



**Data Glacier**

Your Deep Learning Partner

# Exploratory Data Analysis

G2M insight for Cab Investment firm

**Name : Saad Bin Munir**

# Agenda

Executive Summary

Problem Statement

Data Exploration

EDA Analysis

Conclusions

Recommendations

# Problem Statement

- Due to remarkable growth in the Cab Industry in last few years and multiple key players in the market, XYZ (A private firm in US) is planning to invest in Cab industry and as per their Go-to-Market(G2M) strategy they want to understand the market before taking final decision.
- In this project we will use multiple data sets that contain information on 2 cab companies and use that information to identify the right company to make their investment.

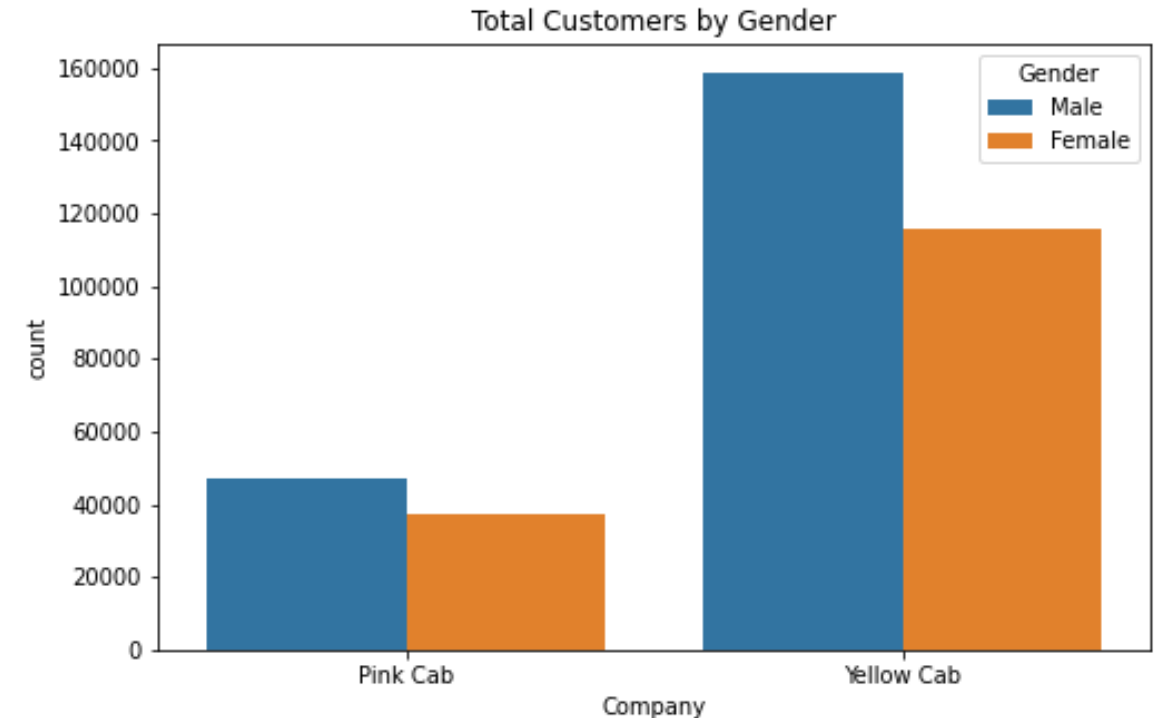
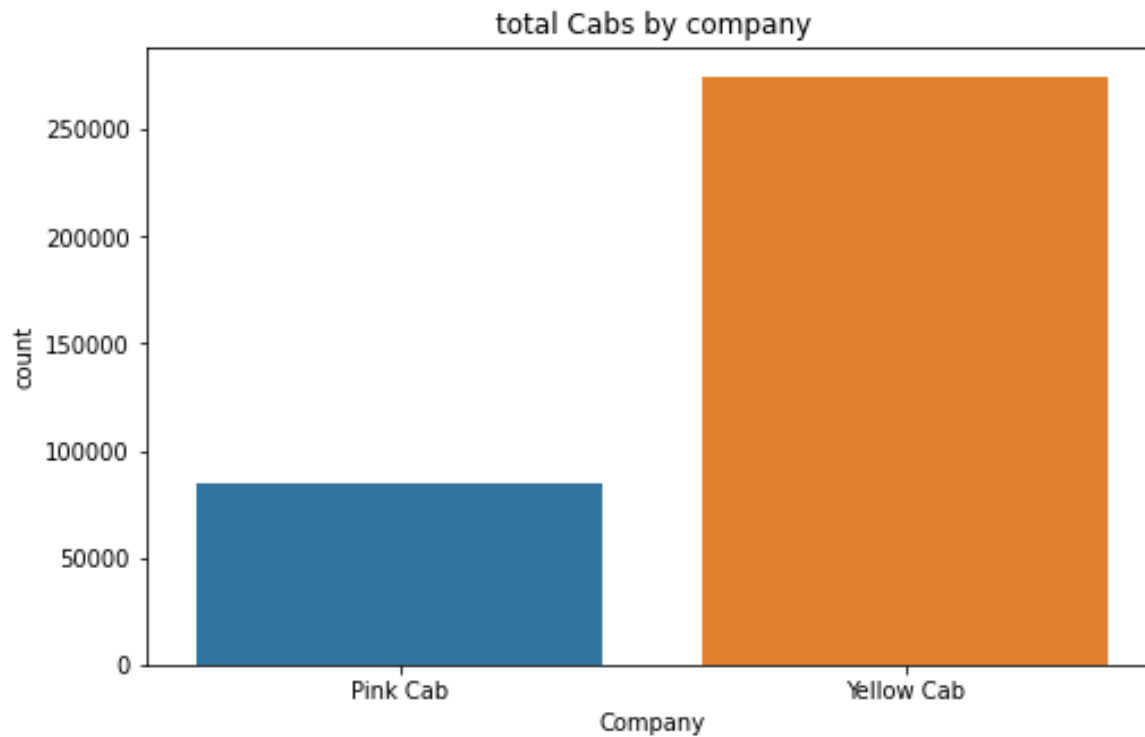
# Data Exploration

- The datasets were already clean and had no duplicate values preset.
- The datasets contains data of three years ranging from 2016 to 2018.
- Univariate outlier analysis was carried out to find out the information about variables.
- All four datasets were combined in meaningful way in order to use them for analysis.

Filename	Number of entries	Number of attributes
<b>Cab_Data.csv</b>	359392	7
<b>Customer_ID.csv</b>	49171	4
<b>Transaction_ID.csv</b>	440098	3
<b>City.csv</b>	20	3

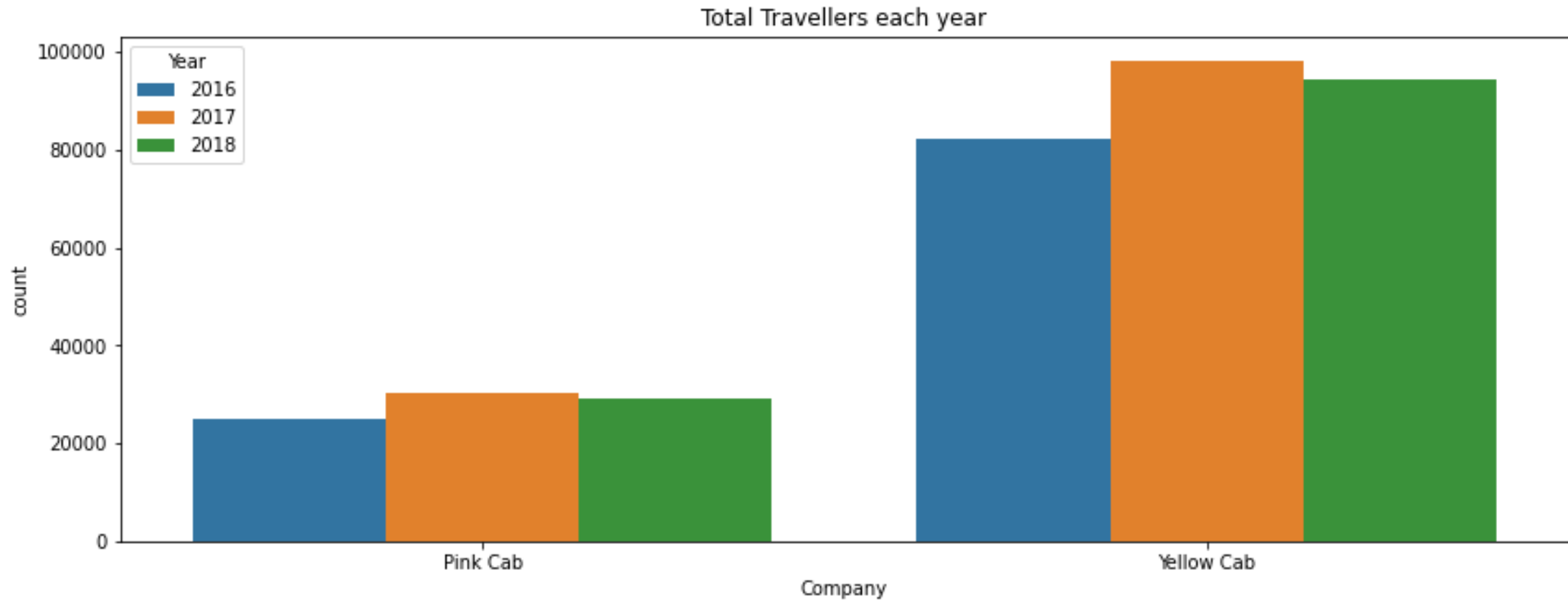
# EDA: Company Performance and Gender

- Most of the travelers prefer yellow cab over pink cab.
- Both companies have more proportion of male travelers.



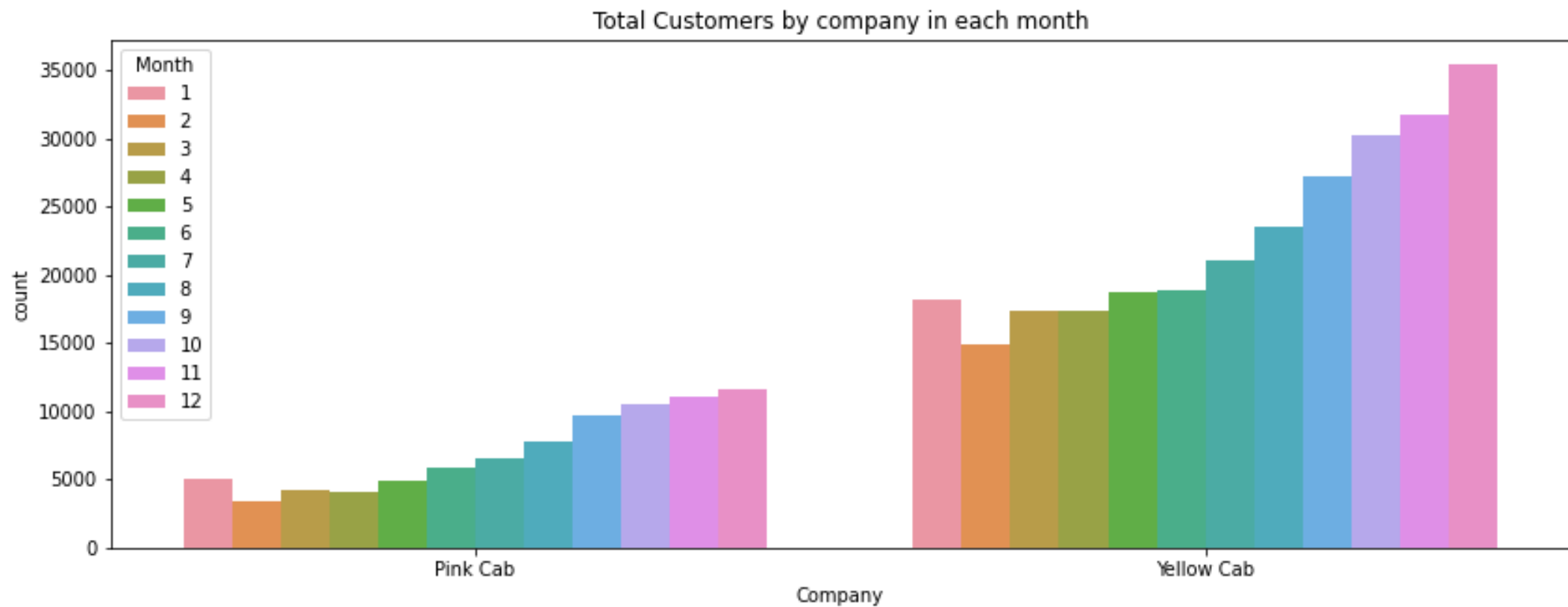
# EDA: Yearly Performance

- Both companies had maximum customers in 2017 and minimum customers in 2016.



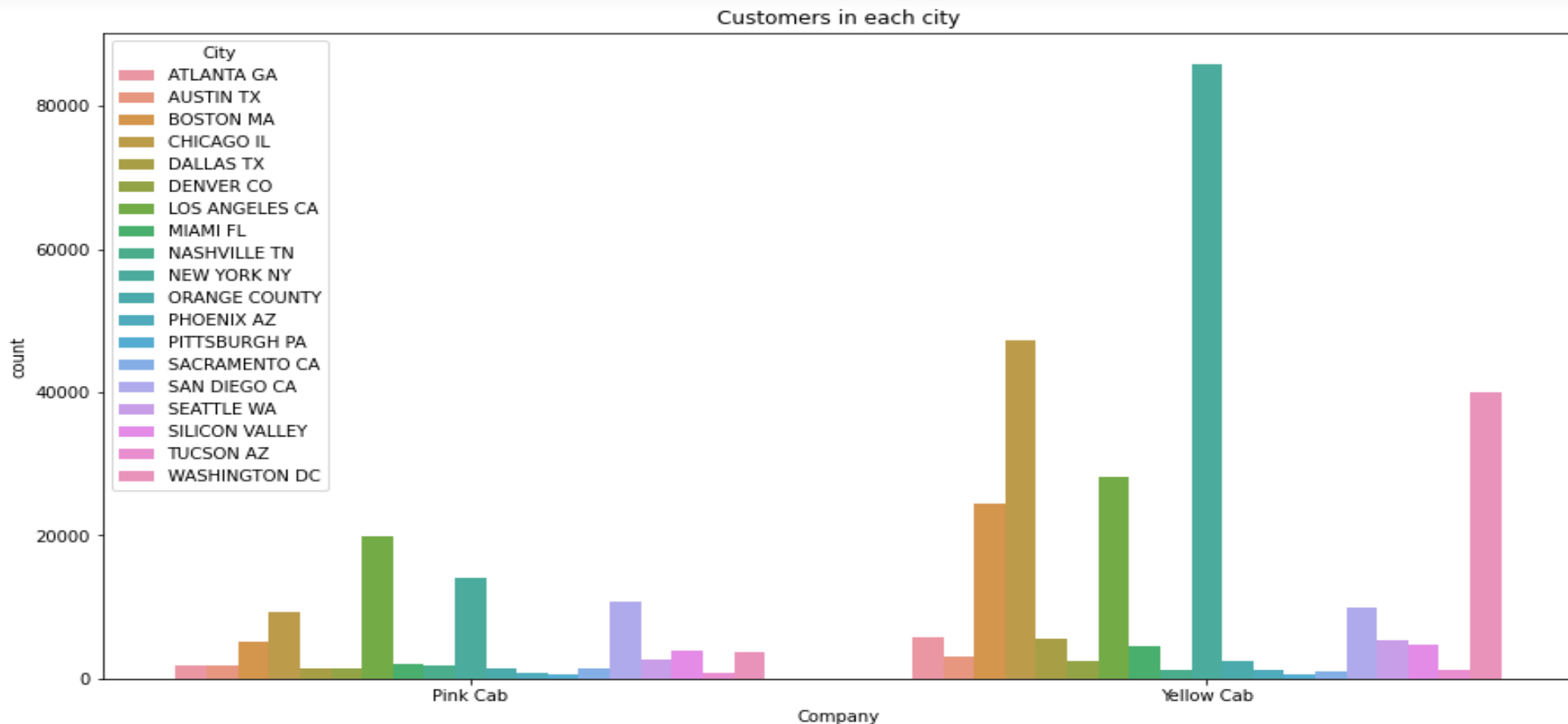
# EDA: Monthly Performance

- The number of travelers increase in both companies with the month.
- The maximum numbers of customers are seen in December for both companies.



# EDA: Locations Overview

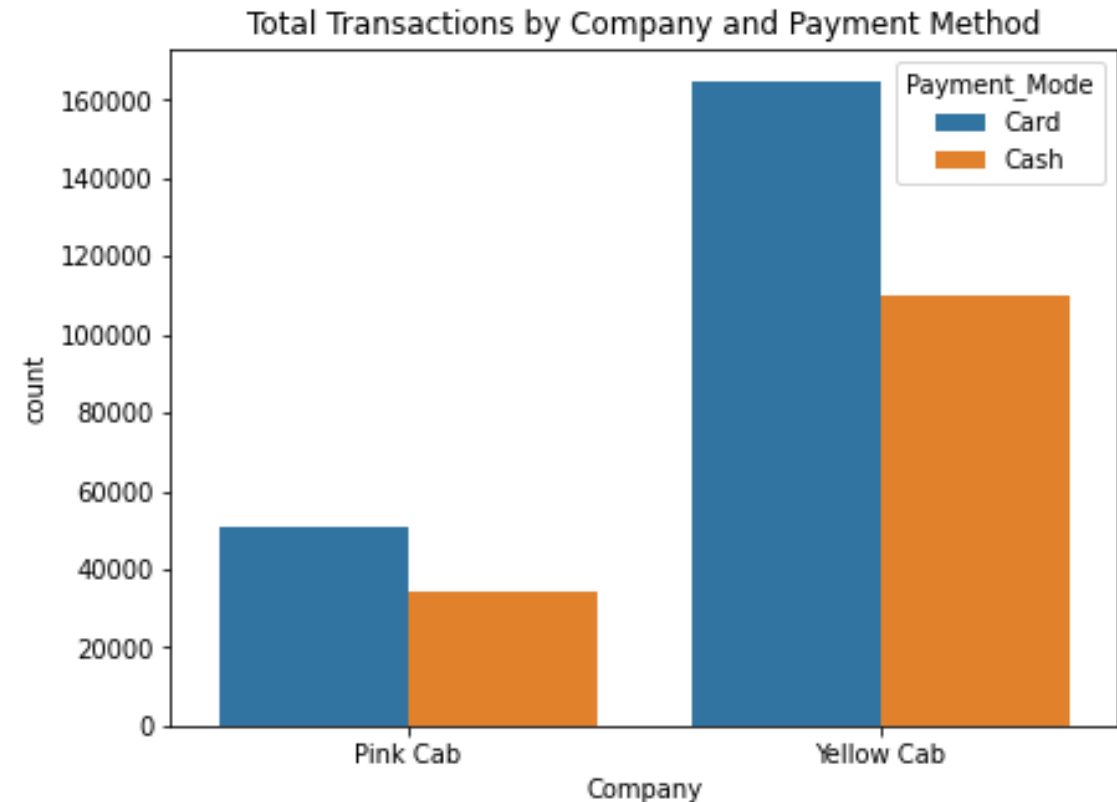
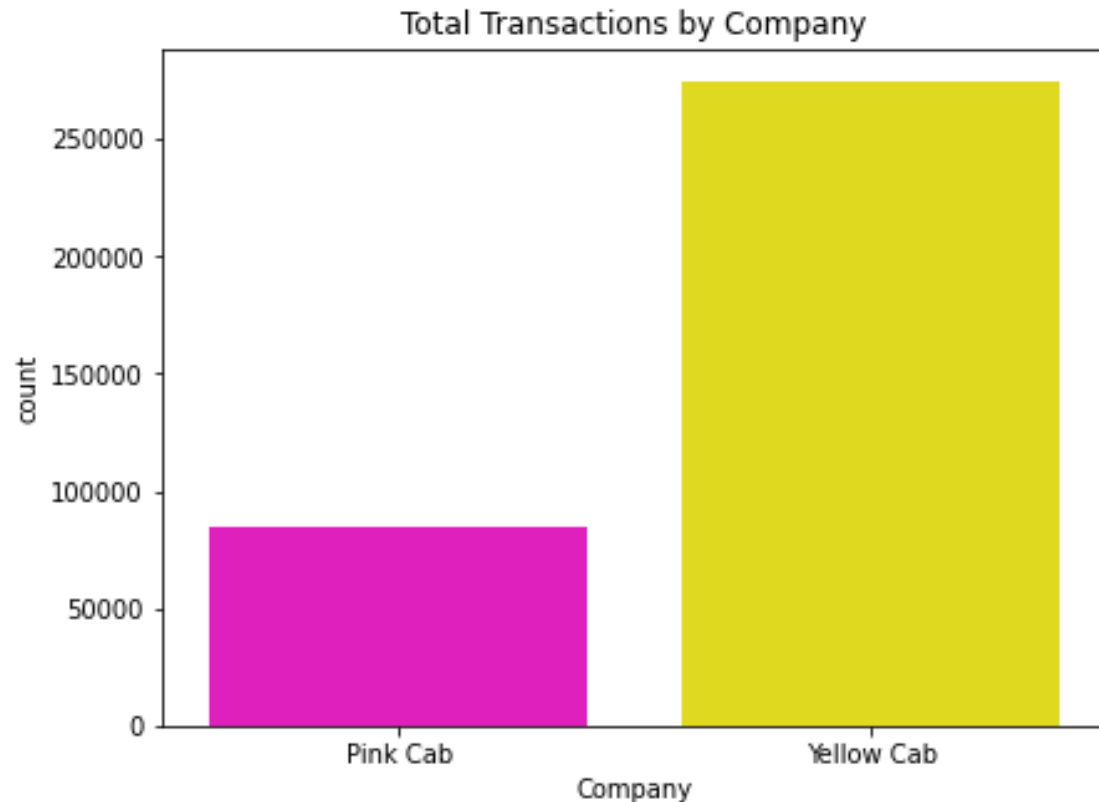
- The Yellow cab is quite popular in New York NY (more than 80000 people travelers) and Chicago IL and Washington DC compared to Pink cab.





# EDA: Transaction Overview

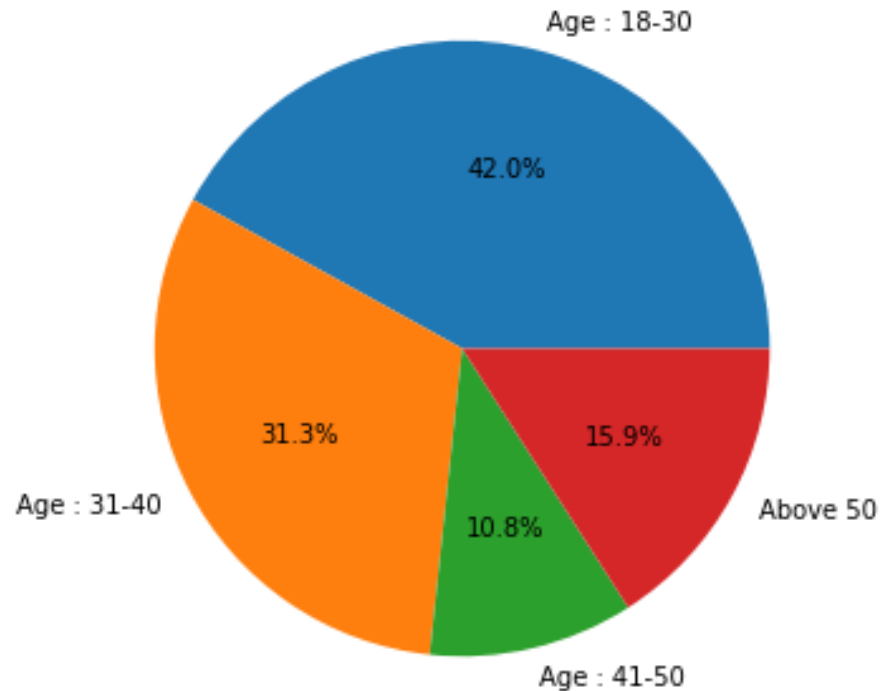
- The yellow cab company has significantly more transactions than pink cab.
- Users of both cab companies prefer to pay using card rather than cash.



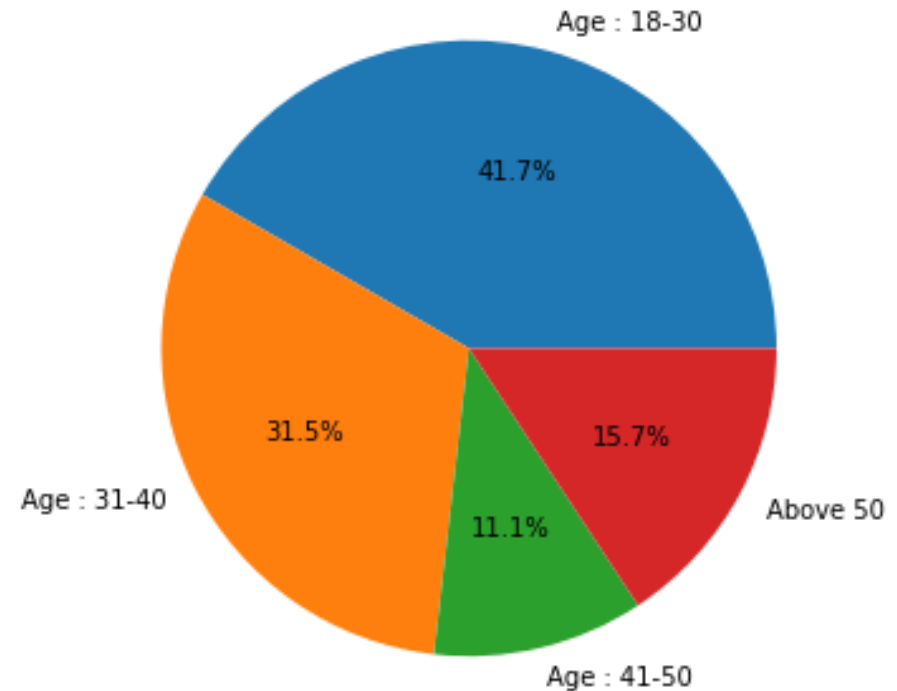
# EDA: Age of travelers

- It is evident that both have almost the same proportions of age groups travelling.
- Most peoples using both cab services are young people i.e., more than 70% are below 40.

Distribution of Pink Cab Travellers by age



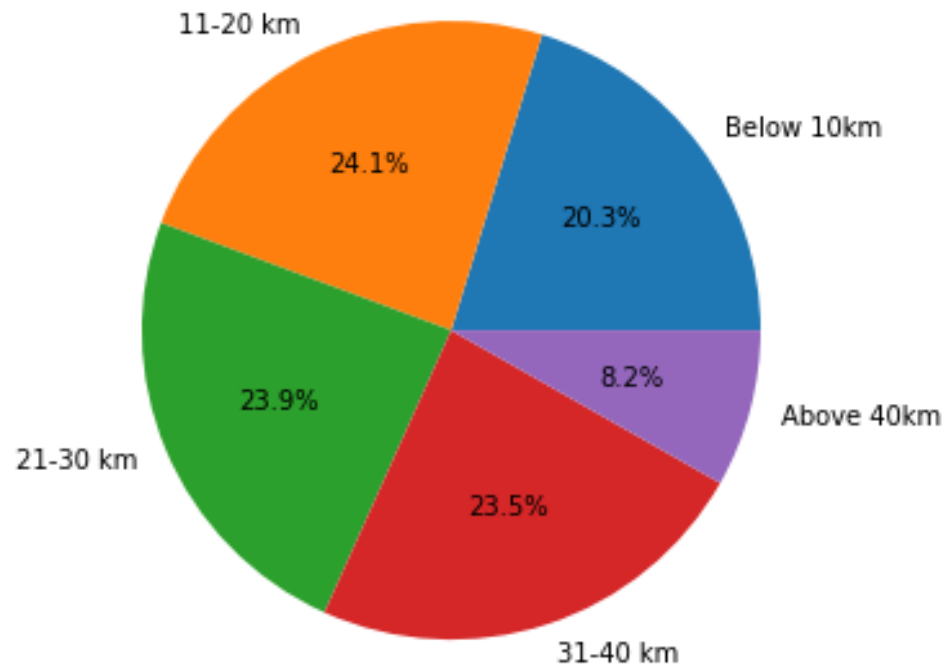
Distribution of Yellow Cab Travellers by age



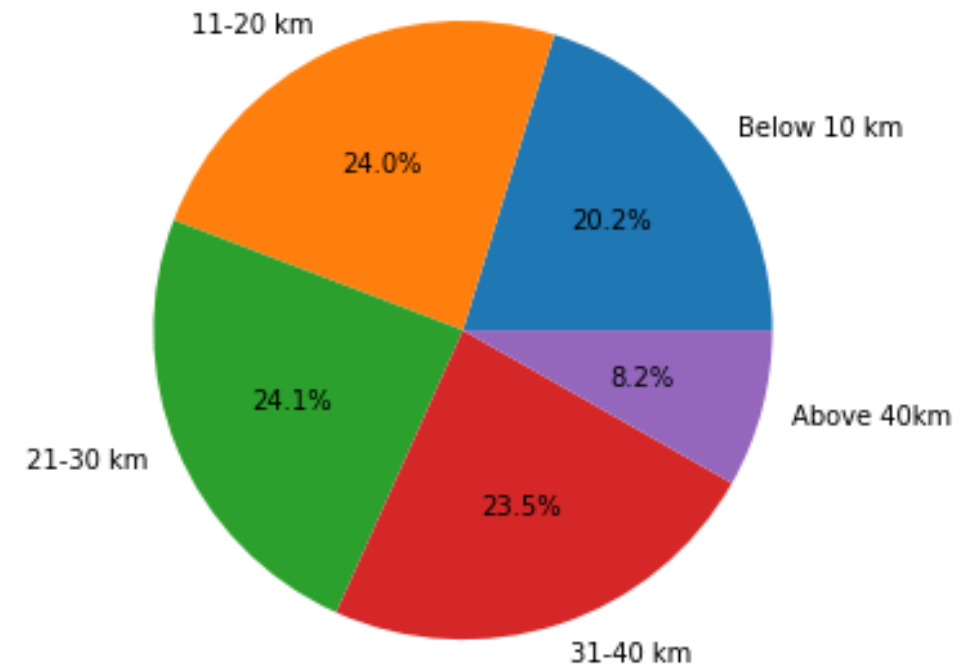
# EDA: Distance Travelled

- It is evident that both have almost the same proportions of distance travelled.

Distribution of Pink Cab Travellers by km travelled



Distribution of Yellow Cab Travellers by km travelled



# Conclusions

- Yellow Cab has more travelers than Pink cab and hence is more profitable.
- There is more proportion of male travelers than female in both companies.
- Users of both companies prefer using card as a payment method.
- New York, Chicago and Washington DC have the maximum number of travelers.
- The majority of cab riders are 40-year-old or younger.
- Typical car rides do not exceed 40 kilometers

# Recommendations

- Yellow is more profitable hence we recommend to invest in that company.
- The ideal time to start the service is towards the end of the years because of more travelers.
- Make sure that the new service accepts the card as a payment method because that is preferable.
- It is recommended to target New York, Chicago and Washington DC because they are good marketplaces.

# G2M insight for Cab Investment firm

Thank You