Data Glacier

Your Deep Learning Partner

EXPLORATORY DATA ANALYSIS PROJECT: G2M INSIGHT FOR CAB INVESTMENT FIRM

KSENIYA KUDZELICH

JUNE 20, 2024

AGENDA

01 PROBLEM STATEMENT

02 APPROACH

03 GENERAL OVERVIEW

04 ANALYSIS OVERVIEW

05 EXPLORATORY DATA ANALYSIS

06 HYPOTHESIS TESTING RESULTS

07 RECOMMENDATIONS

01
PROBLEM STATEMENT

CLIENT:

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:

Provide actionable insights to help XYZ firm identify the right company to make their investment.

02
APPROACH

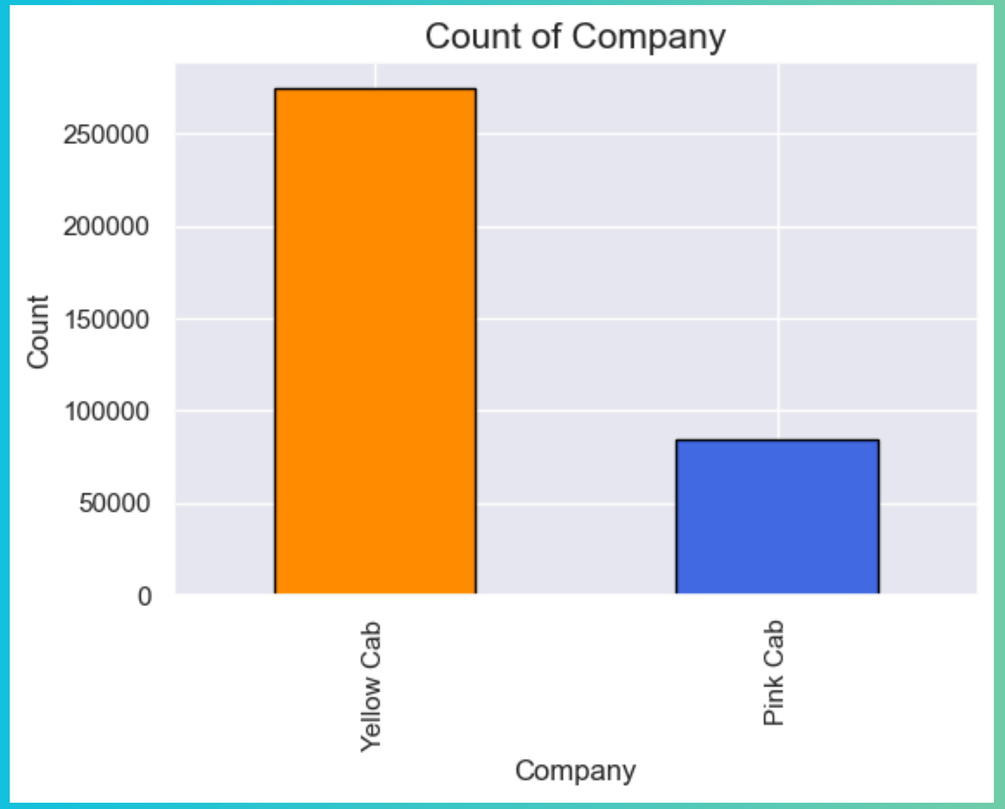
- My strategic approach is to **meticulously analyze** the provided datasets, encompassing *transaction details, customer demographics, transaction-to-customer mappings, and city-specific cab usage statistics*.

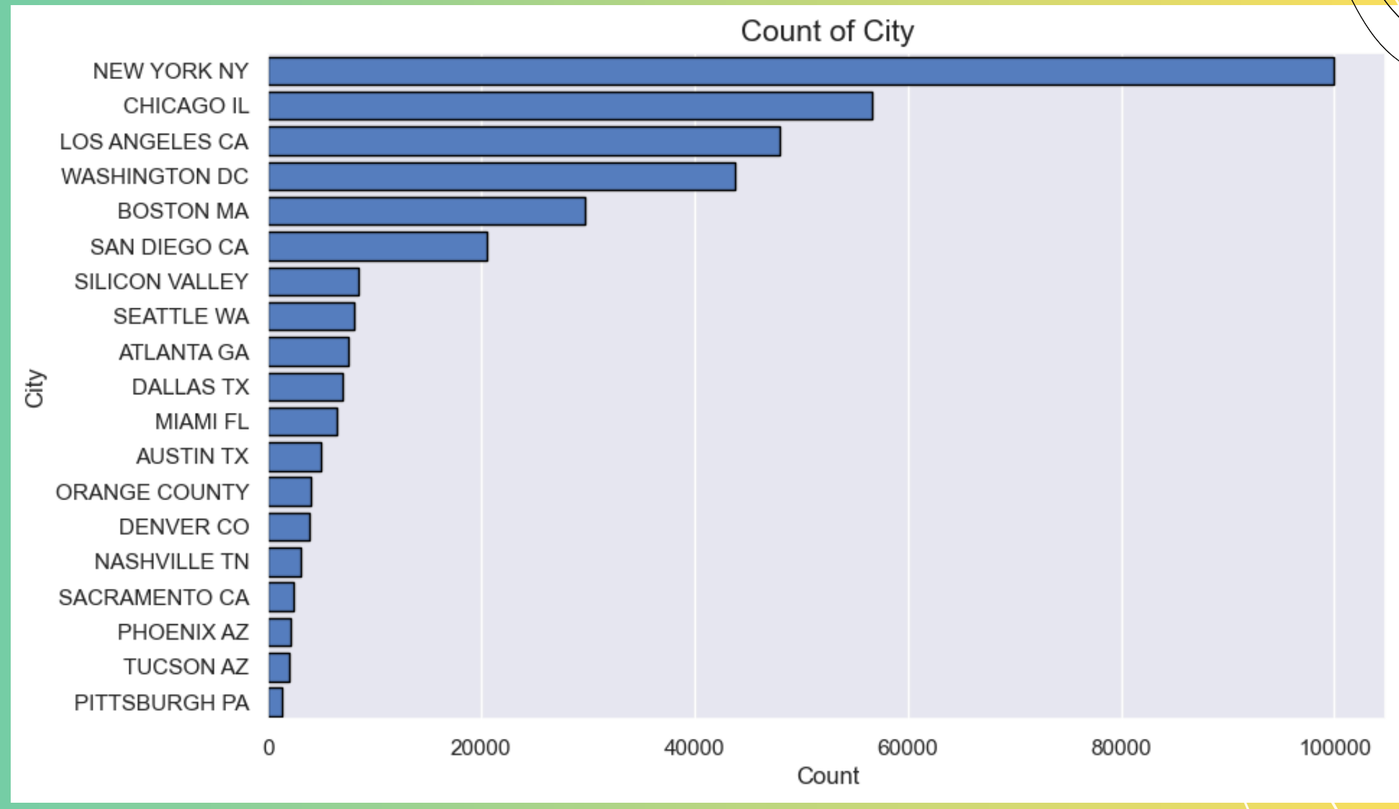
- Through this **comprehensive analysis and hypothesis testing**, I aim to **uncover actionable insights and trends** within the cab industry.

- **Hypothesis testing** will allow to validate assumptions about the factors influencing profitability and market dynamics, providing a robust statistical foundation for my analysis.

03
GENERAL OVERVIEW

GENERAL OVERVIEW OF CAB DATA

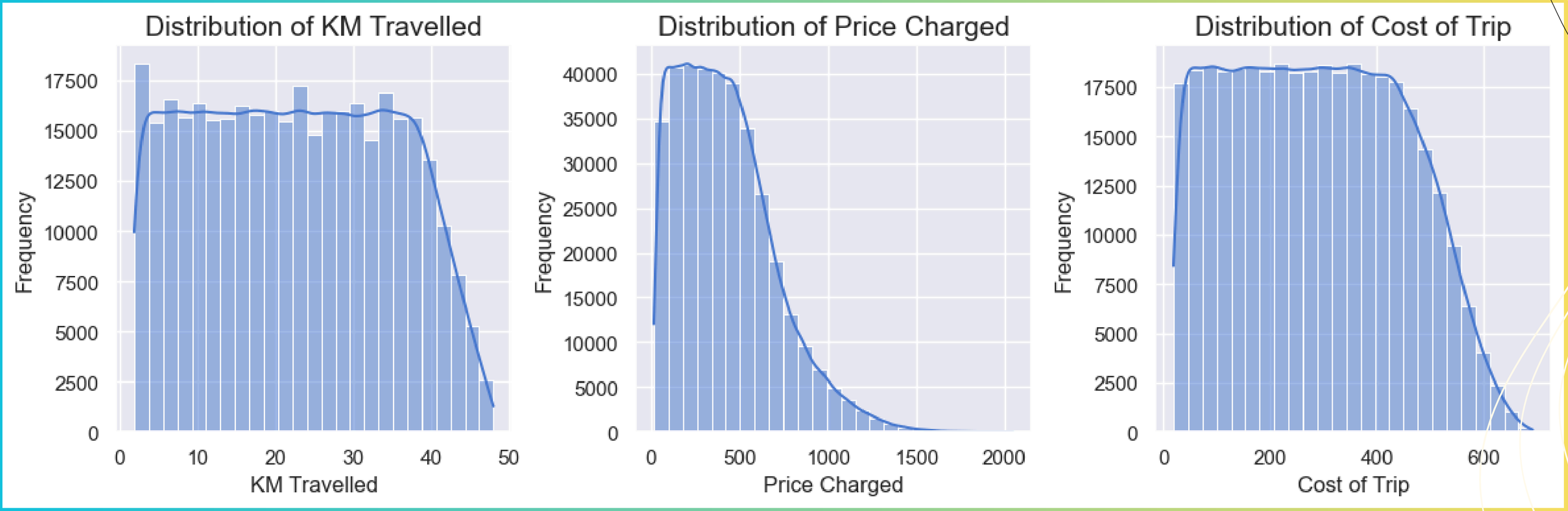




Yellow Cab dominates the market share

New York has the highest number of transactions

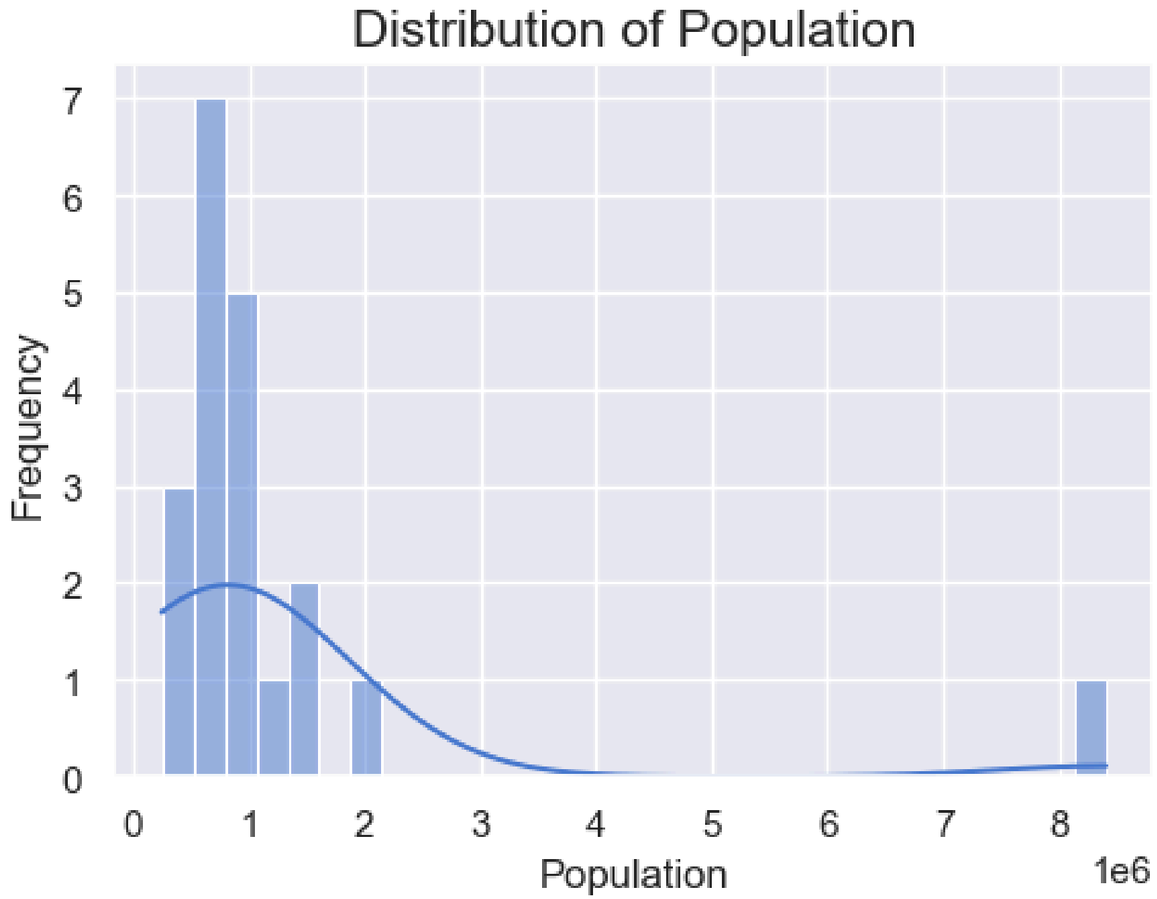
GENERAL OVERVIEW OF CAB DATA

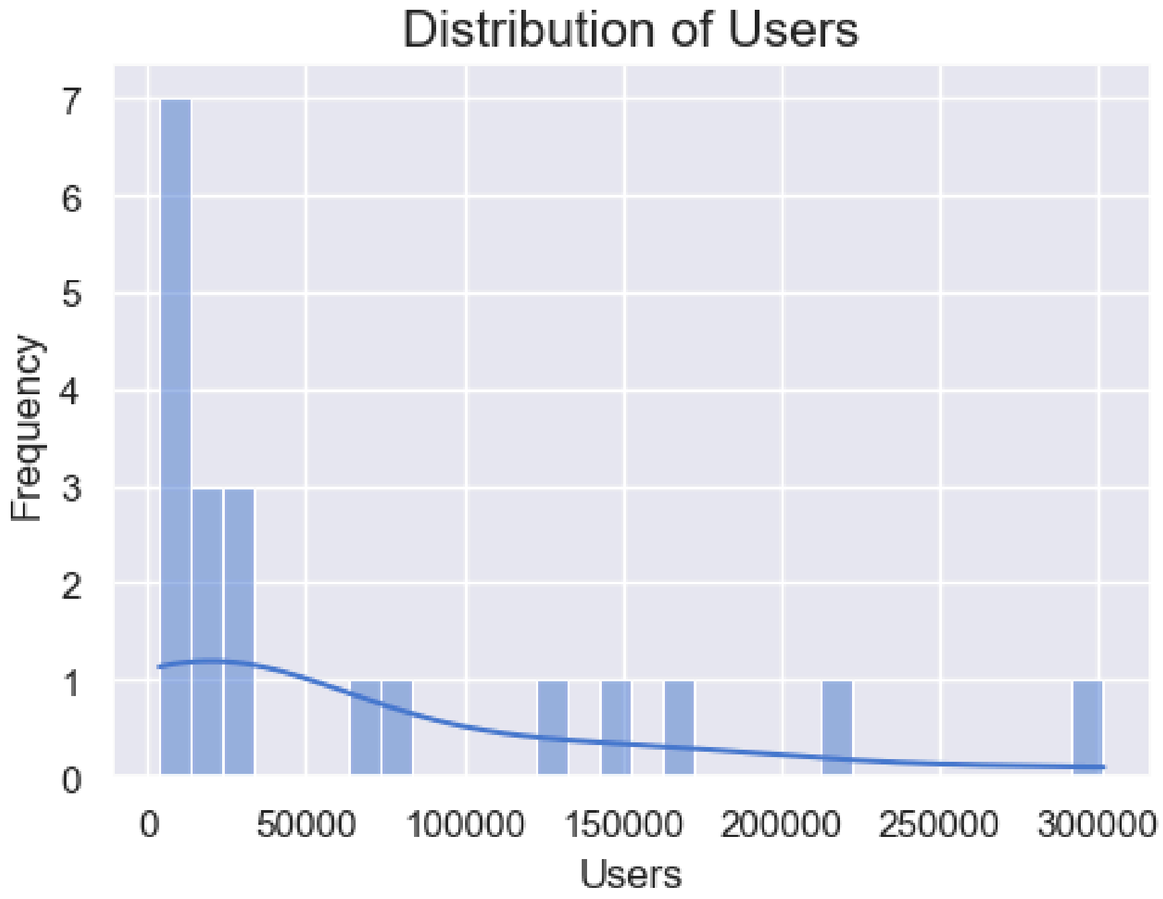


Most of the rides varies from 2 to 48 KM

Price Charged and Cost of Trip indicate a consistent pricing model based on KM Travelled

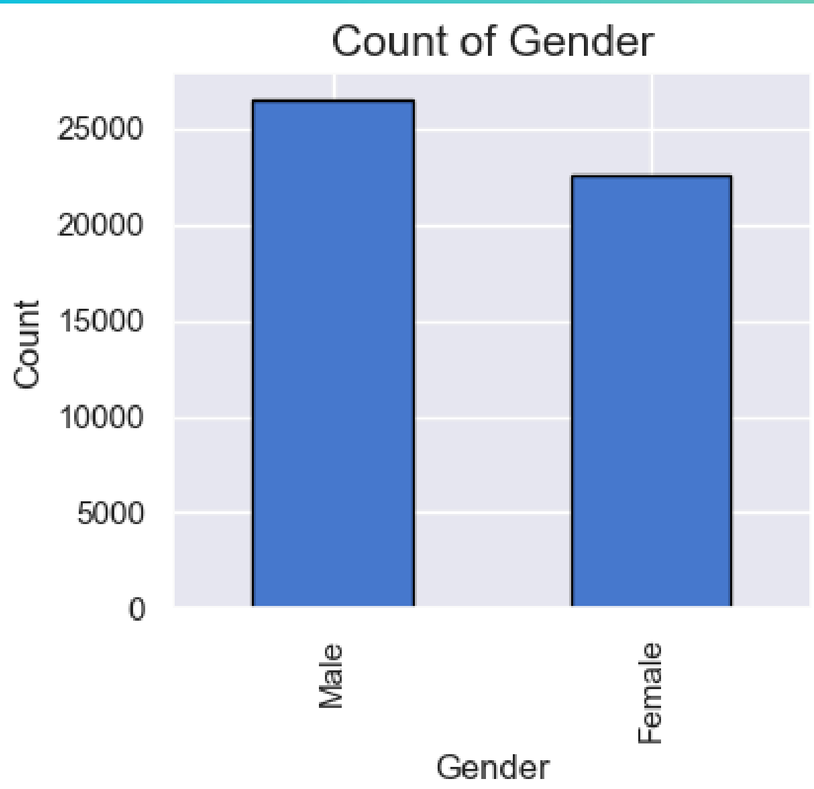
GENERAL OVERVIEW OF CITY DATA

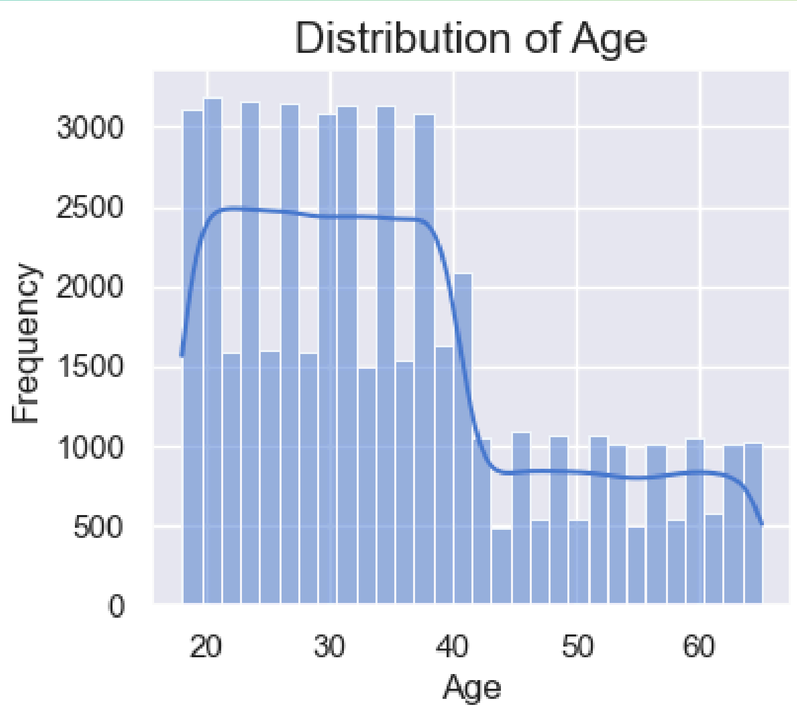


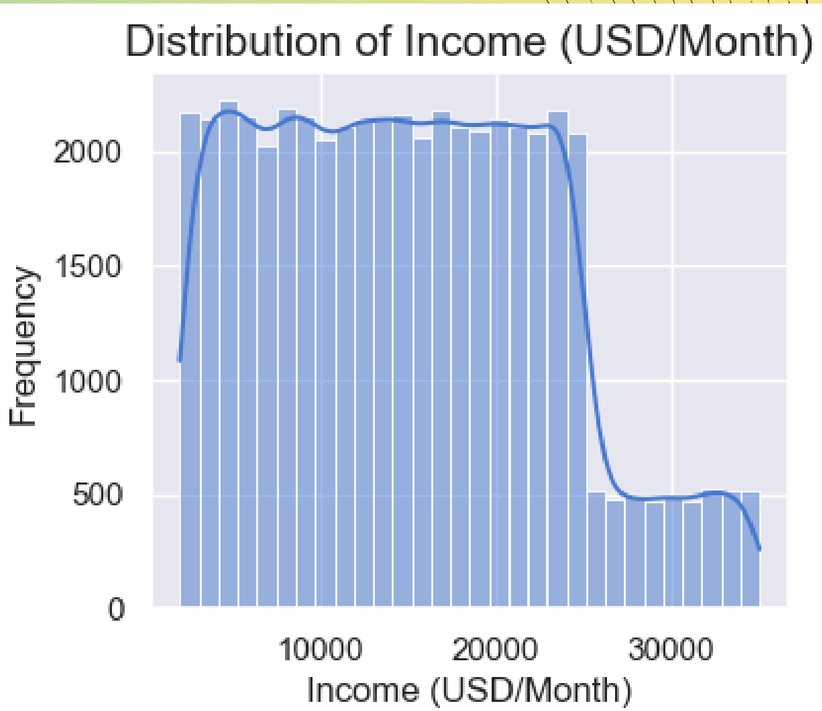


Population and Users show large values in certain cities, likely due to major urban centers

GENERAL OVERVIEW OF CUSTOMER ID DATA





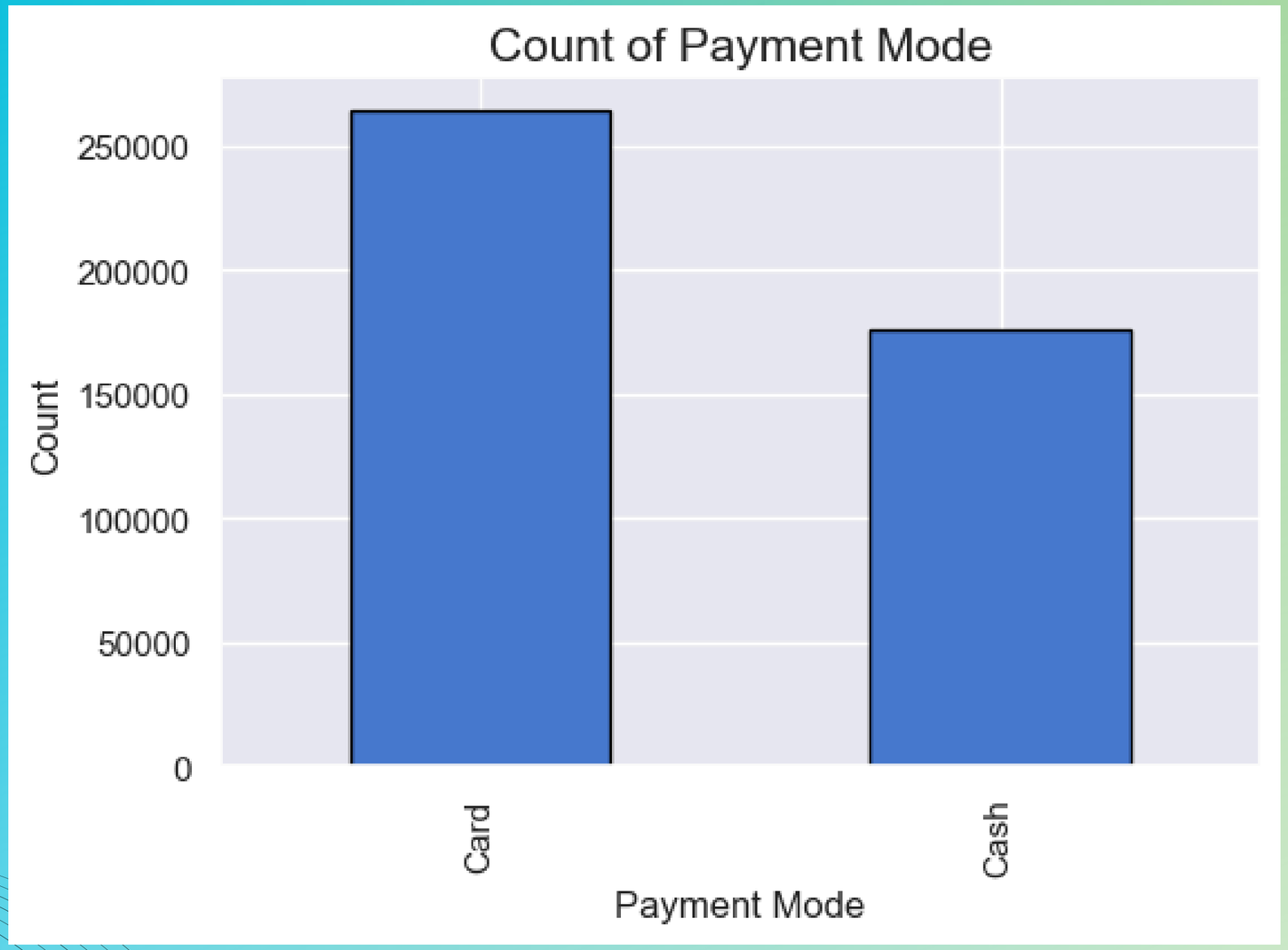


The number of male customers is higher than female customers.

Average age - 35 years

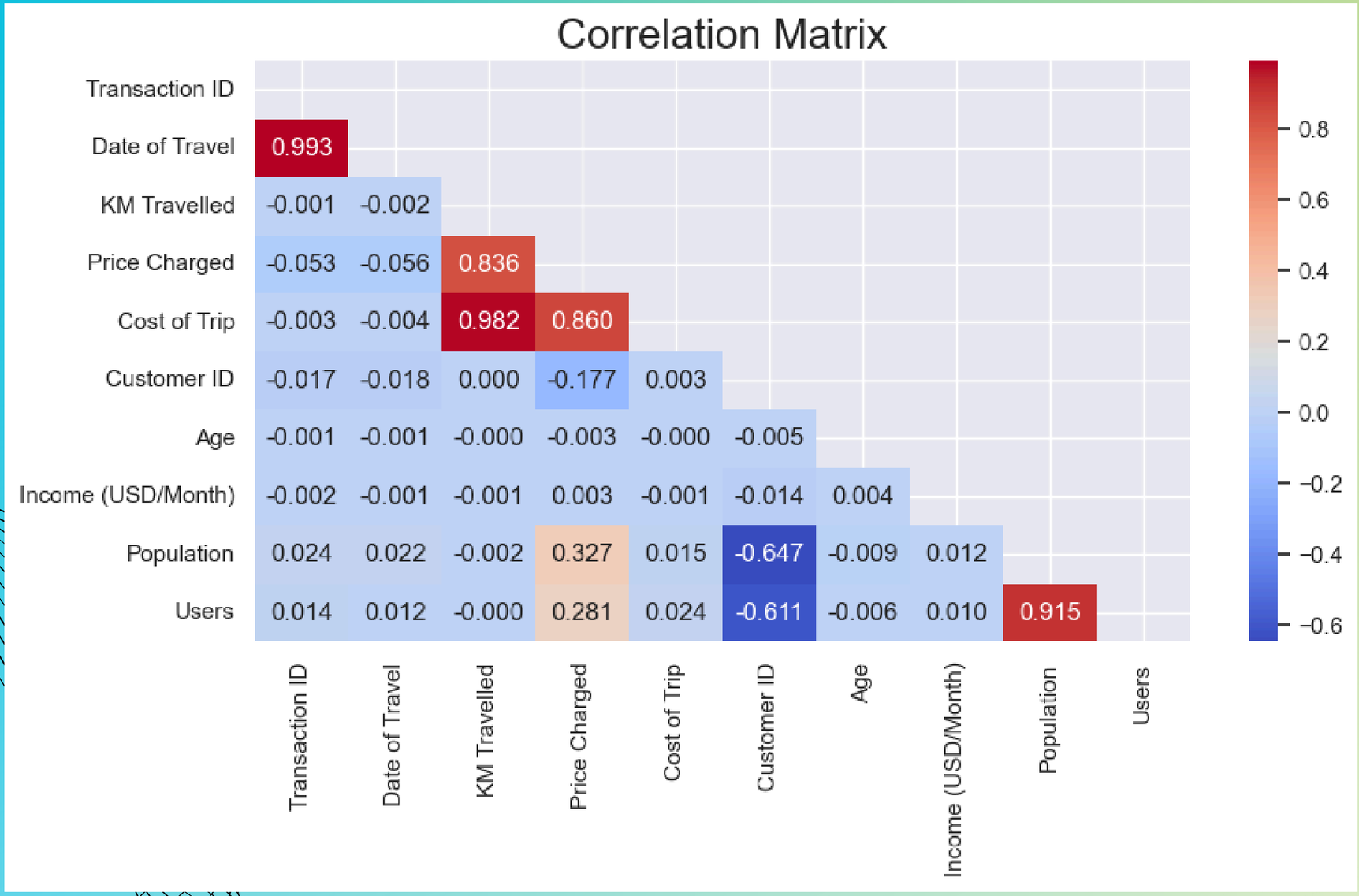
Most customers earn less than $20,000 per month.

GENERAL OVERVIEW OF TRANSACTION ID DATA



Card payments are more common than cash payments

CORRELATION BETWEEN COLUMNS



Strong positive correlation between:

- Population - Users
- Price Charged - Cost of Trip - KM Travelled

04
ANALYSIS OVERVIEW

OVERALL PERFORMANCE

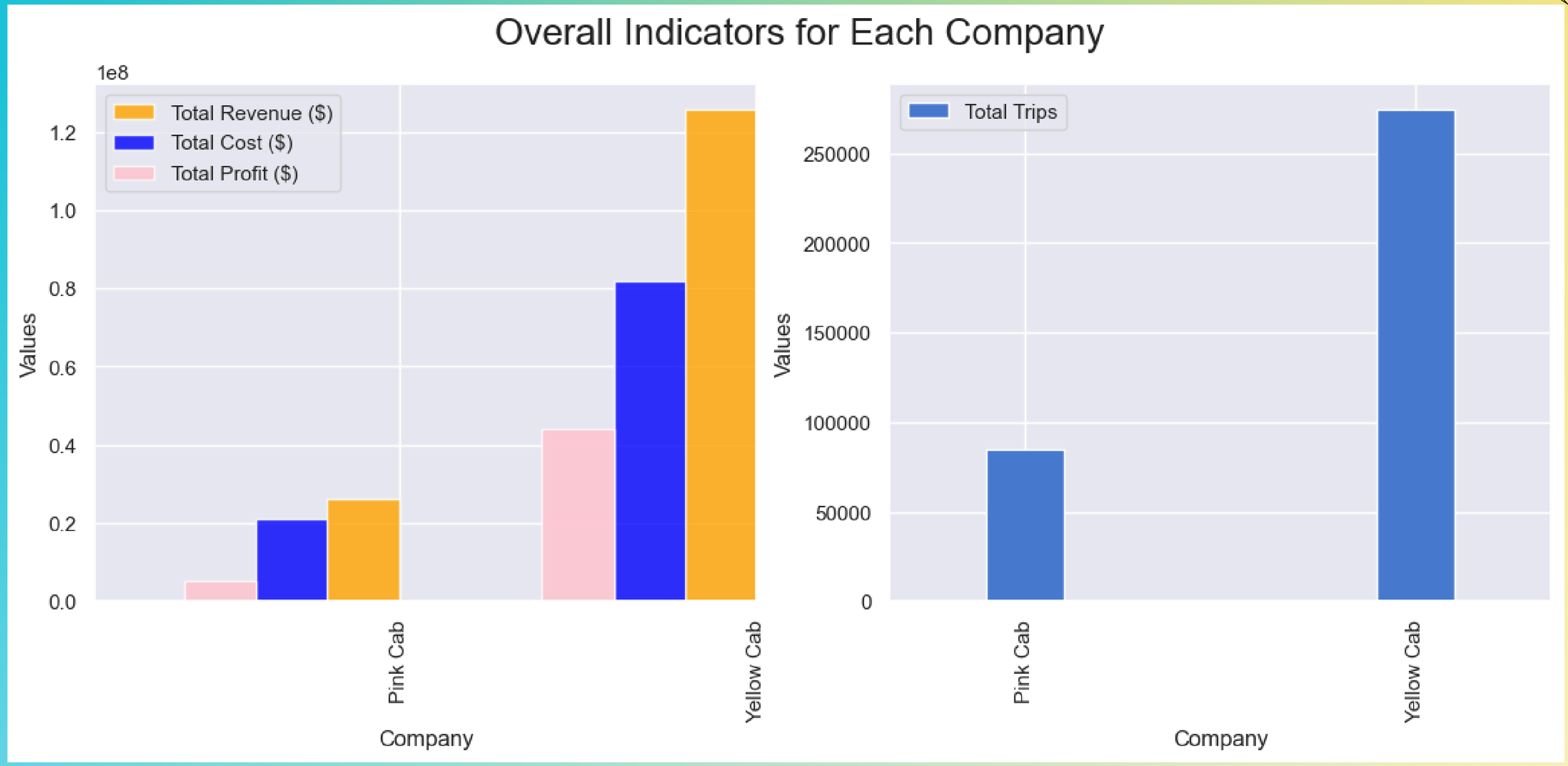
PROFITABILITY ANALYSIS

CUSTOMER BASE

GEOGRAPHICAL ANALYSIS

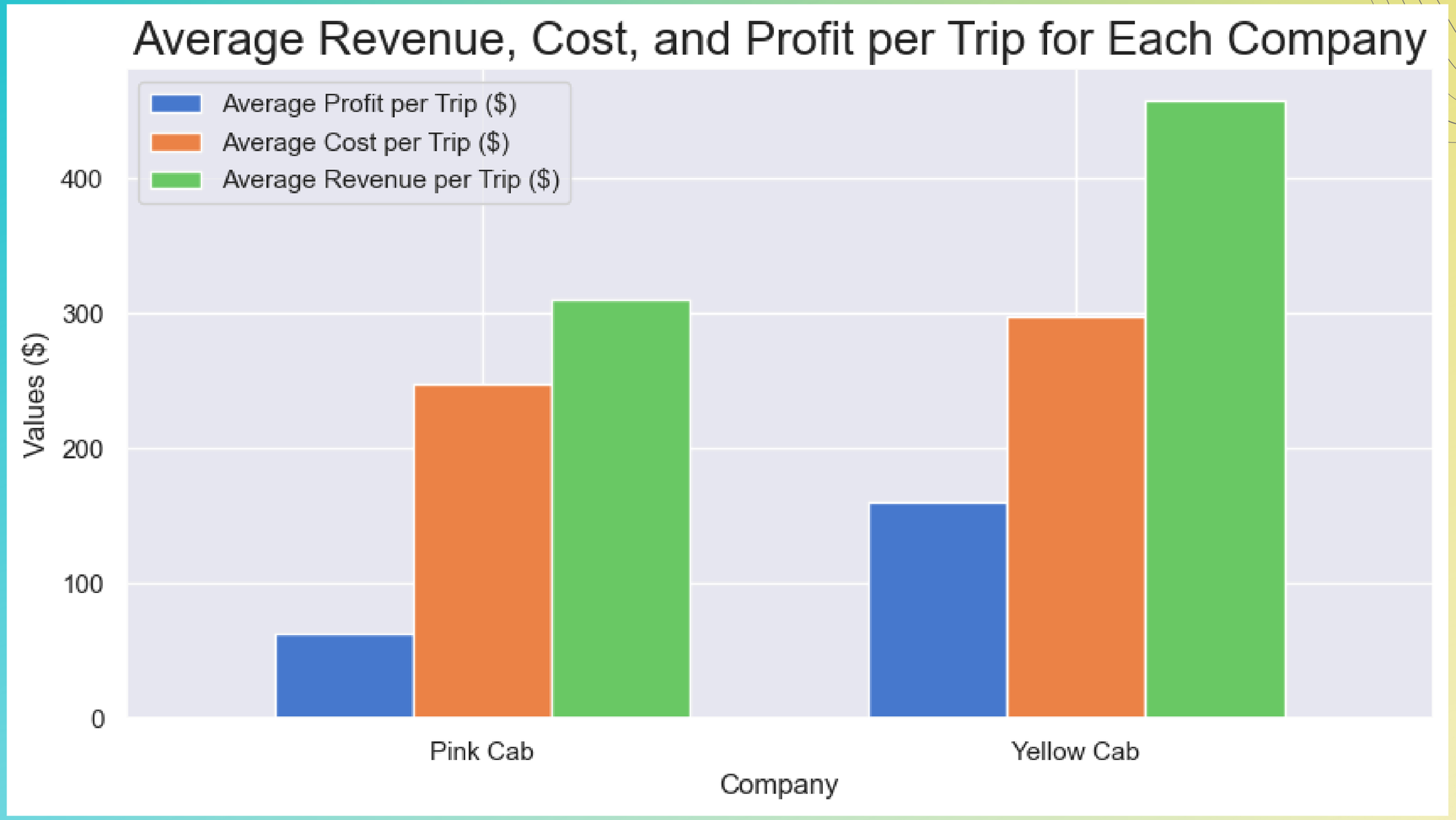
05
EXPLORATORY DATA ANALYSIS (EDA)

01. OVERALL PERFORMANCE



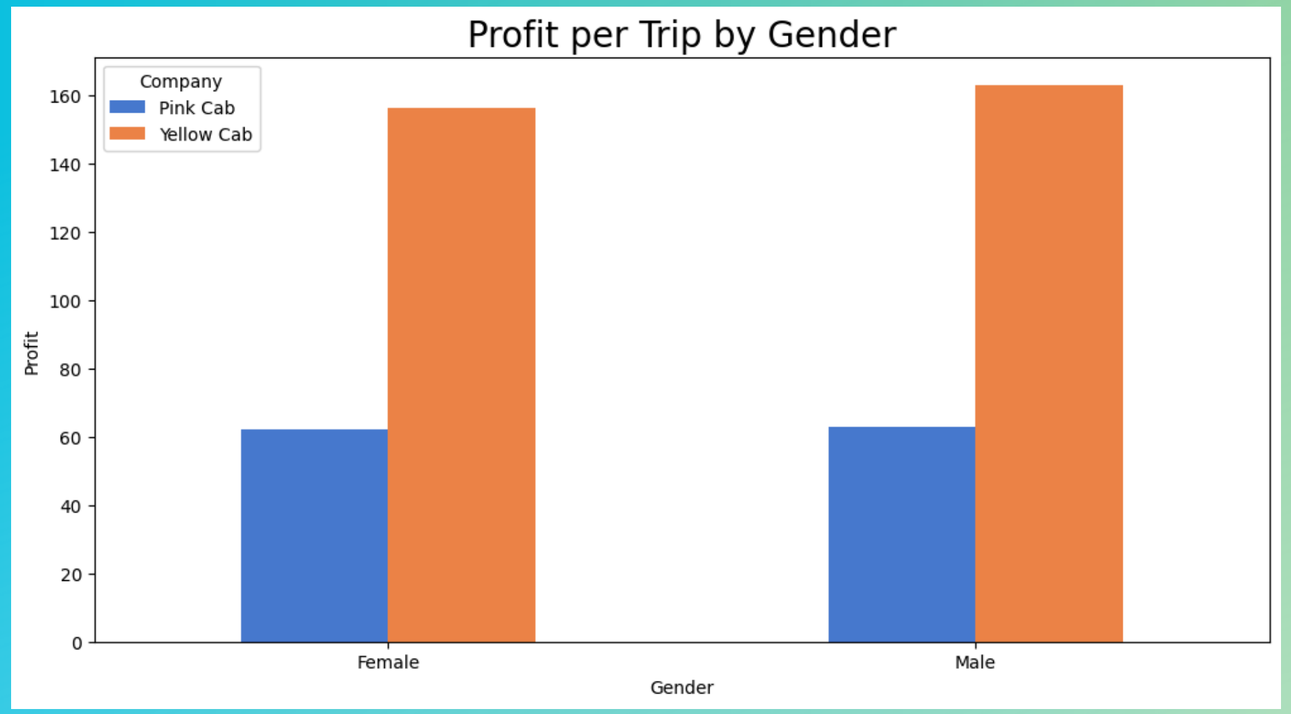
Yellow Cab is clearly superior to the Pink Cab in **all key indicators**

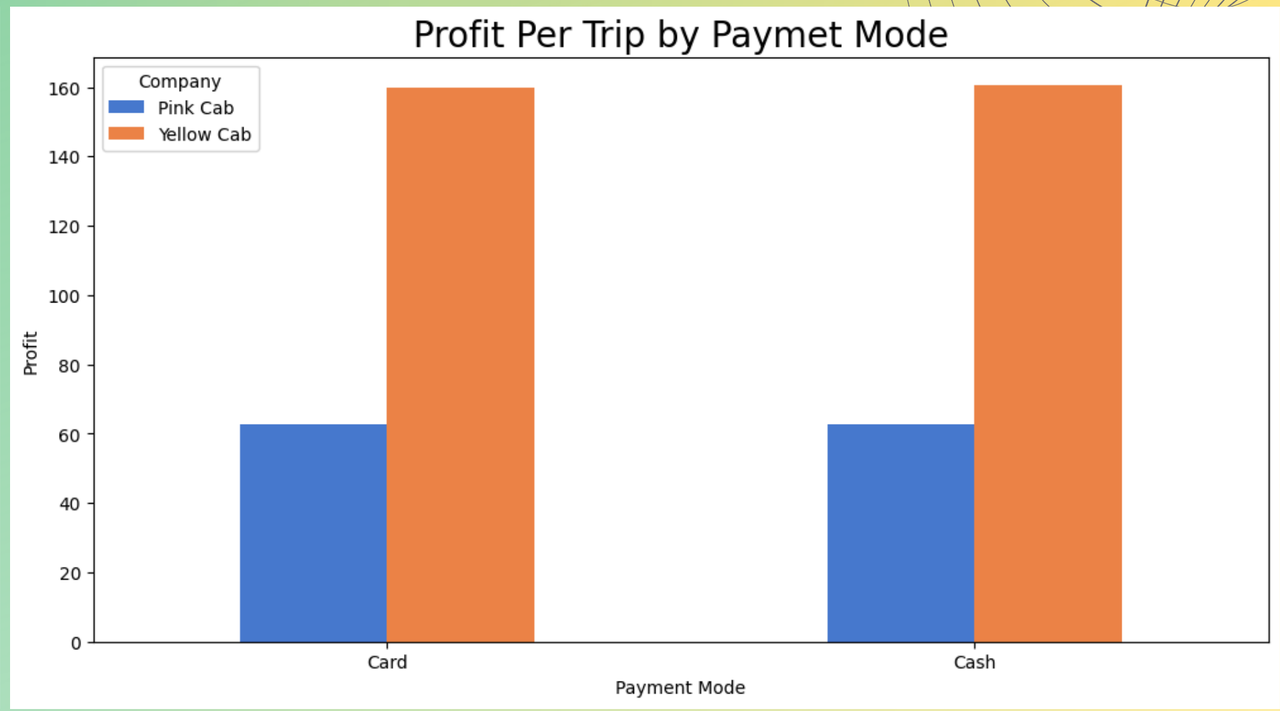
02. AVERAGE REVENUE, COSTS AND PROFIT PER TRIP



All indicators per Trip for Yellow Cab is higher for Pink Cab.

02. PROFIT PER TRIP BY GENDER OR PAYMENT MODE





There is no difference in Profit Per Trip by Gender (hypothesis 1)

There is no difference in Profit Per Trip by Payment Mode (hypothesis 2)

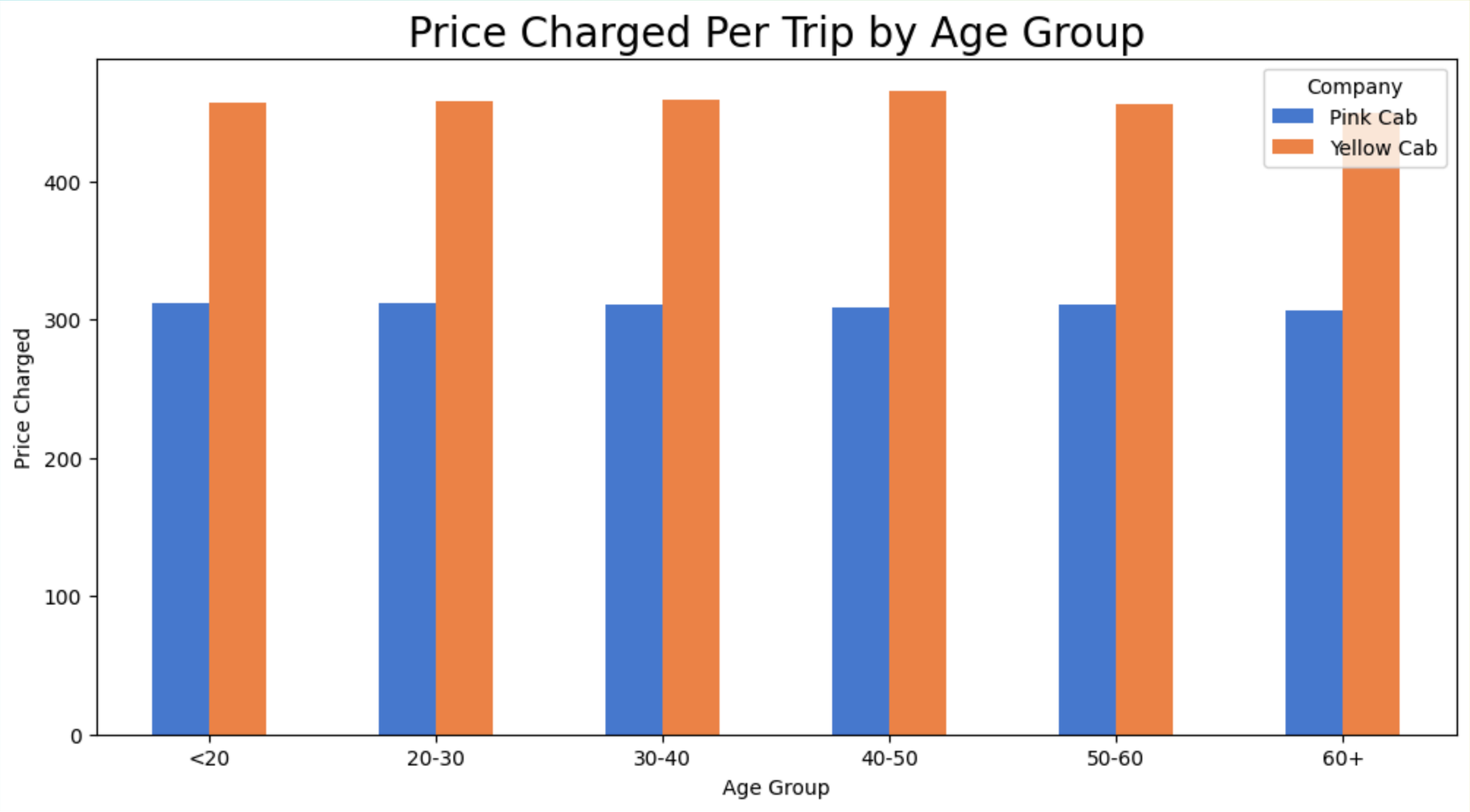
02. AVERAGE REVENUE, COSTS AND PROFIT PER TRIP

- The average *revenue per trip* for Yellow Cab ($458.18) is higher than for Pink Cab ($310.80)
- The average *cost per trip* for Yellow Cab ($297.92) is also higher than for Pink Cab ($248.15)
- The average *profit per trip* for Yellow Cab ($160.26) significantly exceeds Pink's Cab profit ($62.65)
- These data indicate that **Yellow Cab provides a higher Profit on each trip**, despite the higher costs

02. PRICES CHARGED BY COMPANY

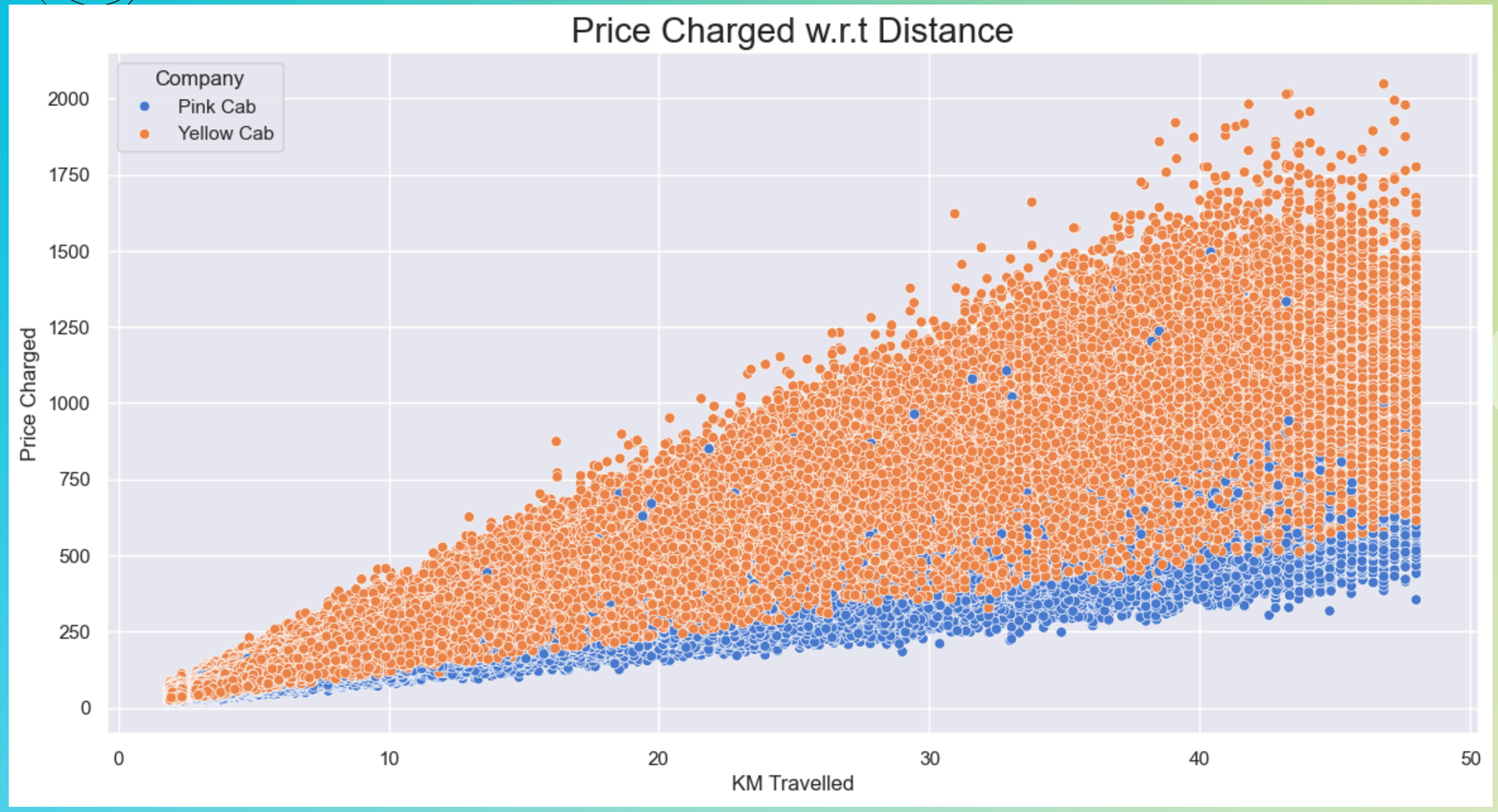
Prices Charged of Yellow Cab in generall is higher than of Pink Cab (median values $425 vs $298)

02. PRICE CHARGED PER TRIP BY AGE GROUP



There is no difference between Price Charged Per Trip by Age Group (hypothesis 3)

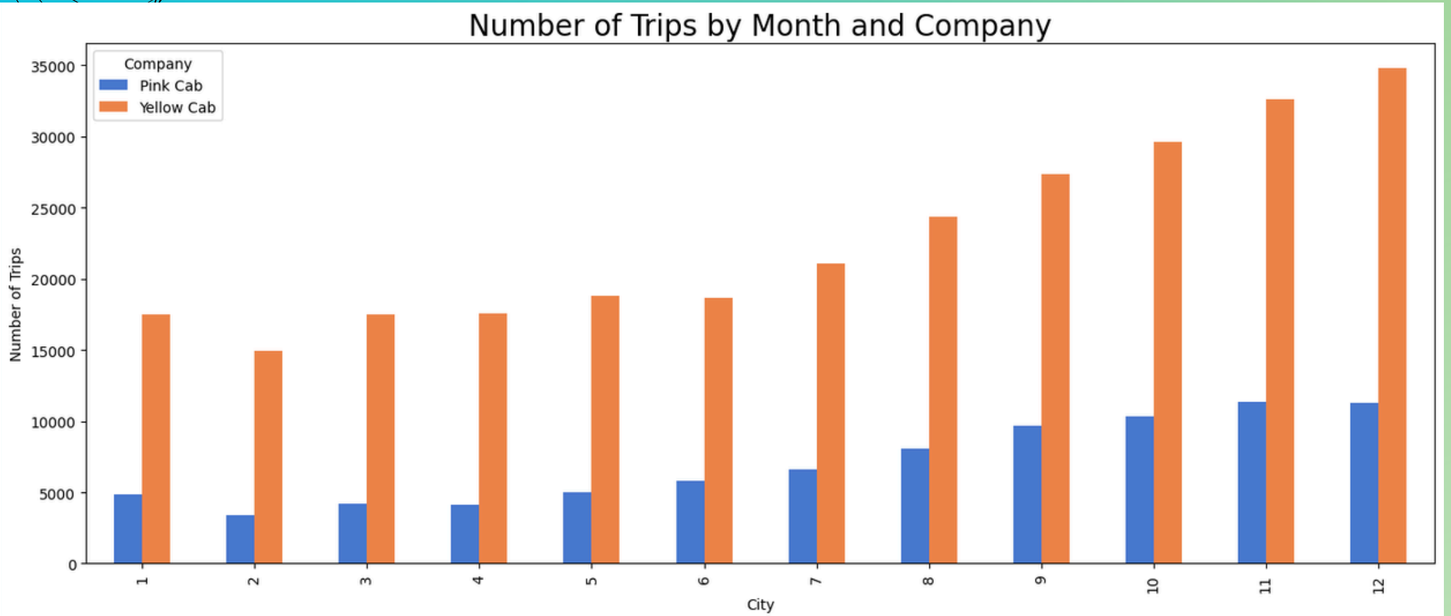
02. PRICE CHARGED VS DISTANCE

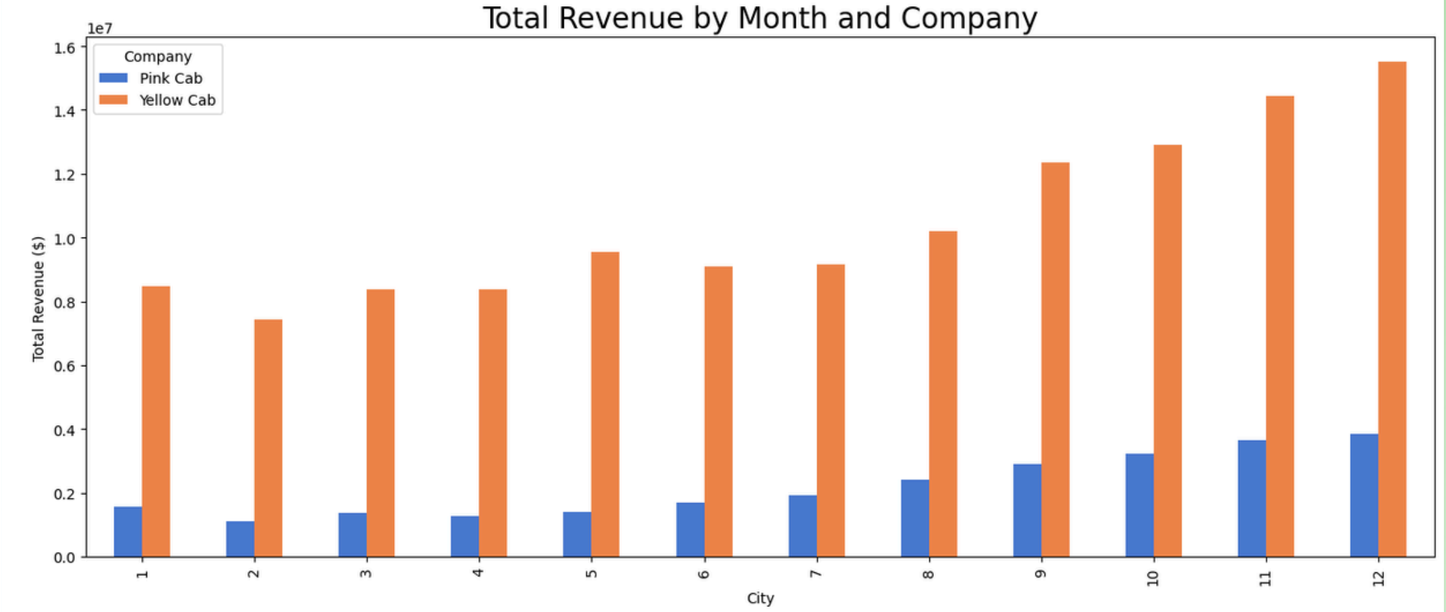


There is a *linear relationship* between KM Traveled and Price Charged

However, **Yellow Cab has higher charges** compared to Pink Cab

02. SEASONAL TRENDS

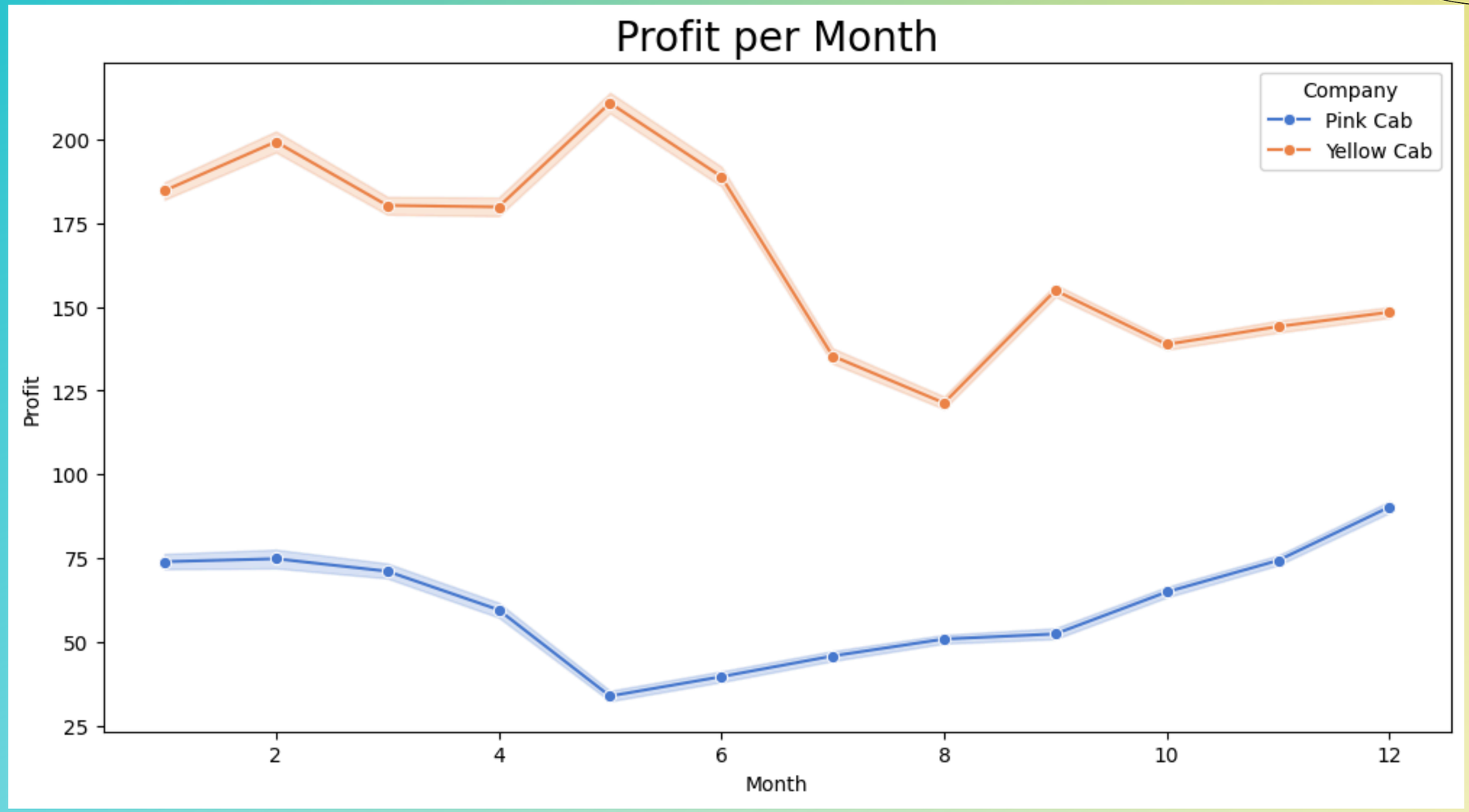




- Both companies show seasonal fluctuations with certain peaks and troughs at different times of the year.
- However, the **Yellow Cab shows more stable growth** compared to the Pink Cab.

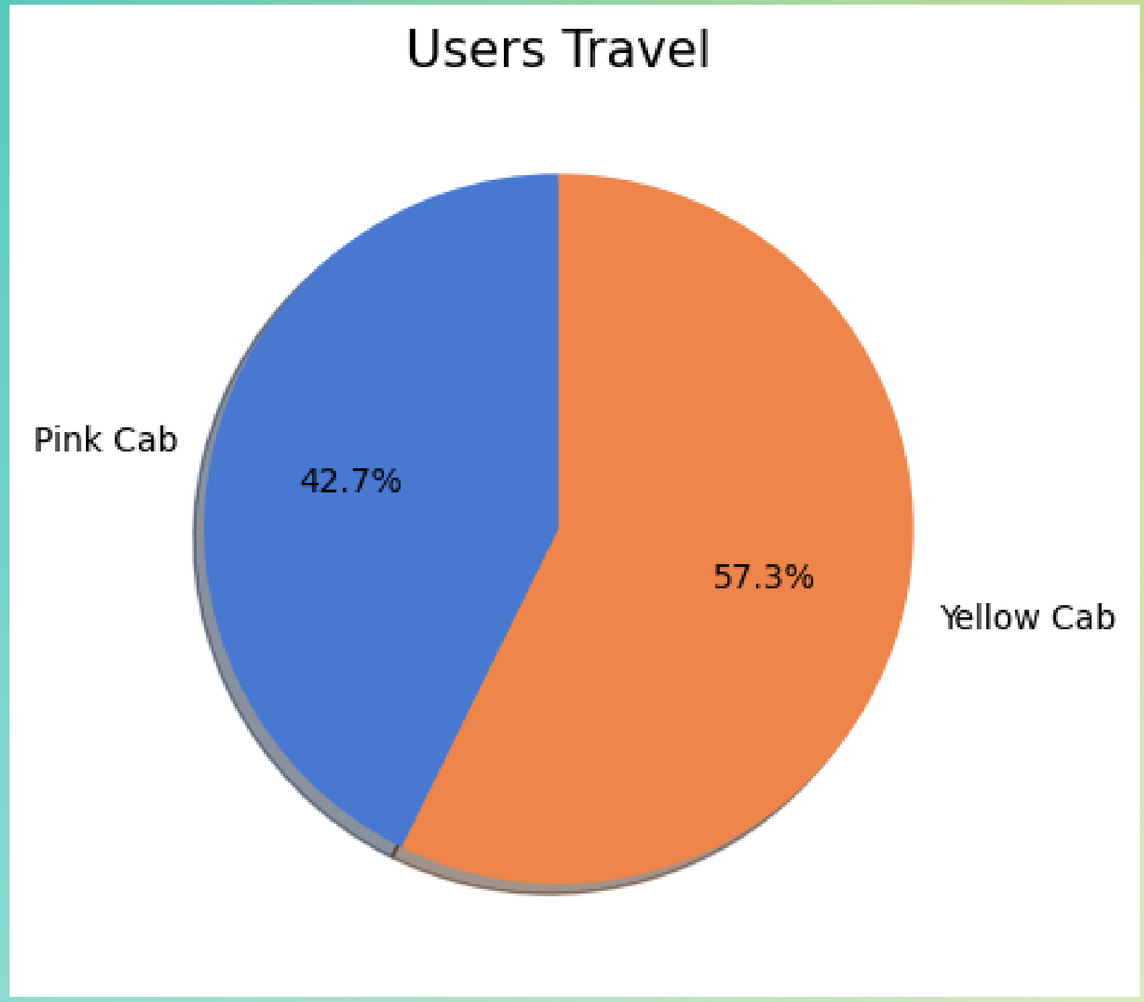
- The revenue of both companies also shows seasonal fluctuations.
- **Yellow Cab has higher revenue indicators during the year.**

02. PROFIT PER MONTH



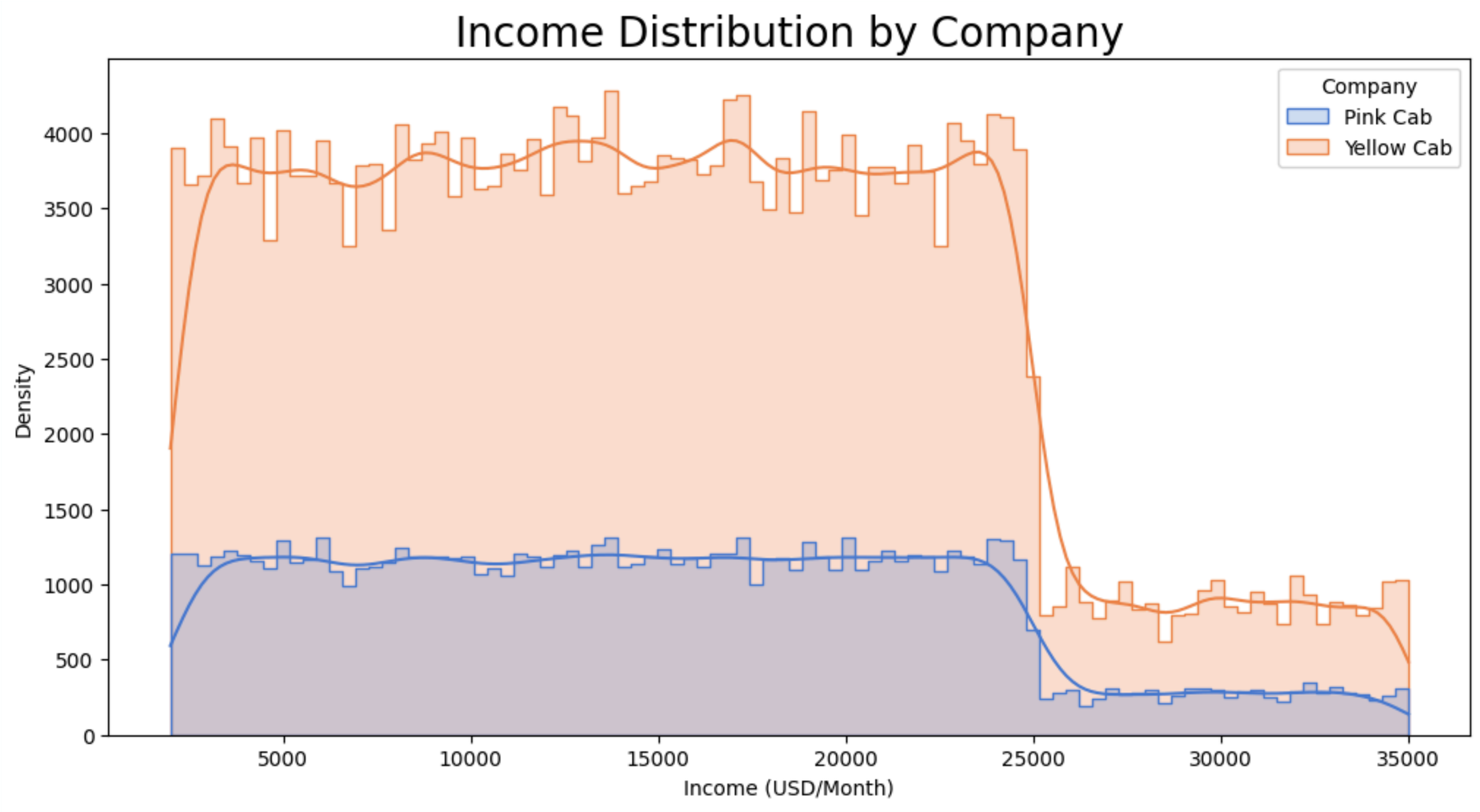
The **total Profit of Yellow Cab is much more** than Pink Cab, which makes sense considering Yellow Cab's market share is about 70%.

03. USERS TRAVEL



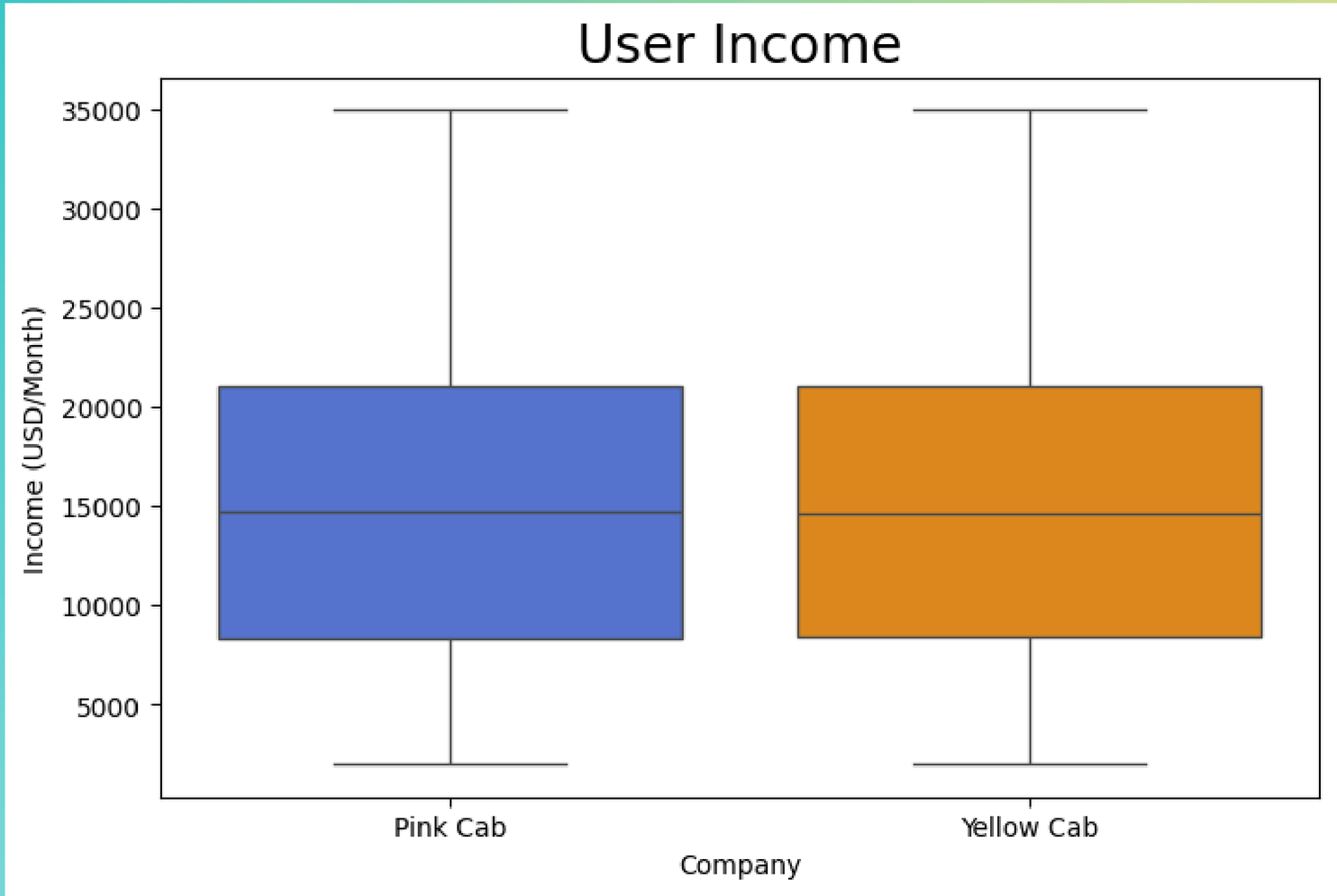
Users are more likely ride by Yellow cab more than by Pink Cab.

03. CUSTOMER INCOME DISTRIBUTION



The income distribution by company shows that both companies serve customers across a wide range of income levels.

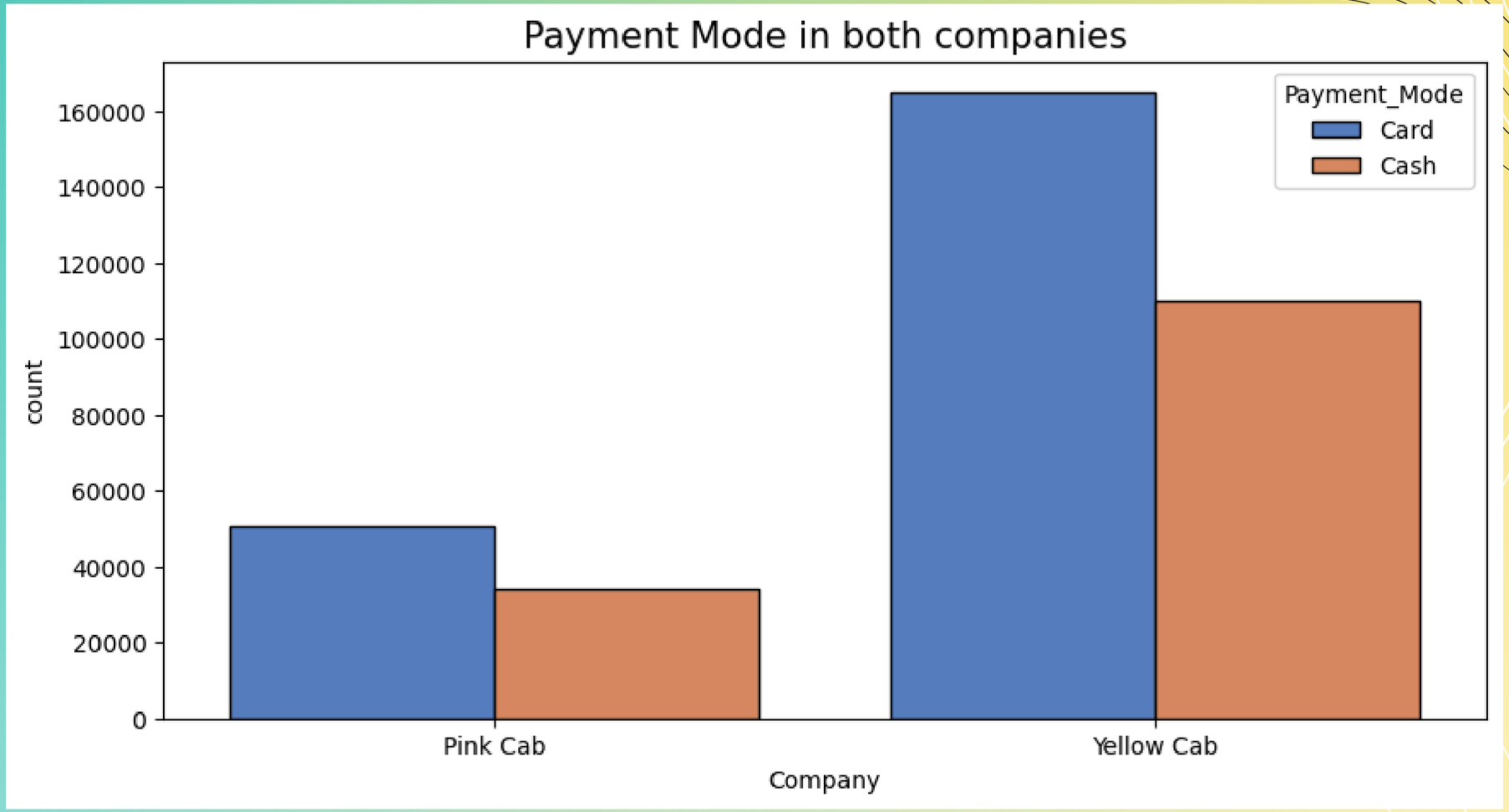
03. AVERAGE INCOME OF CUSTOMERS



The income of the clients of both companies varies widely, but the **median income of Yellow Cab** (15059.05$) **customers is slightly less** compared to Pink Cab (15045.66$). (hypothesis 6)

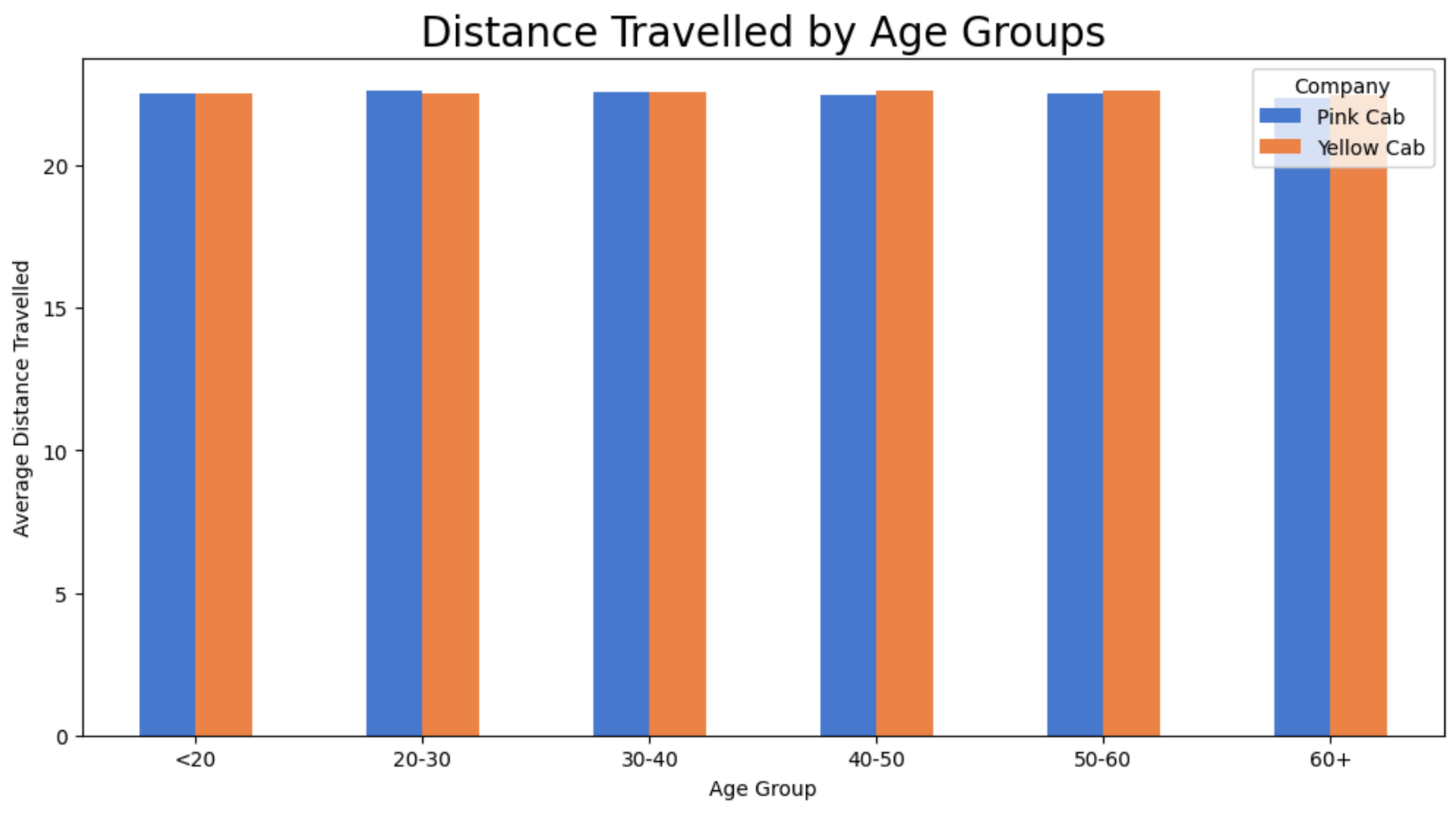
03. PAYMENT MODE

- Pink Cab has a more even distribution between cash and cards.



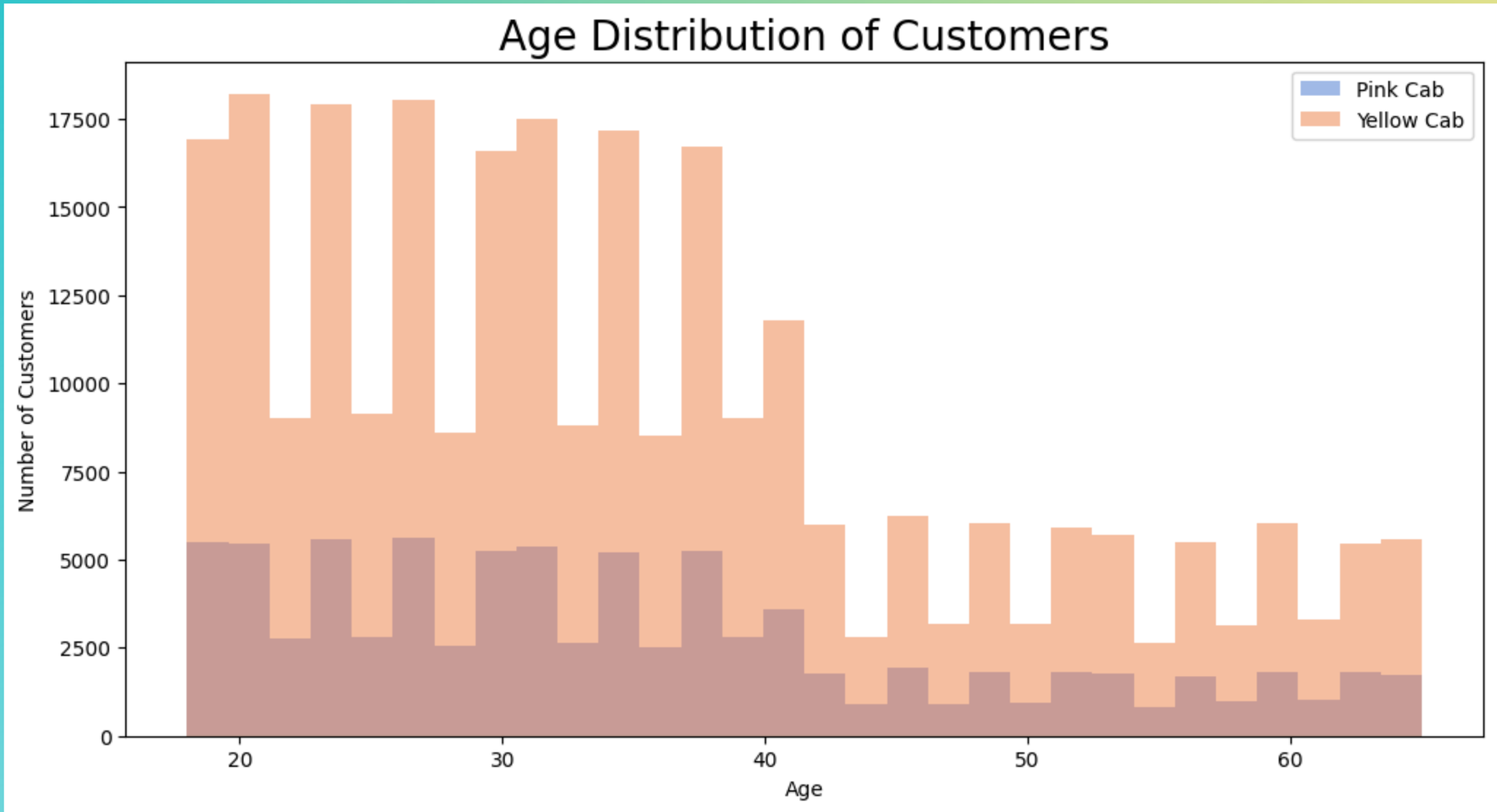
- Both payment methods (cash and card) are used in both companies, however, the use of Yellow Cab cards is significantly higher compared to Pink Cab.

03. DISTANCE TRAVELLED BY AGE GROUPS



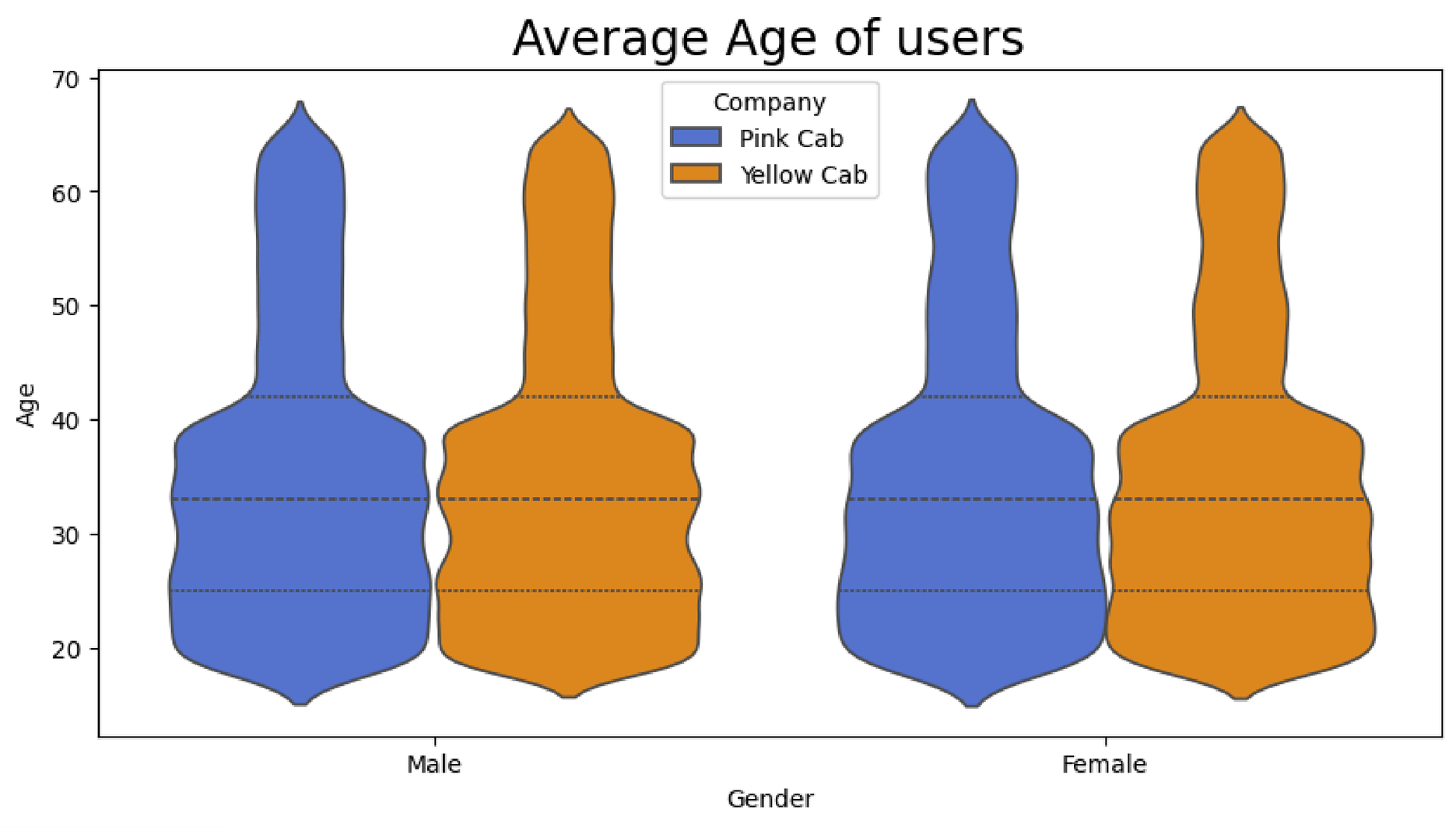
There is no diiference in Distance Travelled between Age Groups in companies. (hypothesis 5)

03. AGE DISTRIBUTION



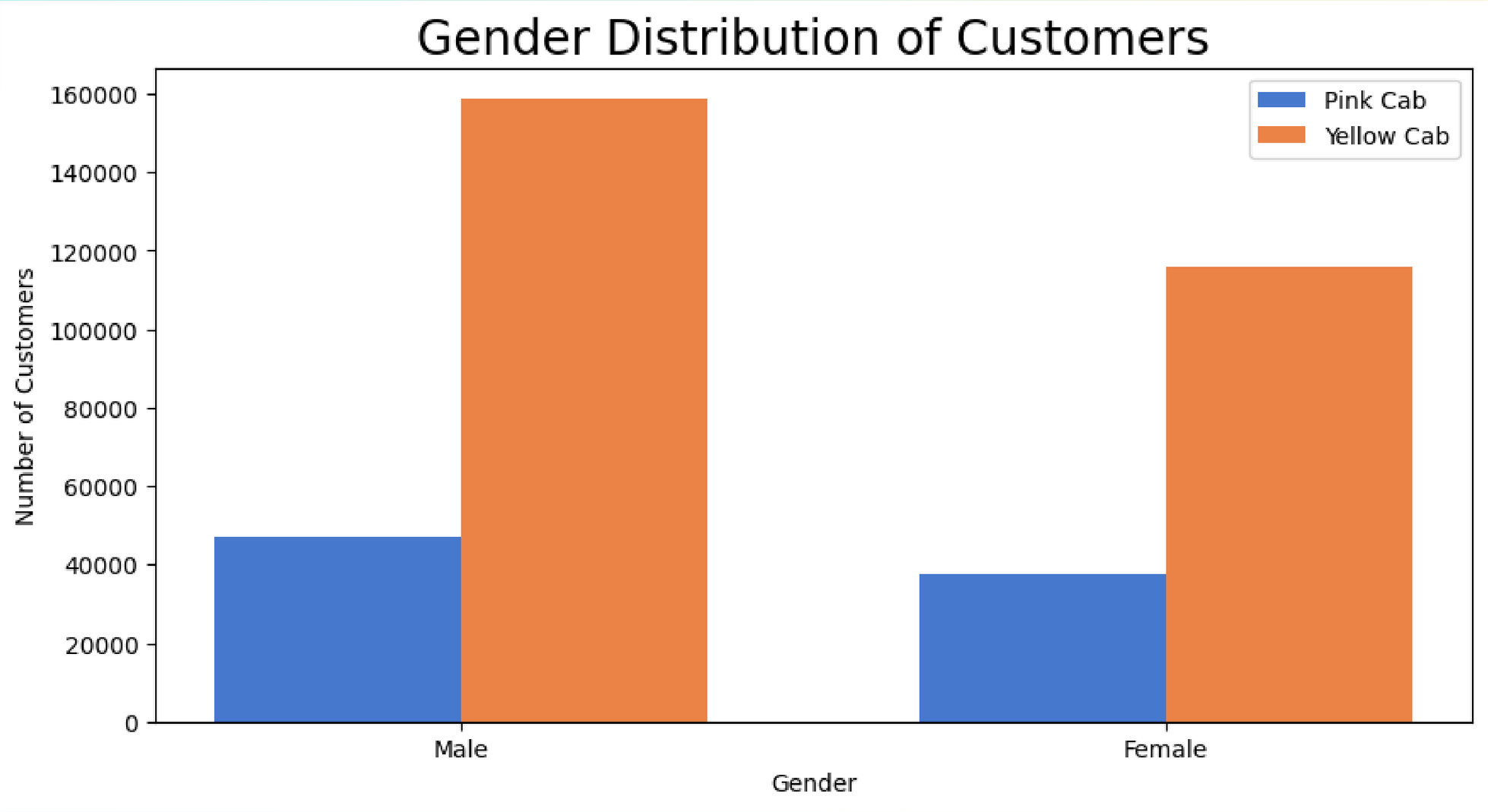
- The age groups of clients are similar for both companies.
- The majority of clients are between the ages of 20 and 40.

03. AVERAGE AGE OF USERS



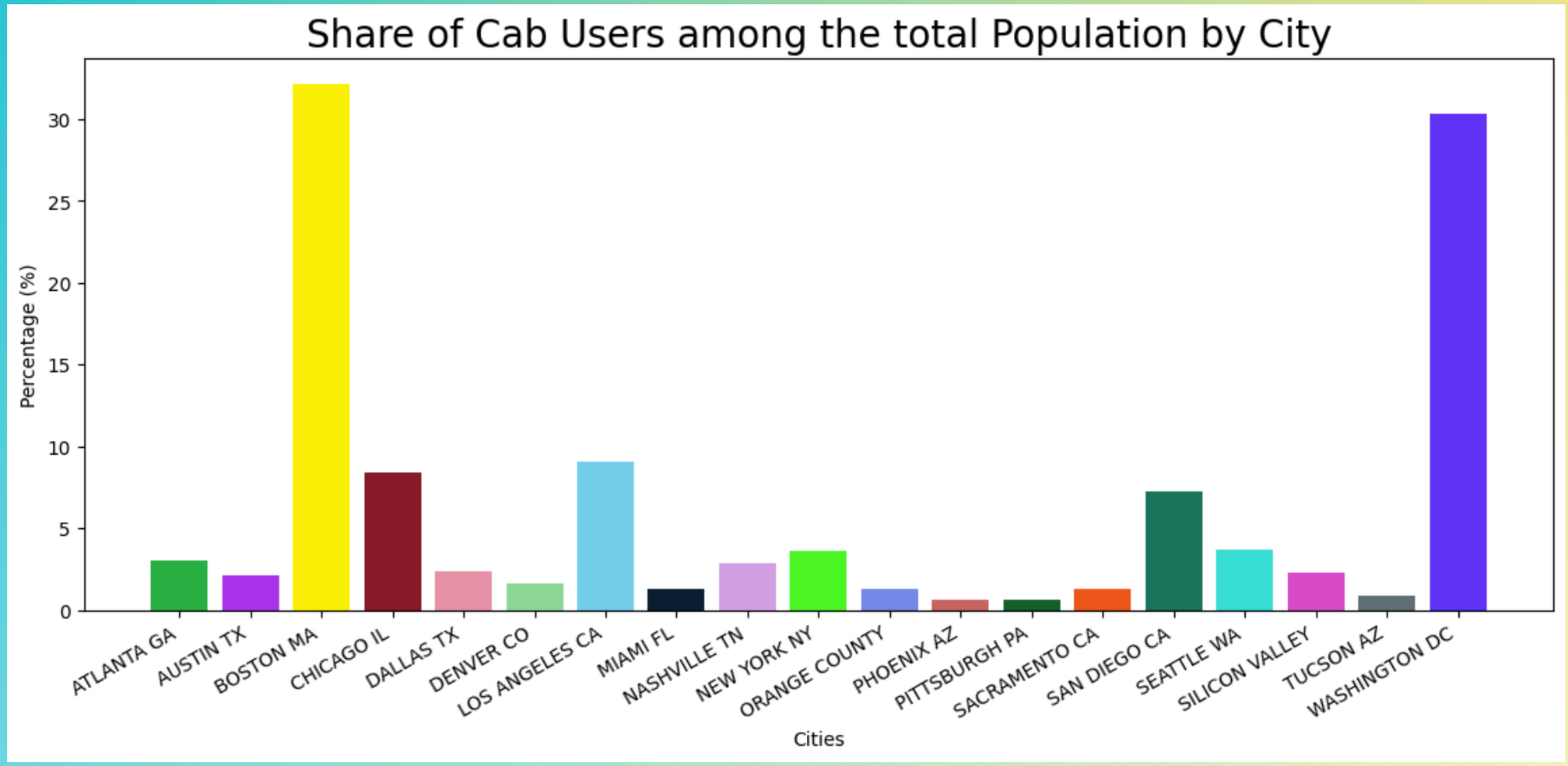
35 - is an average Age of Female and Male who use Cab services

03. GENDER DISTRIBUTION



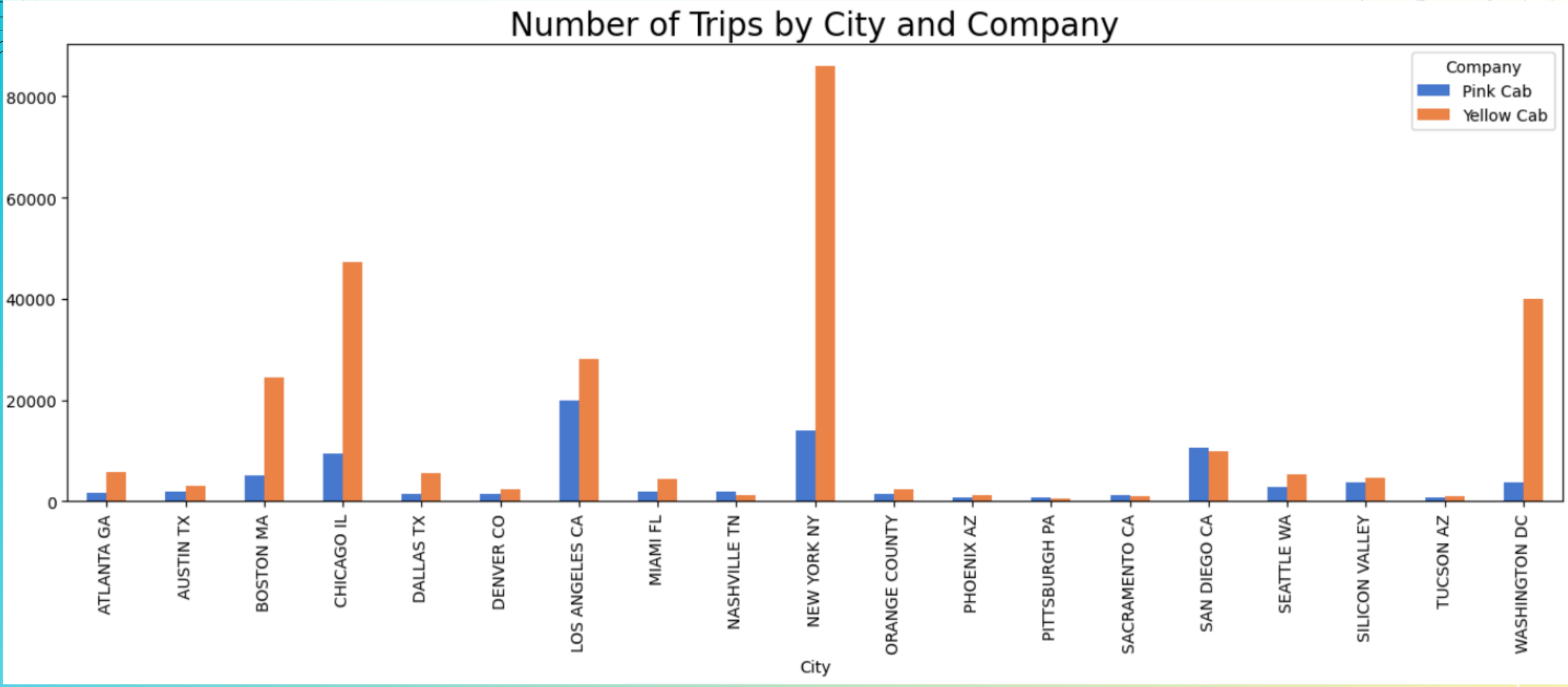
The ratio of men and women is similar, with a slight predominance of men.

03. SHARE OF CAB USERS AMONG THE TOTAL POPULATION BY CITY



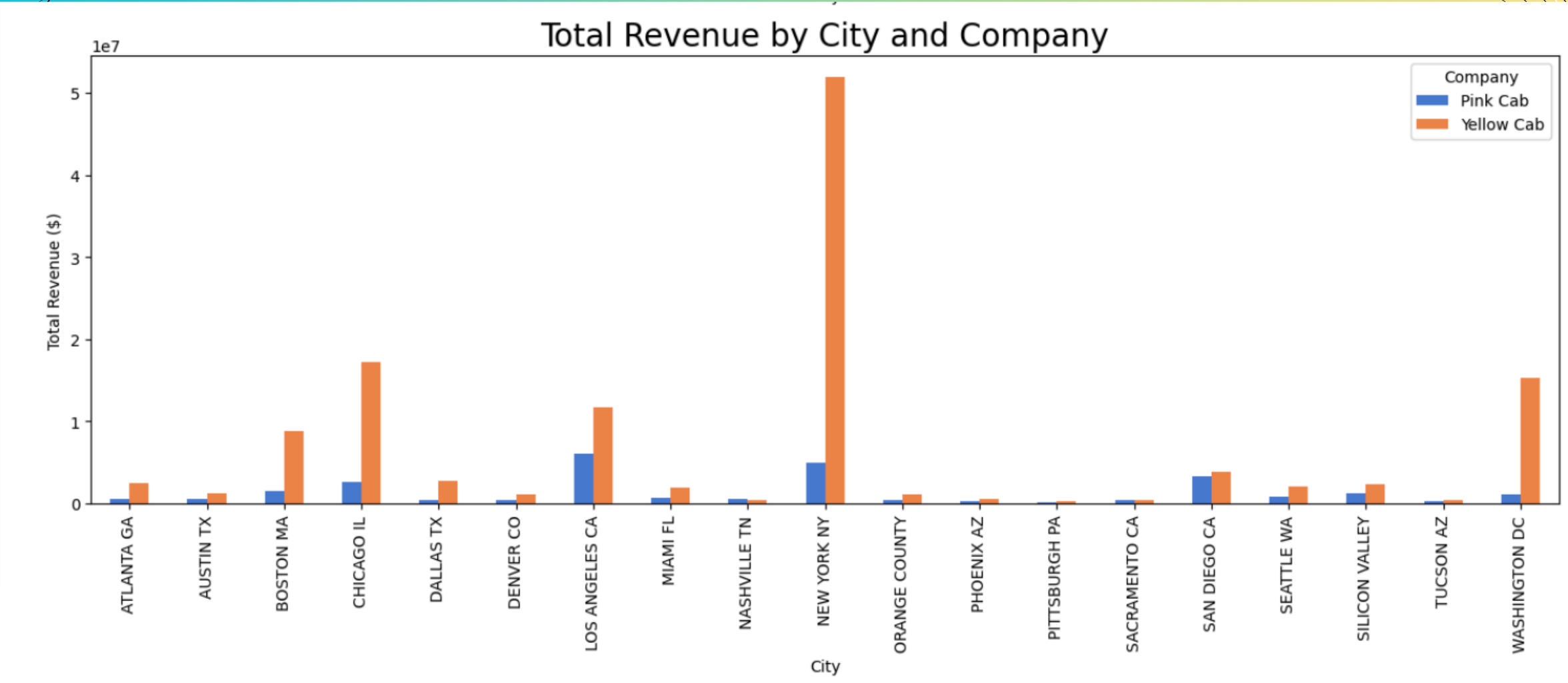
In Washington and Boston more than 30% of population use Cab services

04. NUMBER OF TRIPS BY CITY AND COMPANY



Yellow Cab is the leader in most cities in terms of the number of trips. (hypothesis 4)

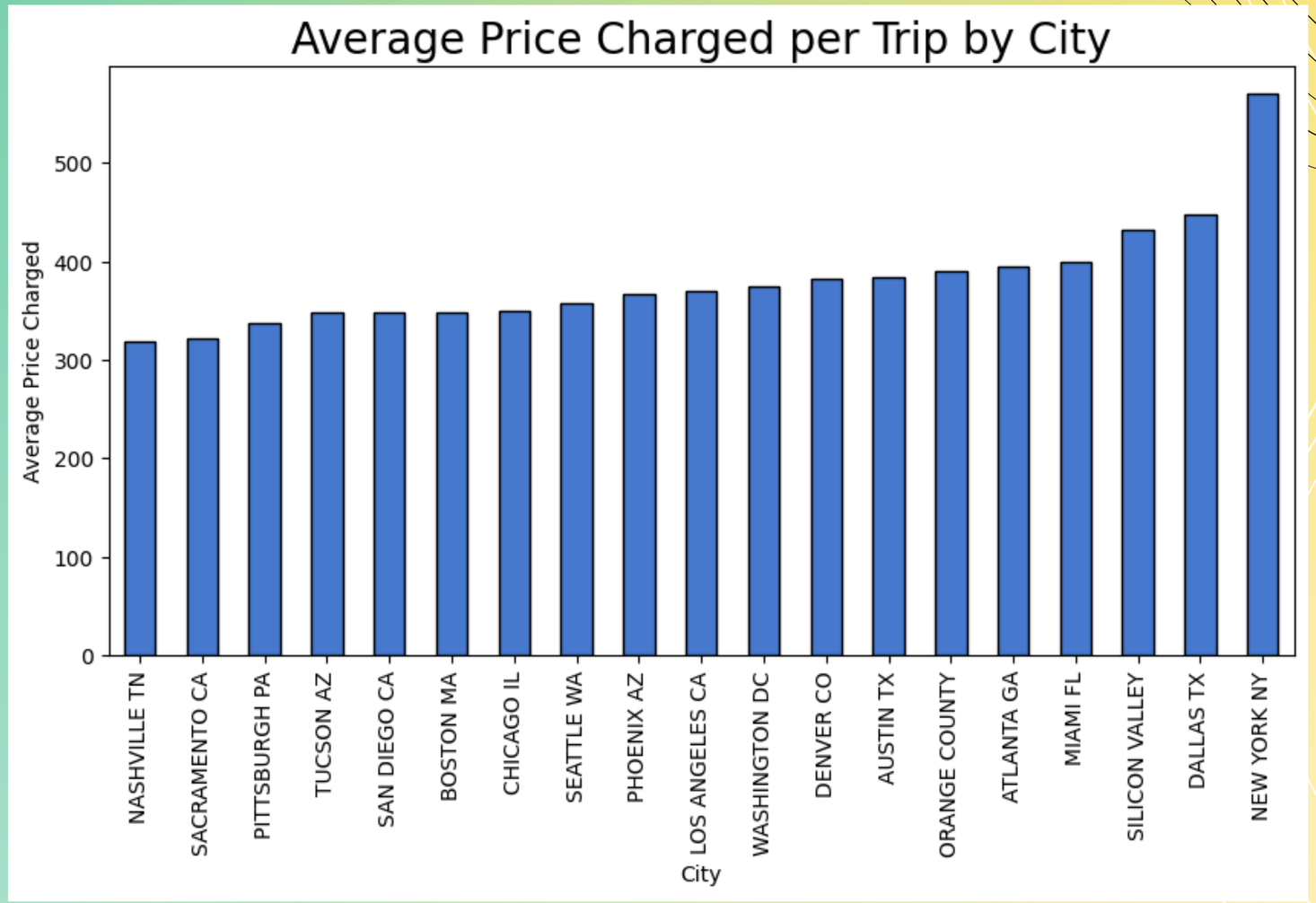
04. TOTAL REVENUE BY CITY AND COMPANY



Yellow Cab also generates more revenue in most cities.

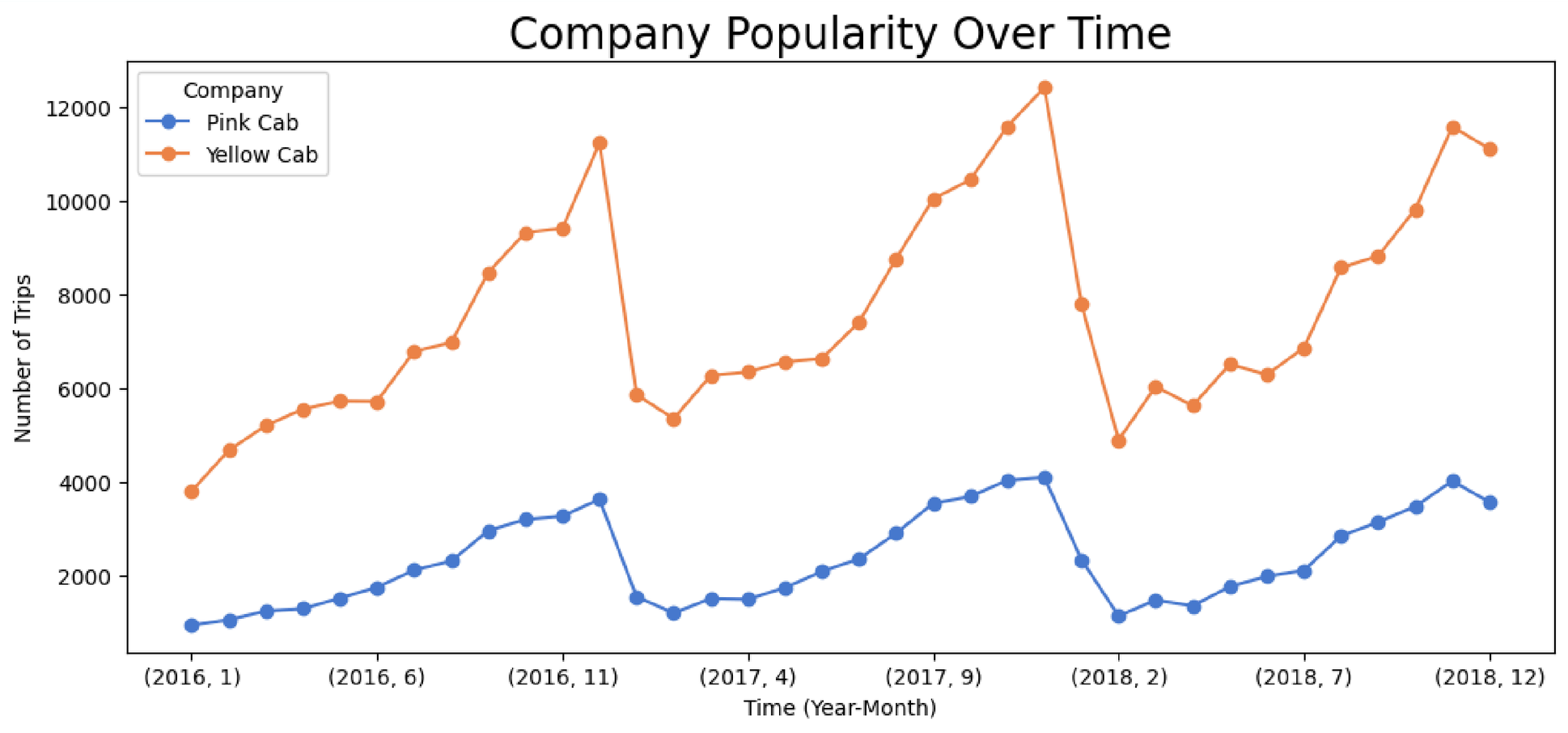
04. AVERAGE PRICE CHARGED PER TRIP BY CITY

There are differences in the average cost of trips between different cities (hypothesis 8).



ADDITIONAL INSIGHTS FROM THE DATA:

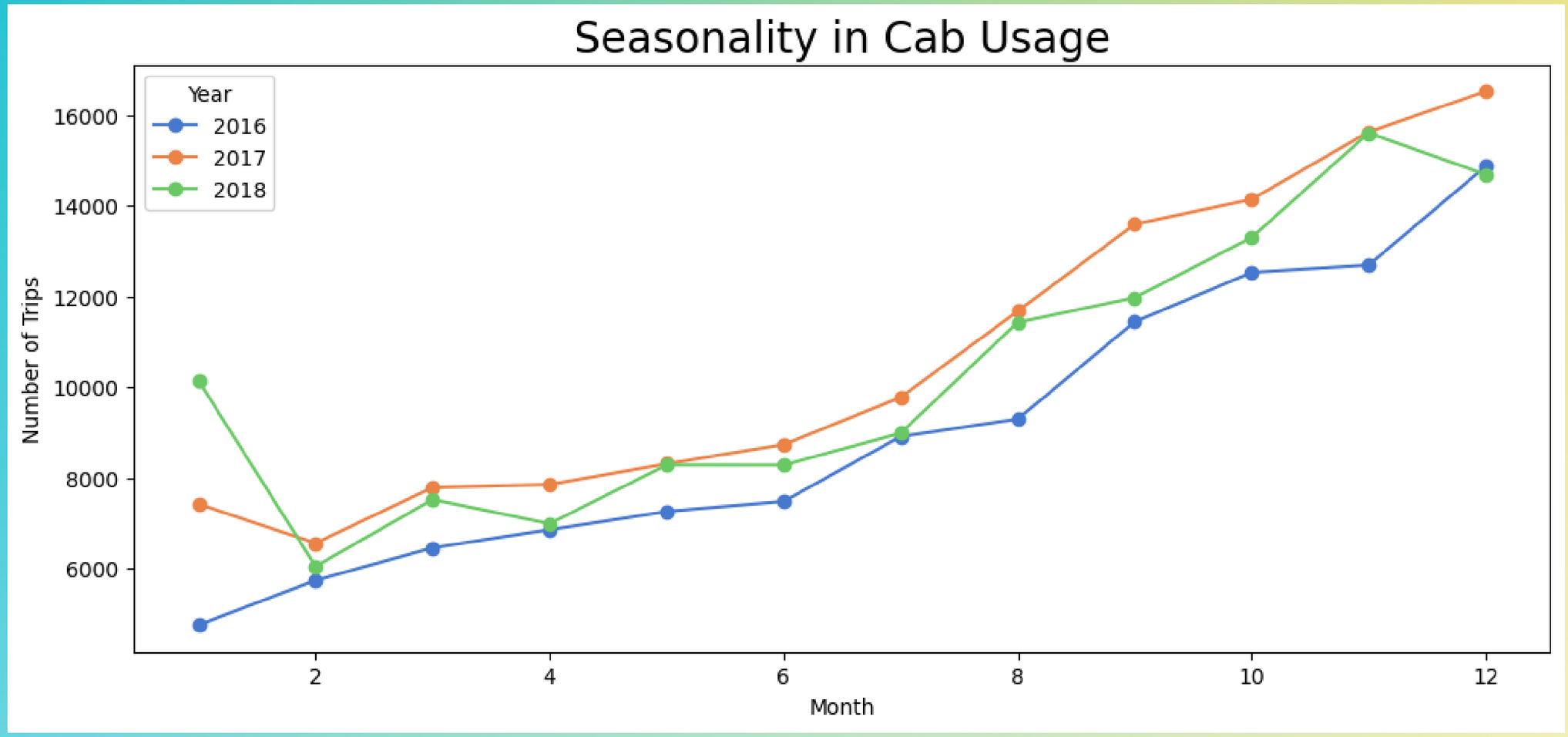
COMPANY POPULARITY OVER TIME



Yellow Cab generally has a higher number of trips compared to Pink Cab **over time.**

ADDITIONAL INSIGHTS FROM THE DATA:

SEASONALITY IN CAB USAGE



The seasonality analysis shows a **clear trend in cab usage over the months**. There are noticeable peaks and troughs, indicating that cab usage varies throughout the year.

06
HYPOTHESIS TESTING RESULTS

HYPOTHESIS 1:

IS THERE ANY DIFFERENCE IN PROFIT BASED ON GENDER

HO : There is no difference in profit regarding gender.
H1: There is a difference in profit regarding gender.

Yellow Cab:

There is a difference in profit regarding gender.

```
Yellow Cab:
size a: 158681, size b: 116000
t-statistic is 10.315494207195322
p-value is 6.060473042494056e-25
We accept alternative hypothesis (H1) for Yellow Cab
```

Pink Cab:

There is no difference in profit regarding gender.

```
Pink Cab:
size a: 47231, size b: 37480
t-statistic is 1.5754642478511207
p-value is 0.115153059004258
We accept null hypothesis (H0) for Pink Cab
```

HYPOTHESIS 2:

IS THERE ANY DIFFERENCE IN PROFIT BASED ON PAYMENT MODE

H0 : There is no difference in profit regarding payment mode.
H1: There is a difference in profit regarding payment mode.

Yellow Cab:

There is no difference in profit regarding payment mode.

```
Yellow Cab
size a: 109896, size b: 164785
t-statistic is 1.050898652643264
p-value is 0.29330606382987284
We accept null hypothesis (H0) for Yellow Cab
```

Pink Cab:

There is no difference in profit regarding payment mode.

```
Pink Cab
size a: 33992, size b: 50719
t-statistic is -0.26625096474899373
p-value is 0.7900465828793286
We accept null hypothesis (H0) for Pink Cab
```

HYPOTHESIS 3:

IS THERE ANY DIFFERENCE IN REVENUE BASED ON AGE GROUP

H0 : There is no difference in revenue regarding age group.
H1: There is a difference in revenue regarding age group.

Yellow Cab:

There is no difference in revenue regarding age group.

```
Yellow Cab
size a: 88163, size b: 86550
t-statistic is -0.8439046674773827
p-value is 0.3987238352474821
We accept null hypothesis (H0) for Yellow Cab
```

Pink Cab:

There is no difference in revenue regarding age group.

```
Pink Cab
size a: 27203, size b: 26556
t-statistic is 0.85122795620289
p-value is 0.3946465277392325
We accept null hypothesis (H0) for Pink Cab
```

HYPOTHESIS 4:

IS THERE ANY DIFFERENCE IN NUMBER OF TRIPS BASED ON CITY

H0 : There is no difference in the number of trips regarding city.
H1: There is a difference in the number of trips regarding city.

Yellow Cab:

There is a difference in the number of trips regarding city.

```
t-statistic is 22.553164620006115
p-value is 6.012869365786724e-75
We accept alternative hypothesis (H1) for Yellow Cab
```

Pink Cab:

There is a difference in the number of trips regarding city.

```
t-statistic is 9.321498670608708
p-value is 2.6442305400816e-26
We accept alternative hypothesis (H1) for Pink Cab
```

HYPOTHESIS 5:

IS THERE ANY DIFFERENCE IN DISTANCE TRAVELLED BASED ON AGE GROUP

H0 : There is no difference in the distance travelled regarding age group.
H1: There is a difference in the distance travelled regarding age group.

Yellow Cab:

There is no difference in the distance travelled regarding age group.

```
Yellow Cab
size a: 88163, size b: 86550
t-statistic is -0.21982562866317218
p-value is 0.8260072158421701
We accept null hypothesis (H0) for Yellow Cab
```

Pink Cab:

There is no difference in the distance travelled regarding age group.

```
Pink Cab
size a: 27203, size b: 26556
t-statistic is 0.5696792188764248
p-value is 0.5688976678230542
We accept null hypothesis (H0) for Pink Cab
```

HYPOTHESIS 6:

IS THERE ANY DIFFERENCE IN INCOME OF CUSTOMERS BASED ON COMPANY

HO : There is no difference in the income of customers regarding the cab company.

H1: There is a difference in the income of customers regarding the cab company.

```
size a: 274681, size b: 84711
t-statistic is -0.427112697888899964
p-value is 0.66929750057506659
We accept null hypothesis (H0) that there is no difference in the income of customers between the companies
```

There is no difference in the income of customers regarding the cab company.

HYPOTHESIS 7:

IS THERE ANY DIFFERENCE IN DISTANCE TRAVELLED BASED ON GENDER

HO : There is no difference in the distance travelled regarding gender.
H1: There is a difference in the distance travelled regarding gender.

Yellow Cab:

There is no difference in the distance travelled regarding gender.

```
Yellow Cab
size a: 158681, size b: 116000
t-statistic is -0.651591243962616
p-value is 0.5146654429411317
We accept null hypothesis (H0) for Yellow Cab
```

Pink Cab:

There is no difference in the distance travelled regarding gender.

```
Pink Cab
size a: 47231, size b: 37480
t-statistic is -0.5008718498401431
p-value is 0.6164626165258722
We accept null hypothesis (H0) for Pink Cab
```

HYPOTHESIS 8:

IS THERE ANY DIFFERENCE IN AVERAGE PRICE CHARGED PER TRIP BY CITY

HO : There is no difference in Average Price Charged per Trip regarding City.
H1: There is a difference in Average Price Charged per Trip regarding City.

Yellow Cab:

There is a difference in Average Price Charged per Trip regarding City.

```
Yellow Cab
t-statistic is 2240.8998127101095
p-value is 0.0
We accept alternative hypothesis (H1) for Yellow Cab
```

Pink Cab:

There is a difference in Average Price Charged per Trip regarding City.

```
Pink Cab
t-statistic is 91.96667541583189
p-value is 0.0
We accept alternative hypothesis (H1) for Pink Cab
```

07
RECOMMENDATIONS

BASED ON THE ANALYSIS, THE FOLLOWING CONCLUSIONS CAN BE DRAWN:

1. Overall Company Performance:

Yellow Cab has higher performance in all key parameters:

- number of trips (+224.1% more than Pink Cab),
- total revenue (+378.1% more than Pink Cab),
- total costs (+289.4% more than Pink Cab),
- and most importantly, profit (+729.4% more than Pink Cab, or 8.29 times more).

BASED ON THE ANALYSIS, THE FOLLOWING CONCLUSIONS CAN BE DRAWN:

2. Average Performance Per Trip:

Yellow Cab generates more profit per trip, despite higher costs, which makes investments in Yellow Cab more attractive in terms of profitability:

- average revenue per trip (+47.4% more than Pink Cab),
- average costs per trip (+20.0% more than Pink Cab),
- average profit per trip (+155.8% more than Pink Cab, or 2.5 times more).

BASED ON THE ANALYSIS, THE FOLLOWING CONCLUSIONS CAN BE DRAWN:

3. Customer Base:

Both companies have a similar distribution by age and gender of customers, however, *Yellow Cab customers have a slightly lower median income* (-0.09% or 14$).

4. Geographical Distribution:

Yellow Cab leads in most cities in terms of number of trips and revenue, which shows wider market coverage and potential for further growth.

5. Seasonal Trends:

Both companies show seasonal fluctuations, but *Yellow Cab shows more stable growth*.

BASED ON THE ABOVE CONCLUSIONS, **INVESTING IN YELLOW CAB** SEEMS TO BE A MORE REASONABLE AND PROMISING DECISION COMPARED TO PINK CAB.


Data Glacier
Your Deep Learning Partner
THANK YOU!