



Data Glacier

Your Deep Learning Partner

Exploratory Data Analysis

G2M Case Study

Francis Ngethe Ngari

6, September 2021

Background: 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.
- ❑ Target: Provide insights to XYZ company to help them make a data informed decision on which is the best company to invest in.
- ❑ In our data we are presented with data from two cab companies, Pink and Yellow cabs.

The analysis process includes these procedures:

- ❑ Data understanding: Checking all datasets provided and understanding its features.
- ❑ Identifying the number of transactions which were processed in different cab companies, then checking also the margins which represent profits.
- ❑ Identifying the most profitable company.
- ❑ Presenting recommendations.

Exploration

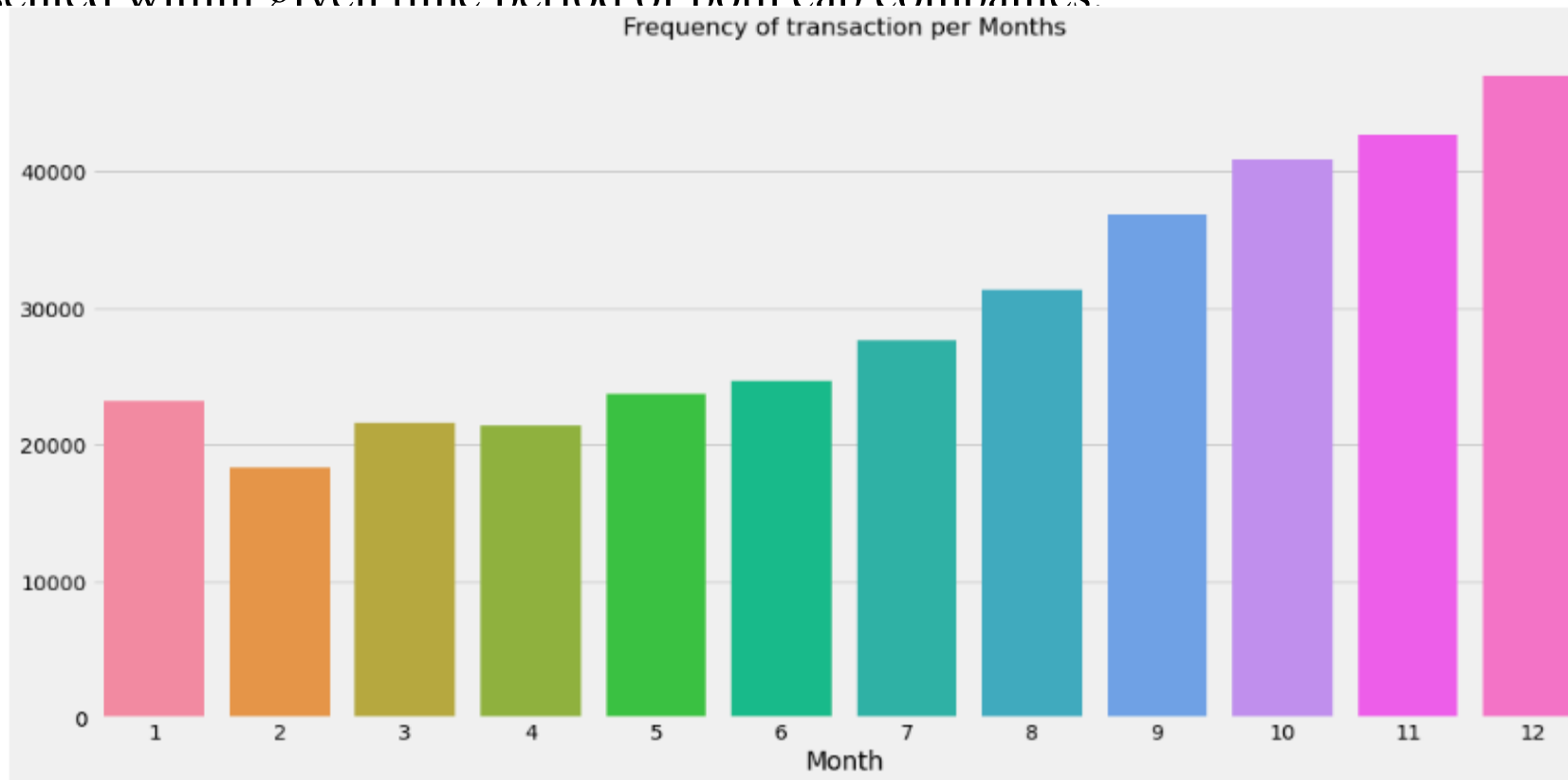
- ❑ Time frame of data is: 2016-01-31 to 2018-12-31.
- ❑ The Date of Travel column in Cab_Data.csv is encoded in Excel serial number hence should be transformed to correct format.
- ❑ There are datasets provided: Cab_Data.csv, Customer_ID.csv, Transaction_ID.csv, City.csv
- ❑ Combining Cab_Data.csv, Customer_ID, Transaction_ID to form a master dataset.
- ❑ Master dataset contains 15 columns and 359392 rows as observations.

Assumptions:

- ❑ Having dropped City.csv, the number of transactions represent the number of users of both cabs and also its features are not unique.
- ❑ It seems there are outliers in Price_Charged feature but later on we discover that some discounts are offered by Yellow cab company.

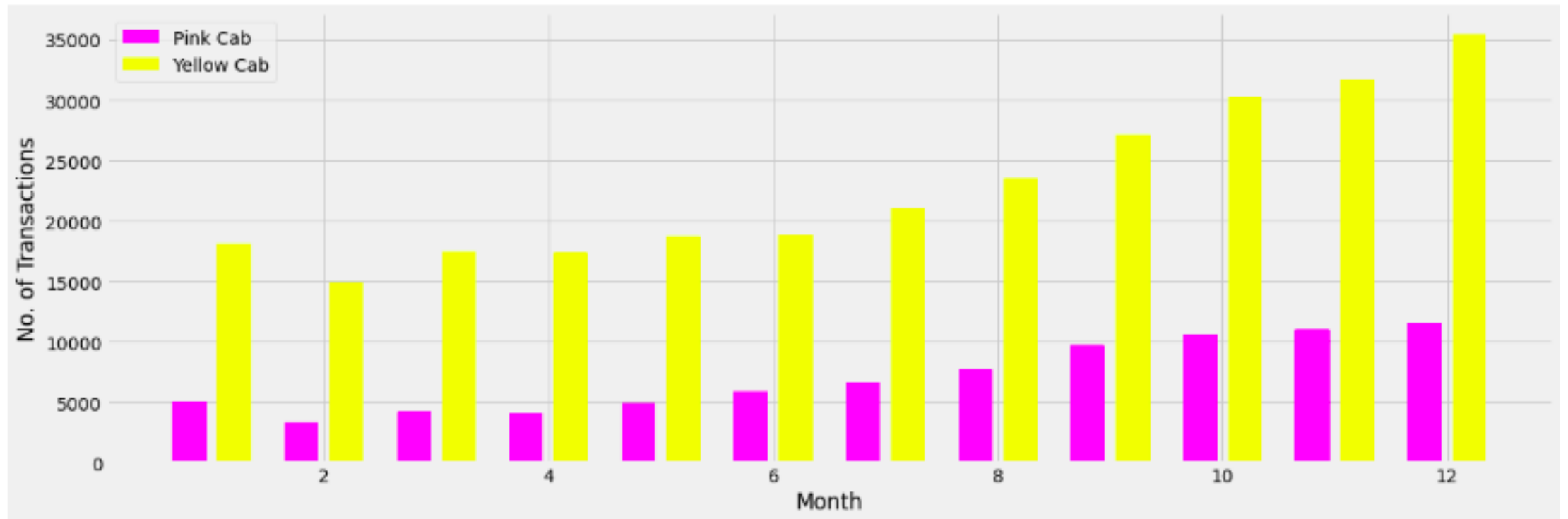
Monthly Transaction Analysis

- ❑ Frequencies of Transactions occurred during different months shows they increased throughout the year except the months of February, march and april which remained stagnant. This is represented within given time period of both cab companies.



Monthly Transaction of both companies

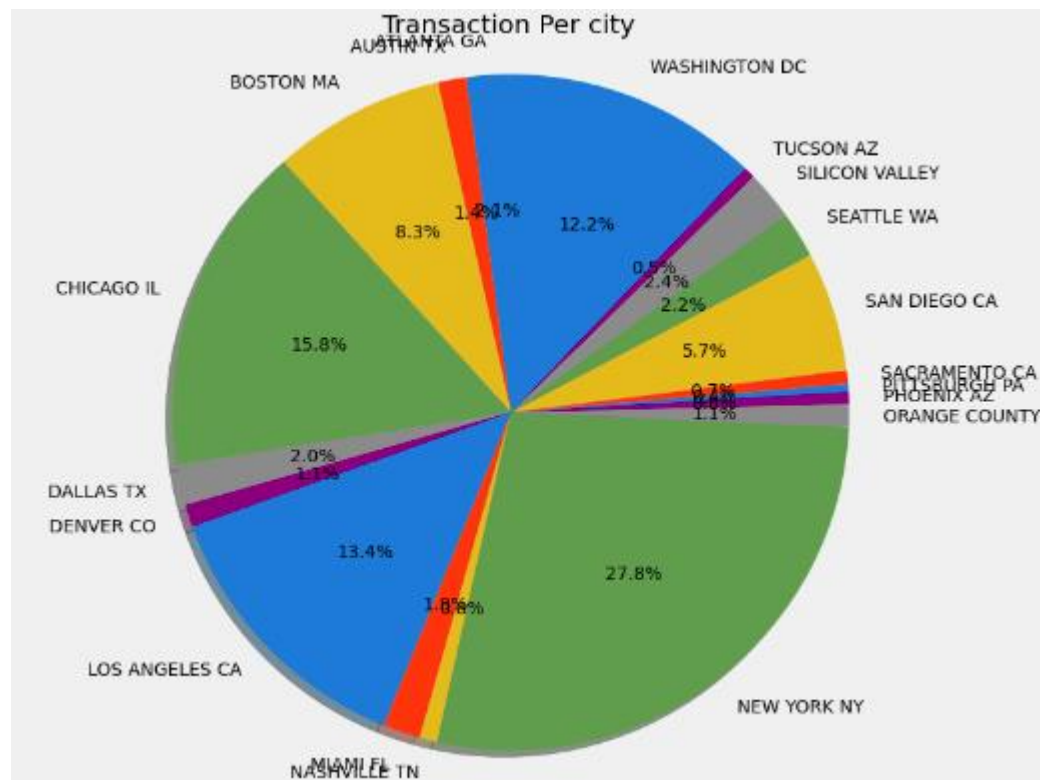
❑ In diagram below it shows Yellow cab has a huge no transactions monthly within the provided



Transaction per city

Observation below:

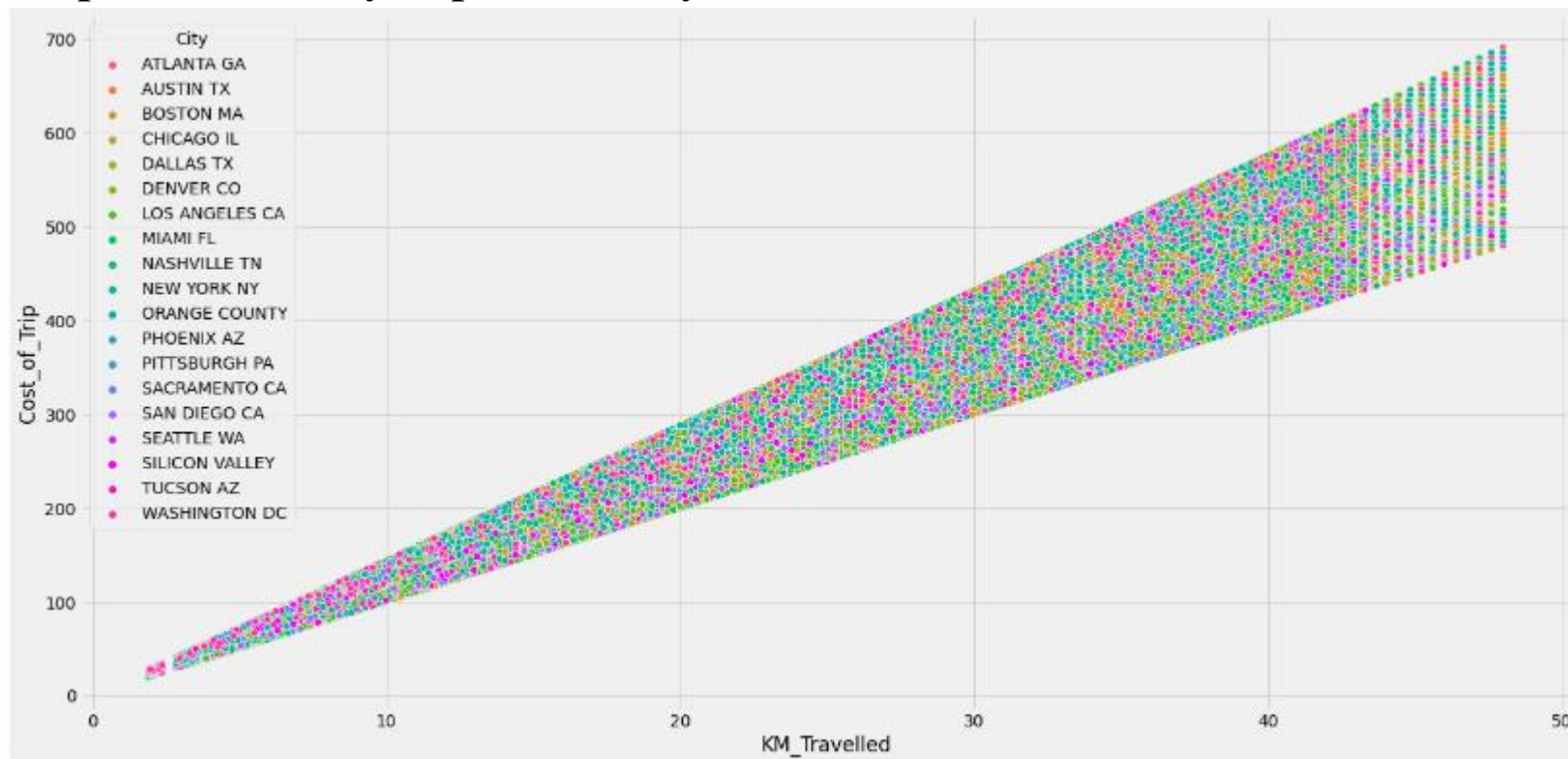
- ❑ Transactions happening in different cities.
- ❑ NY has a high number of transactions.



Cost of trip in relation with Km travelled

Observation below:

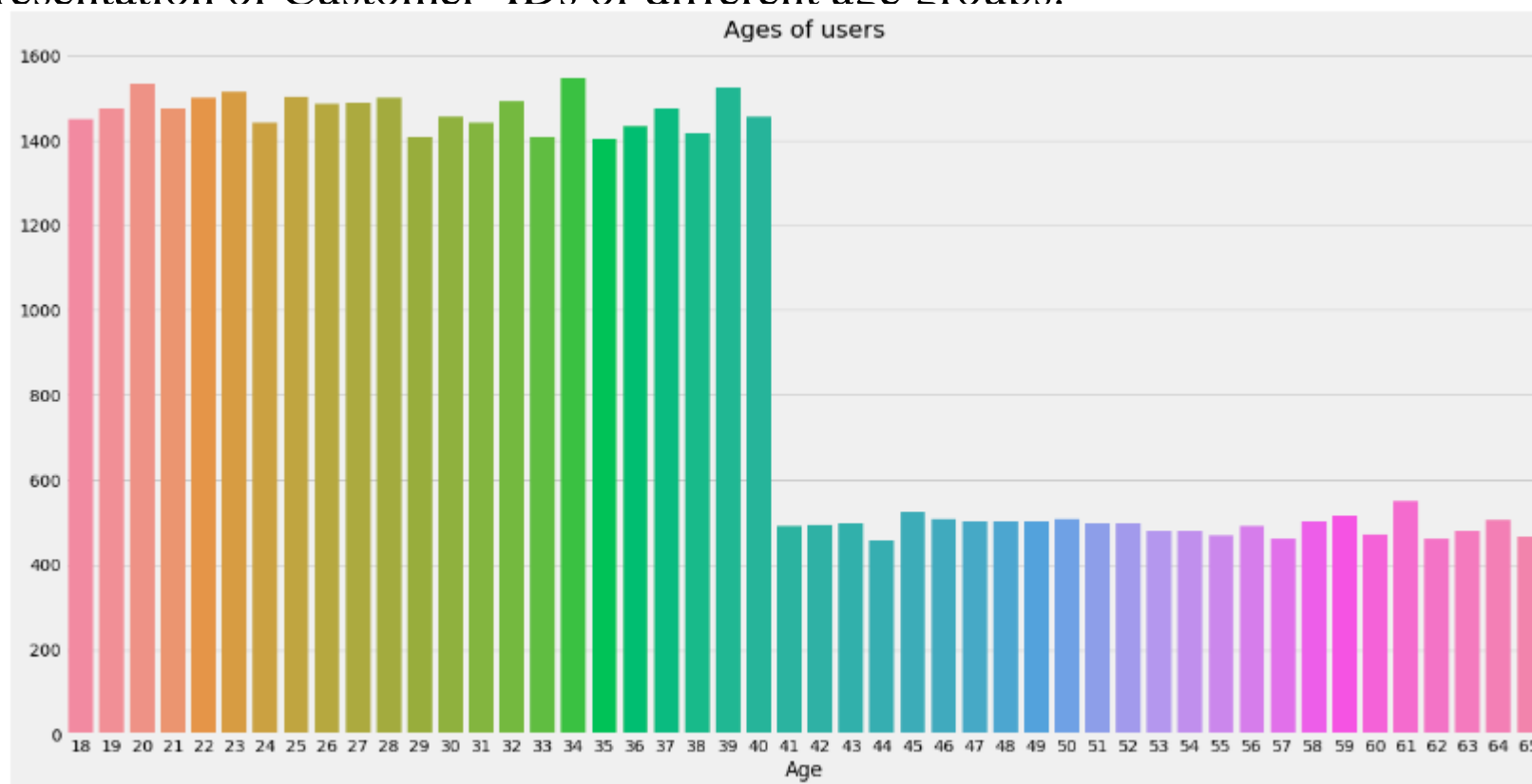
- ❑ Cost of trip is directly proportional to KM travelled.
- ❑ Cost of trip doesn't really depend on city travelled to.



Age share

Observation below:

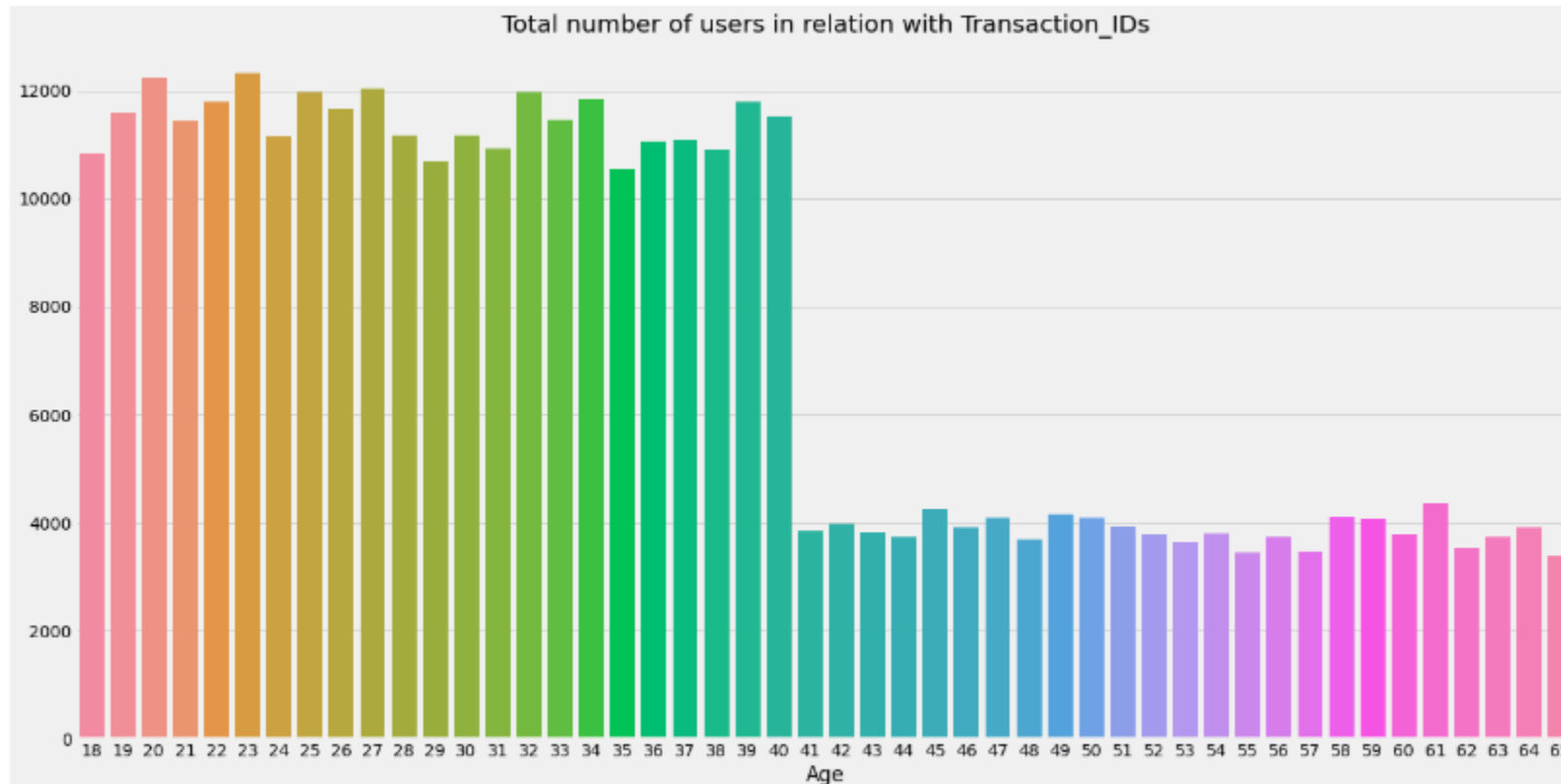
- ❑ Representation of Customer IDs of different age groups.



Relationship of Transactions and Age

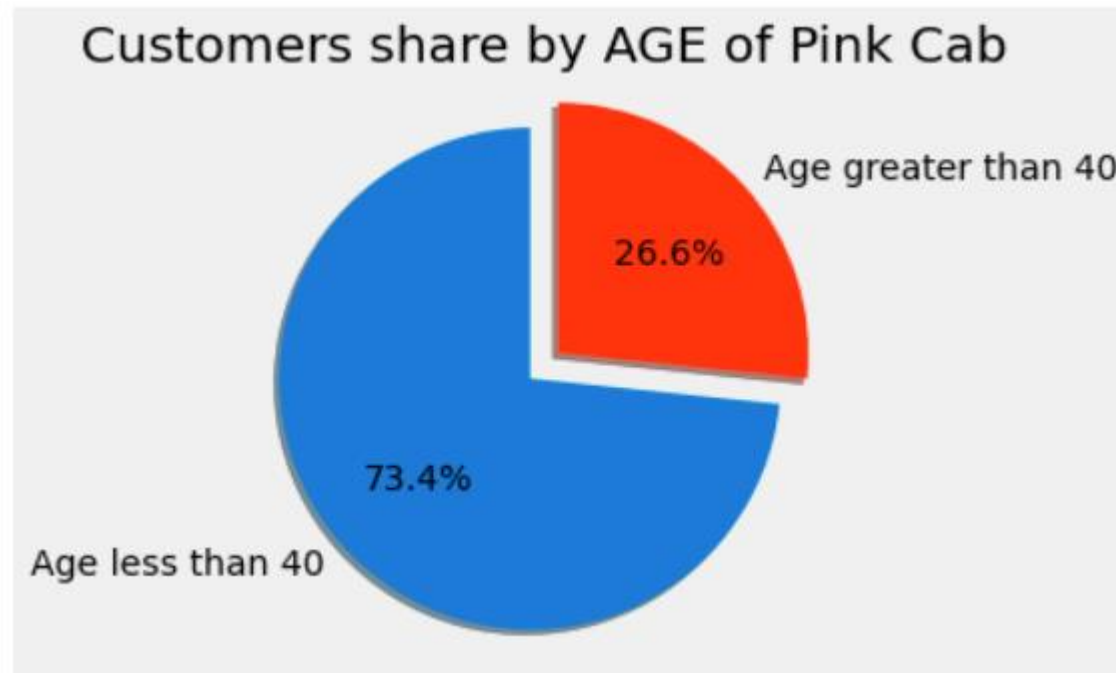
Observations:

- ❑ Transactions of cab usage are high within the ages of 18 - 40 years brackets. The representation show data of both cab companies.

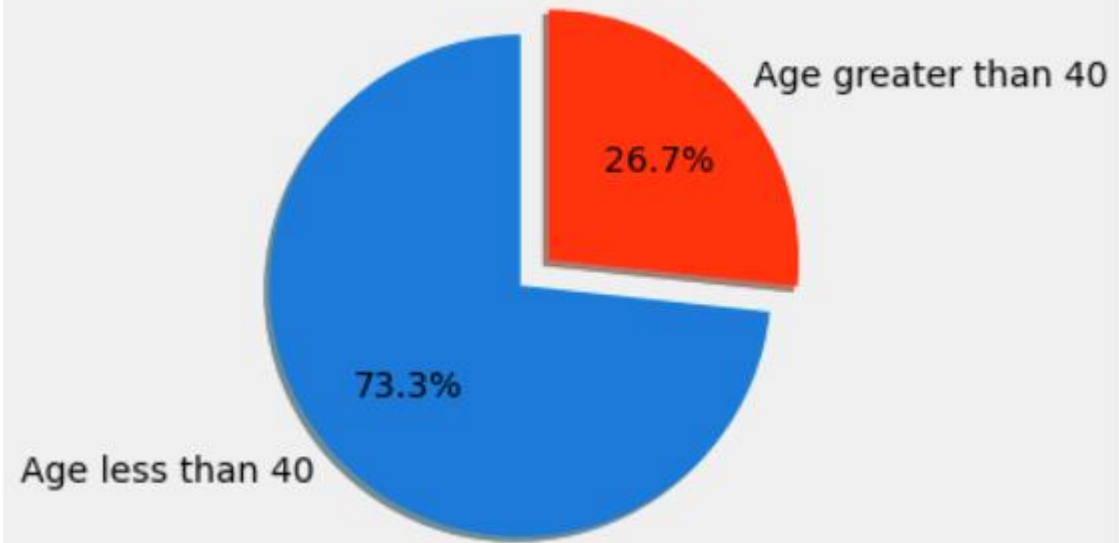


Customer share by age

- ❑ The customer share of Pink cab users in respective of age.
- ❑ The customer share of Yellow cab users.

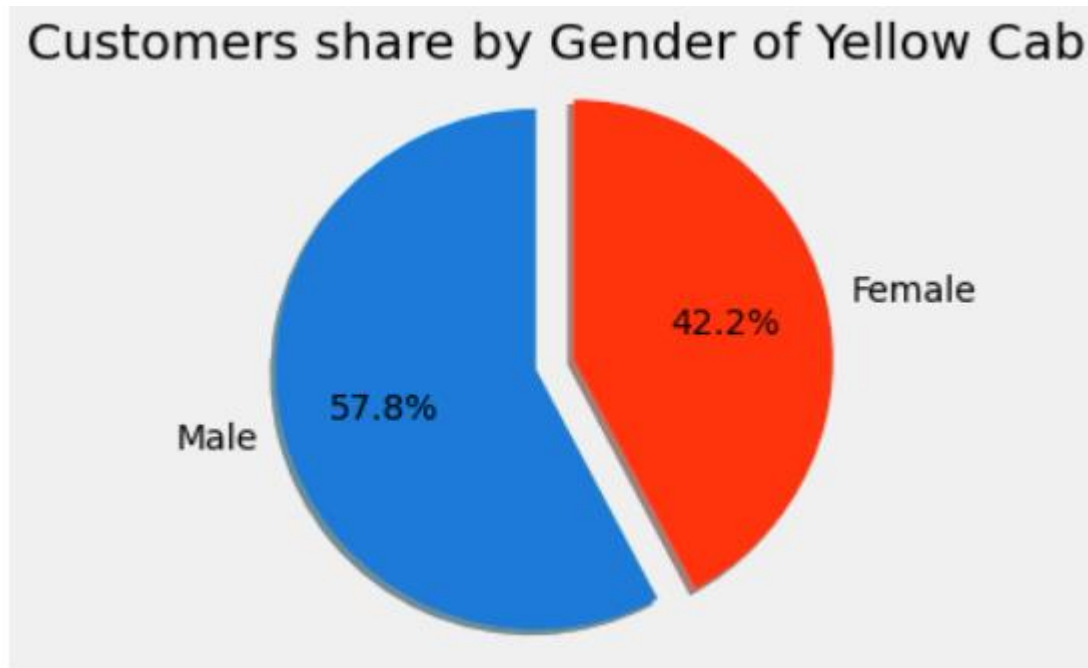


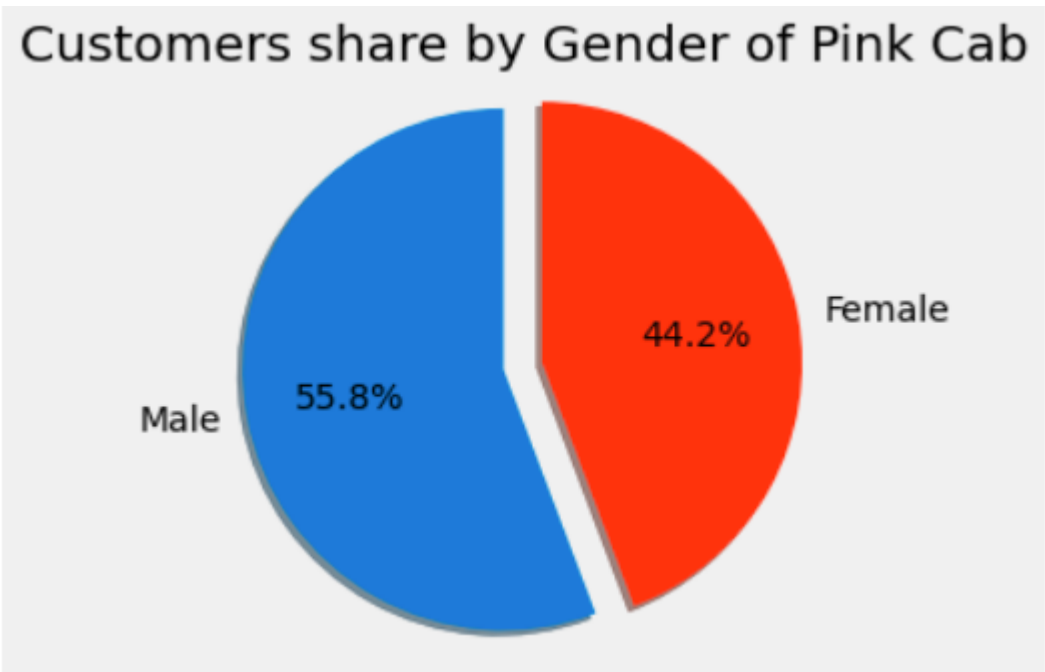
Customers share by AGE of Yellow Cab



Customer share by gender

- ❑ Customer share by gender of both cab companies.

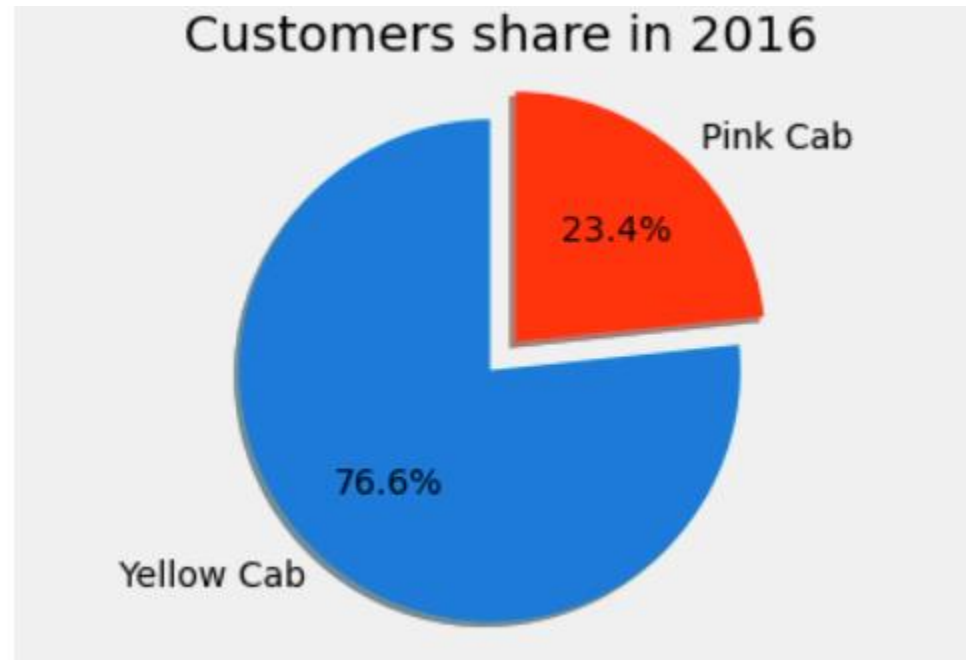




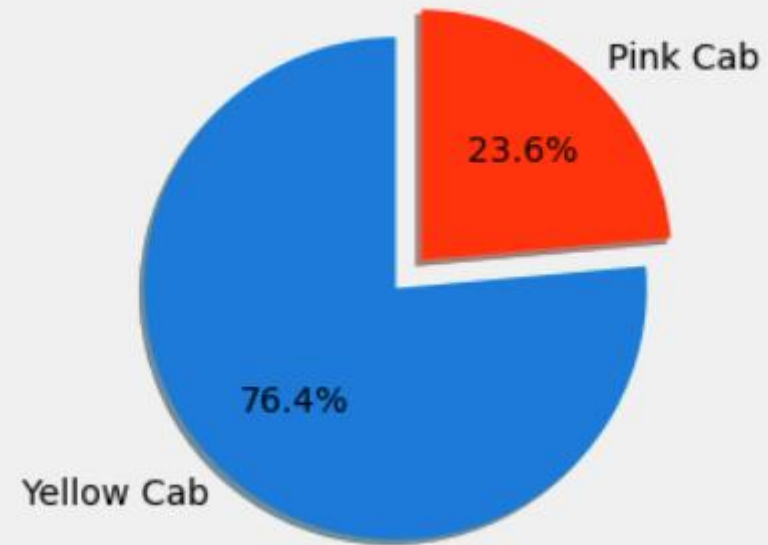
Customer share

Observation below within the three years:

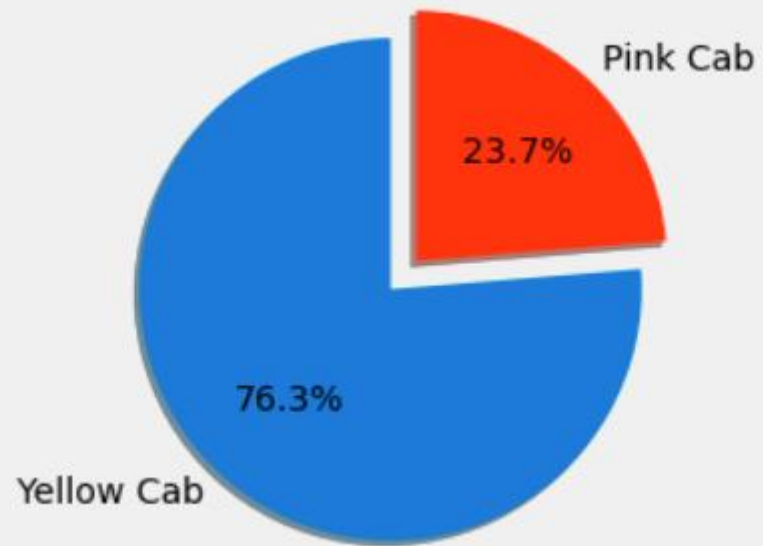
- ❑ From our analysis of the three years Yellow CAb has a high number of user, while Pink Cab has a low customer base but it increases each year within the three year period.



Customers share in 2017

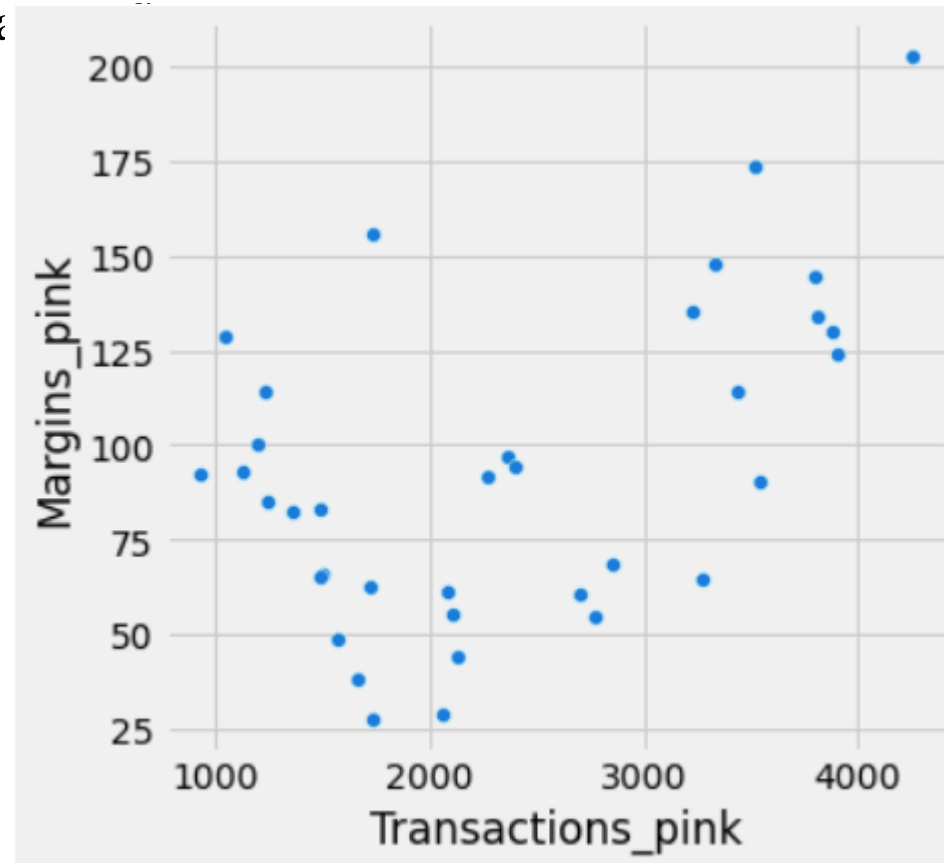


Customers share in 2018



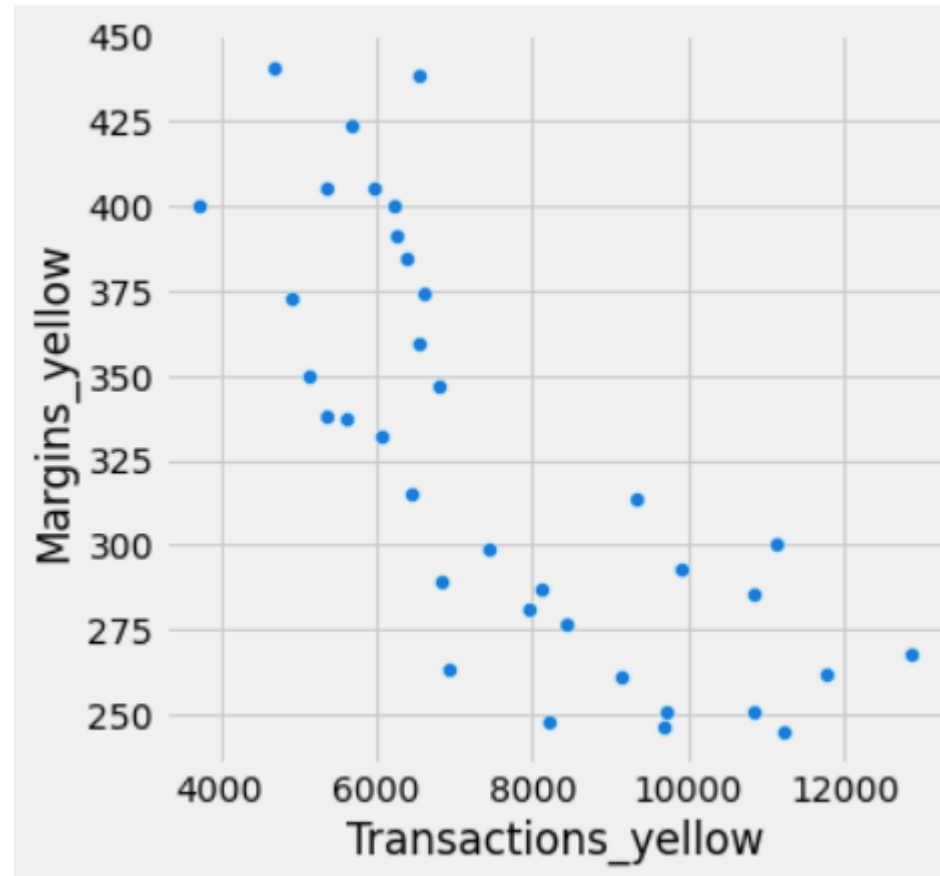
Margins relationship with Transactions

- ❑ Observation below depicts:
- ❑ This shows that Pink Cabs margins increase with increase in number of transactions.
- ❑ Margins can be labelled as



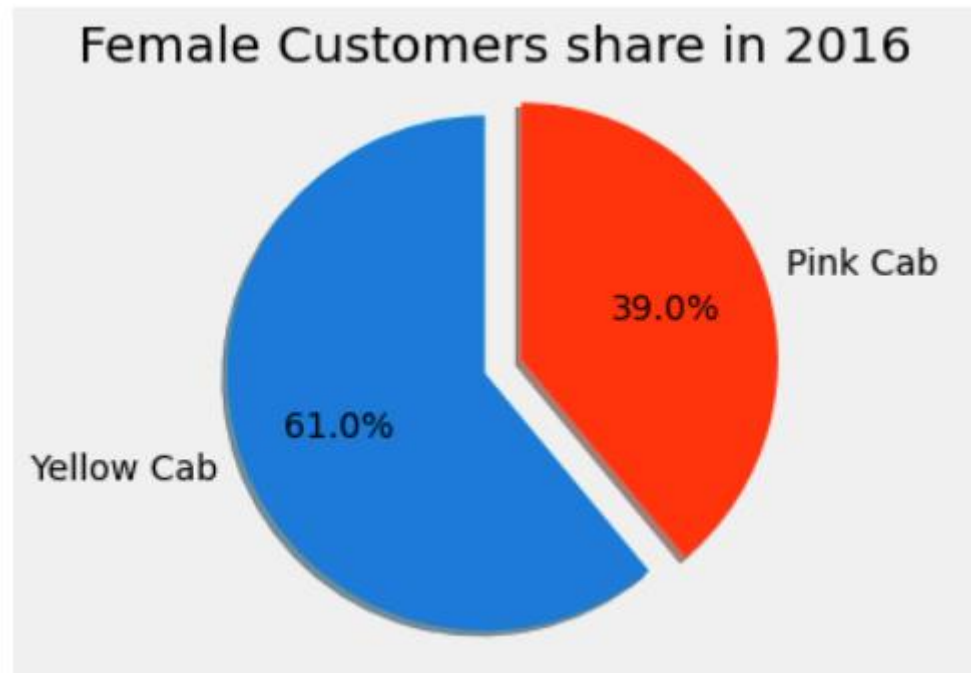
Margins vs transactions yellow cab

- ❑ Observation below rep:
- ❑ This shows that Yellow cab margins decrease with increase of transactions.
- ❑ This means with increase in transactions there a corresponding decrease in price charged and cost per trip.

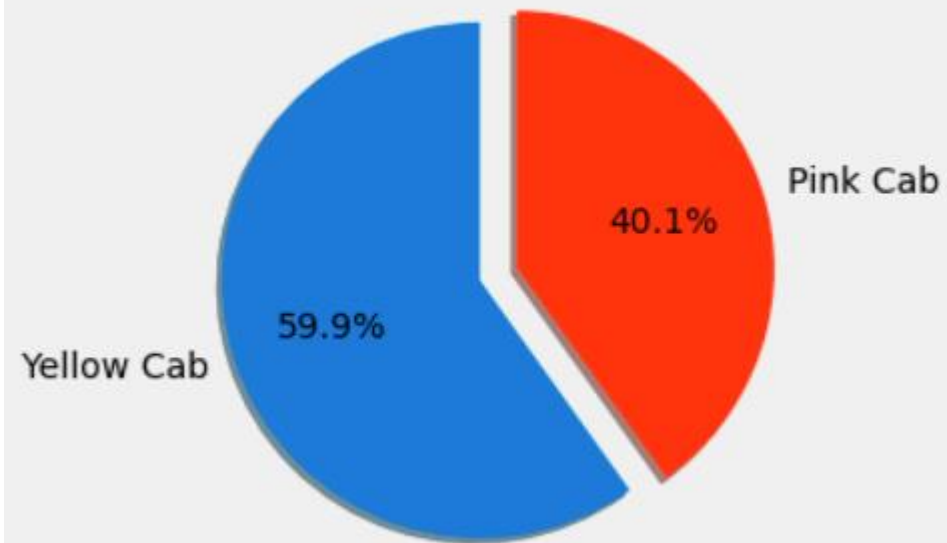


Female customer share yearly

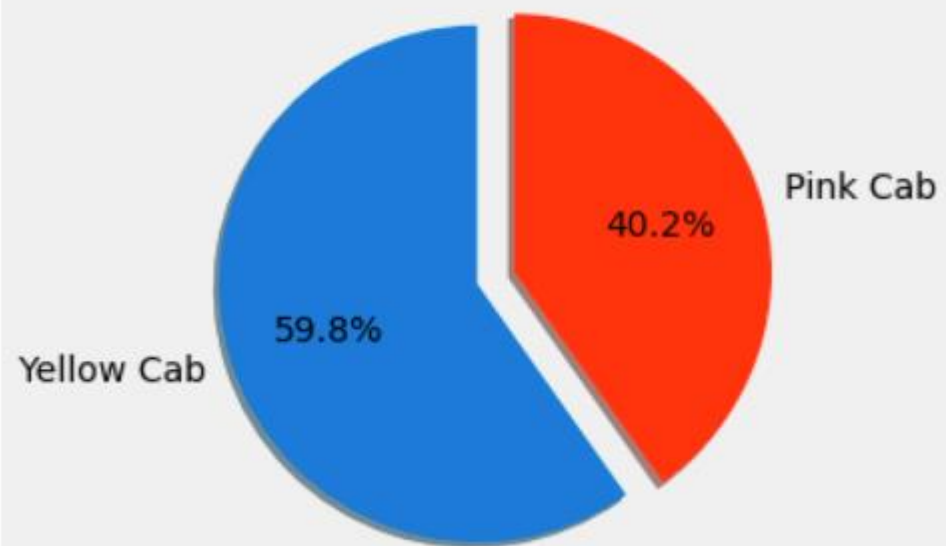
- ❑ Observation below:
- ❑ Yearly the number of female customers increase in Pink cab.



Female Customers share in 2017

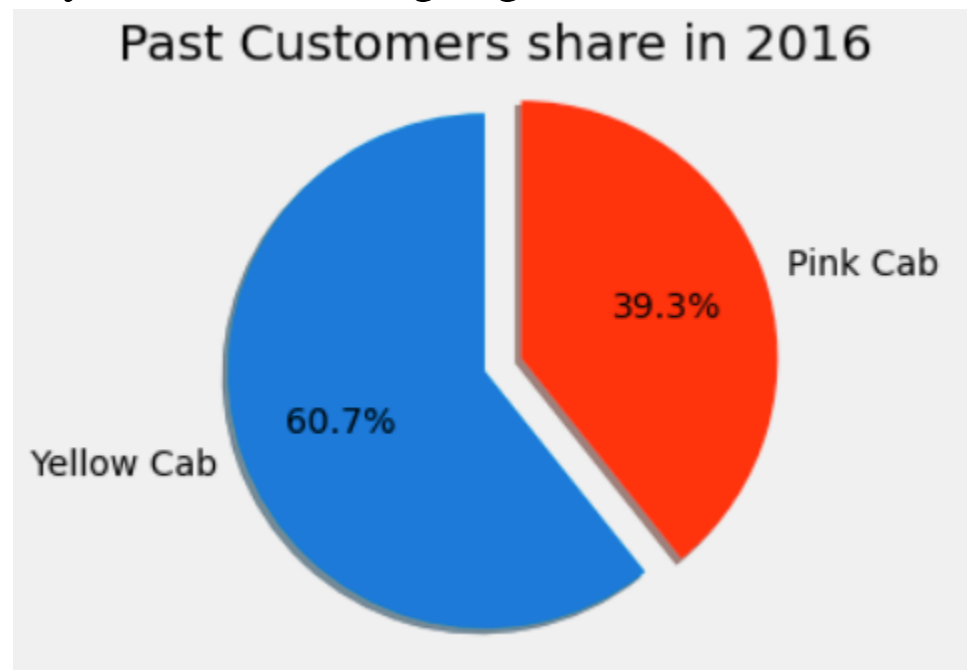


Female Customers share in 2018

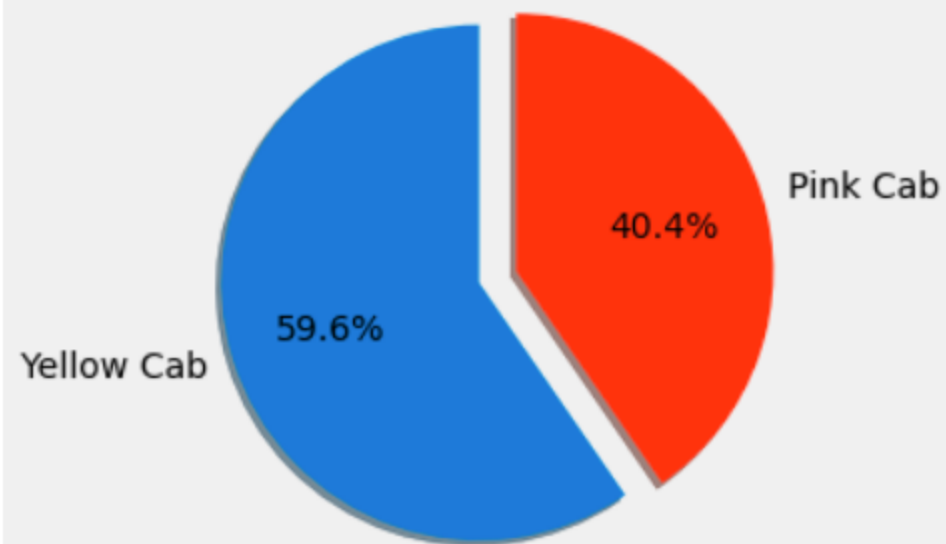


Customer share in relation with age

- ❑ Observation below:
- ❑ Yellow cab is losing elderly customers of ages greater than 45 to Pink cab.



Past Customers share in 2018



Conclusion of Analysis

- ❑ Initially in analysis outliers in price charged could be identified but through further analysis this could be a result of discounts offered by Yellow Cab Company.
- ❑ Yellow Cab offers discount which favour Female and elderly customers.
- ❑ Margins of Yellow Cab decreases with the increase of transactions while Pink Cab margins increases with increase of transactions. The model Yellow Cab usage is favourable this shows during peak periods prices charged or cost per trip rates decrease making it favourable.
- ❑ Pink Cab has a higher number of Young customer base in which shows potential of growth given improvement of their business model.
- ❑ Yellow Cab will maintain a monopoly in the market

On the basis of analysis I recommend that the company should invest in Yellow Cab.

Thank You