Exploratory Data Analysis

< Insight for Cab Invesment Firm >

< 17/08/2022 >

Agenda

- ➢ Problem Statement
- ➢ Insights
- ➢ Data Preparation
- ➢ Data Visualization
- ➢ Recommendation

Problem Statement

Problem : 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.

Task : We are analyzing the Data related to Cab industry to help XYZ make the right investment in the market

After understanding the problem, The work was made through 4 steps :

1. Gaining insights from the Data
2. Data preparation
3. Data Visualization
4. Conclusion and Recommendations

Insights

We have 4 tables that contains 15 columns, and here is the description of them :

```
Data columns (total 15 columns):
 #   Column             Non-Null Count    Dtype
---  ------             --------------    -----
 0   Transaction ID     359392 non-null   int64
 1   Date of Travel     359392 non-null   datetime64[ns]
 2   Company            359392 non-null   object
 3   City               359392 non-null   object
 4   KM Travelled       359392 non-null   float64
 5   Price Charged      359392 non-null   float64
 6   Cost of Trip       359392 non-null   float64
 7   Customer ID        359392 non-null   int64
 8   Gender             359392 non-null   object
 9   Age                359392 non-null   int64
 10  Income (USD/Month) 359392 non-null   int64
 11  Payment_Mode       359392 non-null   object
 12  Population         359392 non-null   int64
 13  Users              359392 non-null   int64
 14  profit             359392 non-null   float64
dtypes: datetime64[ns](1), float64(4), int64(6), object(4)
```

Insights

And here are the main points that we are going to focus on:

- Company name
- Profit
- average

company	km travelled	profit	profit per KM
Pink cab	1911073	5307328	2.777145614
Yellow cab	6199417	44020373	7.100727859

Data preparation

1- Data preparation phase:

- Joining the 4 tables into one final table
- Removing redundant columns (transaction ID)
- Adding extra features (profit)

The final Data Frame :

	Date of Travel	Company	City	KM Travelled	Customer ID	Gender	Age	Income (USD/Month)	Payment_Mode	Population	Users	profit
0	2016-01-10	Pink Cab	ATLANTA GA	30.45	29290	Male	28	10813	Card	8405837	24701	57.3150

Data preparation

And now we have :

- 12 Features(including 9 derived features)
- Timeframe as years from 2016 to 2019
- Total data points :355,032
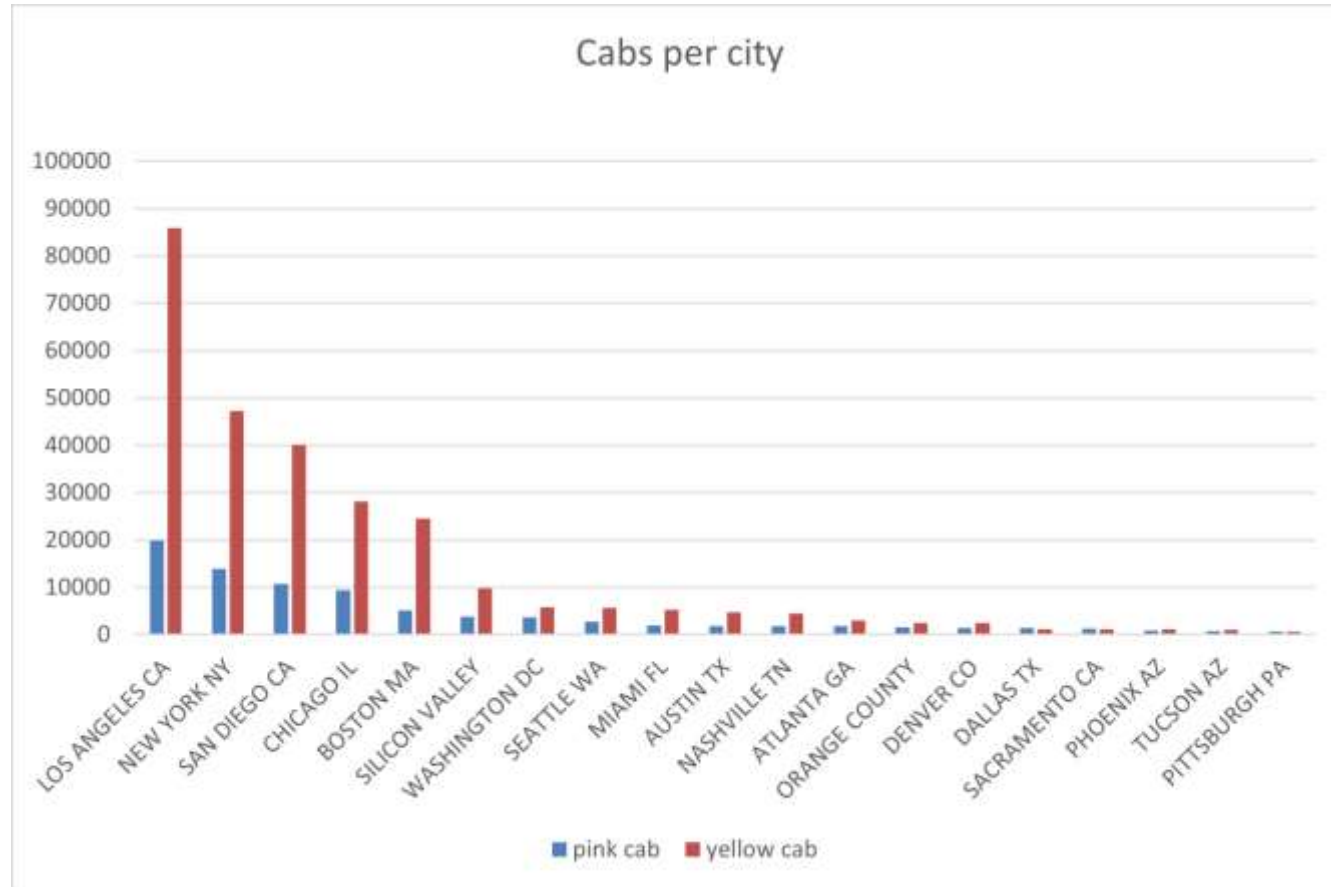
The final Data Frame :

	Date of Travel	Company	City	KM Travelled	Customer ID	Gender	Age	Income (USD/Month)	Payment_Mode	Population	Users	profit
0	2016-01-10	Pink Cab	ATLANTA GA	30.45	29290	Male	28	10813	Card	8405837	24701	57.3150

Data Visualization
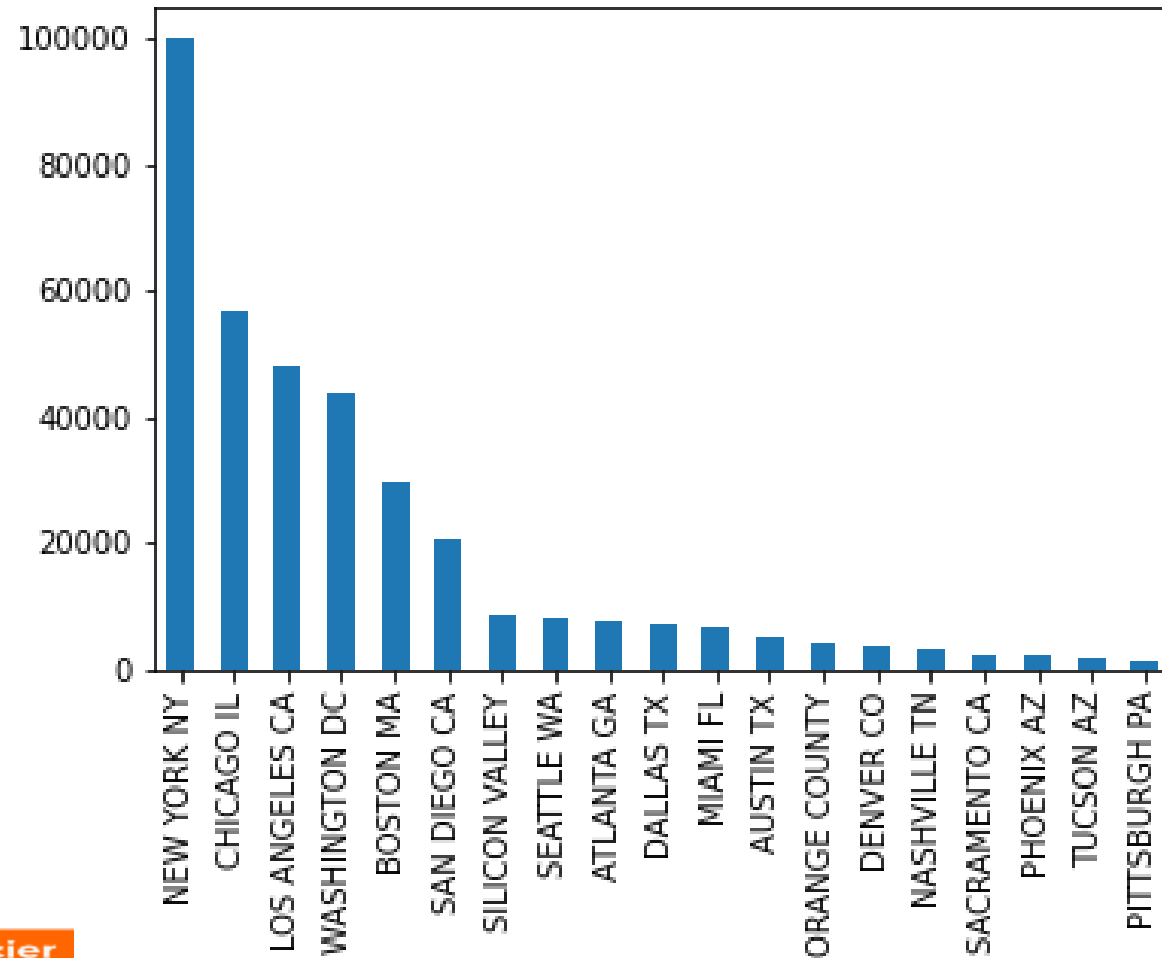
Here is the top cities where cab industry takes place, despite the company's Name
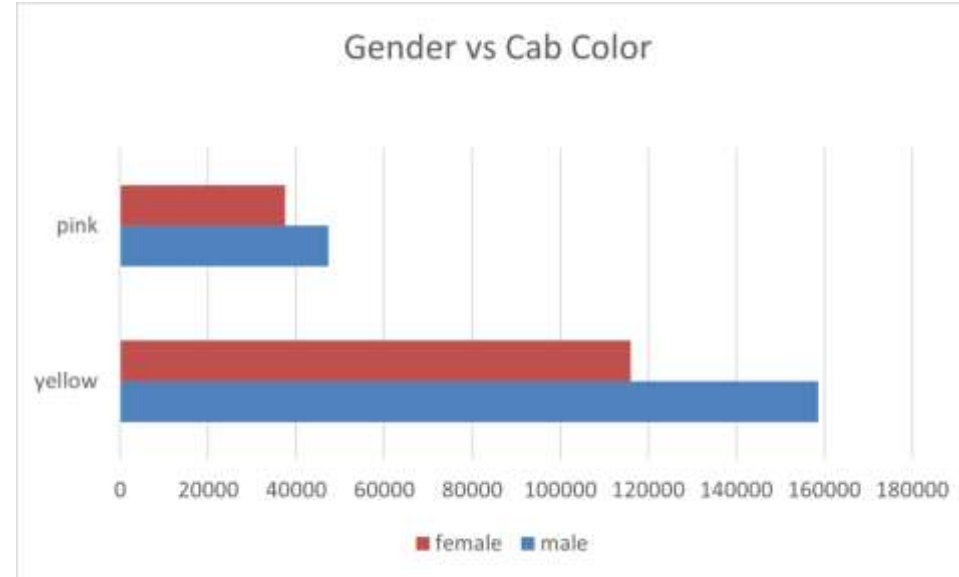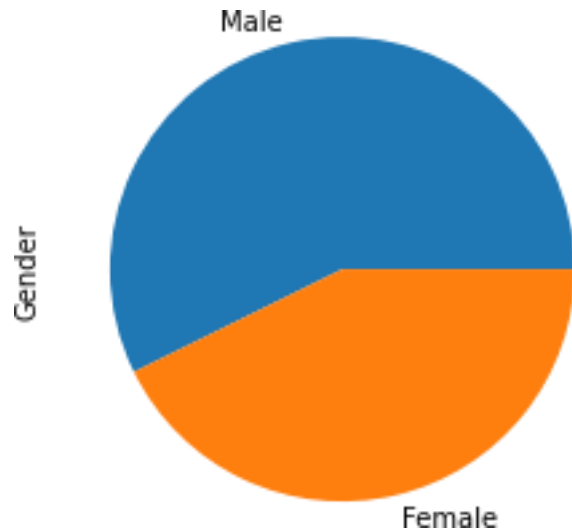
Data Visualization

Cabs per city

And here we have a comparison between the reach of each company(Pink vs Yellow) in the cities we mentioned before
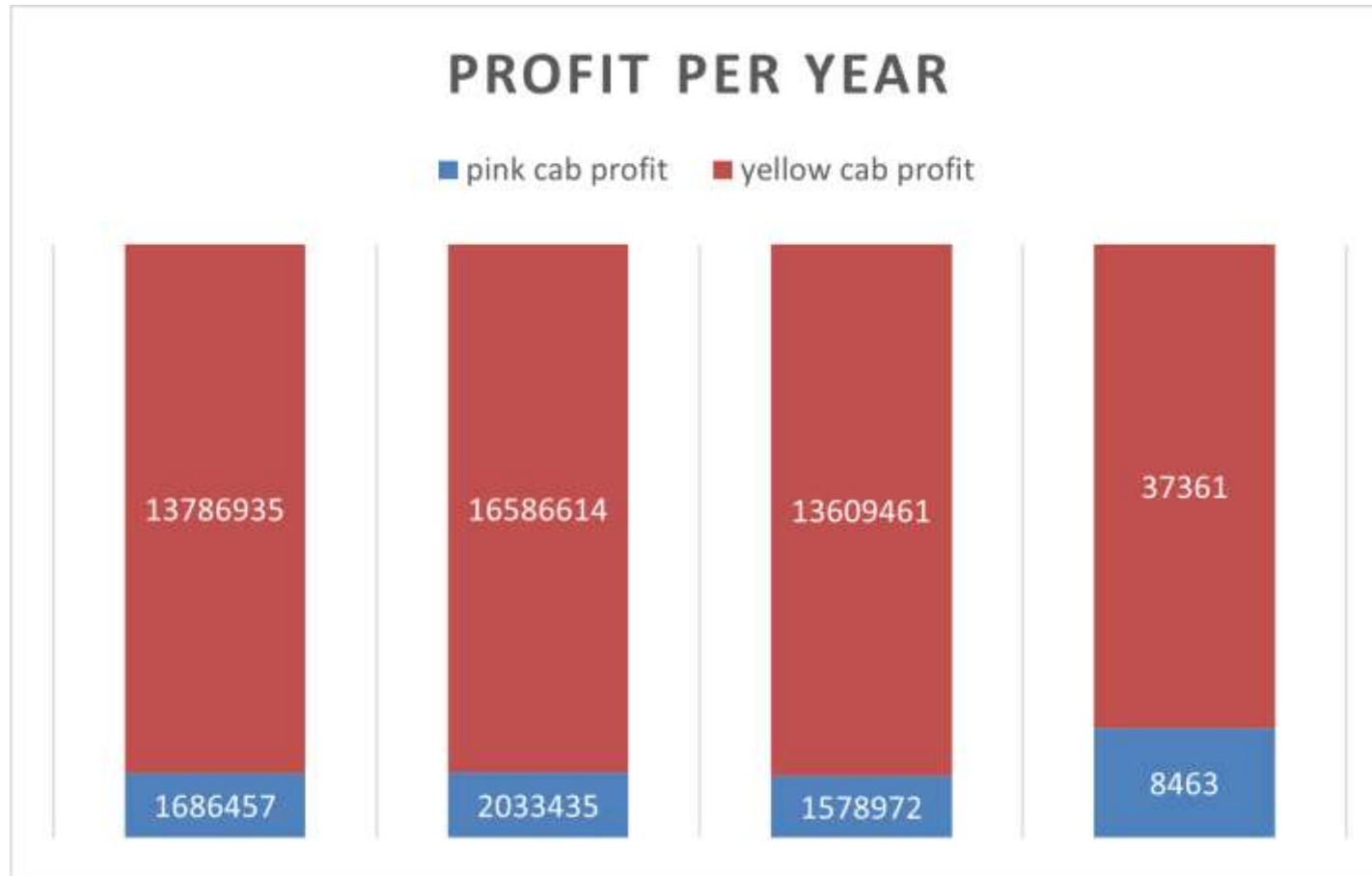
Yellow cab have higher customer reach in all cities except for Dallas ,Sacramento and Pittsburgh

Data Glacier
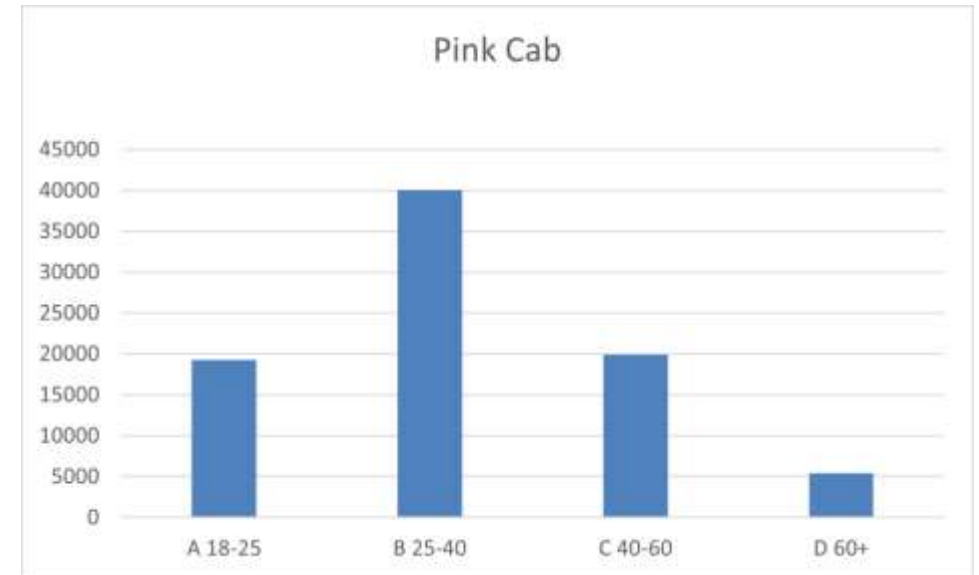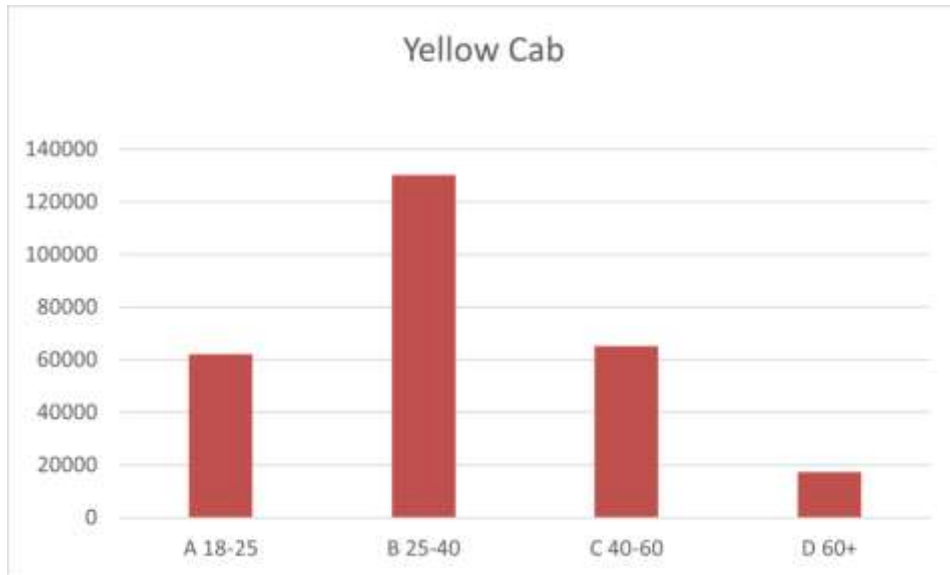
Data Visualization

the male percentage is bigger than the female percentage in both companies but as the graph shows, the female percentage in pink cab companies is almost equal to the male percentage which means : pink cars are more woman- friendly choice.

Data Visualization

PROFIT PER YEAR

■ pink cab profit   ■ yellow cab profit

|			
13786935	16586614	13609461	37361
1686457	2033435	1578972	8463

In terms of Profit per year : Yellow cabs are way more profitable option than the pink cab, we can also see a noticeable growth in pink cab profit in the year 2019

Data Visualization

In both companies, The most common group age in cab industry is group B (25 – 40) and the less one is the group D +60
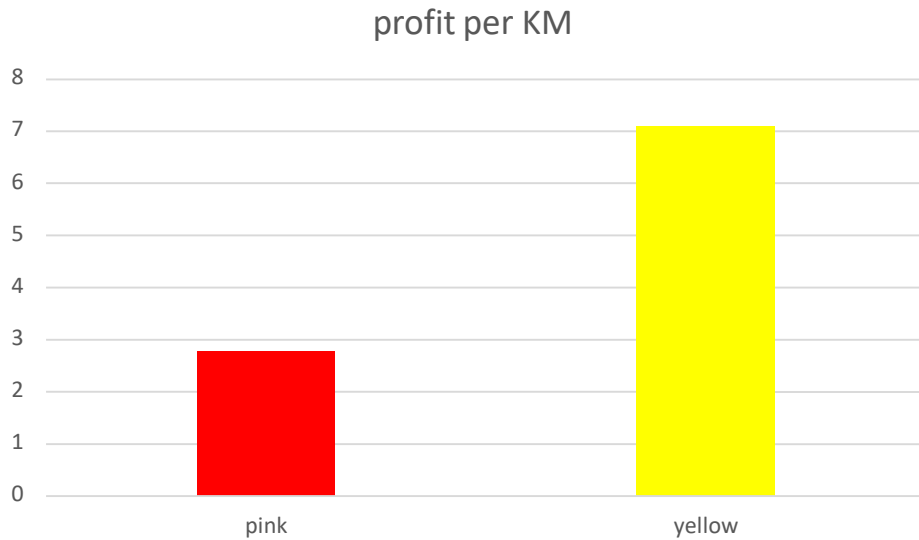
Data Visualization
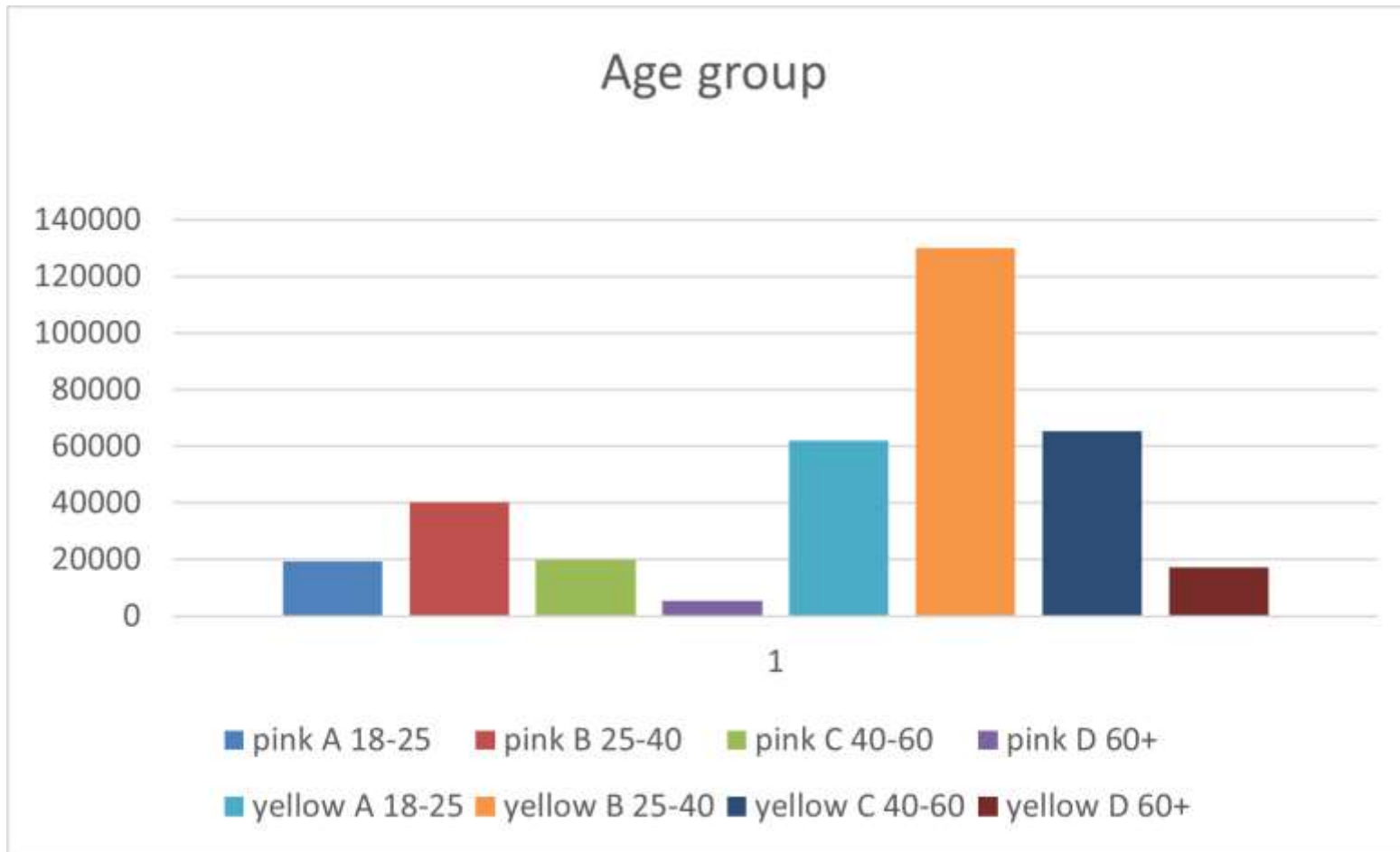
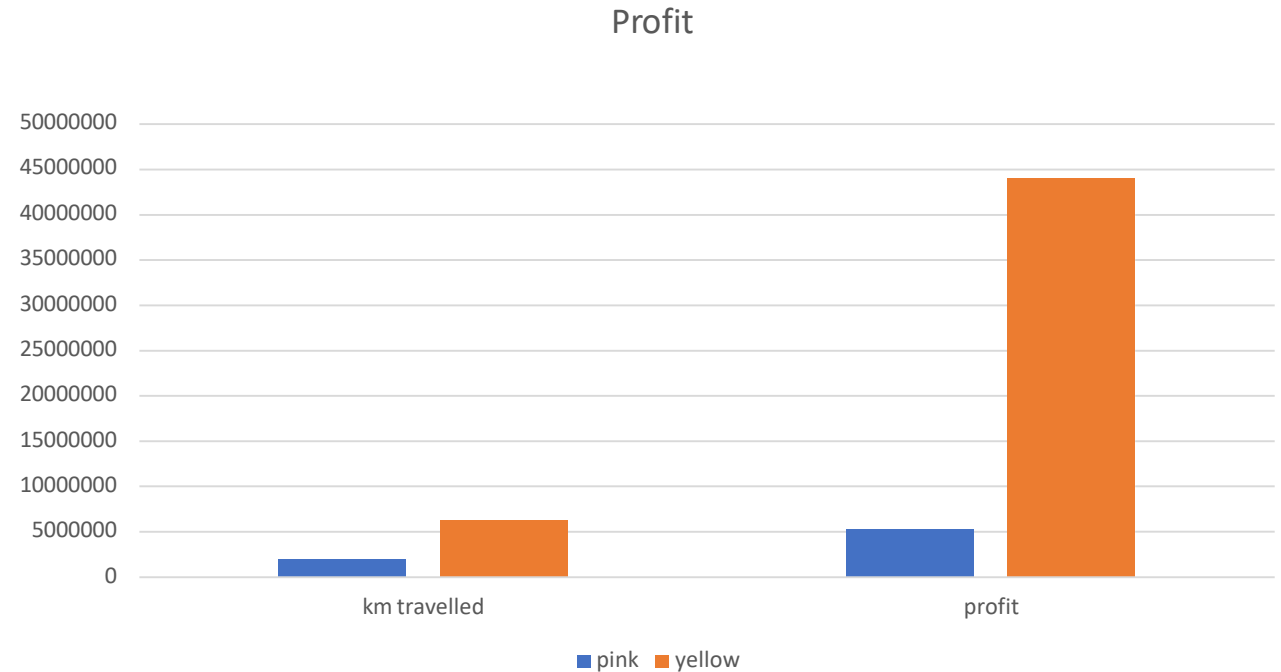In total : the yellow company Group B (25-40) is the best option in terms of profit

Yellow cabs are better than pink in all age groups

The D group in yellow cab is as good as the A group in pink Cab!

Data Glacier

Data Visualization
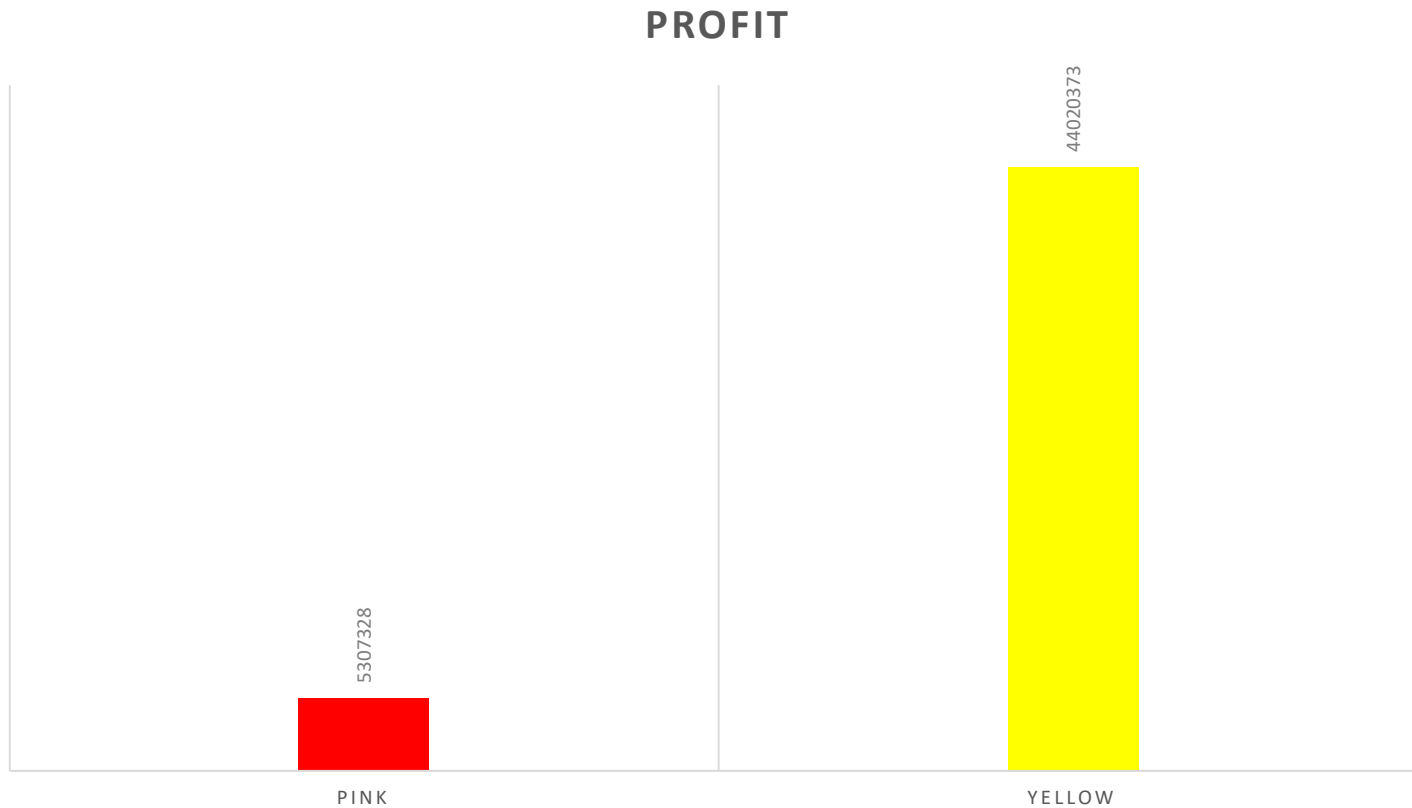
profit per KM

Profit

When analyzing the Profit per KM graph we can tell that the yellow Cab have better ratio than the pink cab

While the number of KM travelled in yellow cab is about 3 times more than the pink Cab, the profits are almost 8 times bigger!!

Data Glacier

Data Visualization

PROFIT

And finally we have the total profit , we can see the huge difference between the 2 companies

Recommendations

After evaluating both companies we concluded that Yellow Cabs are better than Pink in terms of :

Total profit : yellow cabs profits are almost 8 times as the profits of pink cab

Customer Reach : Yellow cab has higher customer reach in 16 cities while Pink cab has higher customer reach only in 3 cities

Age wise Reach : Yellow cabs are better than pink in all age groups especially the 18-25 age group and the +60 age group

Gender : pink company are more woman-Friendly option

Average Profit per KM : Yellow cab's average profit per KM is almost three times the average profit per KM of the Pink cab.

We notice that the yellow cab options is better than pink in all the points mentioned above , and as a result we can say : unless the investment's target is Women, we recommend investing in Yellow cabs.

Thank You