Data Glacier

Your Deep Learning Partner

Exploratory Data Analysis

G2M insight for Cab Investment firm

Archana Devi Ramesh

21st December 2022

Agenda

Problem Statement

Approach

EDA

Recommendations

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 and as per their Go-to-Market(G2M) strategy they want to understand the market before taking final decision.

- Objective: **using your actionable insights to help the firm identify the right cab to make their investment.**

Dataset

The list of datasets provided for the analysis are the following (**Time period of data is from 31/01/2016** to **31/12/2018.)**

Cab_Data.csv – this file includes details of transaction for 2 cab companies

Customer_ID.csv – this is a mapping table that contains a unique identifier which links the customer's demographic details

Transaction_ID.csv – this is a mapping table that contains transaction to customer mapping and payment mode

City.csv – this file contains list of US cities, their population and number of cab users

Approach

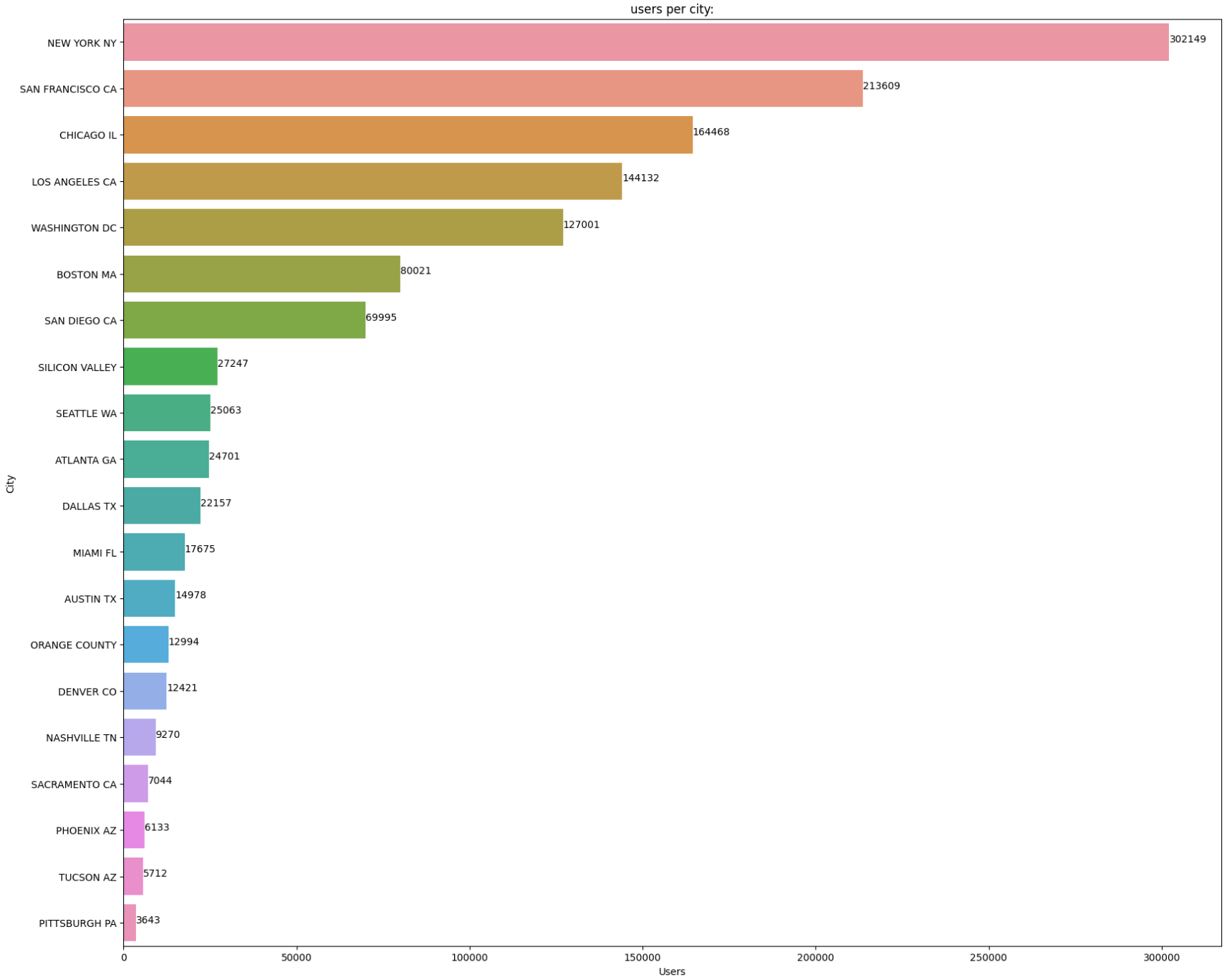
Performing data analysis of the following:

- Dataset understanding and visualization
- Understanding the most users per cab
- Analyzing the expensive and cheapest cab
- Computing the profitable company
- Performing basis hypothesis

EDA begins

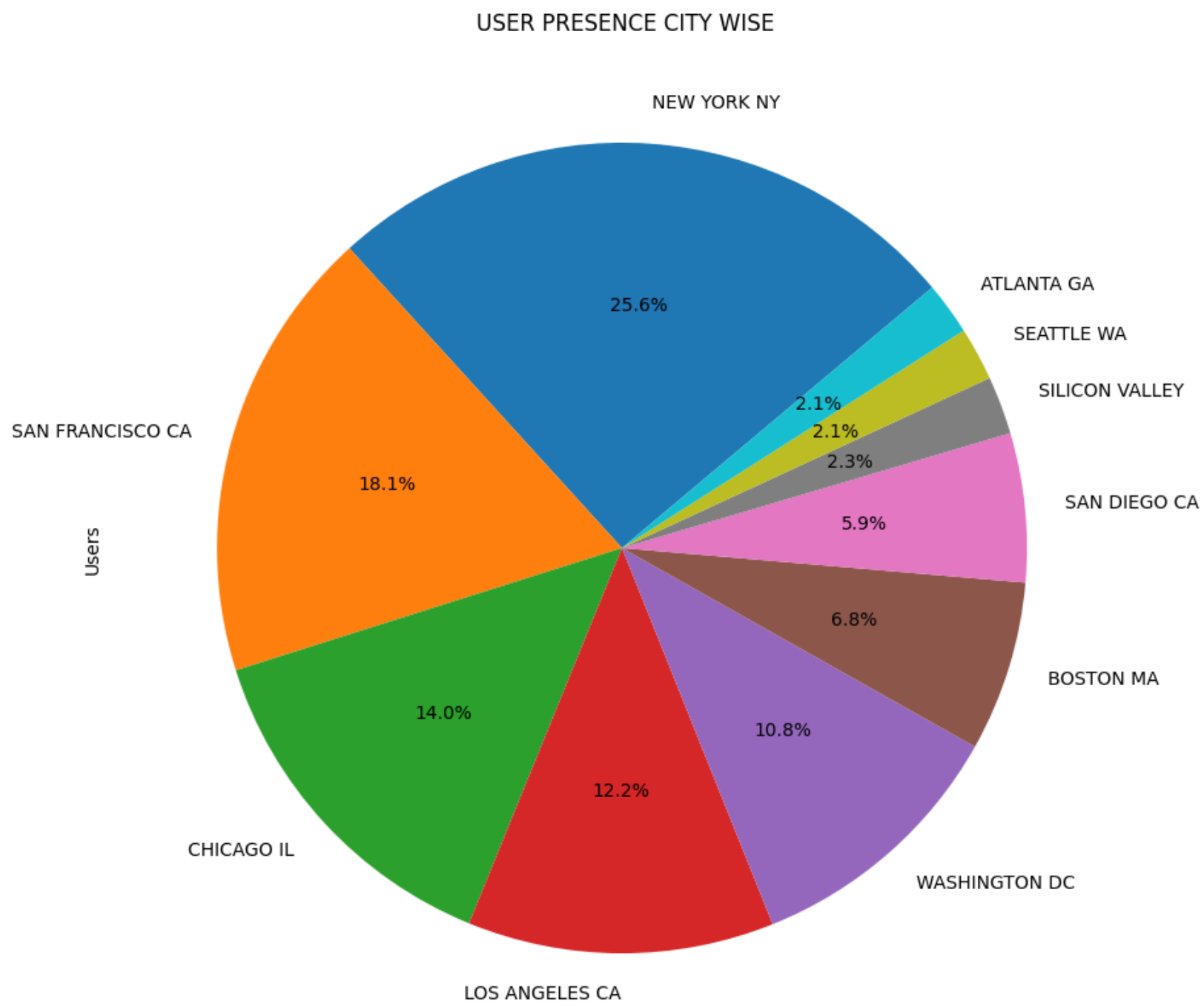
Number of users per city

It is evident from the bart chart that the population of users who used cab are more from New York followed by San Fransisco, Chicago and so on



Data Glacier
Your Deep Learning Partner

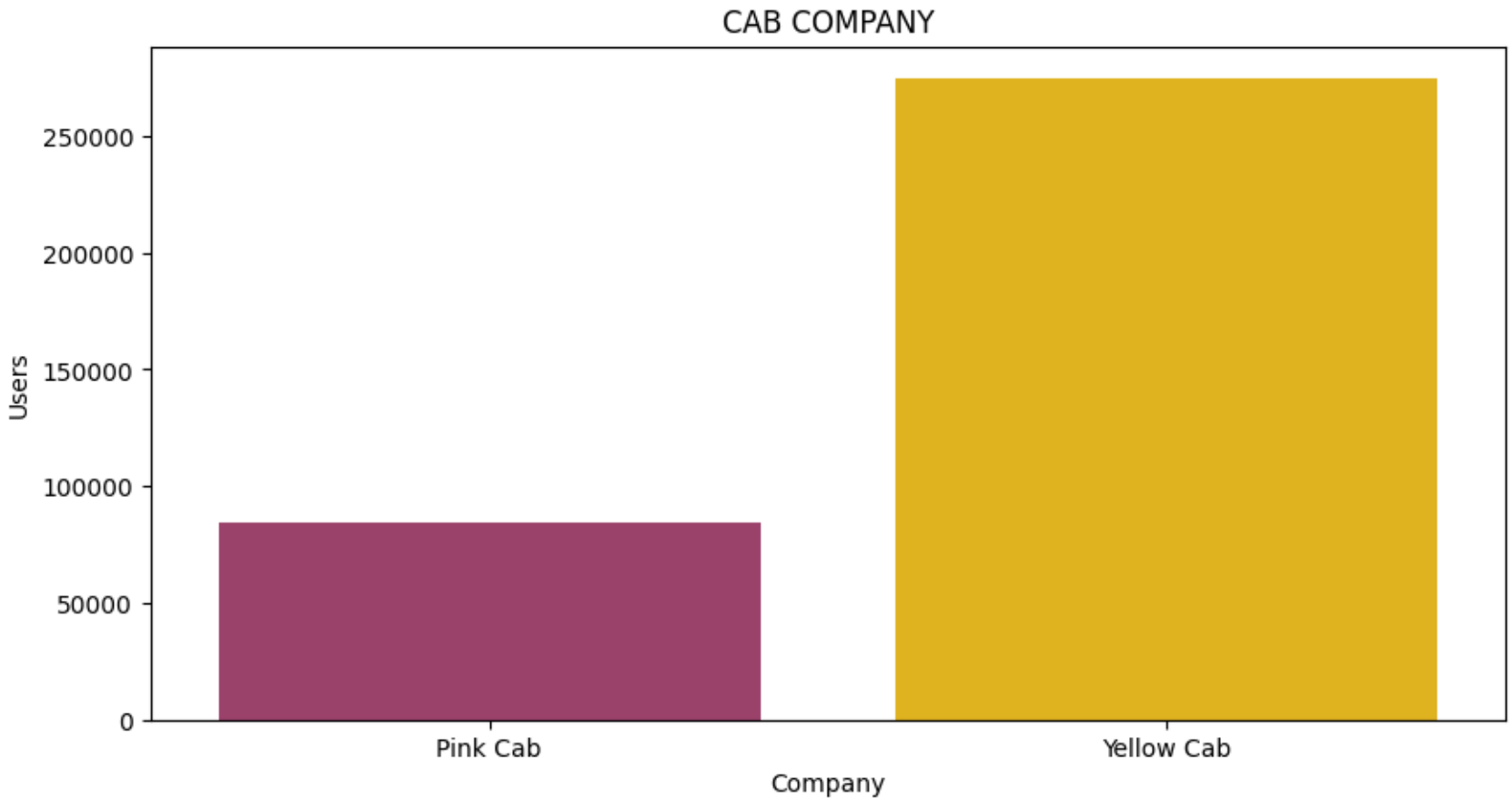
Visualizing the users from the top 10 cites



Data Glacier
Your Deep Learning Partner

Number of users of both companies

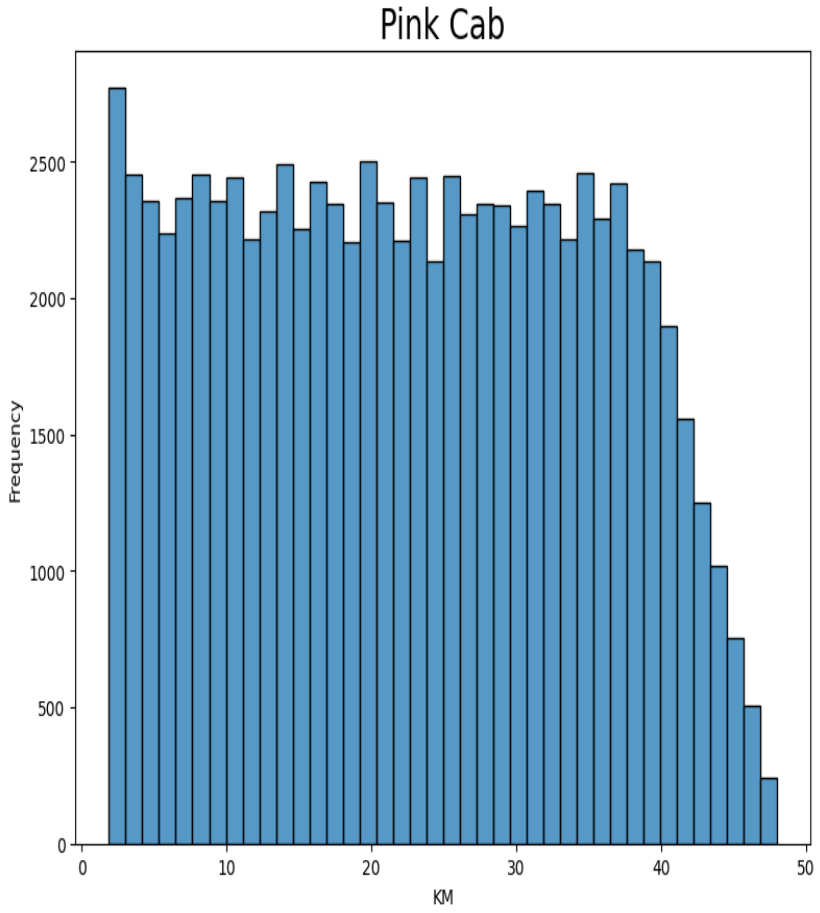
More users have used yellow cab compared to pink cab.

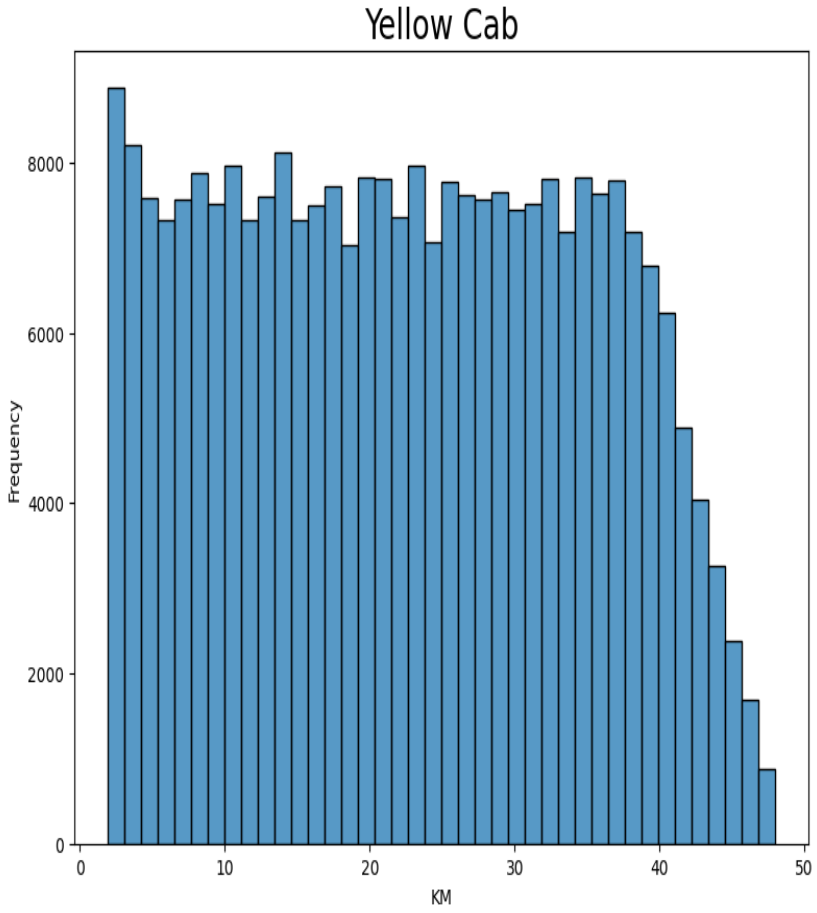


Data Glacier
Your Deep Learning Partner

Number of Kilometers travelled by both cabs

The frequency of KMs travelled are from 2 to 48 by both the cabs, however the frequency distribution is more for Yellow cab compared to pink

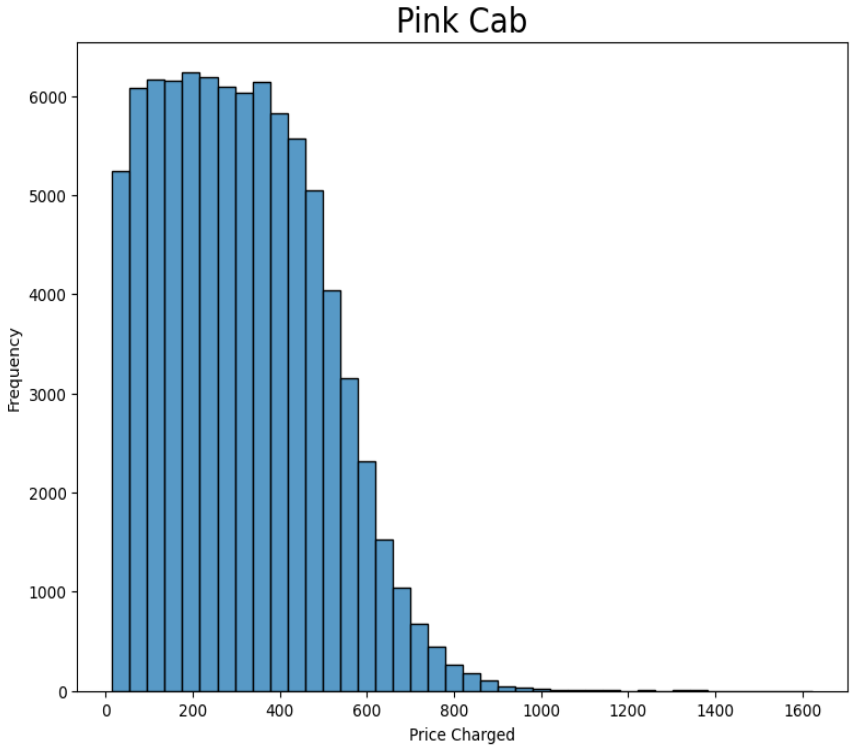


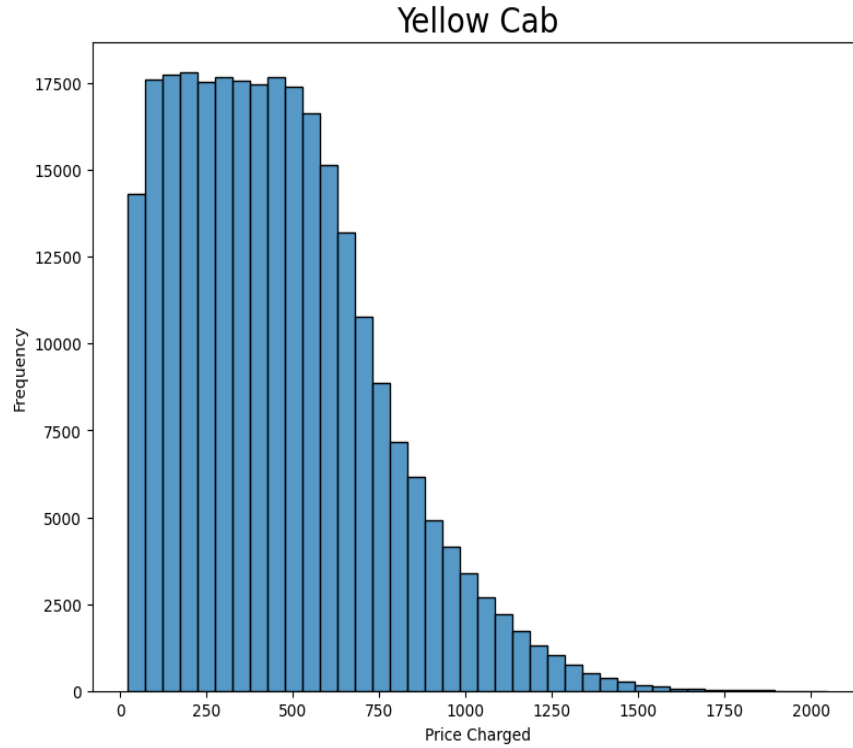


Data Glacier
Your Deep Learning Partner

Comparison on the Price charged by the cabs

The maximum price charged by pink cab is around 1000 dollars, however yellow cab seems to be expensive with the highest charge around 1500 dollars

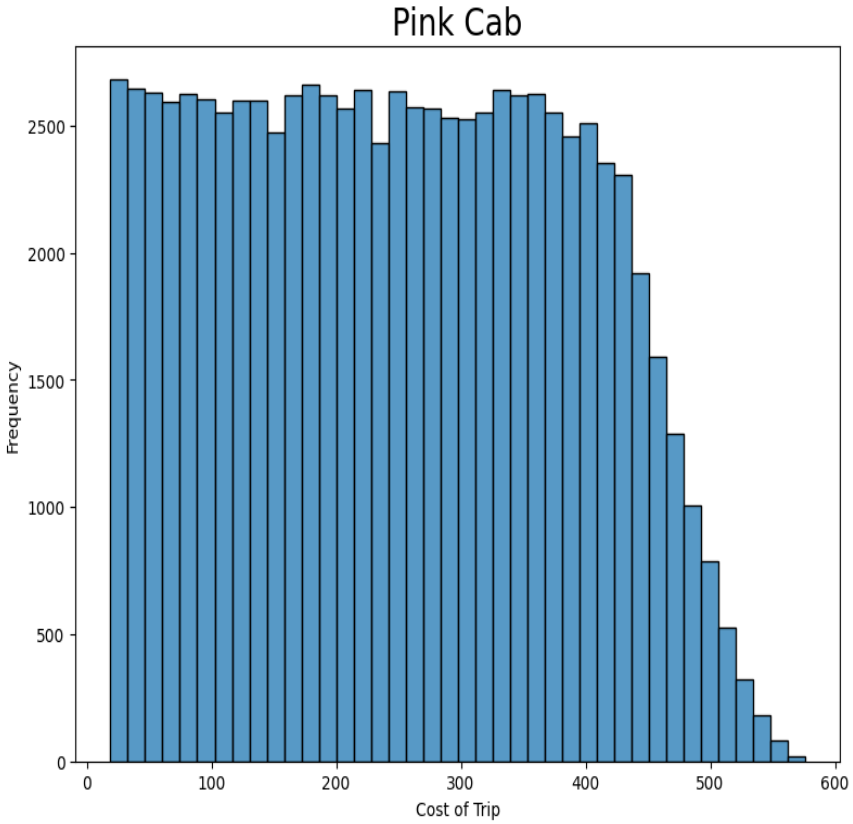


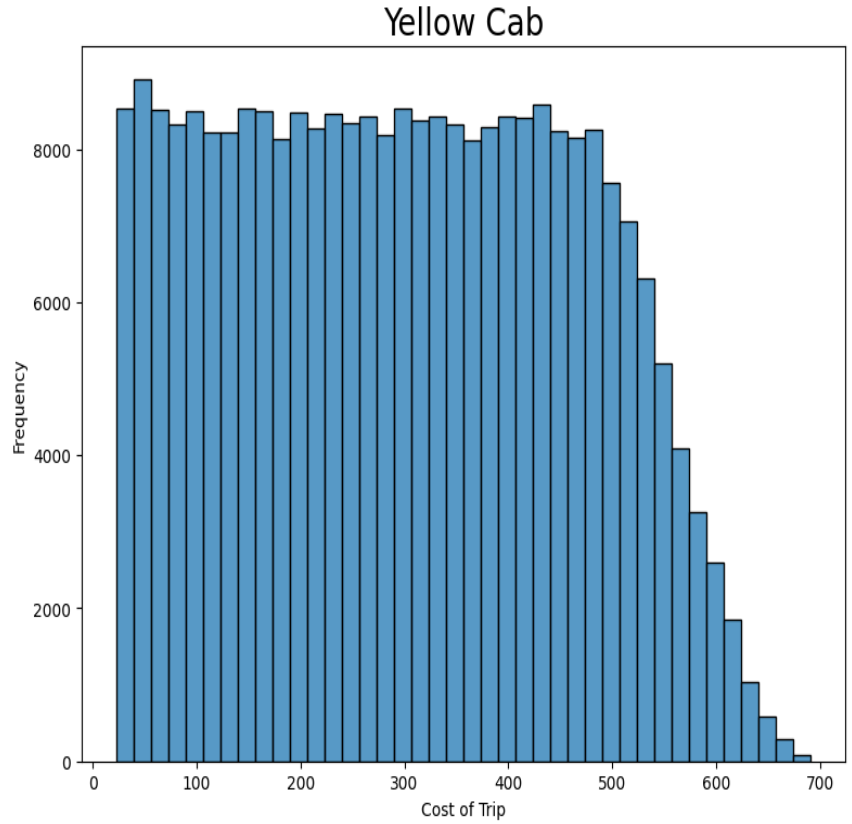


Data Glacier
Your Deep Learning Partner

Cost of trip comparison of the cabs

The maximum price of the cost of trip is within 600 for pink cab , where as for yellow cab it’s high which is within 700

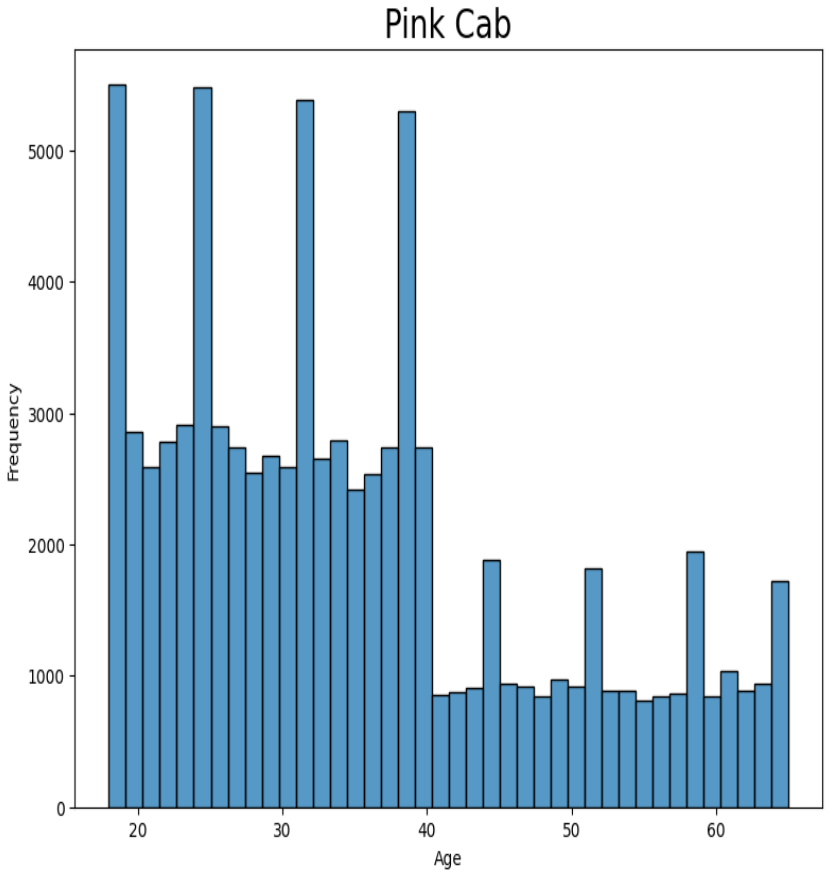


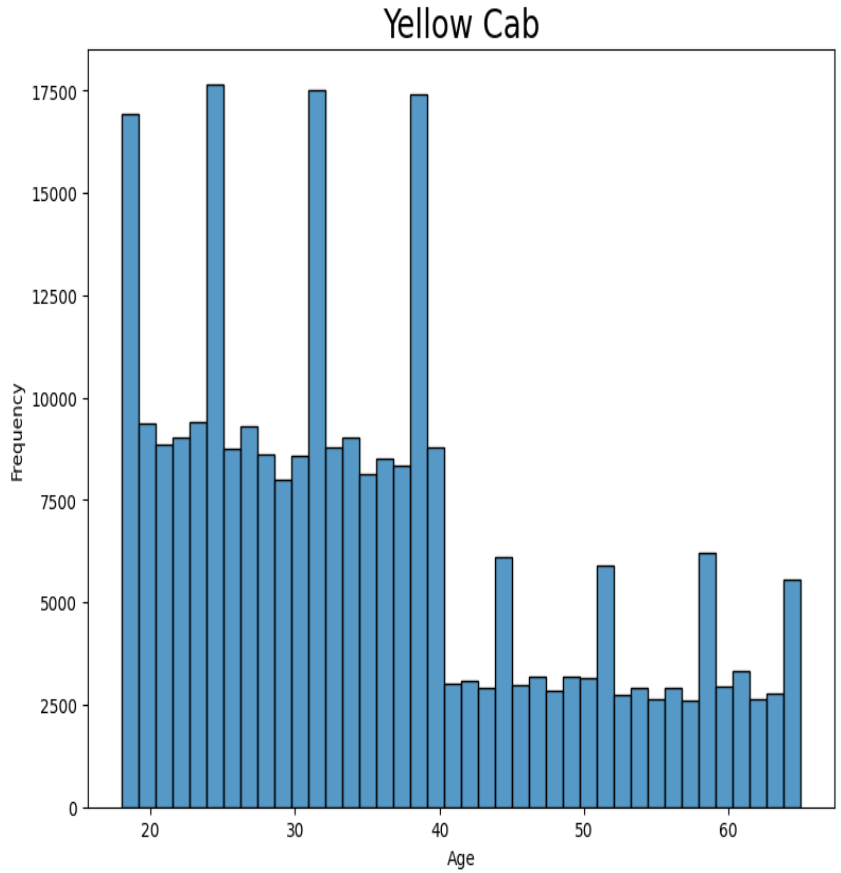


Data Glacier
Your Deep Learning Partner

Age range of users who have travelled in both the cabs

Both cabs have a similar age range of travellers with mostly in the age group of 20 to 40

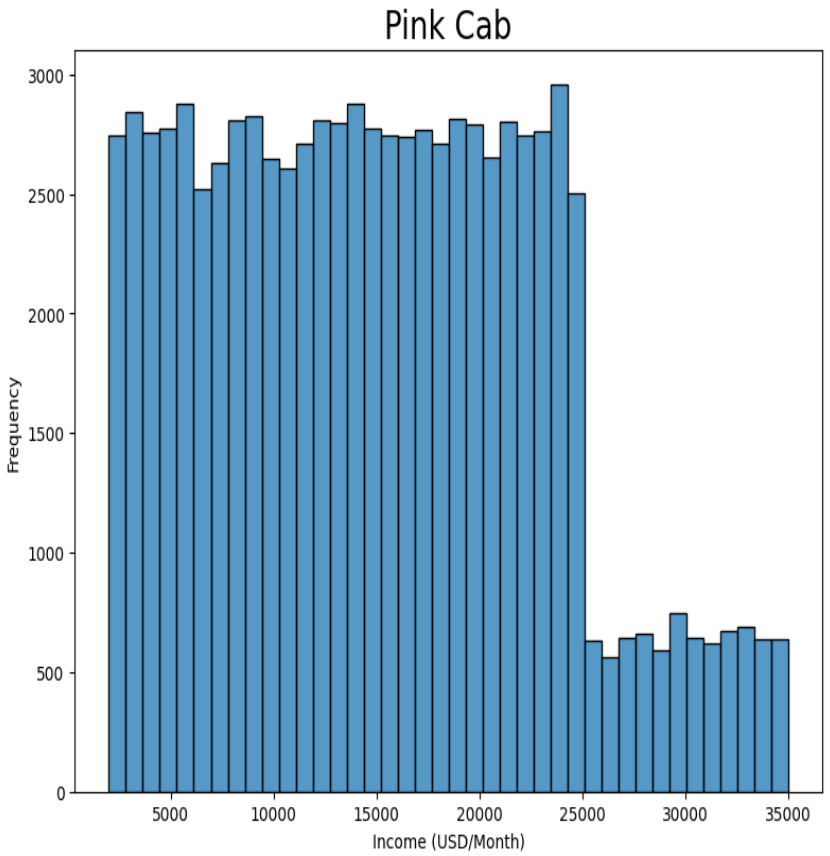


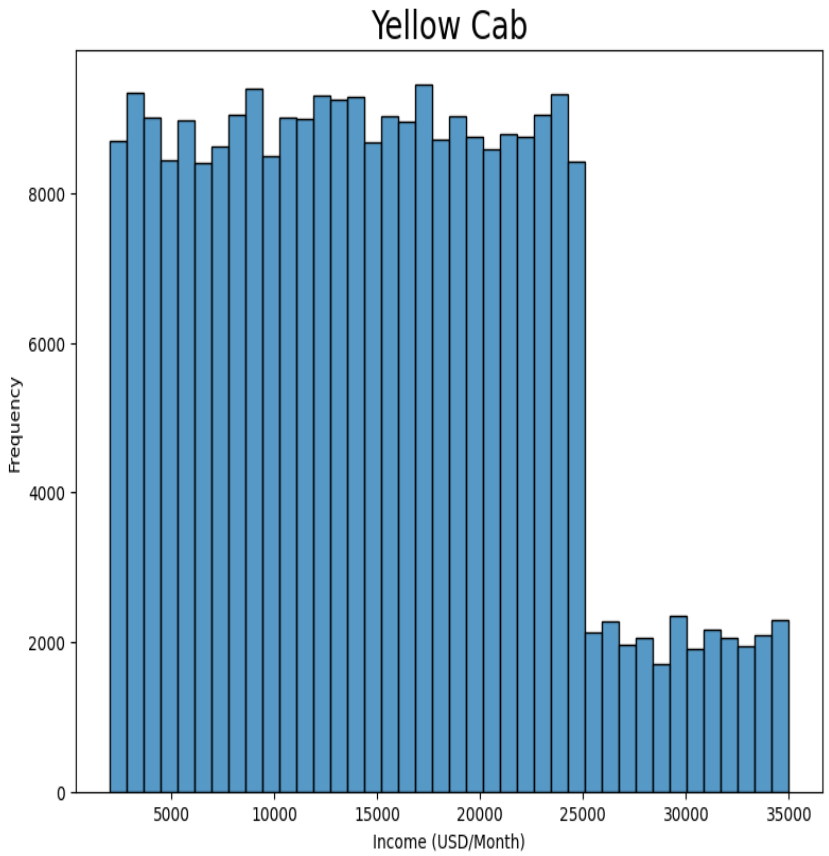


Data Glacier
Your Deep Learning Partner

Income of travellers of both the cabs

Income range of travellers of both the cabs are in the same range

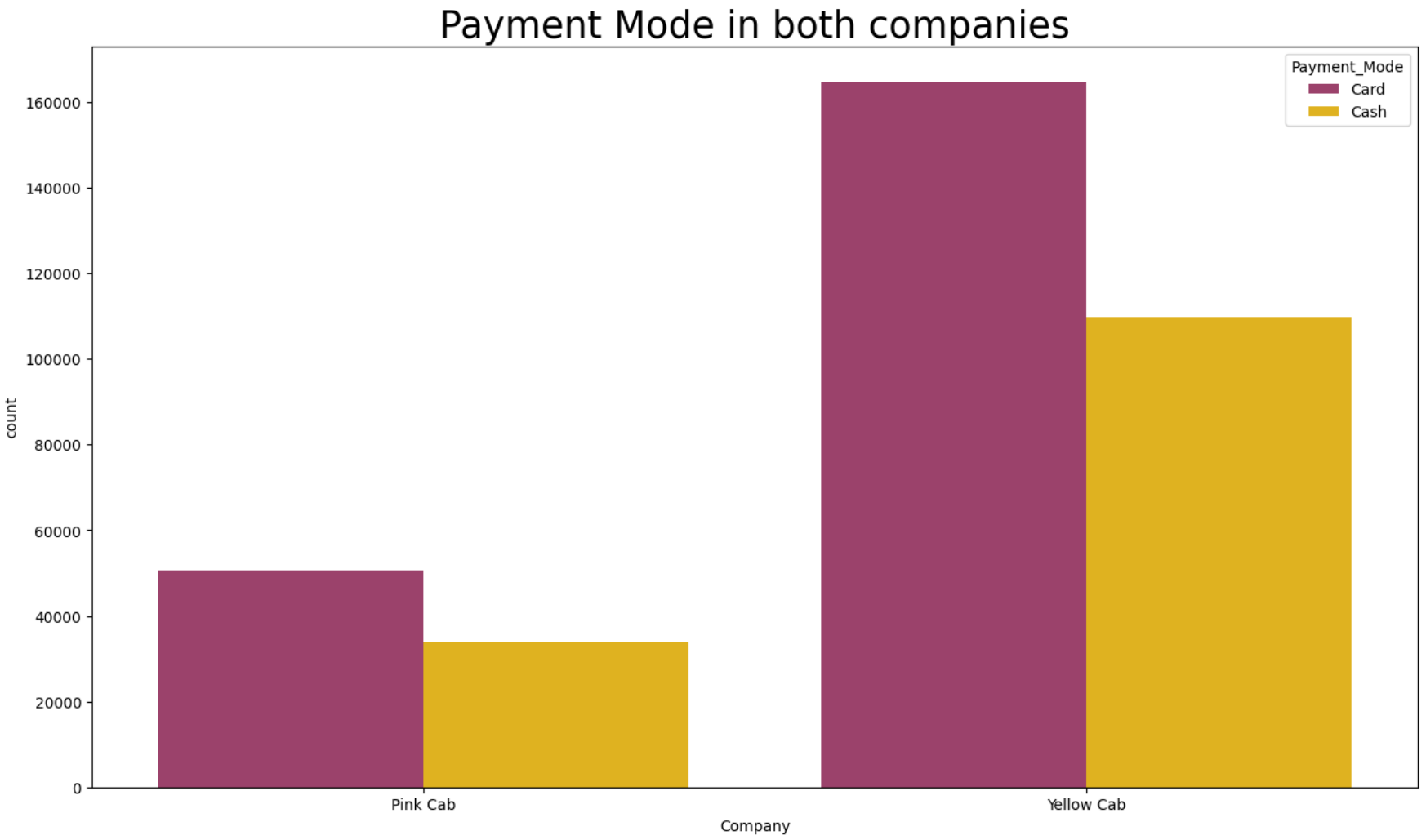




Data Glacier
Your Deep Learning Partner

Mode of payment done by users in both cabs

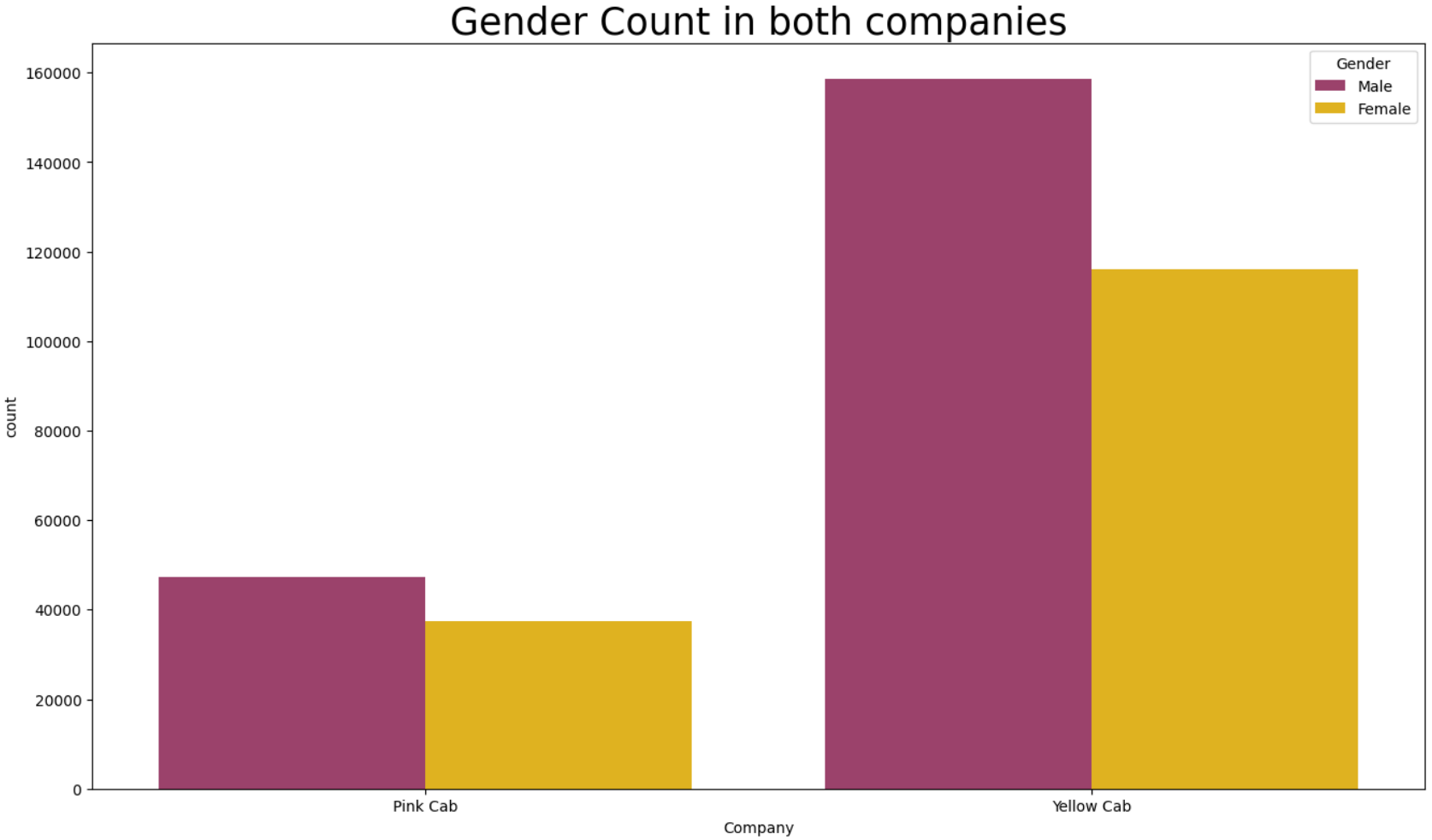
In both the cabs, people have made the payment mostly by card compared to cash



Data Glacier
Your Deep Learning Partner

Gender distribution of cab travellers

In both the cabs, percentage of male travellers are more compared to female.



Data Glacier
Your Deep Learning Partner

Comparison of price charged w.r.t the distance travelled

The scatter plot shows a linear relationship for between the price charged and the km travelled for both the companies

Feature correlation matrix using heatmap

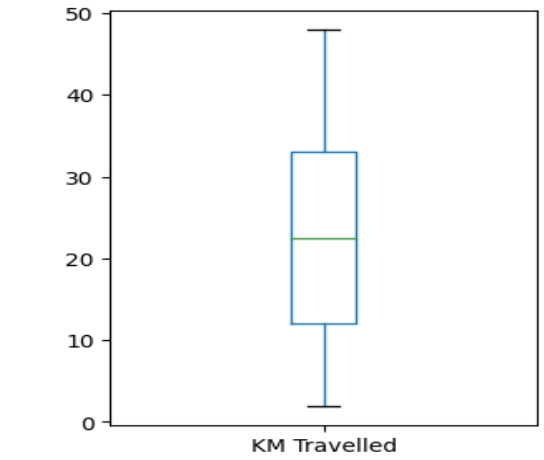
KM Travelled has high correlation with cost of trip and Price Charged

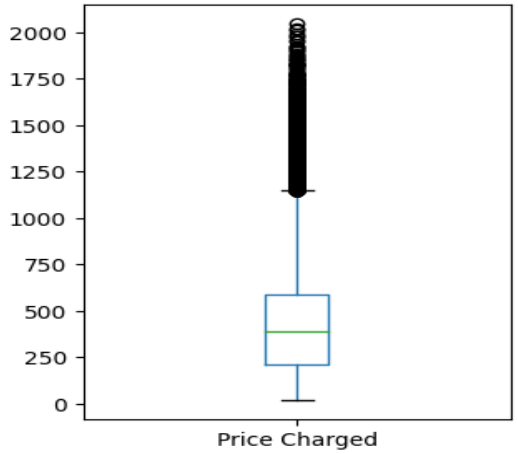
Population and Users are also highly correlated

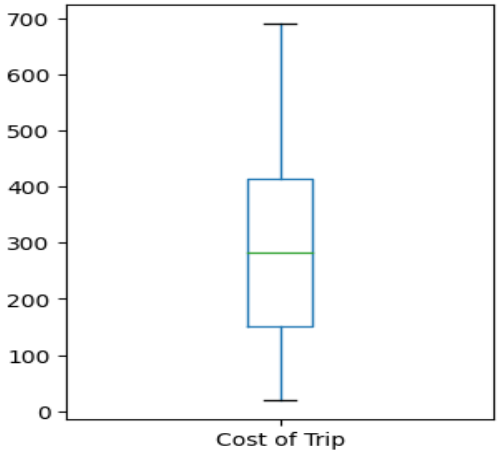
Data Glacier
Your Deep Learning Partner

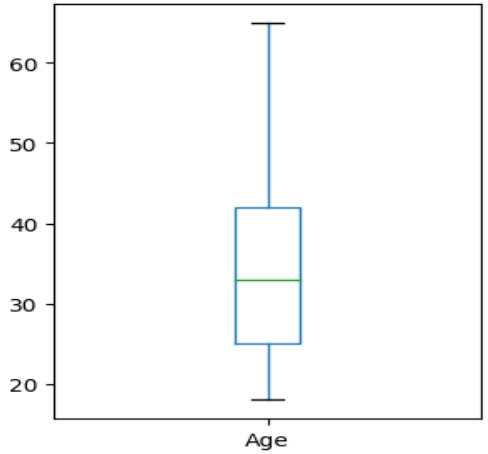
Outlier detection using Boxplots

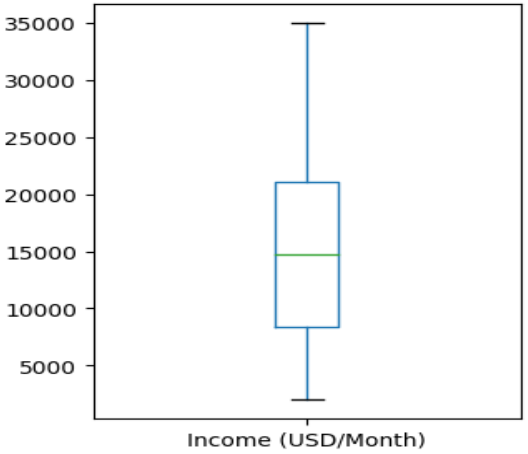
Of all the features, Price Charged has some outliers. Since trip duration details are not available, we are not treating this as outlier.

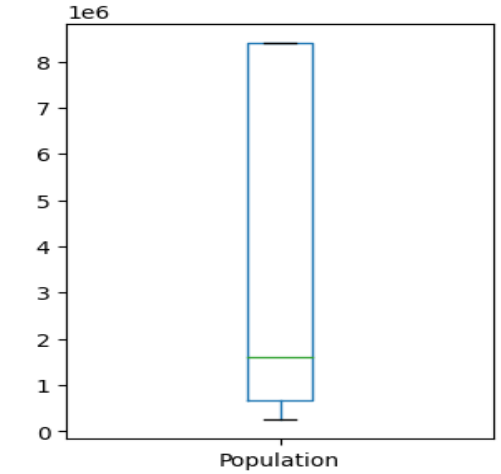


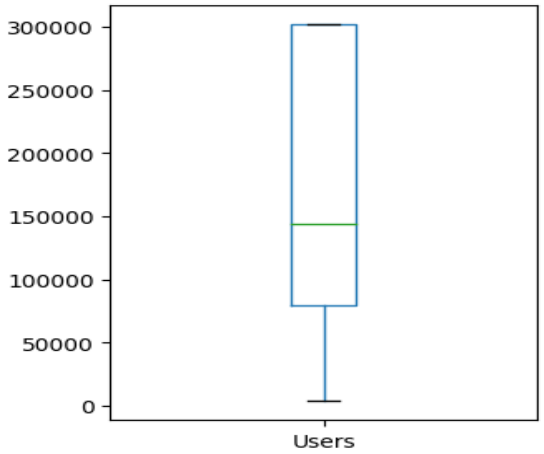








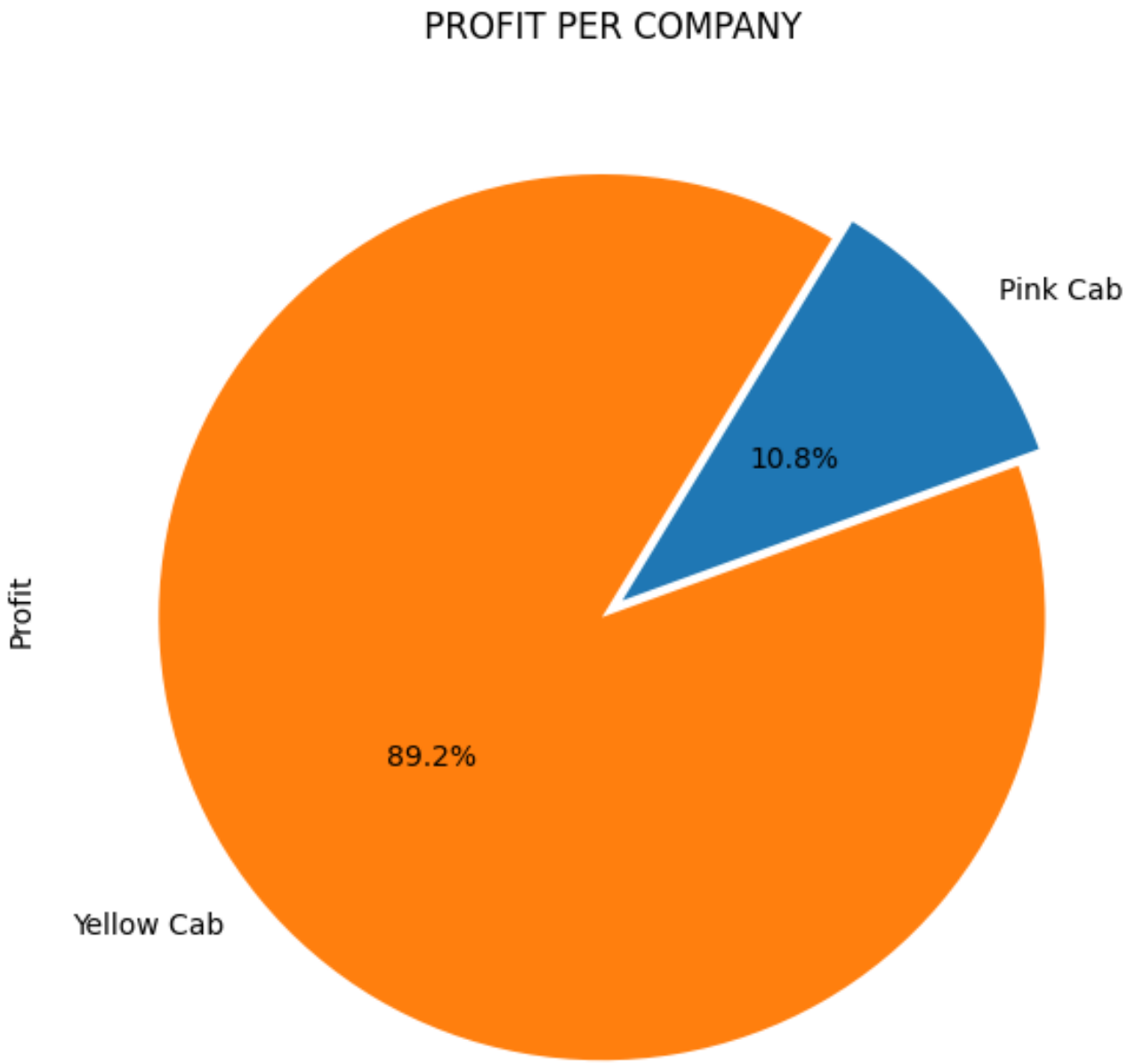




Data Glacier
Your Deep Learning Partner

Profit comparison of both the cabs

Yellow cab has made the highest profit of 89.2% compared to Pink cab

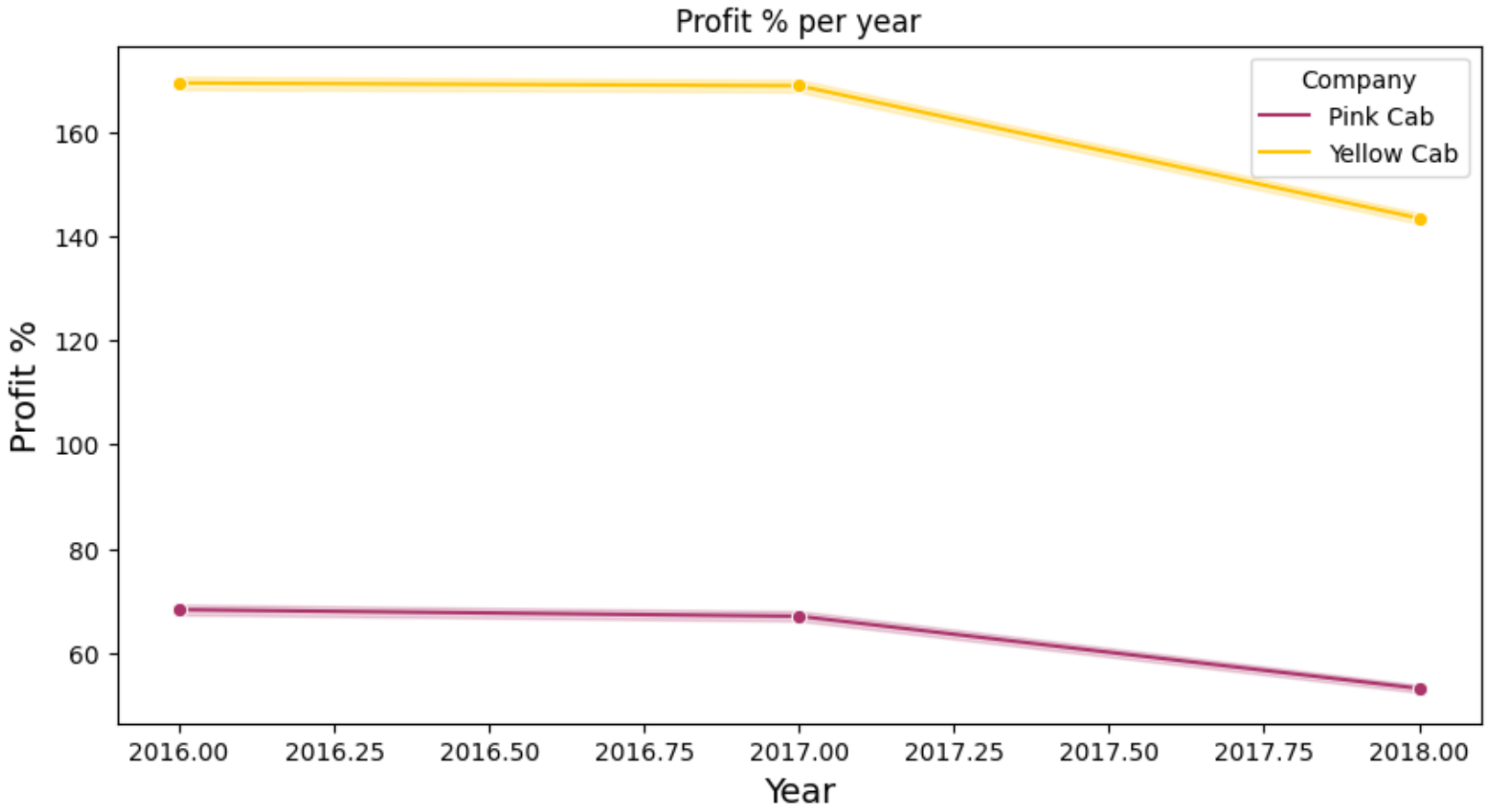


Data Glacier
Your Deep Learning Partner

Profit % per year for the years 2016 to 2018

Both companies have made better profit in 2016 compared to 2018

Compared to YELLOW cab, PINK cab has made better profit in the year 2018



Data Glacier
Your Deep Learning Partner

Profit % per month

YELLOW cab seems to have made profit during the mid of a year

PINK cab has made more profit during the year beginning and end

Data Glacier
Your Deep Learning Partner

Recommendation on Cab Investment

The following observations are made based on the over all analysis

1. Yellow cab owns 89% of the total profit made by both companies

2. Most Users prefer travelling with Yellow cab than Pink cab

3. Yellow cab charges higher than Pink cab

Therefore, I advice the XYZ company to invest in Yellow Cab company.

Data Glacier
Your Deep Learning Partner

Thank You