



Data Glacier

Your Deep Learning Partner

Exploratory Data Analysis

G2M insight for Cab Investment firm

LISUM07 – Andersson Andréé Romero Deza

15-03-2022

Agenda

Background

Executive Summary

EDA

EDA Summary


Recommendations


Background


4 Datasets:

- Date of dataframe: 2016-01-08 to 2018-08-30
- Total data points: 359392

 **Cab_Data.csv** → Include details of transactions of two cabs companies.

 **Customer_ID.csv** → Include customer demographic details.

 **Transaction_ID.csv** → Include transactions and payment mode.

 **City.csv** → List of US cities, their population and cab users.

Agenda

Background

Executive Summary

EDA

EDA Summary

Recommendations

Executive Summary

Punch Statement

XYZ is a private 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.

The Problem

There are two companies cab in the market, you want to know which one to invest in.

The Solution

Analyze the data provided and make a decision based on the results.

The Opportunity

Enter one of the largest and highest paying industries today for technology.

Strategy

Data
Analysis

Recomm
endation

Cab
Selection

Investment

Our Team

Andersson
Romero
*Data
Scientist*



Agenda

Background

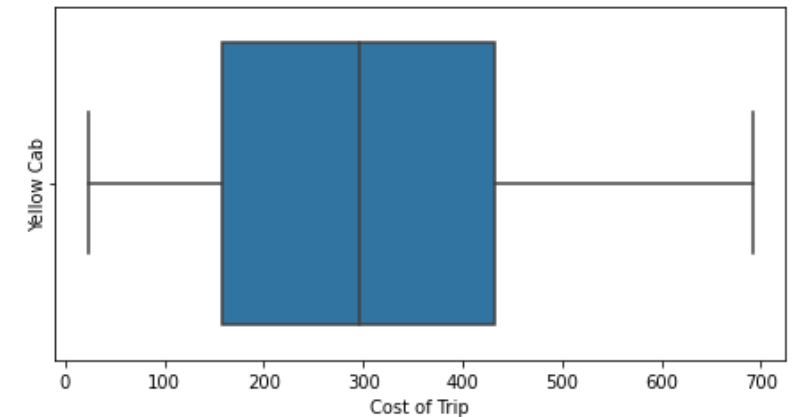
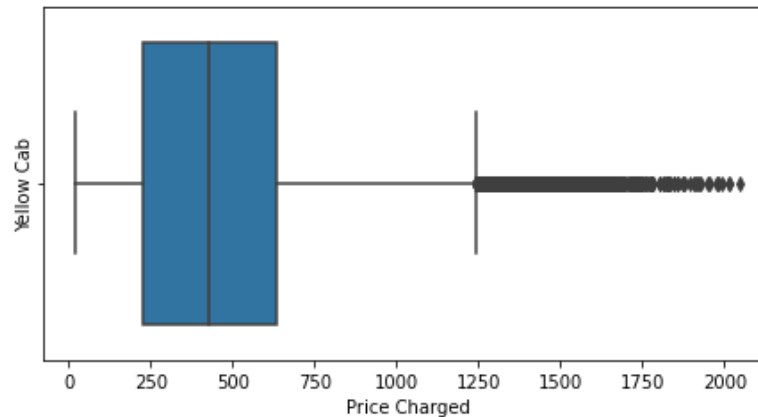
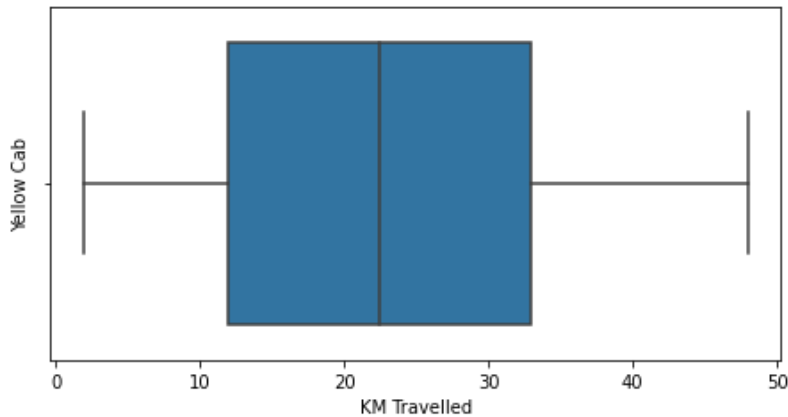
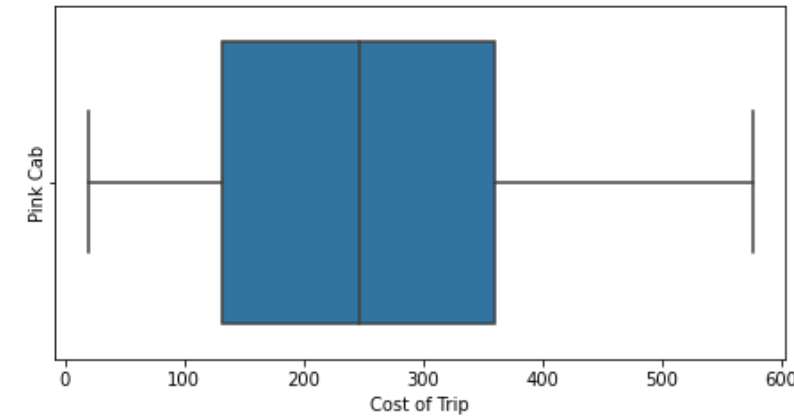
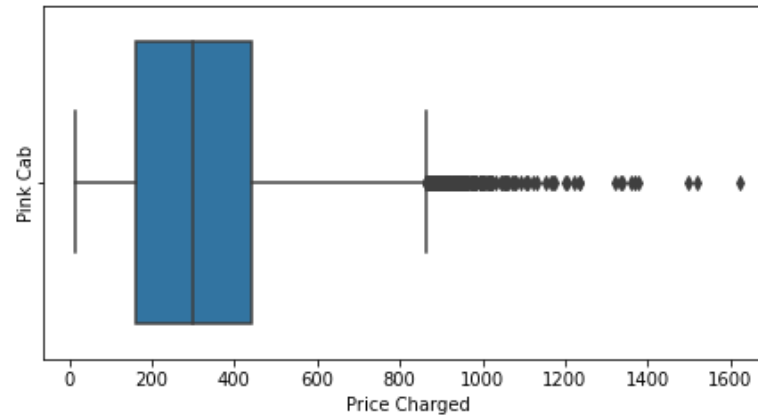
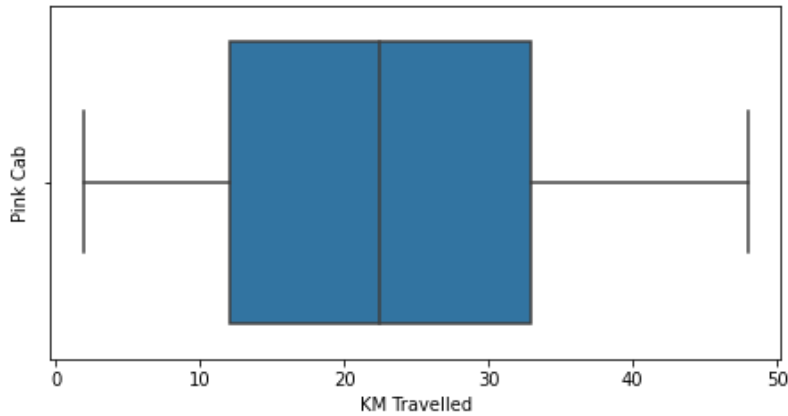
Executive Summary

EDA

EDA Summary

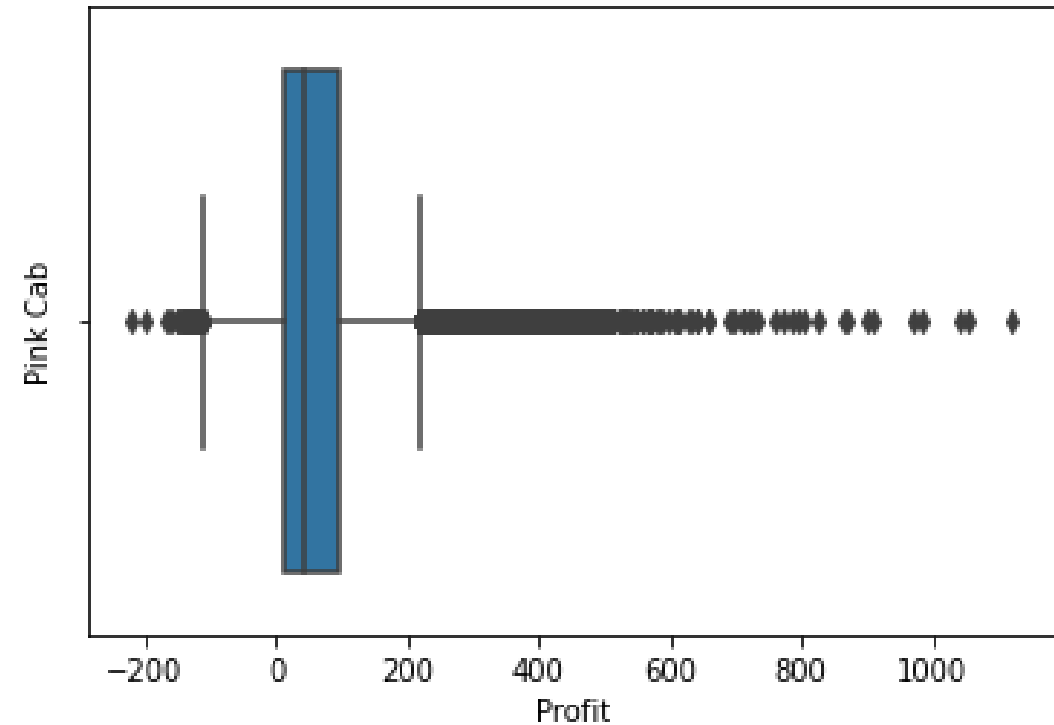
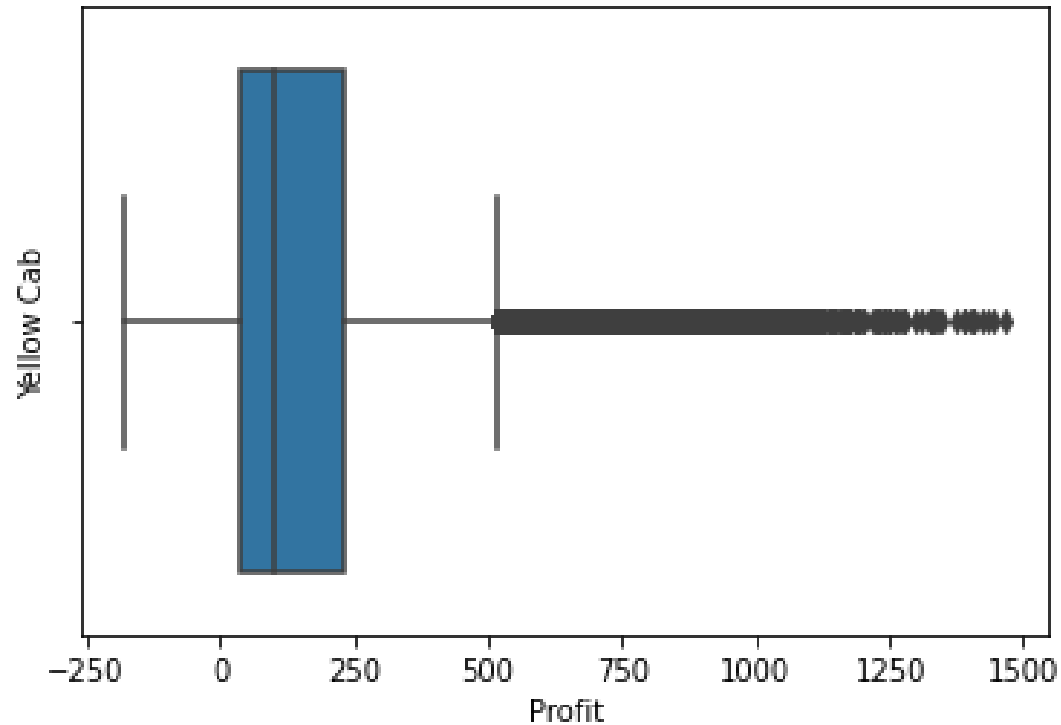
Recommendations

Boxplot Distribution of Variables



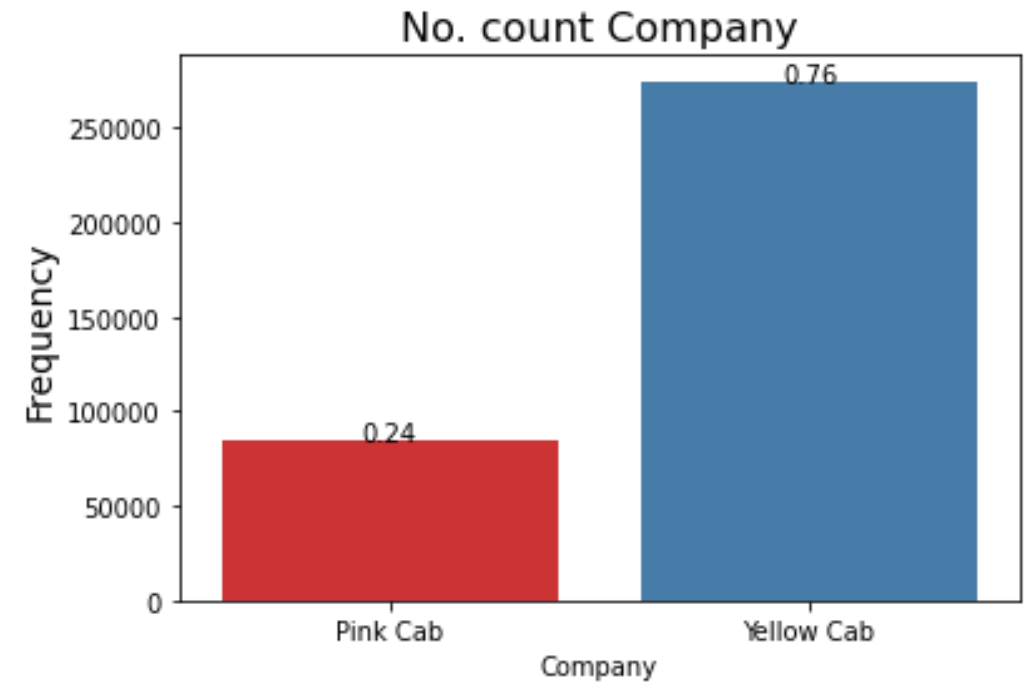
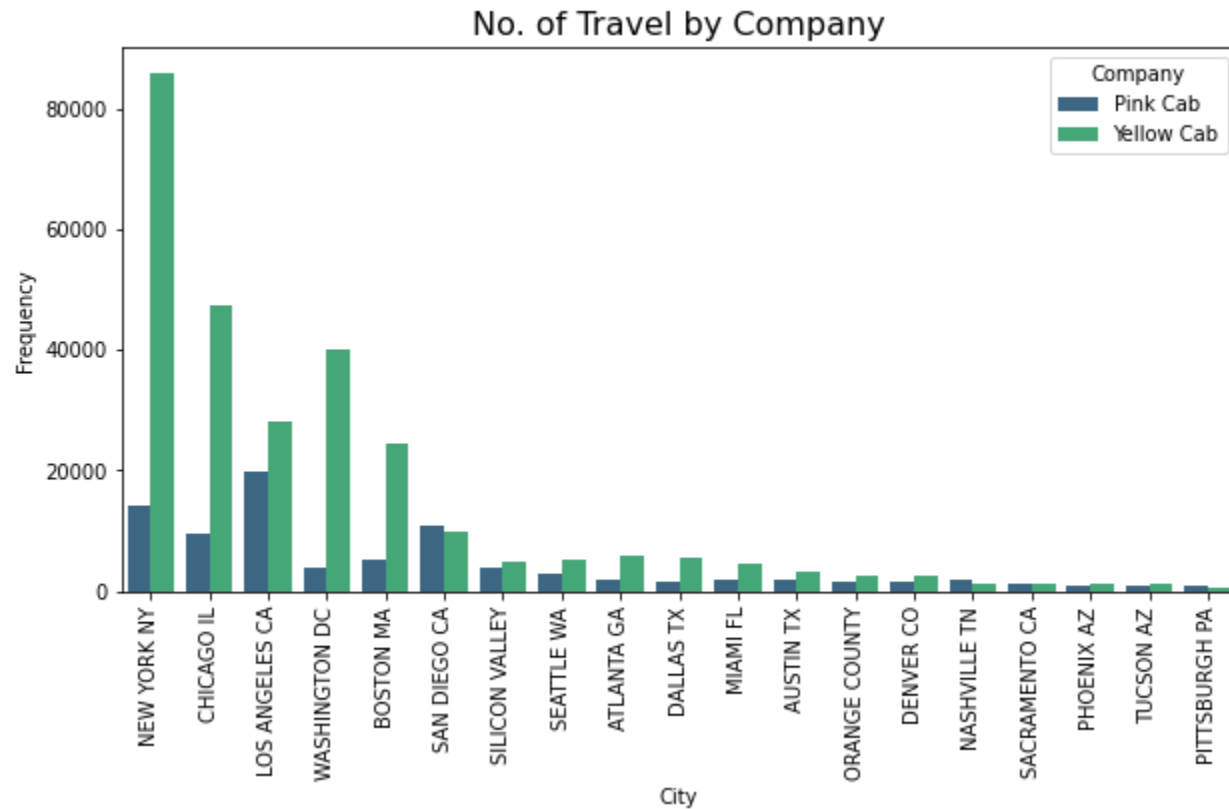
- The amount of KM traveled is very close in both cabs.
- The charged price is higher in the yellow cab.
- The cost is lower in yellow cab than Pink cab.

Boxplot Distribution of Variables



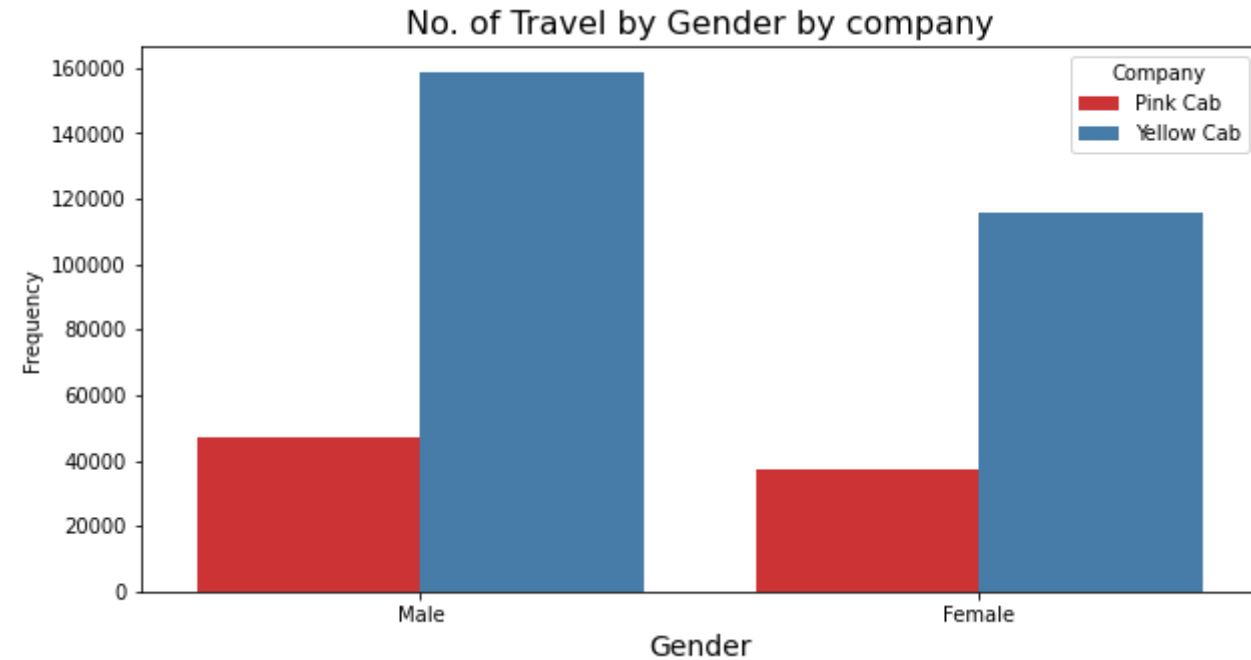
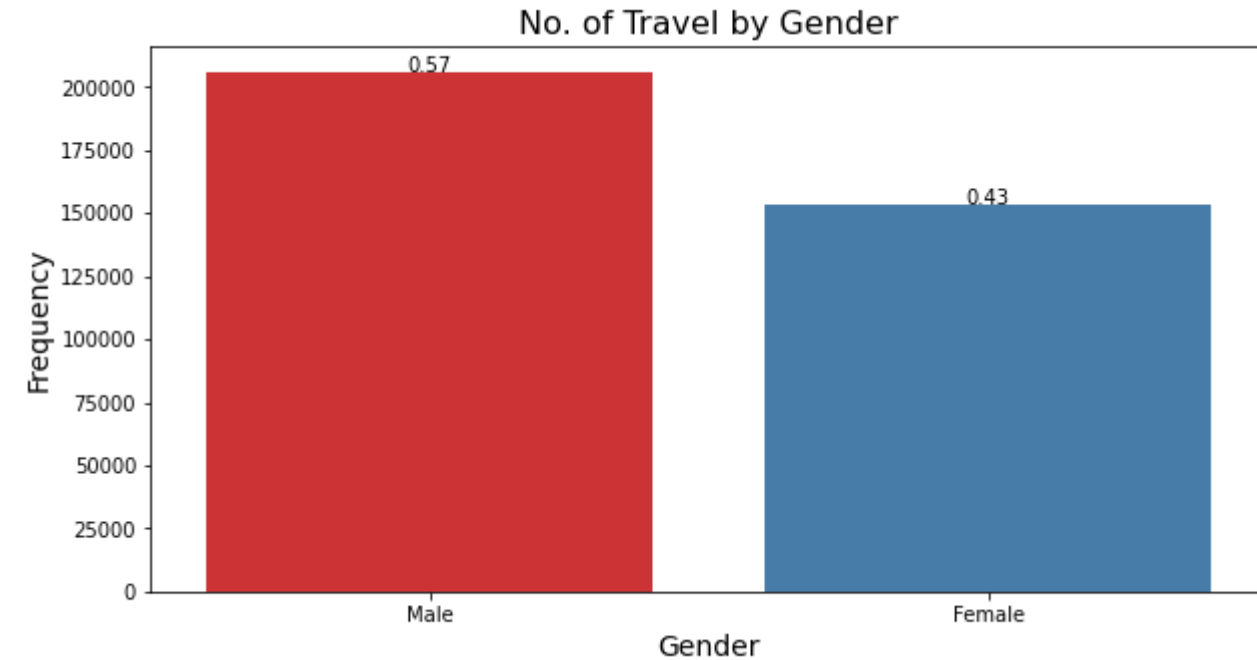
- The gain is higher in yellow cab.
- Pink cab even has losses in relation to the price charged and the cost of the trip.

Travels



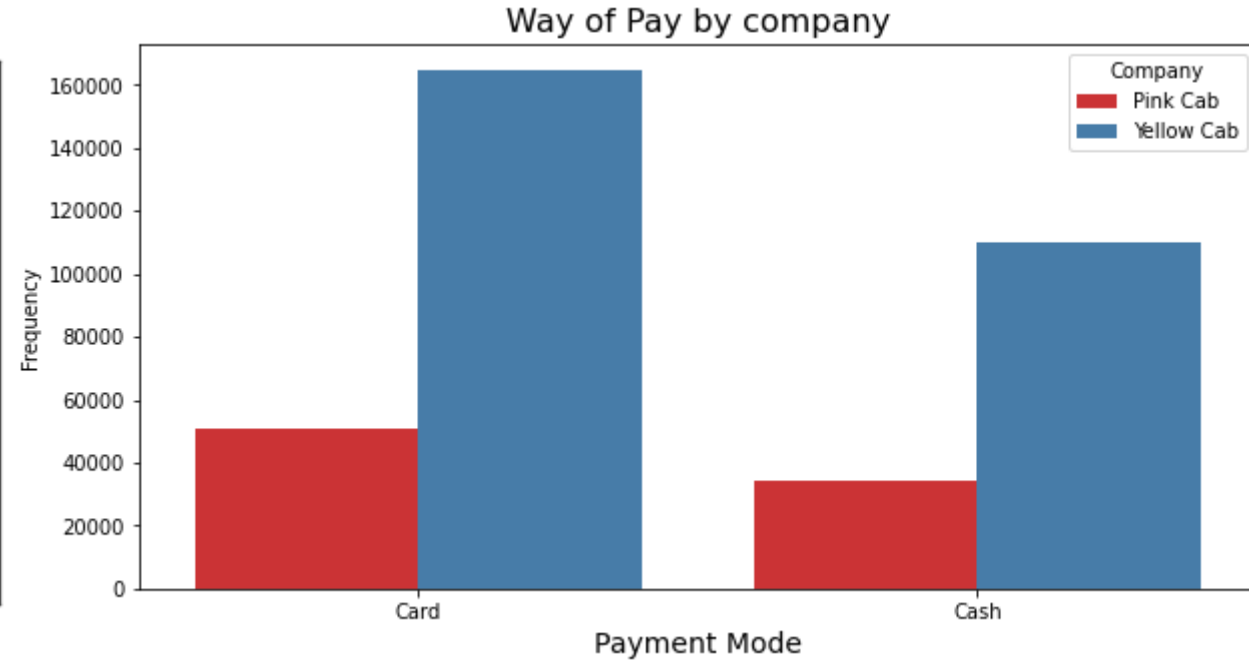
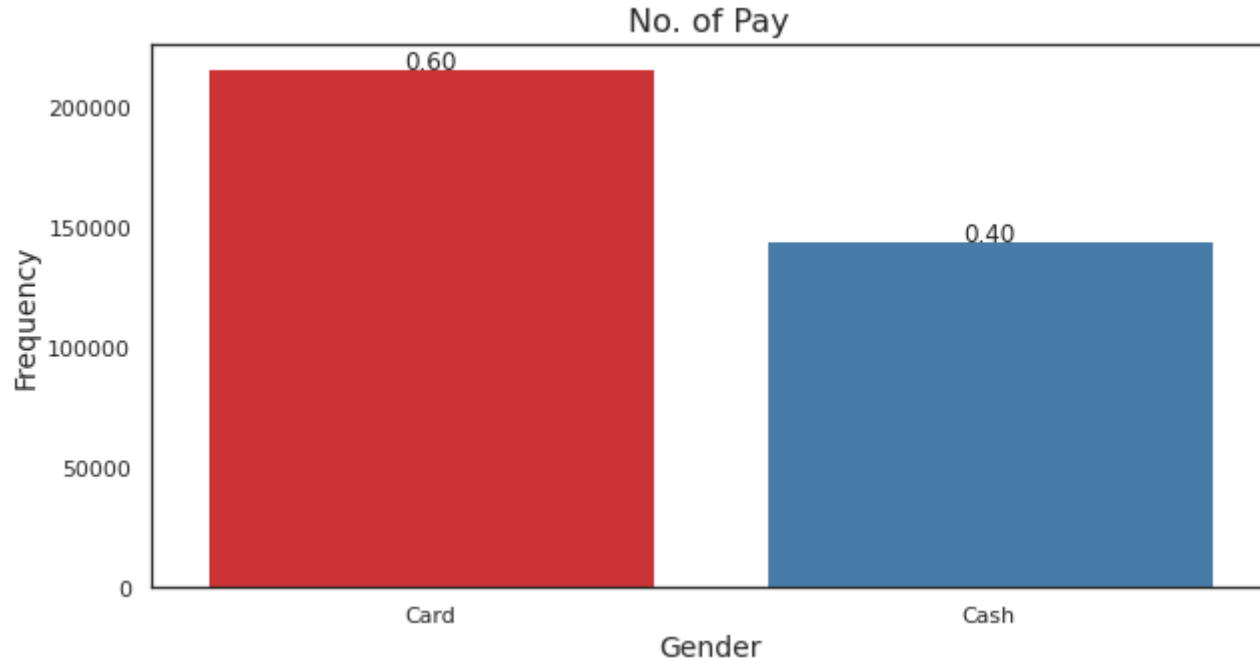
- Yellow cab service has been used more.
- The city that uses the service the most is New York, this service is the yellow cab.
- Pink cab has greater or equal use in cities far from the capital.

Gender



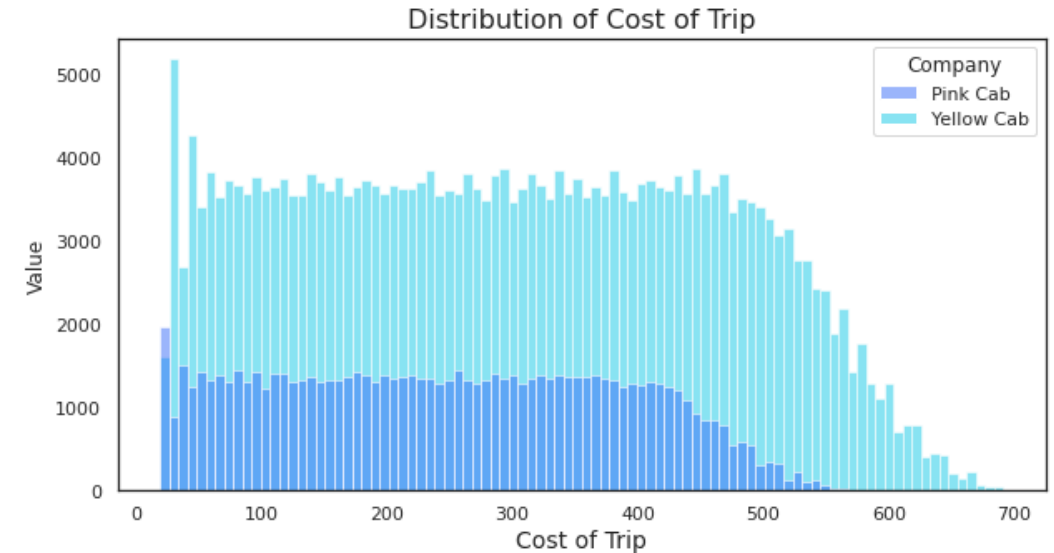
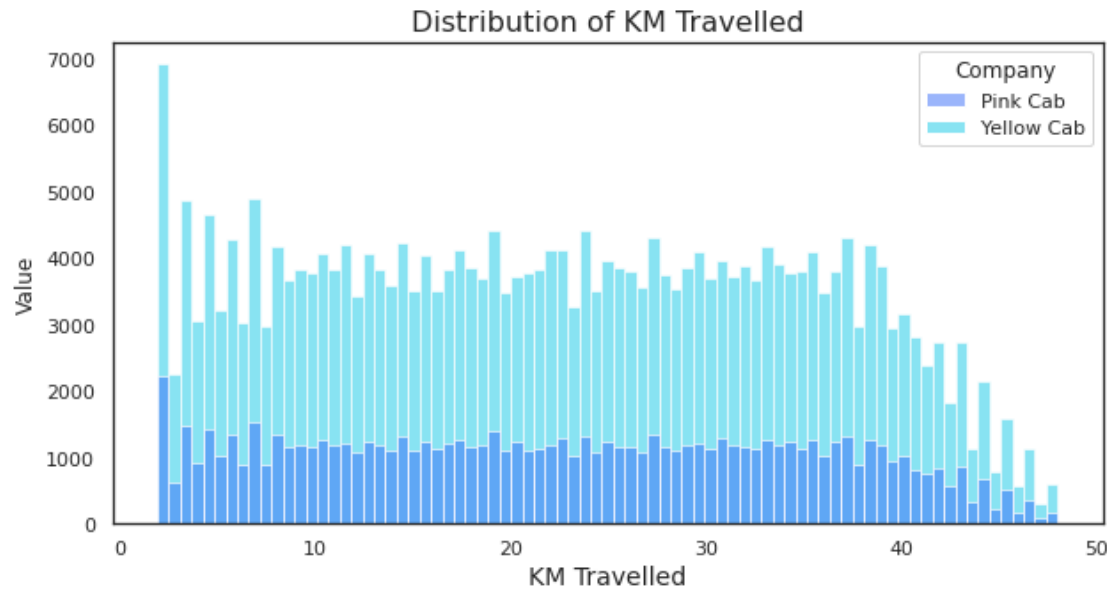
- In general, men use the service more, in both companies.
- The difference is not very big so it can be taken as comparable.
- Regardless of gender, yellow cab is most used.

Pay Mode

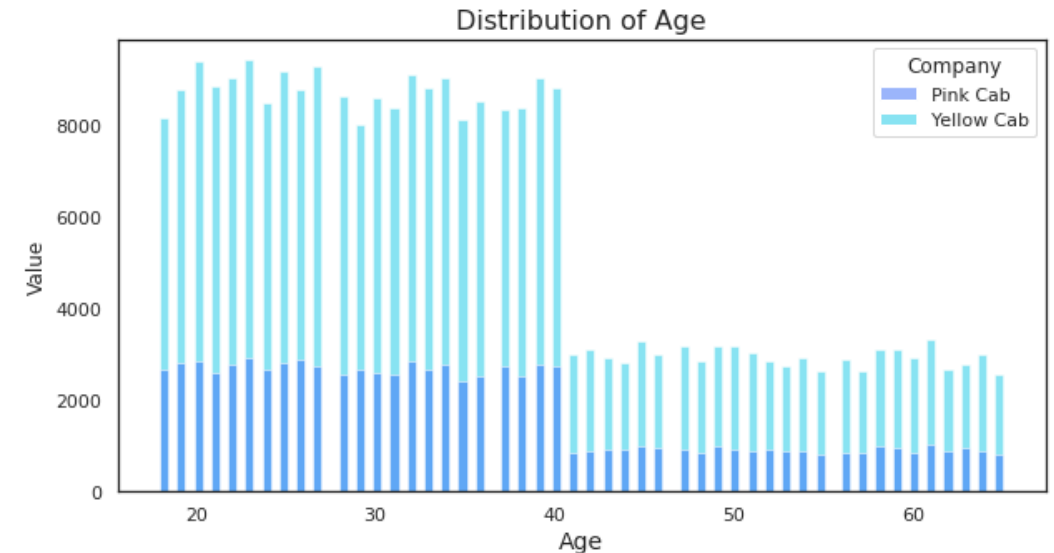


- The predominant form of payment is by card, 6 out of 10 people will pay by card.
- Extrapolating the values, the areas close to the capital have a tendency to pay by card and those far away with physical money.

Distribution

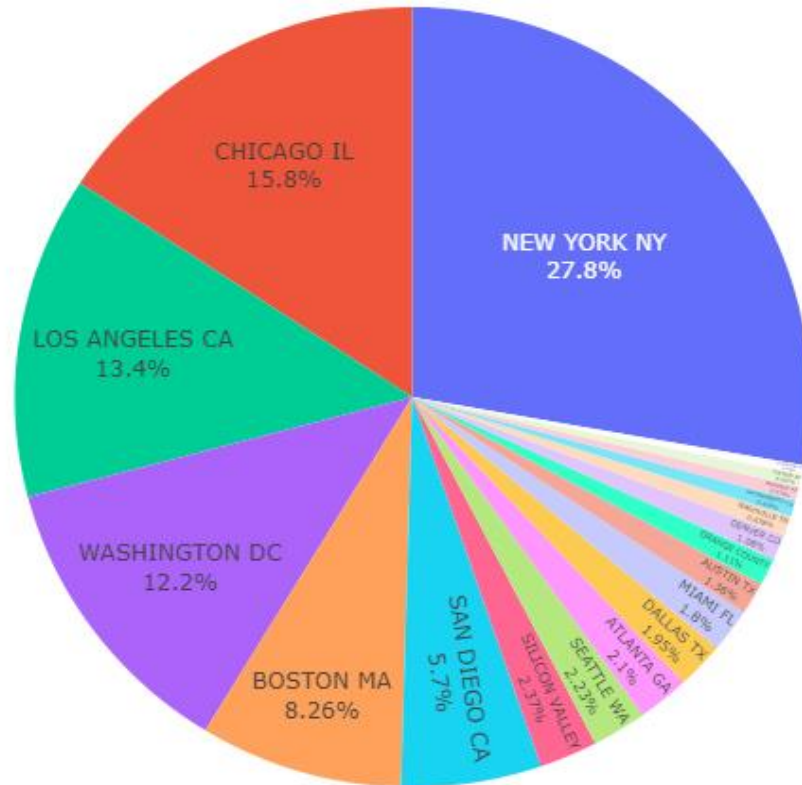


- Usually there are trips between 3 to 40 km, beyond which they are uncommon.
- The cost of the trip for yellow cab ranges from 20 to 500 usd.
- The trip cost for Pink cab ranges from 20 to 450 usd.
- The age range goes from 15 to 40 years, later values are half that of the youngest.

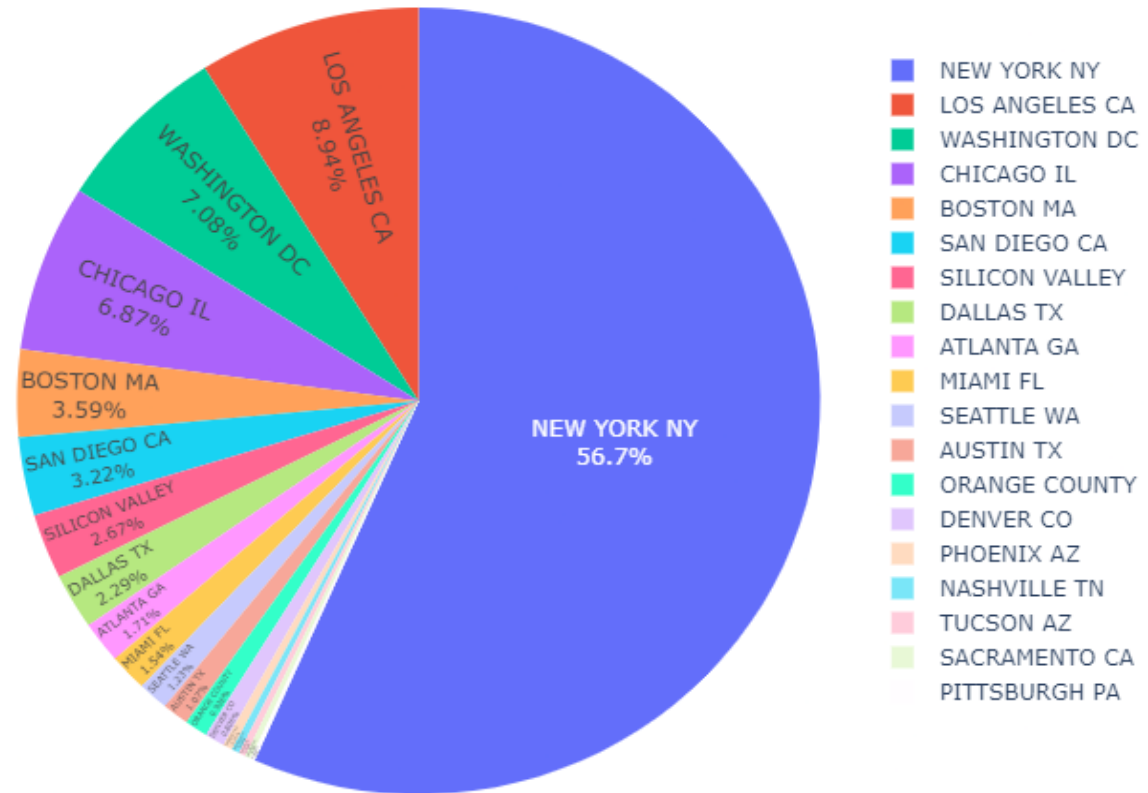


Profit by City/Users

Users by City



Profit by City

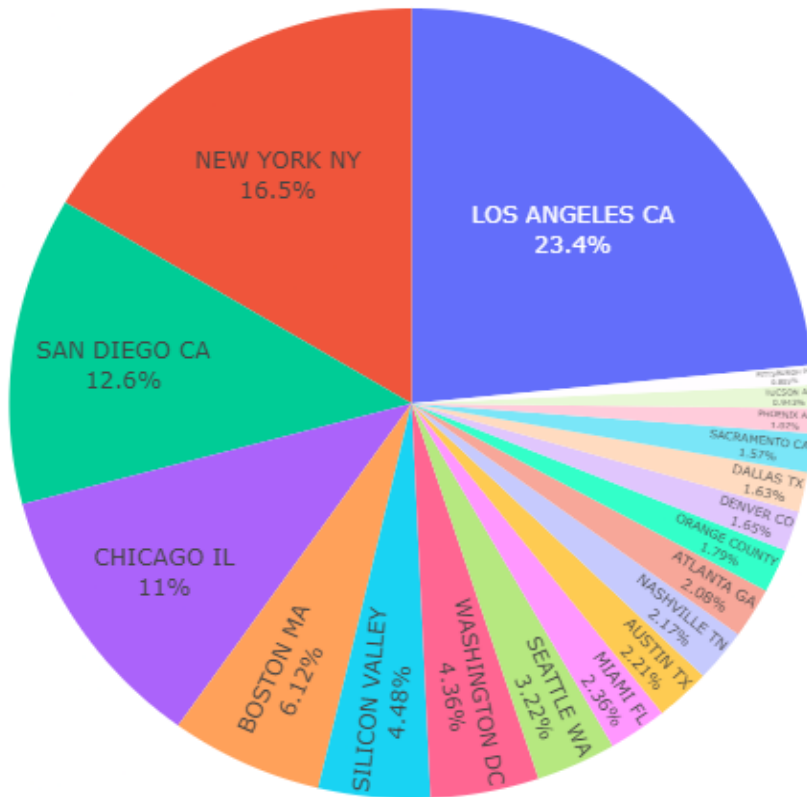


- NEW YORK NY
- LOS ANGELES CA
- WASHINGTON DC
- CHICAGO IL
- BOSTON MA
- SAN DIEGO CA
- SILICON VALLEY
- DALLAS TX
- ATLANTA GA
- MIAMI FL
- SEATTLE WA
- AUSTIN TX
- ORANGE COUNTY
- DENVER CO
- PHOENIX AZ
- NASHVILLE TN
- TUCSON AZ
- SACRAMENTO CA
- PITTSBURGH PA

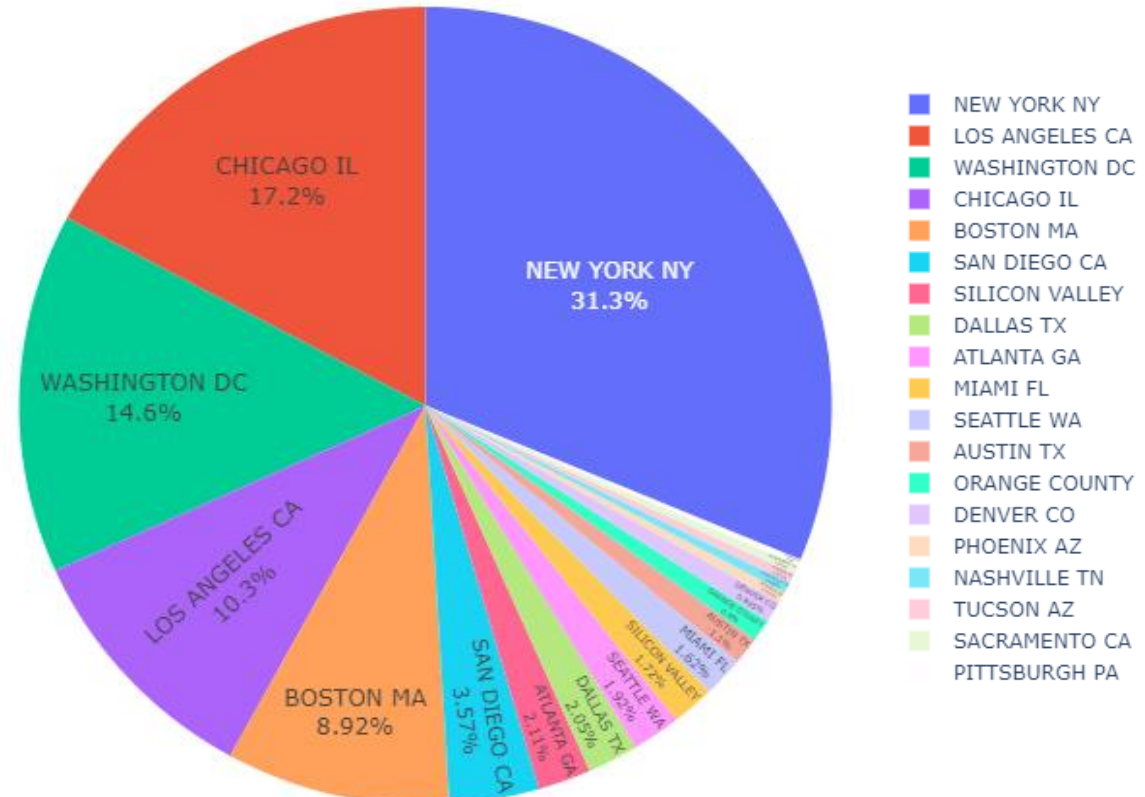
- New York has 28% of the market, it is the highest percentage among cities.
- Extrapolating the values, the highest profit with more than 50%, also belongs to the state of New York.

Transactions by Company

Transaction Pink Cab by City

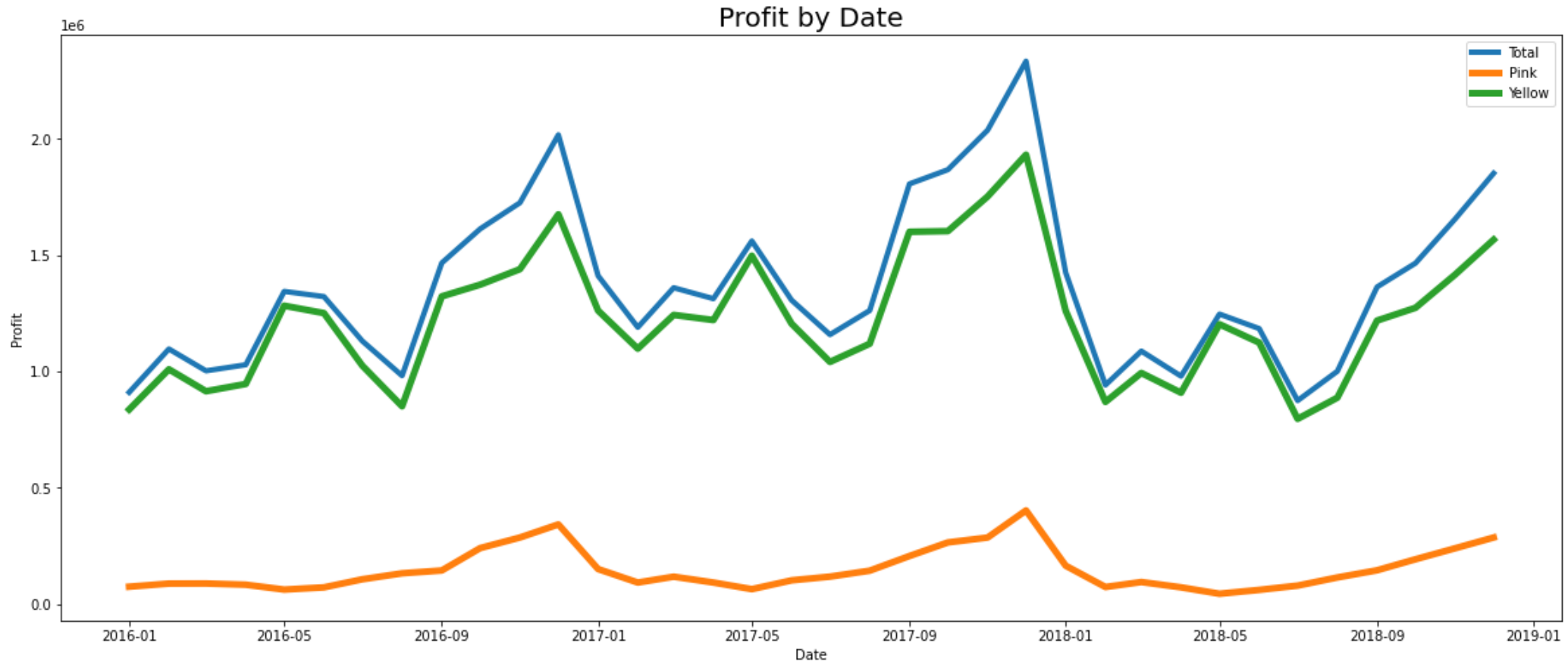


Transaction Yellow Cab by City



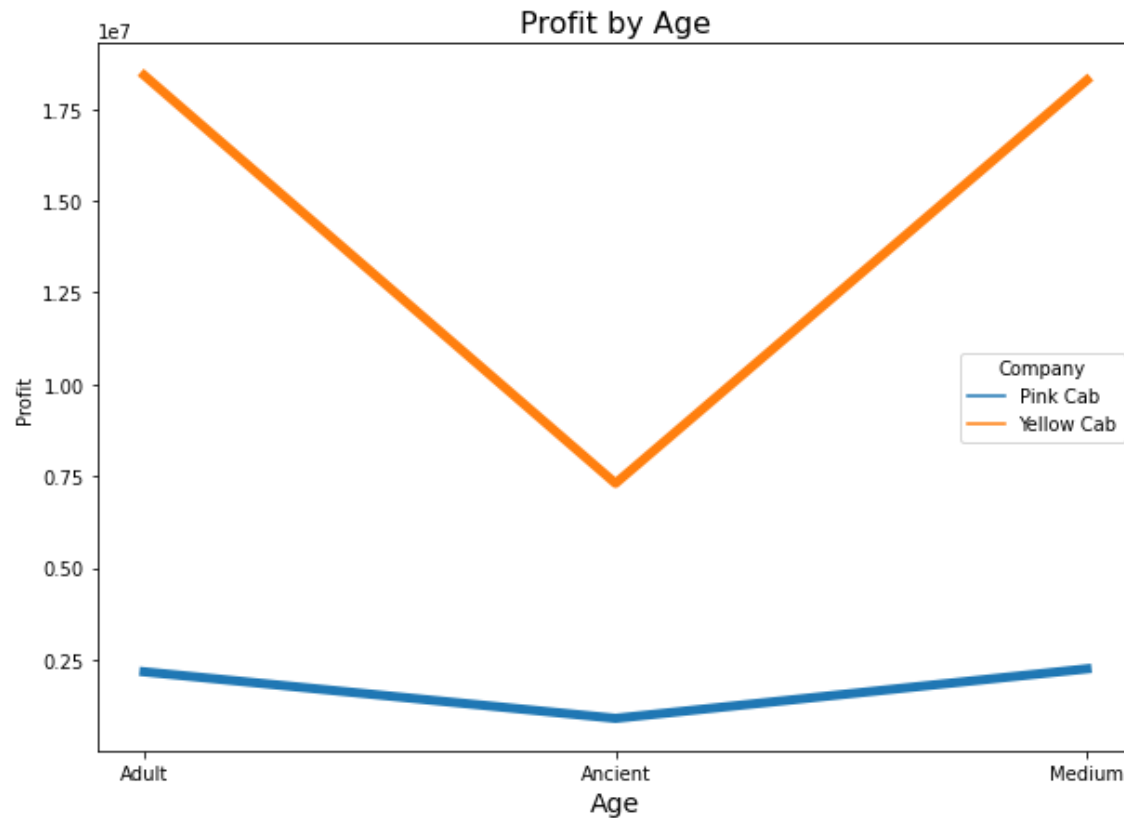
- The number of transactions carried out is in Los Angeles for the Pink Cab with 23.4% and New York for the Yellow Cab with 31.3%.
- Pink Cab is most prevalent in Los Angeles and San Diego, while Yellow Cab is most prevalent in New York.

Profit by Date

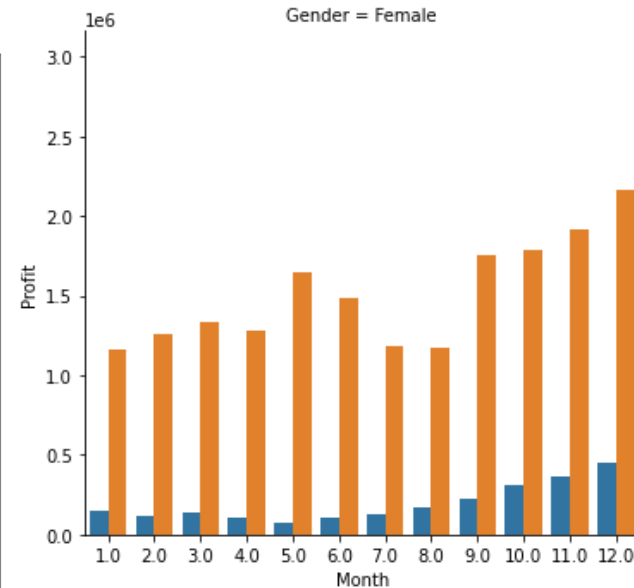


- Yellow Cab has higher gain throughout the timeline.
- The time series is seasonal, with higher gains at the end of the year and lower gains at the beginning.

Profit by Age/Gender

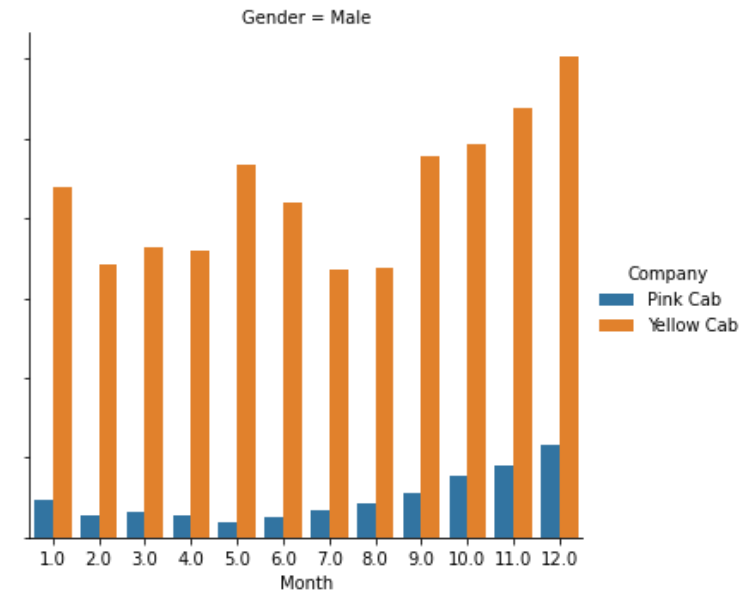


- There are higher profits with the group of people between 18 and 49 years old.

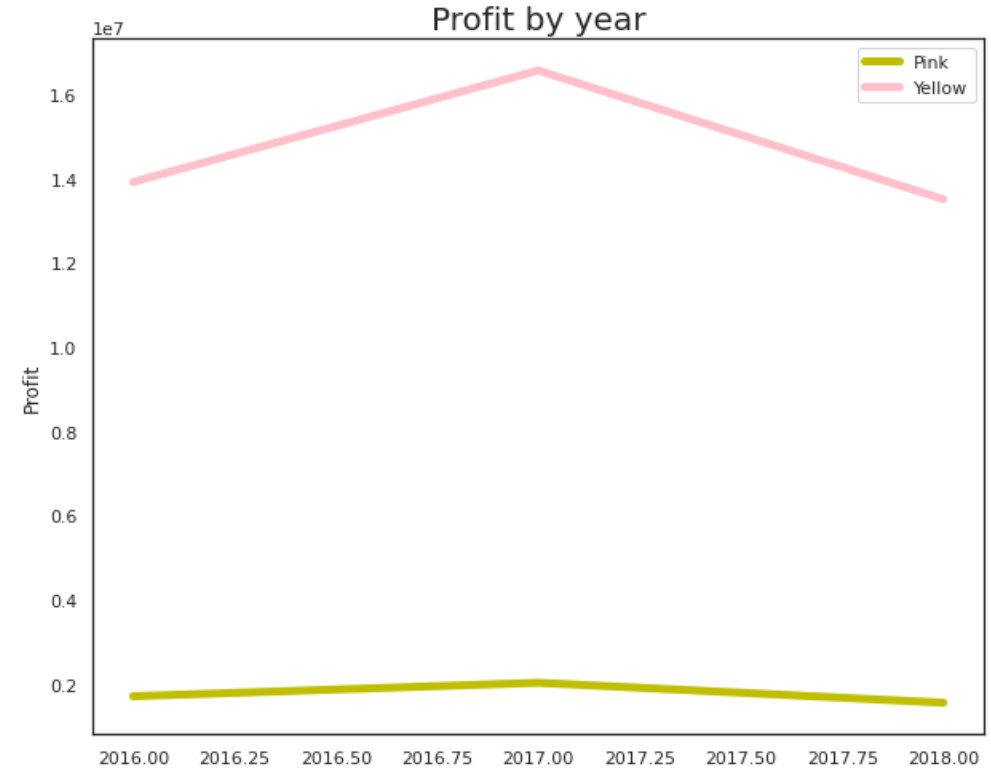


- A higher profit is obtained with the male market.

- Yellow cab has higher earnings in both genders in all months of the year.

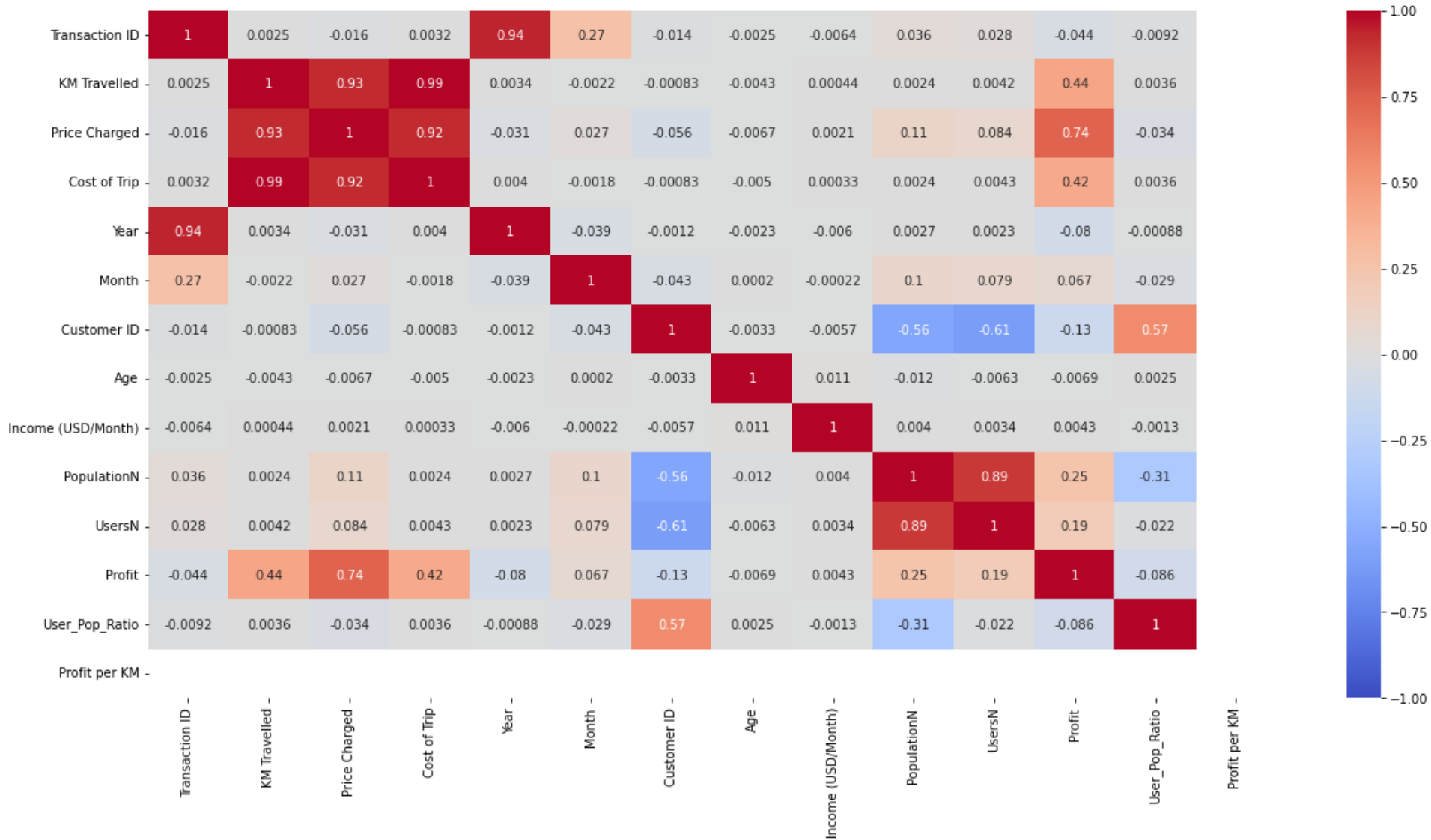


Profit by Month/Year



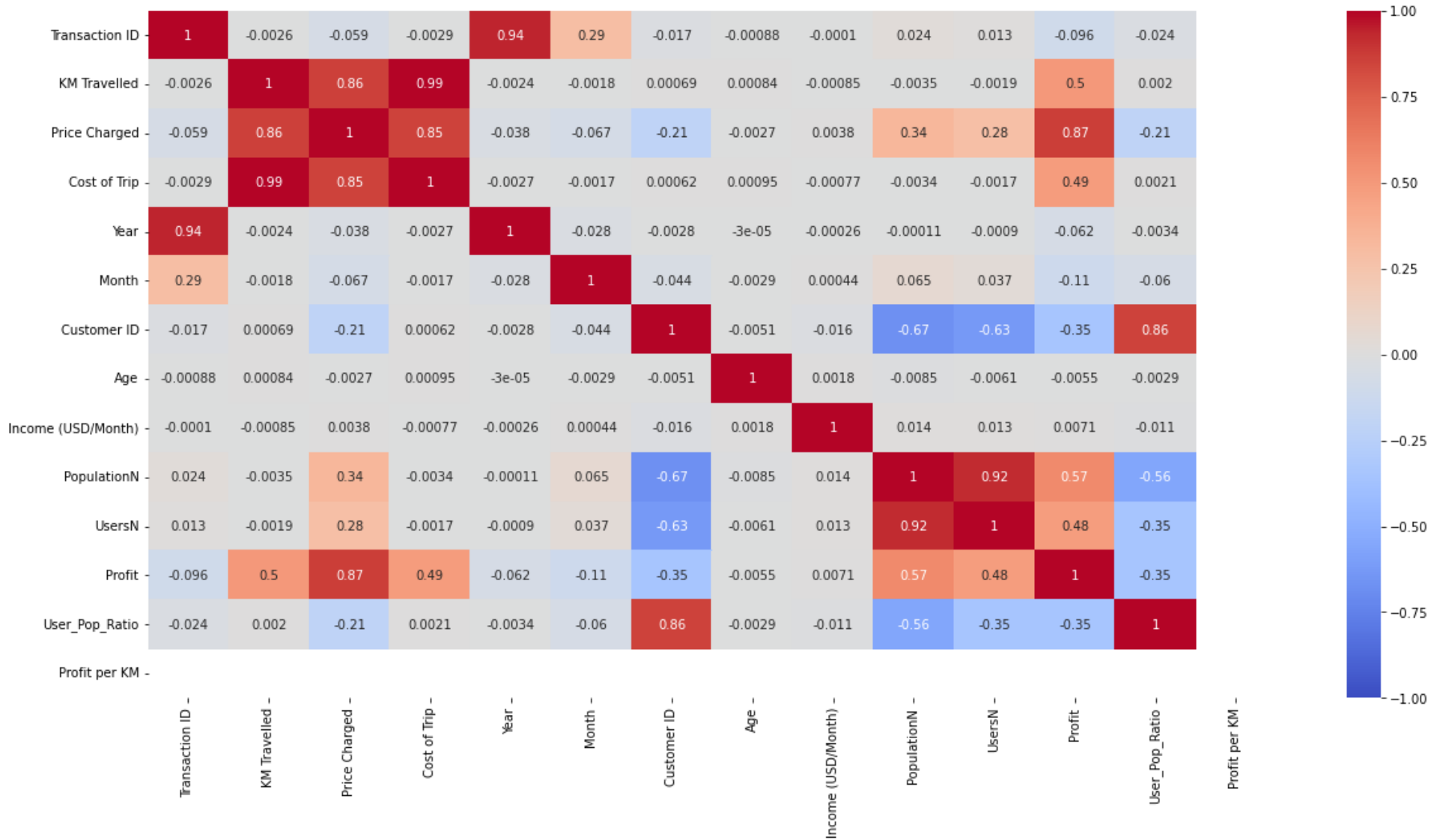
- The biggest profit of the year occurs in the last three months, possibly due to the year-end festivities.
- The lowest earnings are in the first months of the year, due to low economic activity and holidays.
- A higher profit was obtained in 2017, with an increase of 15%. And a decrease in it for the year 2018.

Pink Cab Correlation



➤ The best correlated variables are KM Travelled, Price Charged, Cost of Trip and Profit.

Yellow Cab Correlation



➤ The best correlated variables are KM Travelled, Price Charged, Cost of Trip, Population, Users and Profit.

Agenda

Background

Executive Summary

EDA

EDA Summary

Recommendations

EDA Summary

Yellow Cab	Pink Cab
Yellow cab service is most used in cities with active economies.	Pink cab service is more used in cities far from the capital.
KM Traveled goes from 2 to 48 KM.	KM Traveled goes from 2 to 48 KM.
Yellow cab is more prevalent in New York.	Pink cab is more predominant in Los Angeles.
The service is mainly used by men and not too far away by women.	The service is mainly used by men and not too far away by women.
It maintains a seasonality in earnings on the timeline.	It maintains a seasonality in earnings on the timeline.
The service is mainly used by people under 48 years of age.	The service is mainly used by people under 48 years of age.
Yellow cab had a high increase in its activities and earnings from 2016 to 2017.	Pink cab had a high increase in its activities and earnings from 2016 to 2017.

Hypothesis Testing

- ❖ Does margin remain the same regarding Transaction ID for both yellow cab & pink cab?
 - Pink Cab: We accept null hypothesis that there is no statistical difference
 $P_value = 0.7565393823450782$
 - Yellow Cab: We accept alternate hypothesis that there is a statistical difference
 $P_value = 1.4165376916211611e-08$
- ❖ Does margin remain the same regarding Age>50 for both yellow cab & pink cab?
 - Pink Cab: We accept null hypothesis that there is no statistical difference
 $P_value = 1.0$
 - Yellow Cab: We accept null hypothesis that there is no statistical difference
 $P_value = 1.0$
- ❖ Does margin remain the same regarding pay mode for both yellow cab & pink cab?
 - Pink Cab: We accept null hypothesis that there is no statistical difference
 $P_value = 0.19433870202790374$
 - Yellow Cab: We accept alternate hypothesis that there is a statistical difference
 $P_value = 0.003826321021337554$

Agenda

Background

Executive Summary

EDA

EDA Summary

Recommendations

Recommendations

- Customer Reach : Yellow Cab has a greater range of customers, while the Pink Cab has a shorter range.
- Age wise Reach : Yellow Cab has customers in all age groups and has been found to be popular with the 25-40 age group, and to a lesser extent with the 45+ age group.
- Average Profit per KM: The average gain per KM of the Yellow Cab is almost three times the average gain per KM of the Pink Cab.
- Transaction per year: For Yellow Cab Transaction per year from 2016 to 2018 it is almost twice as much as Pink Cab.
- Margin per Pay Mode: For Yellow Cab, there is a difference in the margin between card payments and cash payments because the percentage of card payments is higher in Yellow Cab compared to Pink Cab.
- Profit Margin: For Yellow Cab, the profit margin is higher per year from 2016 to 2018 compared to Pink Cab.
- Margin per Age: In Yellow Cab there is a difference in the Margin for people over 50 years of age, while in Pink Can there is no difference in the Margin for all age groups.

Thank You