# Exploratory Data Analysis

Cab Industry Analysis

**15/07/2022**

# Agenda

Business Understanding

Data Understanding

Profit Analysis

User Analysis

Trips Trend Analysis

Recommendations

Data Glacier
Your Deep Learning Partner

# Business Understanding

There are two cab companies in USA represented for investment: Yellow Cab and Pink Cab.

The goal is to conduct an analysis in order to identify and select the one company to invest in.

Insights will be explained and justified in the next chapters.

# Data Understanding

4 Datasets were provided for an analysis:

- Cab_Data.csv;
- City.csv
- Customer_ID.csv
- Transaction_ID.csv



Customer_ID —— on Customer ID ——> Transaction_ID —— on Transaction ID ——> Cab_Data

Data Glacier
Your Deep Learning Partner
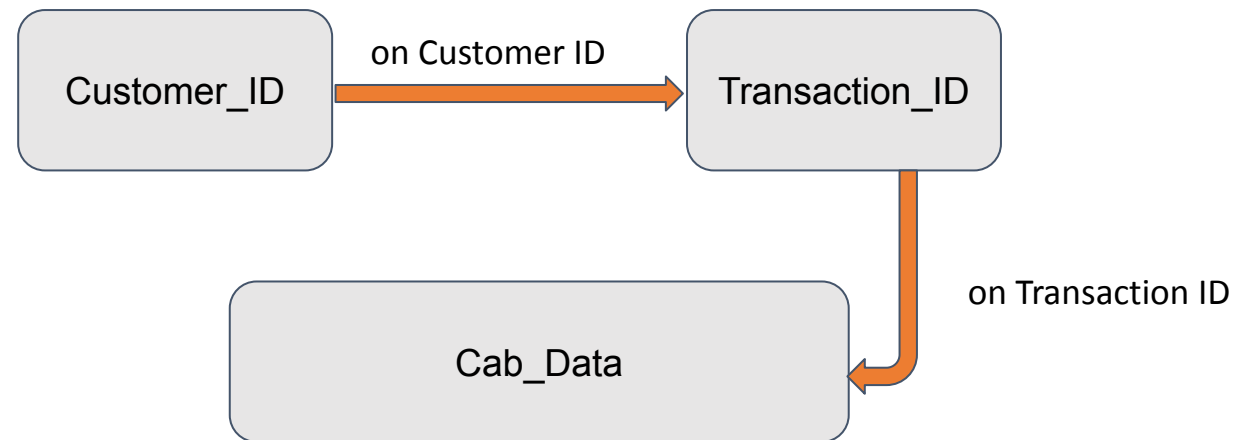
# Data Understanding

The selected datasets didn't contain any missing values, however a few problems were faced needed to be solved:

1. Dataset City.csv had commas in columns separating number instead of dots, making data type 'object' instead of 'integer' thus commas needed to be replaced with dots;

2. Date of Travel column of Cab_Data.csv was represented in excel format, therefore needed to be formatted into python datetime format.
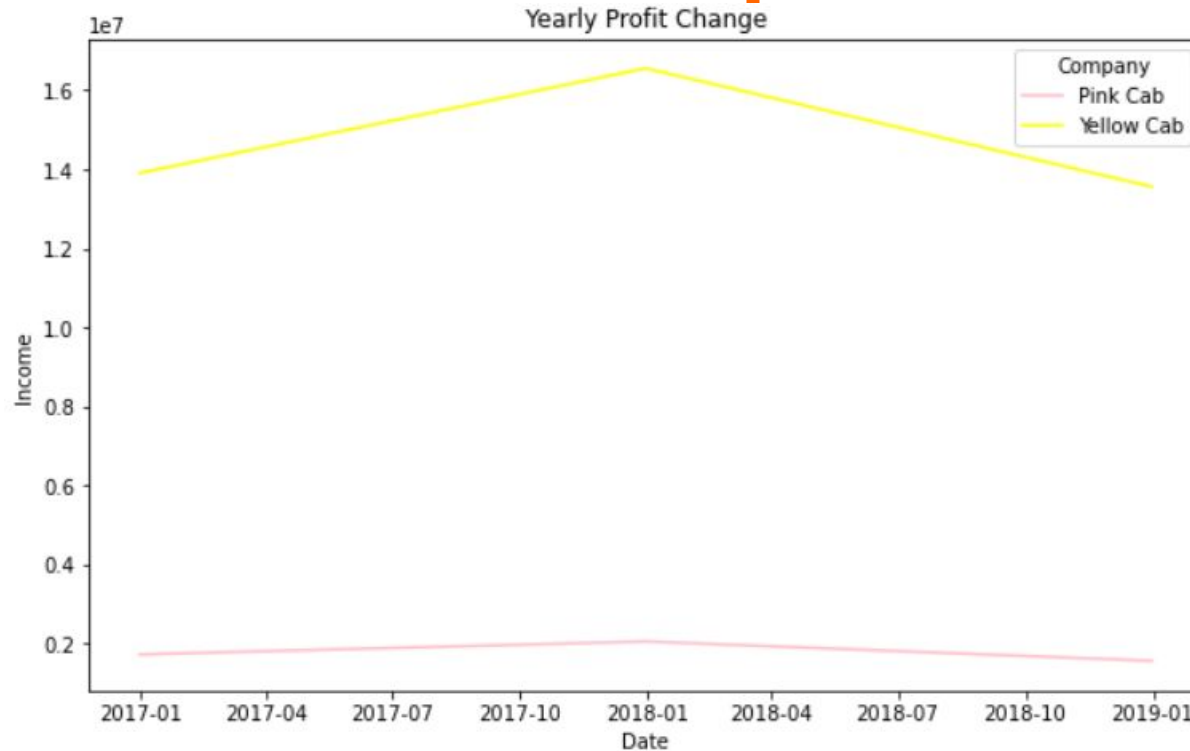
# Profit Analysis of each company

# What company does have a better profit?
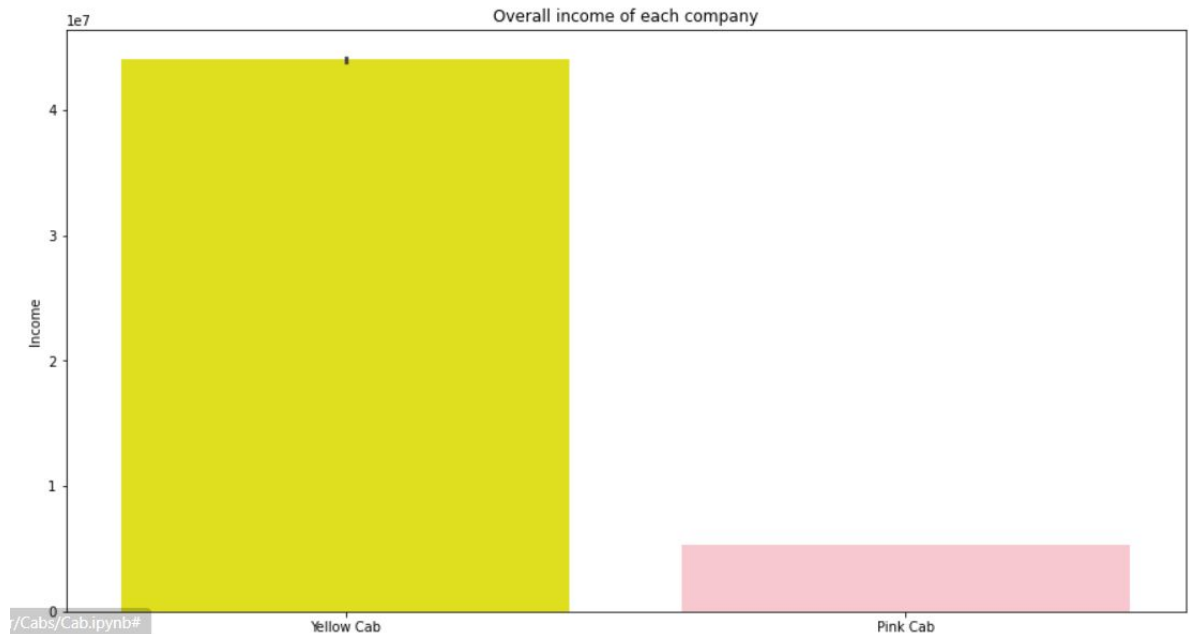


It appeared that Yellow Cab has made profit $44 000 000 since the beginning of 2016, while Pink Cab has made only $5 300 000. The difference of profit is clear on the graph.

# How does the profit change from year to year?


Yearly Profit Change

For both companies 2017 was the most profitable year. However, Yellow Cab's profit is higher.

# What company does charge more per km?



Price Charged Per KM in Different Cities
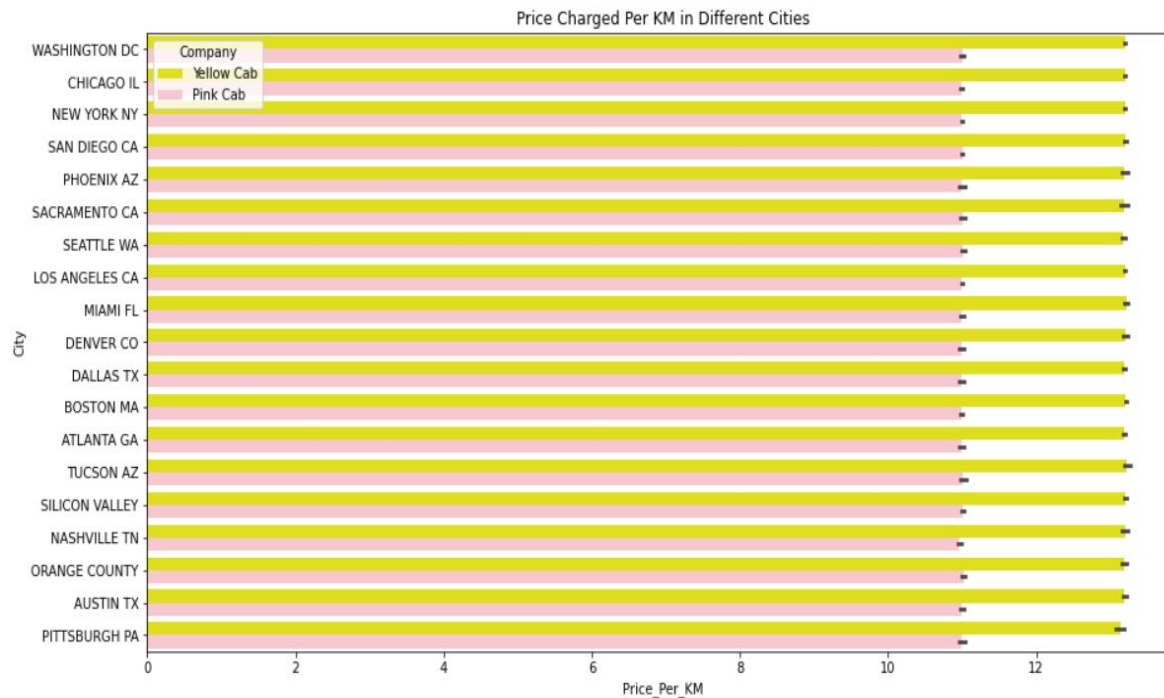
We can understand why Yellow Cab is more profitable than Pink Cab by looking at some details, eg. what company charges more per km.

As initially the dataset didn't contain required feature, we performed feature engineering and created a new feature Price_per_km by dividing Cost of Trip by KM Travelled.

Overall, price per km doesn't differ from city to city. However, Yellow Cab does charge $2 more per km, which brings the average price $13 per km, while Pink Cab charges $11 per km.

# User Analysis of each company

# What are the top-5 cities with the largest population?
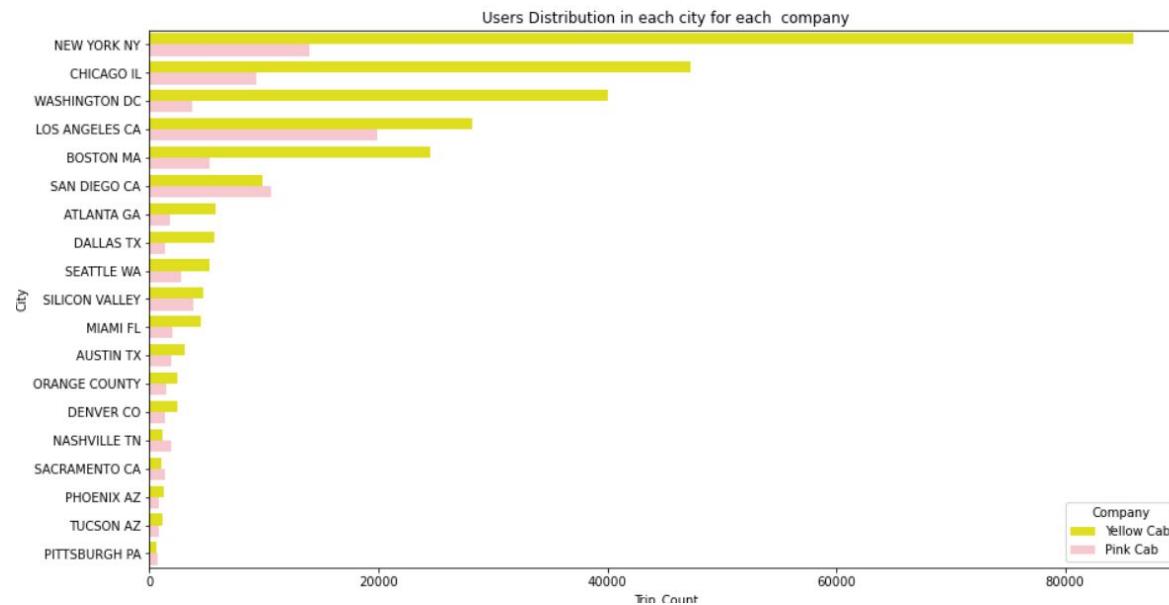

Cities with the highest population
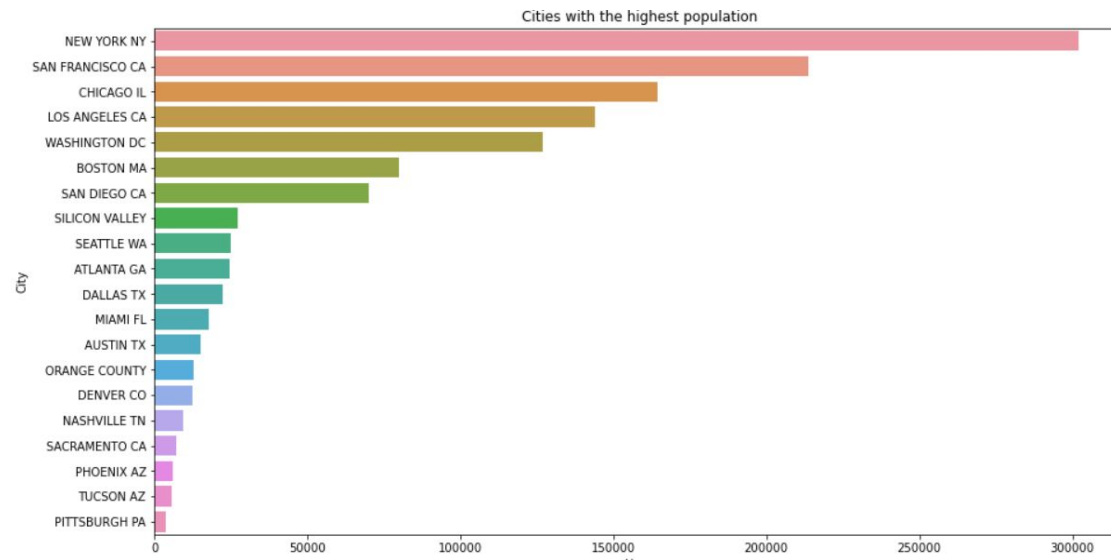
Top 5 cities by number of users:

- New York (NY)
- San Francisco (CA)
- Chicago (IL)
- Los Angeles (CA)
- Washington (DC)

Let's see how users are distributed in these top-5 cities for each company.

# What company does occupy a larger market share?



Users Distribution in each city for each company

It's clear that Yellow Cab controls the market and has done more trips in every city (apart from San Diego) than Pink Cab. In New York, Chicago and Washington Yellow Cab has much larger market share than Pink Cab.

# What company does have more users overall?

| Company | Customer_Count |
|---|---|
| Yellow Cab | 39896 |

| Company | Customer_Count |
|---|---|
| Pink Cab | 32330 |

**Unique customers**

| Company | Customer_Count |
|---|---|
| Pink Cab | 492 |
| Yellow Cab | 10715 |

Yellow Cab has 7 566 more users than Pink Cab. But I also want to see how many unique or loyal customers each company has as I noticed that both companies have customers with the same customer ID represented in both tables.

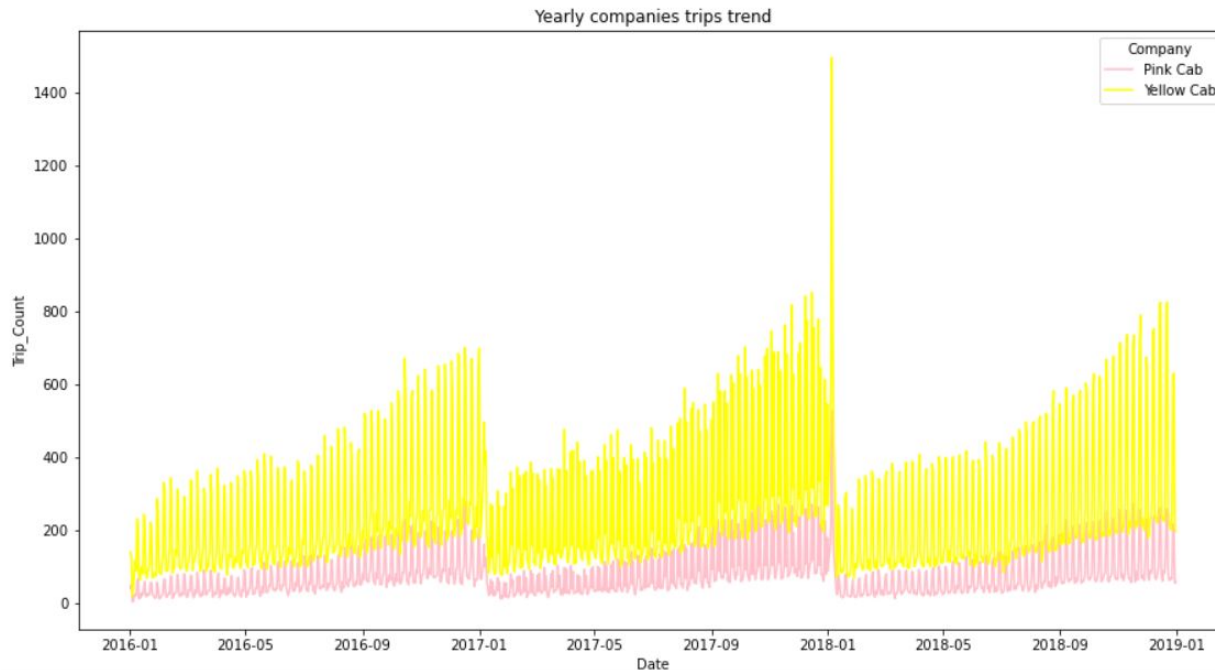While Yellow Cab has 10715 unique customers, Pink Cab has only 492.

# What is a company's target audience?
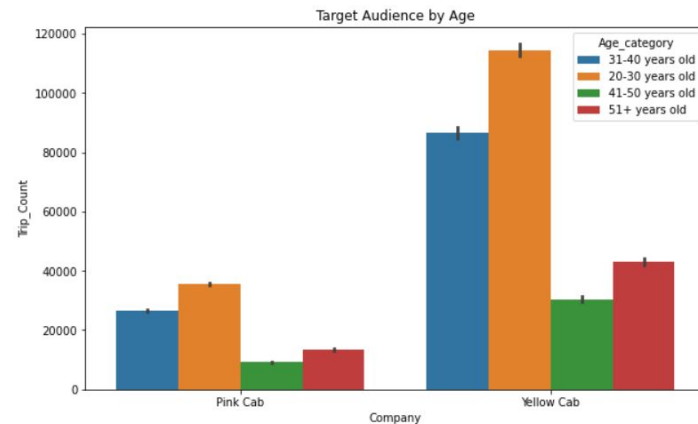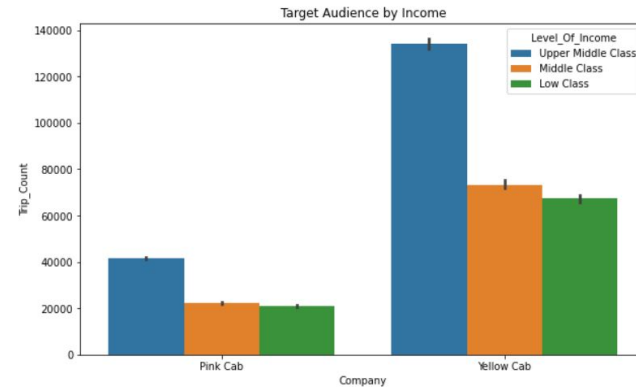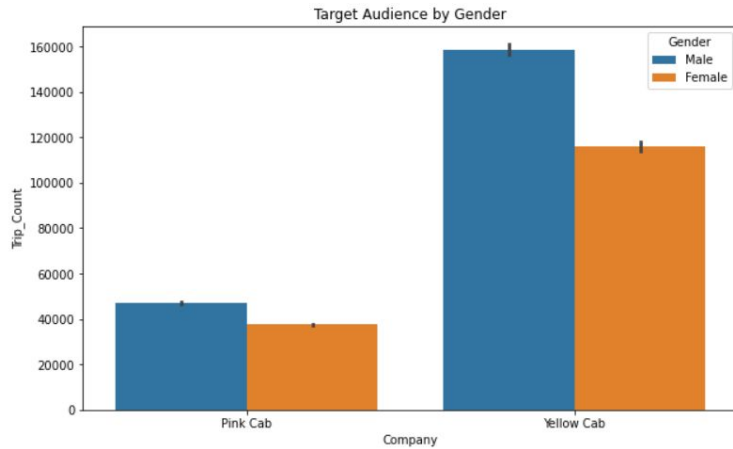


The target audience for both companies: Young male (20-40) years old with an income 15000 dollars per month or higher.

# Trips trend Analysis of each company

# What is a companies' trips trend??



From 1st of January up to the end of year we can observe a constant growth of users and then slump from the beggining of a year. However, after the slump a constant growth can be observed again. It is clear that seasonality trend takes place in number of users.

There's also a huge spike on January 5th 2018, it would be interesting to investigate.

Research shows that on this day a historic bomb cyclone happened which was a cause of high demand in taxi.

# Recommendations for investors

Considering everything mentioned above the recommendation would be to invest in **Yellow Cab** company.

Yellow Cab company has a better and more secure profit, more users and bigger market share overall.

# Thank You