Exploratory Data Analysis

<G2M_insight_for_Cab_Investment_firm >

<07/18/2023>

Executive Summary

Data Glacier
Your Deep Learning Partner

The 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.

Analysis and Hypothesis Test:

- Number of Transactions for each Company;
- Average price charged of two companies;
- Payment method Overview;
- Average profit for both companies each year;
- Gender Distributions;
- Age Distributions;
- Locations Distributions on Profit;
- Users Percentage per City

Problem Statement

Objectives:

XYZ is interested in using our actionable insights to help them identify the right company to make their investment.

Approach

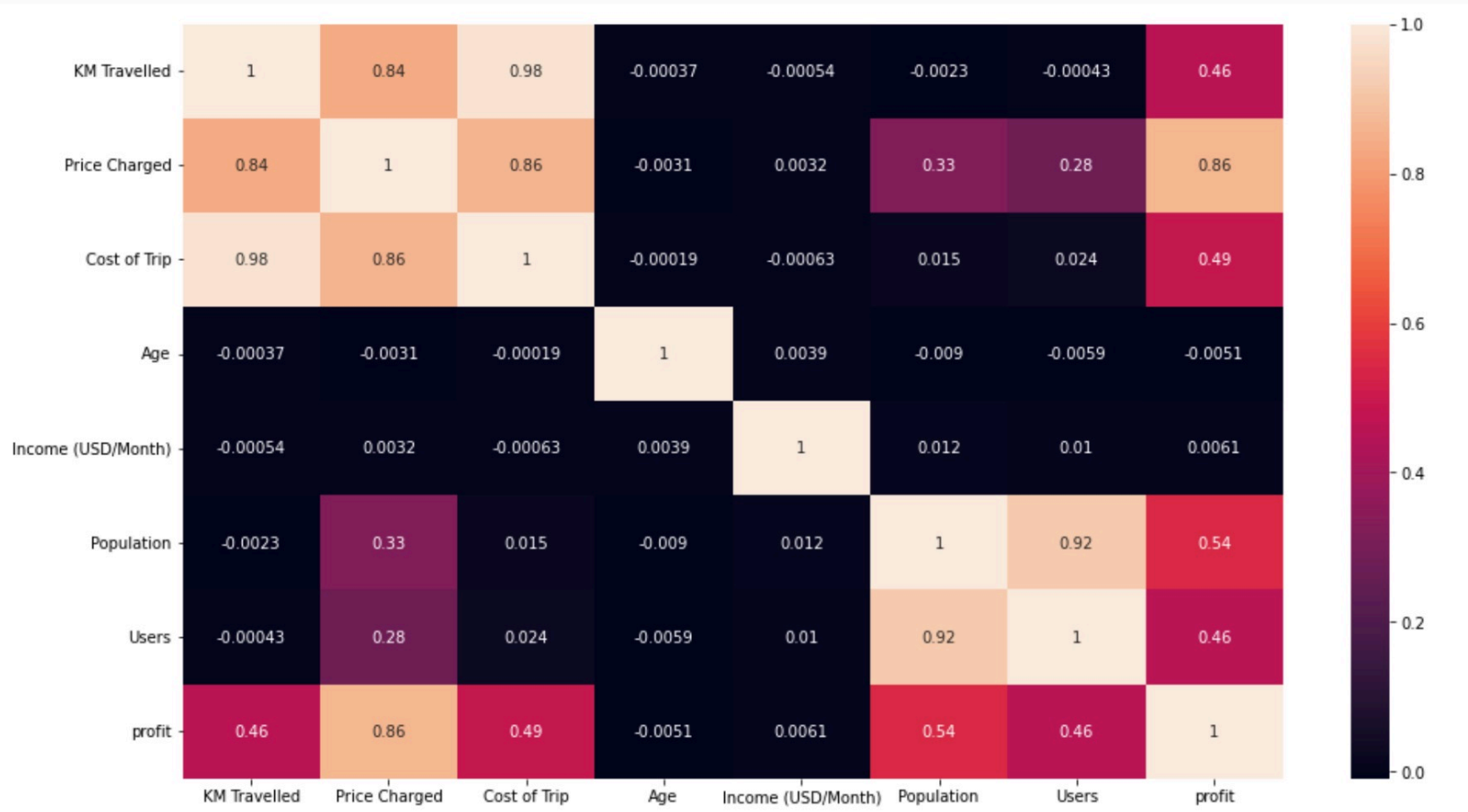
1. Explore Data Information
2. Find Correlations between features
3. Analysis on dataset and set up hypothesis
4. Use visualizations to investigate insights
5. Provide recommendations for investment

Data Information

- **Cab_Data.csv –** this file includes details of transaction for 2 cab companies
- **Customer_ID.csv** – this is a mapping table that contains a unique identifier which links the customer's demographic details
- **Transaction_ID.csv –** this is a mapping table that contains transaction to customer mapping and payment mode
- **City.csv –** this file contains list of US cities, their population and number of cab users

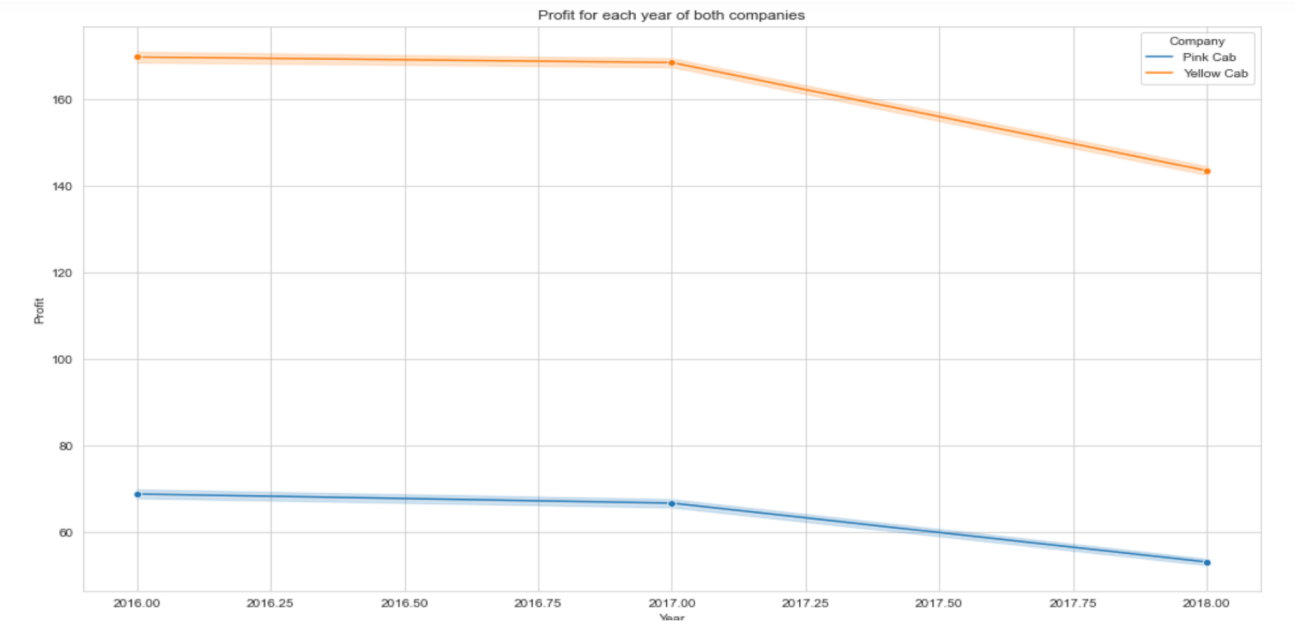
- **Number of features**: 18
- Time period of data is from 31/01/2016 to 31/12/2018.
- **Number of data points:** 359392
- Created new feature **Profit** by using **price charged – Cost of trip**
- **City dataset** contains other cab companies including yellow and pink
- Cash and Card are only payments for both companies

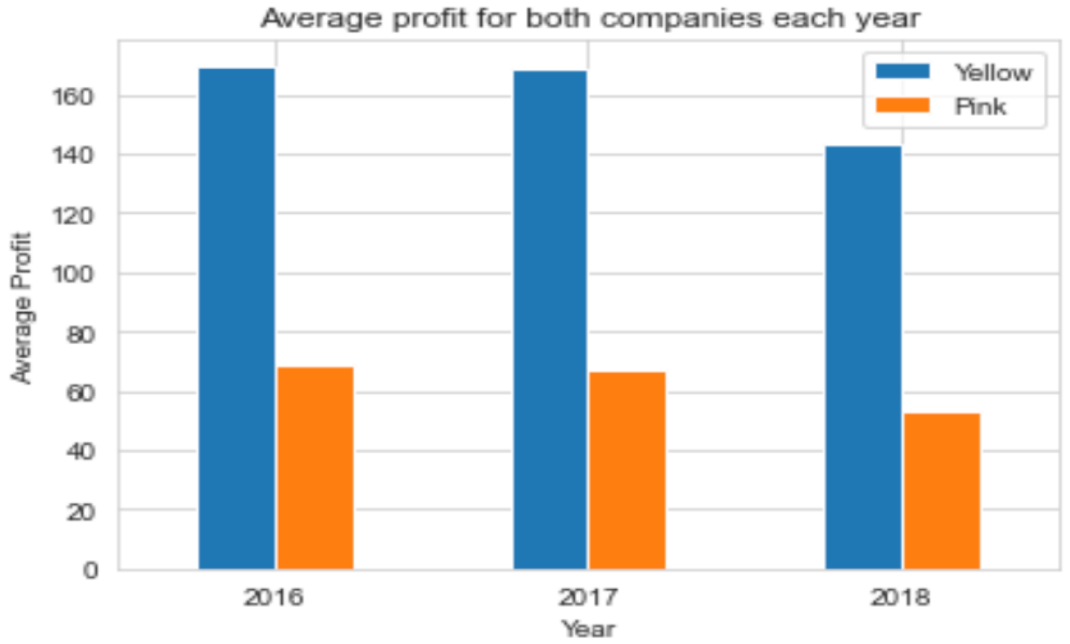
Data Correlation



- There are several high correlations:
  - KM Travelled vs. Price Charged vs. Cost of Trip;
  - Populations vs. Users.
- On the other hand, KM travelled/price charged/ cost of trip doesn't depend on customers' age or income.

Profit Analysis





- The profit decreasing each year from 2016-2018
- The average profit of yellow cab is higher than pink cab each year
- Overall profit of yellow cab is higher than pink cab

Gender Analysis

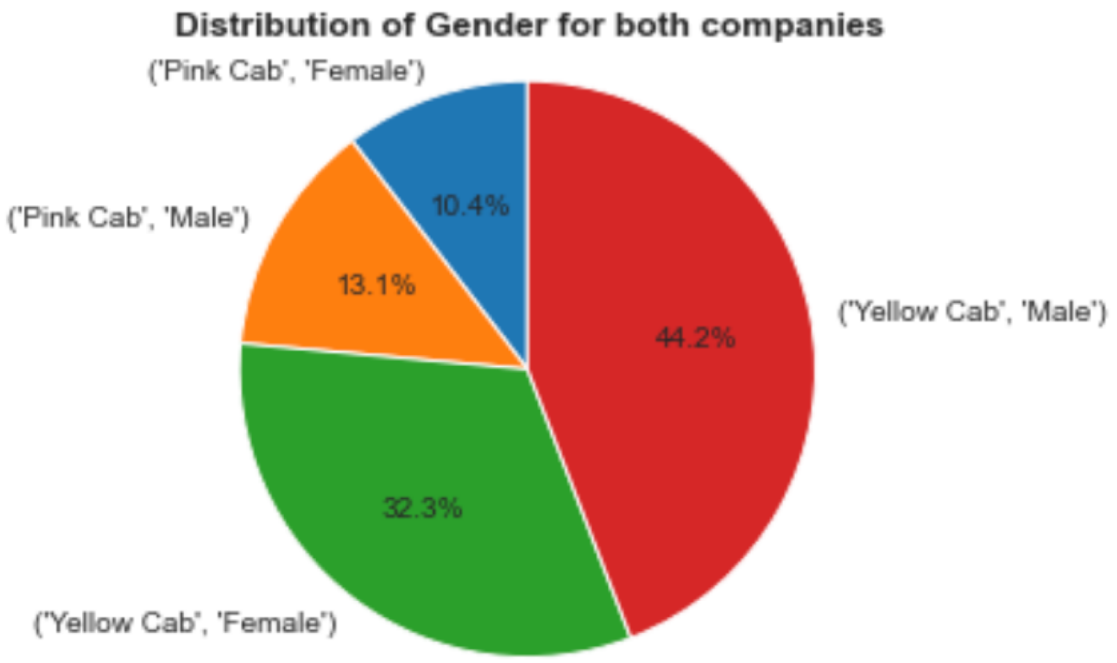
```
Average profit of gender for each company

Company      Gender
Pink Cab     Female      62.180701
             Male        63.026310
Yellow Cab   Female     156.305325
             Male       163.150948
Name: profit, dtype: float64
```

Hypothesis: there is a difference in profit regarding gender for both companies.

```
t-stat for yellow cab: 2.785616066253197
p-value for yellow cab: 0.005345131177690455
```

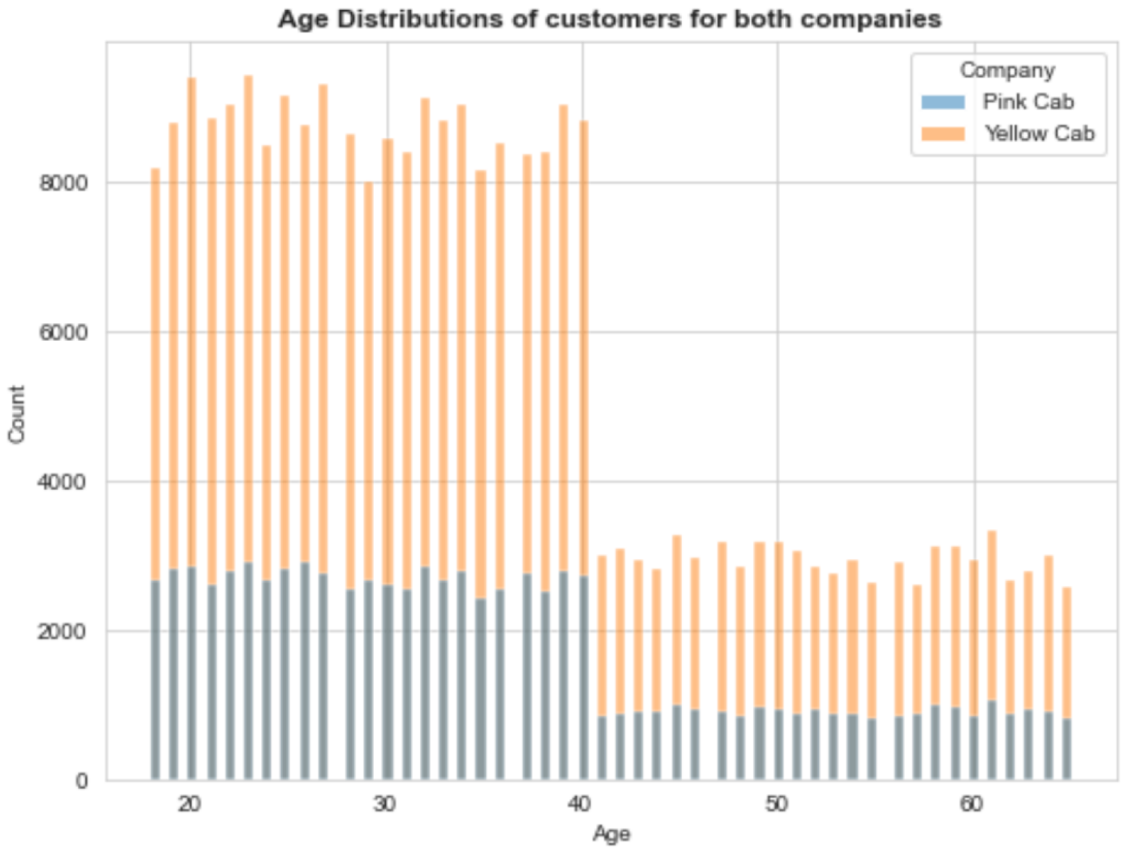
```
t-stat for pink cab: 0.5403746407894268
p-value for pink cab: 0.5889424154257256
```



Since P-value for yeallo cab is less than 0.05 and t-test is > 1.96, we reject the null hypohthesis. Thus, there is a difference in profit regarding gender for yellow cab.

On the other hand, we fail to reject null hypothesis for pink cab. Thus, there is no difference in profit regarding gender for pink cab.

Customers’ Ages Analysis

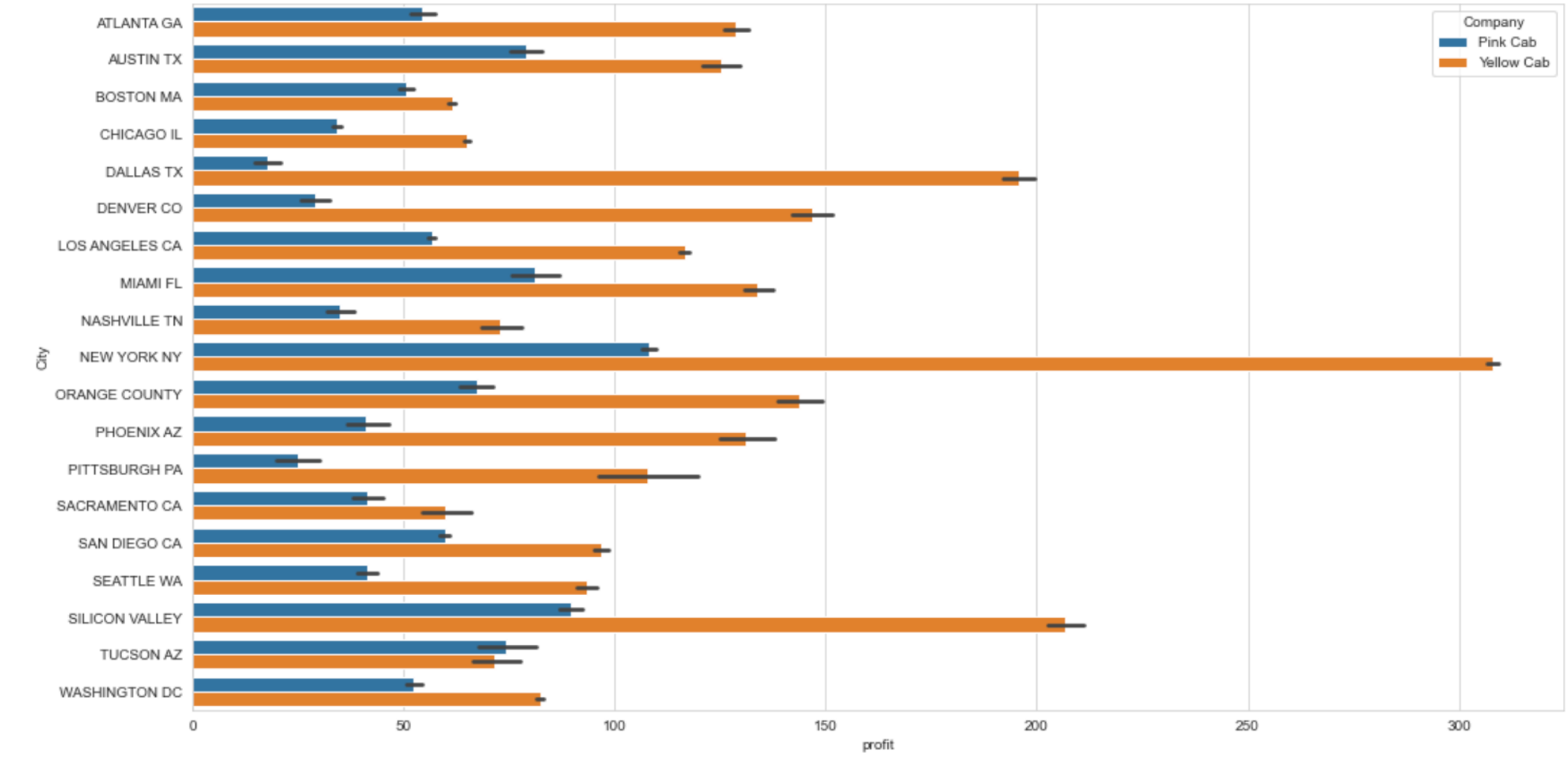


- Most of the customers are in the age group 18-40.
- 20- 30 age group has the most amount for yellow cab.
- Total number of customers of yellow Cab is more than the one of pink cab.

Average Income Analysis

The average income is about 15000 USD/month for both companies' customers.

Location Analysis



New York city has the most Yellow Cab and Pink Cab customers. Dallas and Silicon Valley are the second for the yellow cab.

Recommendations

We recommend investing in yellow cab because:

1. Yellow cab has more users as well as transactions
2. Yellow cab has larger profit margin
3. Yellow cab has more average profit each year.

Thank You