Exploratory Data Analysis

G2M insight for Cab Investment firm

19.07.2022

Agenda

- Executive Summary
- Problem Statement
- Cab industry in US overview
- Description of data
- EDA
- Hypothesis
- Hypothesis results
- Recommendations

Executive Summary - the United States has seen significant growth in the cab industry and the emergence of key players in this market. A private firm in the US wants to understand the market and choose the right company to invest in.

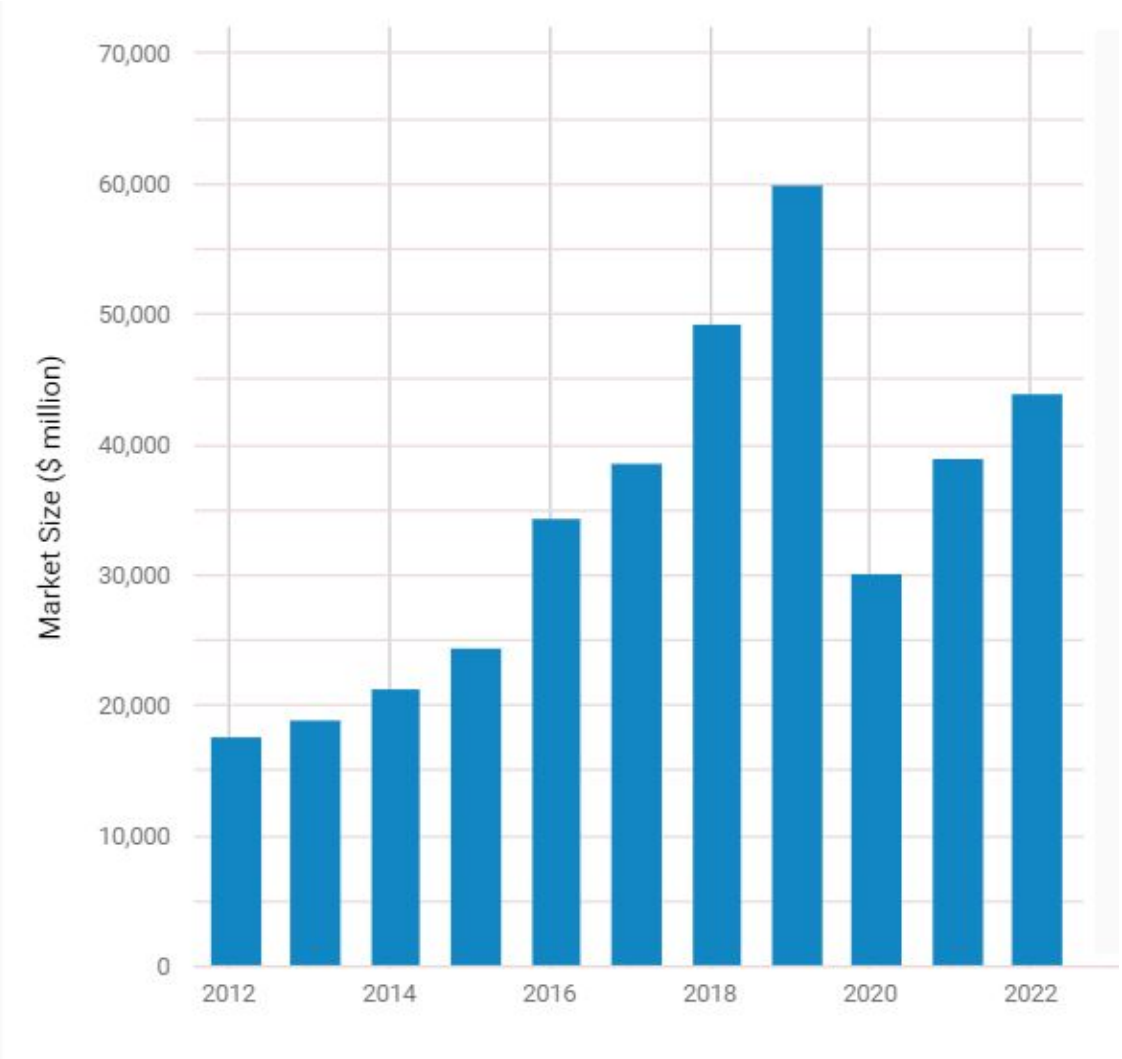
Problem Statement - to analyze companies from the cab industry in the USA and give recommendations on the choice of a company for investment.

Repository: https://github.com/kkalyagina/Data-Glacier.git

Cab industry in US overview

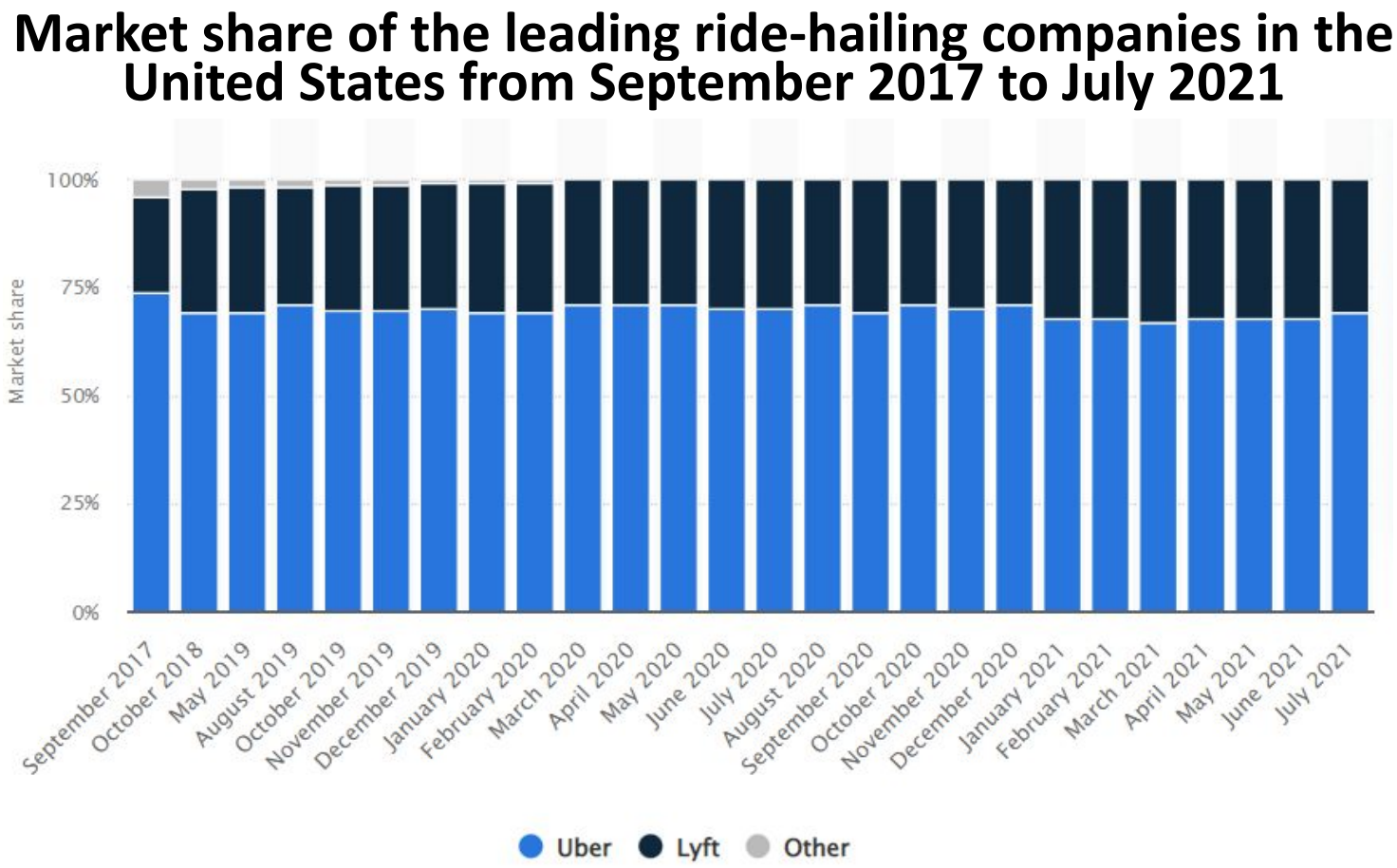
- Market Size: $44bn
- Number of Businesses: 1,453,304
- Industry Employment: 1,455,492
- Taxi Services in the US Annualized Market Size Growth 2017–2022: 2.6%

Demand for industry services has largely flourished over the past five years, as growth in the US economy drove corporate expense accounts higher and prompted everyday consumers to spend more on small conveniences and luxuries, such as taxi.



Data Glacier
Your Deep Learning Partner

Market share of the leading companies in the US

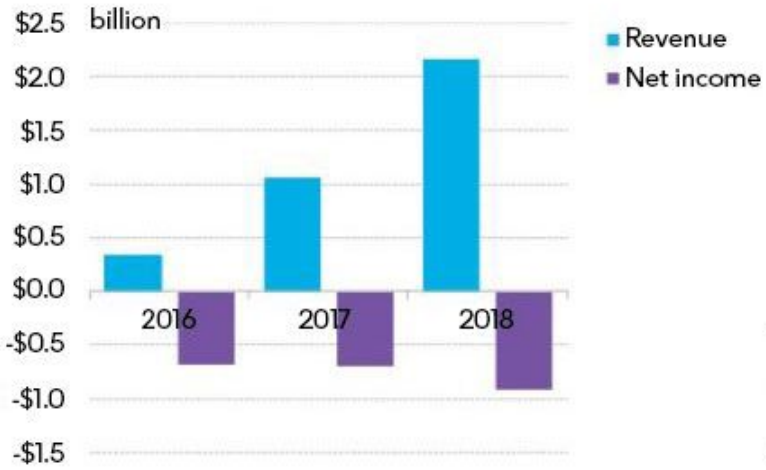


Despite being almost synonymous with the ridesharing industry, Uber's share of the U.S. market has been declining, falling from 74 percent in September 2017 to 69 percent in July 2021. This loss has been to the benefit of Lyft, whose market share stood at 31 percent in that same month.

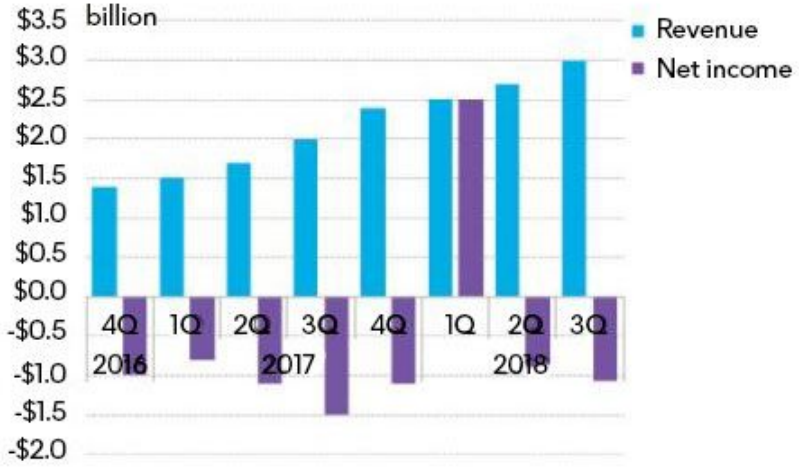
While Uber's U.S. market share may be declining, the company is still growing strongly in terms of revenue and ridership. There are two reasons for this. First, Uber is a global company, whereas Lyft only operates in the North American market. And secondly, the overall size of the ride sharing market in North America is growing, with 36 percent of U.S. adults using ride sharing services in 2018 compared to 15 percent in 2015.

Financial performance

Lyft and Uber revenue and net income



Source: Lyft prospectus



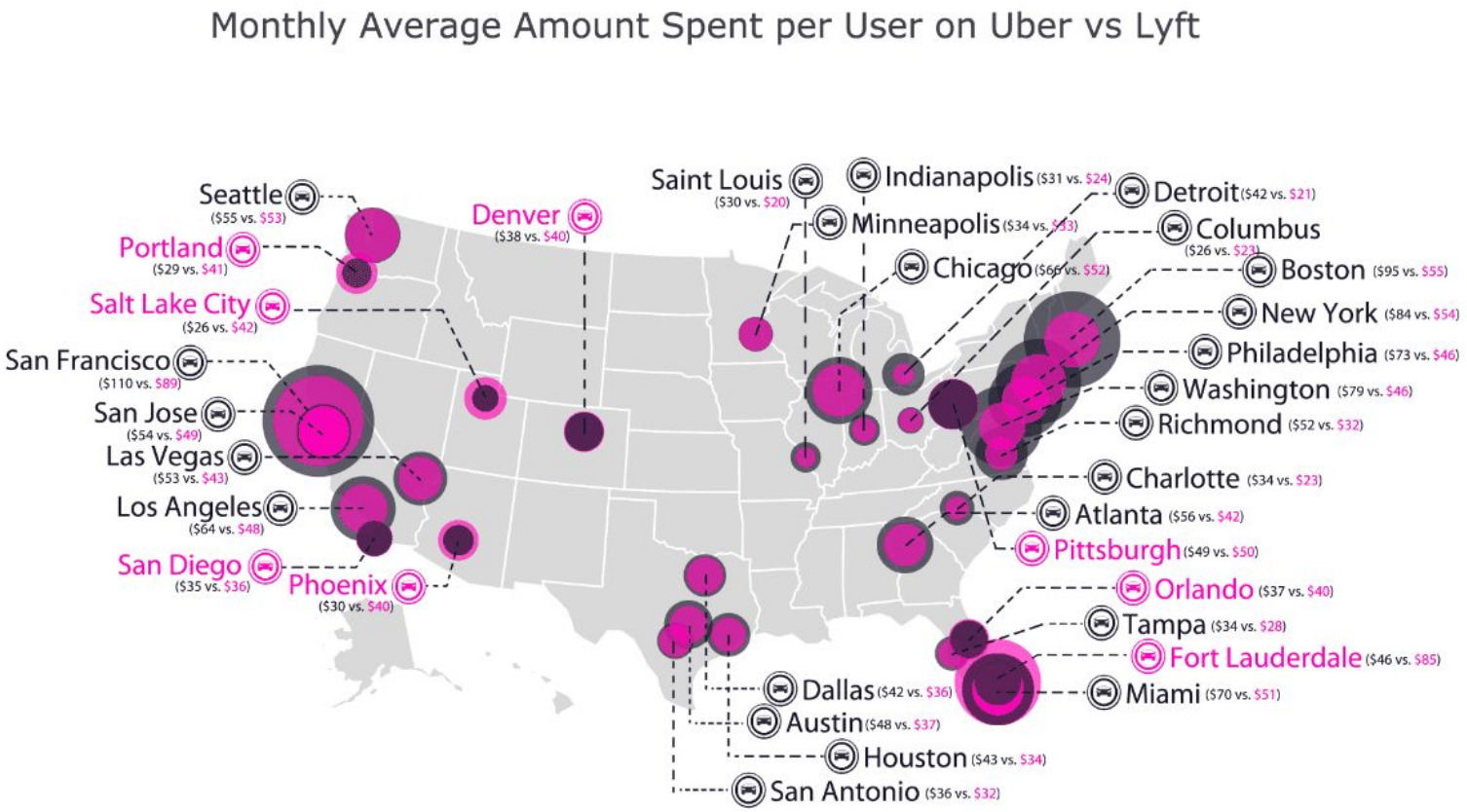
Source: Uber. Note: Figures are preliminary and subject to revision by Uber.

Lyft's prospectus shows revenue reached $2.2 billion in 2018 — 103 percent higher than 2017 — and its net loss totaled $911 million. While the net loss in 2018 was 32 percent wider than in 2017, revenue grew three times faster than the increase in losses, suggesting Lyft is making headway in managing its bottom line.

For the first three quarters of 2018, Uber reported $8.2 billion of revenue and net income of $539 million. Uber's year-on-year revenue growth rate for the first three quarter of 2018 was 58 percent, below that of Lyft's 103 percent for the full year.

While Uber's revenue growth rate appears to be lower than that of Lyft, it may have had more success in controlling its bottom line. Uber's net loss of $1.07 billion in the third quarter of 2018 is 36 percent of revenue in that quarter, while Lyft's net loss of $911 million in 2018 accounts for 42 percent of revenue in 2018.

Cab industry in the cities of the US

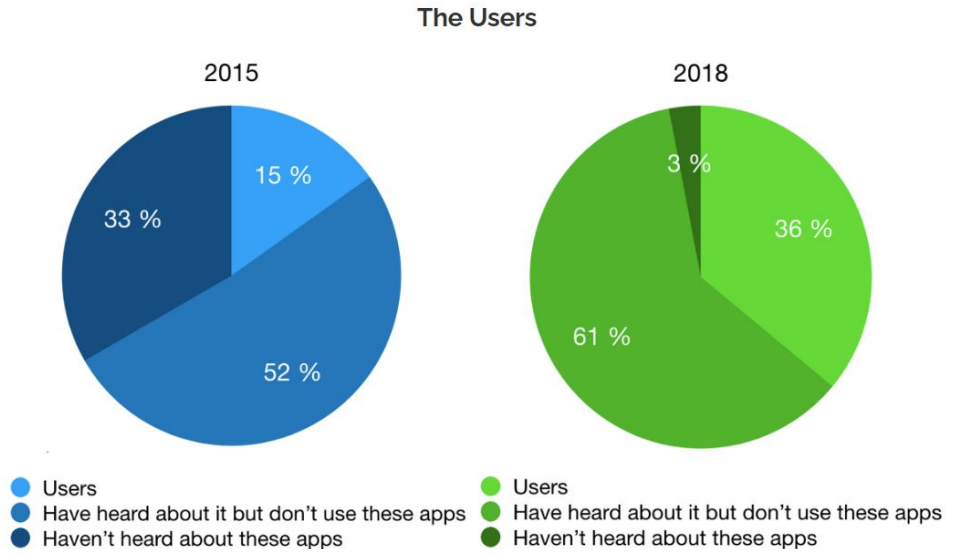


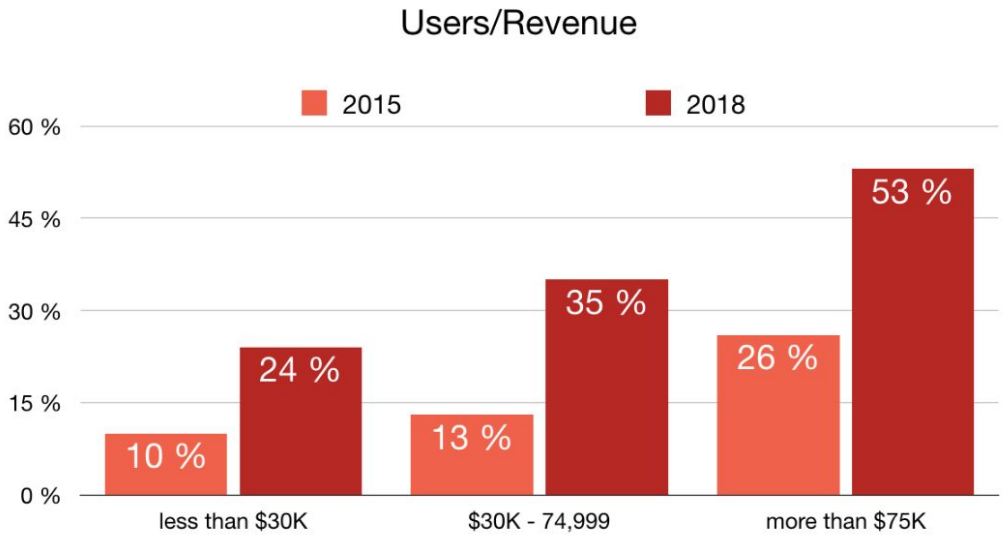
These are some of the cities with the highest monthly spending per user through the Uber app:
San Francisco, CA – $110
Boston, MA – $95
New York, NY – $84
Washington, DC – $79
Philadelphia, PA – $73
Miami, FL – $70
Chicago, IL – $66
Los Angeles, CA – $64

It's also worth noting a couple of cities where Lyft has a high average monthly spending per user. Lyft users in San Francisco $89 per month, and spend $85 per month in Fort Lauderdale, FL.

But for the most part, Uber users spend more each month compared to Lyft riders. Lyft users spent more than Uber users in just 8 of the 32 cities in this study. Those cities are: Salt Lake City (UT), Phoenix (AZ), Portland (OR), Orlando (FL), Denver (CO), Pittsburgh (PA), Fort Lauderdale (FL), San Diego (CA).
In some instances, the difference was marginal, as low as $1 in Pittsburgh and San Diego.

Who are the customers of ride-hailing applications?

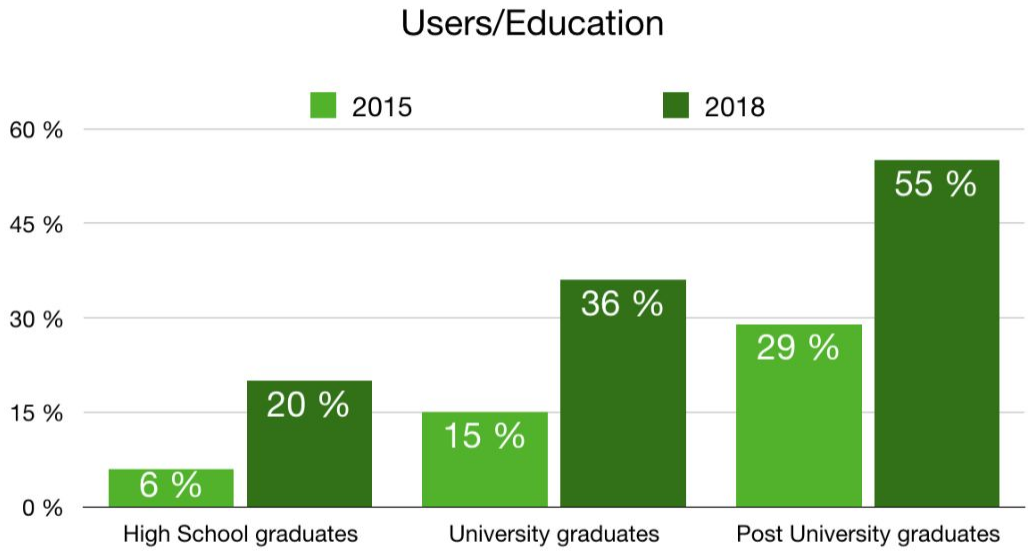




In 2015, only 15% of Americans said that they used applications such as those provided by Uber or Lyft to commute, compared to 36% in 2018.
The percentage of users has more than doubled in four years, and only 3% are unaware of the existence of these applications in 2018.

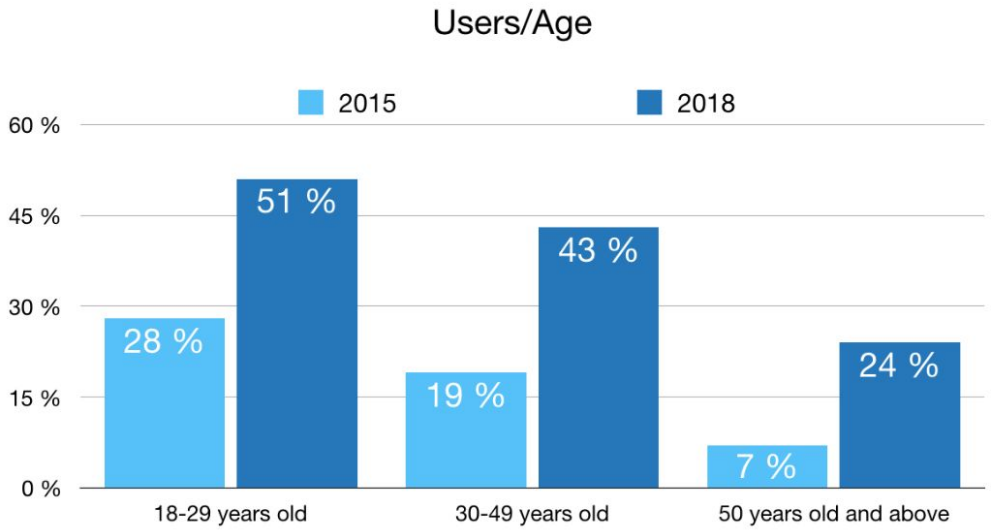
More than 70% of people with a yearly salary of $75,000 or more live in cities. Plus, they are also the most likely to choose to use this mode of transport rather than other potentially slower or more congested modes such as public transport or cycling, for example.

Who are the customers of ride-hailing applications?



In 2018, less than a quarter of American high school graduates use Uber or Lyft while more than a third of university graduates and more than half of post-graduates are using these applications. The more educated people tend to use these applications more often. This is also because many of them live in urban areas where access to this type of service is straightforward and fast, unlike in rural areas where there are very few Uber or Lyft drivers.

The figures have changed significantly, and the user rate has doubled for almost all the categories mentioned above. All age groups are affected by this need to travel



Description of Data

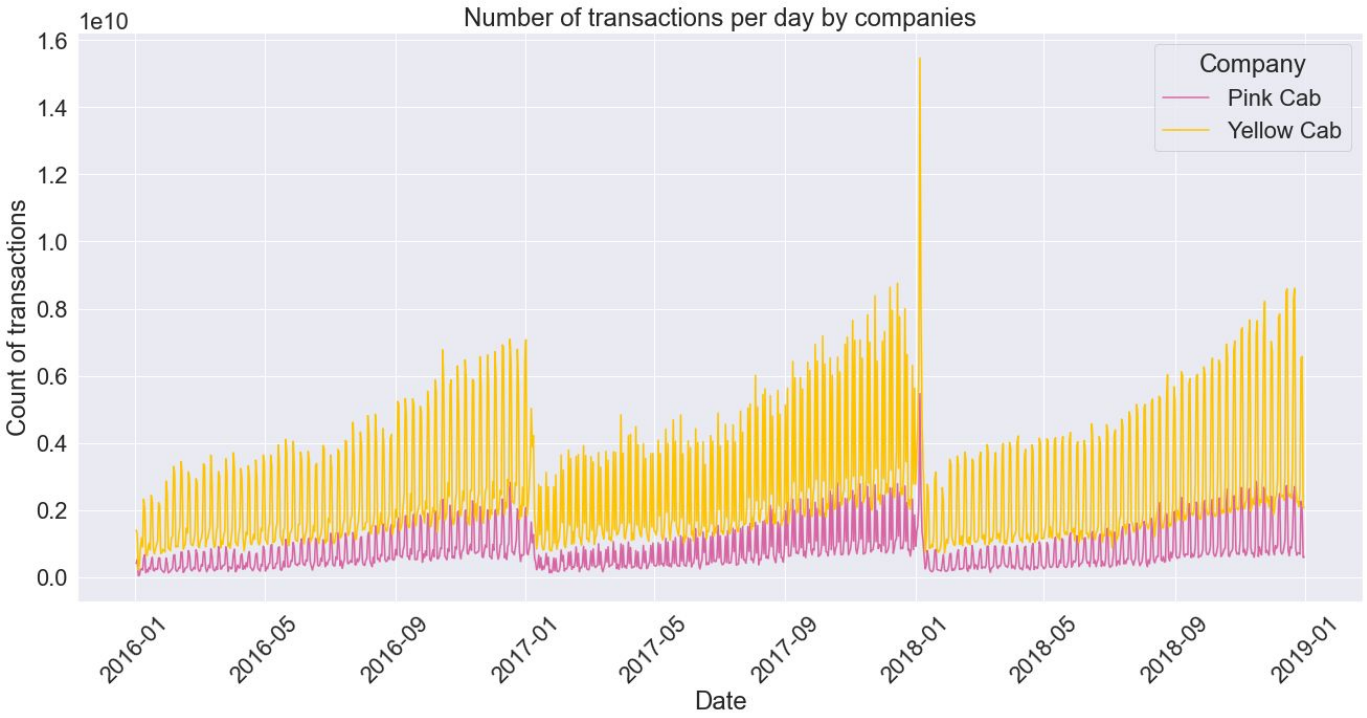
	cab_data
19.2+ MB	**Information about companies' transactions**
int64	Transaction ID
datetime64	Date of Travel
object	Company
object	City
float64	KM Travelled
float64	Price Charged
float64	Cost of Trip

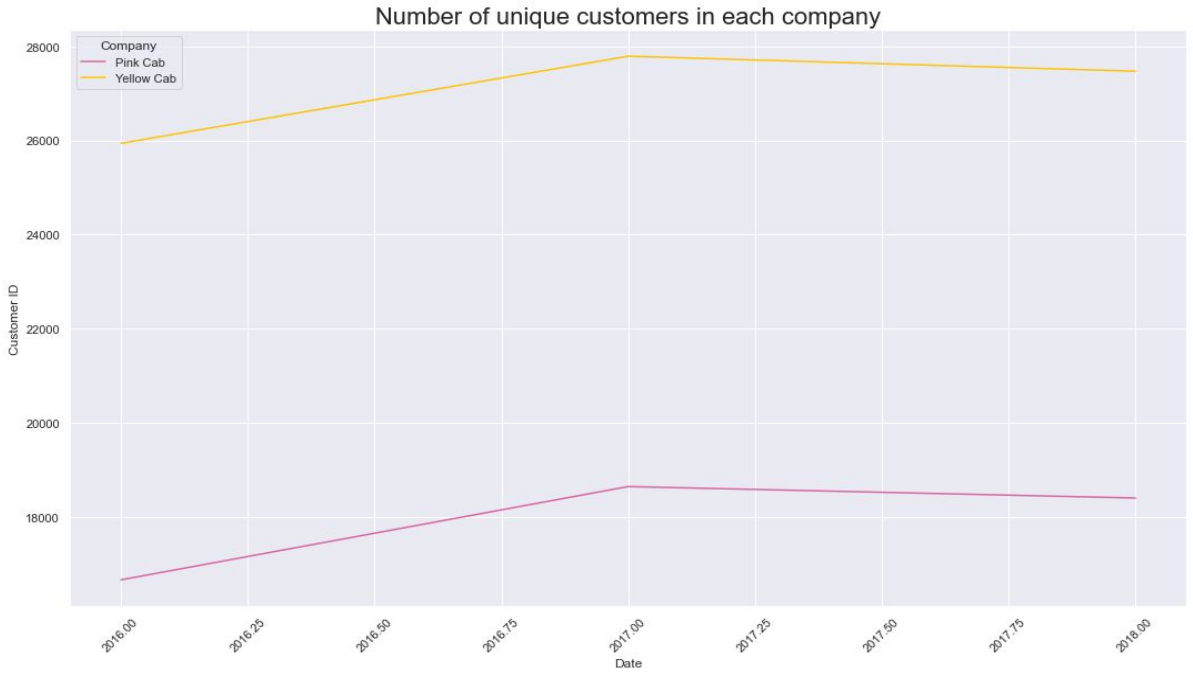
	transaction_id
10.1+ MB	**Information about the method of payment for transactions**
int64	Transaction ID
int64	Customer ID
object	Payment_Mode

	city
448.0+ bytes	**Information about cities**
object	City
int64	Population
int64	Users

	customer_id
1.5+ MB	**Information about customers**
int64	Customer ID
object	Gender
int64	Age
int64	Income (USD/Month)

EDA. Number of transactions and unique users by companies.



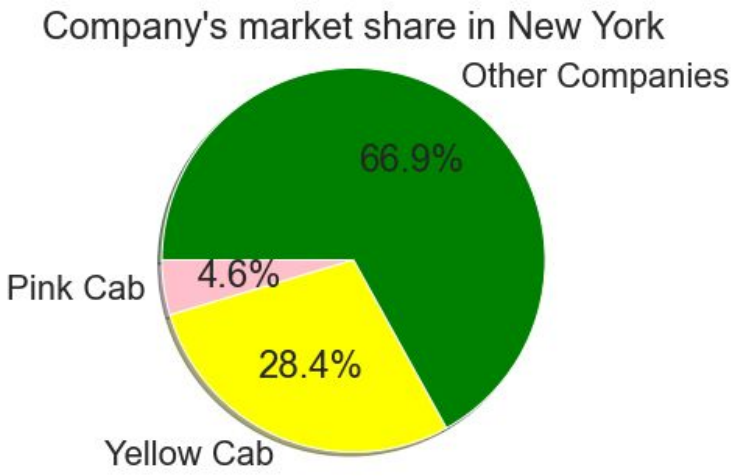


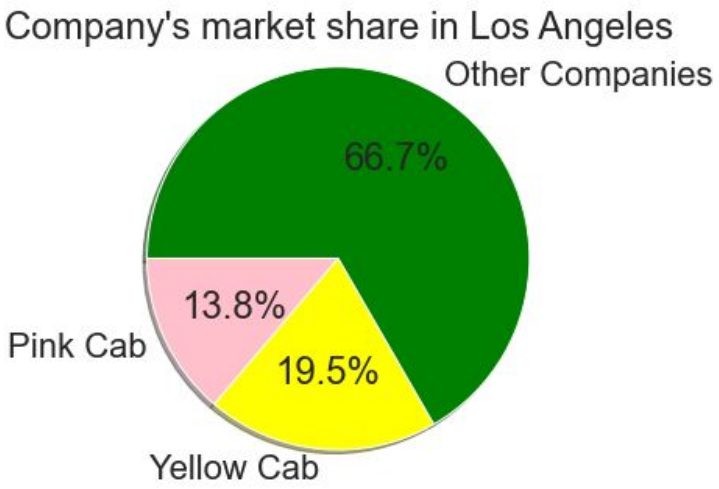
- **Each company's number of transactions increases seasonally**
- **It can be seen that the Yellow Cab company has a lot more transactions compared to the Pink Cab**

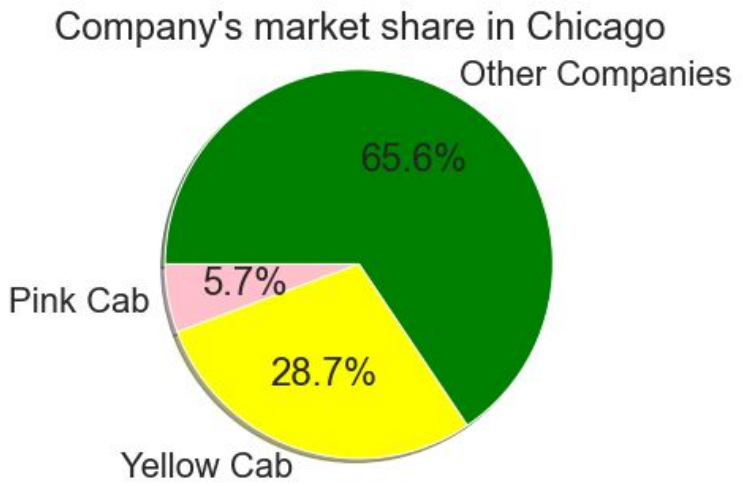
- **The number of unique users increased by 2017 and decreased by 2018 for both companies**

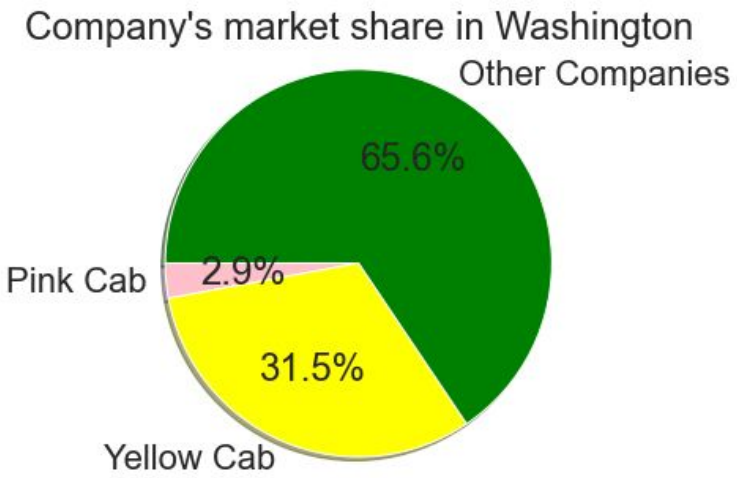
Data Glacier
Your Deep Learning Partner

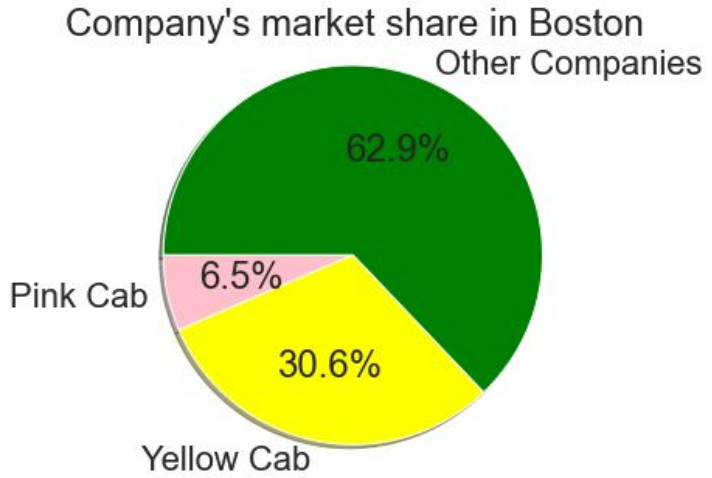
EDA. Company's market share in the cities with the largest number of cab users.

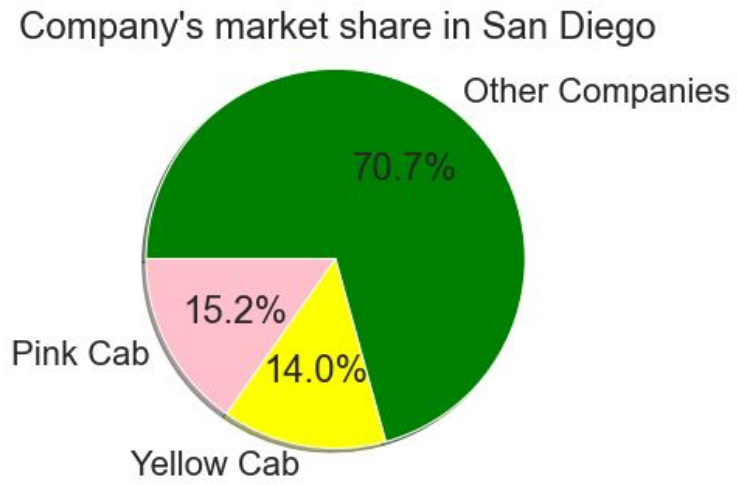








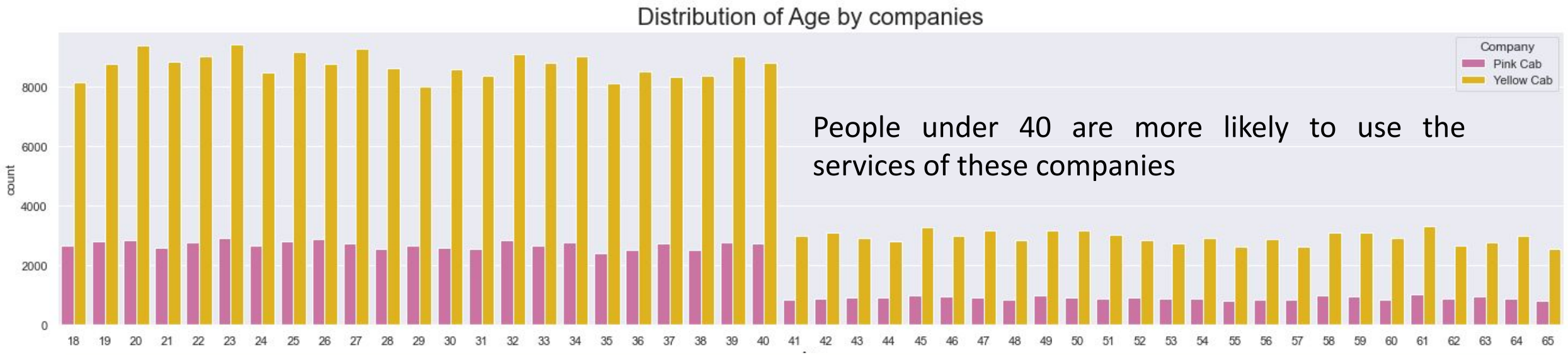




- **~30% of all users are occupied by Yellow Cab, with the exception of San Diego, where Pink Cab outperformed Yellow cab by 1%**
- **In total, in all such cities, both companies together occupy about 30-35% of all users**

Data Glacier
Your Deep Learning Partner

EDA. Age and Gender.



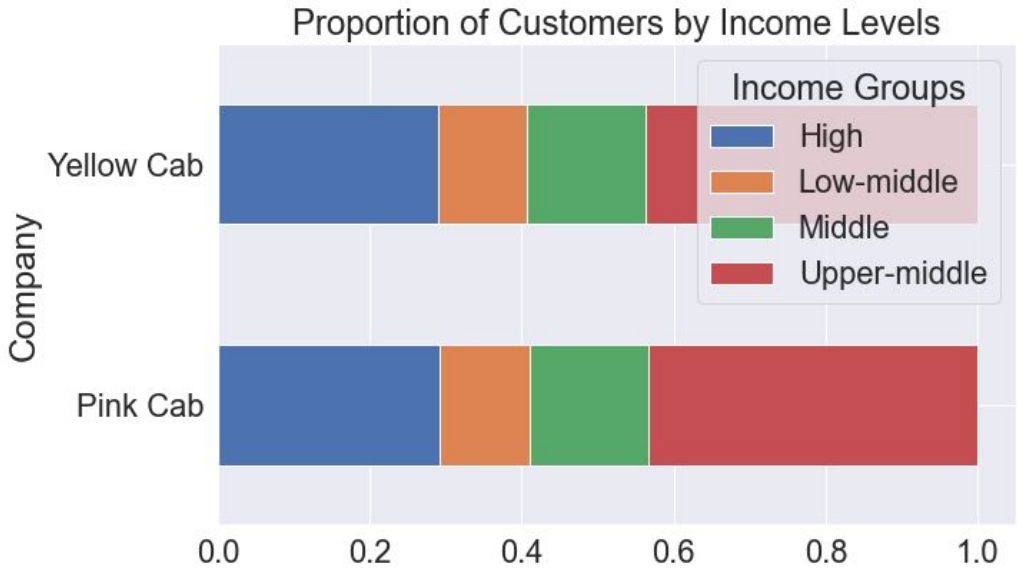
People under 40 are more likely to use the services of these companies

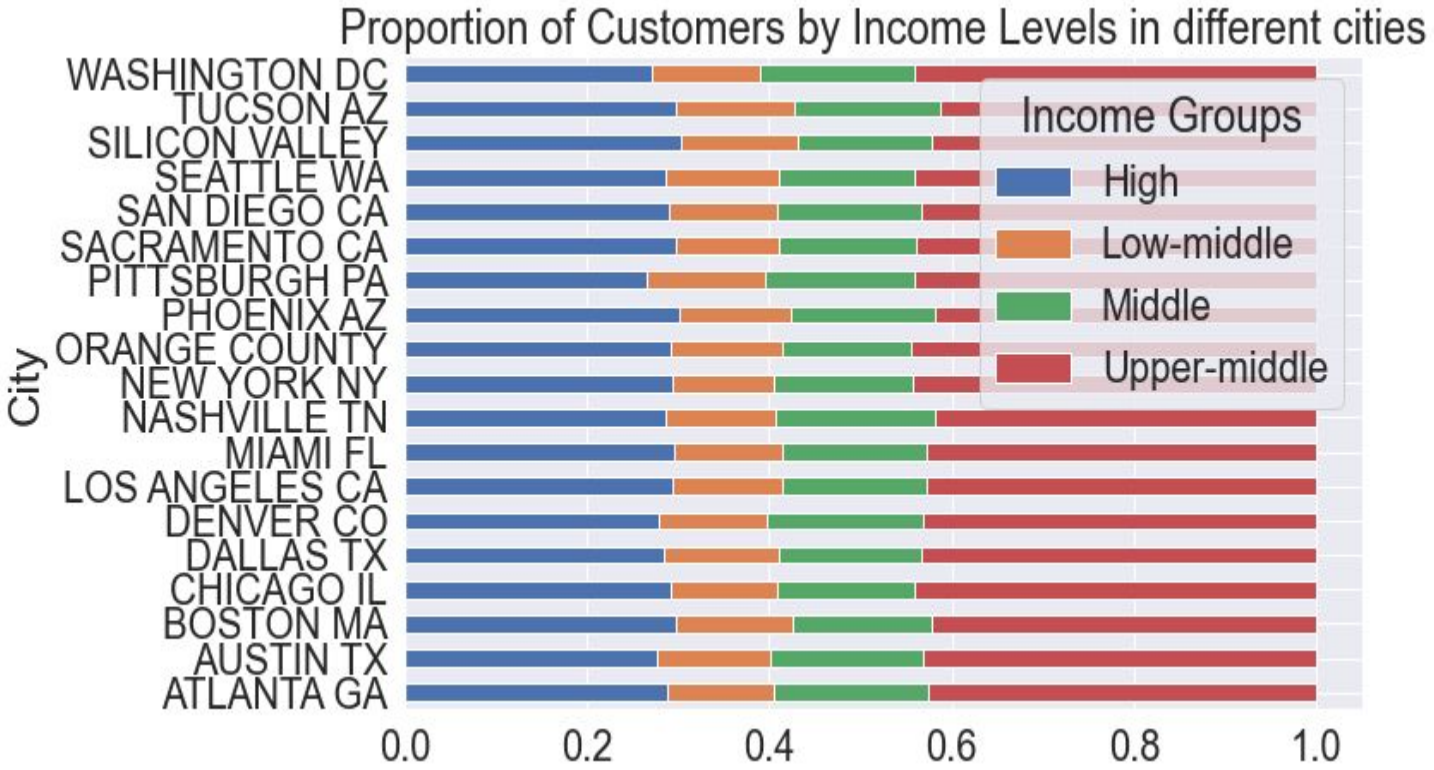
Men use cabs more often, both in general terms and in individual companies

Data Glacier
Your Deep Learning Partner

EDA. Income Level.

'Low-middle' - income <5000 USD/Month
'Middle' - from 5000 to 9000 USD/Month
'Upper-middle' - from 9000 to 20000 USD/Month
'High' - more than 20000 USD/Month





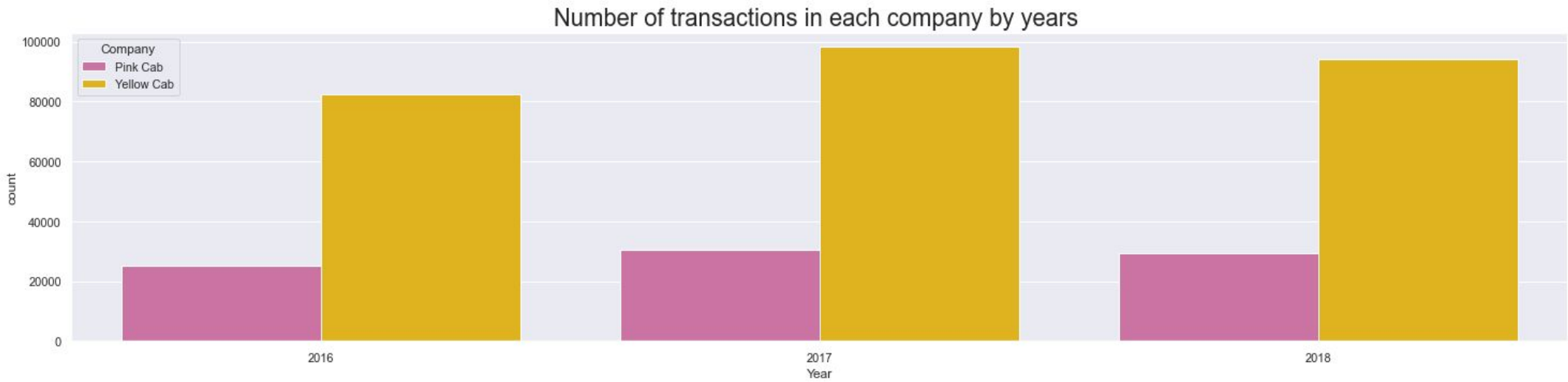
- **The proportions of customer income in the companies are the same**
- **All cities are dominated by high- and upper-middle income customers**

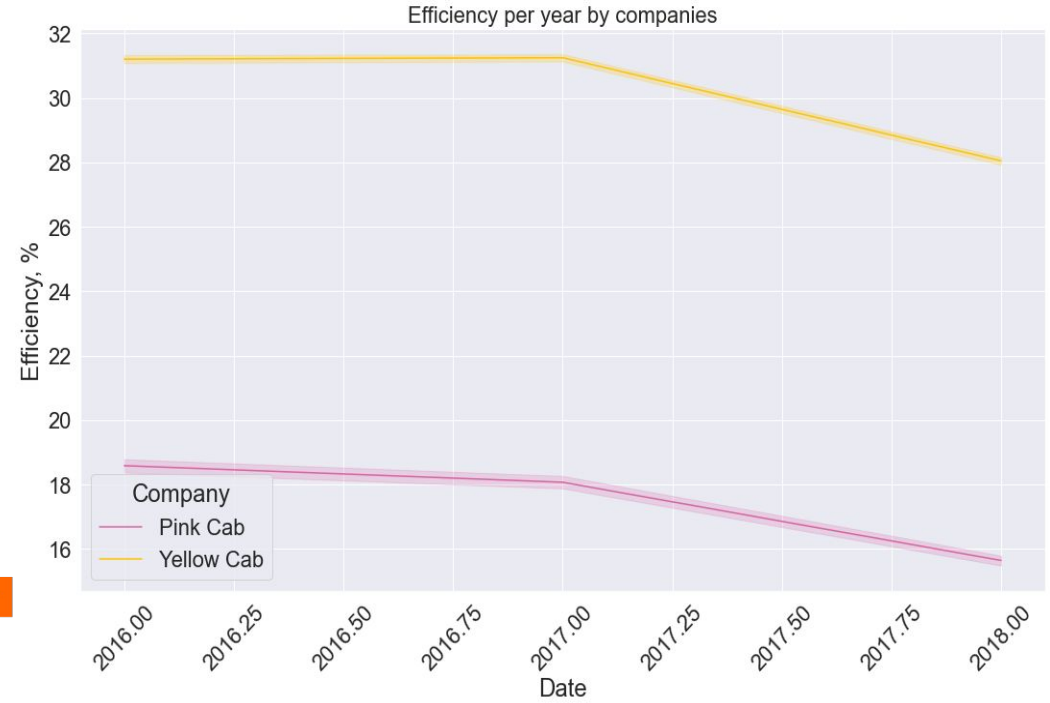
Data Glacier
Your Deep Learning Partner

Hypothesis

1. The more transactions a company has, the higher its efficiency.
2. The more unique customers a company has, the higher its efficiency.
3. The larger the company's revenue, the higher its efficiency.
4. The more transactions a company has in a particular city, the higher its efficiency in that city relative to another company.
5. The larger the population of a city, the more trips there are.
6. The larger the population of the city, the higher the efficiency.
7. The higher the income of the company's customers, the higher its efficiency.
8. When paying for a trip with a bank card, the efficiency of the company is lower than when paying in cash.
9. Loyal customers lead to greater efficiency.
10. The more profit from one trip, the higher the efficiency.

Hypothesis No.1: The more transactions a company has, the higher its efficiency





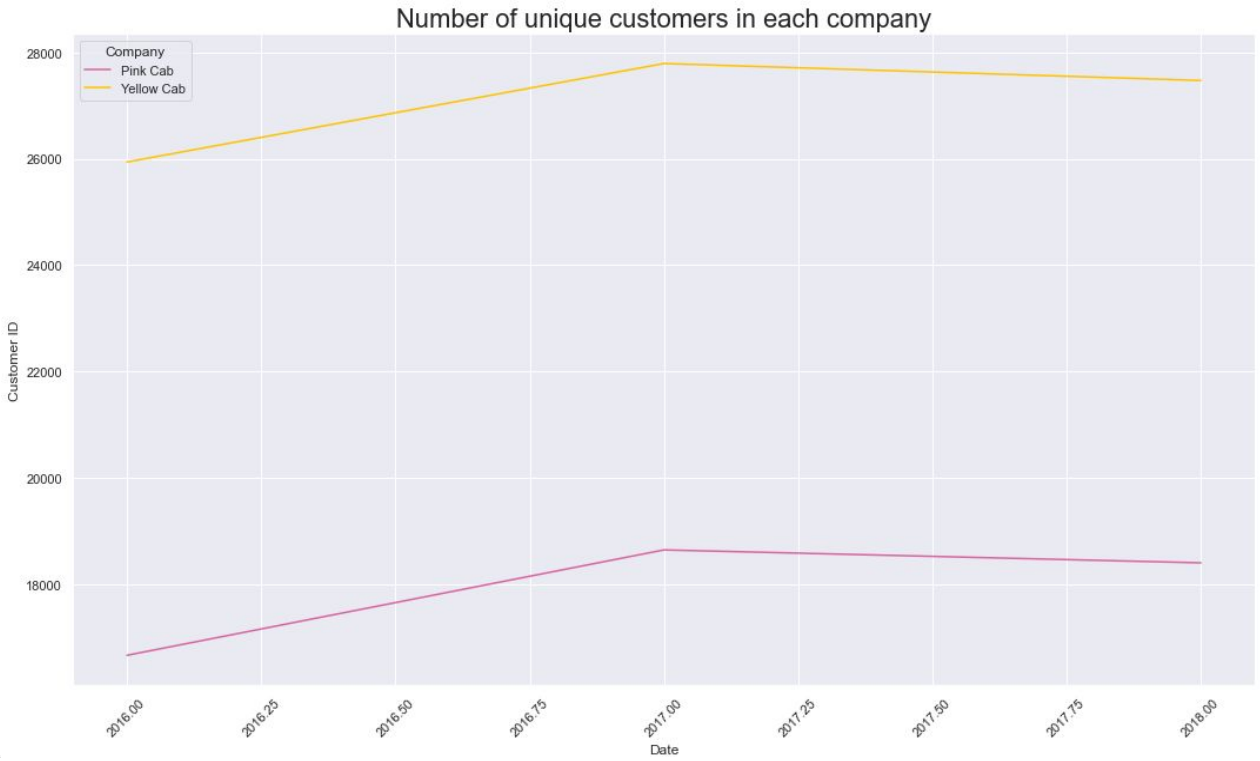
The graphs show that the Yellow Cab company has several times more transactions, and its efficiency is higher than the Pink Cab.

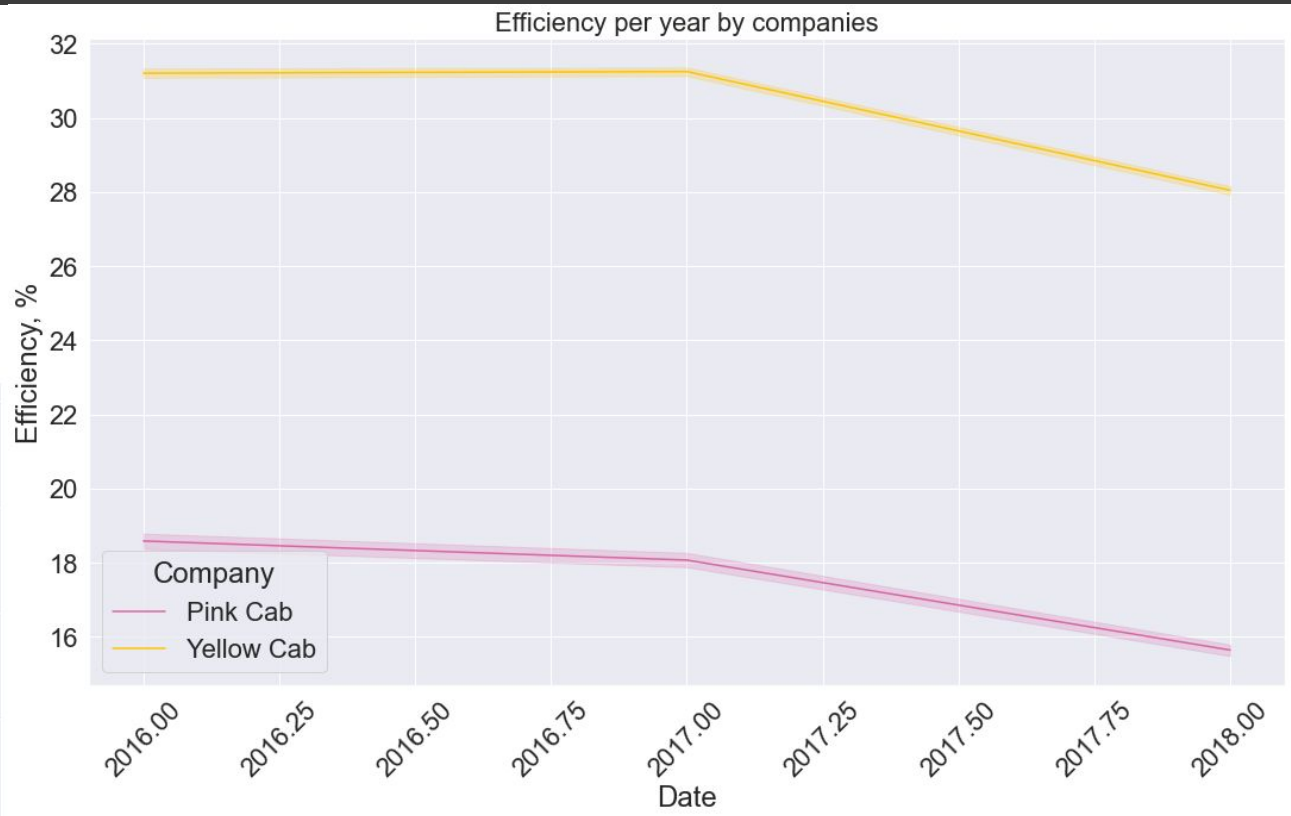
The hypothesis is confirmed

Data Glacier
Your Deep Learning Partner

Hypothesis No.2: The more unique customers a company has, the higher its efficiency.

Yellow Cab has more unique customers and is higher efficiency.

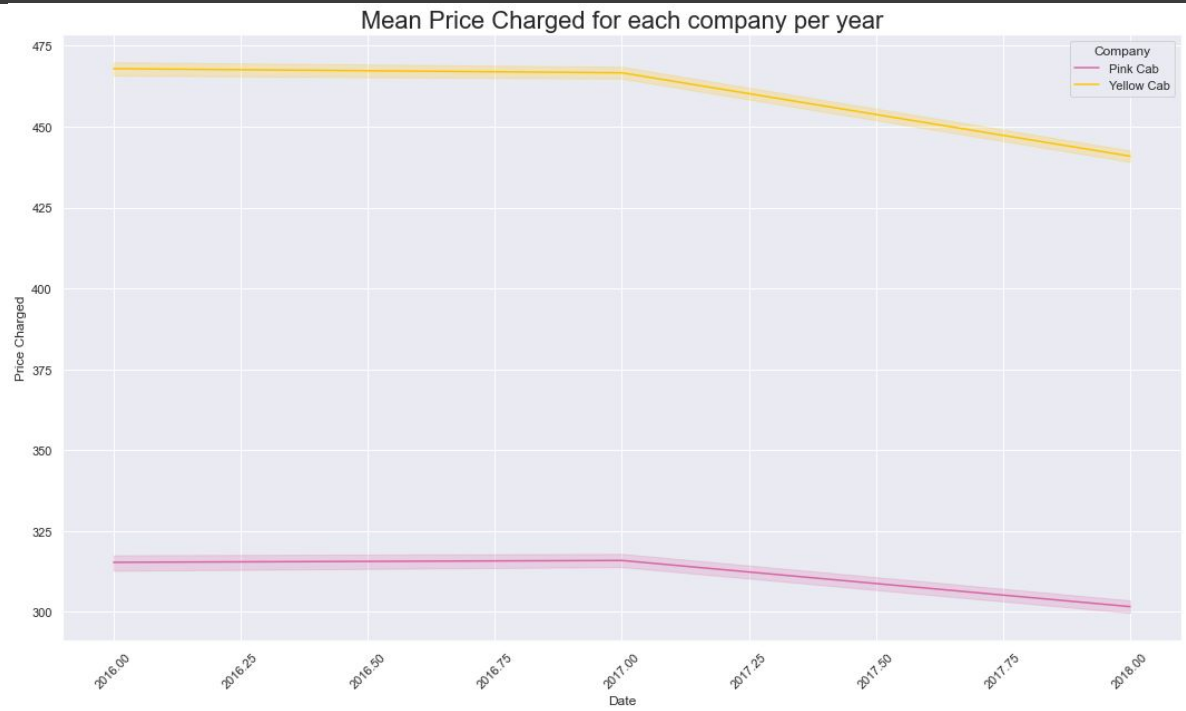


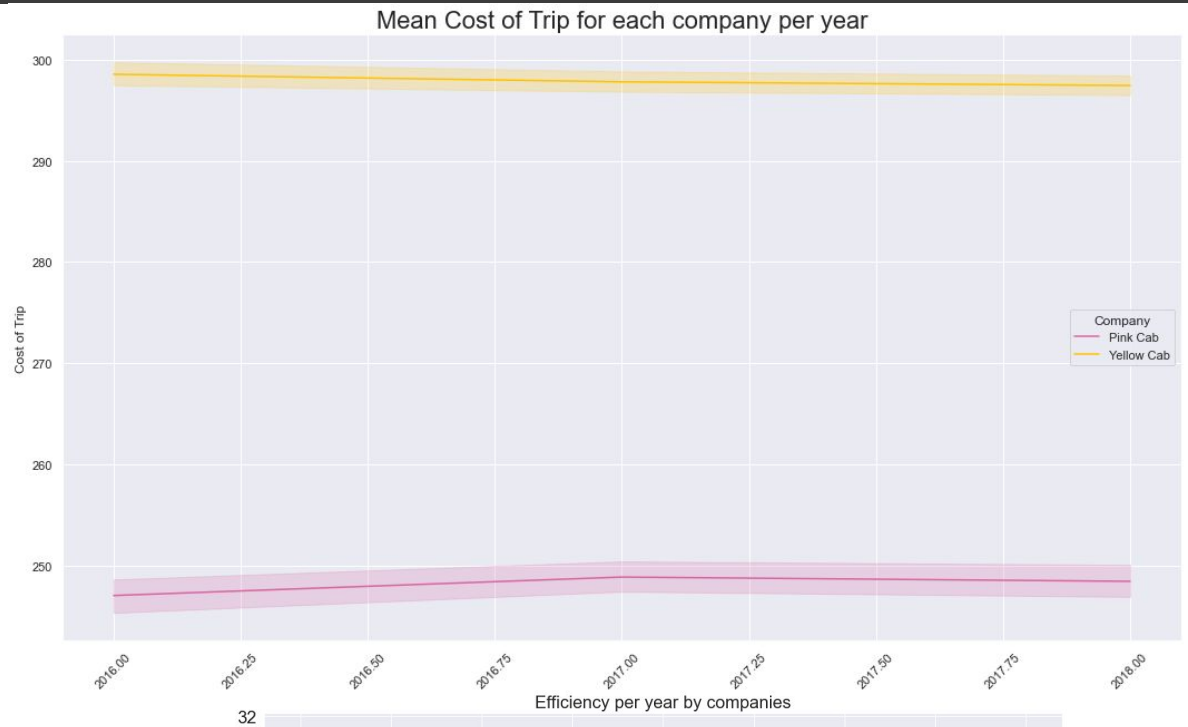


The hypothesis is confirmed

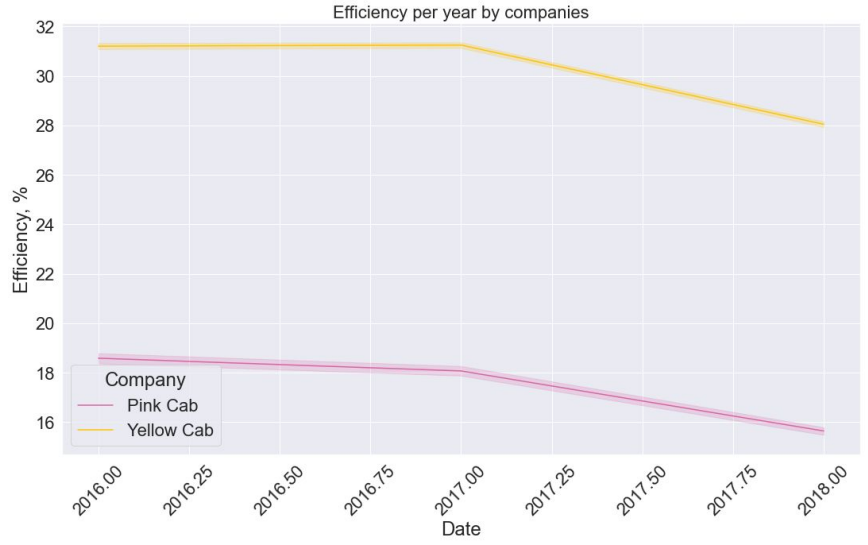
Data Glacier
Your Deep Learning Partner

Hypothesis No.3: The larger the company's revenue, the higher its efficiency





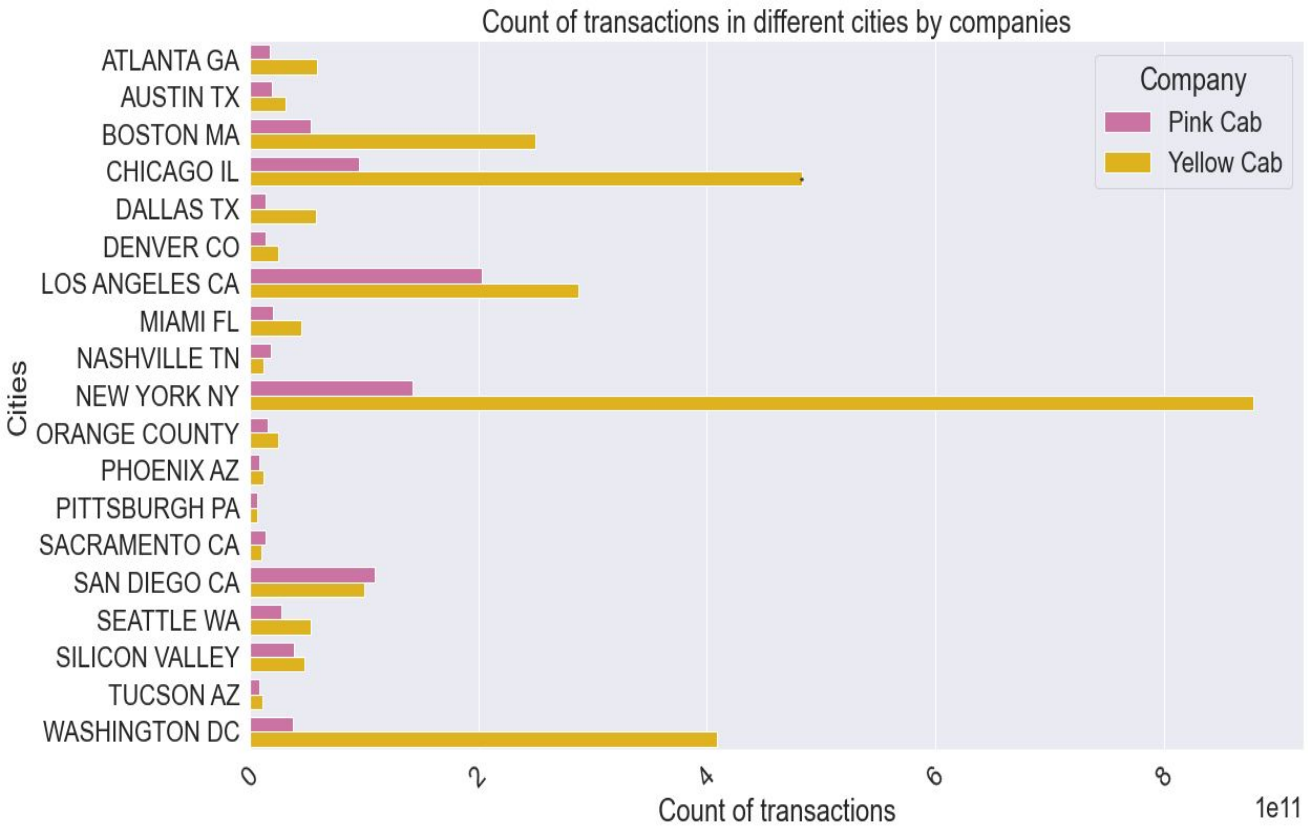
The components of efficiency are the company's income and travel expenses. Income and efficiency behave in the same way, since the cost curve for companies is less volatile.

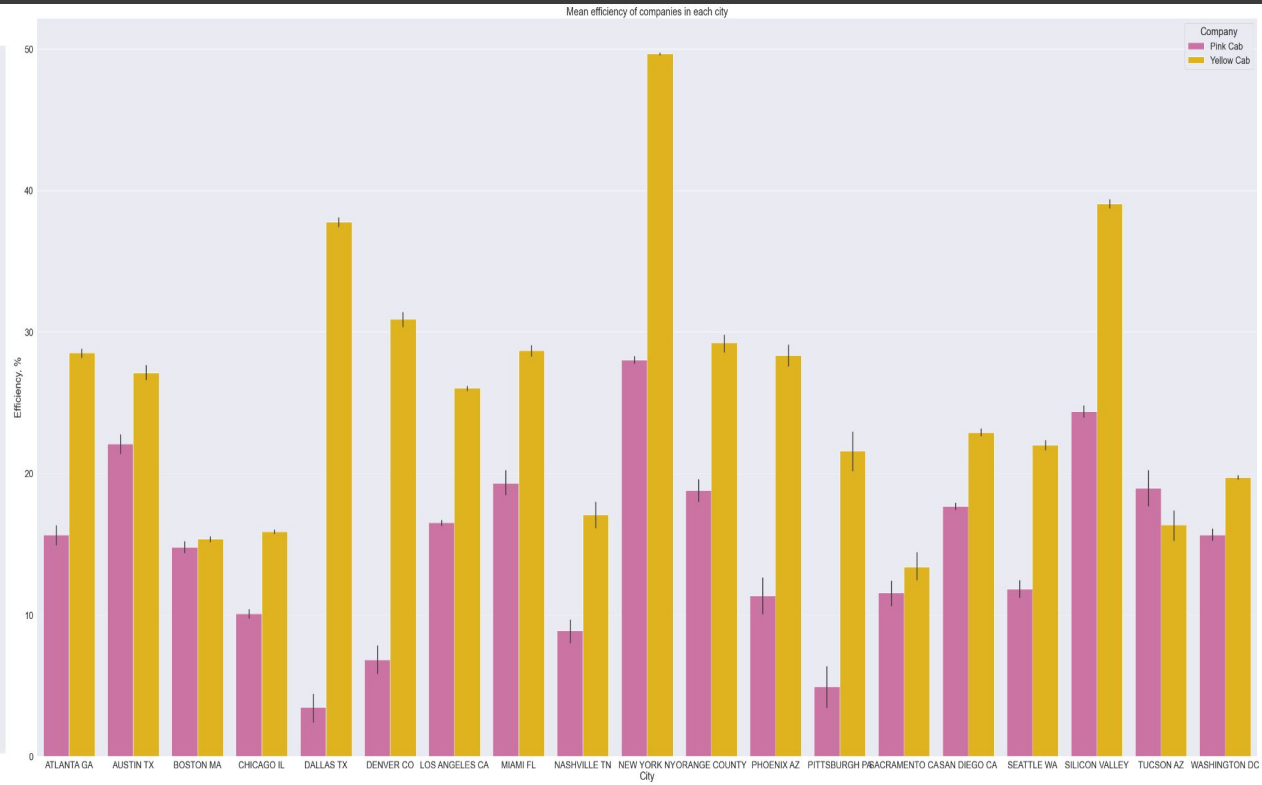


The hypothesis is confirmed

Data Glacier
Your Deep Learning Partner

Hypothesis No.4: The more transactions a company has in a particular city, the higher its efficiency in that city relative to another company



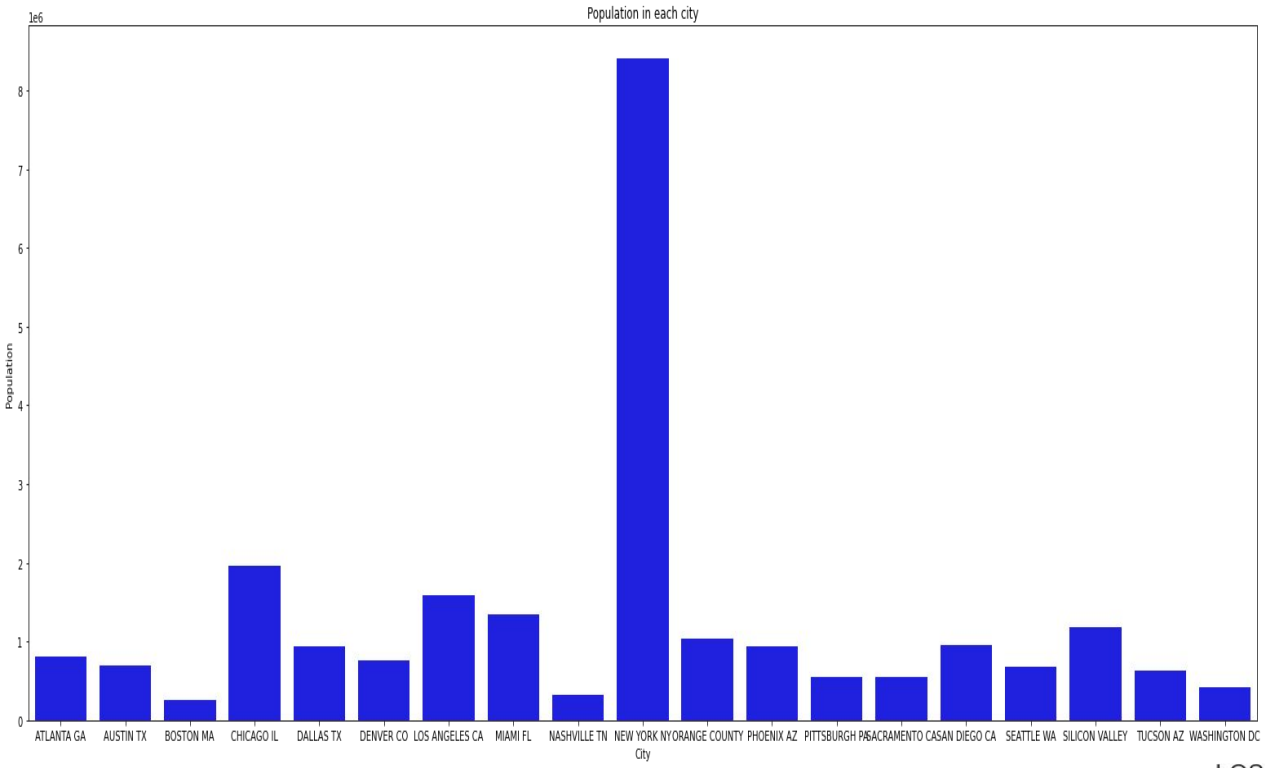


The hypothesis is rejected

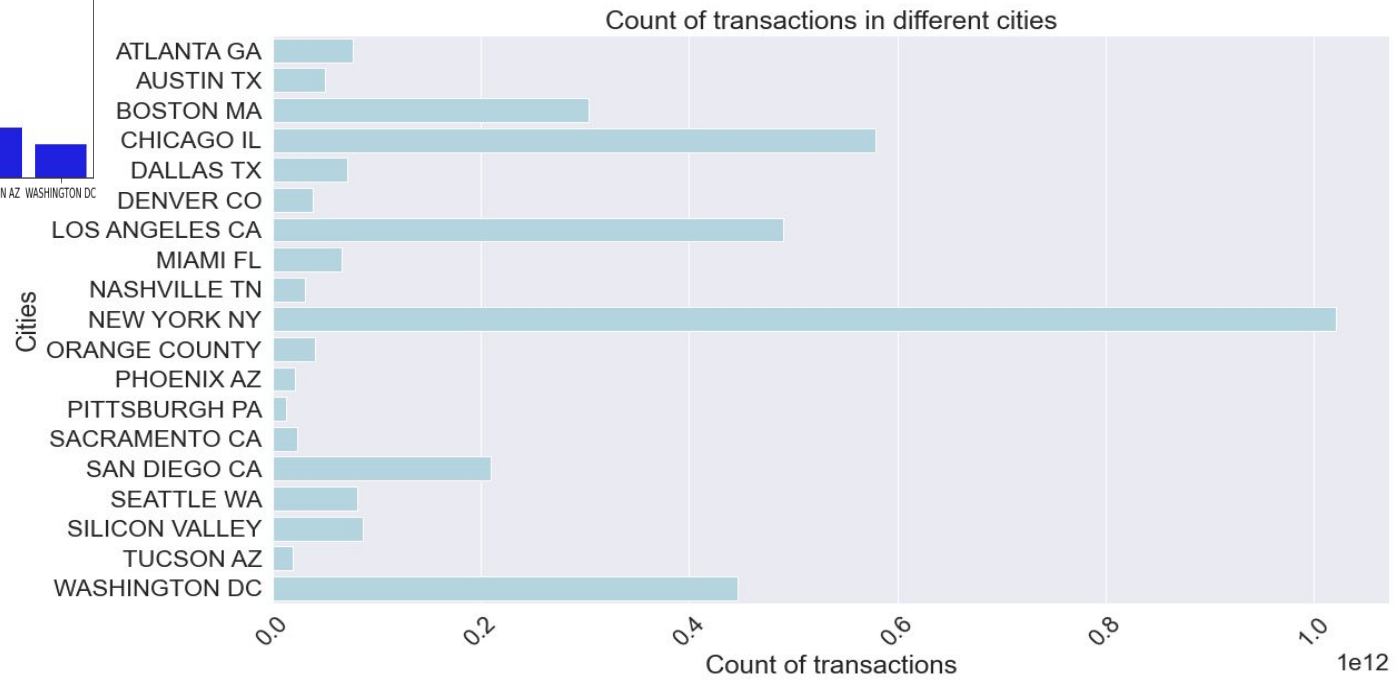
There are cities in which companies have very high efficiency, but they do not have very many transactions. And cities where a lot of trips take place are far from being in the forefront in terms of efficiency. The only exception is New York, as it is the most populated city in the United States.

Data Glacier
Your Deep Learning Partner

Hypothesis No.5: The larger the population of a city, the more trips there are



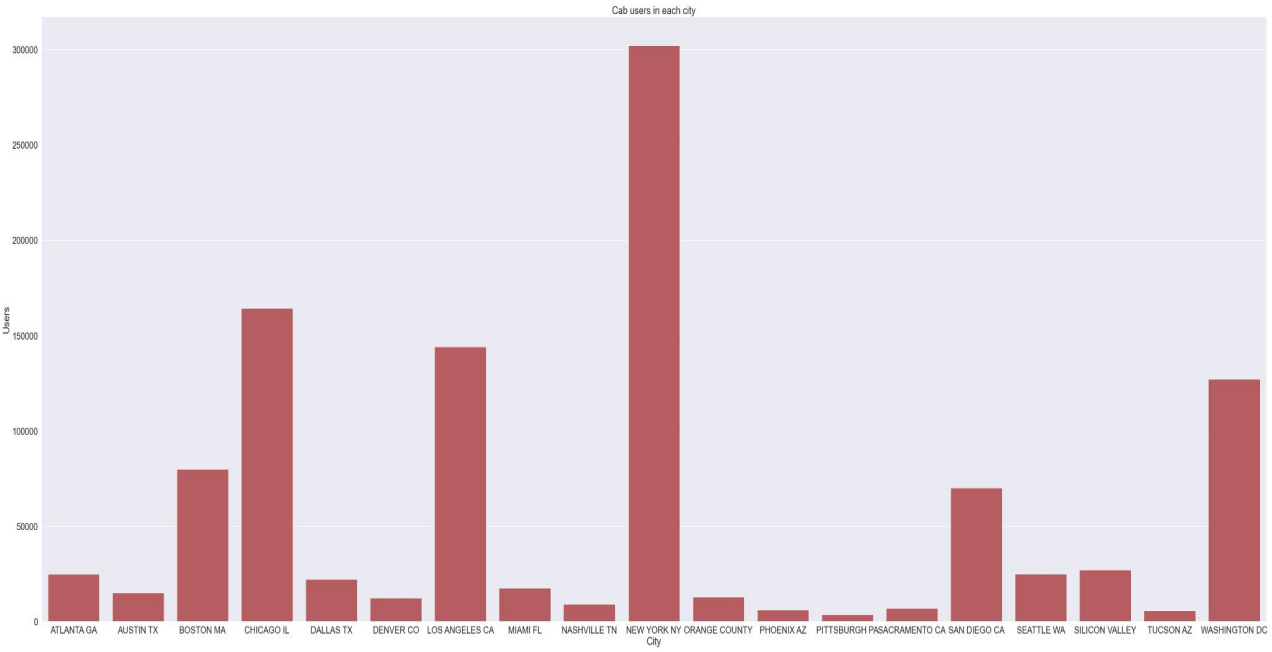
- Miami and Silicon Valley have high population figures, however, the number of transactions is low.
- Boston and Washington, which are in the top five in terms of the number of trips, are the least populated cities in America compared to the others in this row.
- New York, Chicago and Los Angeles follow the trend in this hypothesis, but these are only exceptions to the rule, judging by the chart.



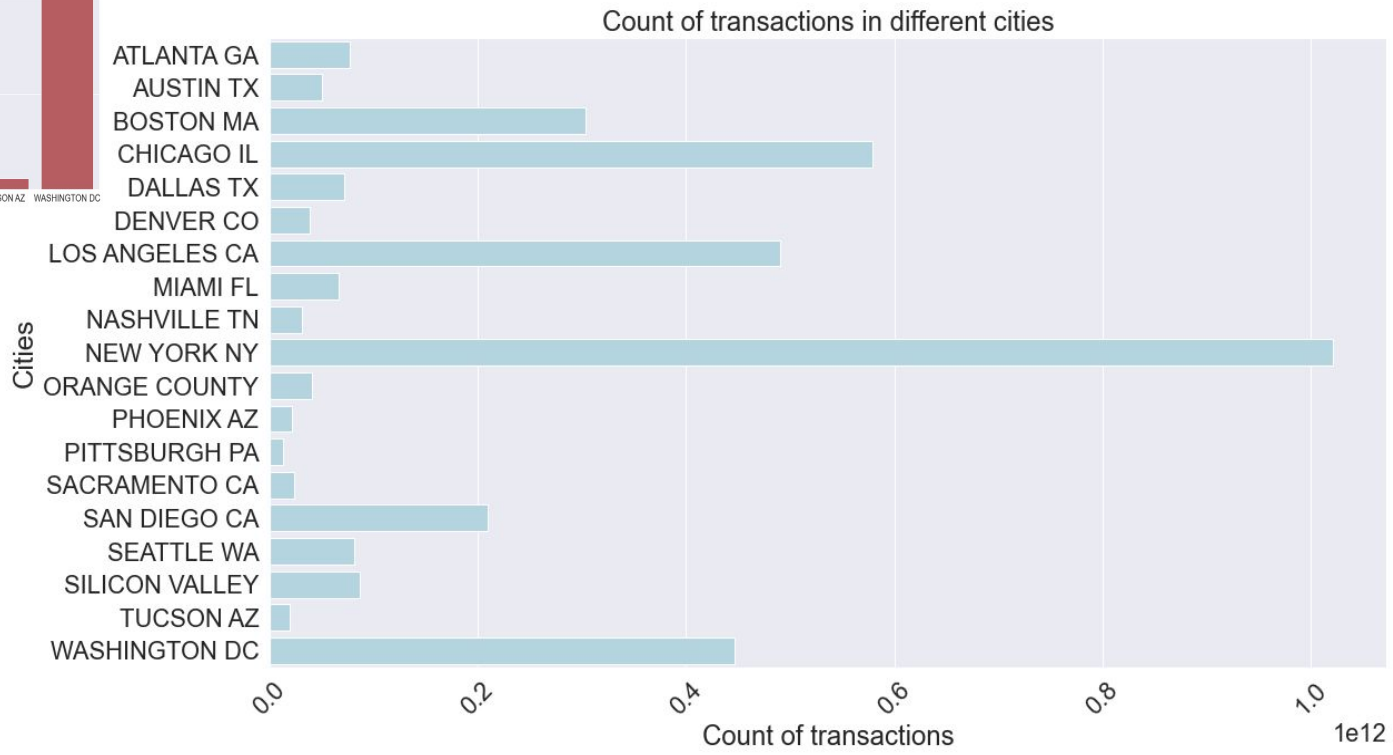
The hypothesis is rejected

Data Glacier
Your Deep Learning Partner

Alternative hypothesis: The larger the cab users in a city, the more trips there are

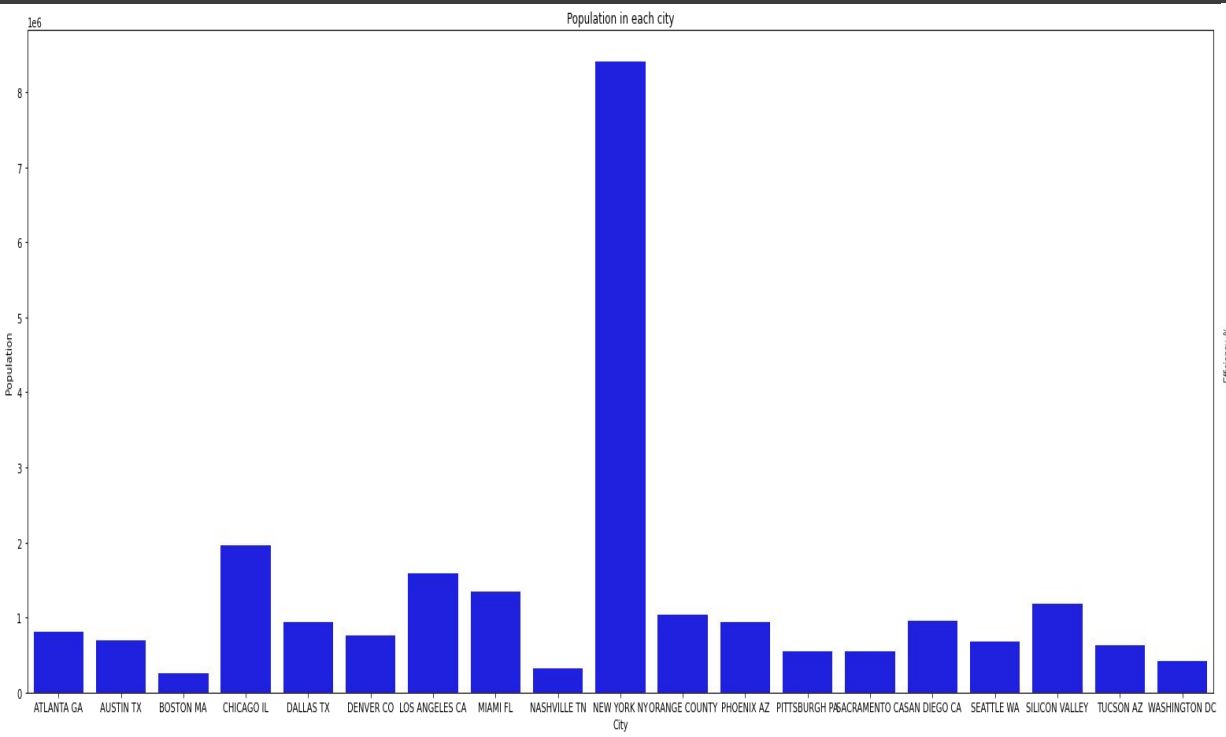


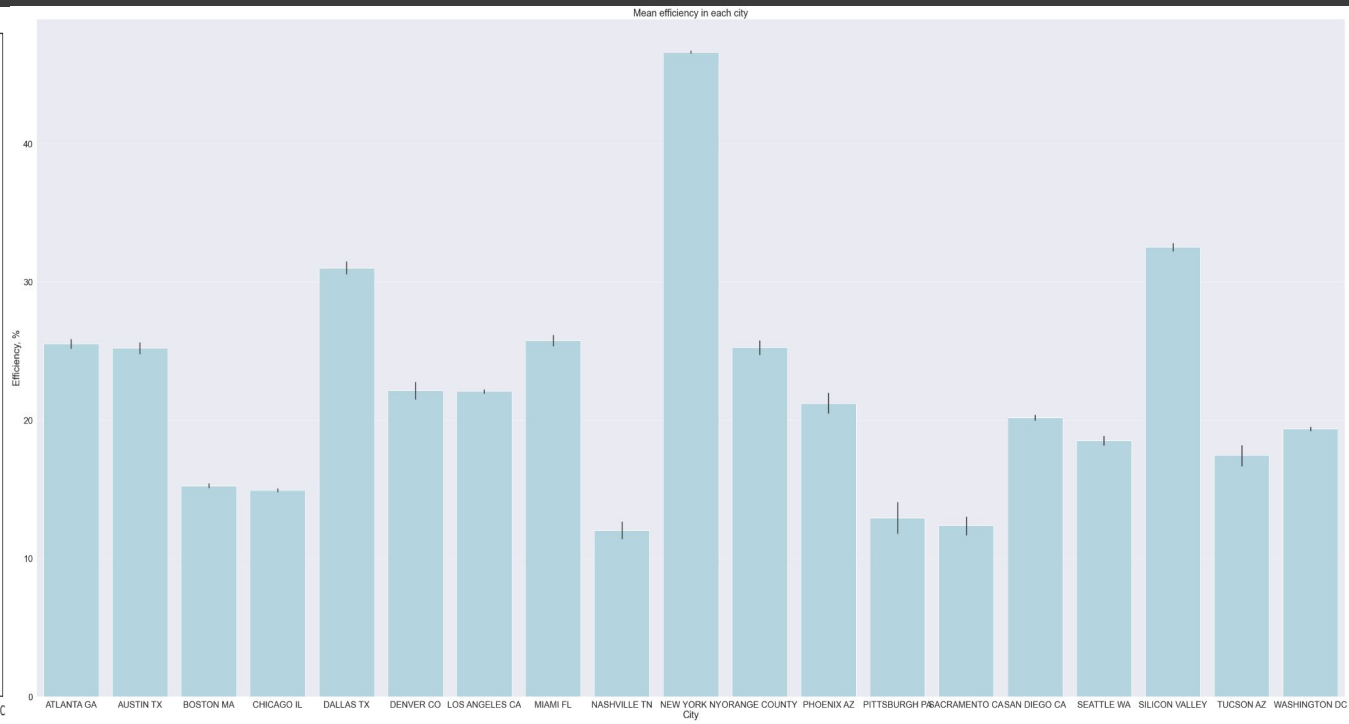
The best relationship can be found between the number of cab users in total by city and the number of transactions.



Data Glacier
Your Deep Learning Partner

Hypothesis No.6: The larger the population of the city, the higher the efficiency





Population is not as strong a factor influencing efficiency as the average income of clients (next slide).

The hypothesis is rejected

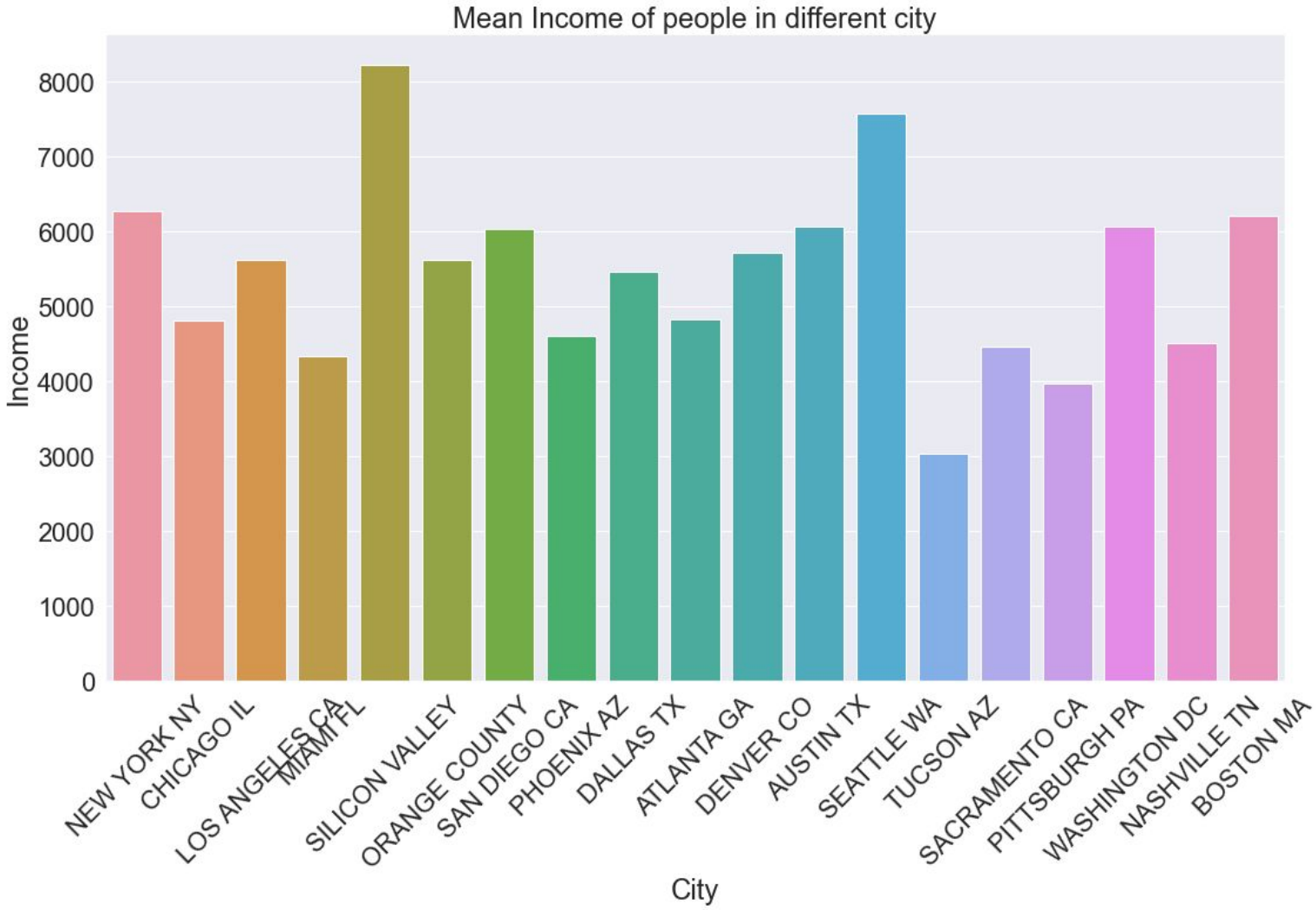
Data Glacier
Your Deep Learning Partner

Alternative hypothesis: The larger the average Income of people in the city, the higher the efficiency of cab companies

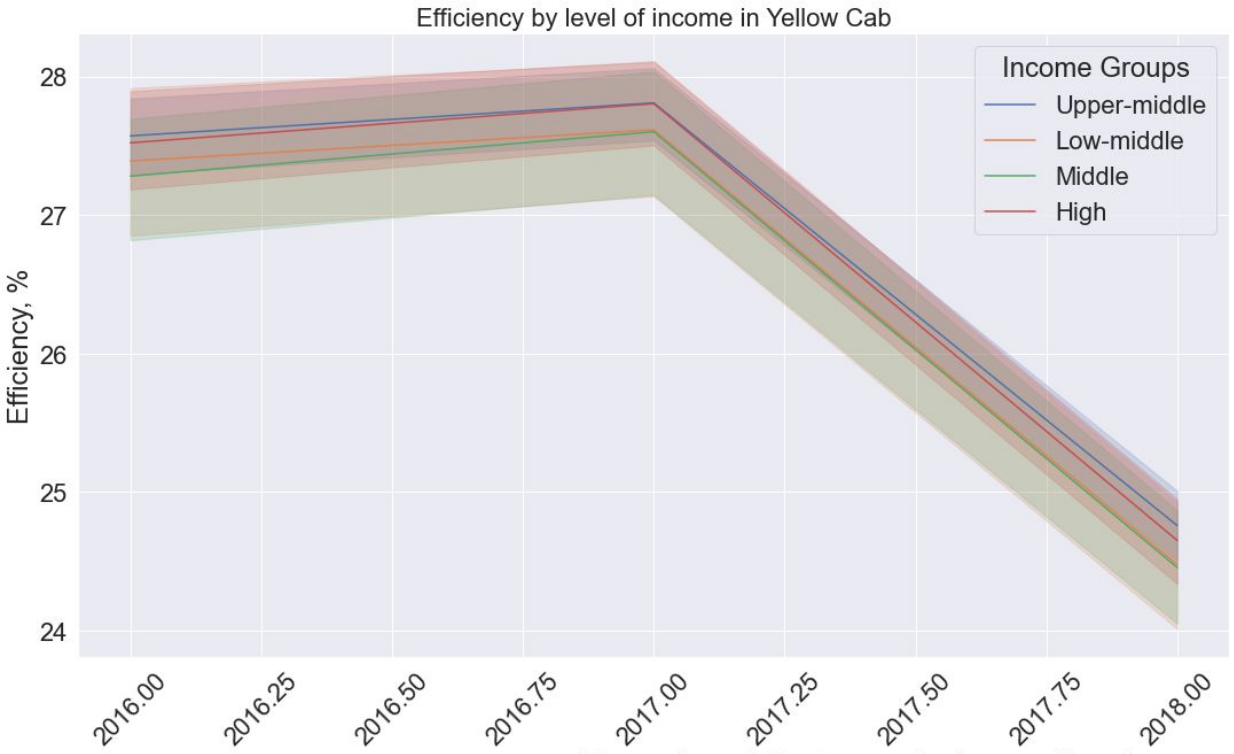
New dataset with info about average income among cities:

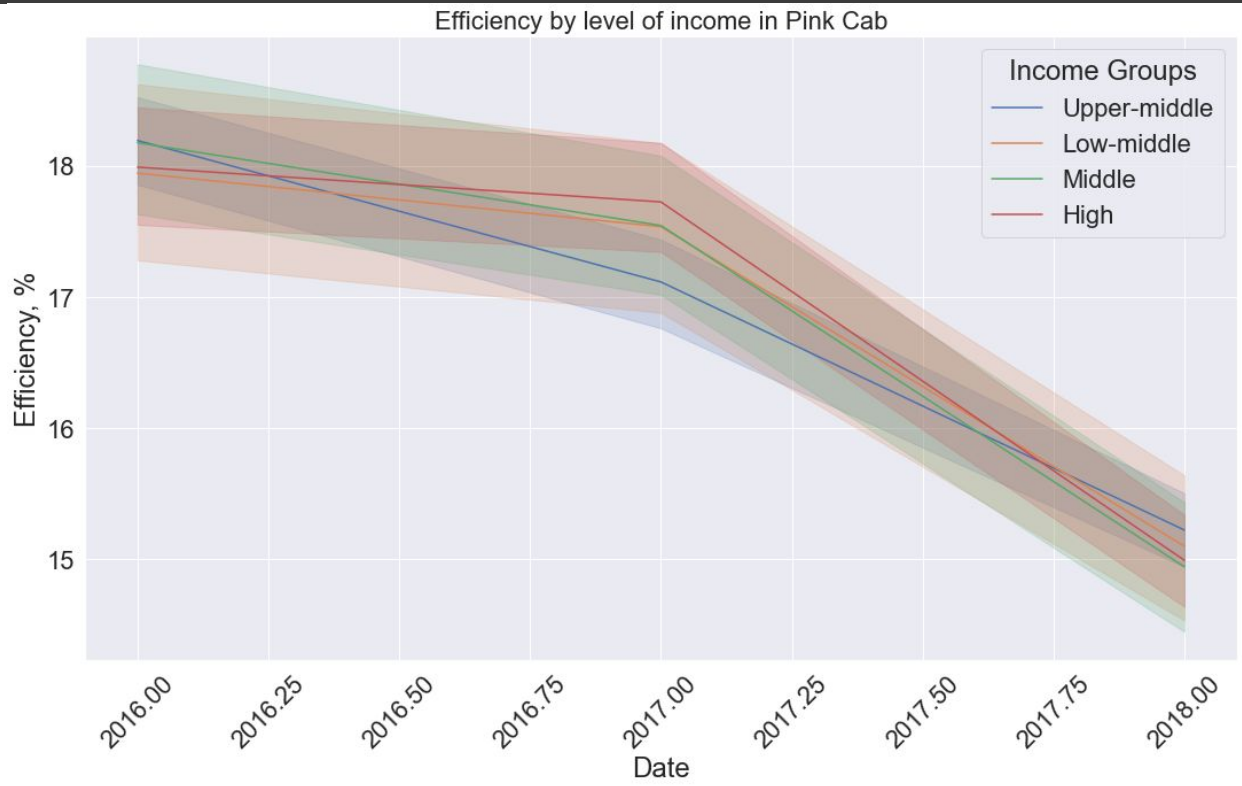
	income
432.0+ bytes	**Information about the average income in cities**
object	City
int64	Income

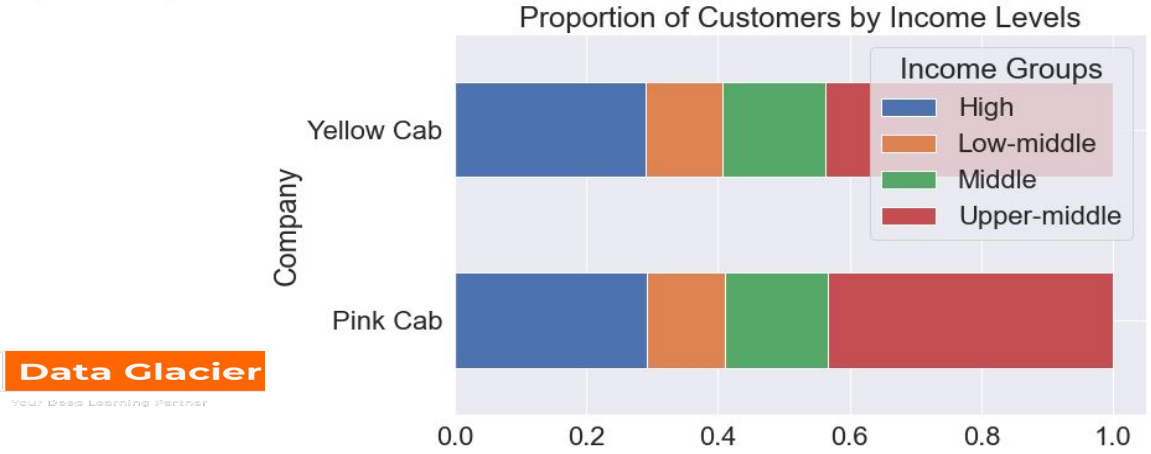
Cities with a high average income have higher efficiency.



Hypothesis No.7: The higher the income of the company's customers, the higher its efficiency





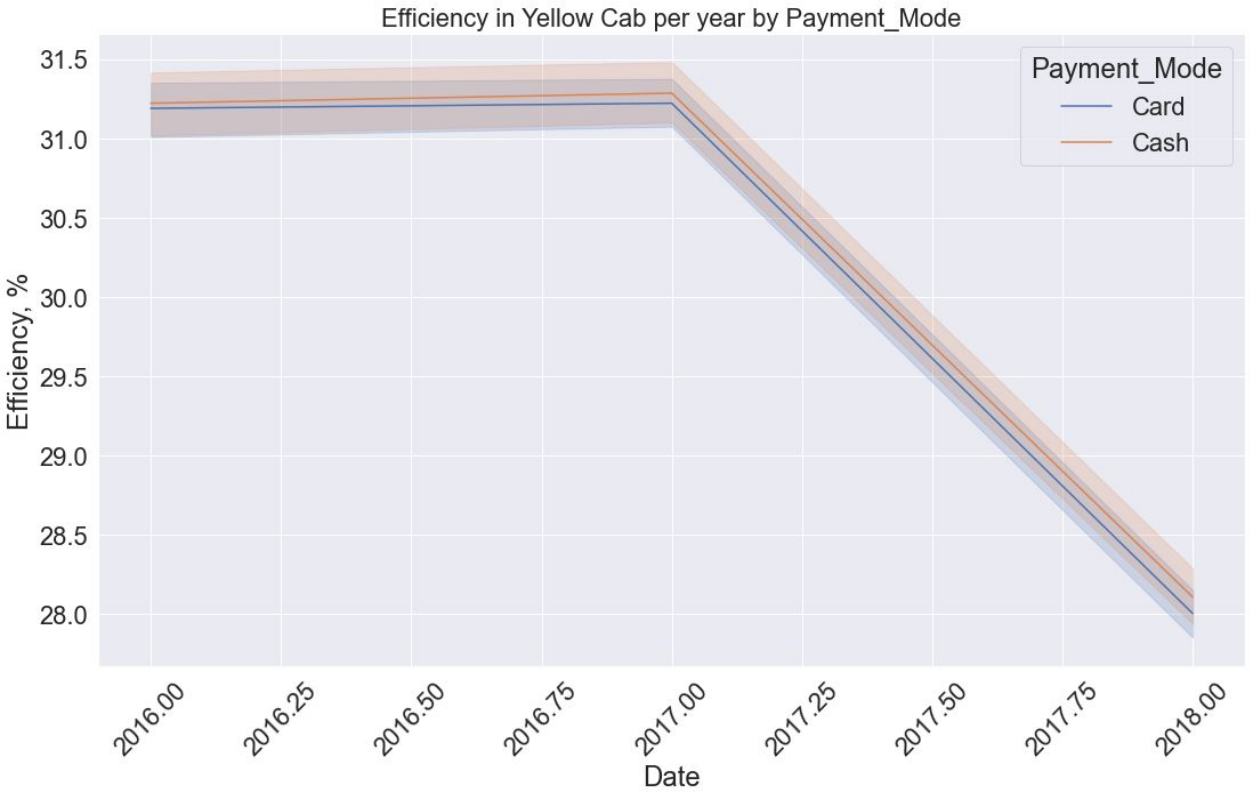


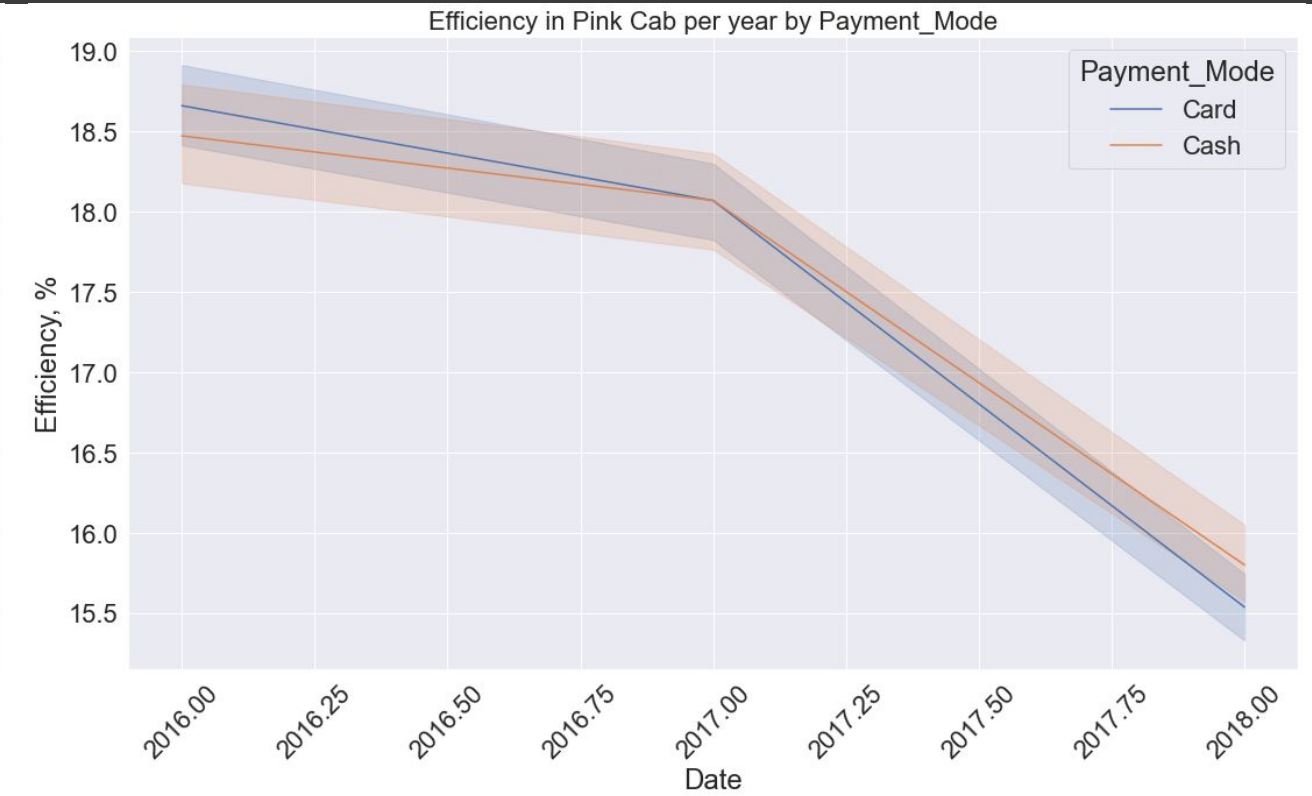
There is no difference in efficiency between different income groups of customers among these 2 companies. This is because the last graph shows that the distribution among clients of companies with different income levels does not differ.

The hypothesis is rejected

Data Glacier
Your Deep Learning Partner

Hypothesis No.8: When paying for a trip with a bank card, the efficiency of the company is lower than when paying in cash



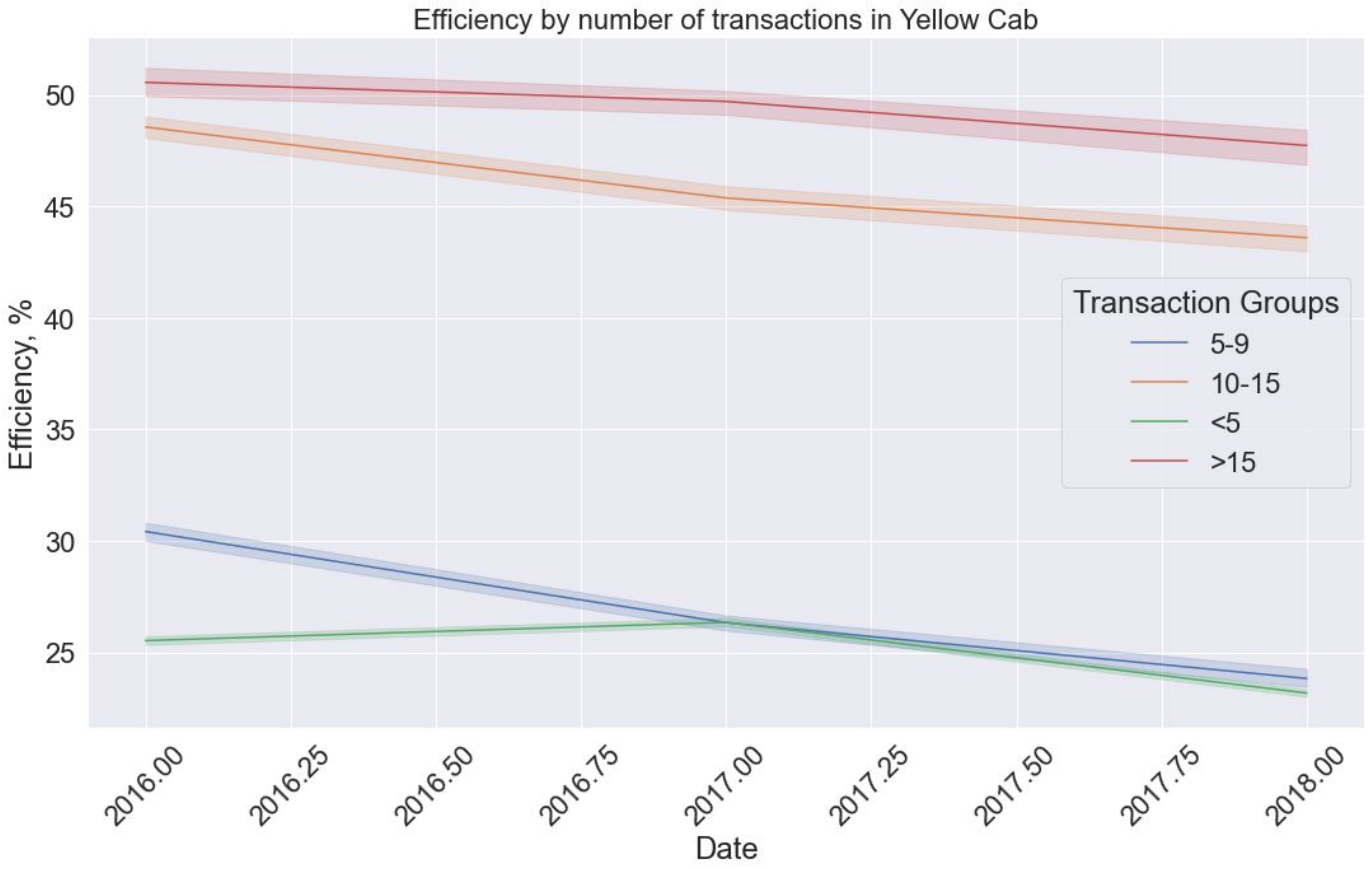


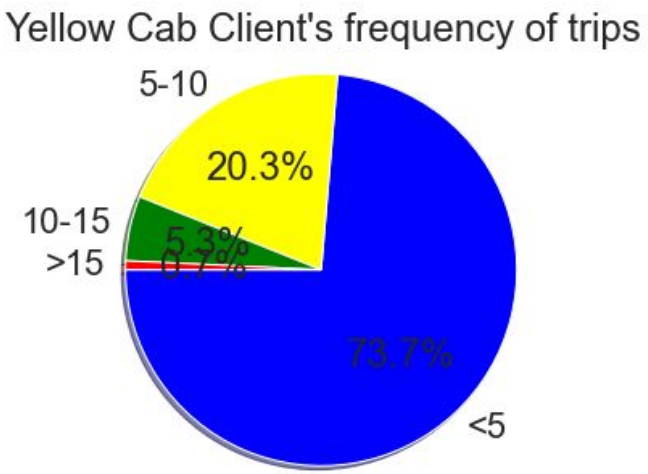
Since companies incur transaction costs when paying with a card, their efficiency becomes lower. At the beginning of 2016, there are doubts about this hypothesis. This is due to the fact that in the Pink Cab company during this period there is an opposite situation and the efficiency of payment by card is higher than when paying in cash.
My guess: the cab drivers were cheating.

The hypothesis is confirmed

Data Glacier
Your Deep Learning Partner

Hypothesis No.9: Loyal customers lead to greater efficiency (Yellow Cab)



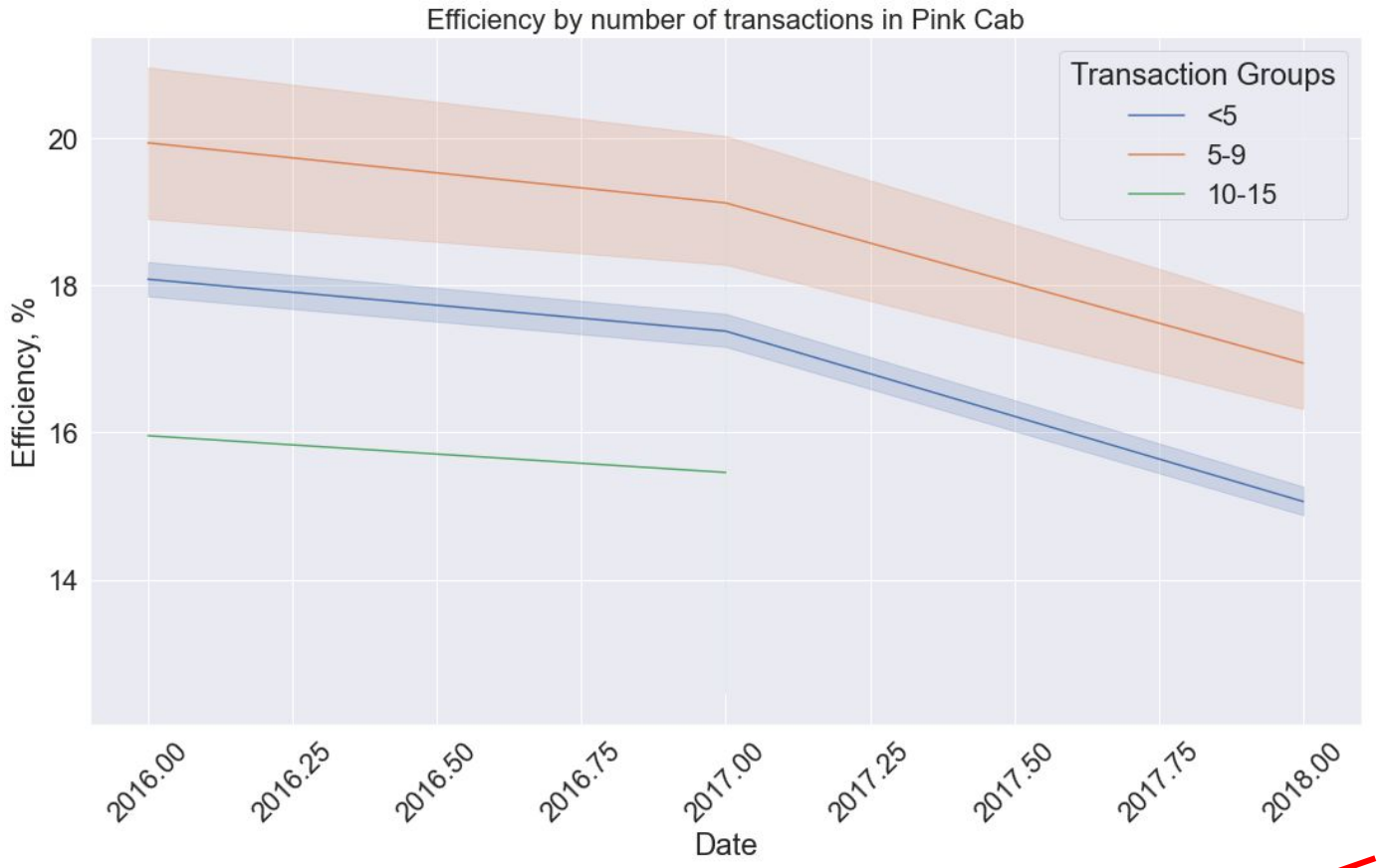


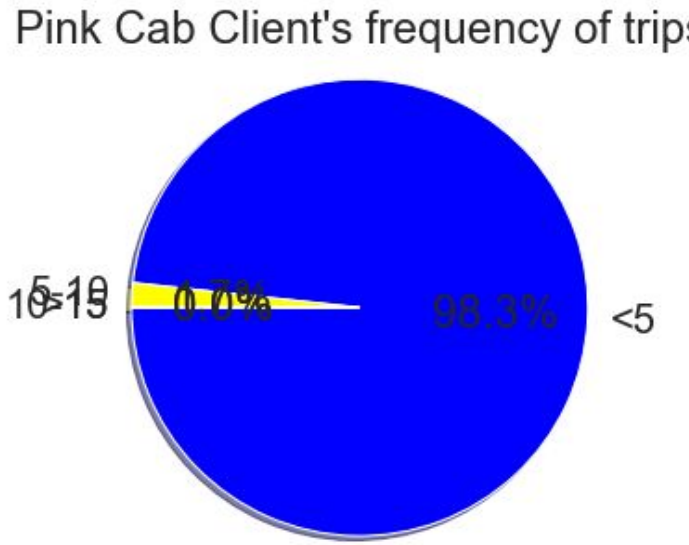
The Yellow Cab data confirms the hypothesis and the more trips made by customers, the more efficiency they bring to the company.

The hypothesis is confirmed

Data Glacier
Your Deep Learning Partner

Hypothesis No.9: Loyal customers lead to greater efficiency (Pink Cab)



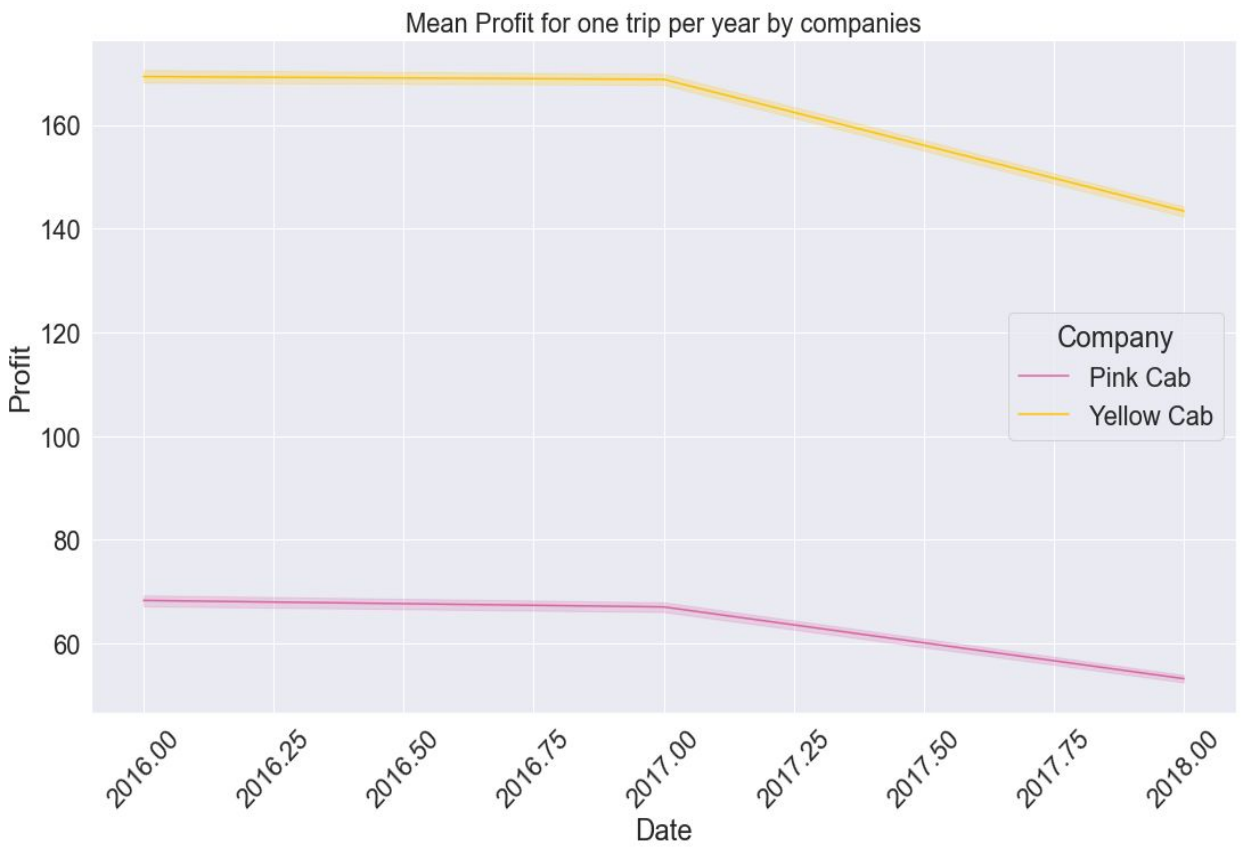


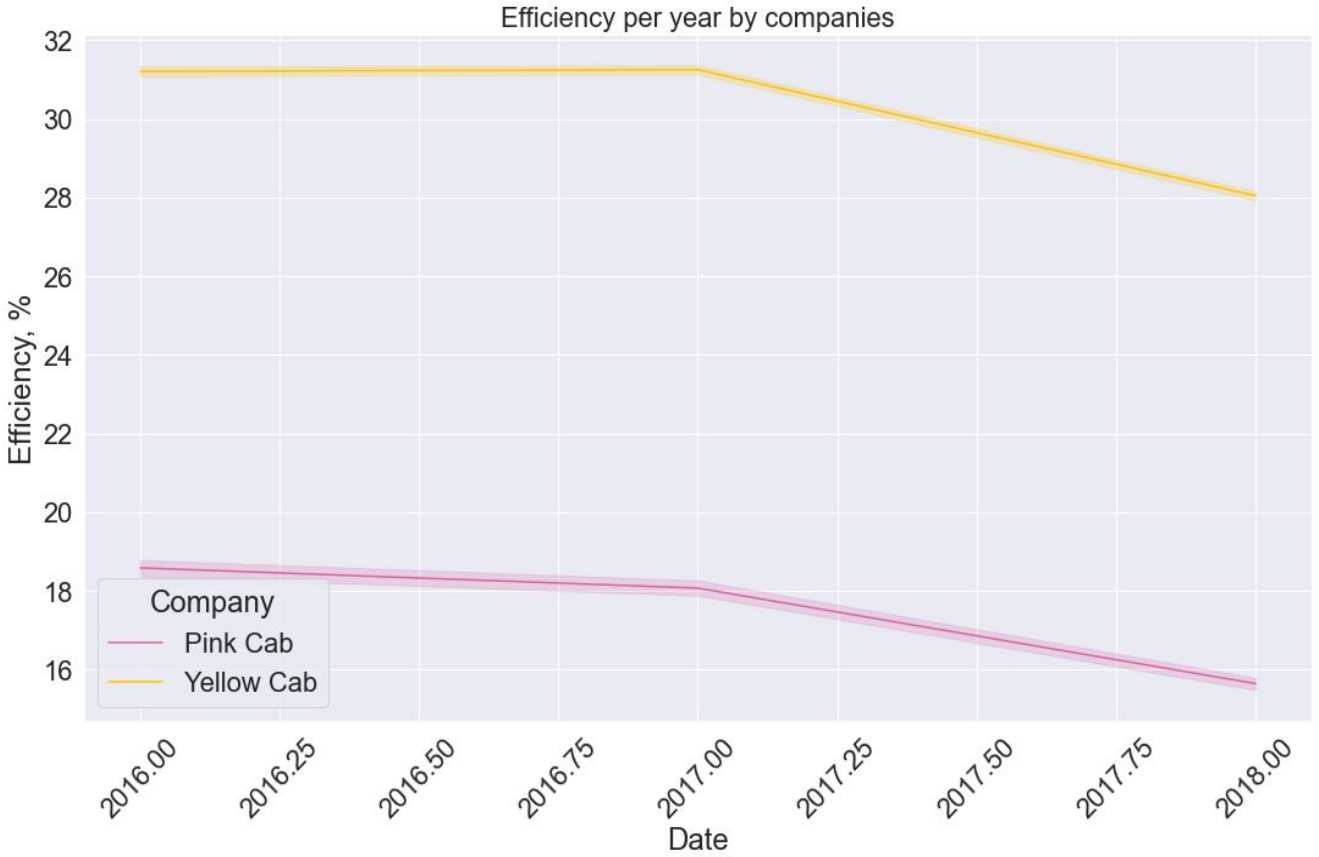
The hypothesis is rejected

The hypothesis is rejected in the case of Yellow Cab, but we believe that since Pink Cab's repeat customer rate is less than 2%, these results can be ignored when confirming the hypothesis.

Data Glacier
Your Deep Learning Partner

Hypothesis No.10: The more profit from one trip, the higher the efficiency.





The hypothesis is confirmed

Data Glacier
Your Deep Learning Partner

Recommendations

1. The data shows us that the 2016-2018 market is expanding both in terms of demand and supply. This is due to a sharp slowdown in the growth of the number of transactions, which indicates the emergence of new companies in the industry and the reduction in the cost of a trip, as well as the attempts of the Yellow Cab and Pink Cab companies to maintain their existing share of customers by lowering prices. As we can see, the efficiency of the Yellow Cab remains above the market average even against the backdrop of a price reduction, so the Yellow Cab is more investment attractive from this point of view.
2. Looking at the efficiency by city, we saw that the efficiency is higher in those cities where the average income of people is higher. The Yellow Cab company has more market share in such cities. Therefore, the acquisition of such a company will allow doing business in cities with a large percentage of the solvent population.
3. Yellow and Pink Cabs account for 30% of the total number of cab users in large cities, while Yellow Cabs account for about 25%, so since we confirmed the hypothesis that the number of customers increases efficiency, it is more profitable to invest in this company than in Pink Cab.
4. The Yellow Cab company has not only a large market share, but also a large number of repeat customers, namely more than 25% of customers who have made a trip more than 5 times. In addition, the analysis of efficiency by group of loyal customers showed that customers who make more trips provide more efficiency for the company, which is also the growth potential of the Yellow Cab company.
5. Against the backdrop of declining revenue, it is important for the company to optimize its costs. After rising costs in 2017, the Yellow Cab company found resources to smoothly reduce costs, which also indicates a greater potential for the company's efficiency.

The investment recommendation is to buy a Yellow Cab company.

References

BloombergNEF URL: https://about.bnef.com/

Lyft’s prospectu URL: https://investor.lyft.com/financials-and-reports/sec-filings/default.aspx

Uber Investor URL: https://investor.uber.com/financials/default.aspx

Statista URL: https://www.statista.com/

Intro the minds URL: https://www.intotheminds.com/blog/en/

Thank you!