G2M Case Study

Virtual Internship

Xiaoying Yang
11-Jun-2022

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.

- Objective : Provide actionable insights to help XYZ firm in identifying the right company for making investment.

The analysis has been divided into four parts:

- Data Preprocessing

- Exploratory Data Analysis

- Results and Discussion

Data Preprocessing

- All four dataset were merged
- Original Variables: 14 Features
- Timeframe of the data: 2016-01-31 to 2018-12-31
- Total data points :359,392

- The profit was calculated as Price charged minus Trip Cost and identify as gross profit with other variables constant.

Cab_Data.csv – this file includes details of transaction for 2 cab companies

Total number of observations	<359392>
Total number of features	<7>
Base format of the file	<.csv >
Size of the data	<20.2 MB>

Customer_ID.csv – this is a mapping table that contains a unique identifier which links the customer's demographic details

Total number of observations	<49171>
Total number of features	<4>
Base format of the file	<.csv >
Size of the data	<1 MB>

Transaction_ID.csv – this is a mapping table that contains transaction to customer mapping and payment mode¶

Total number of observations	<440098>
Total number of features	<3>
Base format of the file	<.csv >
Size of the data	<8.58 MB>

City.csv – this file contains list of US cities, their population and number of cab users

| Total number of observations	<20>
Total number of features	<3>
Base format of the file	<.csv >
Size of the data	<759 Bytes>

Data Preprocessing

- Missing Value: None
- Duplicates: None
- Outliers: Assume Priced Charged and income are not outliers, they are all reasonable values
- Correlations: KM Traveled, Price Charged and Cost Trip are all highly correlated with each other, all make sense
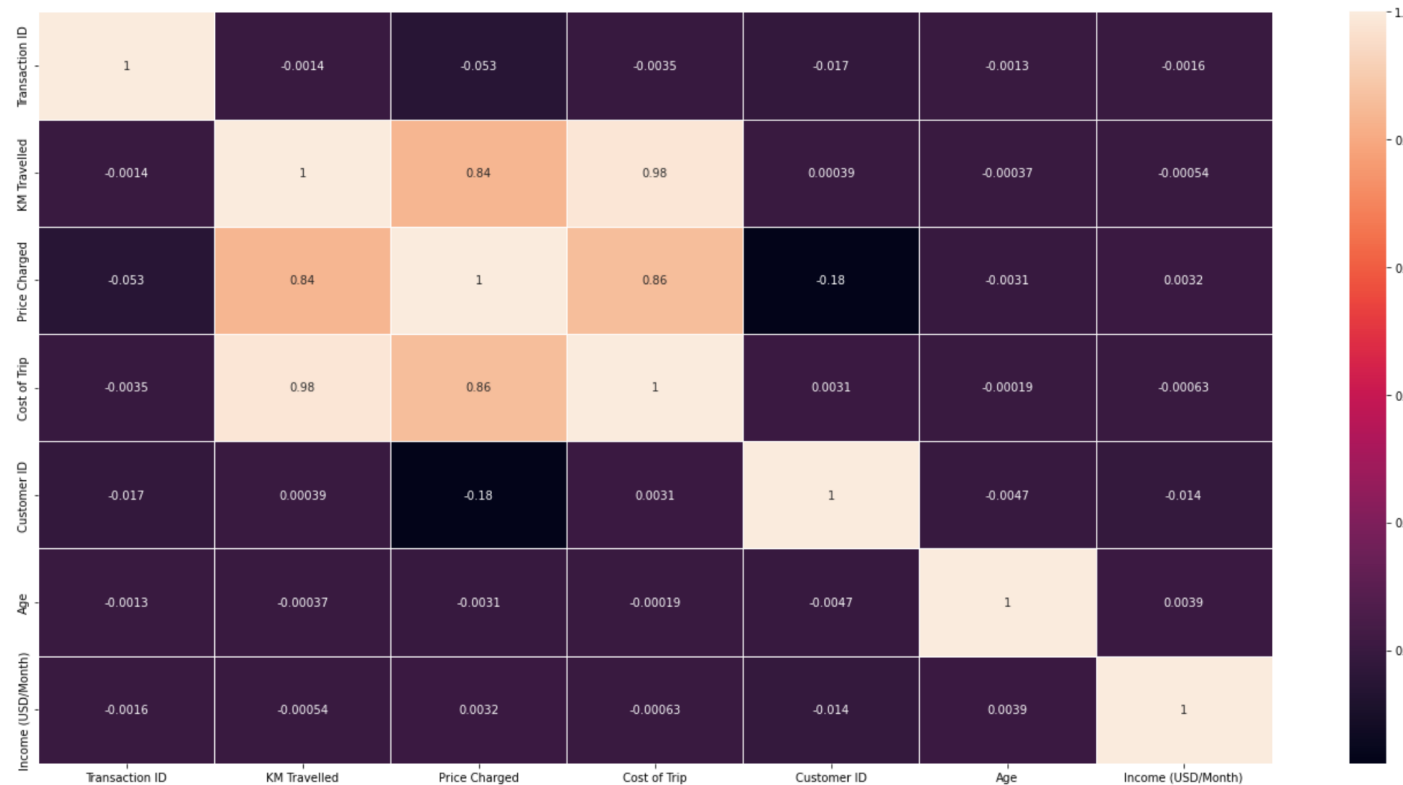
EDA

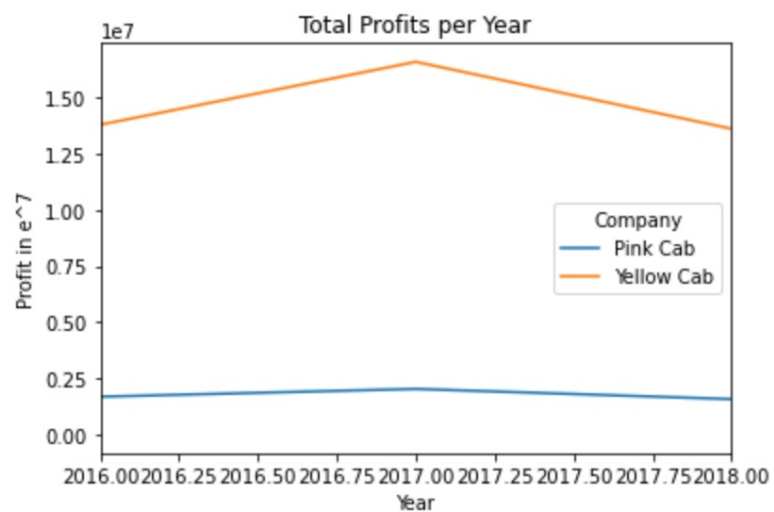Total Number of Customers
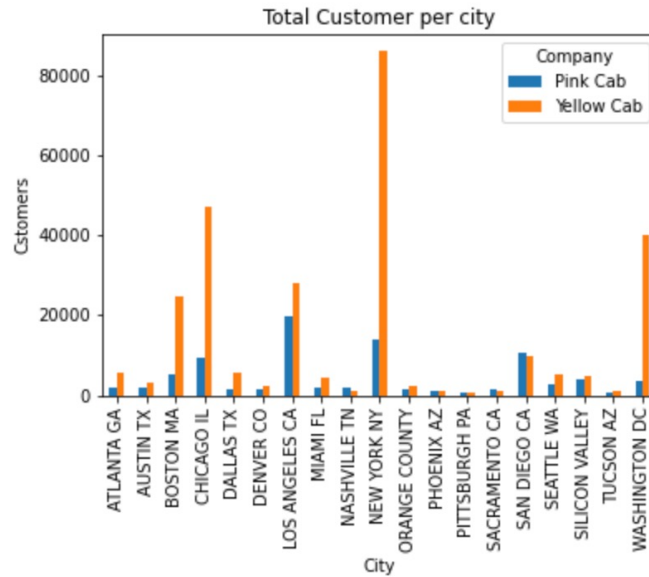
	Customer ID
Company	
Pink Cab	84711
Yellow Cab	274681

Yellow Cab has higher number of customers
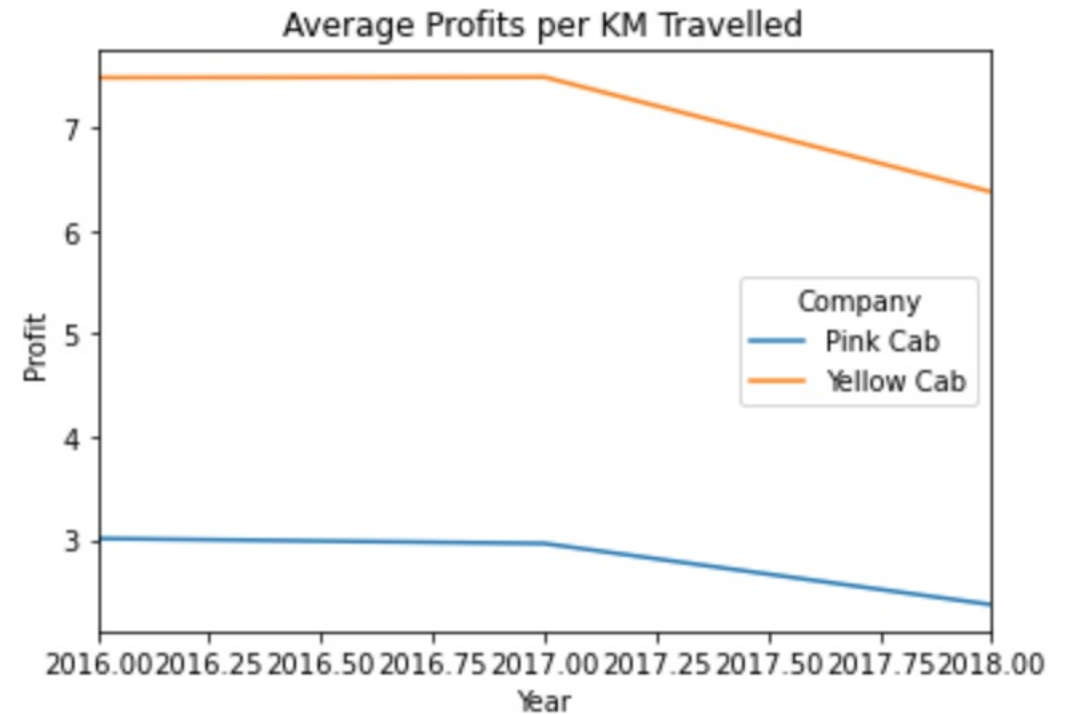
EDA

Total Profits
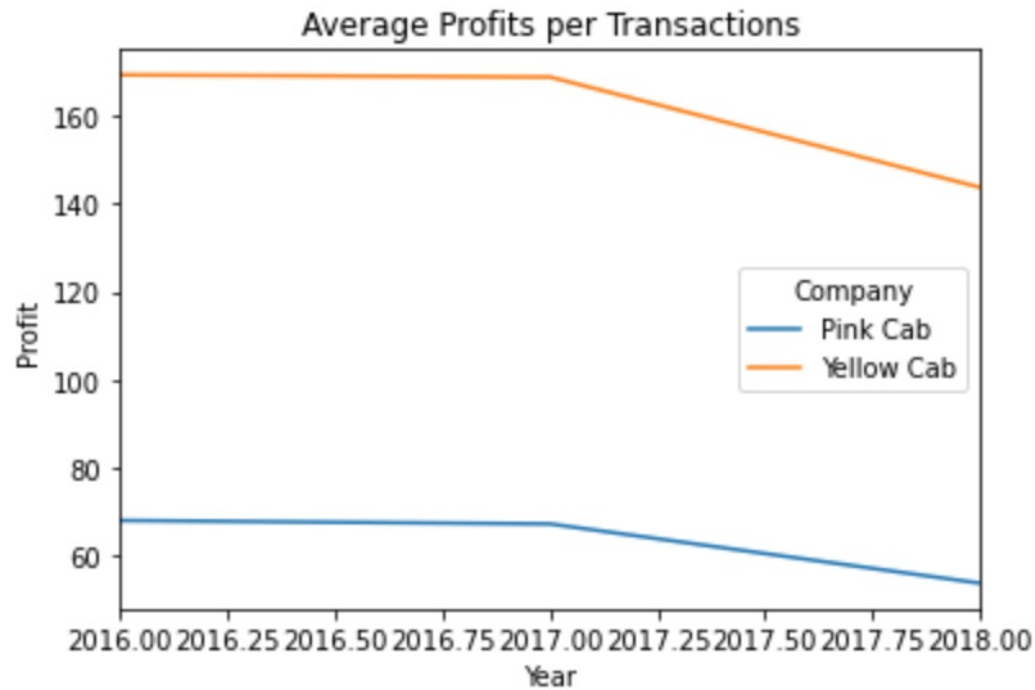

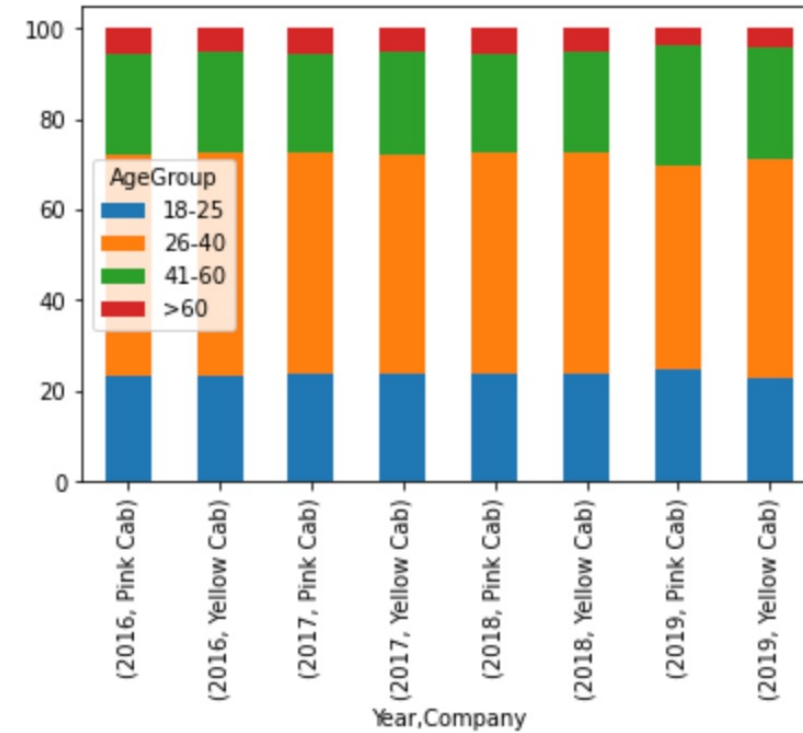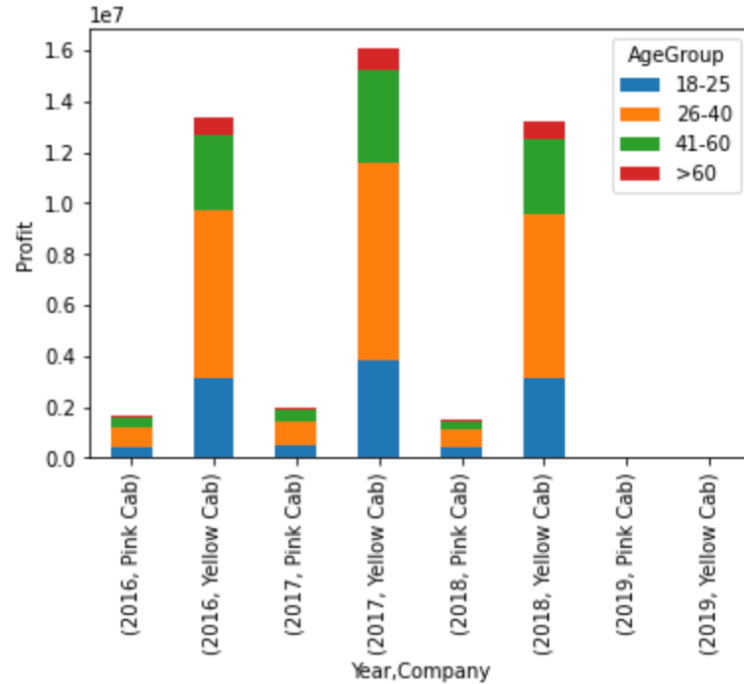| Company	Profit
Pink Cab	5.307328e+06
Yellow Cab	4.402037e+07

Yellow Cab has and higher profit in the dataset period

EDA

The average profits per KM travelled and average profits per transaction of Yellow Cab's were all approximately 2.5 times higher than Pink Cabs'
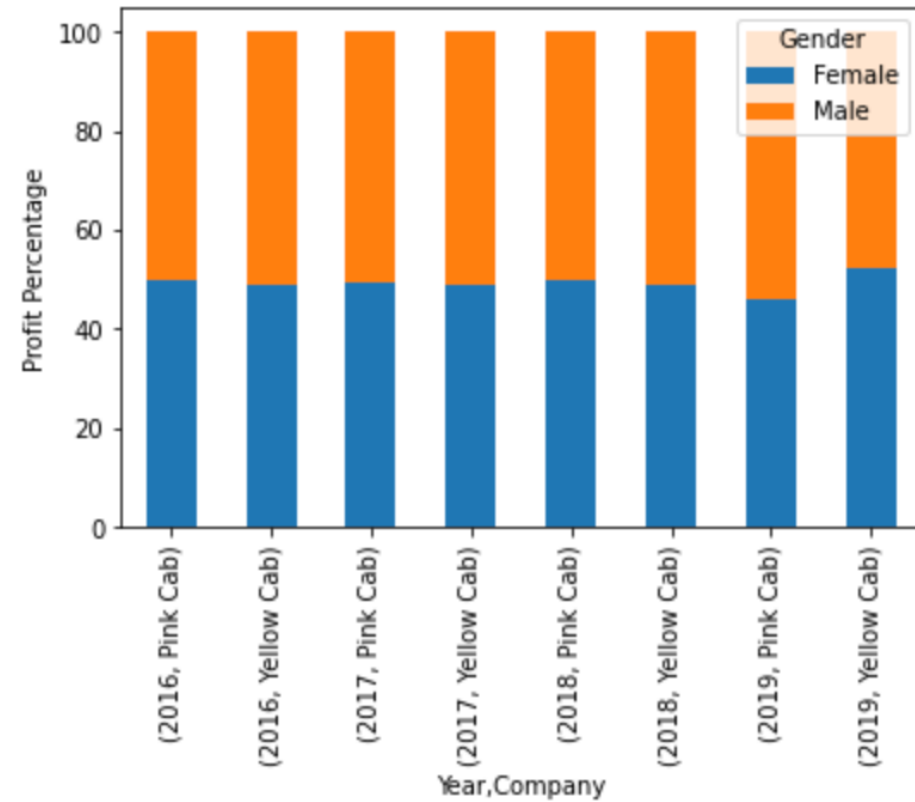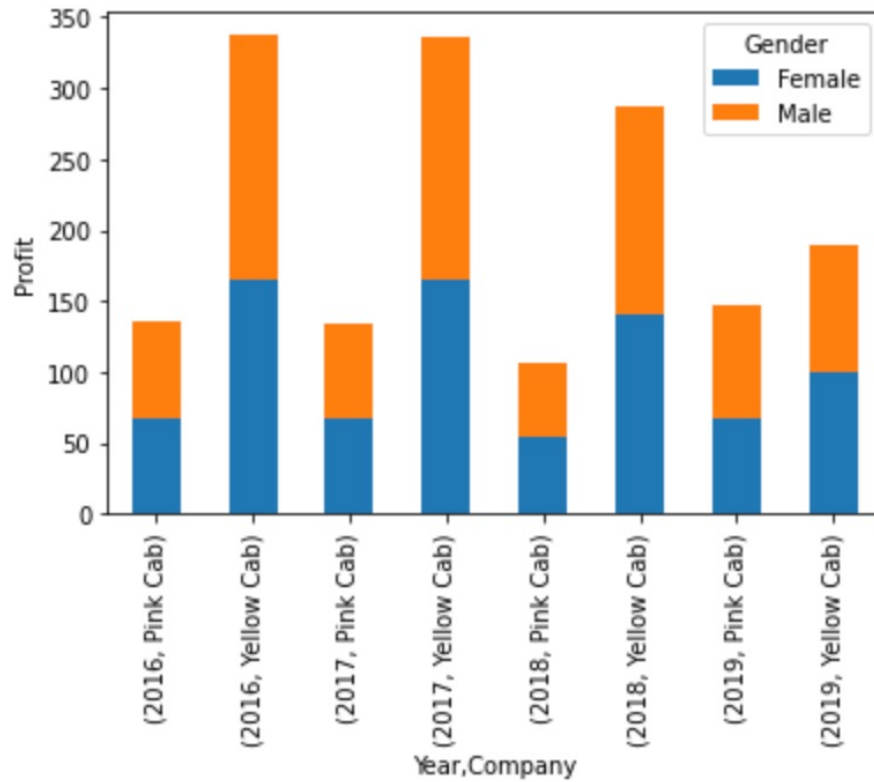
EDA

Age Group's influence for profits are approximately similar in two companies and they occupied with young-middle-aged adults(26-40) majority and followed by young adults (18-25) and middle-aged adults (41-60). And the total profits are related to the number of customers.

EDA

Gender's influence for profits are approximately similar in two companies and they occupied half and half

EDA
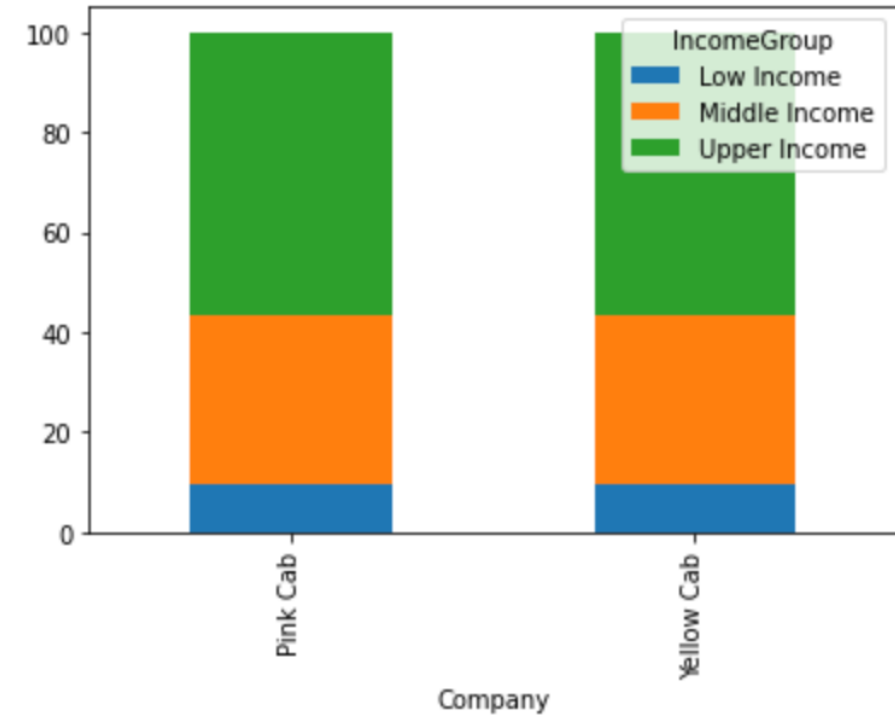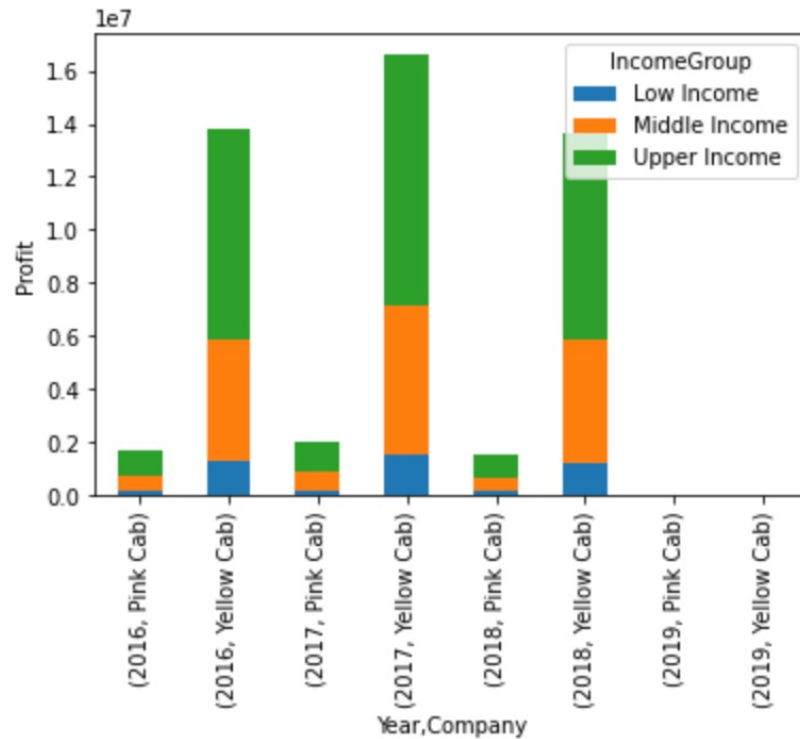
Income groups' influence for profits are approximately similar in two companies with upper-class income the most, followed by the middle-class income and then low incomes. The proportion of the three categories of people in the two companies is approximately equal

Results and Discussion

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

- **Total Customers:** Yellow cab has 3.5 times customers than Pink Cab and it also has higher customers in almost all cities especially in modern cities such as New York, Chicago, Boston and DC.

- **Total Profits:** Yellow has approximately 8 times total profits than Pink Cab in three years, every year the profits is much higher than Pink Cab's, even though in 2018 compared to 2017 there was a decline, but still up to 6 times more

- **Average Profit per KM/Transaction:** Yellow cab's average profit per KM and per transaction is almost three times the average profit per KM and transactions of the Pink cab.

- **Age wise Reach :** Age Group's influence for profits are approximately similar in two companies and they occupied with young-middle-aged adults(26-40) majority and followed by young adults (18-25) and middle-aged adults (41-60). And the total profits are related to the number of customers.

- **Gender wise Reach :** Gender's influence for profits are approximately similar in two companies and they occupied half and half

- **Income wise Reach :** Income groups' influence for profits are approximately similar in two companies with upper-class income the most, followed by the middle-class income and then low incomes. The proportion of the three categories of people in the two companies is approximately equal

On the basis of above point , we will recommend Yellow cab for investment.

Thank You