



**Data Glacier**

Your Deep Learning Partner

# Exploratory Data Analysis

<G2M insight for Cab Investment firm>

<15th October, 2024>

# Background –G2M(cab industry) case study

---

- XYZ is a private equity firm based in the U.S. It is considering an investment in the cab industry due to its impressive growth over recent years and the presence of several key competitors in the market.
- **Objective:** To provide actionable insights that will assist XYZ in identifying the most suitable company for investment. (*Yellow or Pink Cab Company*)

The analysis is structured into *four sections*:

- Data Preparation, Understanding and Exploration
- Data Analysis: Identifying the most profitable cab company
- Build a model to predict profits
- Investment recommendations

# Data Preparation, Understanding and Exploration

---

- The dataset consists of **17 features and 35 derived features**.
- The data spans from **31/01/2016 to 31/12/2018**.
- The total number of data points is **359392**.
- The given four files have been used for the analysis.

## **\*\*Assumptions:\*\***

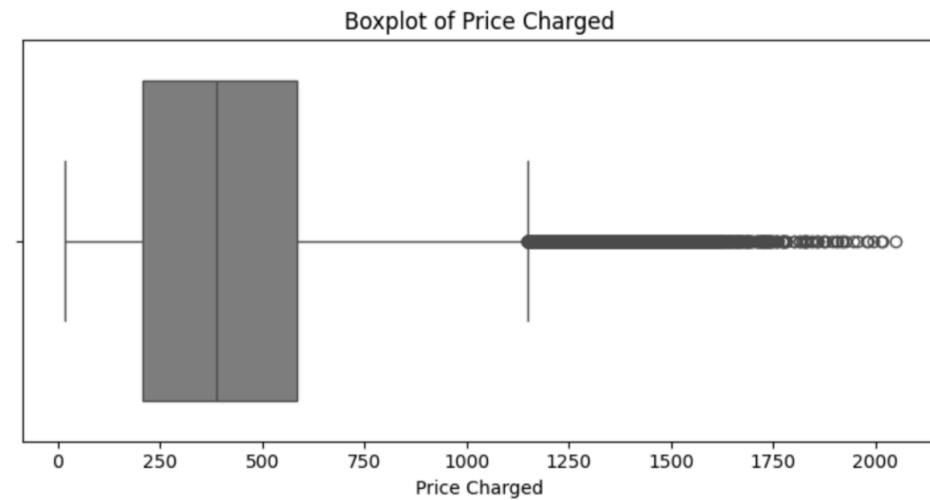
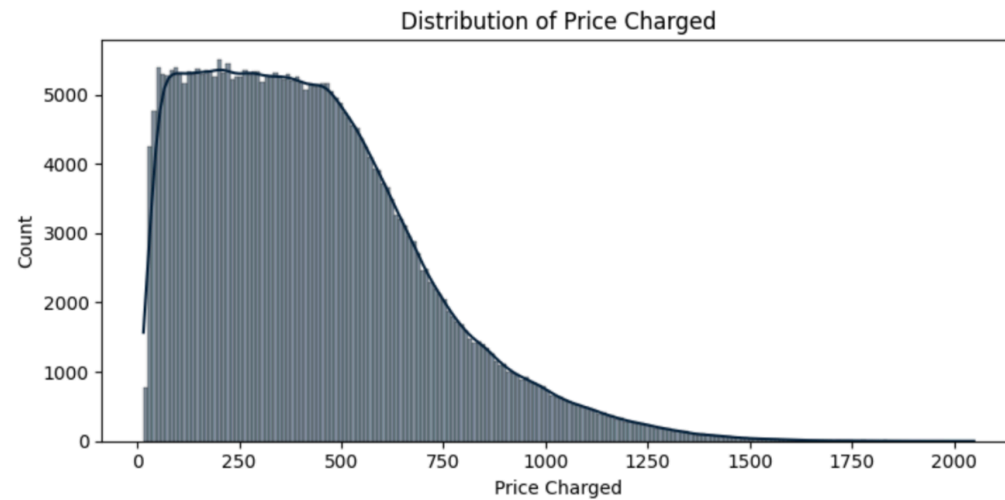
- Since an MLR regression model is attempted to be built, assumptions around No Multicollinearity, Homoscedasticity, No pattern in error and Error normally distributed around mean 0 are made.
- Outliers are present in the Price\_Charged feature; however, further study is to be conducted to understand if these are valid outliers.
- Profit from rides is calculated by holding other factors constant, using only the Price\_Charged and Cost\_of\_Trip features for profit calculations.
- The "Users" feature in the city dataset is interpreted as the number of cab users in that city. We assume this includes users of other cab services, such as Yellow and Pink cabs. The following is used to calculate the proportion of users using Users and Population considering other factors as constant.
- Based on the availability of cabs, considering Yellow Cab Company is older than the Pink Cab Company (in reality, one must study their start date and analyse their growth graph to truly gauge their numbers and spread.)

## **\*\*Data Preparation:\*\***

Most work is conducted using Python on the .ipynb notebook, slight modifications have been made using Excel.

# Data Analysis

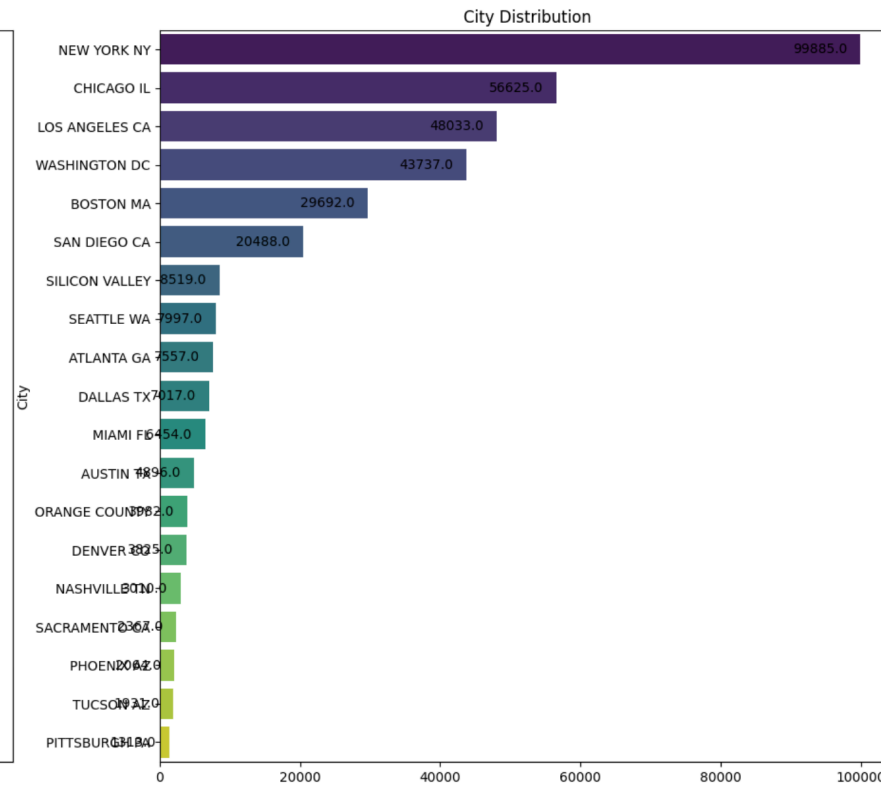
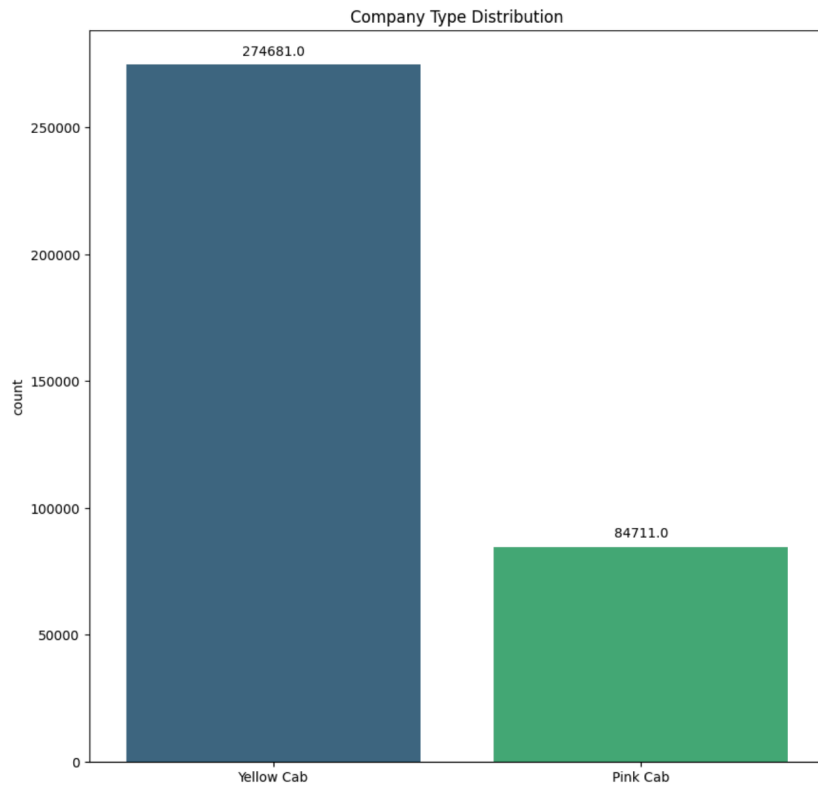
Price Charged  
Skew : 0.87



## Exploration Scope:

It is important to understand that the Price variable is heavily skewed. However, it is common to have many lower-paying trips and comparatively fewer higher-paying trips. Further, in the analysis, we shall explore this subject more.

## • Data Analysis: II



### Gender (Total)

*Female*

2.046195e+07

*Male*

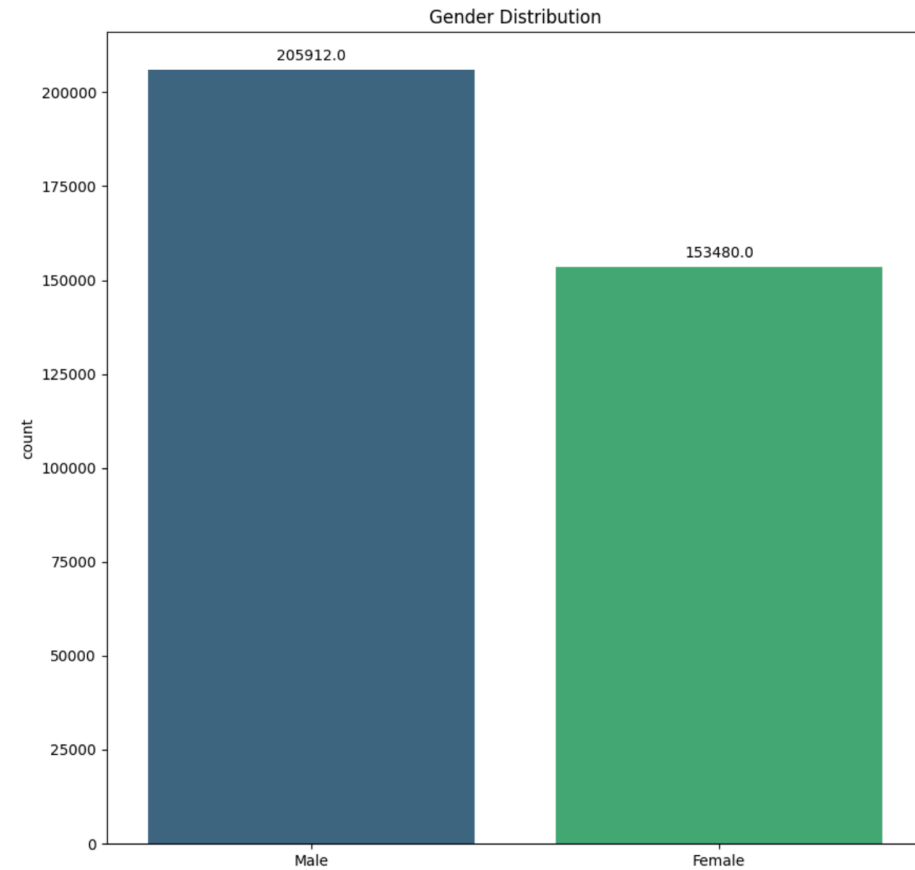
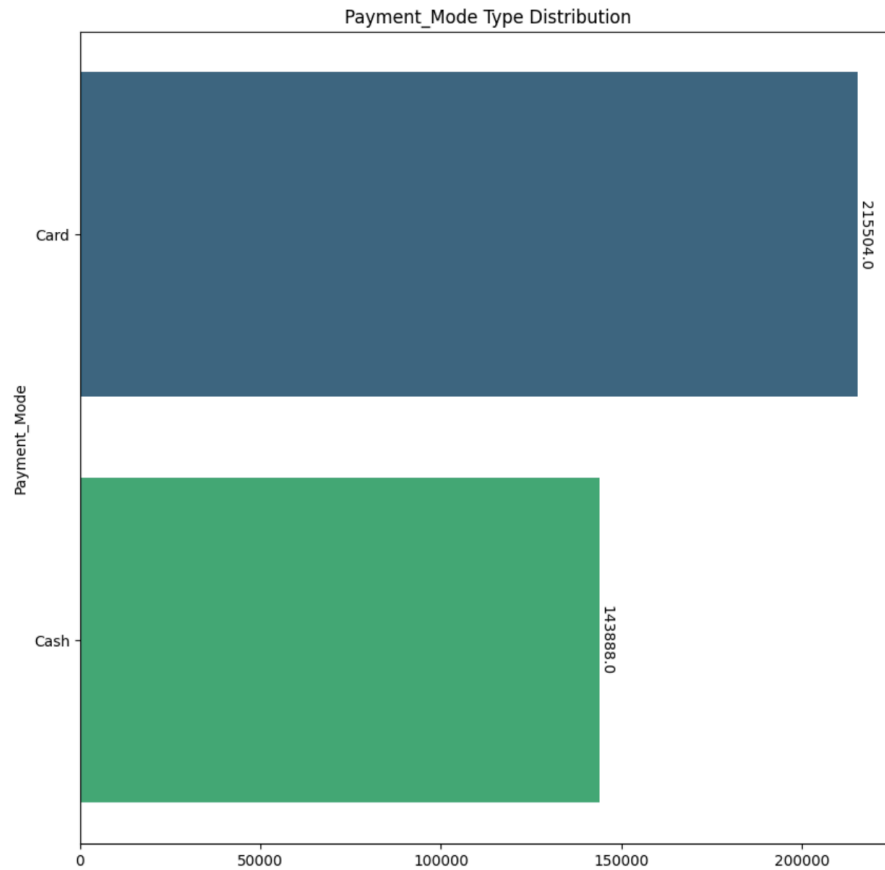
2.886575e+07

*For reference:*

City	Users_prop
NEW YORK NY	35.945141
CHICAGO IL	84.121260
LOS ANGELES CA	90.362794
MIAMI FL	13.198622
SILICON VALLEY	23.137561

- The above plots suggest that the **Yellow cab** company has **more** cabs (274681) compared to the Pink cab company (84711).
- Most **widely preferred** across **New York**, followed by Chicago, Los Angeles and so on to **least preferred** in **Pittsburgh**.
- Another important takeaway point is that one must consider the proportion of users compared to the number for it greatly depends on the population size. A high proportion indicates that out of X people, so many of them use cabs for transportation building clientele.

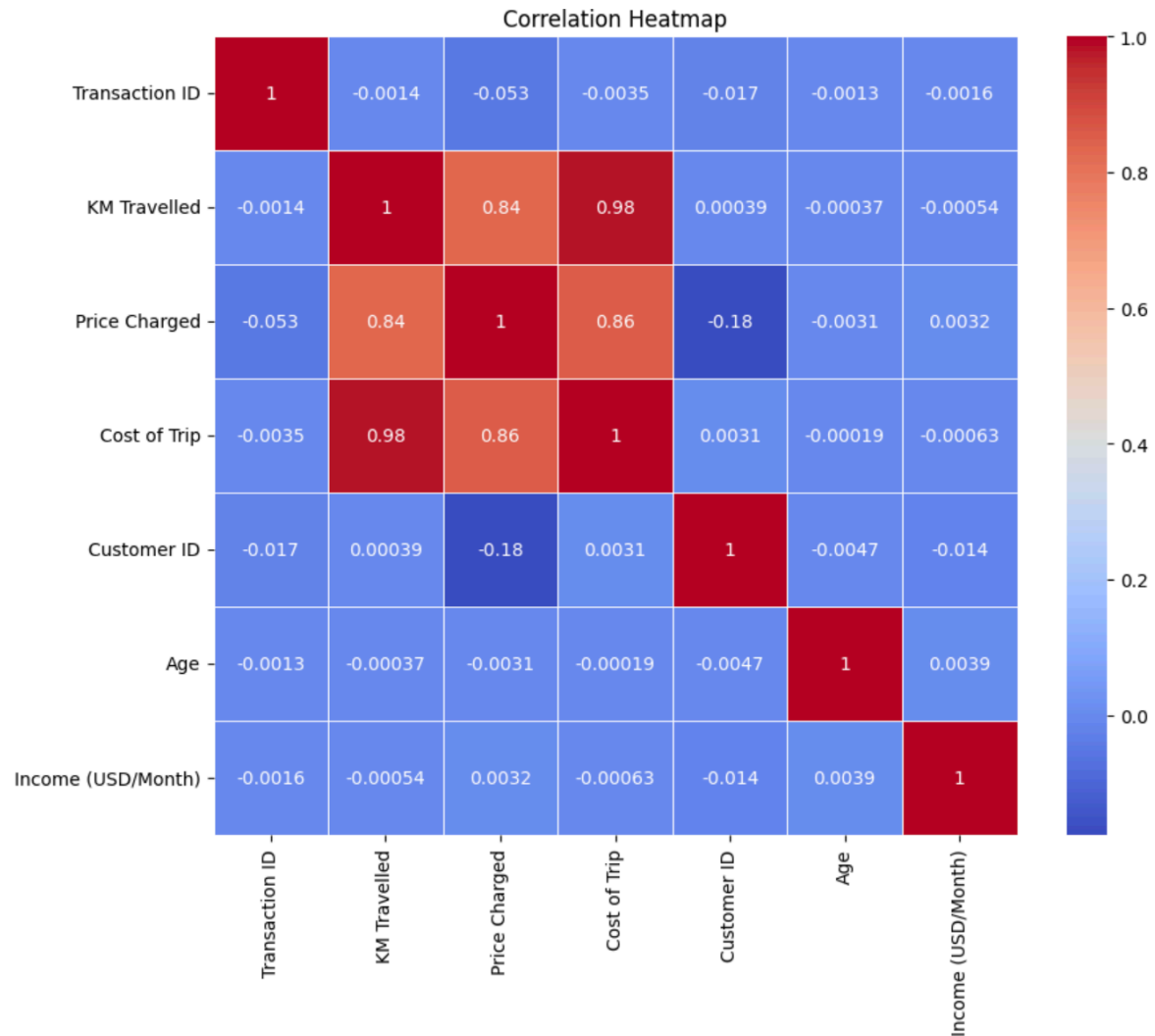
# • Data Analysis: III



The above plots suggest that in general, the **better-preferred way to make payments by customers is Card compared to Cash.**

Also, more men utilise the service in comparison to women.

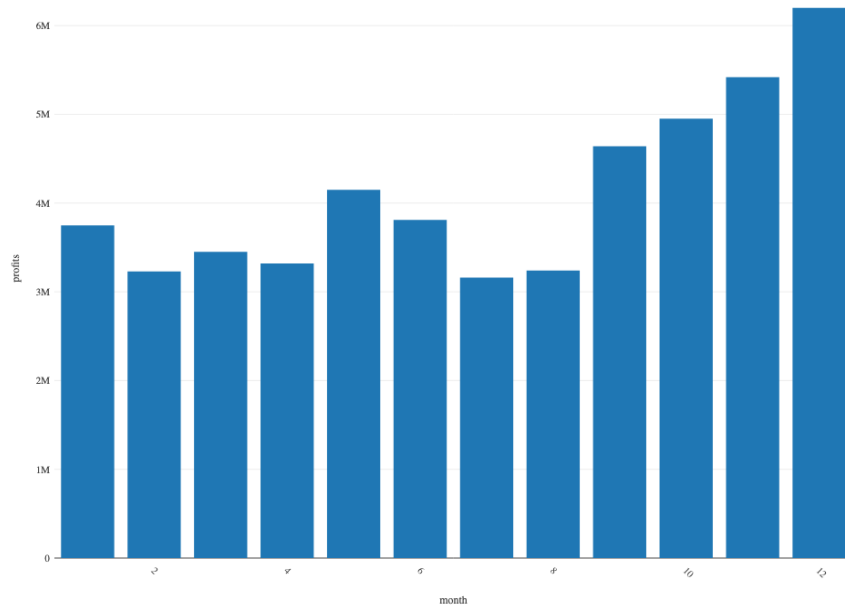
## • Data Analysis: IV



As expected, a very strong correlation can be seen between Kilometers travelled, Price Charged, and Cost of Trips incurred. As **one-factor** increases, the **other factor** increases.

# Data Analysis: identifying the most profitable cab company

## Understanding the Monthly Profits Made



month	
1	3.746490e+06
2	3.228020e+06
3	3.450311e+06
4	3.321361e+06
5	4.152157e+06
6	3.813054e+06
7	3.163447e+06
8	3.244522e+06
9	4.636818e+06
10	4.946949e+06
11	5.419647e+06
12	6.204925e+06

1. Can **CLIMATE** influence cab users?

2. Are **HOLIDAY SEASONS** and **TOURISM** factors causing increased cab usage?

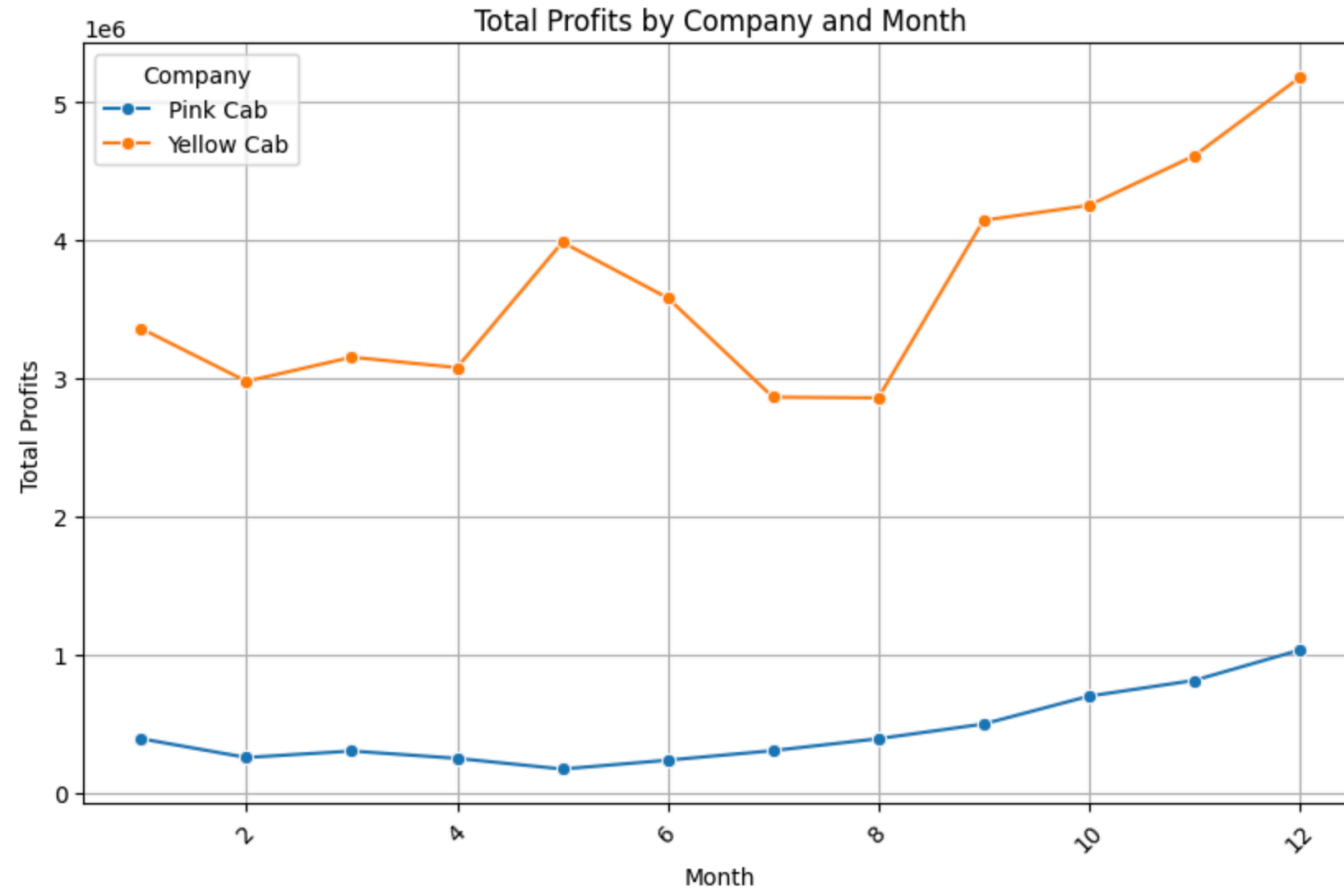
The months of May (Summers) and those following September to December (Winters) suggest more profits.

A relationship with extreme hot or cold conditions can be studied and also, these months are prone to more travelling, and therefore frequent use of cab services?



# Data Analysis: identifying the most profitable cab company

Understanding the **Monthly Profits** Made **Company Wise**



The **YELLOW CAB COMPANY** is seen to make more profits across all months of all years.

**year Company**

2016 Pink Cab 1.713511e+06

**Yellow Cab 1.392700e+07**

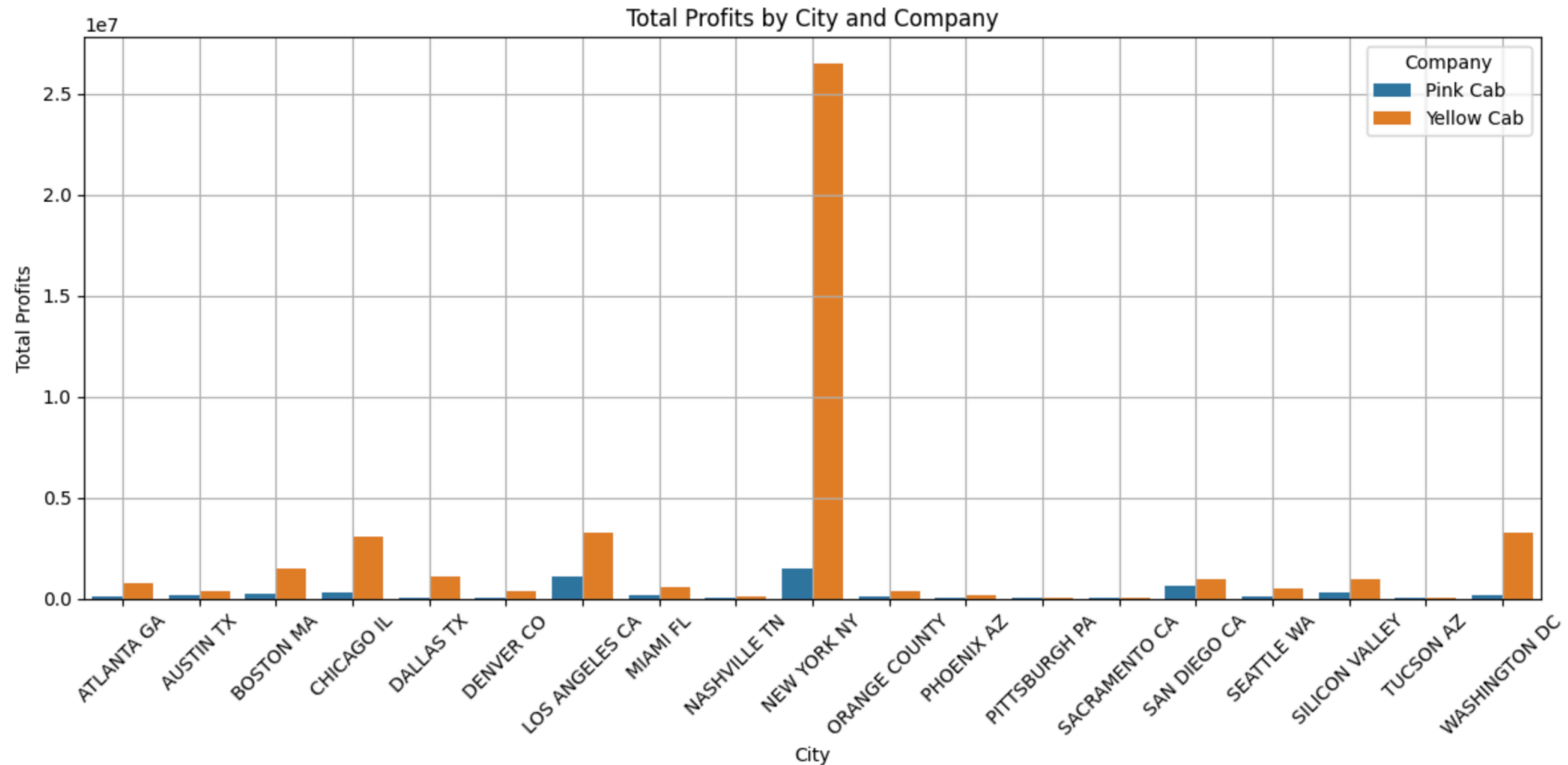
2017 Pink Cab 2.033655e+06

**Yellow Cab 1.657598e+07**

2018 Pink Cab 1.560162e+06

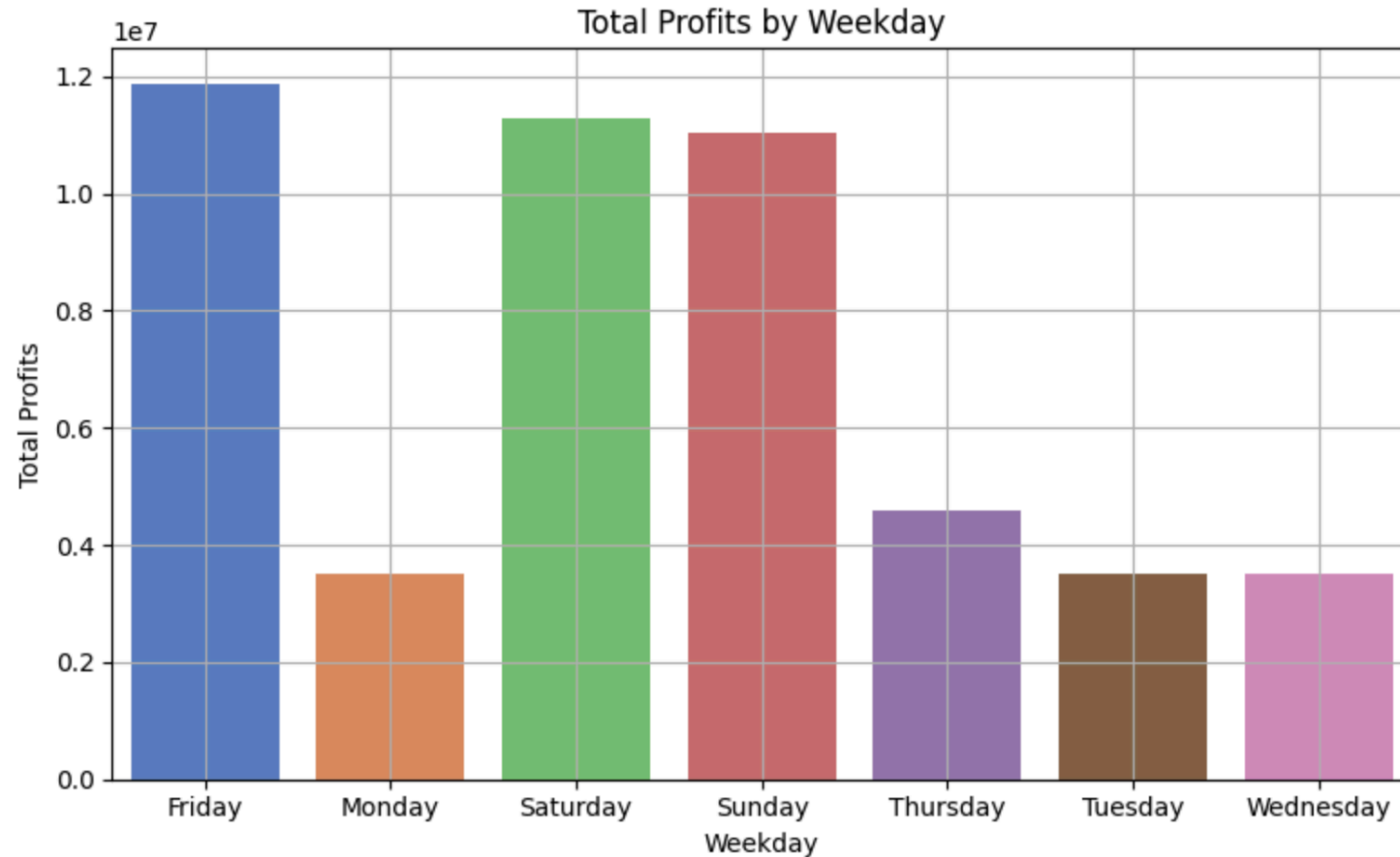
**Yellow Cab 1.351740e+07**

# Data Analysis: identifying the most profitable cab company



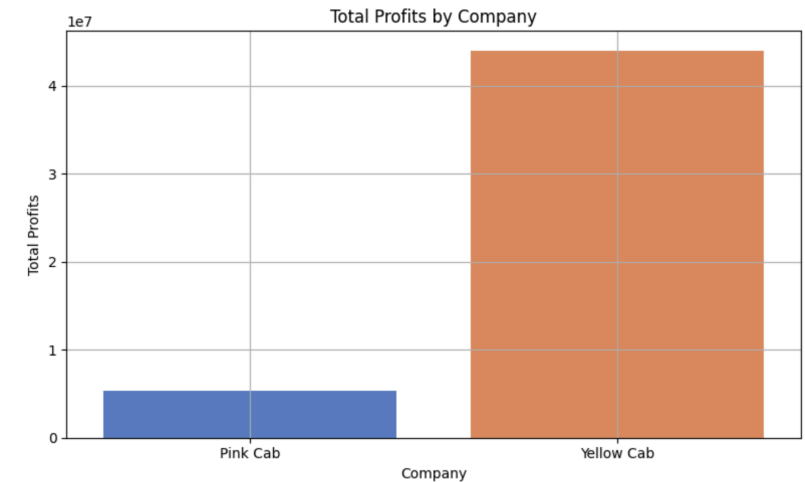
The **YELLOW CAB COMPANY** is most preferred in a city where cabs are frequently used and available.

# Data Analysis: identifying the most profitable cab company



**Weekends show peaking in users- Friday, Saturday, Sunday.**

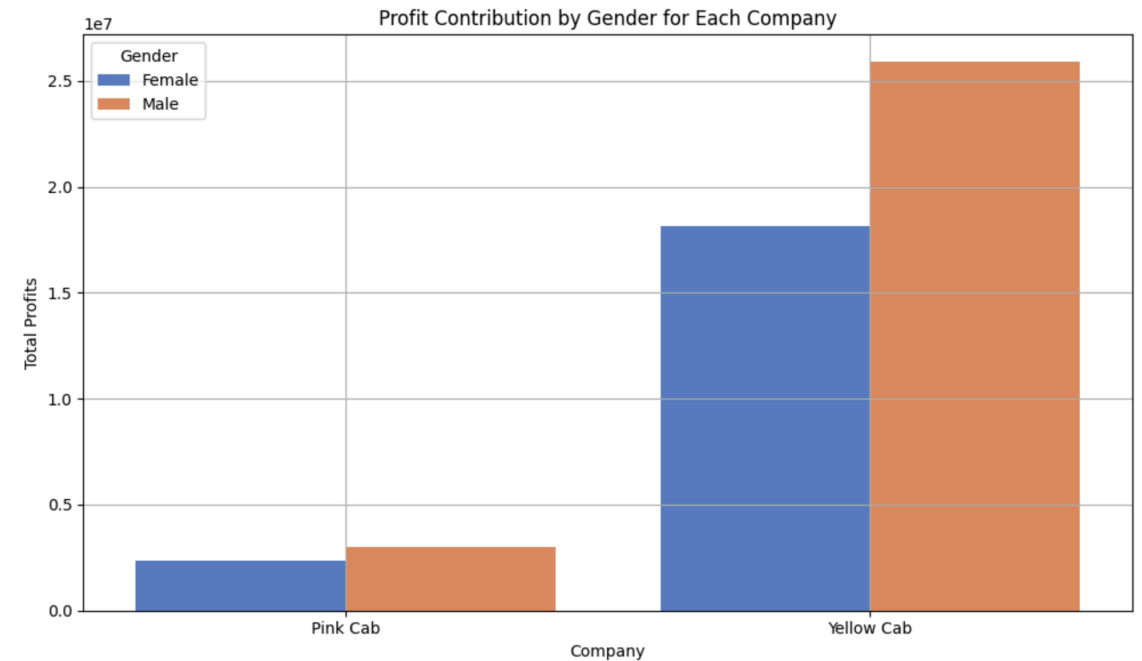
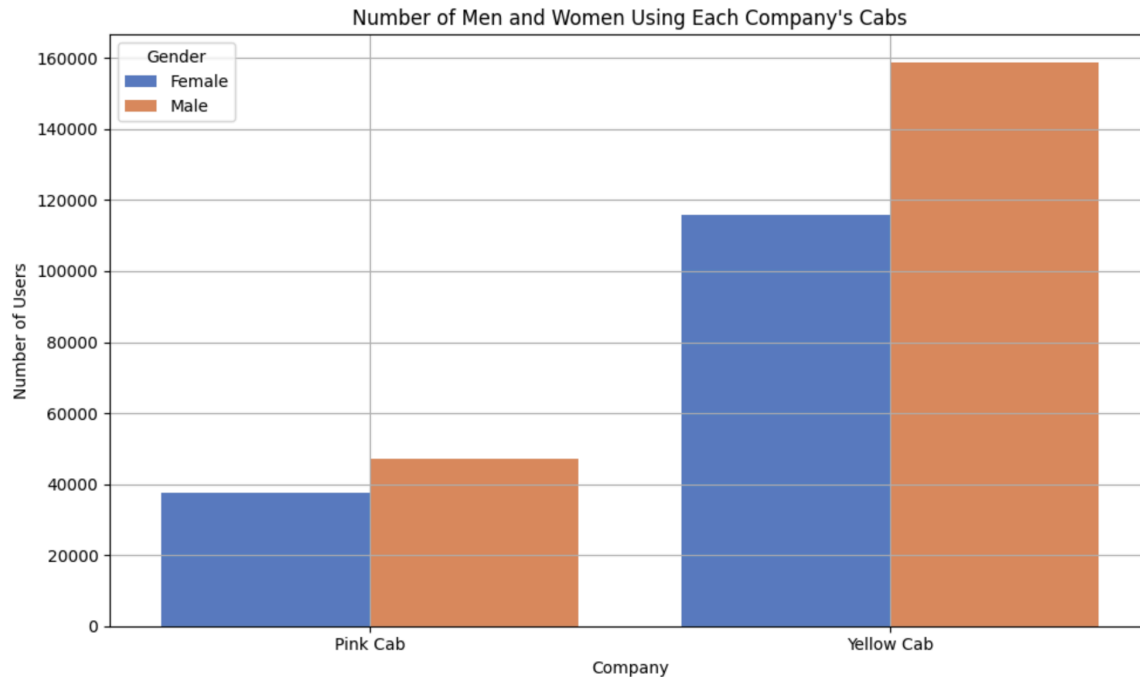
Does this mean the Yellow cab company has better accessibility and network?



**The numbers suggest YES!**

# Data Analysis: identifying the most profitable cab company

## GENDER and PROFIT Analysis



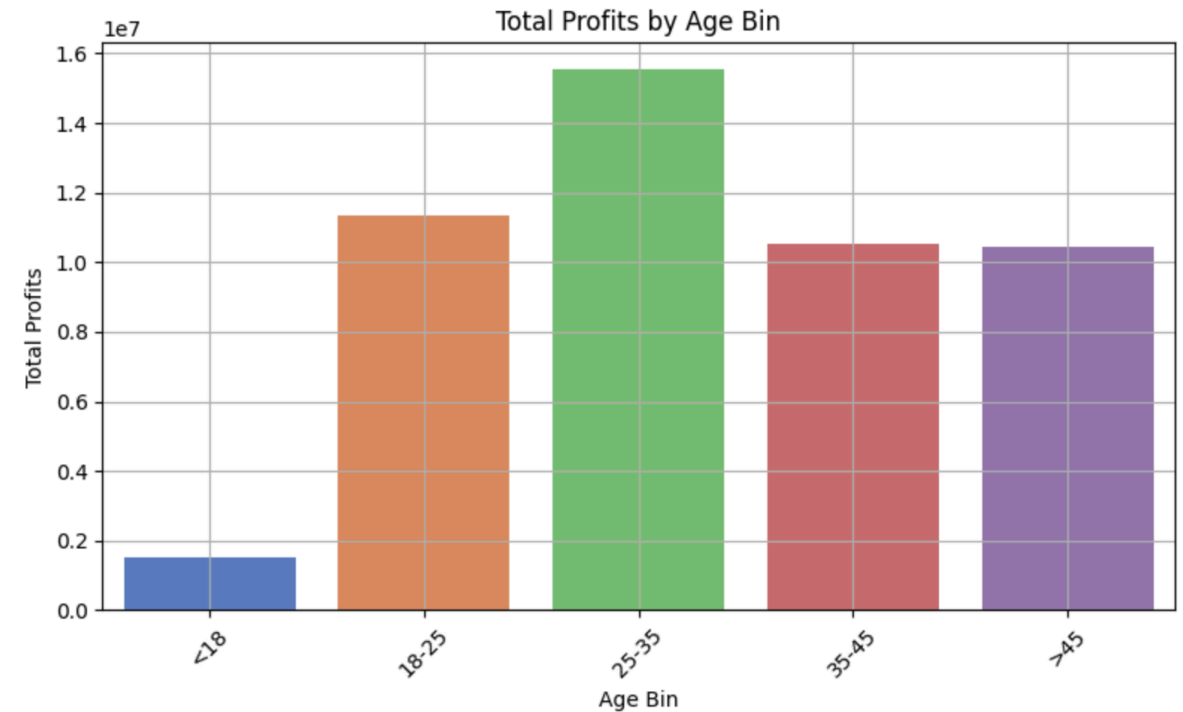
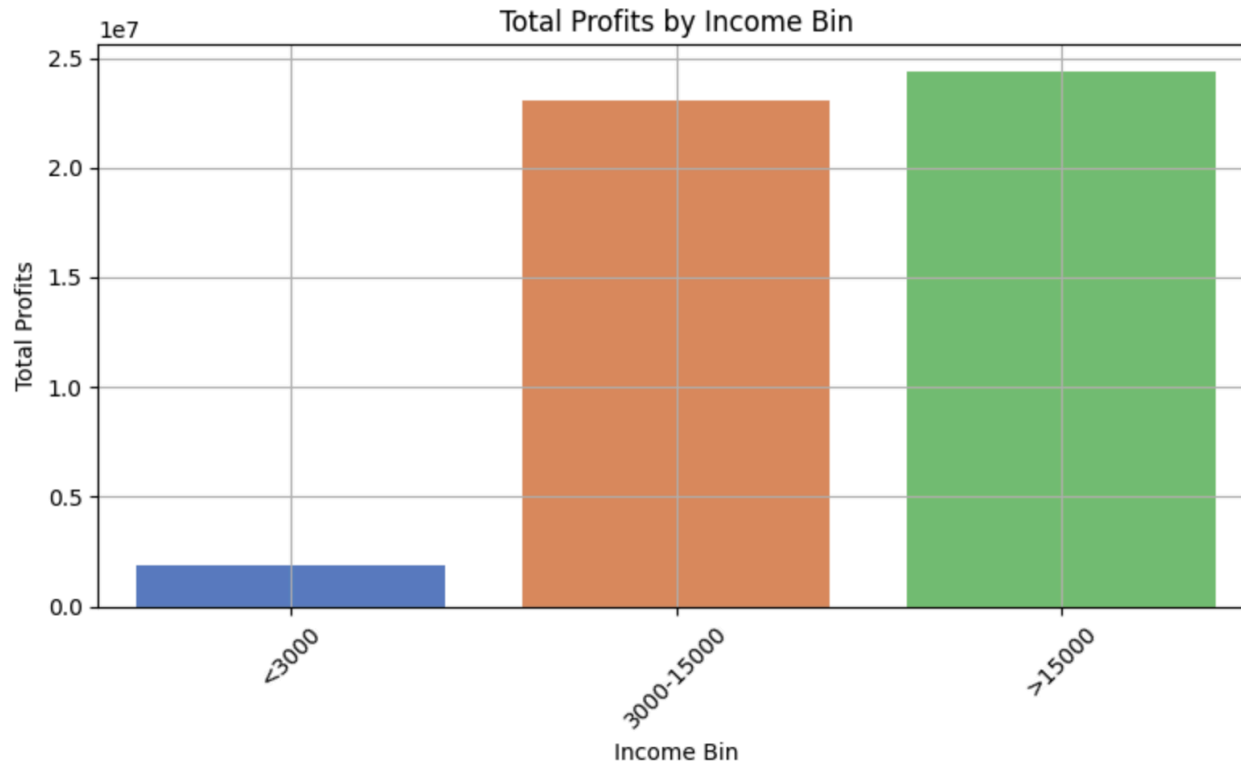
GENDER  
COUNT

The numbers suggest that in general, **profits are more drawn in by men** and overall stats show that **Yellow Cab Company benefits more** than the Pink Cab Company.

PROFITS  
MADE

# Data Analysis: identifying the most profitable cab company

## UNDERSTANDING THE ROLE OF INCOME AND AGE ON PROFIT



As expected, people who **earn more spend more** on cabs.  
For age groups, the **spending of those between 18-35 is more**. This could be associated with their earning power, commute needs, work travels etc.

# Data Analysis: identifying the most profitable cab company

Are there any regulars for these cab companies?

---

Company	Company	Customer ID	
Pink Cab	Pink Cab	<b>8120</b>	<b>18</b>
		6159	17
		7927	17
		8595	17
		7340	16
		8474	16
		8915	16
		6416	15
		7764	15
		7938	15
Yellow Cab	Yellow Cab	<b>494</b>	<b>47</b>
		1360	47
		1803	47
		636	46
		126	45
		903	45
		2766	45
		992	44
		1070	44
		1673	44

While both companies are seen to retain customers, Yellow Cab company does a better job. The same customer has vouched for their service by booking rides 47 times and consistency is seen in numbers that follow the highest number.

# Data Prediction

## Can profits be predicted?

A model in the making!

To help predict profits, a predictive model can be built as shown as part of the notebook considering gender, income, age, geography and more.

More factors such as time of day, purpose of travel, climatic conditions and others can be considered by collecting relevant data for these can highly impact the profits.

# Recommendations

---

Upon evaluation of both cab companies based on the following factors and concluded that **Yellow Cab** *outperforms* Pink Cab:

- **Offering loyalty schemes:** As seen, the yellow cab company has better clientele and users prefer to rebook the ride. Better customer service can be offered by building loyalty schemes such as weekly offers, monthly passes, offers based on distance and more.
- **Understanding gender roles and attracting more women riders:** While it is understood that the population of men is more than women, men generally tend to own their transport and therefore cabs are in general more preferred by women. The trend of men continuing to book more rides must be investigated to offer better service to women and retain better numbers from their demographic.
- **Age-wise Reach:** While it is seen that adults prefer cab rides, the younger generation can be targeted and be made better offers, such as community pooling or college drop-off services at least in peak demand seasons.
- **Holiday Packages:** The company can identify peak periods like the months of May, and Sept-Dec and can plan more availability of cabs, better pricing deals and make more profits for the number of users also tends to grow.

Based on the findings, the Yellow Cab company is better preferred to invest. Furthermore, these recommendations must be taken into consideration as areas for improvement and discussion while investment options are weighed.



# Thank You

By  
Vuppu Sree Manasa