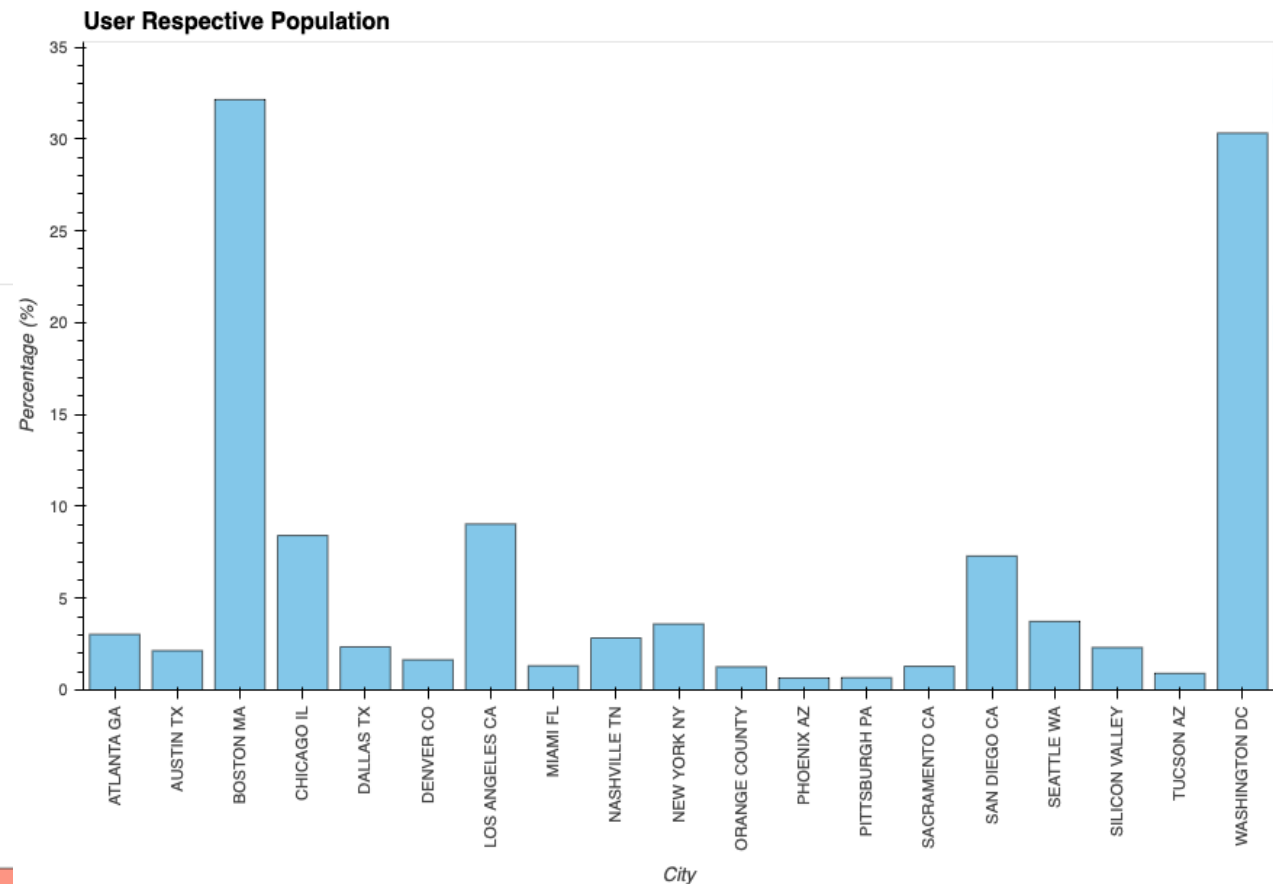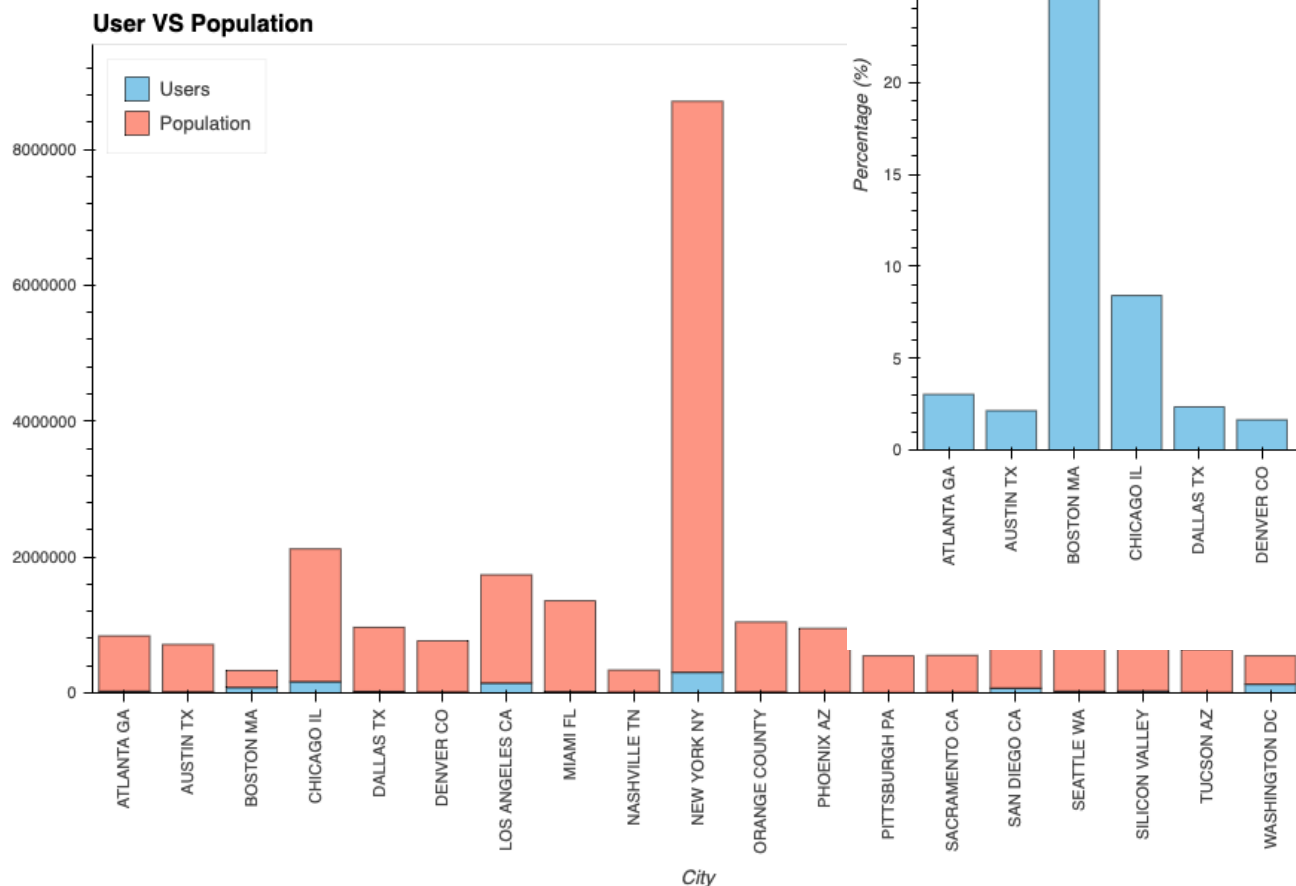**User Income Distribution by Gender by Company**
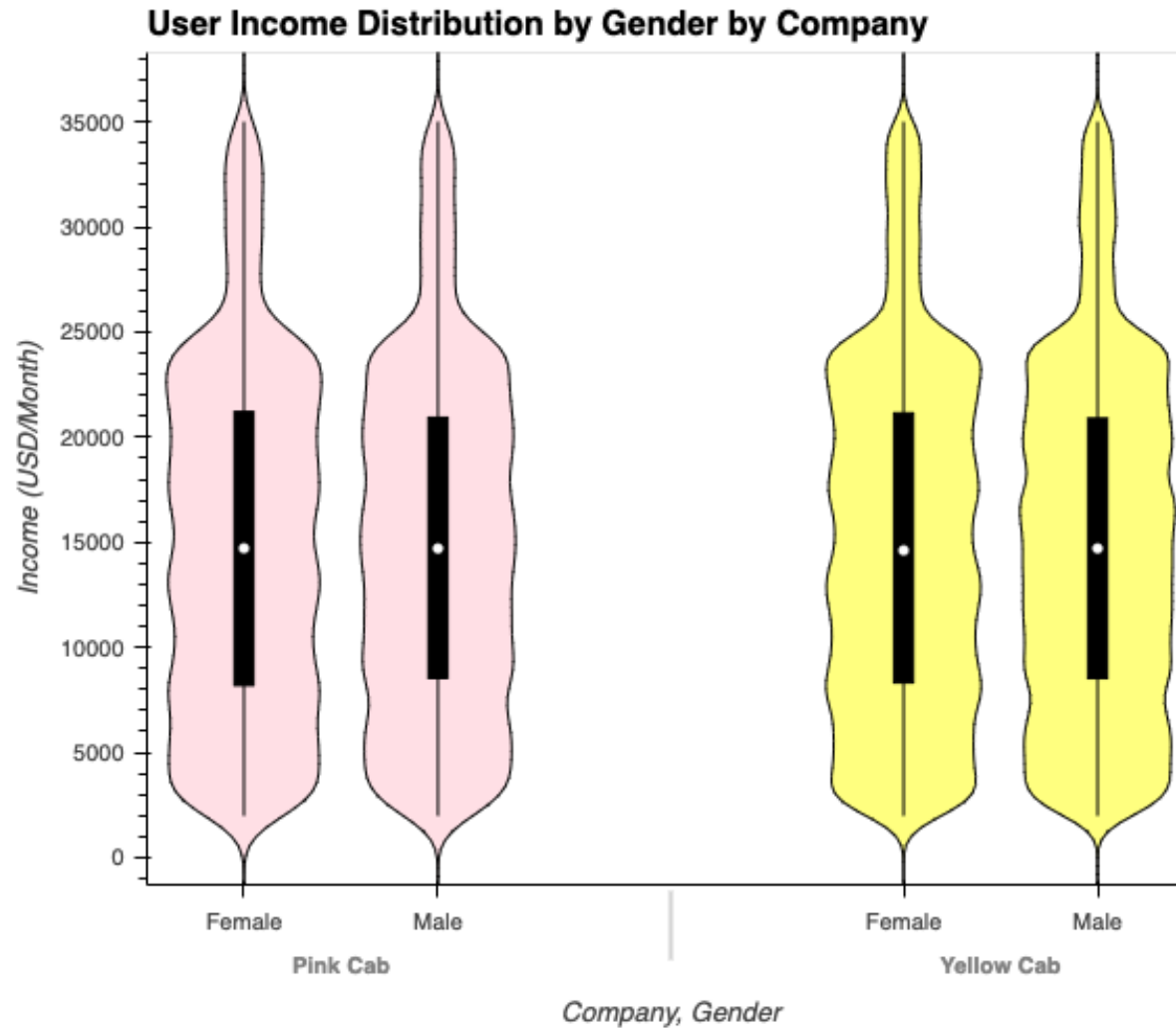
Both companies have similar income distributions in each gender groups

The majority of customers' income fall under 3000 - 24000

Income in both companies and gender groups are heavily weighted below it's mean

EDA

User VS Population

User Respective Population

Cab users only make up a minimal portion of the total population in each city

Only (6 out of 19 cities) New York, Chicago, Los Angeles, DC, Boston, and San Diego have more obvious amount of users comparing to other cities
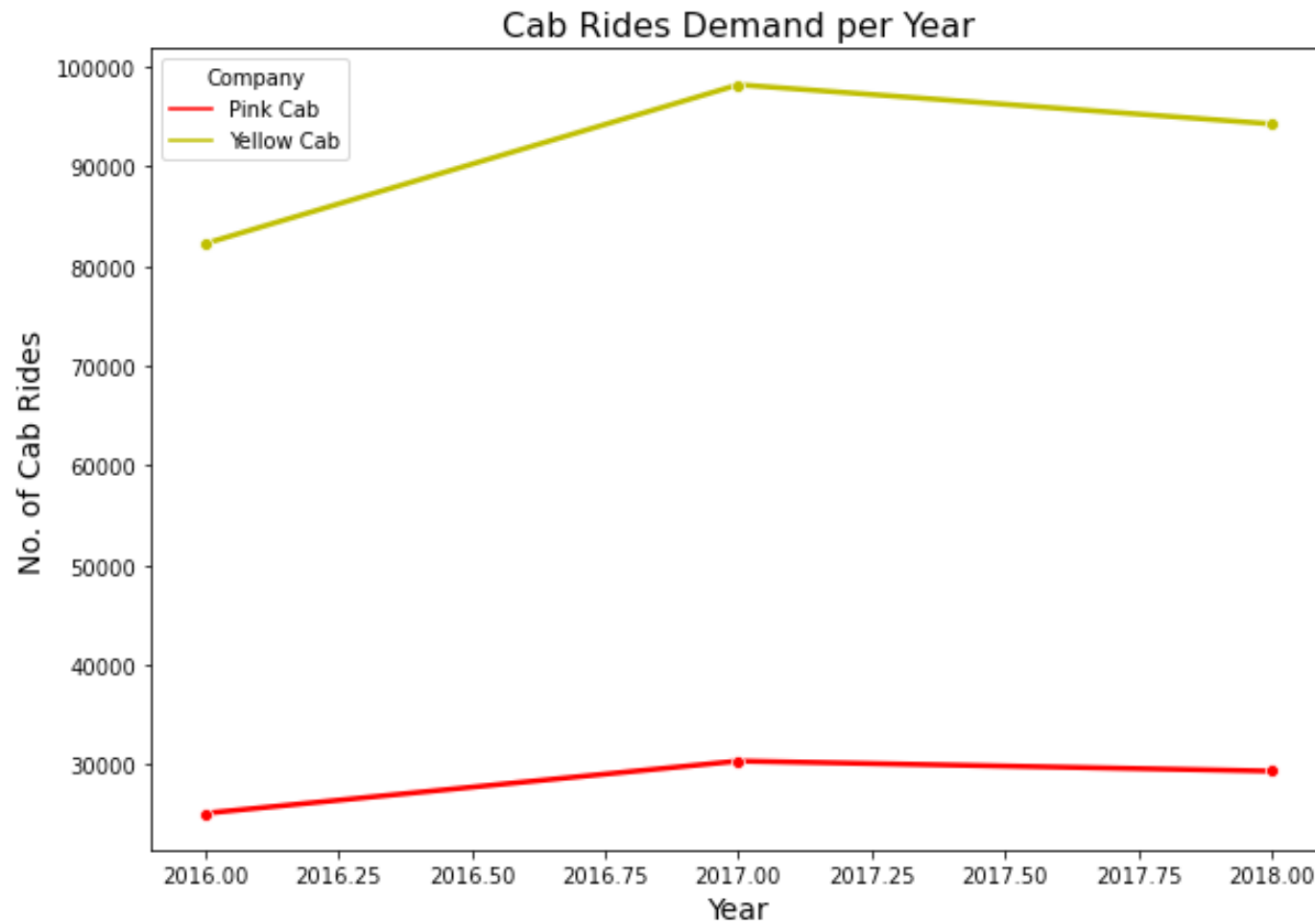
Even though New York have more users, there are around 30% Boston and DC population use cab services

Data Glacier
Your Deep Learning Partner

EDA
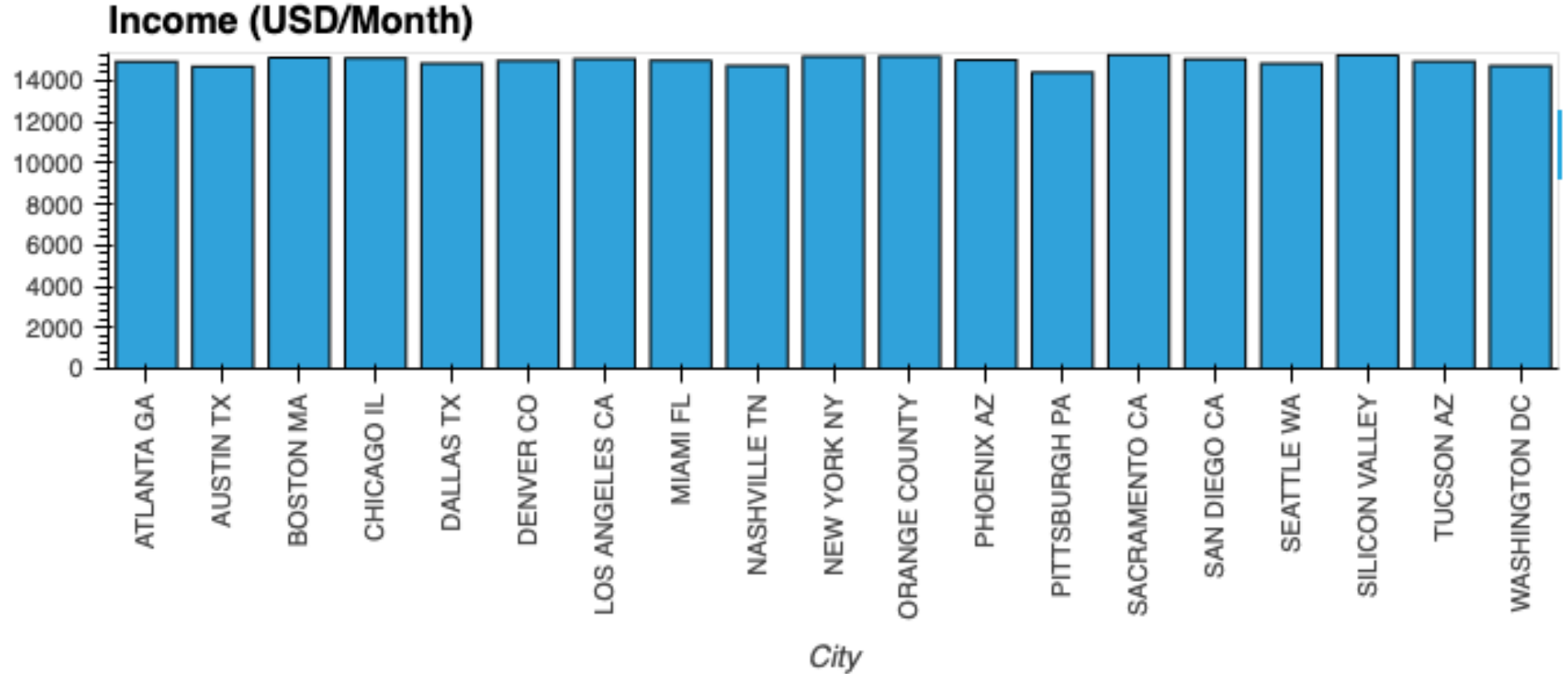
Regardless of demands, cab users' average income level are around 14000 in the list of cities
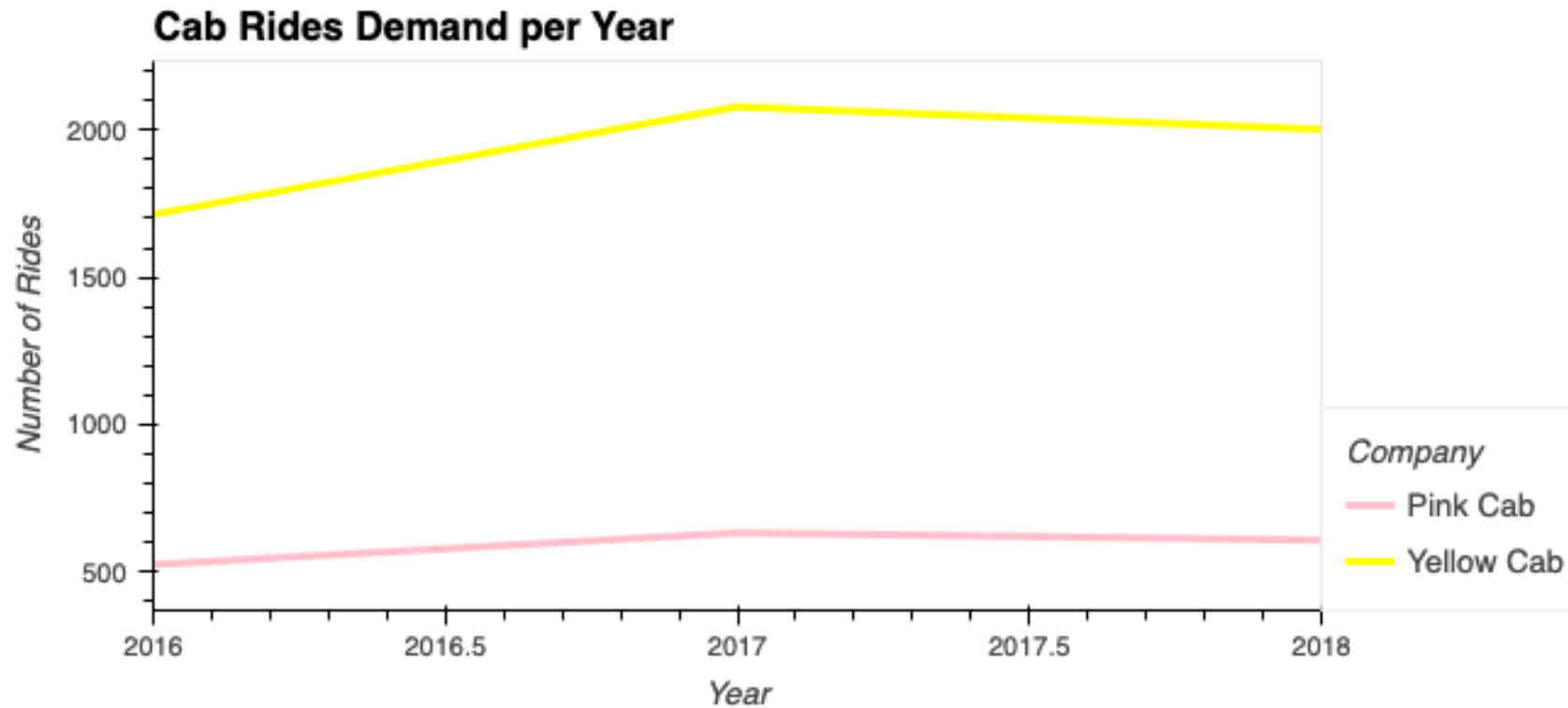
EDA

Cab Rides Demand per Year

The overall demand for cab service increased for both companies from 2016 to 2017 followed by a down trend towards 2018

The demand for Yellow cab service increased higher than Pink cab; it also decreased faster than Pink cab

Pink cab had a steady increase in demand from 2016 to 2017 and kept a stable demand record to 2018
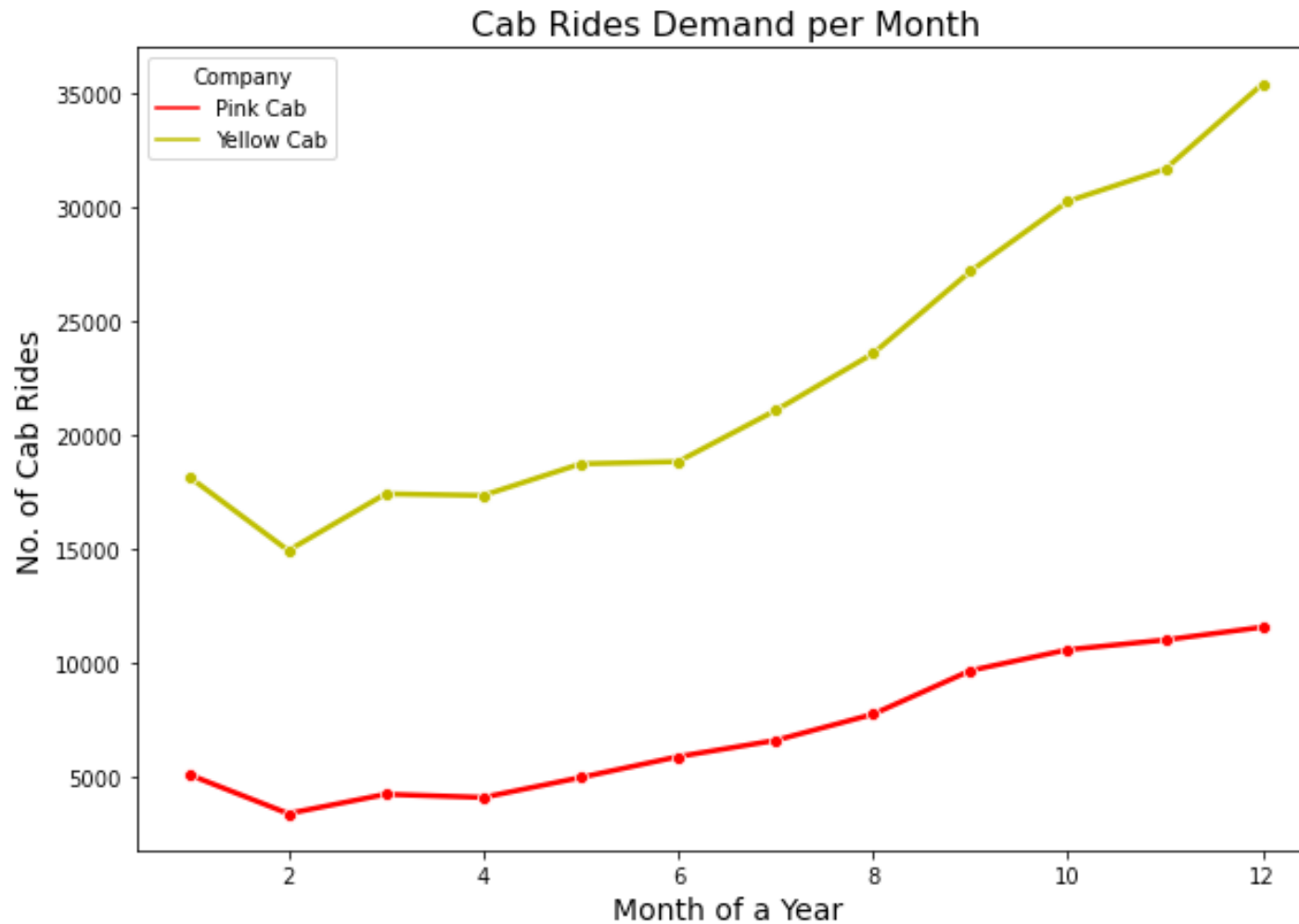
EDA

Overall, Yellow cab has higher demand and the demand for its service increased faster than Pink cab

Demand for Yellow cab increased high in: Atlanta, Austin, Denver, Los Angeles, Miami, Orange County, Phoenix, Silicon Valley, Tucson; Demand decrease slow in: Denver, Chicago, DC

Demand for Pink cab increased high in: Austin, Denver, Los Angeles, Orange County, Phoenix, Silicon Valley, Tucson; Stable demand in: Atlanta, Boston, Chicago, Dallas, Miami, New York, Seattle, DC

Demand for `Pink` cab higher than Yellow cab: Nashville, Pittsburgh, Sacramento, San Diego
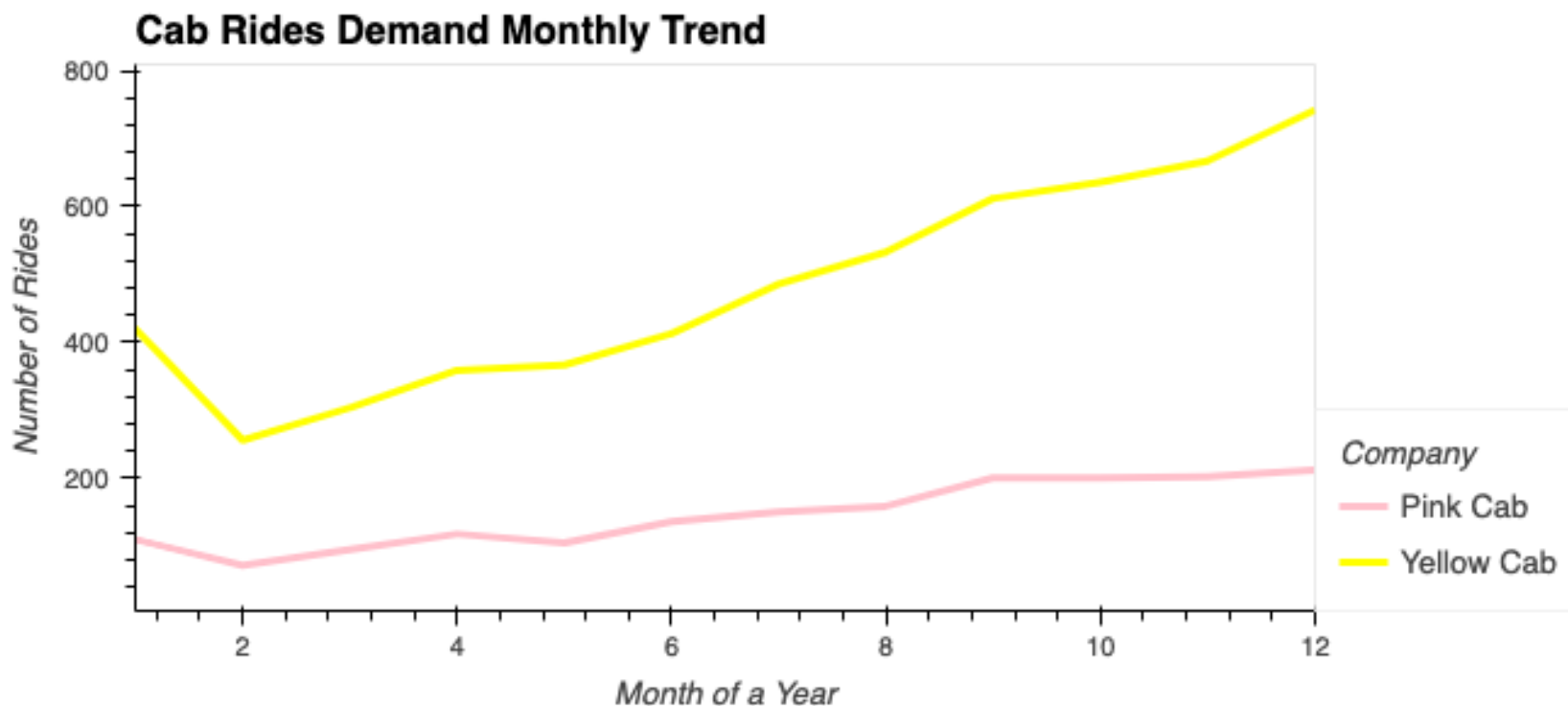
EDA

Cab Rides Demand per Month

The overall demand trend for cab service increases faster through summer to winter and drops around spring

The demand trend for Yellow cab service increased higher than Pink cab; it also decreased faster than Pink cab in spring

Pink cab had a steady increase in demand from summer to winter and drops slightly during spring

EDA

Cab Rides Demand Monthly Trend

Overall, Yellow cab has higher demand and the demand for its service increased faster than Pink cab throughout the year

Demand for Yellow cab increased high in: Atlanta, Austin, Denver, Los Angeles, Miami, New York, Orange County, Phoenix, Silicon Valley, Tucson; Demand decrease slow in: Chicago, DC

Demand for Pink cab increased high in: Austin, Denver, Los Angeles, Orange County, Phoenix, Silicon Valley, Tucson; Stable demand in: Atlanta, Boston, Chicago, Dallas, Miami, New York, DC

Demand for Pink cab higher than Yellow cab: Nashville, Pittsburgh, Sacramento, San Diego

Demand for Pittsburgh picks up in spring when it's still cold

Demand for `San Diego` and `Tucson` peaks in Fall when it's still hot

EDA
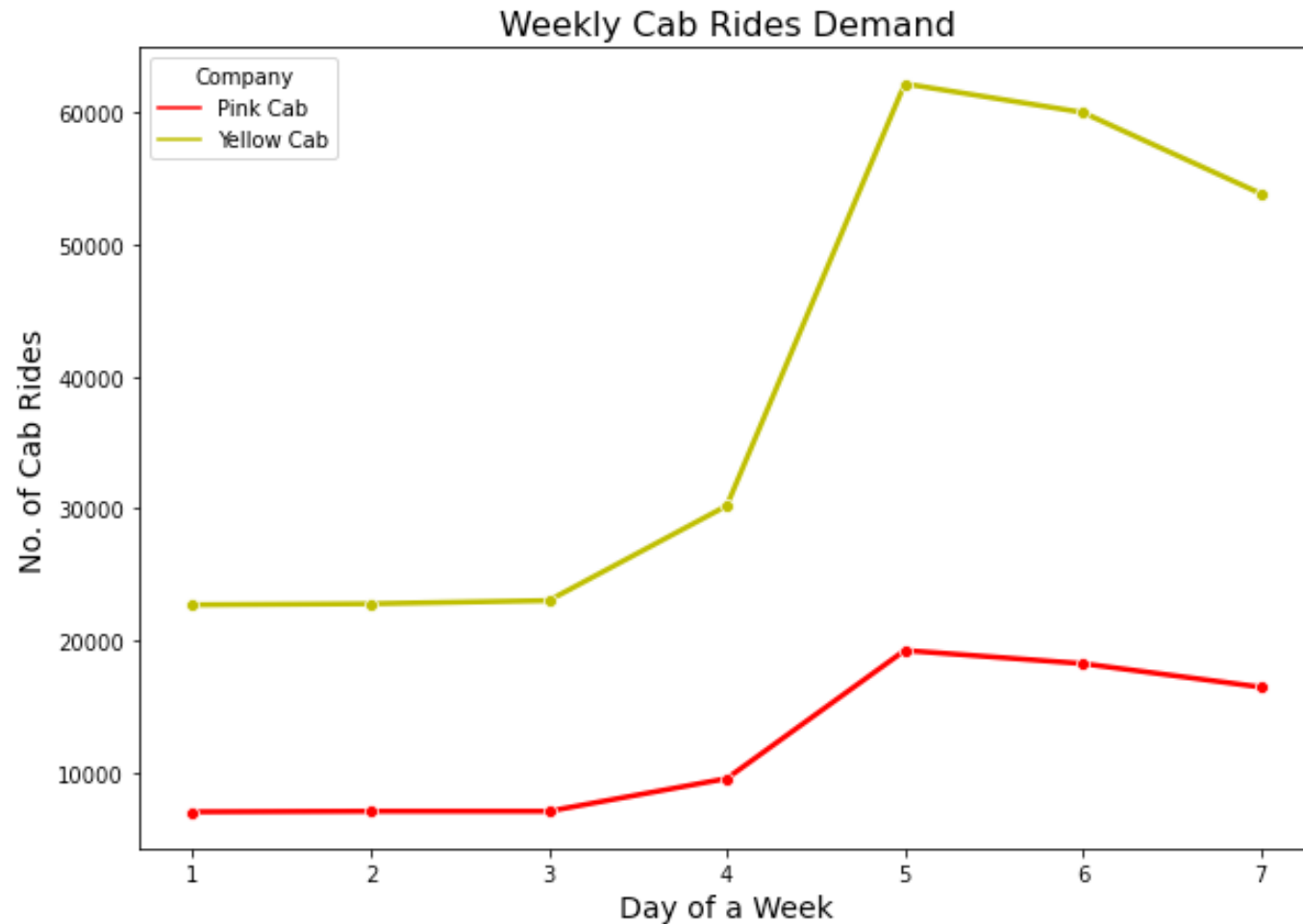
The overall demand trend for cab service picks up during the weekend

The demand trend for Yellow cab service increased higher than Pink cab; it also decreased faster than Pink cab on Sunday

Pink cab had a steady increase in demand from Thursday and drops slightly on Sunday

EDA
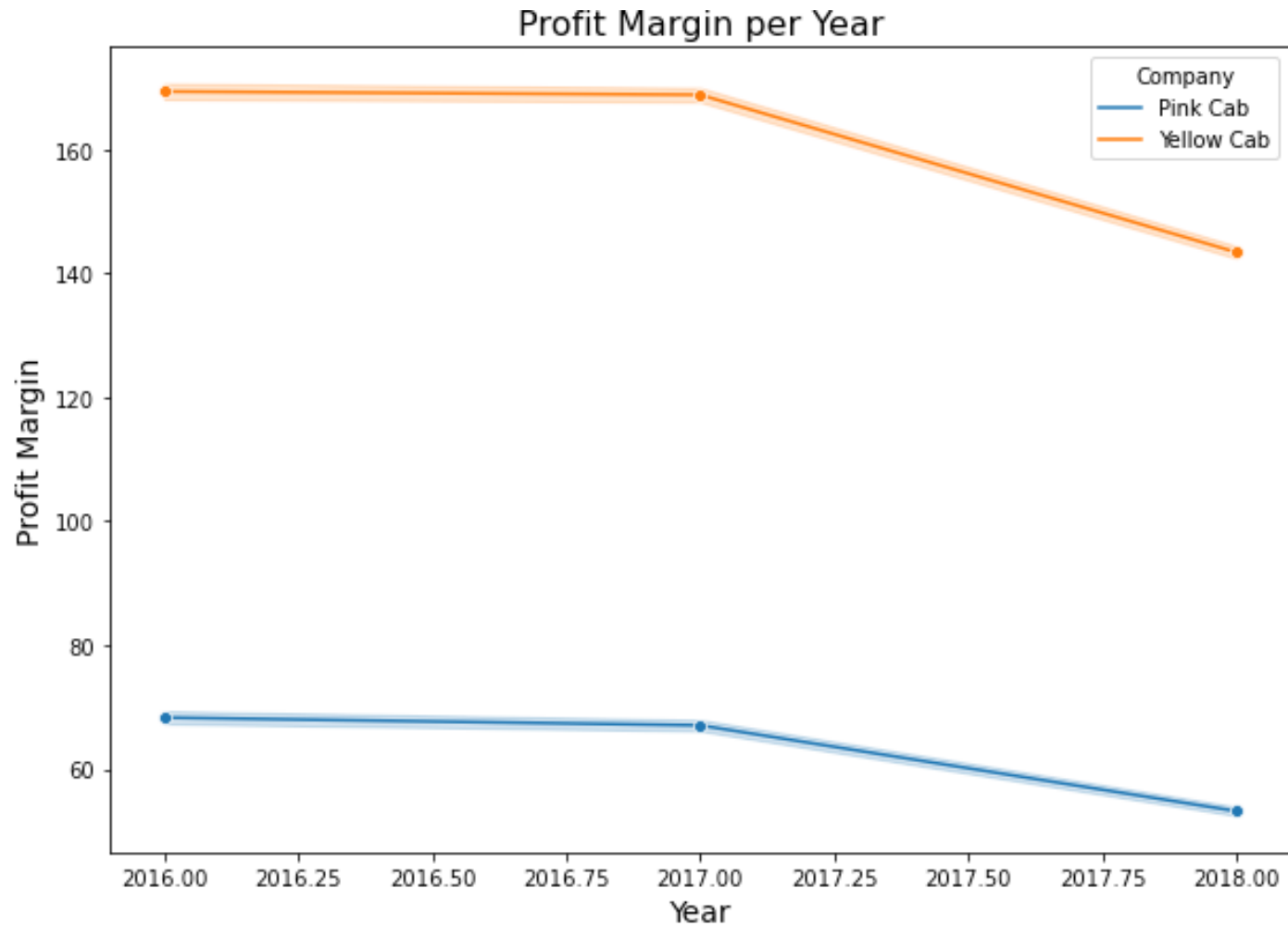
The overall profit for cab service decreases from 2017 to 2018

The profit margin for Yellow cab decreases faster than Pink cab; although its profit is still higher in general

Pink cab had a steady decrease in profit margin

EDA

Profit Margin per Month
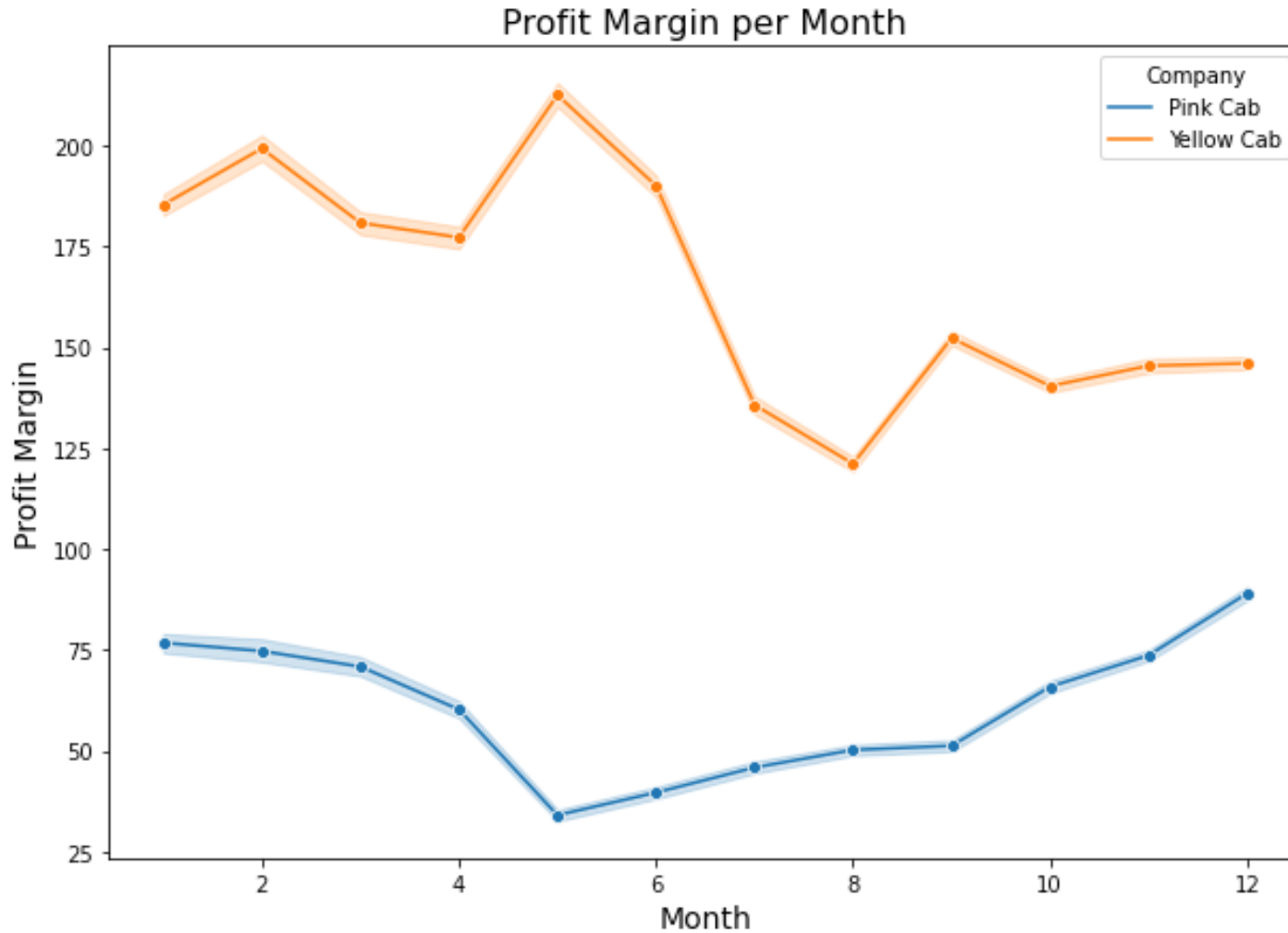
It seems customers takes cabs more often during winter times or when it's colder for both companies

The profit margin for Yellow cab increases during winter and spring and reaches its highest profit margin in May, then drops during summer and reaches its lowest in August

Pink cab had a steady increase in profit throughout Fall and Winter and drops slowly in Spring and reaches its lowest profit margin in May

It seems that it cost Pink cabs' more to run during spring

It seems that Yellow cabs' dominate customer groups are in cities have long summers and it cost more to run during the summer which comply with analysis in previous sections
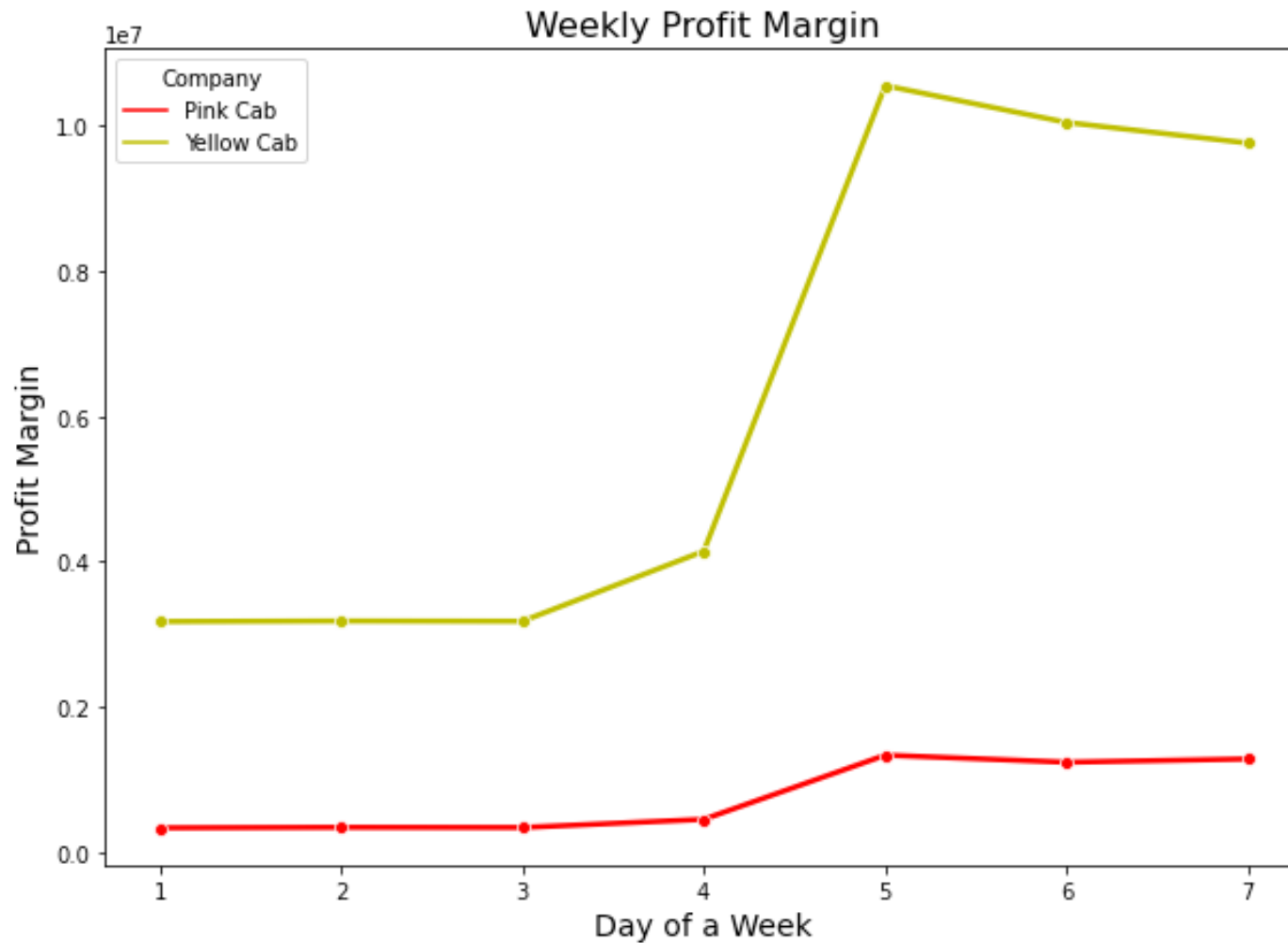
Data Glacier
Your Deep Learning Partner

EDA

The overall demand trend for cab service picks up during the weekend

The demand trend for Yellow cab service increased higher than Pink cab; it also decreased faster than Pink cab on Sunday

Pink cab had a steady increase in demand from Thursday and drops slightly on Sunday

EDA SUMMARY

If buy Yellow cab:

Pros:

Very established business
Longer operating history
Larger capacity
Broader geographical service area
Bigger customer base
Higher average profit margin

Cons:

Cost more to buy
Many old cars
Too many cars for regular maintenance to ensure service quality
High labor and management costs

EDA SUMMARY

If buy Pink cab:

Pros:

    Cost less to buy

    Younger company easier to reshape

    Newer equipment and lower maintenance costs

    Lower cost per KM to run

    Lots of room to improve its profit margin, capacity, location

Cons:

    Too many future plans to implement

    Taking risks to carry out plans

    Not as established

If buy Yellow cab:

Suggestions:

 Find other sources of costs that drive down profit margin

 Decrease number of old cars

 Target higher income age group since they travel further provides higher profit margin

 Decrease cab numbers during the summer since it cost more to run in summer

 Increase cab numbers in cities with increasing demand for both cabs: Atlanta, Austin, Denver, Los Angeles, Miami, Orange County, Phoenix, Silicon Valley, Tucson

 Increase cab numbers in cities with slower demand decrease: Denver, Chicago, DC

If buy Pink cab:

Suggestions:

 Find other sources of costs that drive down profit margin

 Emphasize it's cheaper to ride Pink or change marketing strategy to charge more

 Increase the charges to customers in all cities especially in NY, Silicon Valley, Dallas

 Increase the charges in long distance travel

 Increase number of cabs could be competitive as Pink cabs generally cost less per km travelled

 Target higher income age group since they travel farther

 Increase cab numbers in cities with increasing demand for both cabs: Atlanta, Austin, Denver, Los Angeles, Miami, Orange County, Phoenix, Silicon Valley, Tucson

 Increase cab numbers in cities with slower demand decrease: Denver, Chicago, DC

 Rising charging rate during the weekend, increase cab numbers during the weekend

Thank You