



Data Glacier

Your Deep Learning Partner

G2M Insight for Cab Investment Firm

Virtual Internship

Janesh Hasija

31-July-2024

Agenda

- **Problem Statement**
- **Data Information**
- **Exploratory Data Analysis (EDA)**
- **Hypothesis tests**
- **Recommendations**

Problem Statement

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 Understanding
- Forecasting profit and number of rides for each cab type
- Finding the most profitable Cab company
- Recommendations for investment

Data Information

The dataset contains four different csv files. These datasets were combined in the 'all_data.csv' file after reviewing.

Cab_Data.csv : This file includes details of transactions for 2 cab companies (9 features).

Customer_ID.csv : This is a mapping table that contains a unique identifier that links the customer's demographic details (4 features).

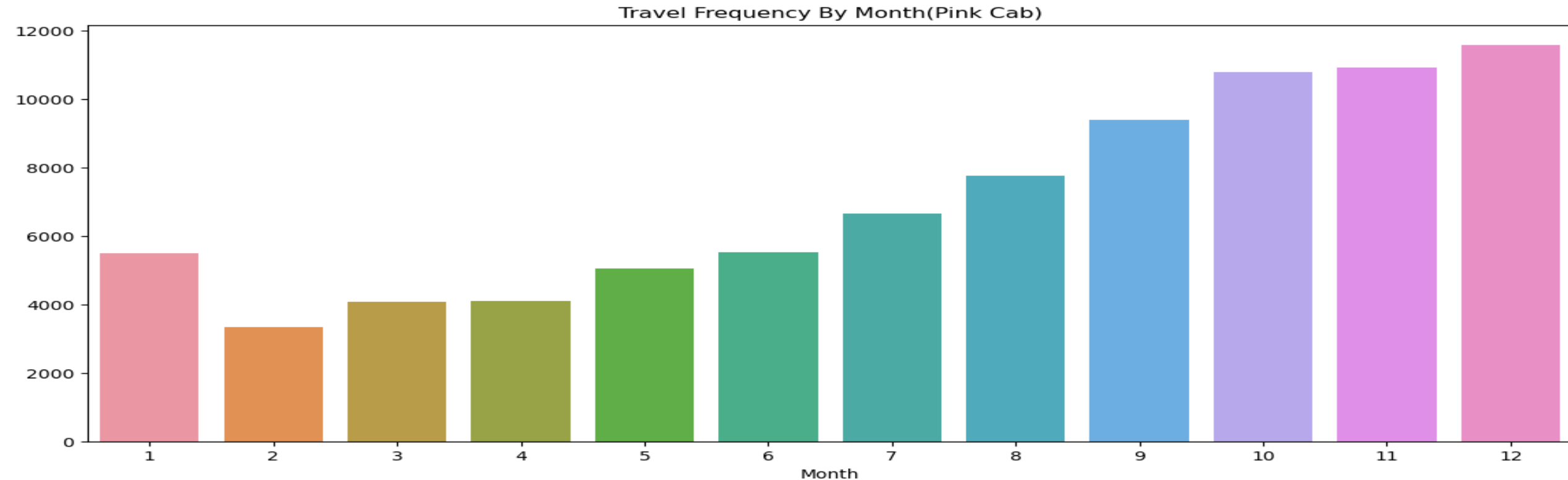
Transaction_ID.csv : This is a mapping table that contains transaction to customer mapping and payment mode (3 features).

City.csv: This file contains a list of US cities, their population, and the number of cab users (3 features).

All_data.csv: This file is a combination of 4 files. The dataset is combined based on the "Customer ID" and "Transaction ID" features. Also, new features such as profit, state have been added to the dataset (16 features).

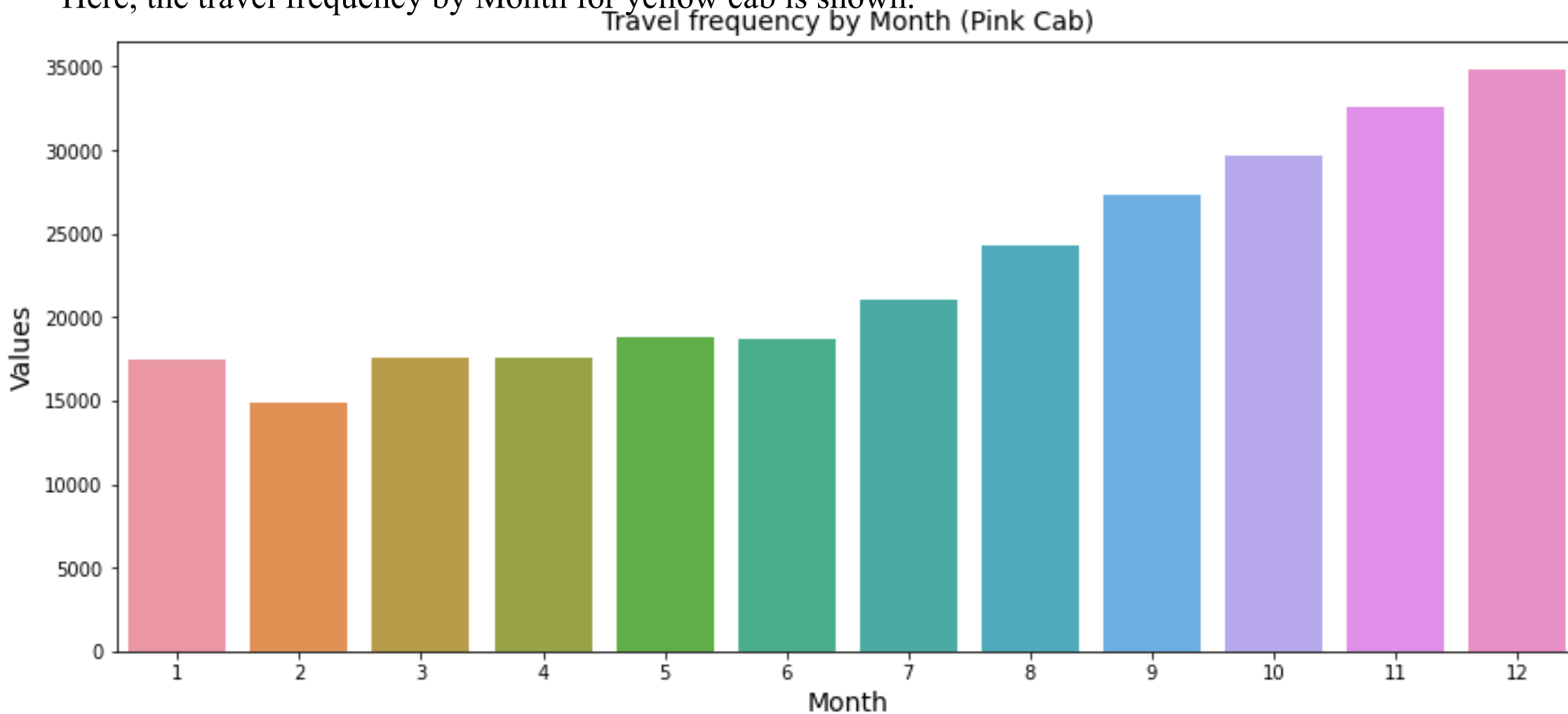
Exploratory Data Analysis (EDA)

Here, the travel frequency by Month for pink cab is shown.



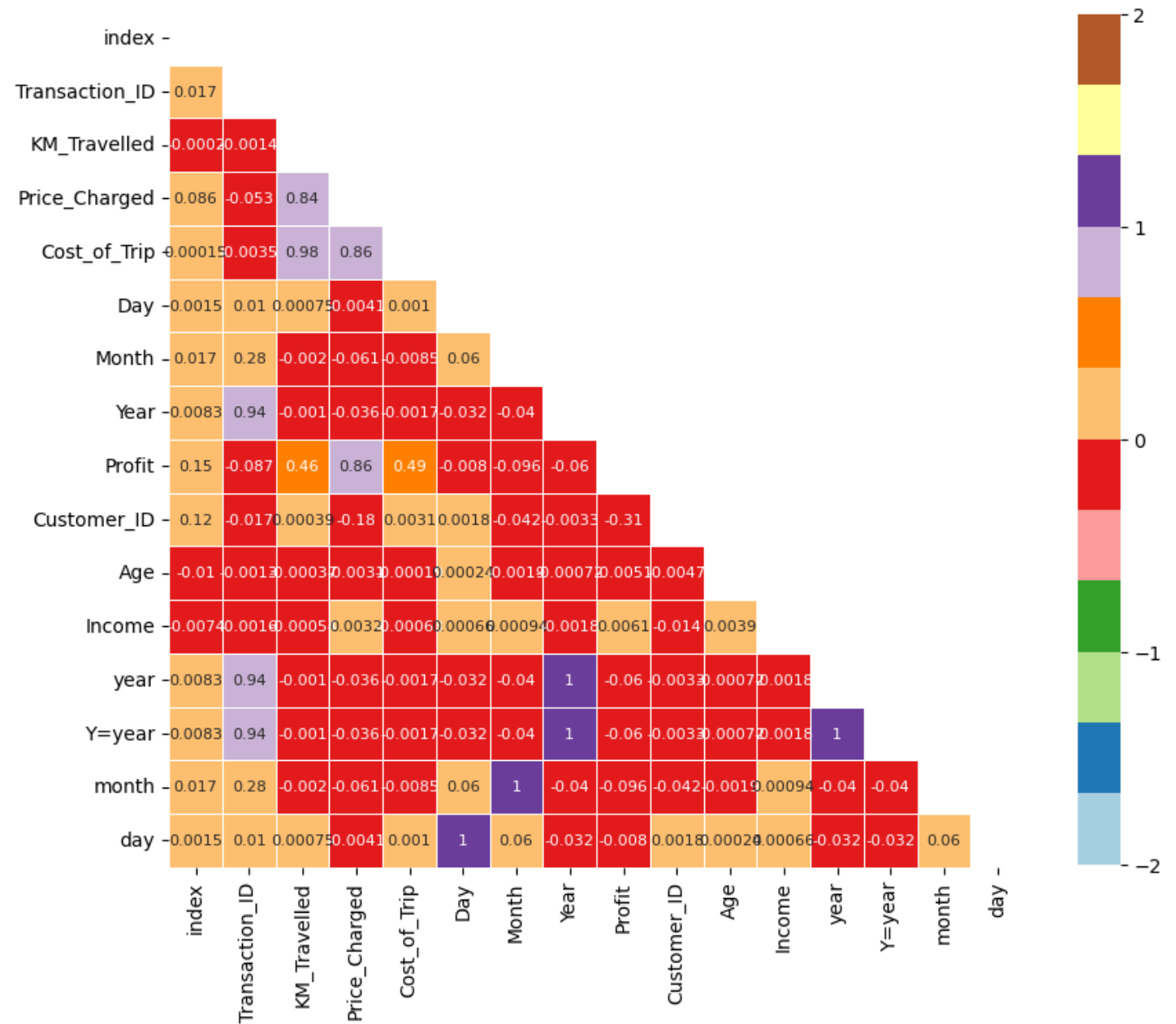
Exploratory Data Analysis (EDA)

Here, the travel frequency by Month for yellow cab is shown.



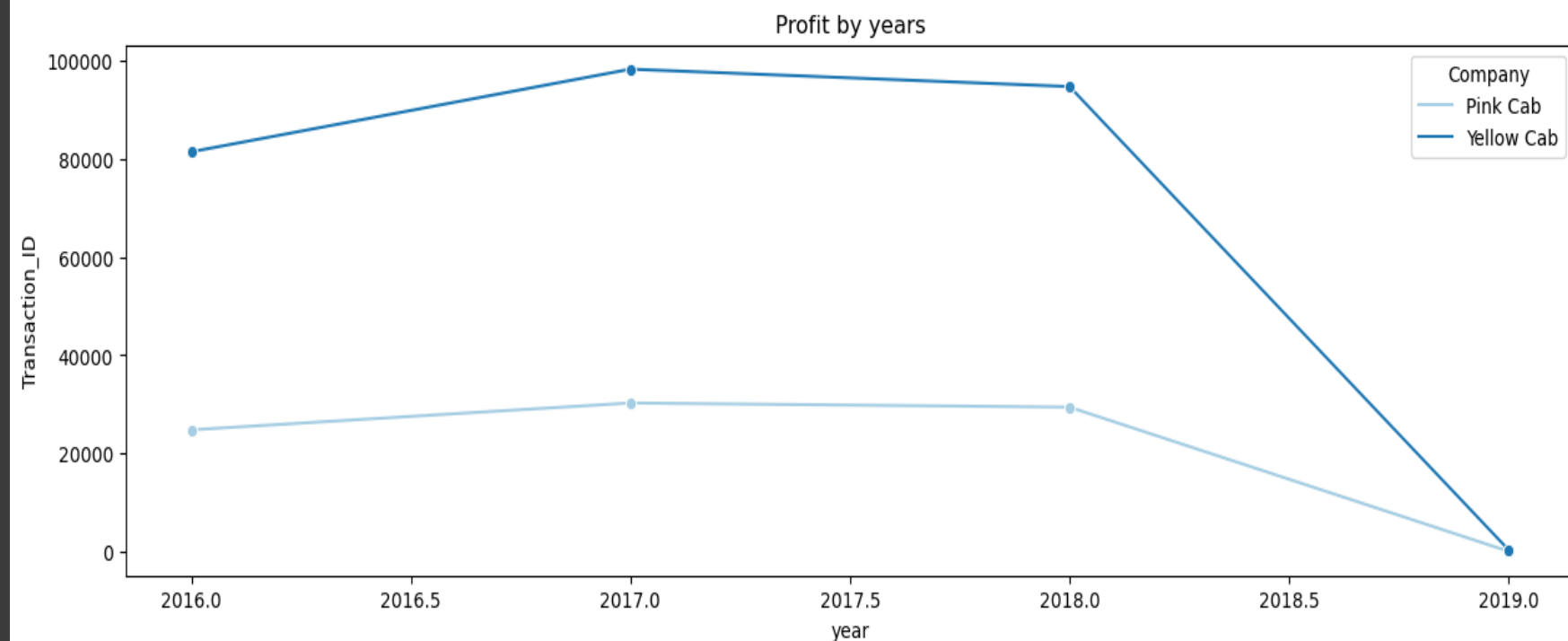
Exploratory Data Analysis (EDA)

- The two features with the strongest correlation in the dataset are 'KM_Travelled' and 'Cost of Trip' with 0.98.
- It is seen that there is a strong positive correlation with 0.92 correlation between 'Year' and 'Transaction_ID' features.



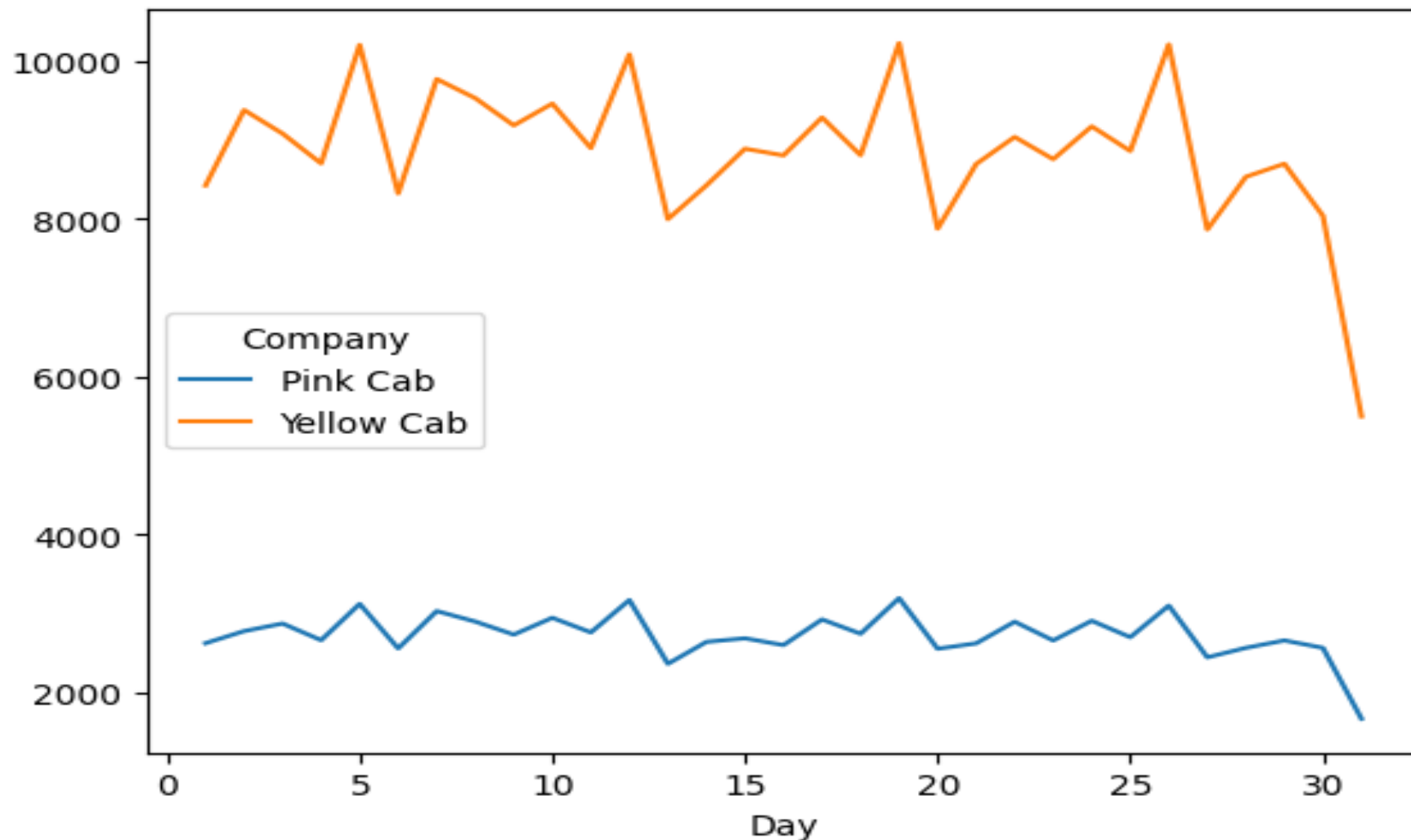
Exploratory Data Analysis (EDA)

- Profits of cab companies on a monthly and annual basis



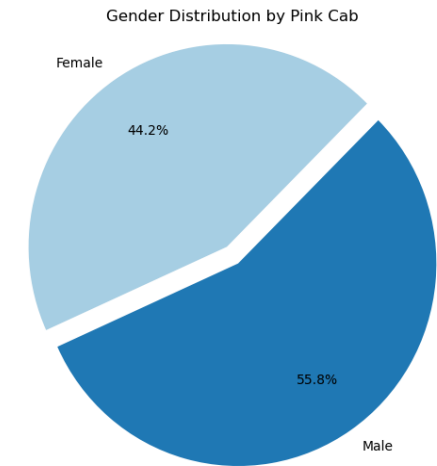
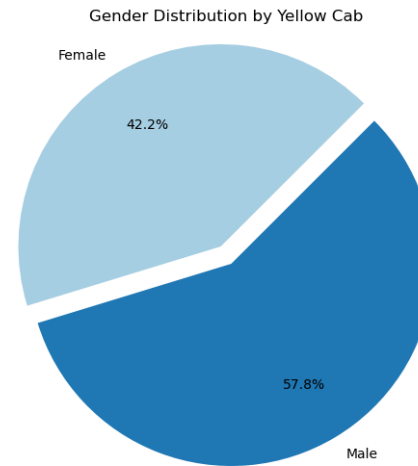
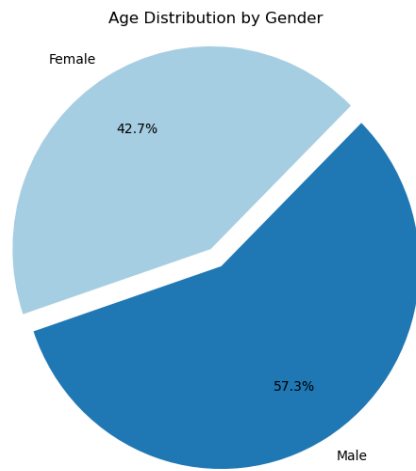
Exploratory Data Analysis (EDA)

It is seen that the demand is higher on the third day of the week for both companies.



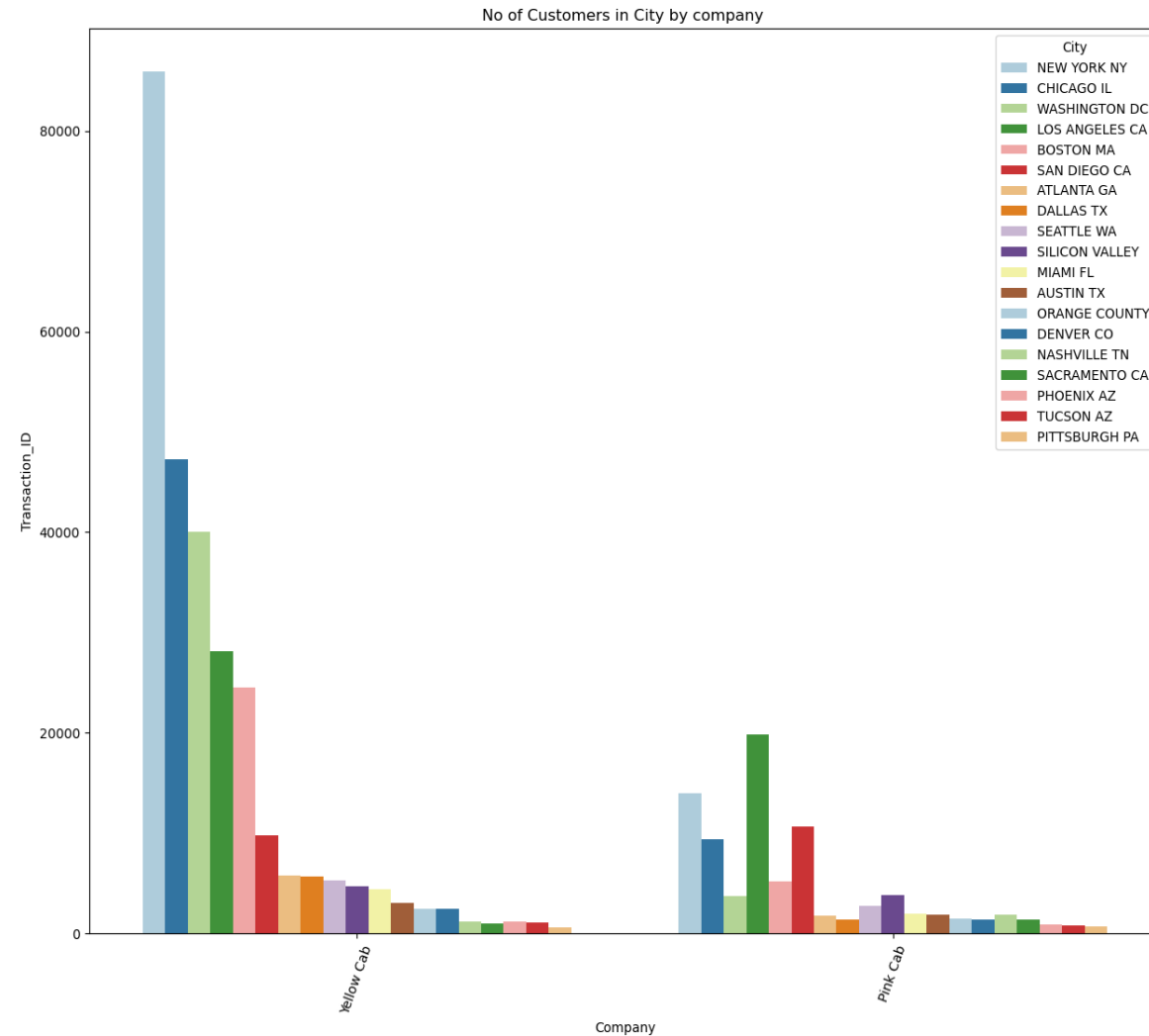
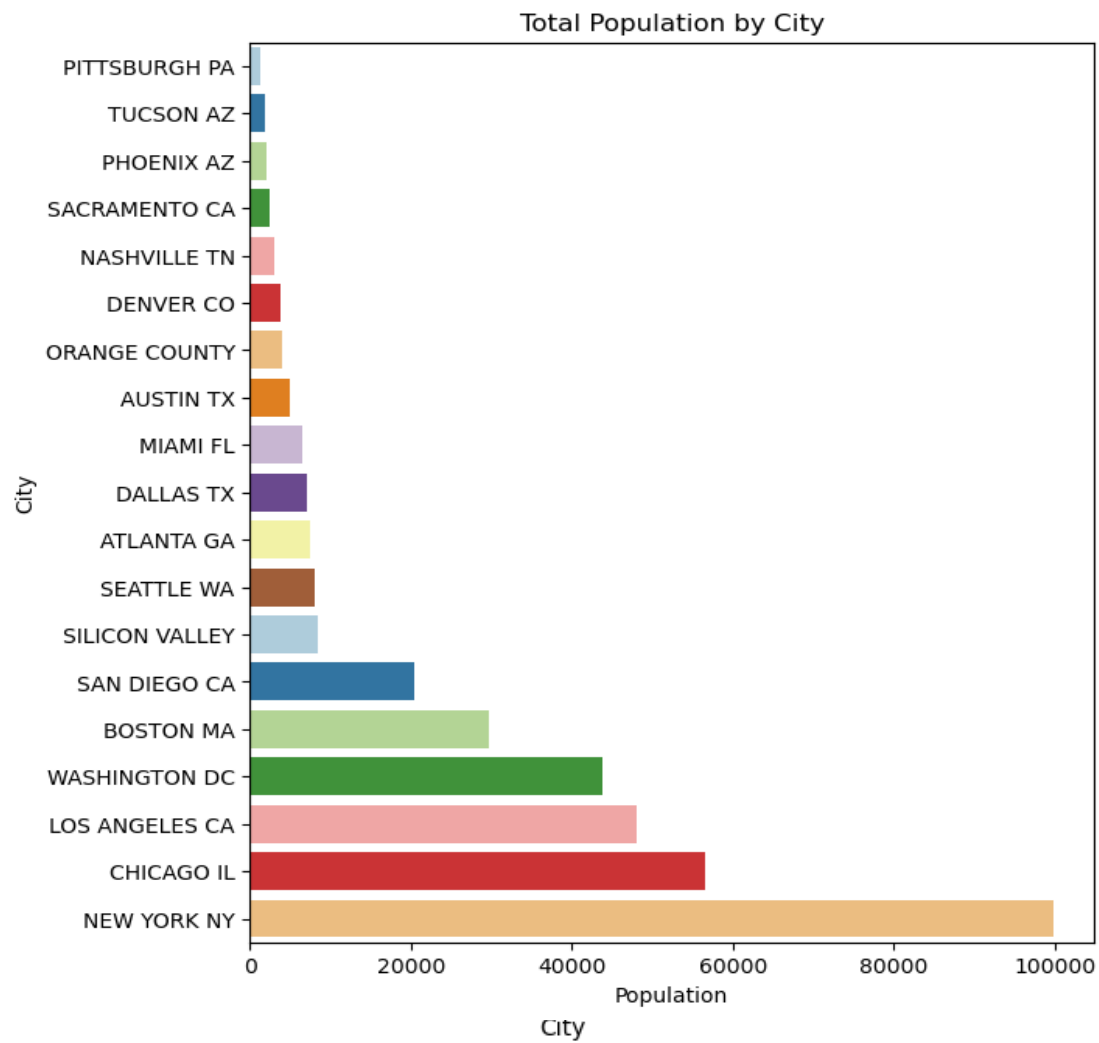
Exploratory Data Analysis (EDA)

Gender distribution by cab company.



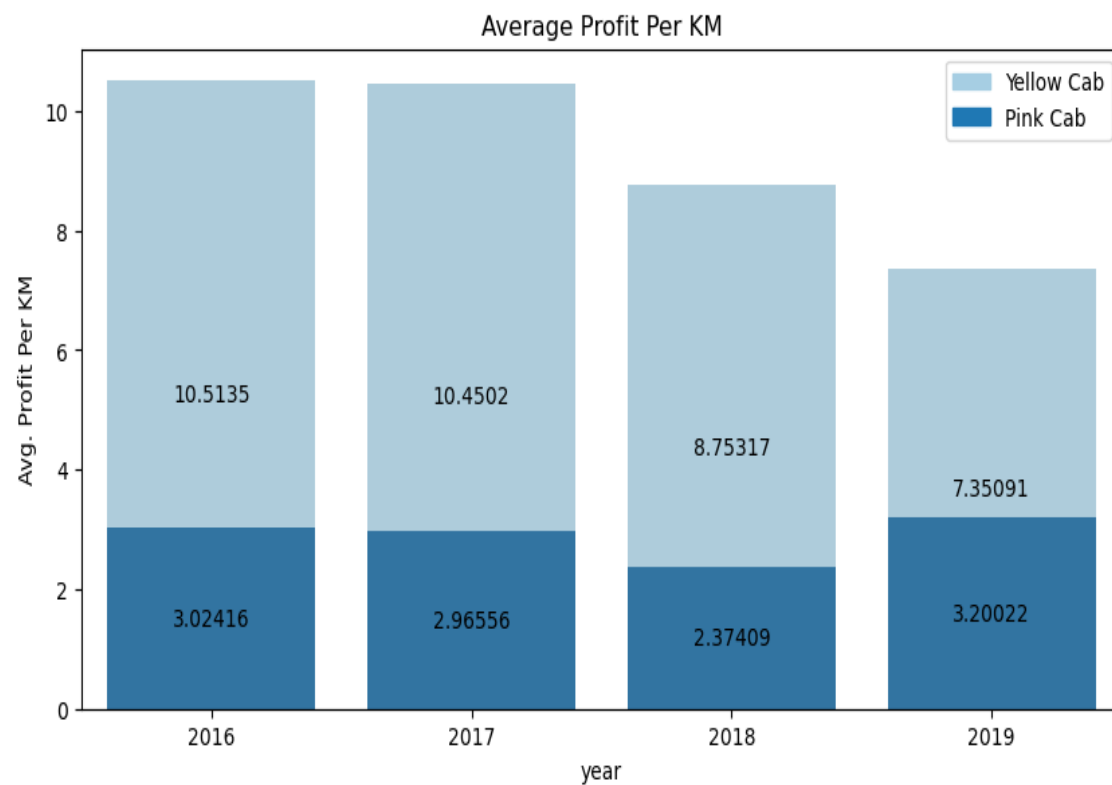
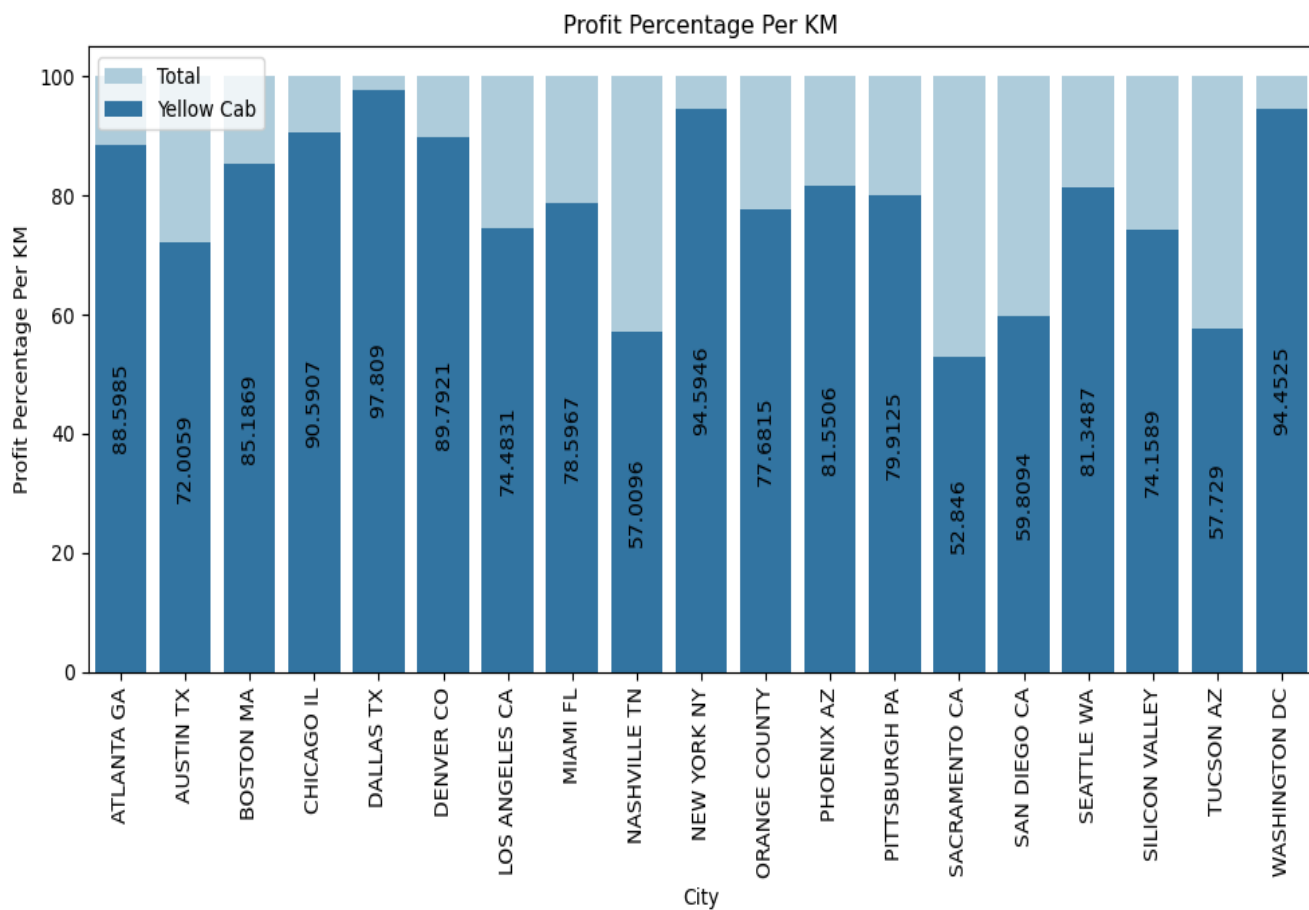
Exploratory Data Analysis (EDA)

P



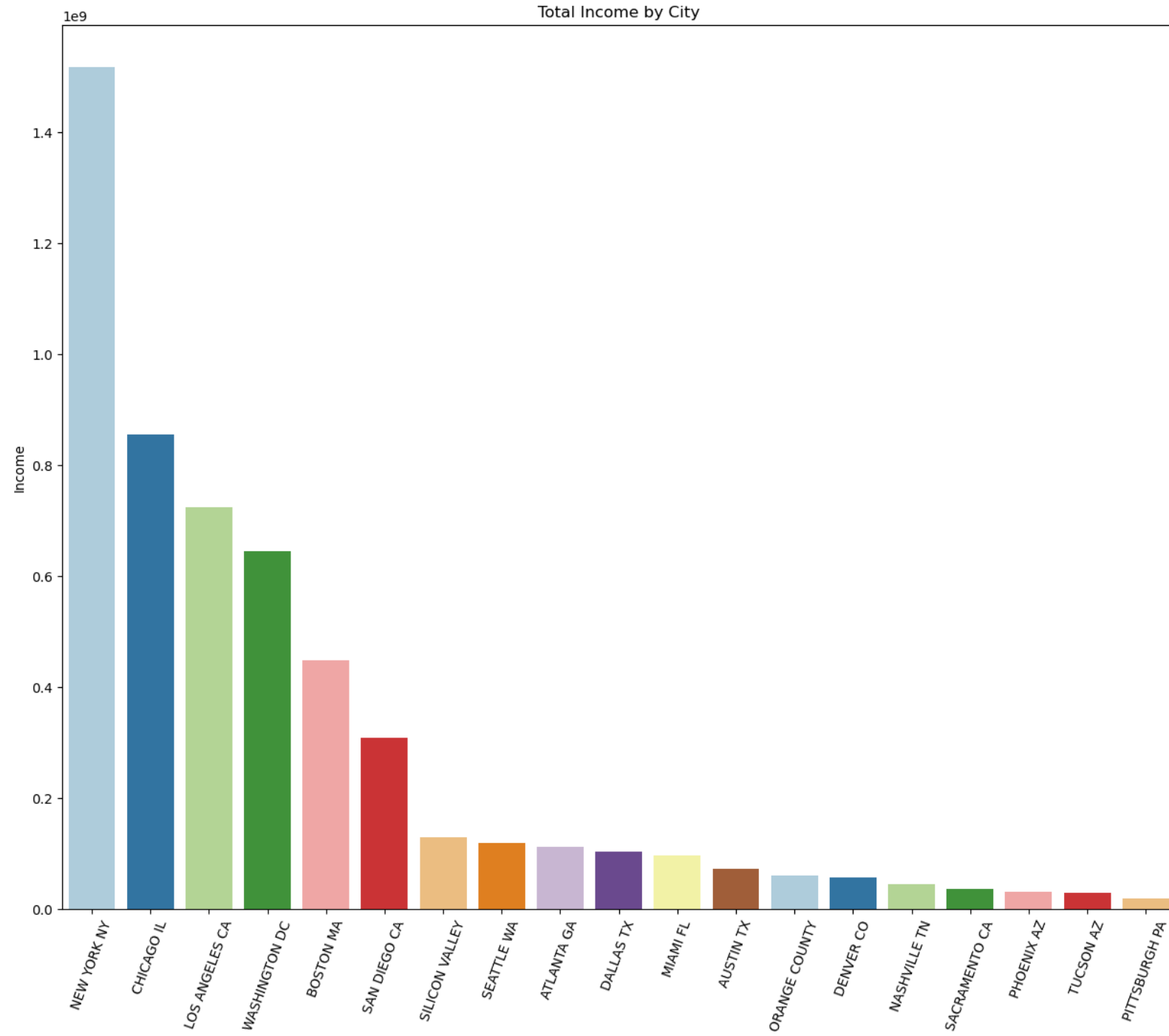
Exploratory Data Analysis (EDA)

Profit distribution of cab companies by cities and years.



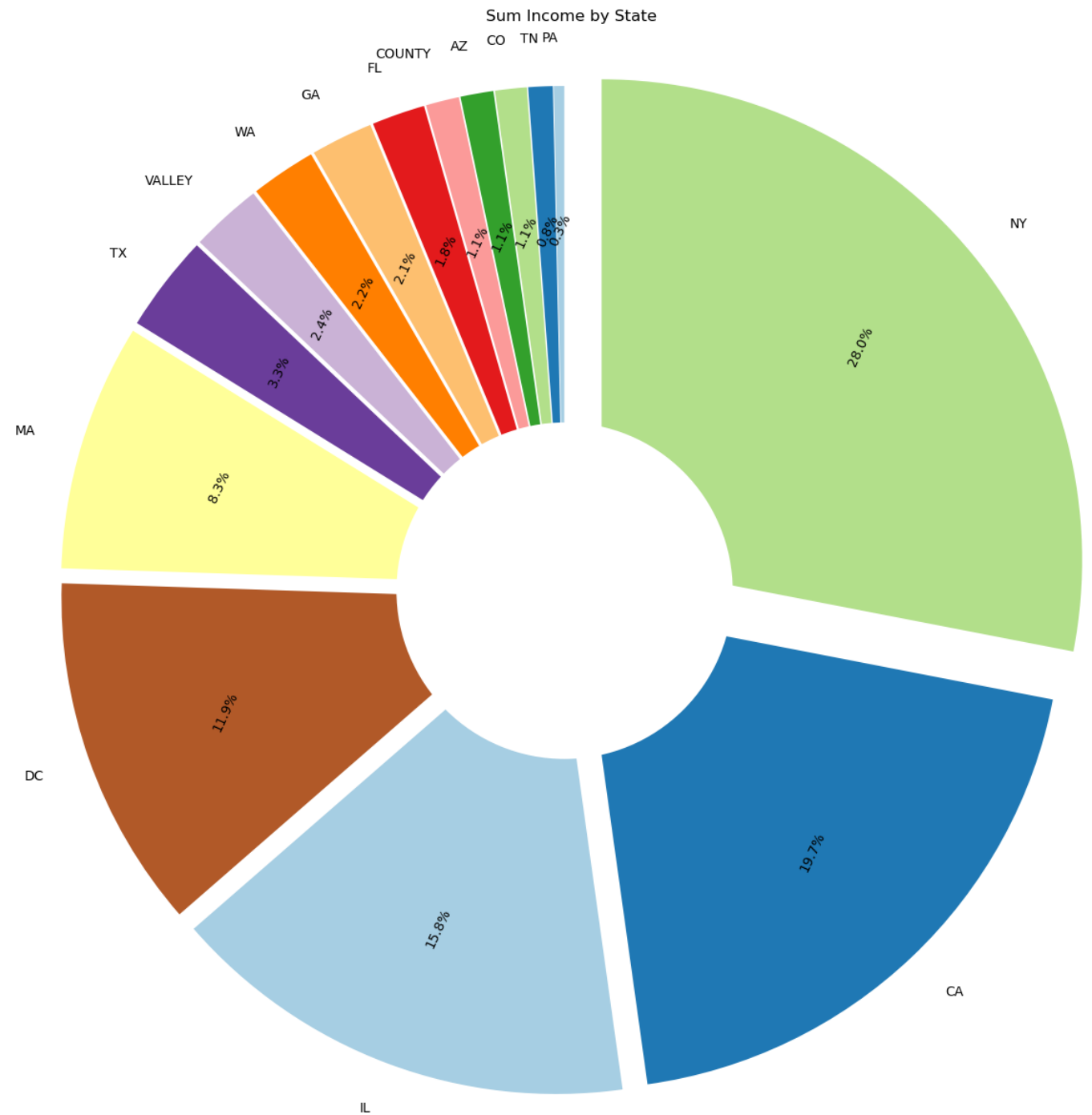
Exploratory Data Analysis (EDA)

- Maximum total income is new york city.



Exploratory Data Analysis (EDA)

- Income distribution by state.



Hypothesis tests

Hypothesis 1:

H0:Gender has not effect on company profit

H1:Gender has effect on company profit

'Pink Cab'

t_statistics: [0.02786795]

p_value: [0.98029821]

Reject Alternative Hypothesis(H1)

'Yellow Cab'

t_statistics: [-0.08386648]

p_value: [0.94080144]

Reject Alternative Hypothesis(H1)

Hypothesis 2:

H0:Payment mode has not effect on company profit

H1:Payment mode has effect on company profit

'Pink Cab'

t_statistics: [-0.03834307]

p_value: [0.97289731]

Reject Alternative Hypothesis(H1)

'Yellow Cab'

t_statistics: [0.17875066]

p_value: [0.8746019]

Reject Alternative Hypothesis(H1)

Hypothesis tests

Hypothesis 3:

H0:The variable Year has not a positive correlation with mileage

H1:The variable Year has a positive correlation with mileage

'Pink Cab'

t_statistics: [-0.66243408]

p_value: [0.57581736]

Reject Alternative Hypothesis(H1)

'Yellow Cab'

t_statistics: [-0.42772975]

p_value: [0.71050085]

Reject Alternative Hypothesis(H1)

Hypothesis 4:

H0:Age has not effect on company profit

H1:Age has effect on company profit

'Pink Cab'

t_statistics: [0.13340069]

p_value: [0.90608835]

Reject Alternative Hypothesis(H1)

'Yellow Cab'

t_statistics: [-0.50355161]

p_value: [0.66456451]

Reject Alternative Hypothesis(H1)

Hypothesis tests

Hypothesis 5 :

H0:The variable Income has not a positive correlation with mileage

H1:The variable Income has a positive correlation with mileage

```
t_statistics: [-0.75691848]
p_value: [0.52811574]
Reject Alternative Hypothesis(H1)
t_statistics: [0.19035288]
p_value: [0.86660314]
Reject Alternative Hypothesis(H1)
```

Recommendations

As a result, it is seen that the yellow cab is better than the pink cab.

When both cab companies are compared, it is seen that the yellow cab is much more successful

- average profit per km,
- user,
- income,
- customer retention,
- customer access.

Thank You