Data Glacier

Your Deep Learning Partner

G2M insight for Cab Investment firm

Exploratory Data Analysis

Presentation by Mazen Hawwa

Internship batch LISUM04

11th October 2021

Introduction and Problem Statement:

Problem summary:

XYZ is a private firm in the US that wants to make an investment in the Cab industry.

Objective:

We need to analyze the given datasets and derive actionable insights to determine which of the 2 cab companies should be invested in by XYZ company.

Available Data Sets:

Data Description:

1- Cab_Data.csv – Main data file: transactions for 2 cab companies over 3 years.

2- Customer_ID.csv – mapping table: contains a unique identifier which links the customer's demographic details

3- Transaction_ID.csv – mapping table: contains transaction to customer mapping and payment mode.

4- City.csv – mapping table: contains list of US cities, their population and number of cab users

Data Processing and Summary:

Approach:

1 – After checking for and handling duplicated and missing values in all datasets, they were all joined together to form a master data set with comprehensive information on transactions, customers, individual trips, and the cities in which they occurred.

2 – **6** new features were created to offer new insight into the data and help with our analysis.

Data summary:

359384 – Data Entries

21 features – including **6** created features

Date range from **2016.1.4** to **2019.1.2**

Overview of analysis:

Categories:

1. Examining data ranges and features – processing the data.

2. Company vs Company analysis

3. Analysis of market factors and conditions affecting both company operations.

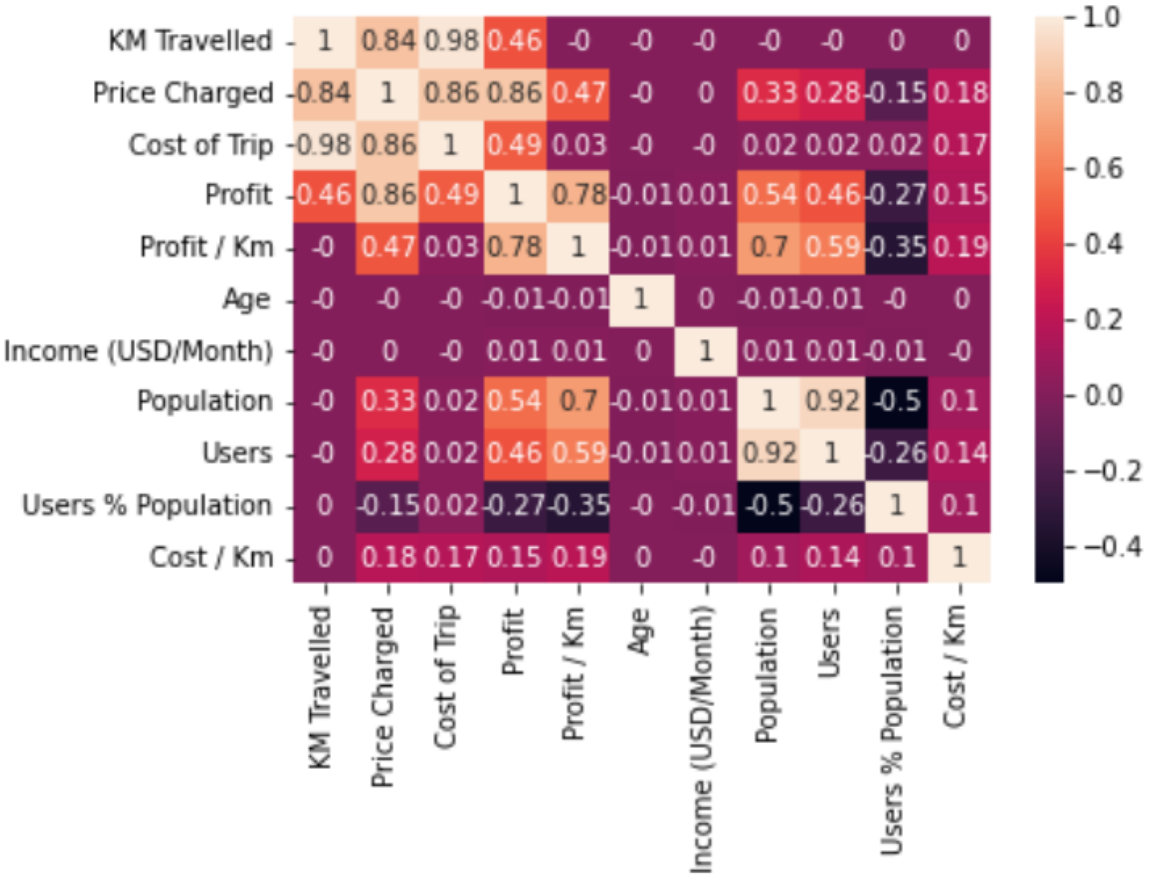
4. Conclusion and recommendation

Overview and understanding our data:

Brief Correlation examination:

1. Cost of Trip correlates almost perfectly with the KM Traveled at **0.98**.
2. Price Charged correlates less strongly (**0.84**) with the KM Traveled
3. weak correlation value of **0.46** between profit and KM Traveled
4. negative weak correlation of KM Traveled with 'Profit / Km' (**-0**)

The above implies that that for longer trips, we have reduced profitability



Overview and understanding our data:

Brief Correlation examination – Part 2:

1. Profit values have a moderate correlation (**0.54**) with the size of the population of a city.
2. Profit has slightly weaker correlation with the number of registered Cab users in a city (**0.46**)
3. 'Profit / Km' has a strong correlation (**0.7**) with the size of the population of a city and significant correlation with the number of registered users in a city at **0.59**.

We can conclude that operating in bigger cities is more profitable.

Other Observations:

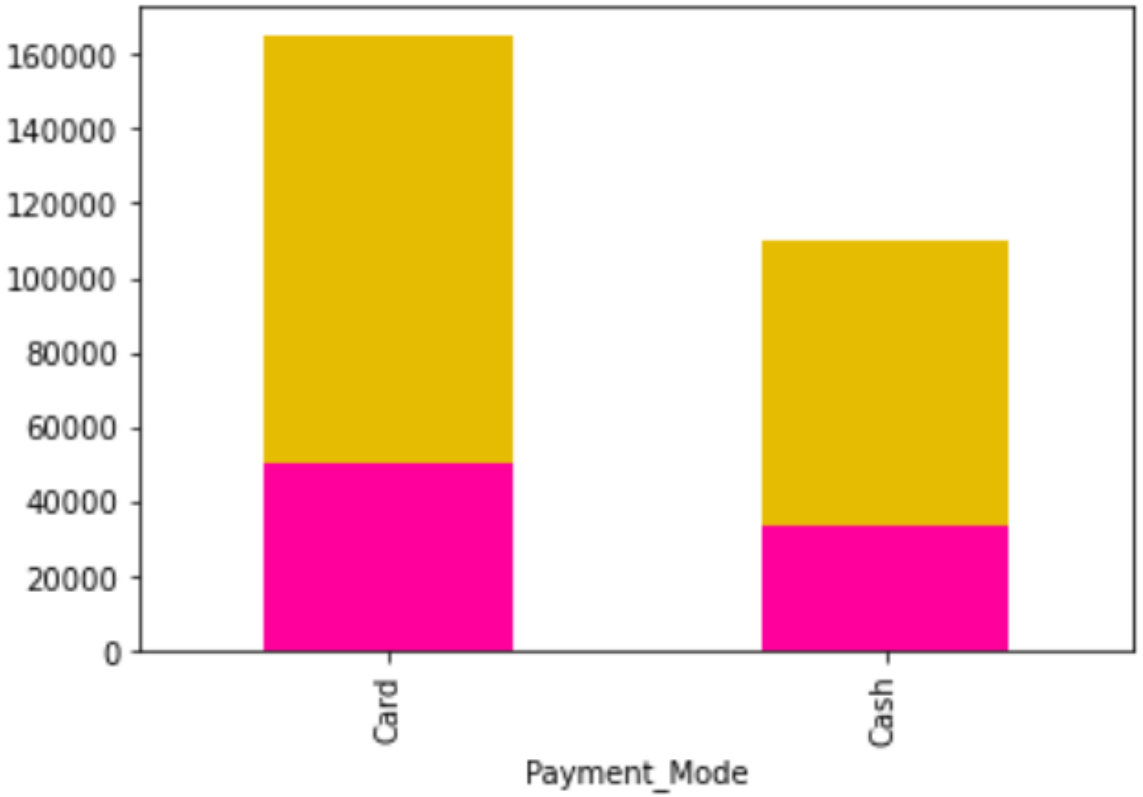
From the correlation table:

We can also see that the % of registered cab users in a city is moderately negatively correlated (**-0.5**) with the size of the population, so we can infer that larger cities have a smaller % of their population as registered Cab users.

Other Observations 2:

Payment Modes:

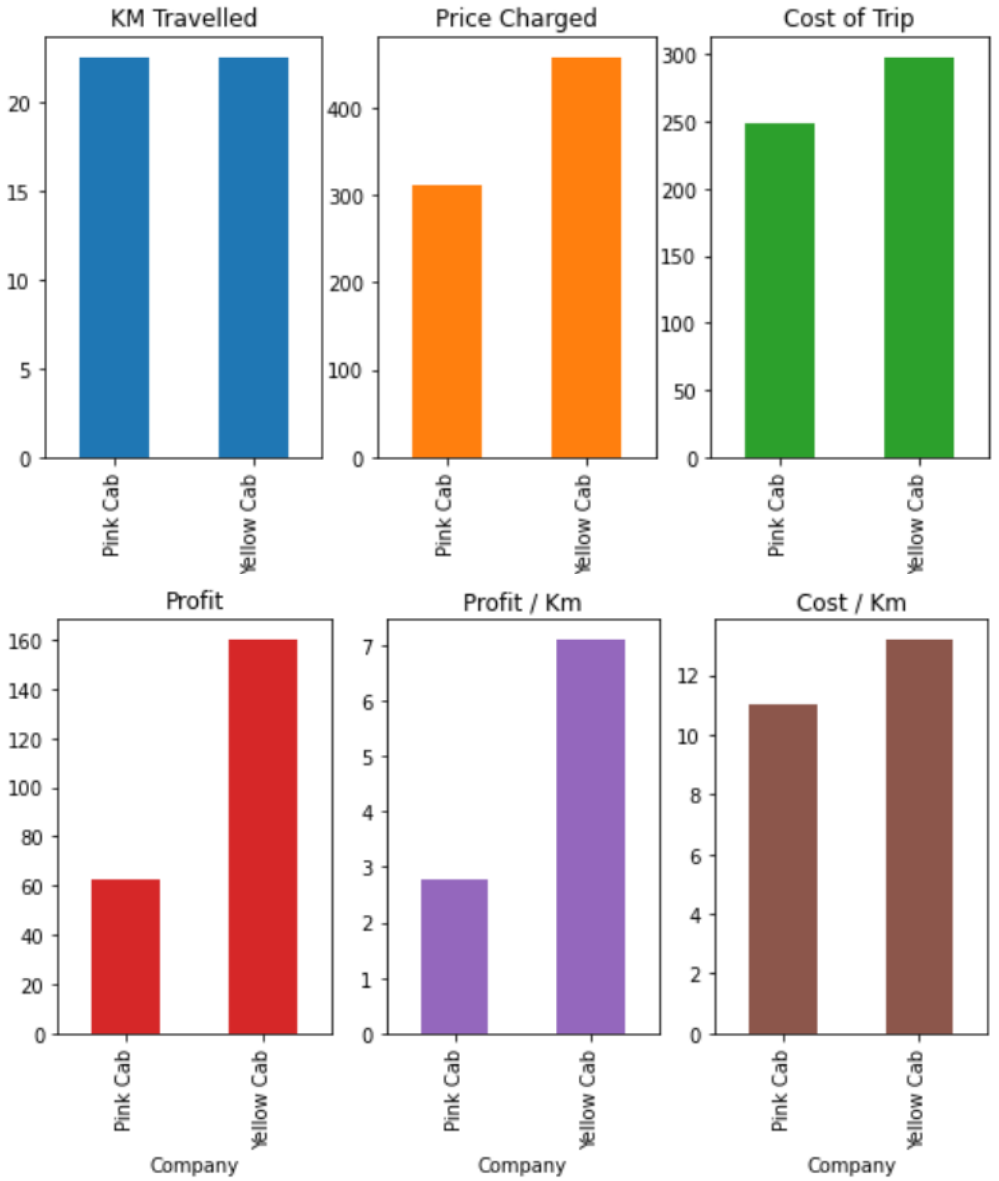
We can see from the data that both companies receive roughly **40%** of payments in cash while the rest is received by card.



Pink Cab vs Yellow Cab Comparisons:

Checking Average Values for several features:

1. We can see that both companies cover roughly an average of **22.6km** per trip.
2. Average charge per trip for yellow cab is **$458** while average charge per trip of pink cab is **$310** - which is a big difference
3. Average cost per trip was also higher for yellow cab (**$297**) compared to average cost for pink cab of **$248**.
4. Average profit per trip is much higher for yellow cab at **$160** vs **$62** for pink cab.



Pink Cab vs Yellow Cab Comparisons 2:

Checking Average Values for several features:

5. Average profit per km is much greater for yellow cab at **$7.1** vs **$2.7** for pink cab as expected.
6. Average cost per km is lower for pink cab at **$11** per km vs **$13.2** as expected from point 3 previously.

- We can see from the first set of visualizations that Yellow cab is much more profitable than Pink cab company despite pink cab being more efficient and having lower cost per trip than yellow cab.

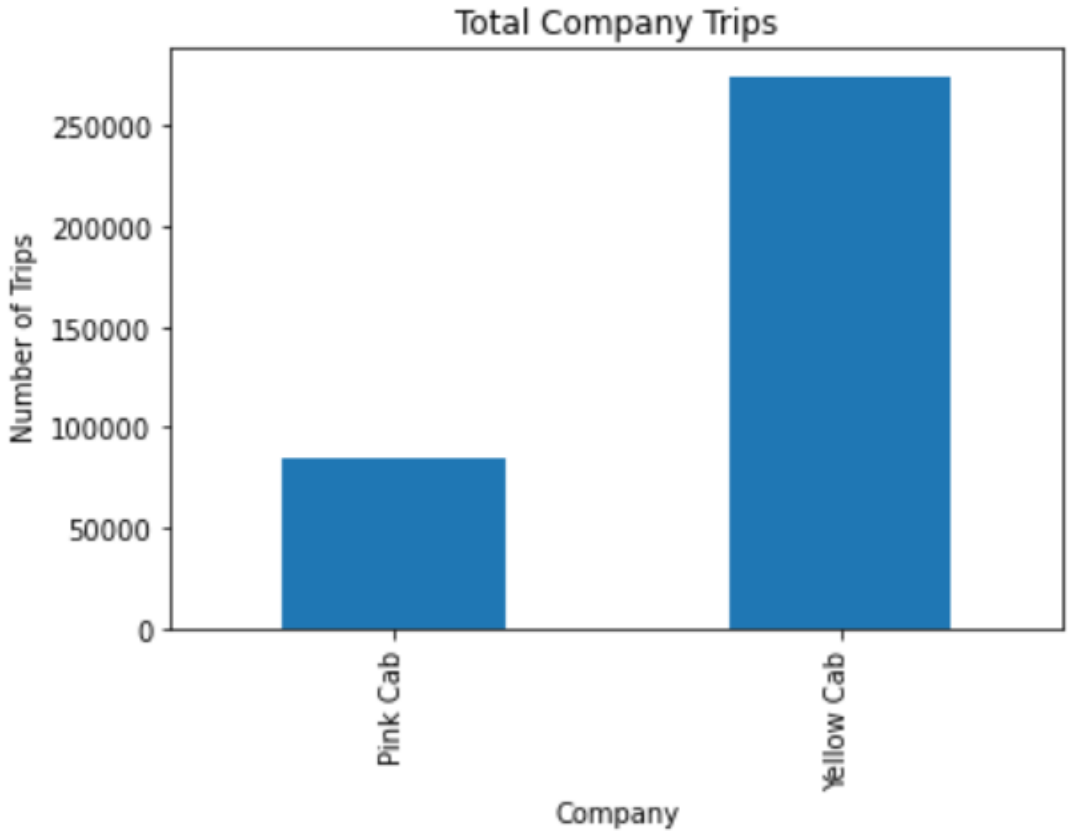
- Yellow cab is more than twice as profitable as pink cab on average per trip and per km traveled

Pink Cab vs Yellow Cab Comparisons 3:

Total Company trips in 3 years:

Yellow Cab total trips: **274,674**
Pink Cab total trips: **84,710**

Yellow Cab Company has more than **3** times total trips as Pink Cab.

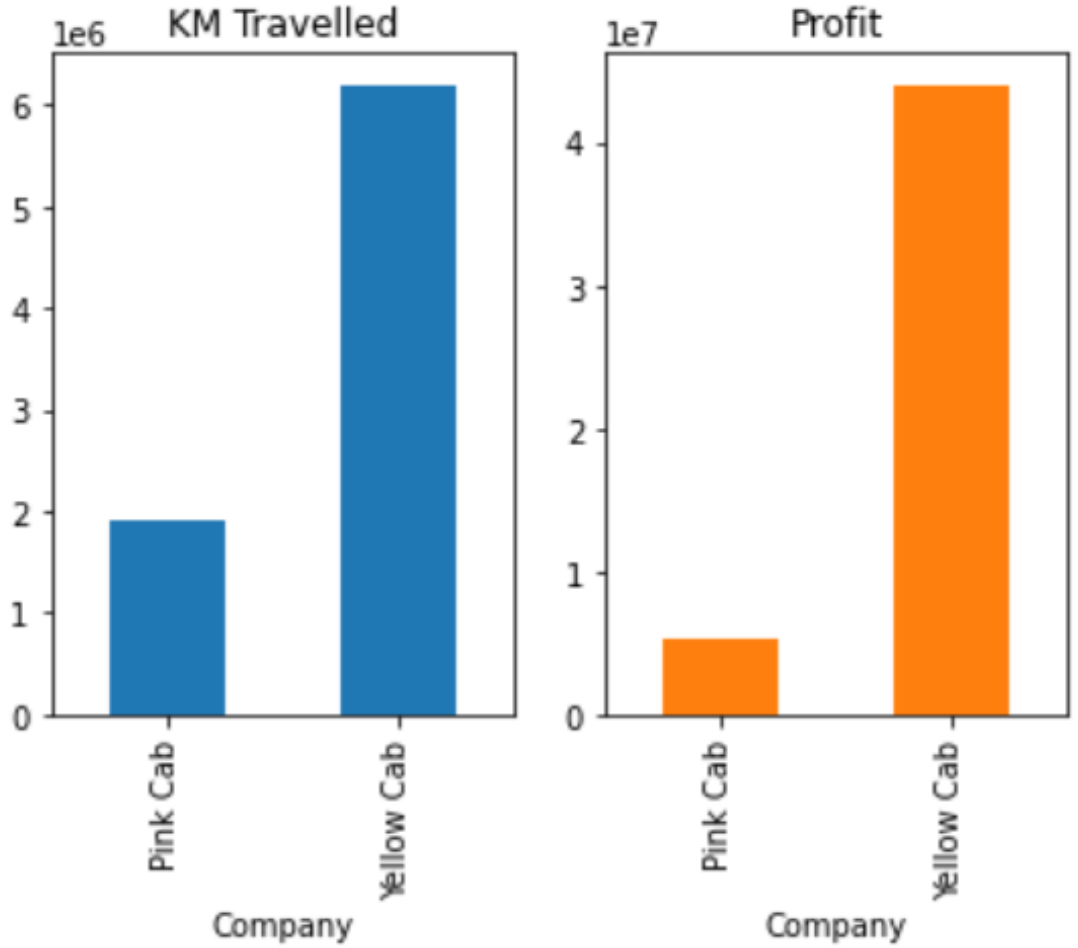


Pink Cab vs Yellow Cab Comparisons 4:

Total KM traveled and Total Profit:

- Total km traveled for yellow cab company more than **3** times the total km covered by pink cab company, with **6,199,216** km covered by yellow cab vs **1,911,036** km for pink cab company.

- Yellow cab company generated **8.29** times the profit of pink cab company over the 3 year period - with yellow cab generating **44 million** dollars in profit vs **5.3 million** dollars generated by pink cab company over the same period.



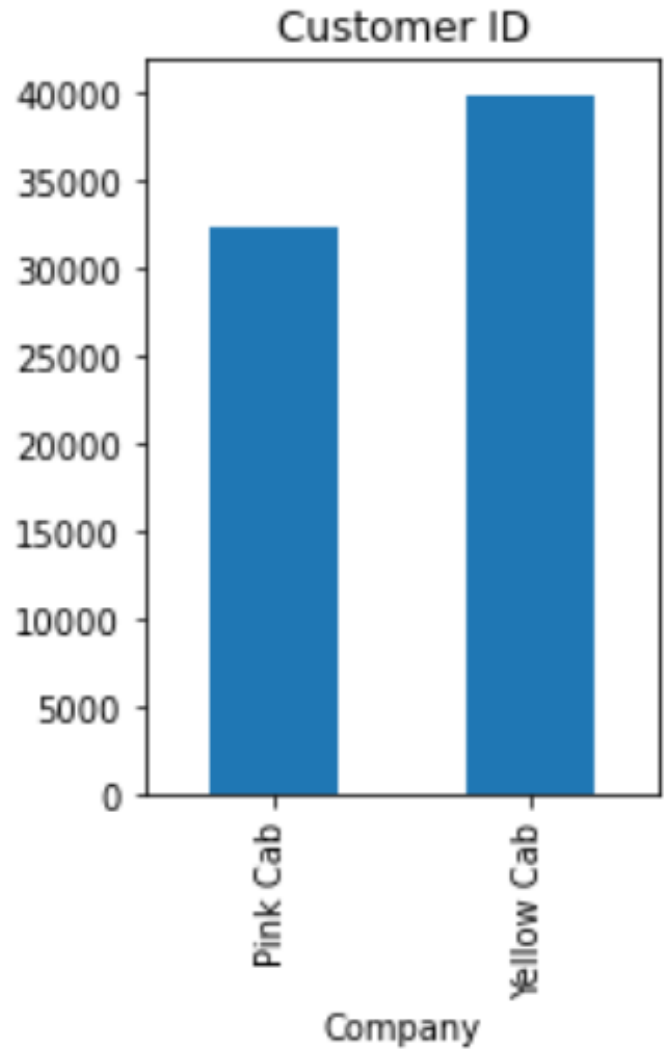
Pink Cab vs Yellow Cab Comparisons 5:

Number of unique users served by each company:

Users served by Pink Cab: **32,330**
Users served by Yellow Cab: **39,896**
Out of **46,148** unique users in our data set

- Pink cab has served approximately **70%** of the total users included in our data set

- Yellow cab has served approximately **86%** of the total users included in our data set



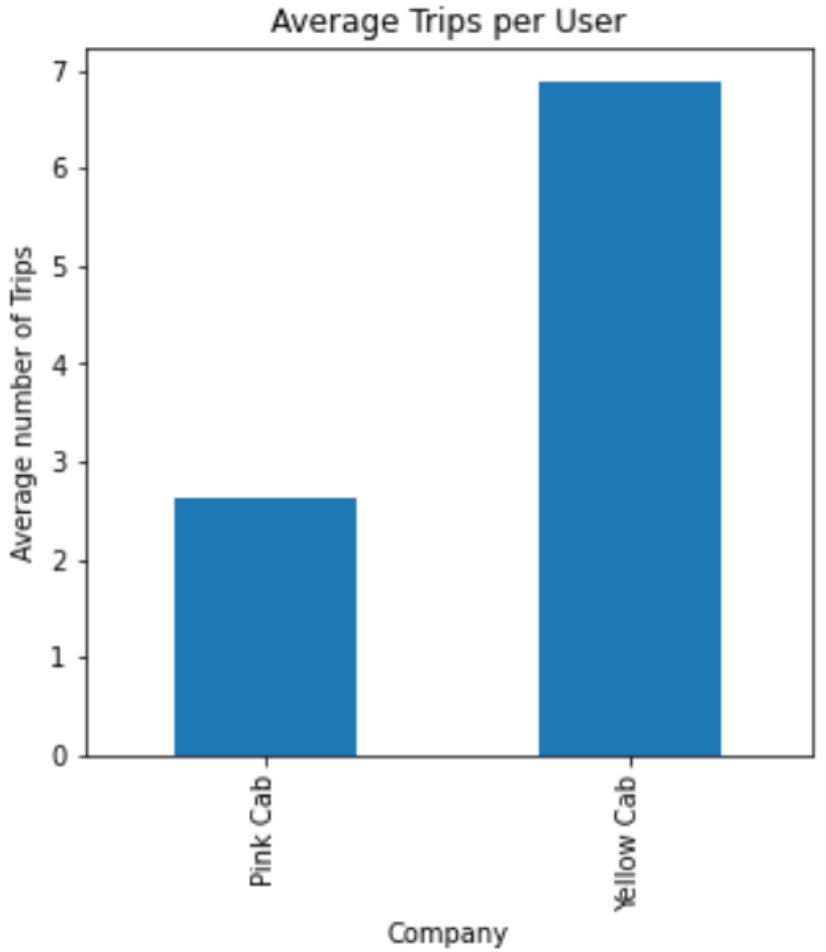
Pink Cab vs Yellow Cab Comparisons 6:

Average number of times a user used each cab company over the past 3 years:

Pink Cab was hired **2.620167** times on average by each one of their users.

Yellow Cab was hired **6.884750** times on average by each one of their users.

On average, each user of Yellow Cab used their services **2.63** times more than Pink Cab users used Pink Cab services.



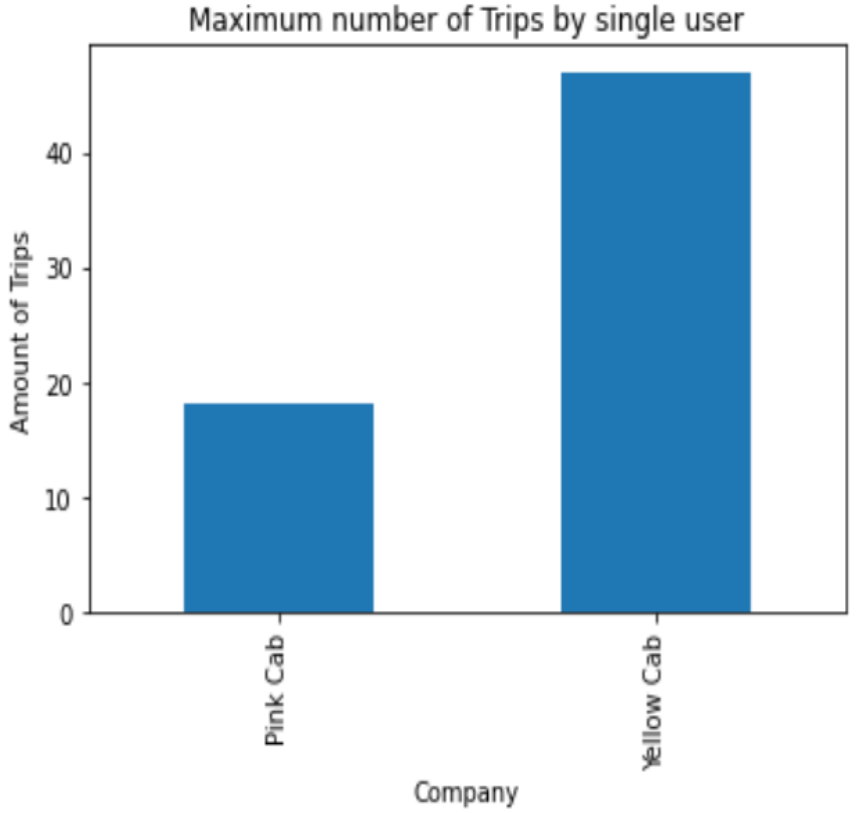
Pink Cab vs Yellow Cab Comparisons 7:

Maximum number of times a user hired each cab company over the past 3 years:

The maximum number of times Pink Cab was hired by a single user is **18** times.

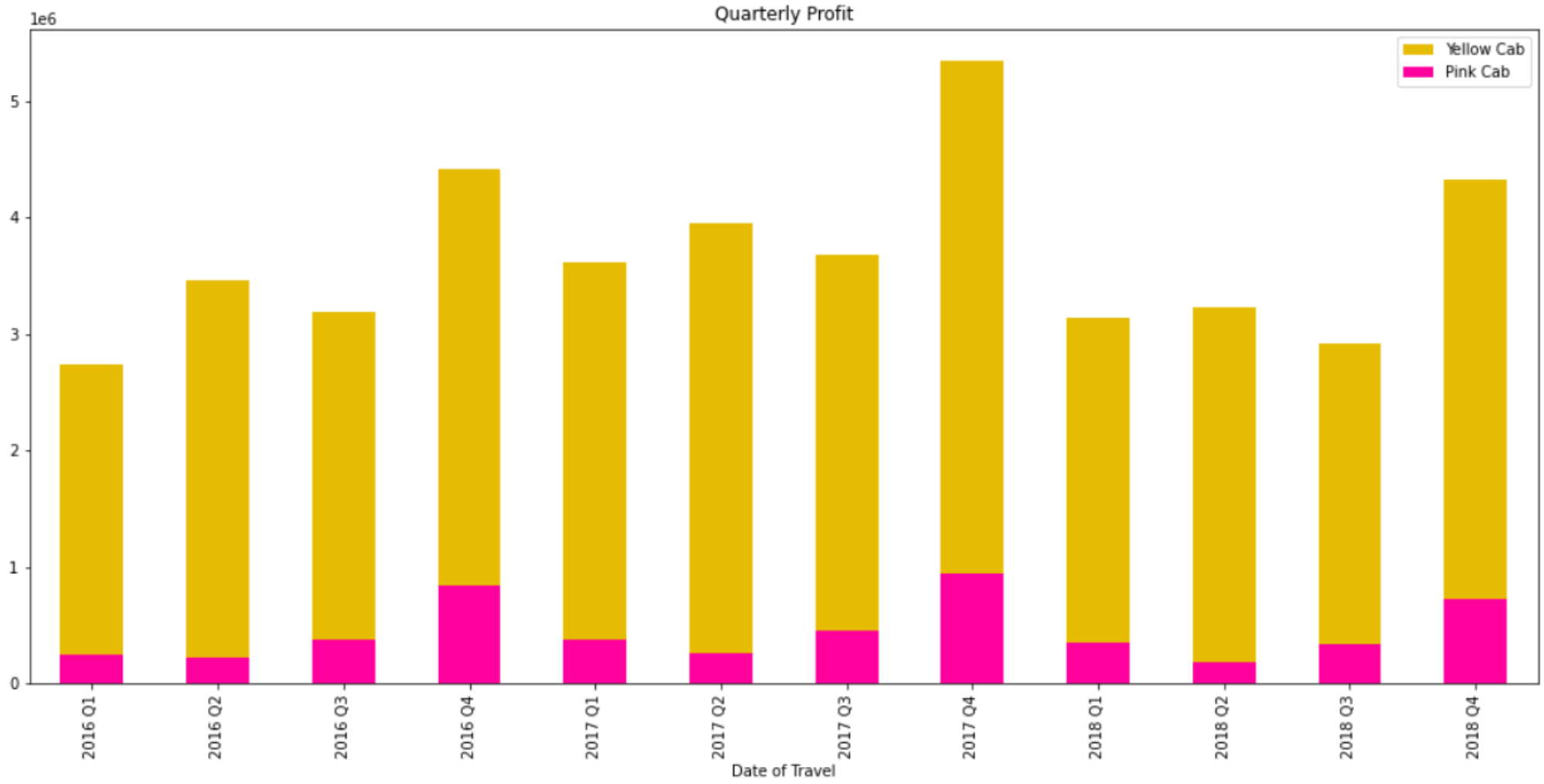
The maximum number of times Yellow Cab was hired by a single user is **47** times.

Yellow Cab maximum single user hires is **2.61** times Pink Cab maximum single user hires.



Pink Cab vs Yellow Cab Comparisons 8:

Profit by quarter:



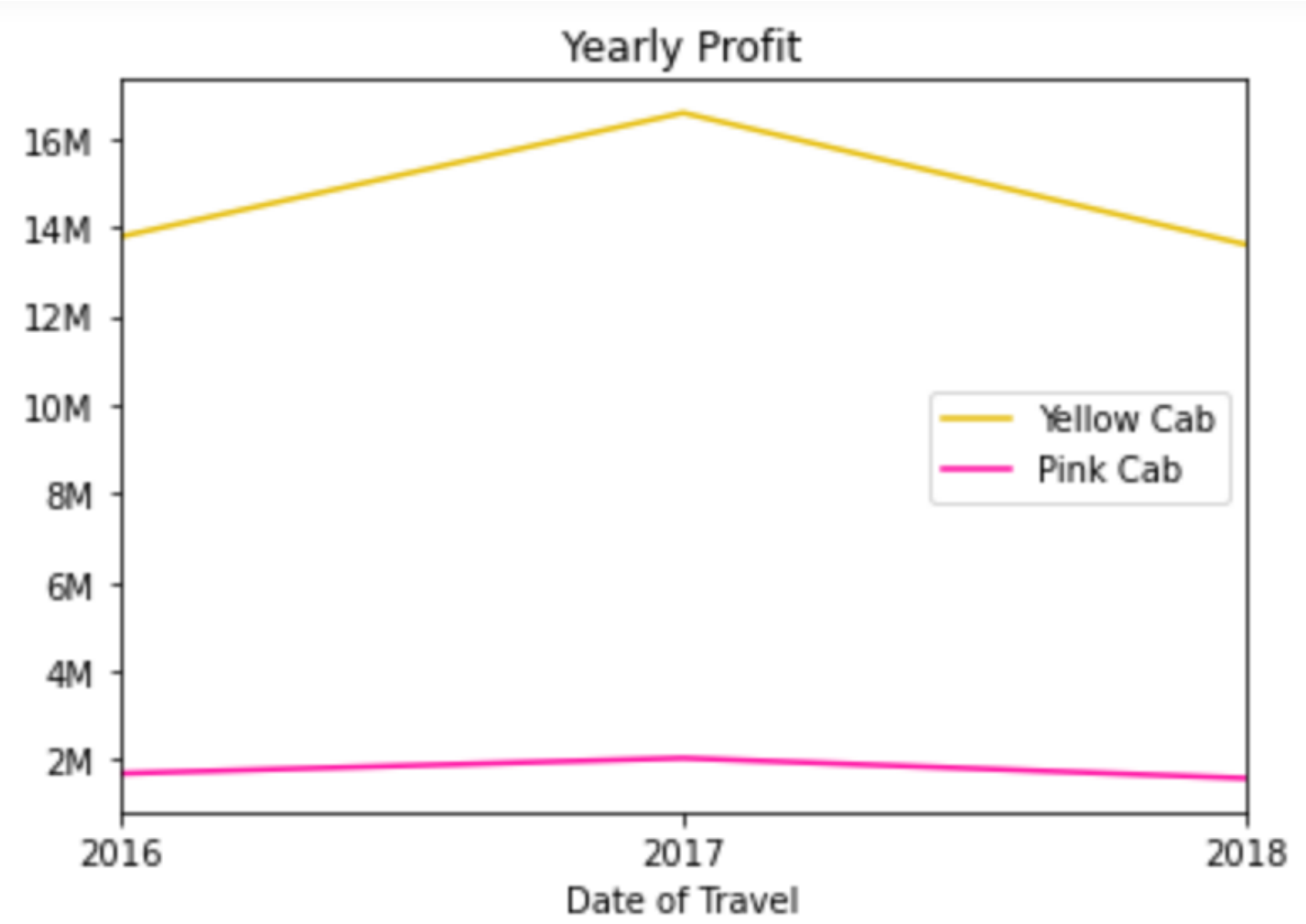
Pink Cab vs Yellow Cab Comparisons 9:

Profit by quarter 2:

- We can see that the quarterly profit of Yellow Cab is always much higher than pink cab.
- For quarter 2 of every year, Yellow Cab profits increase while Pink Cab profits decrease
- For quarter 3 of every year Yellow Cab profits decrease while Pink Cab profits increase
- The best quarter of each year for both companies' profits are Q4
- The second best quarter for Yellow cab every year is Q2, while the second best quarter for Pink Cab Company is Q3.

Pink Cab vs Yellow Cab Comparisons 10:

Yearly Profit:



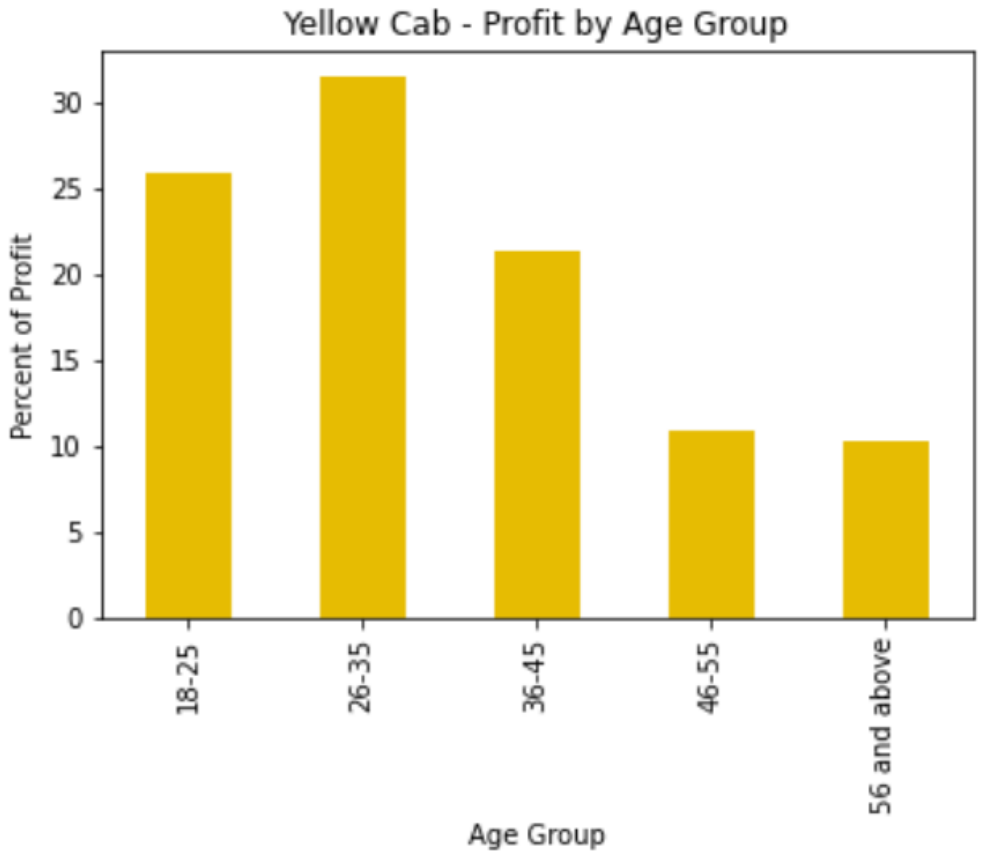
Pink Cab vs Yellow Cab Comparisons 11:

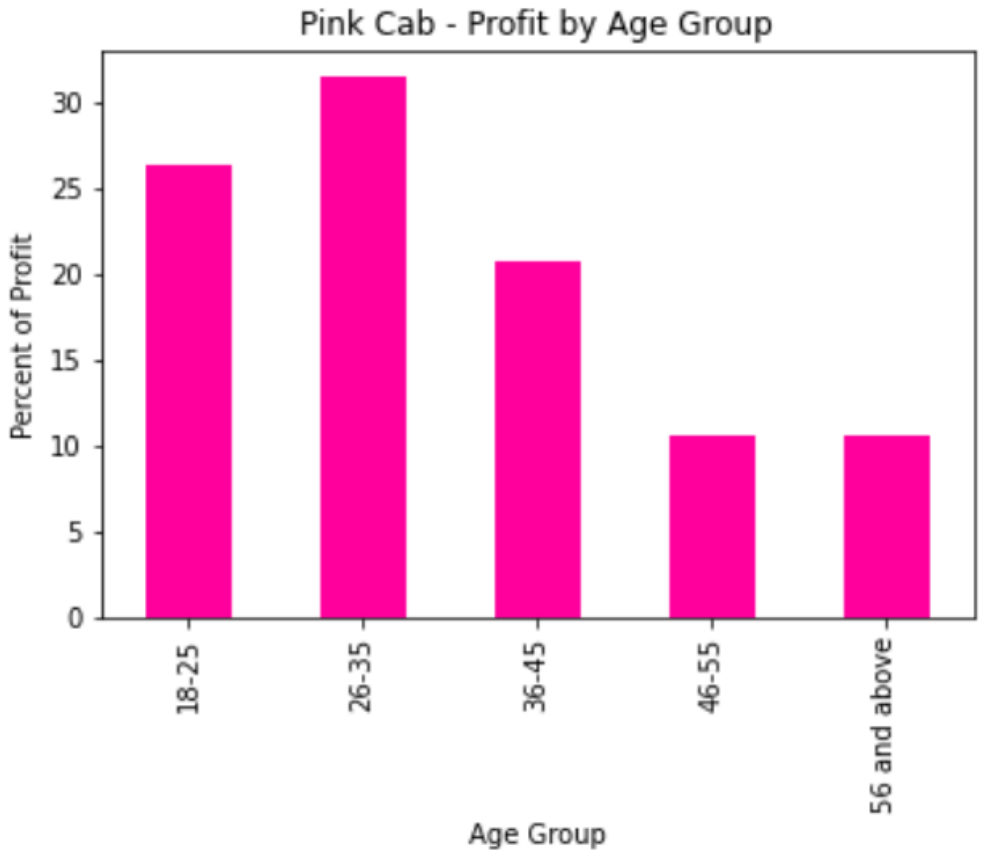
Yearly Profit 2:

- We can see that **2017** was the most profitable year for both companies.
- Pink Cab increased their profit by roughly **20%** in **2017** compared to the previous year, but their profit decreased by roughly **23%** in **2018** compared to the previous year.
- Yellow Cab also increased their profit by roughly **20%** in **2017** compared to previous year, but their profit decreased by roughly **18%** in **2018** compared to the previous year, which is a better percentage compared to Pink Cab.
- Yellow Cab yearly profits are much higher than Pink Cab yearly profits for all years.

Pink Cab vs Yellow Cab Comparisons 12:

Percentage Profit by age group:





Pink Cab vs Yellow Cab Comparisons 13:

Percentage Profit by age group 2:

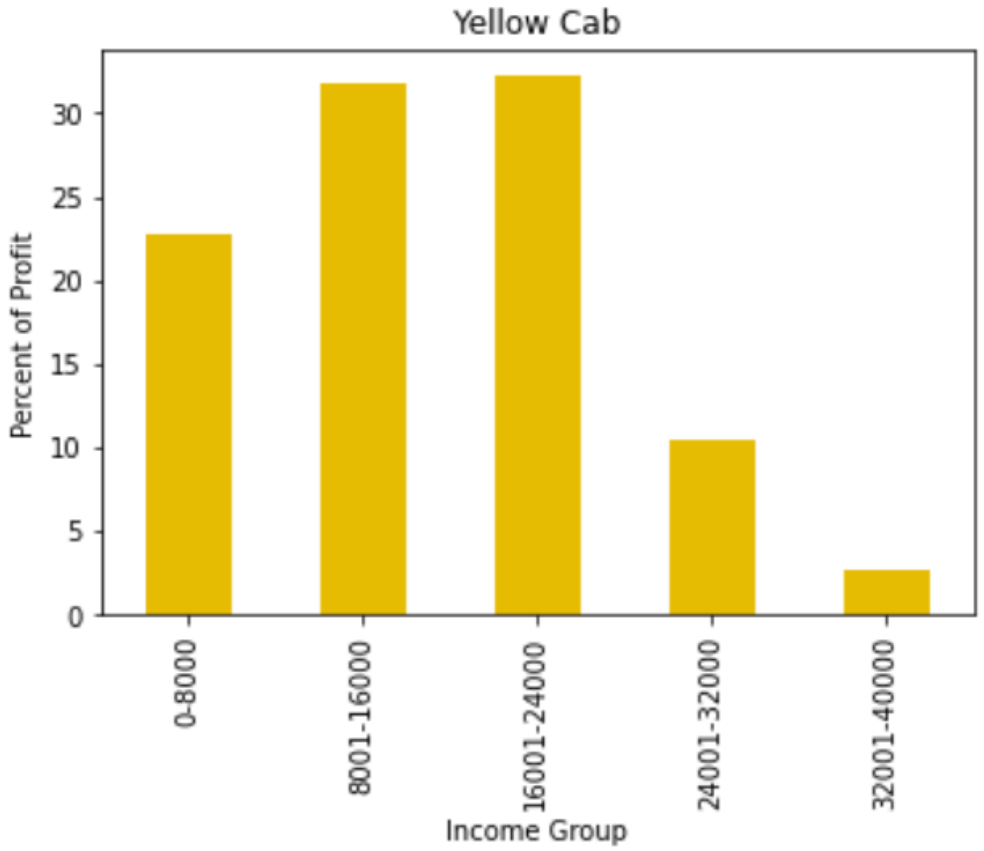
- We can see that both cab companies share a very similar distribution of their profit across their users' age groups.

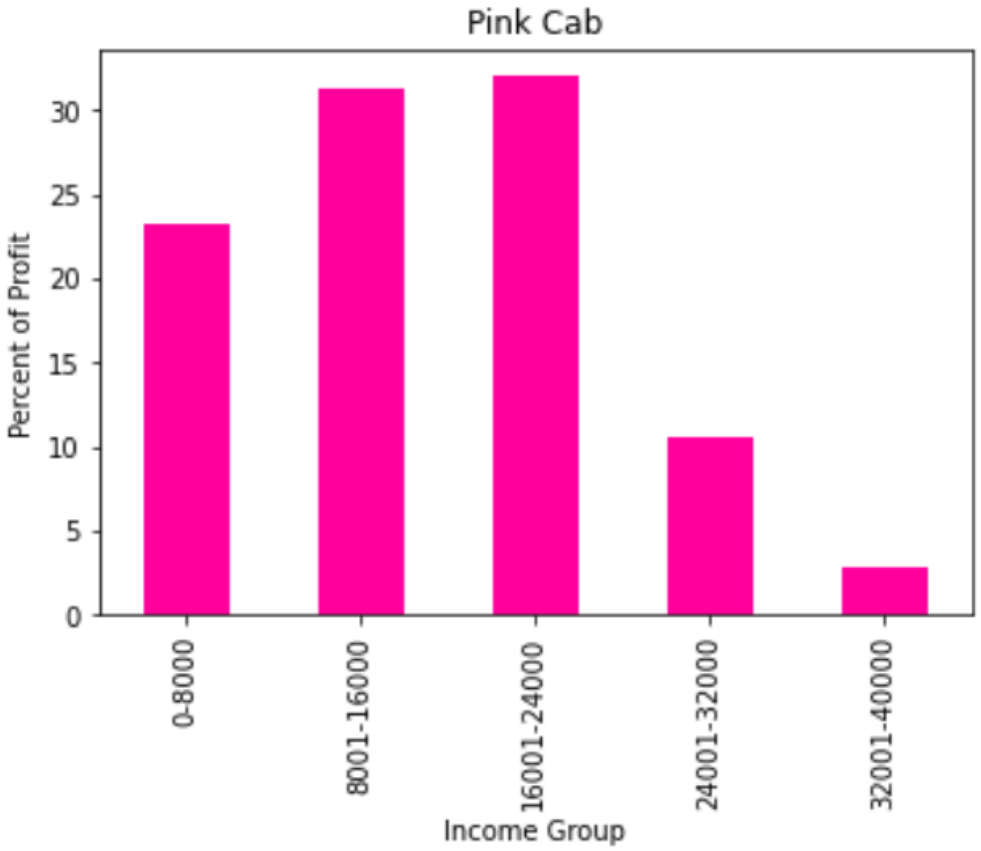
- The biggest age group contributing to both companies' profits are the **26-35** age group, followed by the **18-25** age group and then by the **36-45** age group.

- Users below **46** years old account for roughly **75%** of the profit of both cab companies.

Pink Cab vs Yellow Cab Comparisons 14:

Percentage Profit by income bracket:





Pink Cab vs Yellow Cab Comparisons 15:

Percentage Profit by income bracket 2:

- We can see from our analysis that both companies have almost identical user income segments

- User income segments with the highest contribution to profit are the **16k-24k** and then the **8k-16k** income brackets, and then the users with an income of **0-8k**, together these 3 give us around **86%** of the profit for each company.

Recommendations:

After analyzing the available data and comparing the performance of both companies, we can conclude the following:

- **Profits:** Yellow Cab yearly and quarterly profits are much higher than Pink Cab's respective profits. Yellow cab company generated **8.29** times the profit of Pink cab company over the 3 year period.
- **Average profit per km** is much greater for Yellow cab at **$7.1** vs **$2.7** for Pink cab.
- **Average profit per trip** is much higher for yellow cab at **$160** vs **$62** for pink cab.
- **Returning customer:** On average, each user of Yellow Cab used their services **2.63** times more than Pink Cab users used Pink Cab services.
- **Customer Reach:** Pink cab has served approximately **70%** of the total users included in our data set while Yellow cab has served approximately **86%** of the total users included in our data set.
- **Total Trips and Total Distance:** Yellow Cab Company has more than **3** times total trips and covered roughly **3** times the total distance compared to Pink Cab.

We can conclude based on our analysis results above that Yellow Cab would be a much better target for investment by XYZ company.

Thank You