



Data Glacier

Your Deep Learning Partner

Exploratory Data Analysis

G2M Case Study

15 June - 2022

CONTENTS

1. Executive Summary
2. Problem Statement
3. Approach
4. EDA
5. Recommendations

Executive Summary – G2M(cab industry) case study

- XYZ is a private equity 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.
- The explanatory data analysis was done by combining all 4 csv data sets through the common columns. By combining these data sets, I was able to do an EDA to find out which cab company is a better investment for XYZ firm.
- I found that *Yellow cab* has a higher number of users in every city, and therefore the revenue was higher than the *Pink cab* in every city on average.
- In the year 2016, the profit generated by *Yellow cab* in the 19 cities, on average was higher than the profit generated by the *Pink cab*
- For 2017 and 2018, the general outcome was also the same
- The *Yellow cab* generated a higher revenue and profit each year. Therefore, our suggestion for XYZ company is to invest in the *Yellow cab*.

The analysis has been divided into four parts:

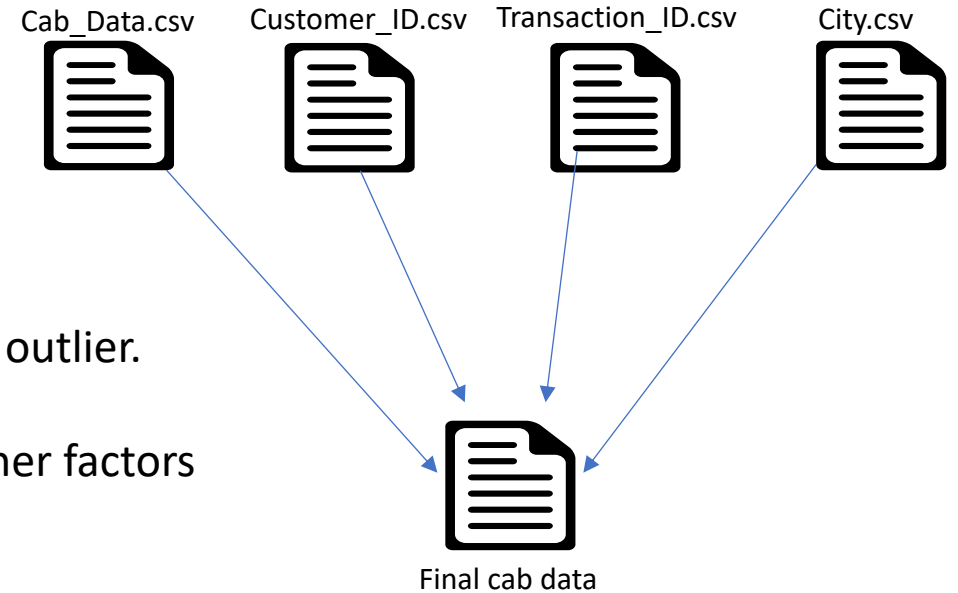
- Data Understanding
- Forecasting profit and number of users for each cab type
- Finding the most profitable Cab company
- Recommendations for investment

Data Exploration & Approach

- Rows & Columns: (359392, 15)
- Timeframe of the data: 2016-01-31 to 2018-12-31

Assumptions:

- Outliers are present in Price_Charged feature but due to unavailability of trip duration details ,we are not treating this as outlier.
- Profit of rides are calculated in the *Final Cab data* by keeping other factors constant and using *Price charged* and *Cost of trip*.
- The number of users is grouped in terms of year and the company to show the difference between *Yellow cab* and *Pink cab*.



Descriptive Statistics

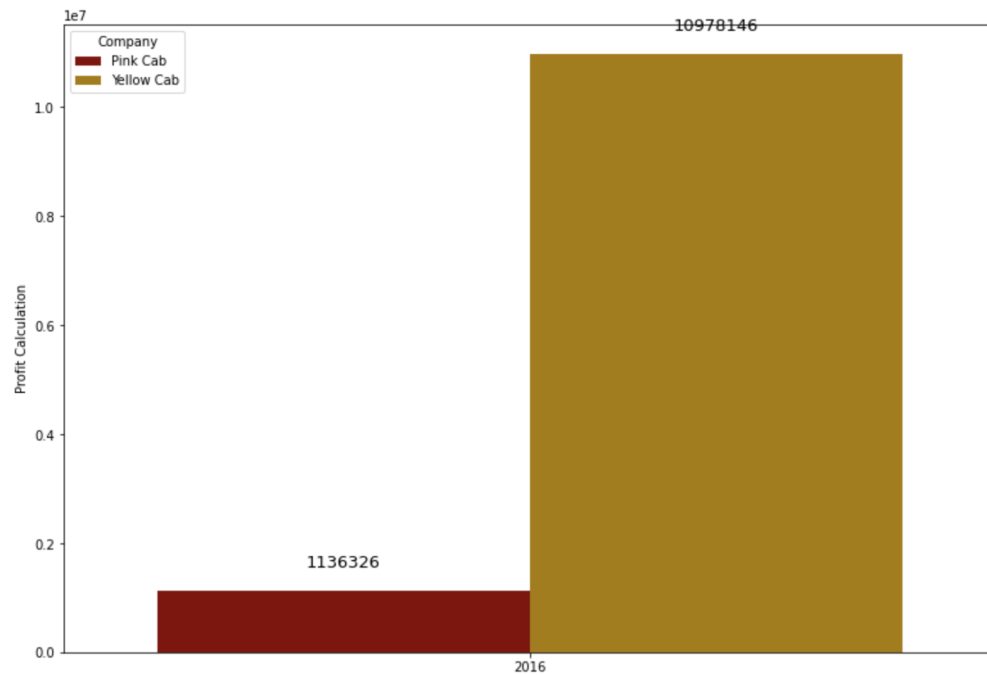
Final cab data

	Transaction ID	KM Travelled	Price Charged	Cost of Trip	Customer ID	Age	Income (USD/Month)	Profit
count	3.593920e+05	359392.000000	359392.000000	359392.000000	359392.000000	359392.000000	359392.000000	359392.000000
mean	1.022076e+07	22.567254	423.443311	286.190113	19191.652115	35.336705	15048.822937	137.253198
std	1.268058e+05	12.233526	274.378911	157.993661	21012.412463	12.594234	7969.409482	160.311840
min	1.000001e+07	1.900000	15.600000	19.000000	1.000000	18.000000	2000.000000	-220.060000
25%	1.011081e+07	12.000000	206.437500	151.200000	2705.000000	25.000000	8424.000000	28.012000
50%	1.022104e+07	22.440000	386.360000	282.480000	7459.000000	33.000000	14685.000000	81.962000
75%	1.033094e+07	32.960000	583.660000	413.683200	36078.000000	42.000000	21035.000000	190.030000
max	1.044011e+07	48.000000	2048.030000	691.200000	60000.000000	65.000000	35000.000000	1463.966000

Descriptive Statistics

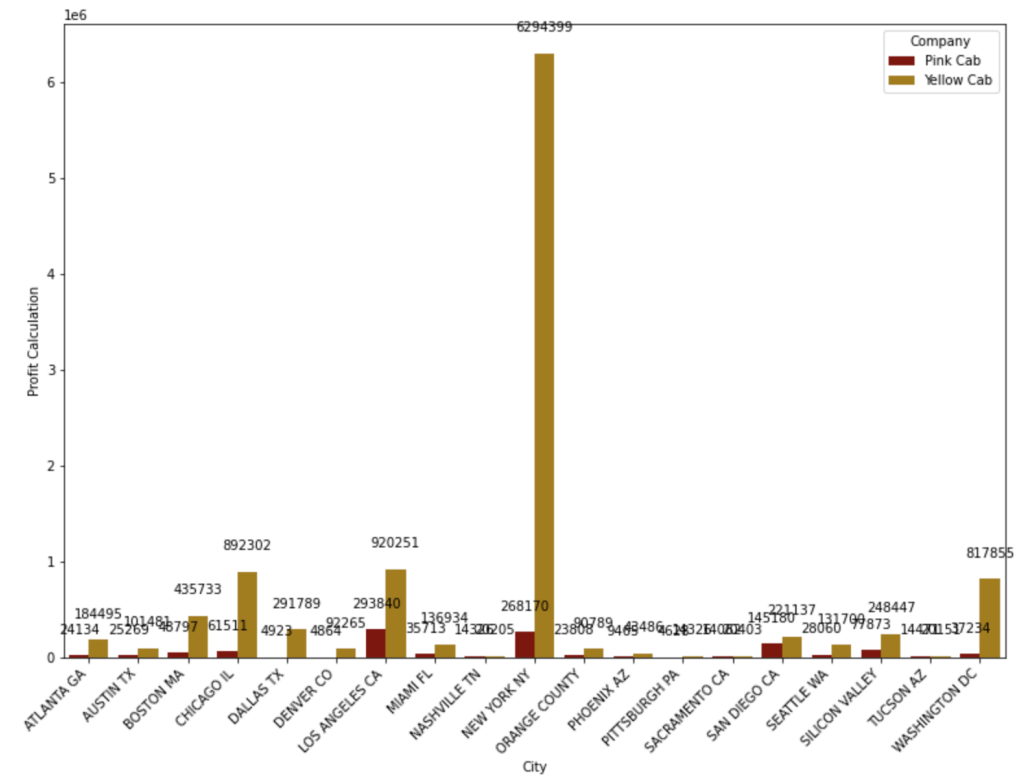
Profit Analysis for 2016

Profit for each company



Profit in 2016: grouped by company

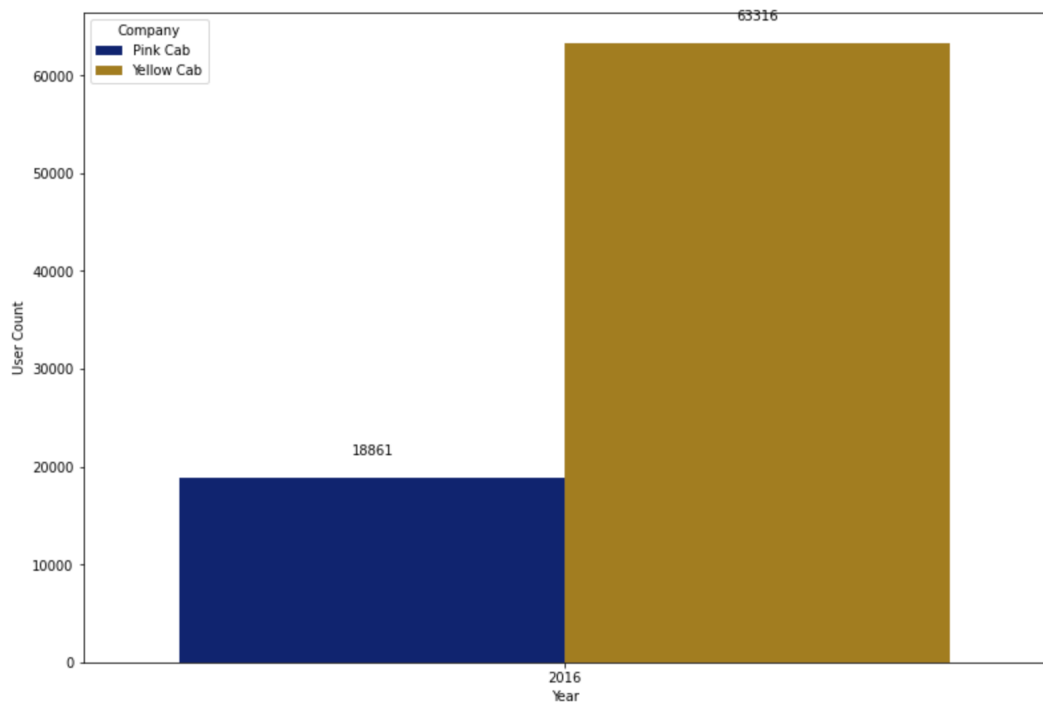
Profit for each city and company



Profit in 2016: grouped by company & city

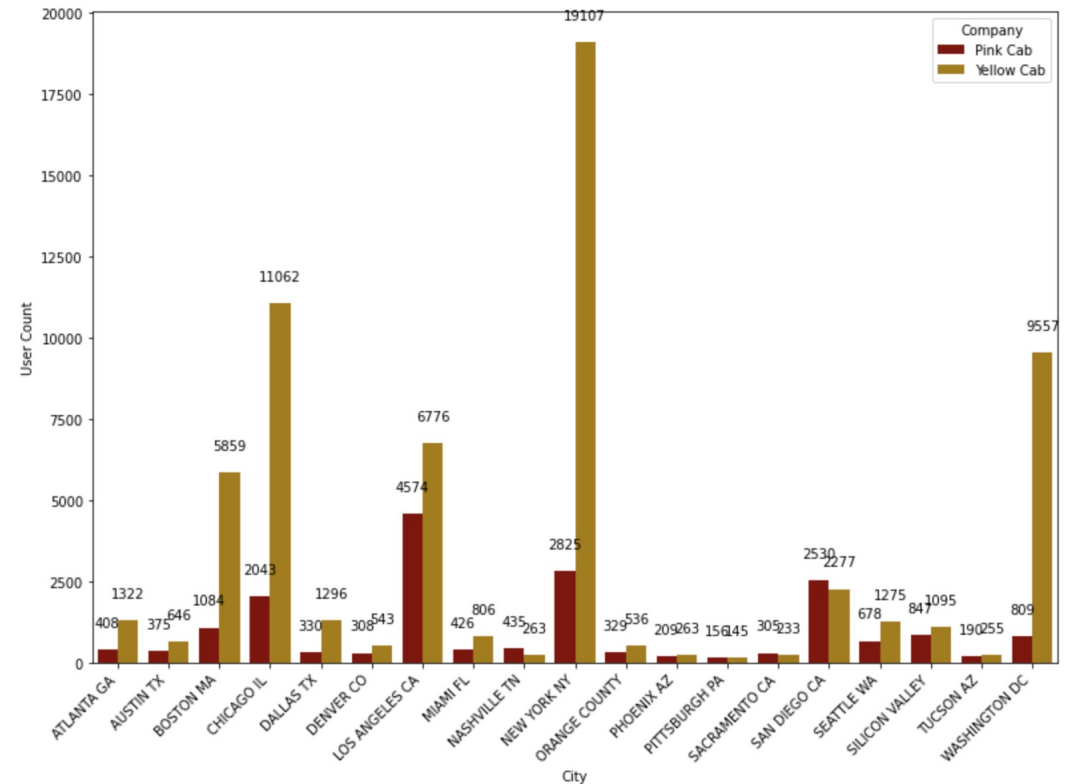
User Analysis for 2016

Users for each company



User count in 2016: grouped by company

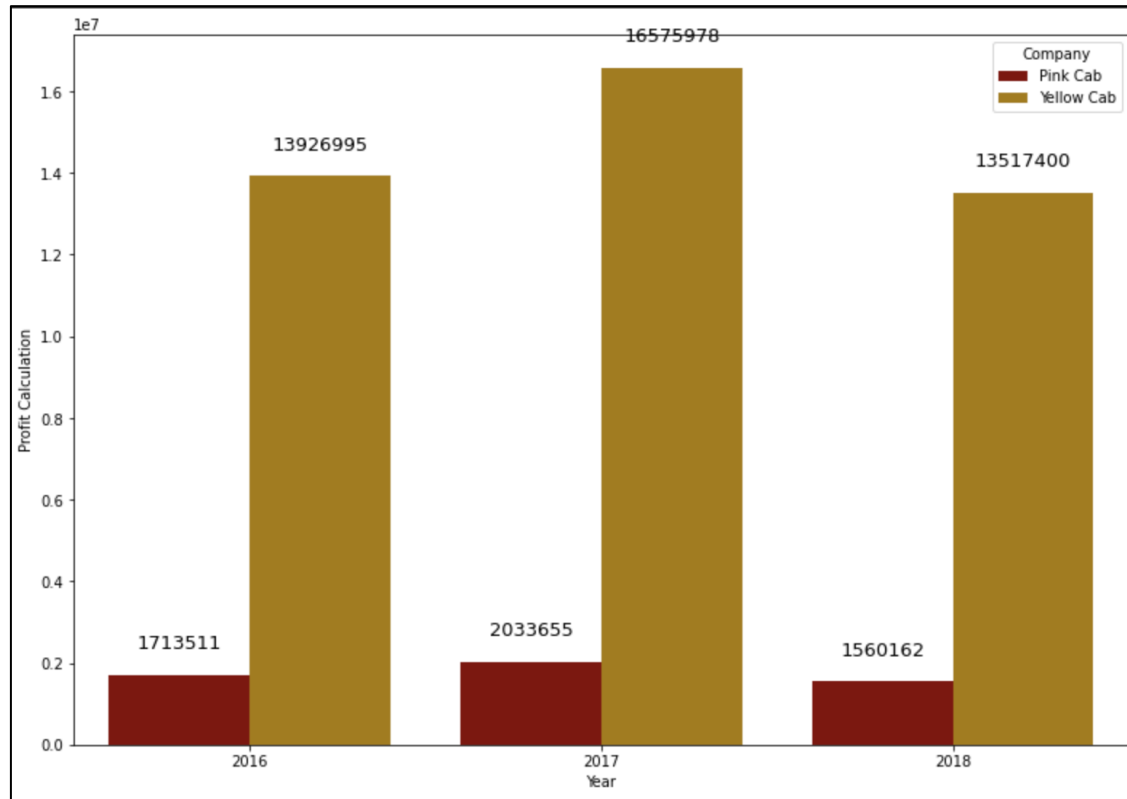
Users for each city and company



User count in 2016: grouped by company & city

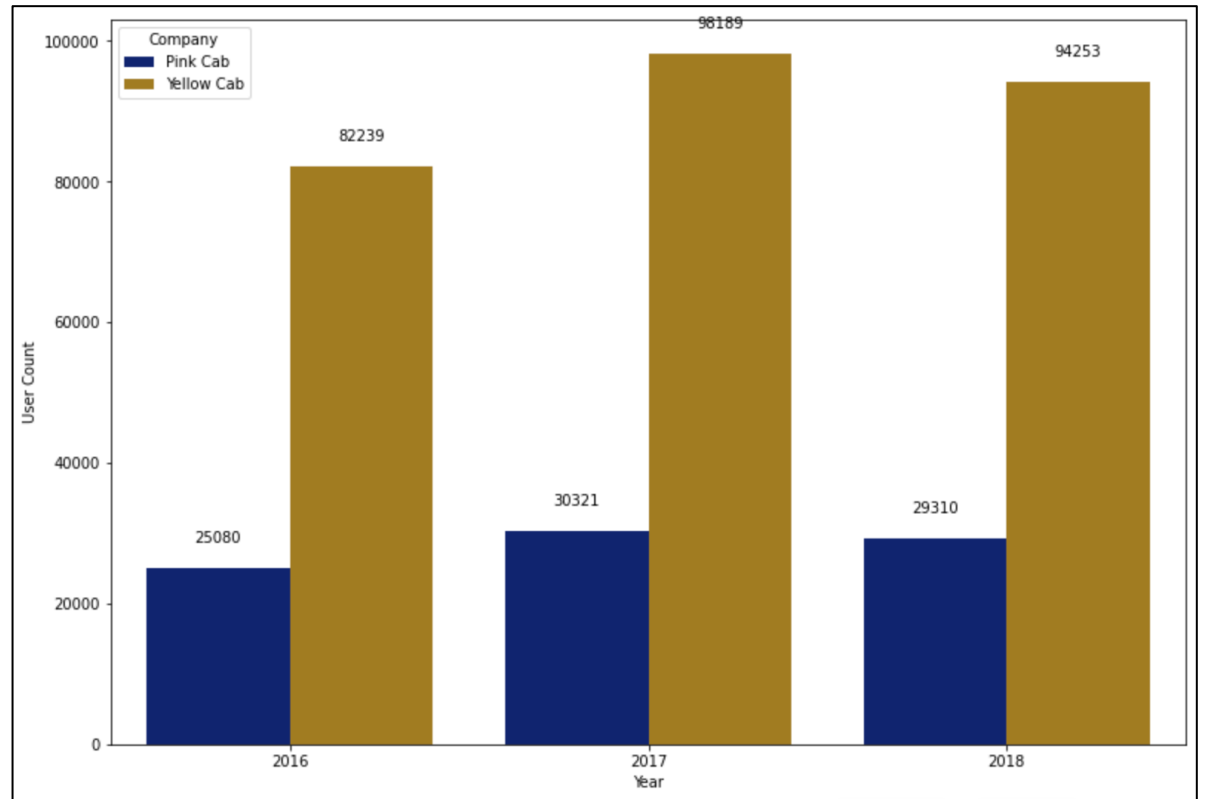
Profit & User Analysis for all 3 years

Profit grouped by company and year



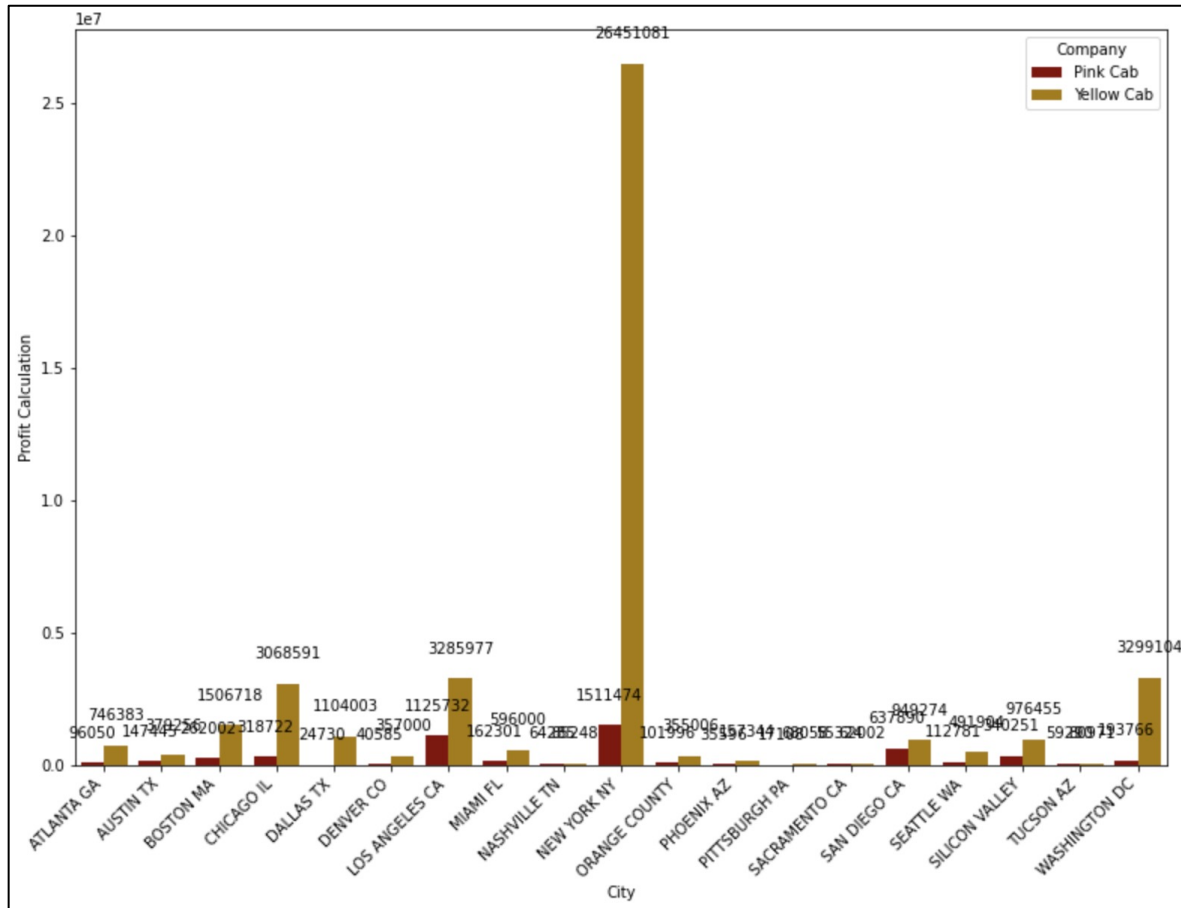
Total Profit: grouped by company & year

Users grouped by company and year



User count: grouped by company & year

Profit Analysis for all 3 years by city



Total Profit: grouped by company & city

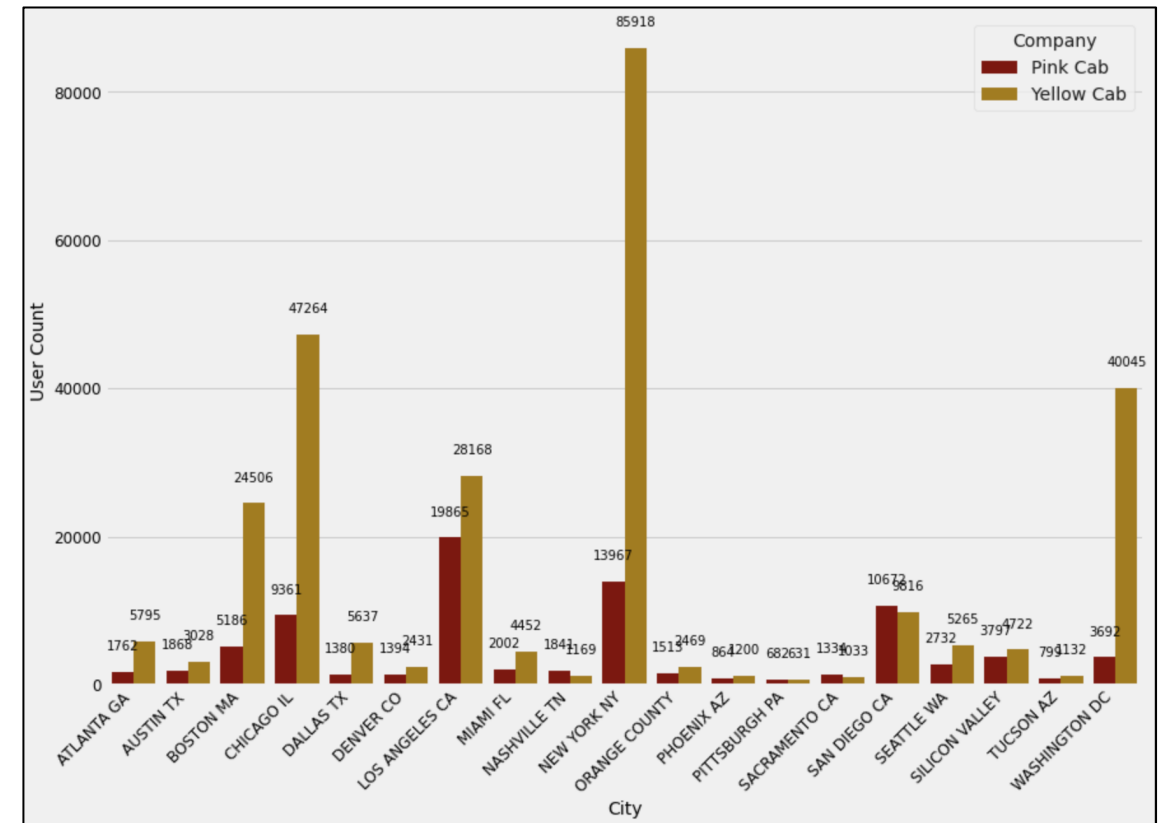
Profit grouped by company and age group

- Yellow cab company outperforms the Pink cab company in terms of profit
- The profit for the Yellow cab company is much higher in each city when compared to the Pink cab company
- The data combines all three years

User Analysis by city

Profit grouped by company and city

- **Yellow cab** company outperforms the **Pink cab** company in terms of number of users
- The user count for the **Yellow cab** company is much higher in each city when compared to the **Pink cab** company, except for San Diego
- The data combines all three years

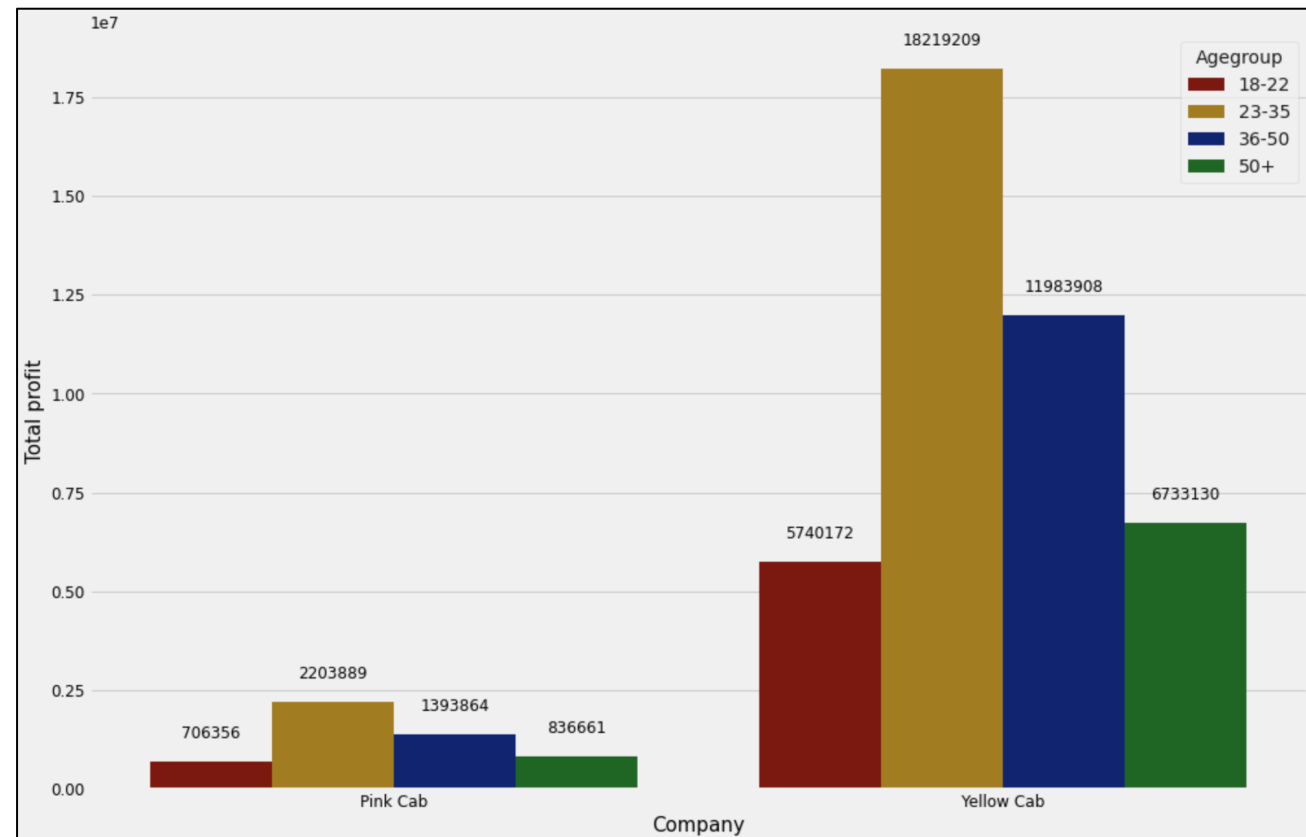


User count: grouped by company & city

Profit Analysis by age group

Profit grouped by company and age group

- Created labels for ages between 18-22, 23-35, 36-50, 50+
- The profit for the **Yellow cab** company is much higher for all of the age groups, specially 23-25
- The data combines all three years

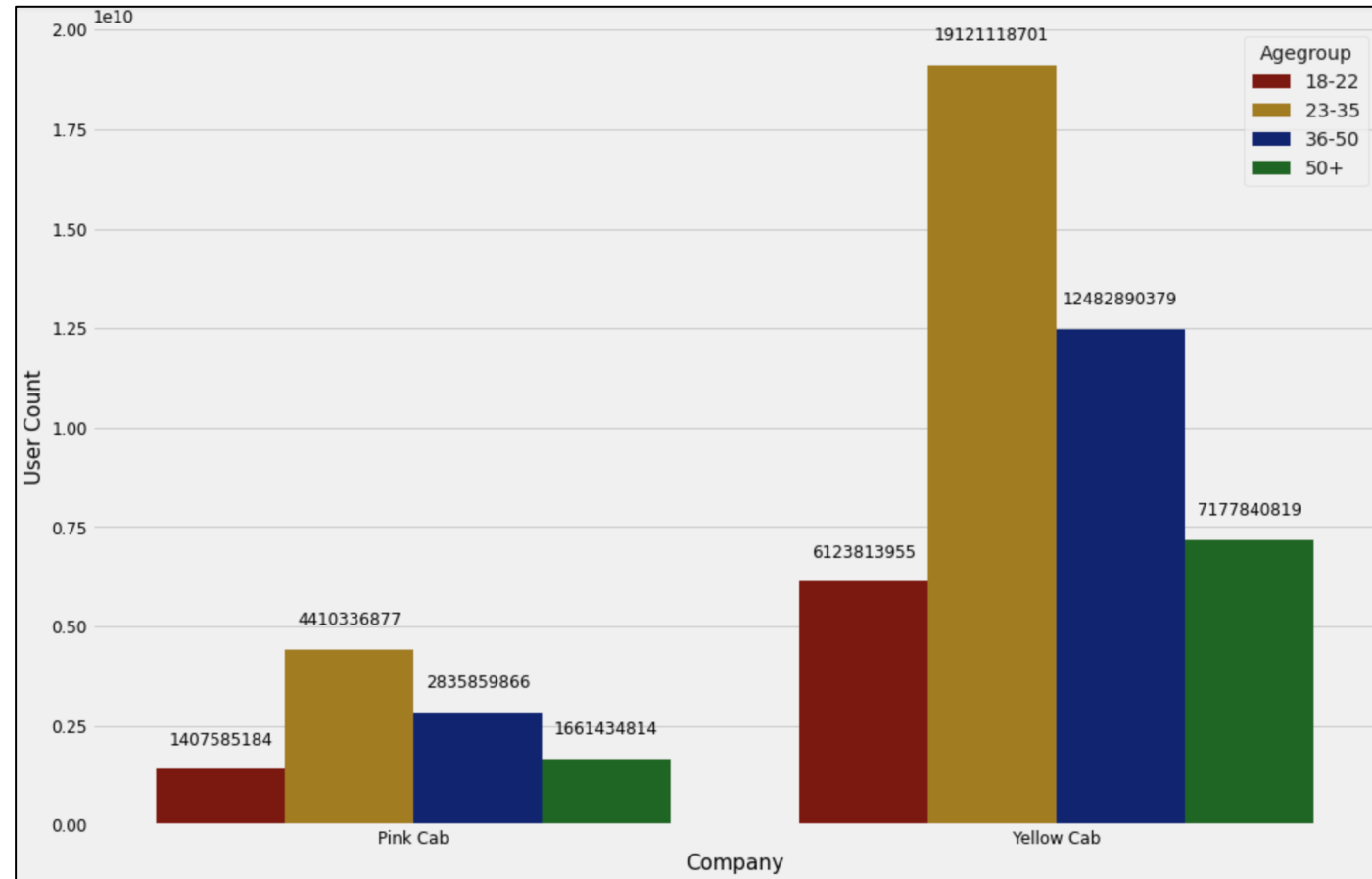


Total Profit: grouped by age group & company

User Analysis by age group

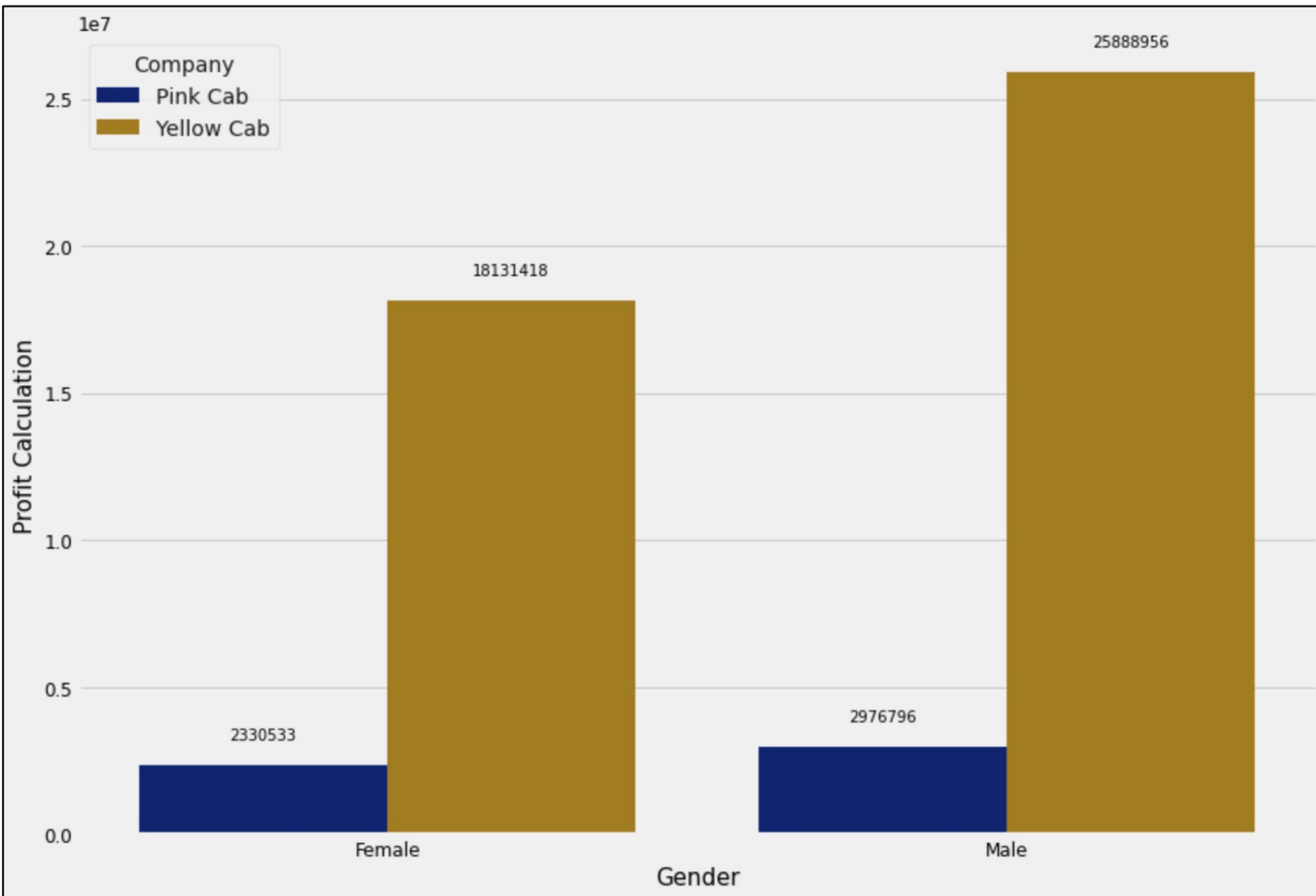
User grouped by company and age group

- Created labels for ages between 18-22, 23-35, 36-50, 50+
- The user count for the **Yellow cab** company is much higher for all of the age groups, specially 23-25
- The data combines all three years



User Count: grouped by age group & company

Gender wise Profit Analysis



Total Profit: grouped by age group & company

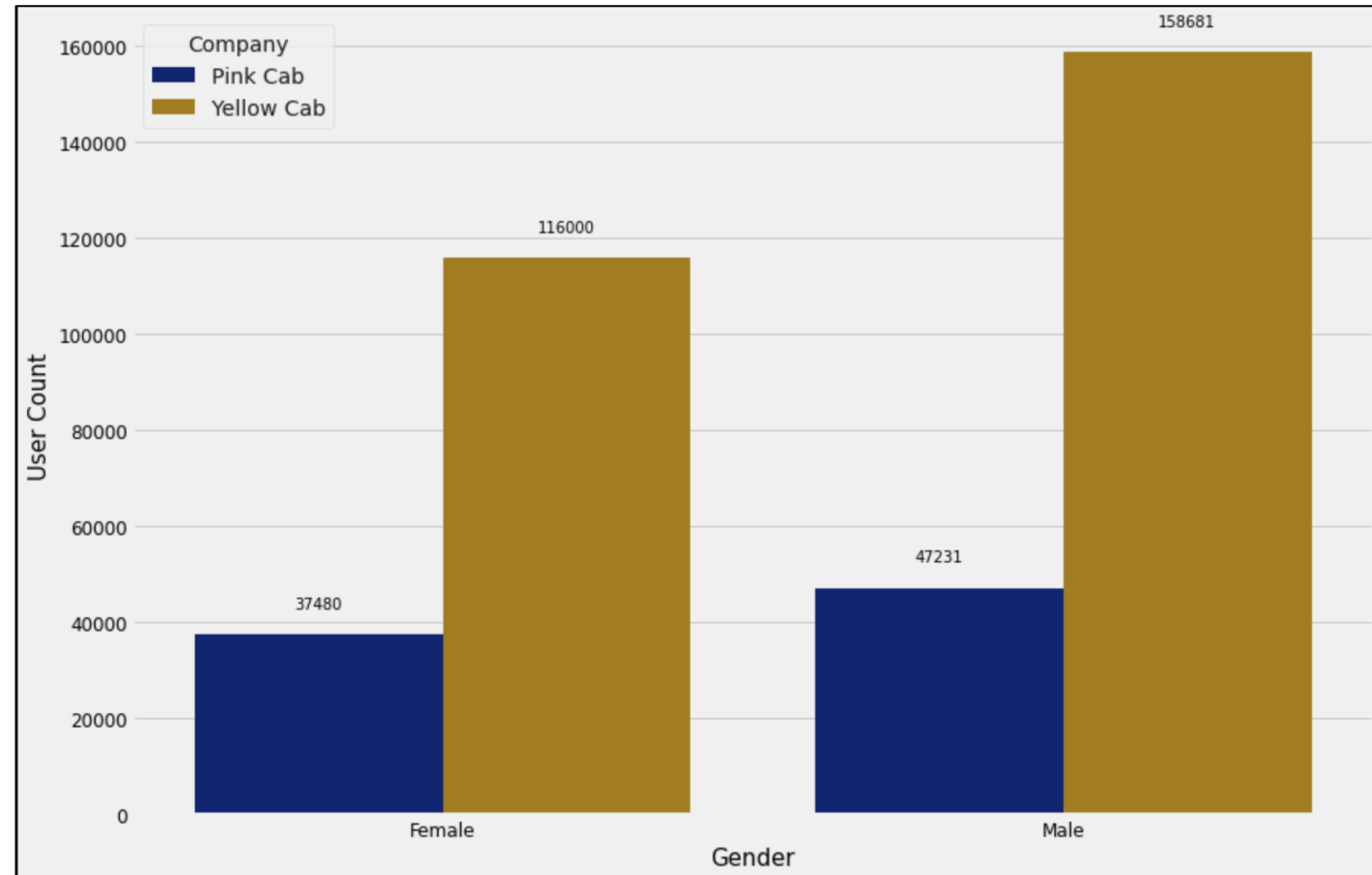
Profit grouped by Company and Gender

- Grouped the data based on company and Gender
- The profit for the **Yellow cab** company is much higher for both Male and Female clients
- For the **Pink cab**, the profit contribution is the same for Male and Female
- The data combines all three years

Gender wise User Analysis

User grouped by company and Gender

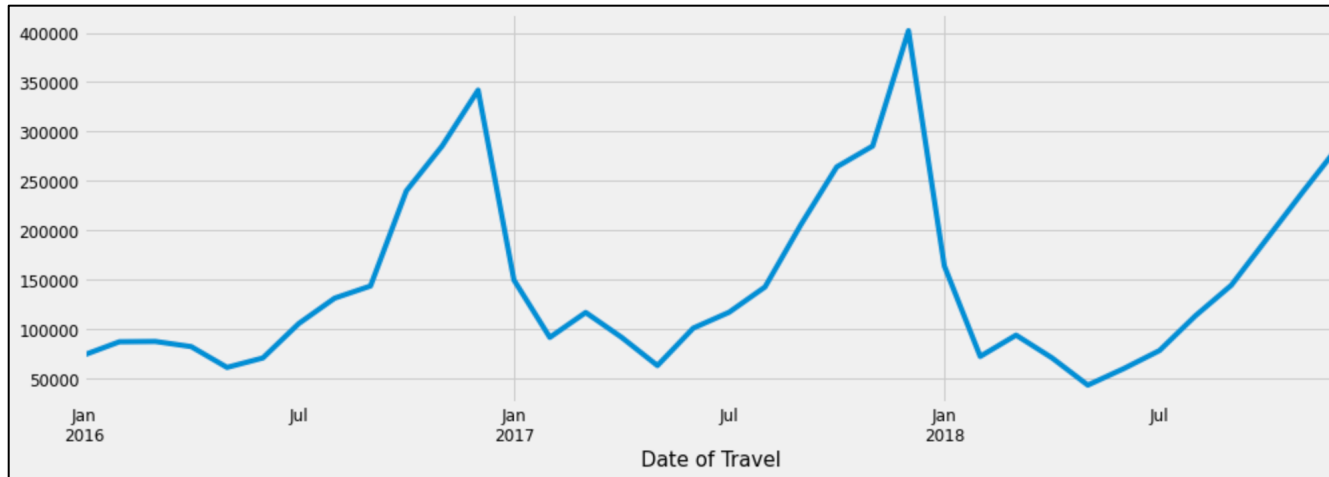
- Grouped the data based on company and Gender
- The users for the **Yellow cab** company is much higher for both Male and Female clients
- For the **Pink cab**, the user contribution is the same for Male and Female
- The data combines all three years



Total Users: grouped by age group & company

Profit & User Analysis

PINK CAB



Profit grouped by company and city
for **PINK CAB**

Profit: grouped by Date of Travel

Profit grouped by company and city
YELLOW CAB

YELLOW CAB



User count: grouped by Date of Travel

Recommendations

We have evaluated both the cab companies on following points and found Yellow cab better than Pink cab:

- **Customer Reach** : Yellow cab has a much higher customer reach in 18 cities while Pink cab has higher customer reach in only 1 city. We have also observed that Yellow cab is doing good in covering other cab users as compared to Pink cab.
- **Gender wise Reach**: We have analyzed this in two cases: by calculating the profit for each company based on gender and by calculating the number of users of each company based on gender. Yellow cab was higher in both the cases.
- **Age wise Reach** : Yellow cab has a higher customer reach in all age groups ranging between 18-22, 23-35, 36-50 and 50+. It's been observed that it's even popular in 60+ age group as equally as its in 18-25 age group.
- **Profit Analysis**: Yellow cab's profit, on average is almost eight times the profit of the Pink cab (for all three years).
- **Ride count and Profit** : Both the companies show a huge difference in the profit and the number of users. Yellow cab's number of users is much higher than that of the Pink cab company. There is a direct correlation between the profit and the user count. Therefore, Yellow cab also shows a much higher profit.

After the above analysis, we can confirm that the company in which XYZ should invest in is the Yellow Cab company. This is mainly due to its high profit generation ability due to a large number of users when compared to the pink cab.

G2M Investment Strategy

Thank You