



**Data Glacier**

Your Deep Learning Partner

# Exploratory Data Analysis

## <G2M Insight for Cab Investment Firm>

<6/20/2022>

# Agenda

Executive Summary

Problem Statement

Approach

EDA

EDA Summary

Recommendations



**Data Glacier**

Your Deep Learning Partner

# Executive Summary

Purpose: Find which company is performing better and is a better investment opportunity for XYZ?



**Data Glacier**

Your Deep Learning Partner

# Questions

Which company has more users?

Where do those users mainly reside?

How much money(profit) does each company earn?

Does each company have the same rate for expenses?

What is the demographic of the customers?



**Data Glacier**

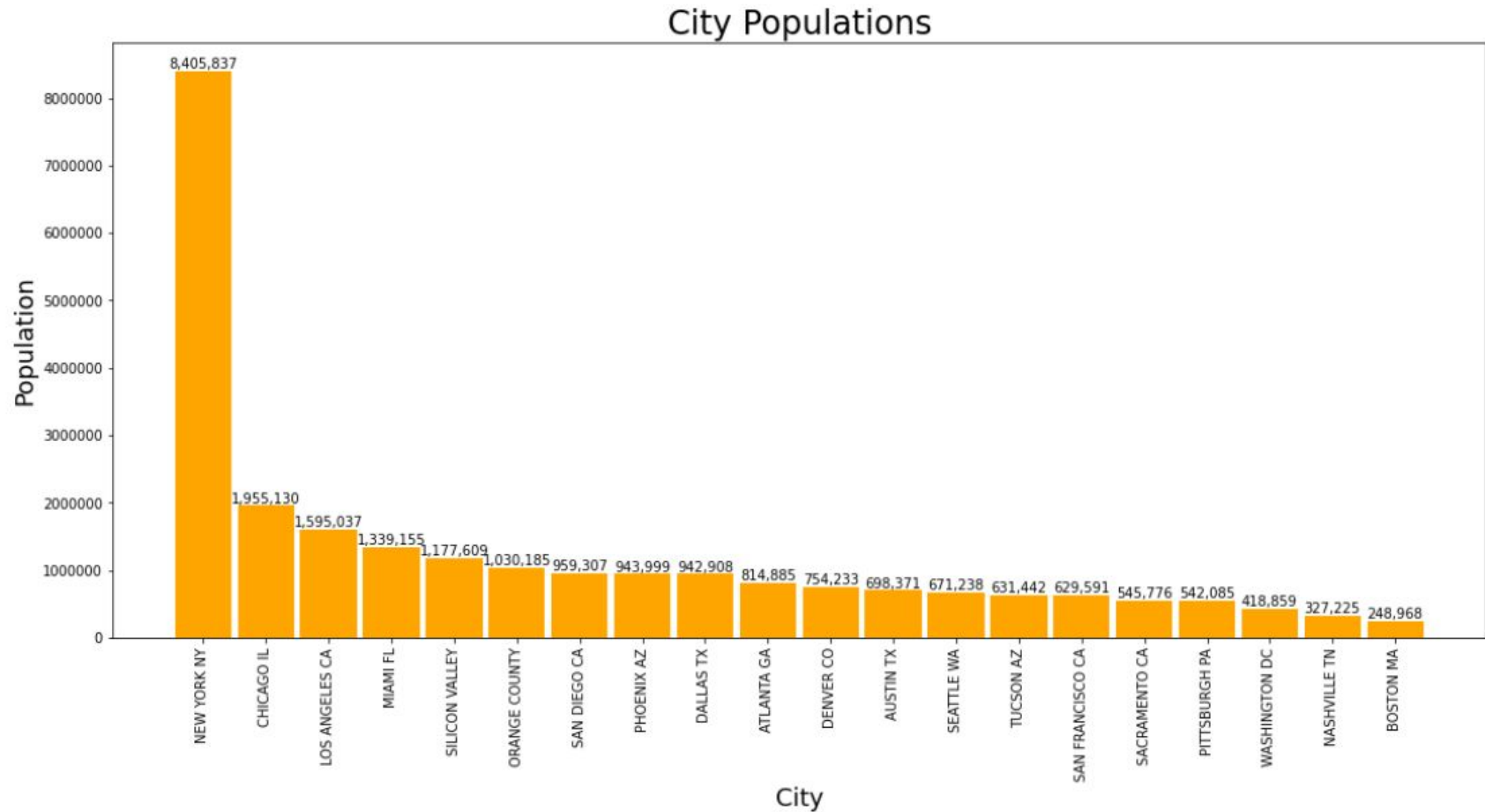
Your Deep Learning Partner

# Master Dataset Info

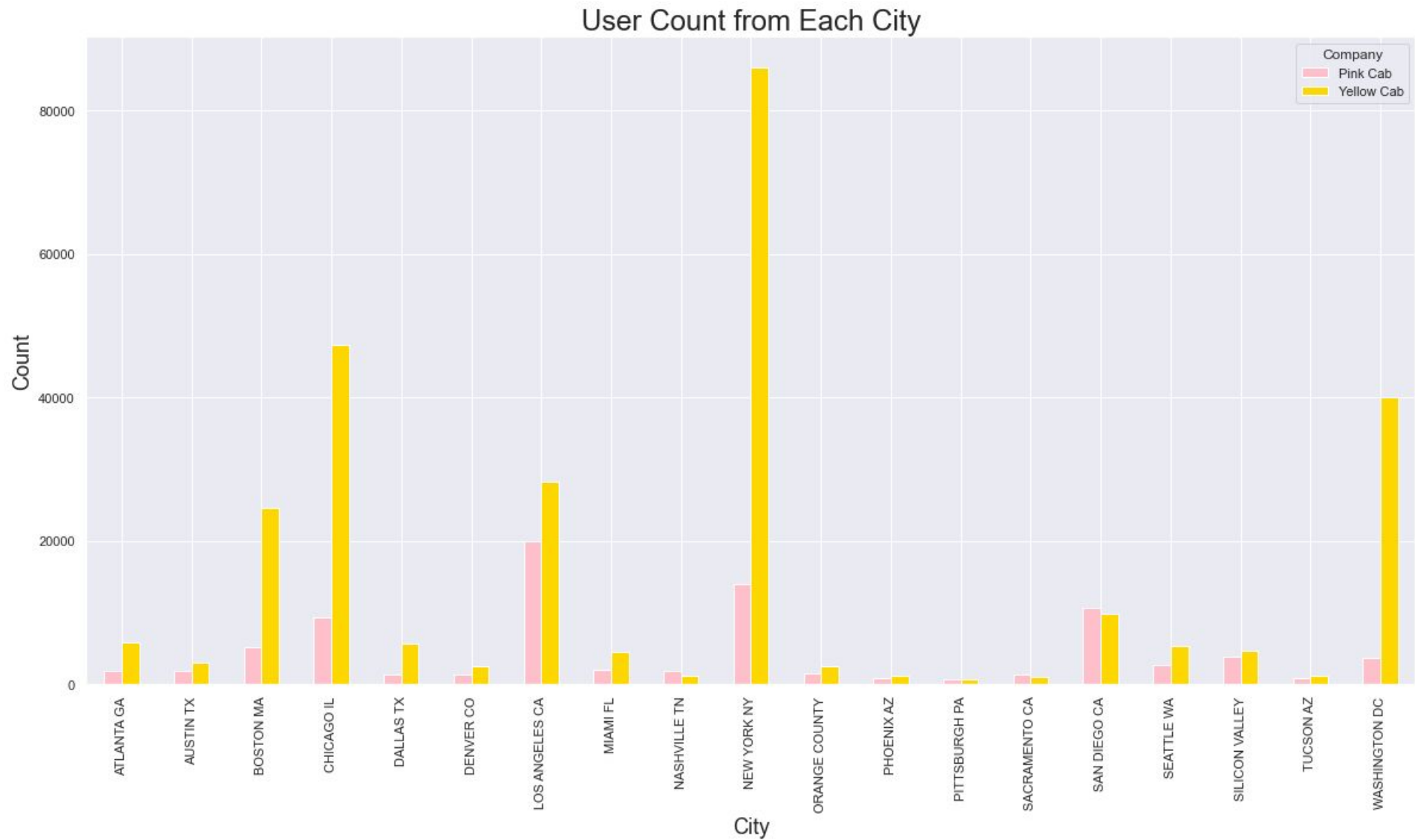
<b>Total number of observations</b>	359392
<b>Total number of files</b>	4
<b>Total number of features</b>	14
<b>Base format of the file</b>	csv
<b>Size of the data</b>	42,293 KB

```
master.head()
```

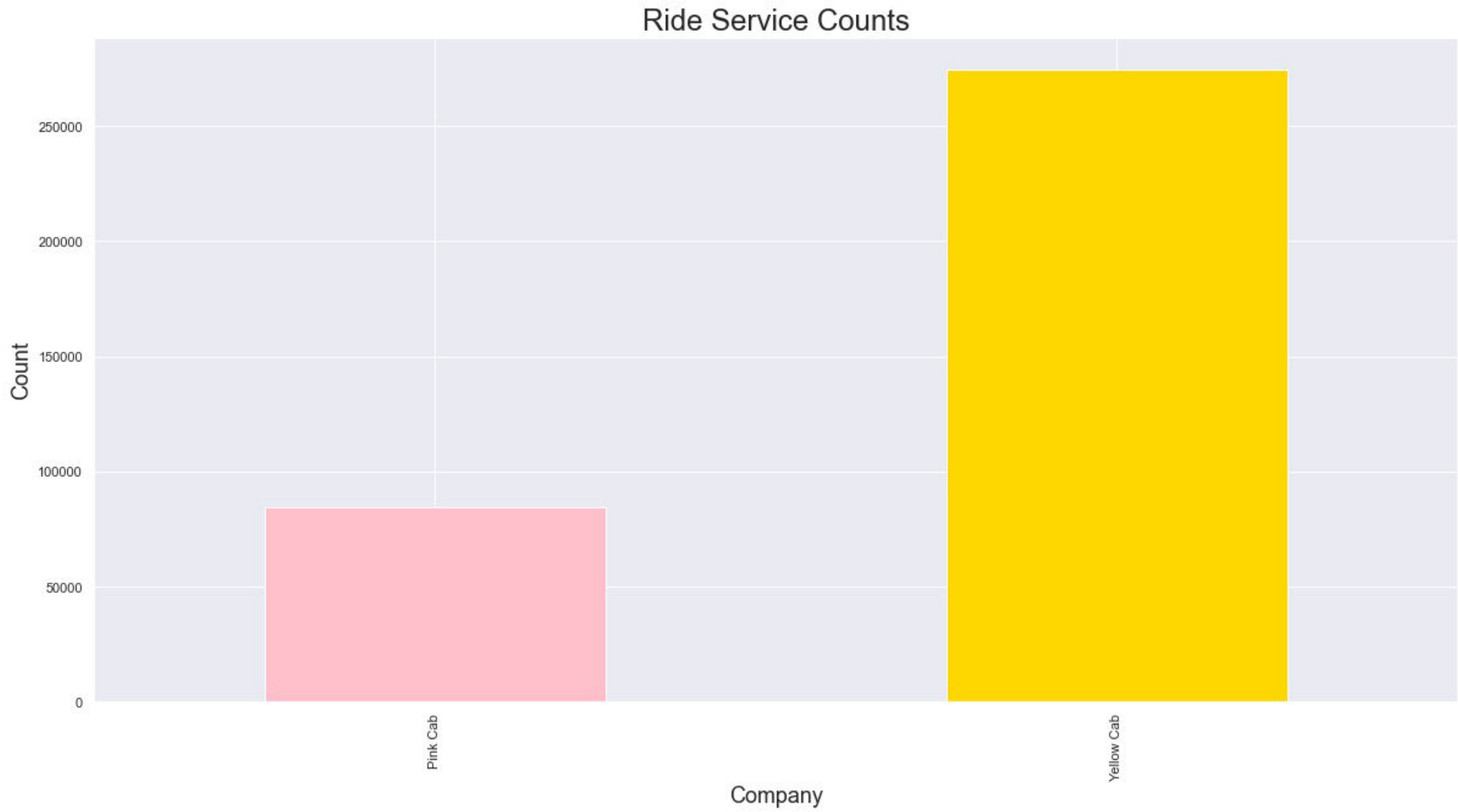
	Transaction ID	Date of Travel	Company	City	KM Travelled	Price Charged	Cost of Trip	Customer ID	Payment_Mode	Population	Users	Gender	Age	Income (USD/Month)
0	10000011	08-01-2016	Pink Cab	ATLANTA GA	30.45	370.95	313.6350	29290	Card	814,885	24,701	Male	28	10813
1	10351127	21-07-	Yellow	ATLANTA	26.19	598.70	317.4228	29290	Cash	814.885	24.701	Male	28	10813



The most populated cities in order is New York, Chicago, Los Angeles, Miami, and Silicon Valley.

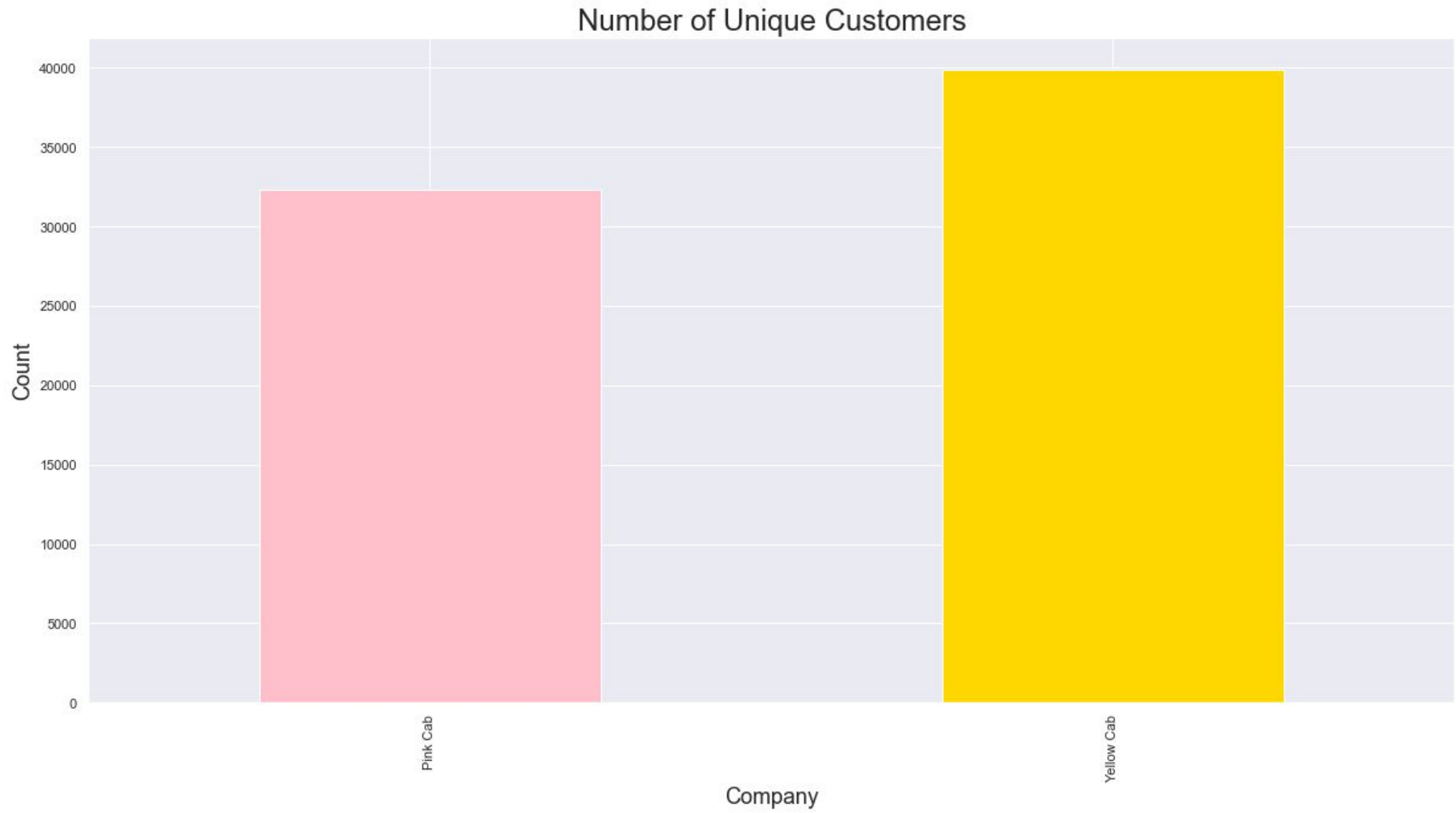


Yellow Cab has more users than Pink Cab in the top 5 most populated cities of the ones included.

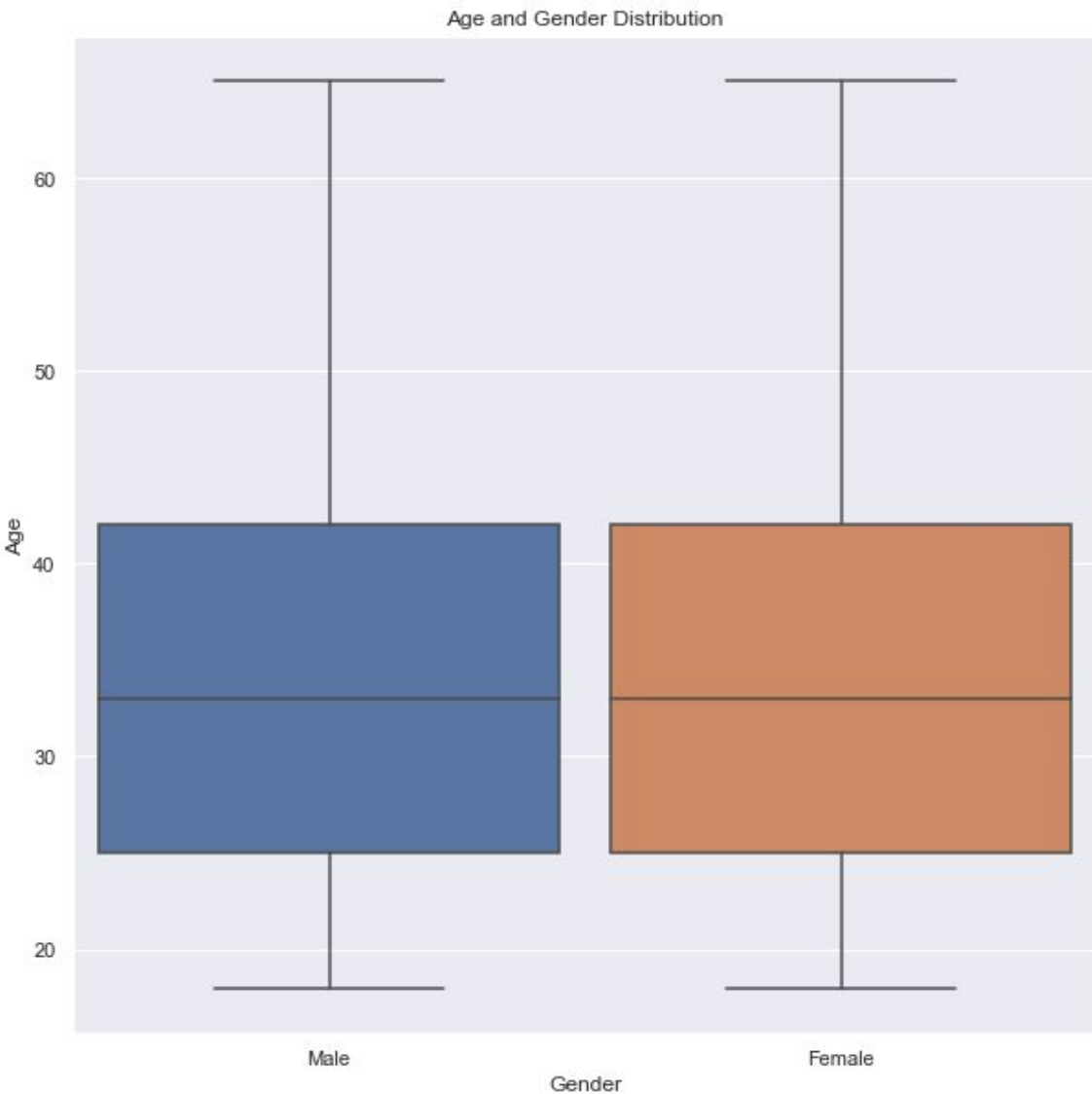


Yellow Cab has provided more rides than Pink Cab.



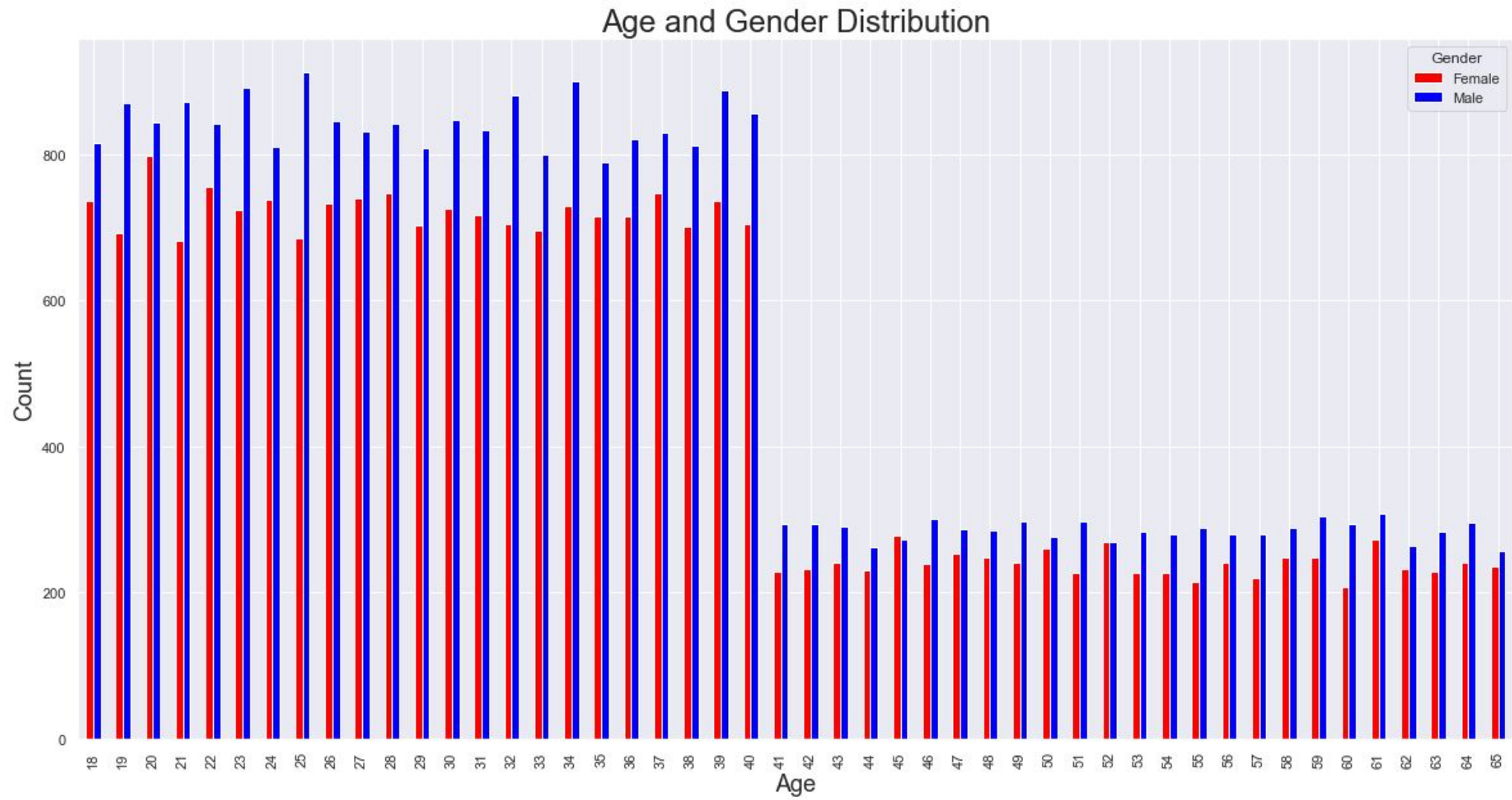


Yellow Cab has more unique customers.

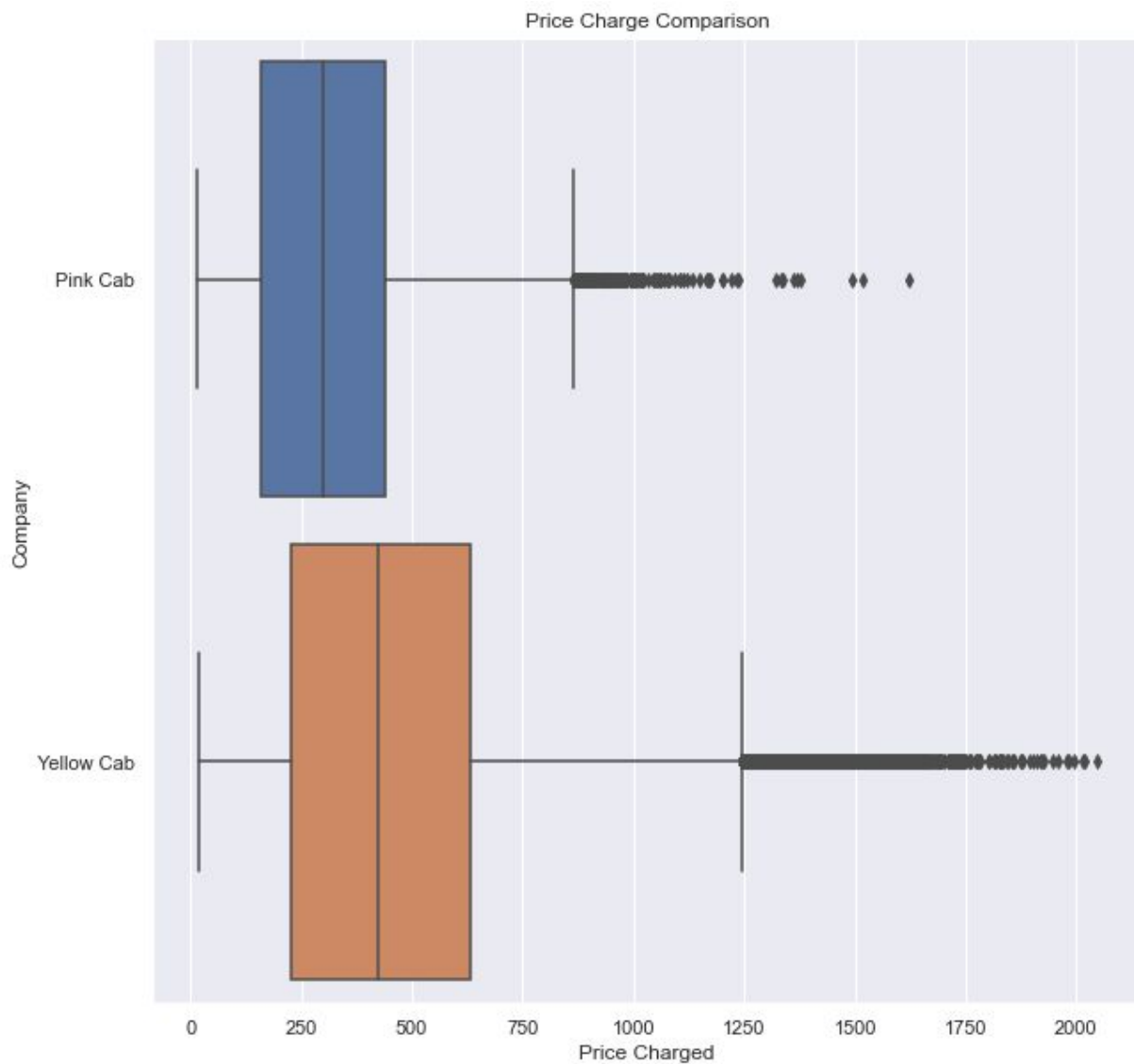
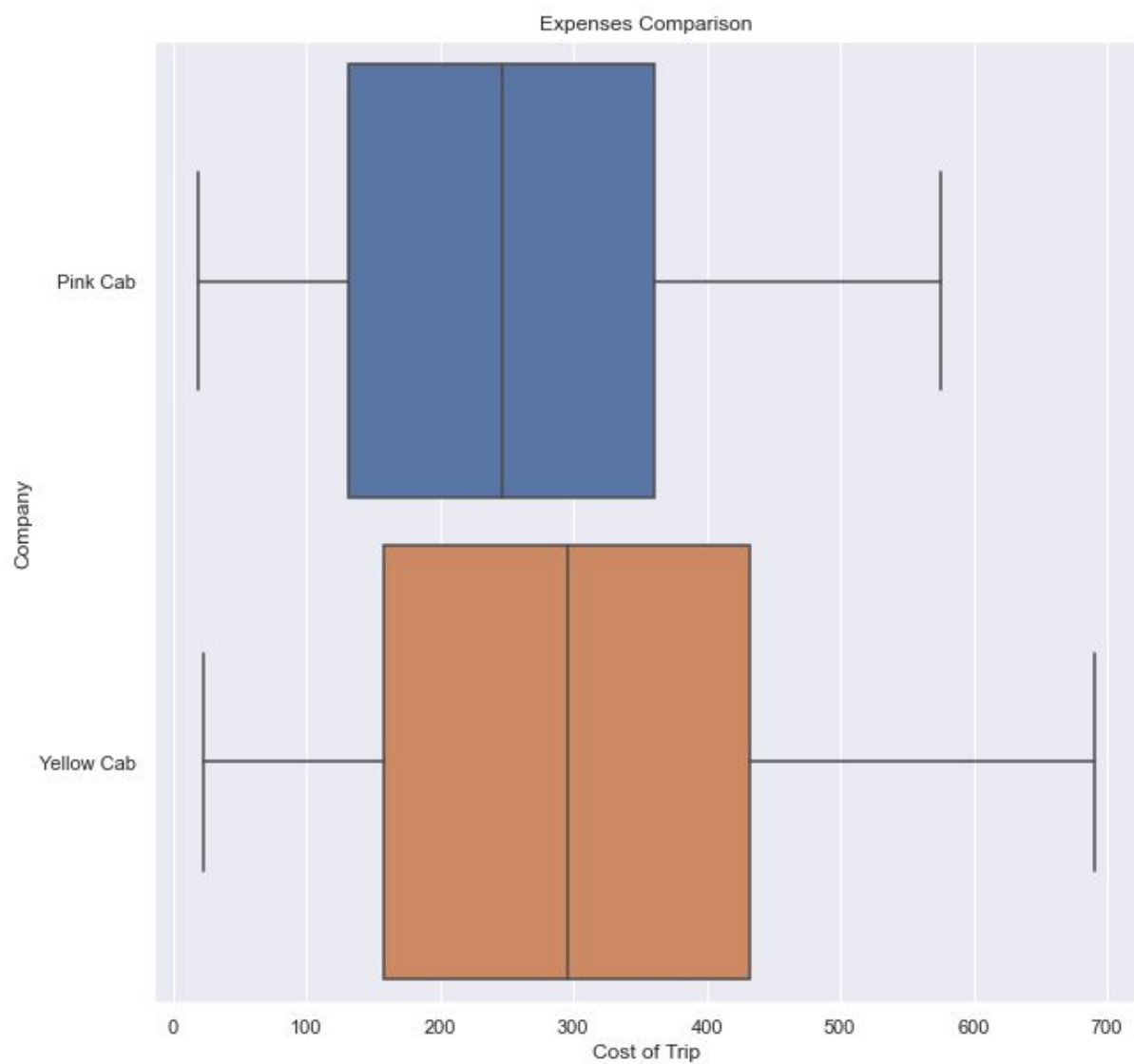


	count	mean	std	min	25%	50%	75%	max
Gender								
Female	153480.0	35.287608	12.620318	18.0	25.0	33.0	42.0	65.0
Male	205912.0	35.373300	12.574664	18.0	25.0	33.0	42.0	65.0

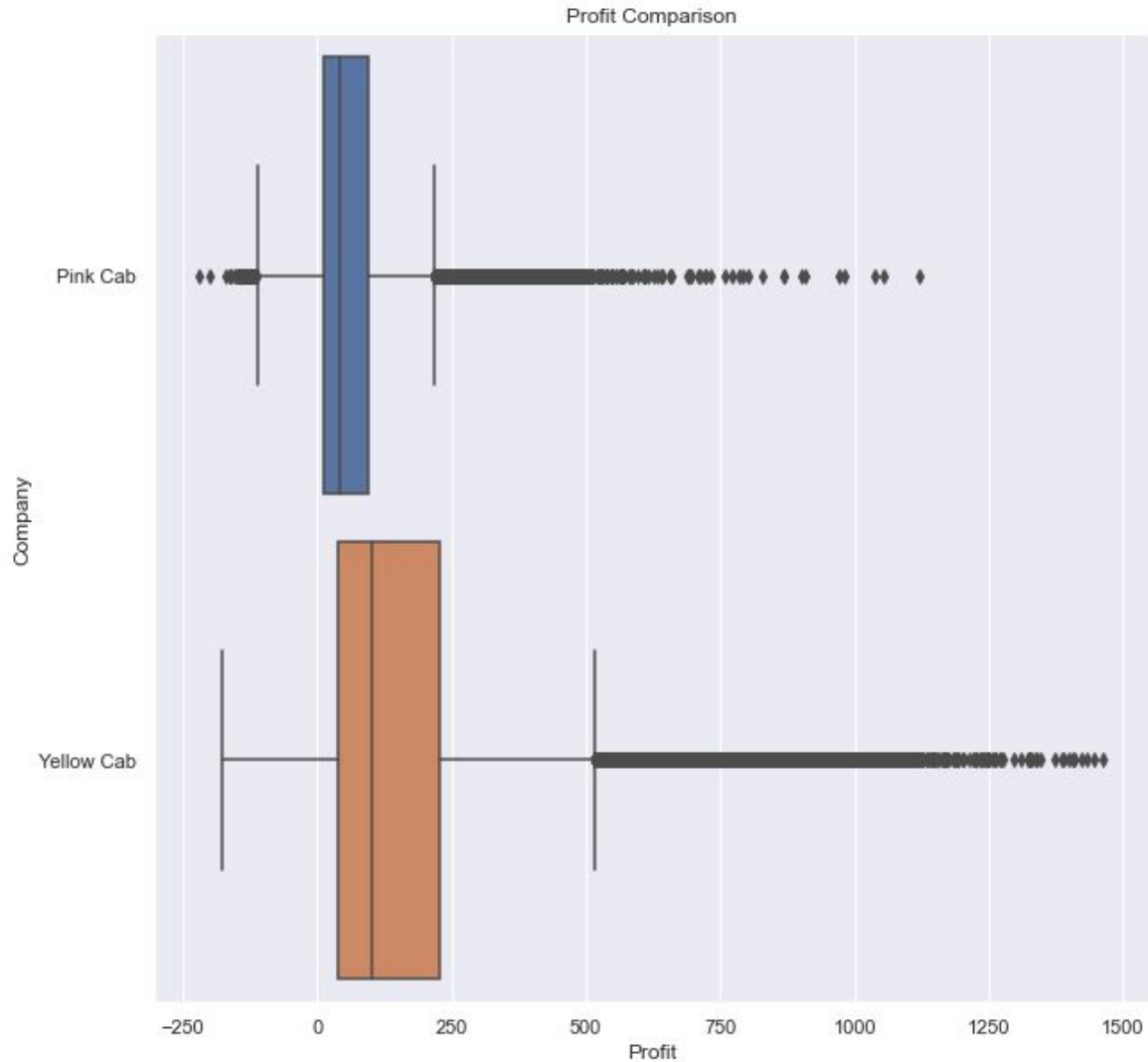
The Male and Female distributions are about the same. 50% of the demographic fall between the ages of 25 and 42.



A majority of the users are at the age of 40 or below.



Yellow Cab spends more money per ride but also charges more.



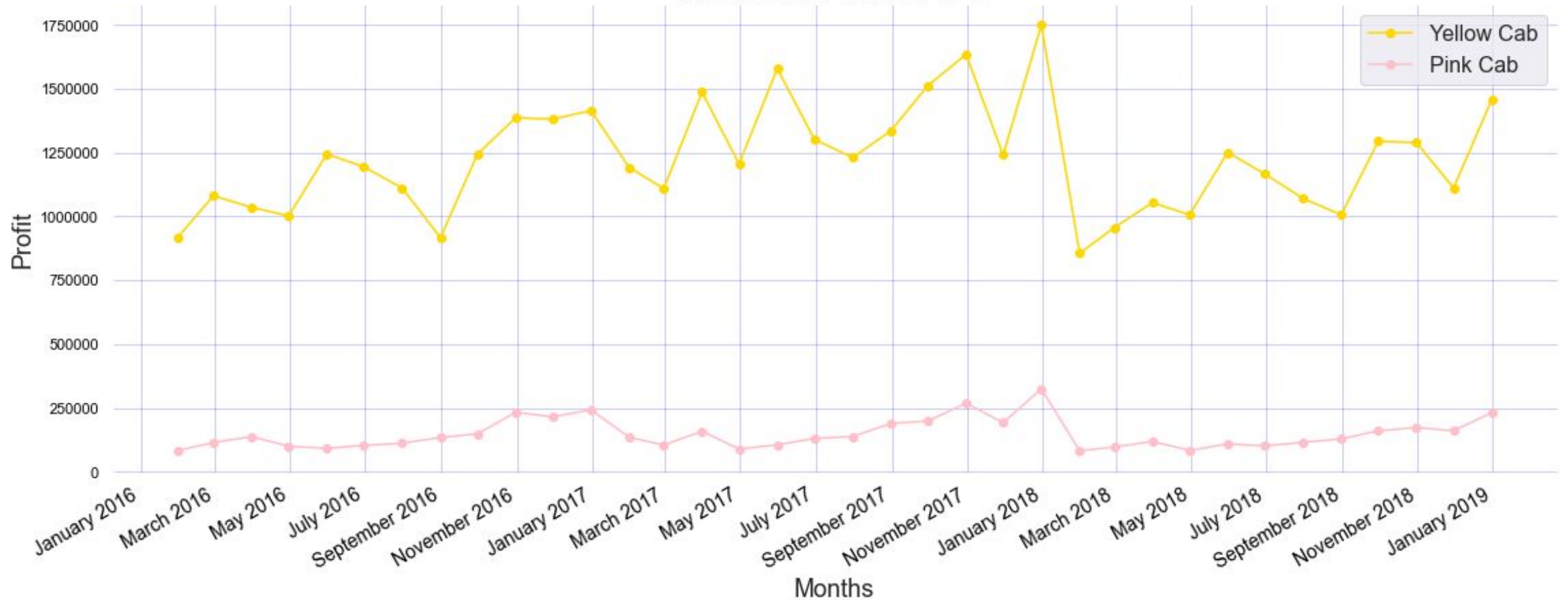
	count	mean	std	min	25%	50%	75%	max
Company								
Pink Cab	84711.0	62.652174	77.590378	-220.0600	11.206	41.5	93.9650	1119.480
Yellow Cab	274681.0	160.259986	171.823833	-176.9308	37.180	102.0	228.6704	1463.966

About 50% of the rides from Yellow Cab earn about \$37.18 to \$228.67.

About 50% of the rides from Pink Cab earn about \$11.20 to \$93.96.

The average profit per ride from Yellow Cab is more than double that of Pink Cab.

## MONTHLY PROFITS



Yellow Cab has more profits in every month compared to Pink Cab.

# Summary

- Yellow Cab company has much a wider customer pool in terms of unique users(different individual users).
- Yellow Cab has provided more rides in almost all of the cities that are shown.
- The most populated cities in order is New York, Chicago, Los Angeles, Miami, and Silicon Valley.
- They have provided more than 6 times the rides in New York city compared to Pink Cab.



**Data Glacier**

Your Deep Learning Partner

# Summary

- Yellow Cab has earned much more money per month than Pink Cab from 31/01/2016 to 31/12/2018 in ALL of the months.
- The average profit per ride from Yellow Cab(\\\$160.25) is more than double that of Pink Cab(\\\$62.65).
- The main demographic for customers are at the ages of 40 or below. Half of the demographic for customers are at the ages of 33 to 42.
- The age distributions for both companies are similar based from the boxplot.



**Data Glacier**

Your Deep Learning Partner



# EDA Recommendation

After reviewing and analyzing the data that has been provided, it is clear that Yellow Cab is a better investment than their competition in Pink Cab.



**Data Glacier**

Your Deep Learning Partner

# Thank You



**Data Glacier**

Your Deep Learning Partner