
- G2M Case Study
- Virtual Internship

- **20-Sept-2022**

# Background –G2M(cab industry) case study

**Problem 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

**Objective**: Draw an analysis based on the data sets given to to help XYZ company to decide which Cab company would be a better investment.

**Strategy**:

- Understanding the data provided.
- Visualizing the provided data to get a better understanding.
- Finding the company which will provide most profit to the XYZ company
- Investment recommendation

# Understanding the Data

- 24 Features( including 9 derived features)
- Timeframe of the data: 2016-01-31 to 2018-12-31
- Total data points :355,032
- I combines the Transaction_ID, Cab_Data, Customer_ID into one dataset and City as a different dataset.

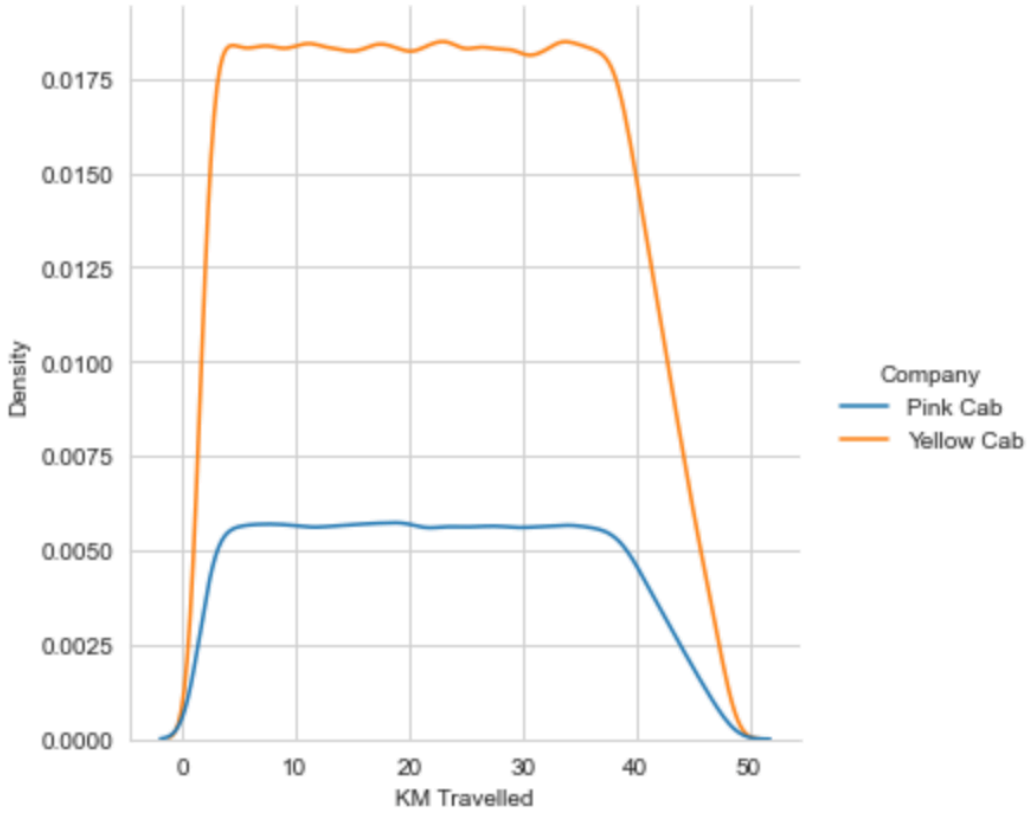
- Cab_Data.csv
- Transaction_ID.csv
- Customer_ID.csv
- City.csv

- **I made a few assumptions while analyzing this data which are as follows:**
- The cab users in the city dataset does not just use the yellow and pink cab company's cabs there are other cab companies too in these cities.
- A few outliers are ignored, as the complete information is missing.

- **Approach:**
- My approach for the Exploratory Data Analysis was to first visualize each feature and then try to analyze the data I have collected and draw meaningful insights.

# Analyzing different features of the given data

## 1. Analysis of Kilo Meters Travelled



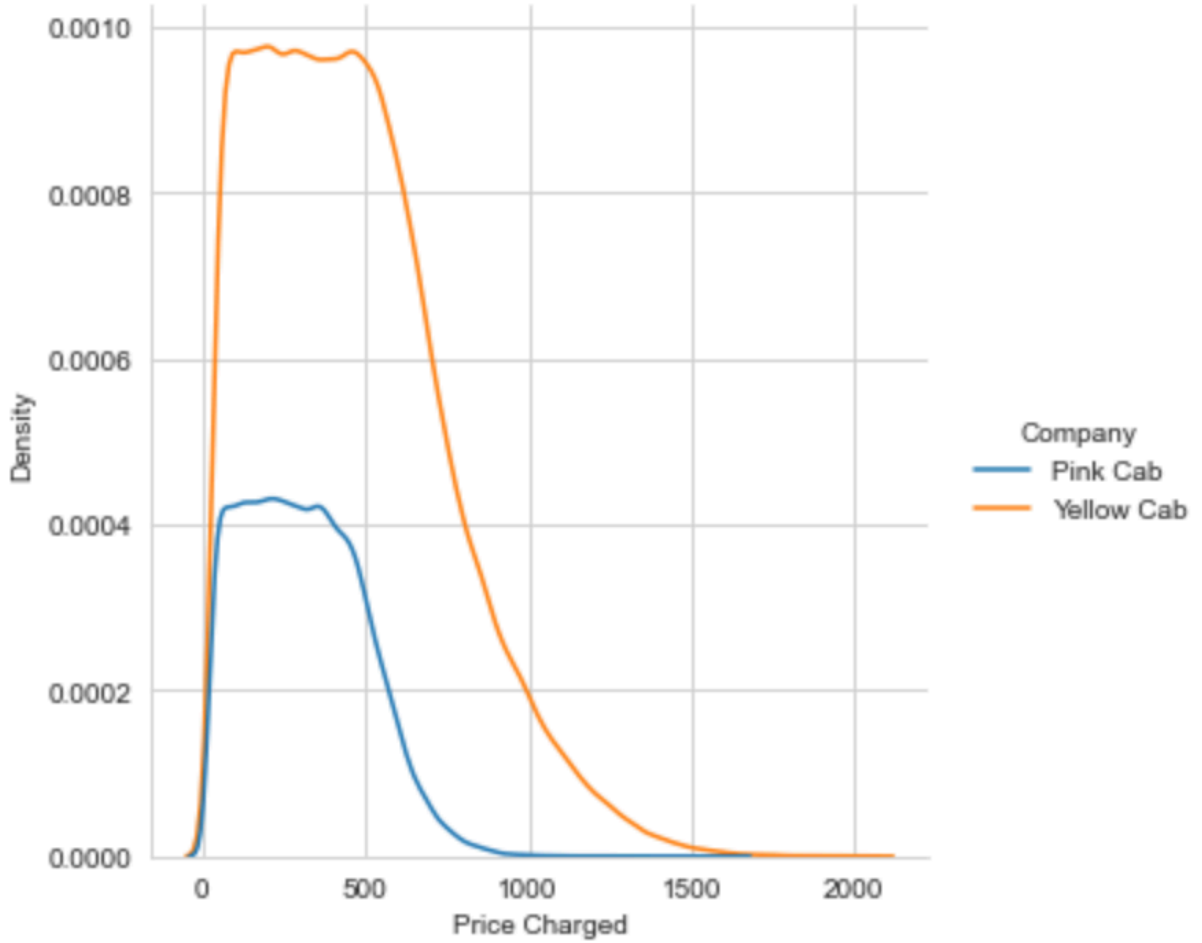

- From the adjacent Probability Density Function we can easily determine that distance travelled by Yellow Cabs are far more than the Pink Cabs. However, the information is not enough to draw any conclusions yet.

- KM Travelled By each Cab

# Analyzing different features of the given data

## 2. Analysis of Price Charged

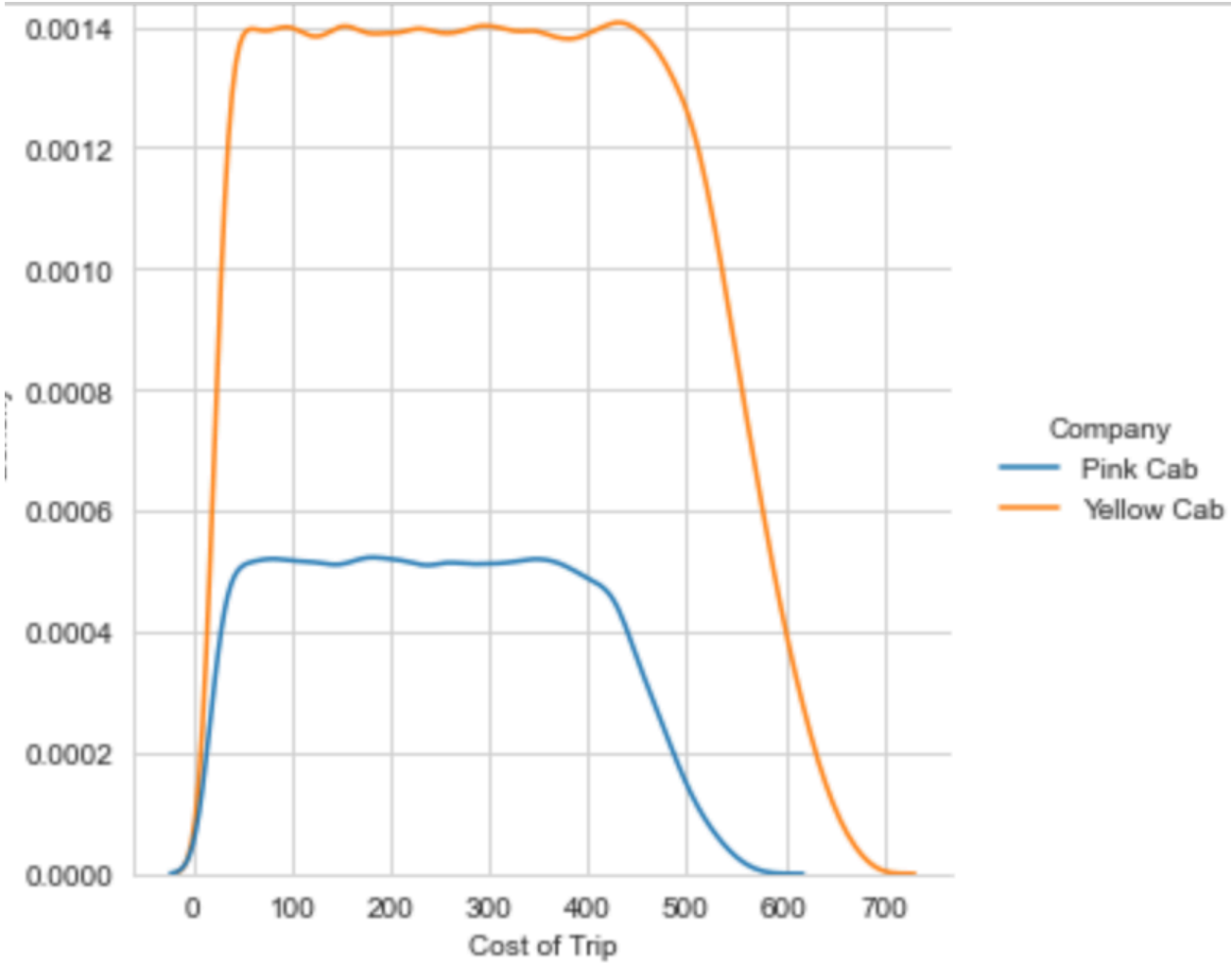
- From the adjacent Probability Density Function we can easily determine that the price charged of yellow cabs is far more than that of pink cab so technically people should prefer pink cabs over yellow cab. This we will explore in the coming slides.



# Analyzing different features of the given data

### 3. Analysis of Cost of Trip

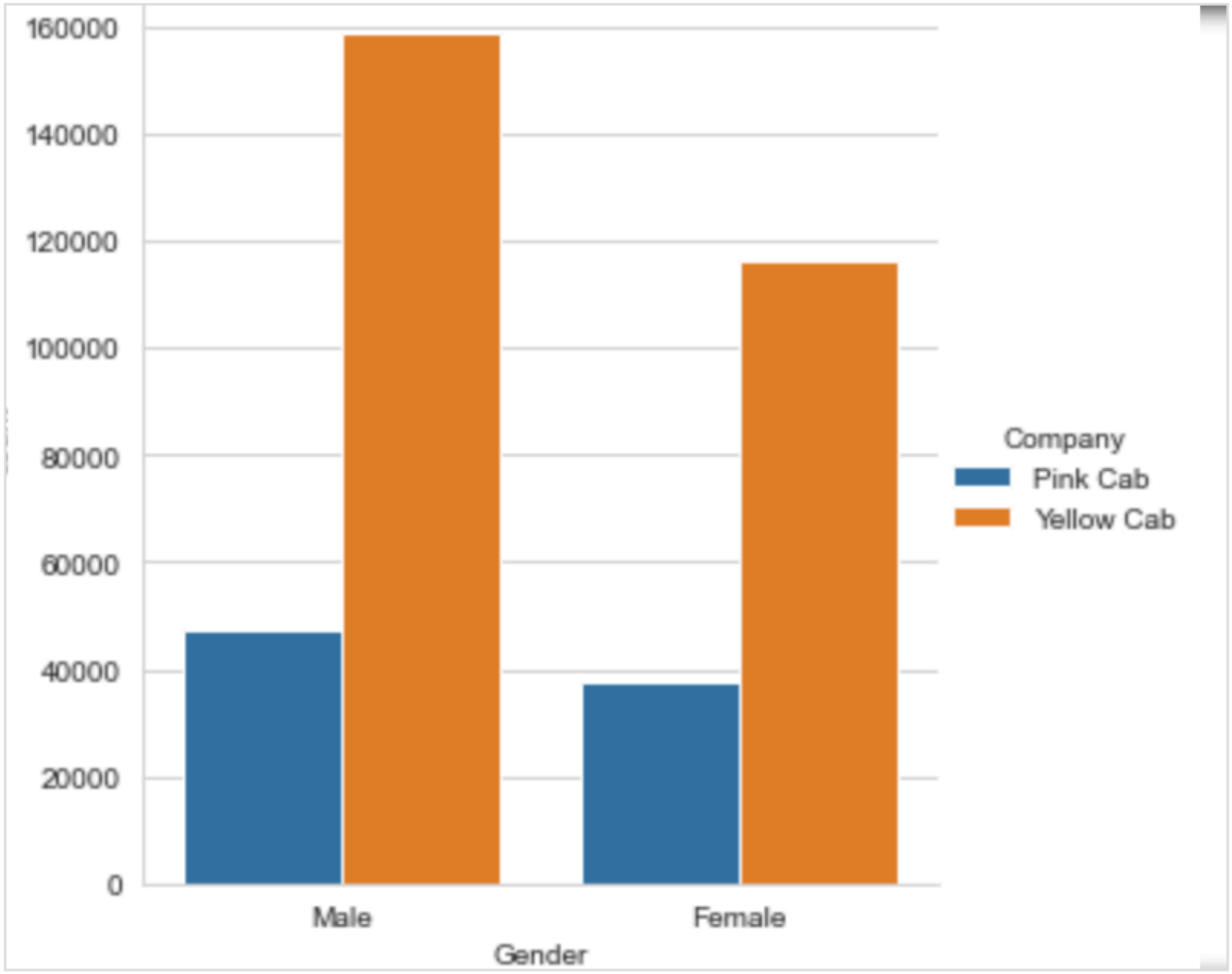
- From the adjacent Probability Density Function we can easily determine that the cost of trip of yellow cabs is far more than that of pink cab. However, it might be just because yellow cabs charge higher than pink cabs.



# • Analyzing different features of the given data

### 4. Analysis of Gender

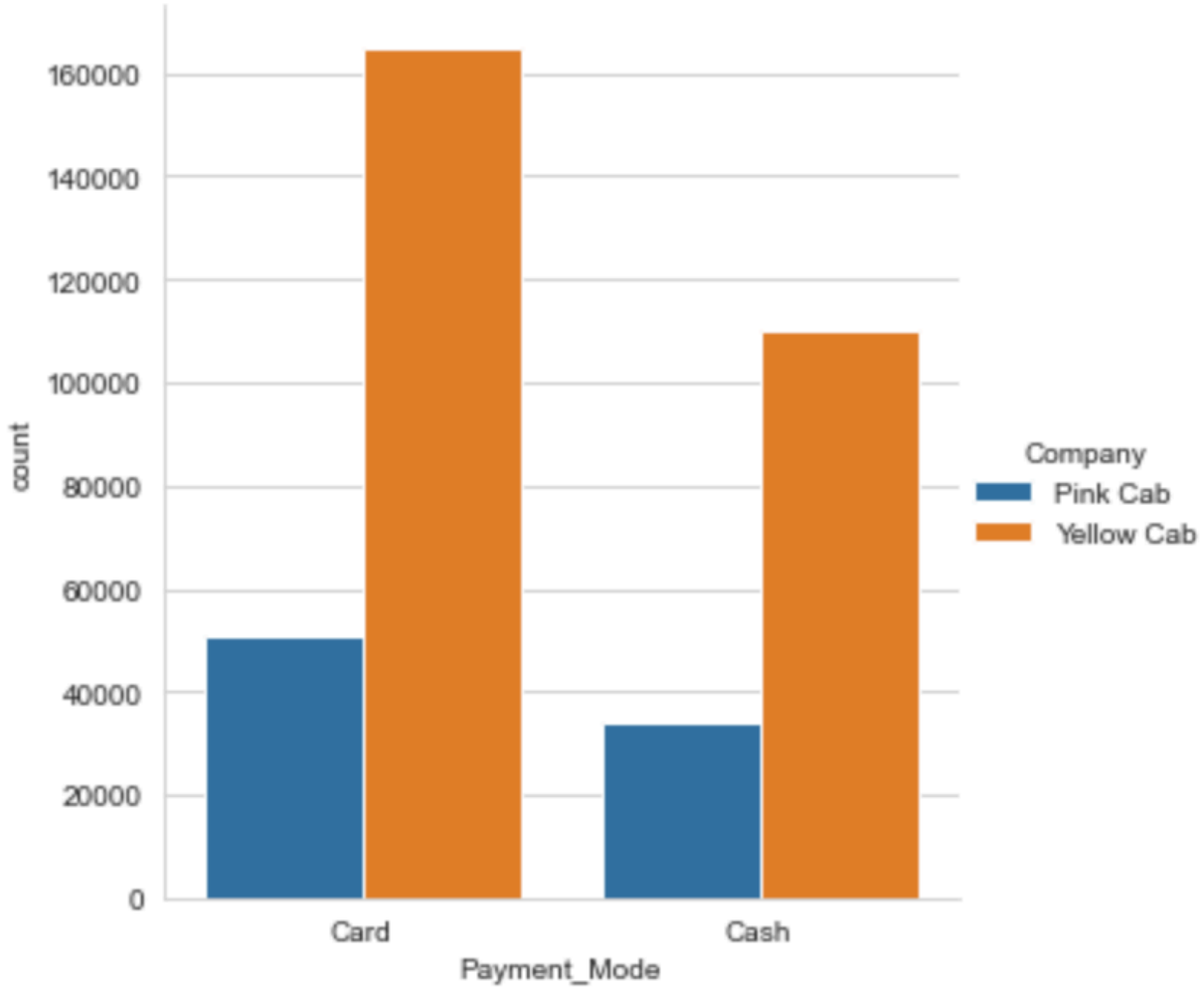
- From the adjacent categorical graph we can determine that male and females both prefer yellow cabs over pink cabs.



# Analyzing different features of the given data

## 5. Analysis of Payment Mode

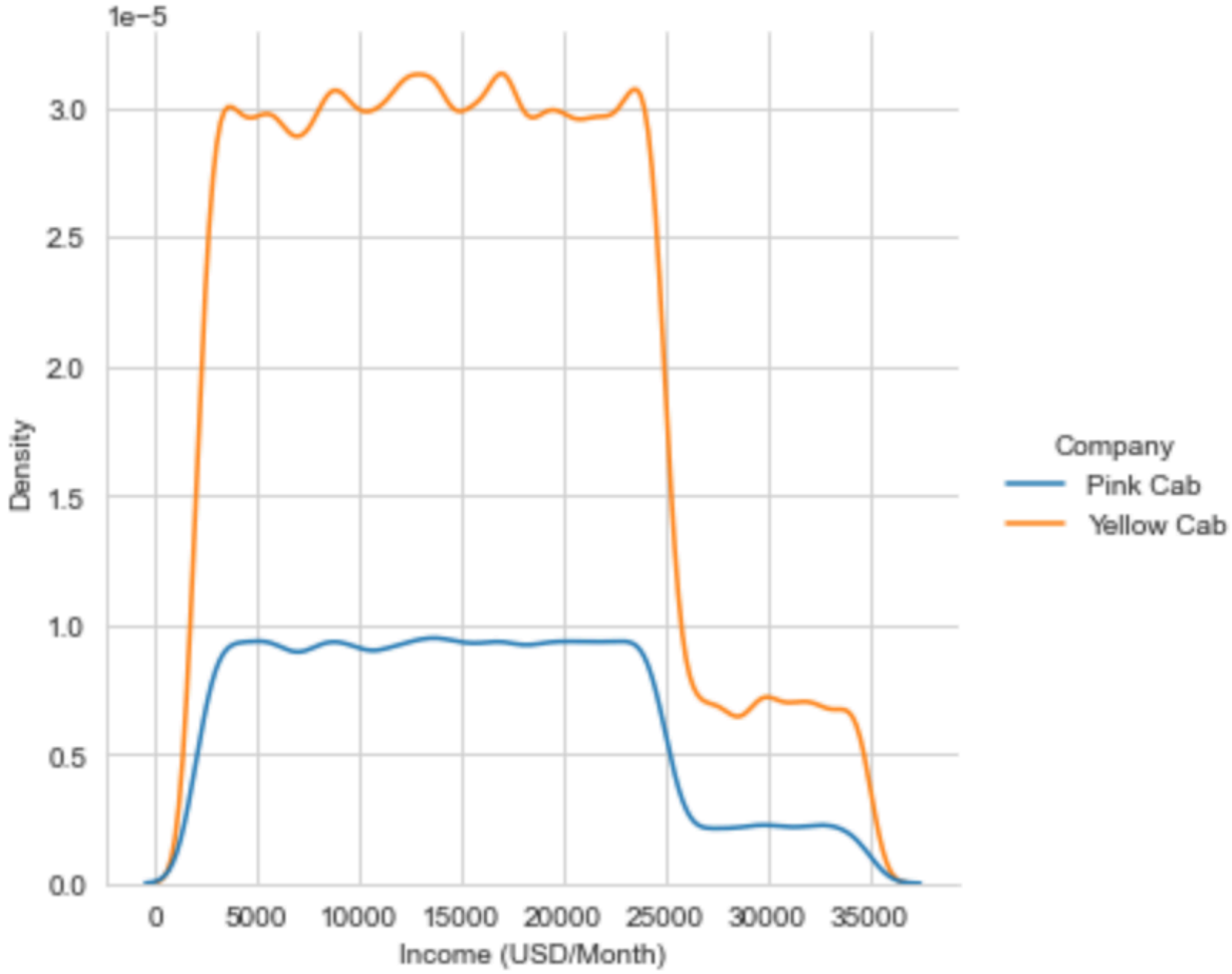
- From the adjacent categorical graph we can determine that for both the payment modes more payments were made for yellow cabs than for pink cabs.



# Analyzing different features of the given data

### 6. Analysis based on Income

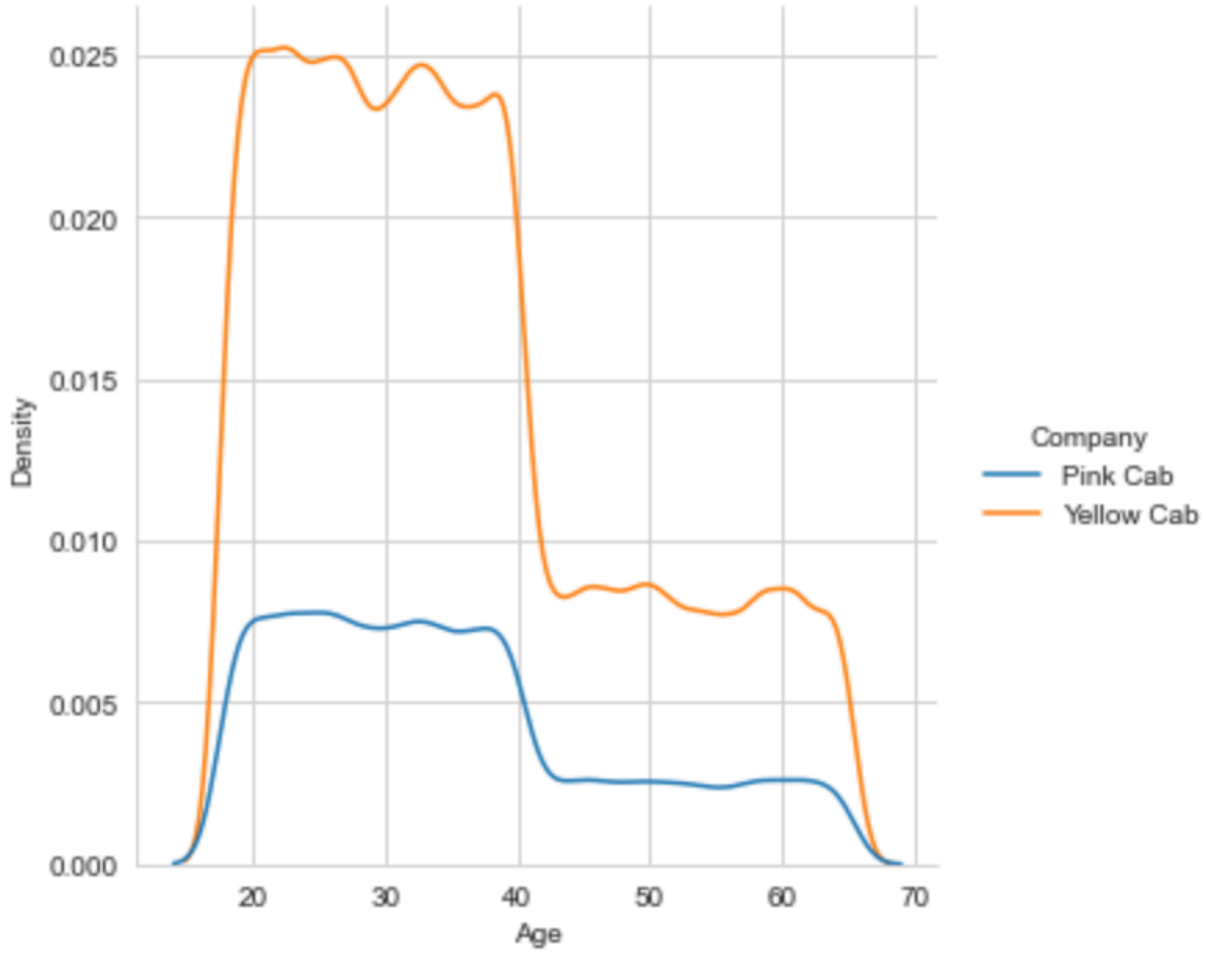
- From the adjacent Probability Density Function we can easily determine that the income of people traveling via yellow cabs is much more than those of traveling via pink cabs.



# Analyzing different features of the given data

## 6. Analysis based on age

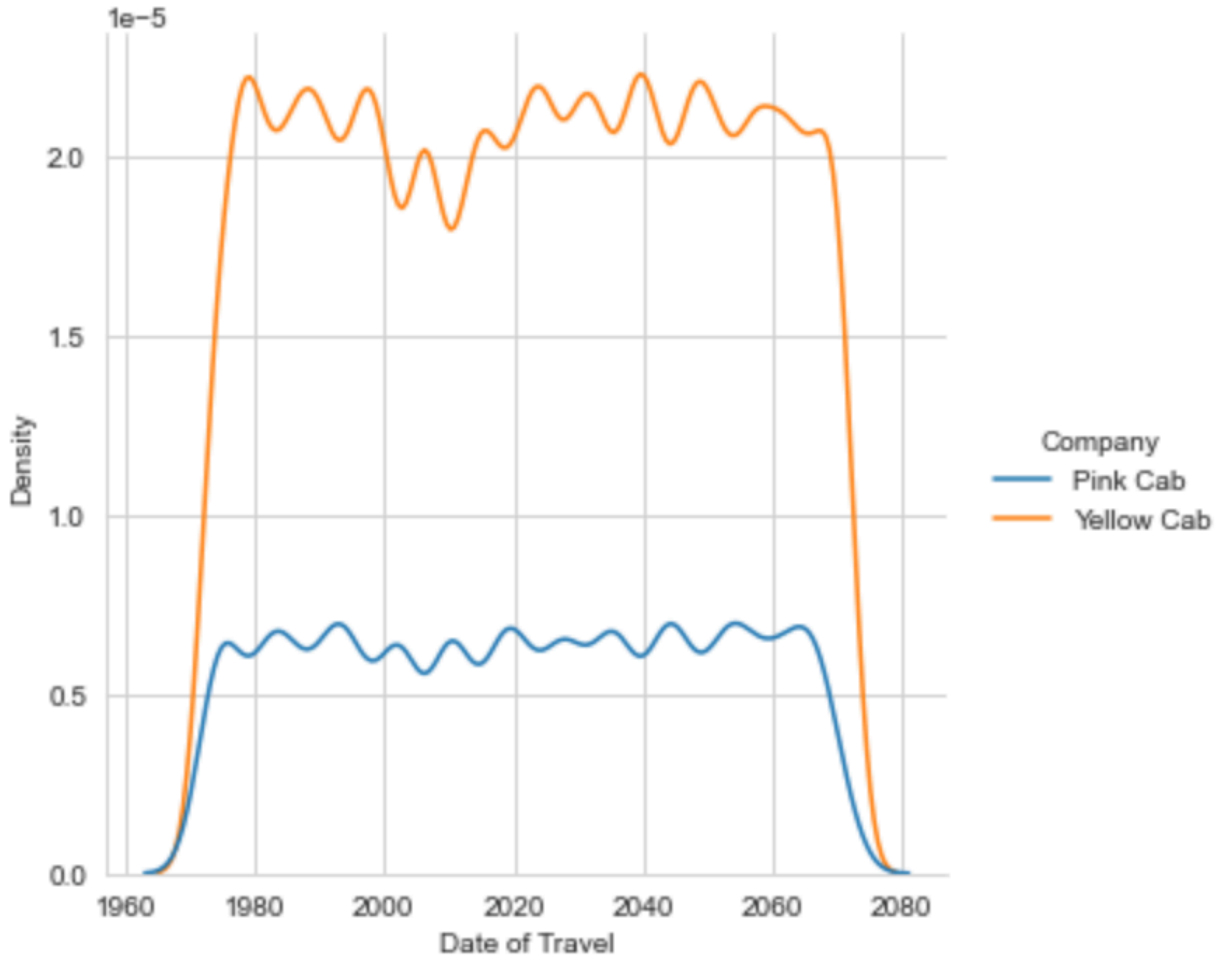
- From the adjacent Probability Density Function we can easily determine that:
- The income of people traveling via yellow cabs is much more than those of traveling via pink cabs.
- The people between age 15- 40 are more inclined to take cabs than people over age 40.



# Analyzing different features of the given data

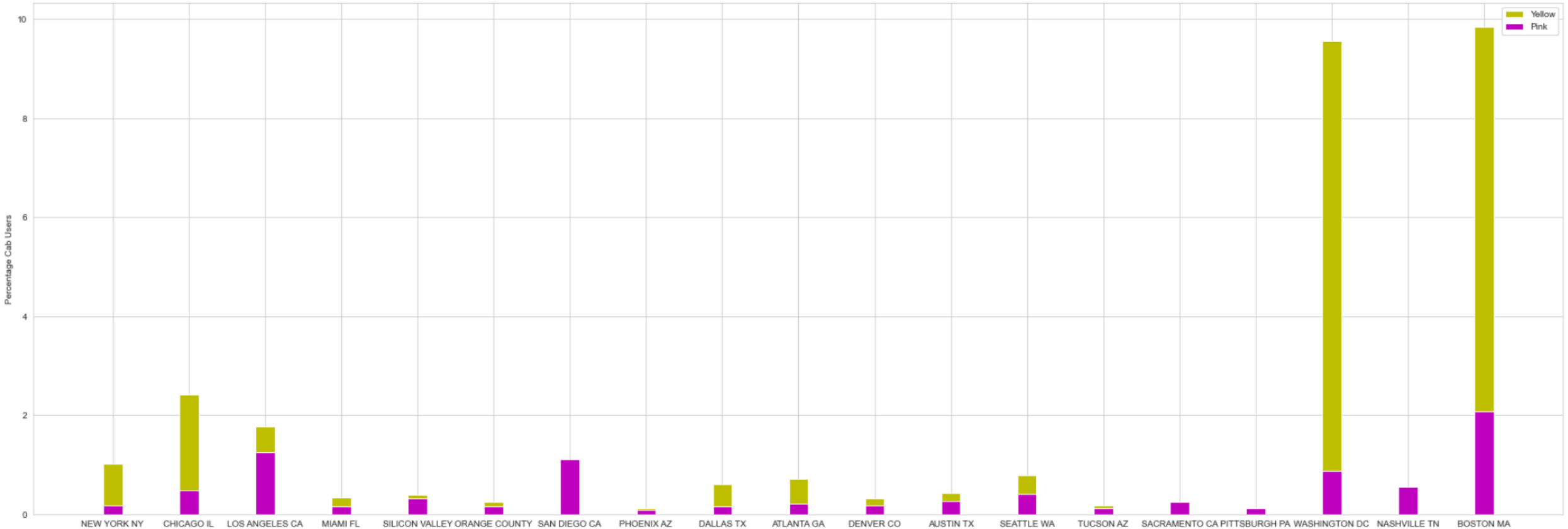
### 7. Analysis based on Date

- From the adjacent Probability Density Function we can easily determine that yellow cabs are travelled more often than pink cabs.
- We can also observe that there is a dip in cab usage of yellow cabs after 2021-2016 however it is still higher than pink cabs.



# Analyzing different features of the given data

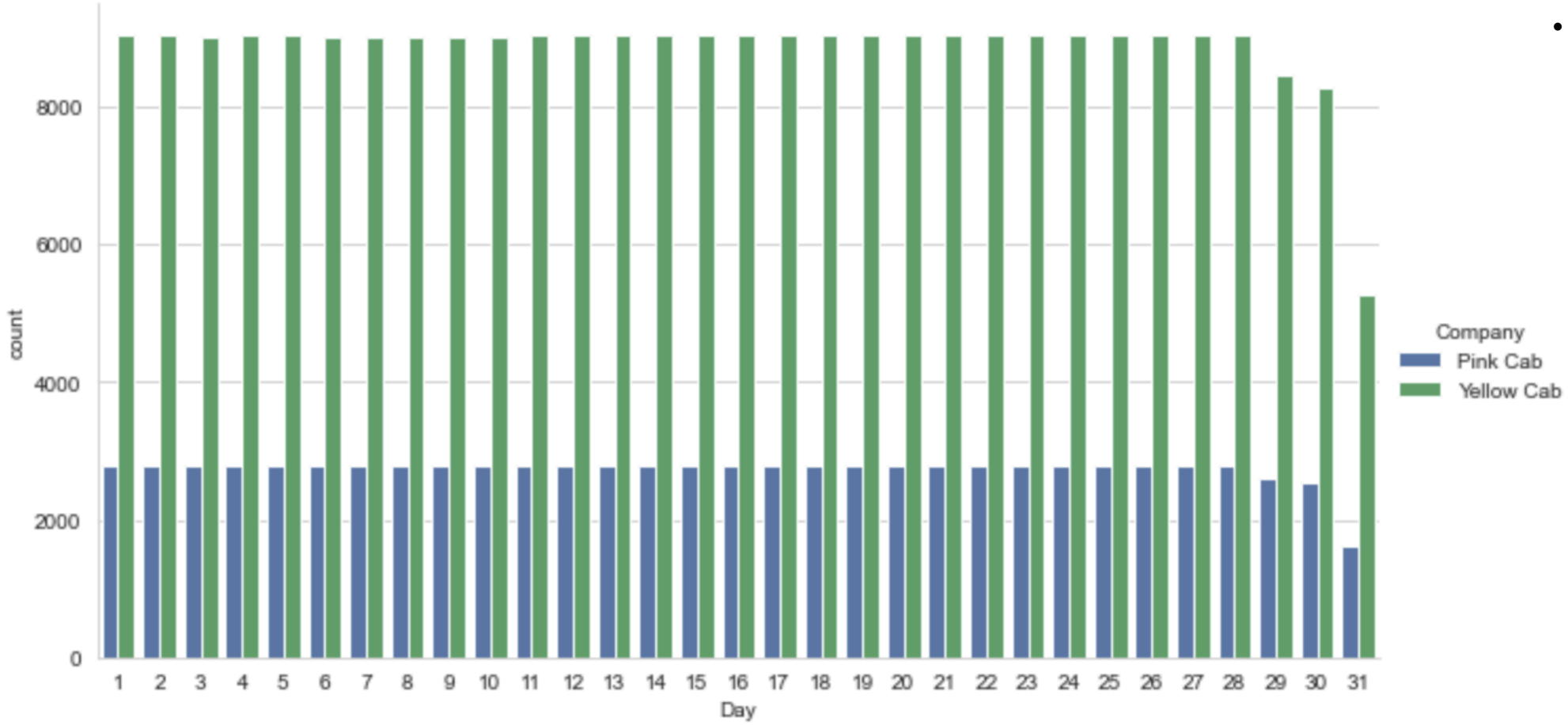
## 8. Analysis Based on Percentage of Users




- From the percentage graph we can confirm that percentage of yellow users in more than that of pink users except for San Diego , Sacramento, Pittsbu and Nashville.

# • Analyzing different features of the given data

## 8. Analysis Based on Day of the Month




- We can observe from the adjacent figure that there is considerable drop in number of rides for 31st of the month.

# EDA Summary

- From the above analysis we can conclude the following:
  1. Analysis of kilo meters travelled we can conclude that yellow cab has traveled much more than pink cabs.
  2. Analysis of Cost pf Trip and Price Charged we can see that it is much more for yellow cab than pink cab.
  3. Analysis of gender we observe that both genders prefer yellow cabs over pink cabs.
  4. Analyzing the payment method we observe for both methods people prefer yellow cabs over pink.
  5. Analyzing the age the income of the users we observe, users of all age and incomes prefer yellow cabs over pink cabs.
  6. Analyzing the dates of travel we observe, usage of yellow cabs is much higher than pink cabs.
  7. Though the price of yellow cabs is way higher than pink cabs there are more yellow cab users than pink cabs
  8. When we plot the data for users in each city, we observe except for Nashville and San Diego users in all other cities yellow cabs over pink cabs, it can be made more clear once we know the percentage of users in each city.

# Recommendation

**Conclusion Drawn:**

1. Though the price of yellow cabs is way higher than pink cabs there are more yellow cab users than pink cabs hence, the profit earned on investing in yellow cabs will be higher.
2. Also people form all age groups and all genders prefer the yellow cabs and hence yellow cabs will likely be preferred if they are introduced at some new cities. Therefore, there is a higher chance of succeeding in new market

**On the basis of the analysis and the conclusions drawn we recommend Yellow Cabs will be a far better investment than Pink Cab Company**

- Shreya Dwivedi

# Thank You

Data Glacier
Your Deep Learning Partner