Exploratory Data Analysis

EDA_G2G

18/03/2022

Agenda

Executive Summary

Problem Statement

Approach

EDA

EDA Summary

Recommendations

Data Glacier

Your Deep Learning Partner

Background –G2M(cab industry) case study

• 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
• Finding the most profitable Cab company
• Finding the various customer base segments for the cab companies
• Finding the company that spreads diseases more rapidly
• Recommendations for investment

Data exploration

Summary:

- 19 Features
- Timeframe of the data: 2016-01-31 to 2018-12-31
- Total data points :359,352

Assumptions:

- Outliers are present in Price_Charged feature but due to

unavailability of trip duration details ,we are not treating this as outlier.

- Profit of rides are calculated keeping other factors constant and only Price_Charged and Cost_of_Trip features used to calculate profit.

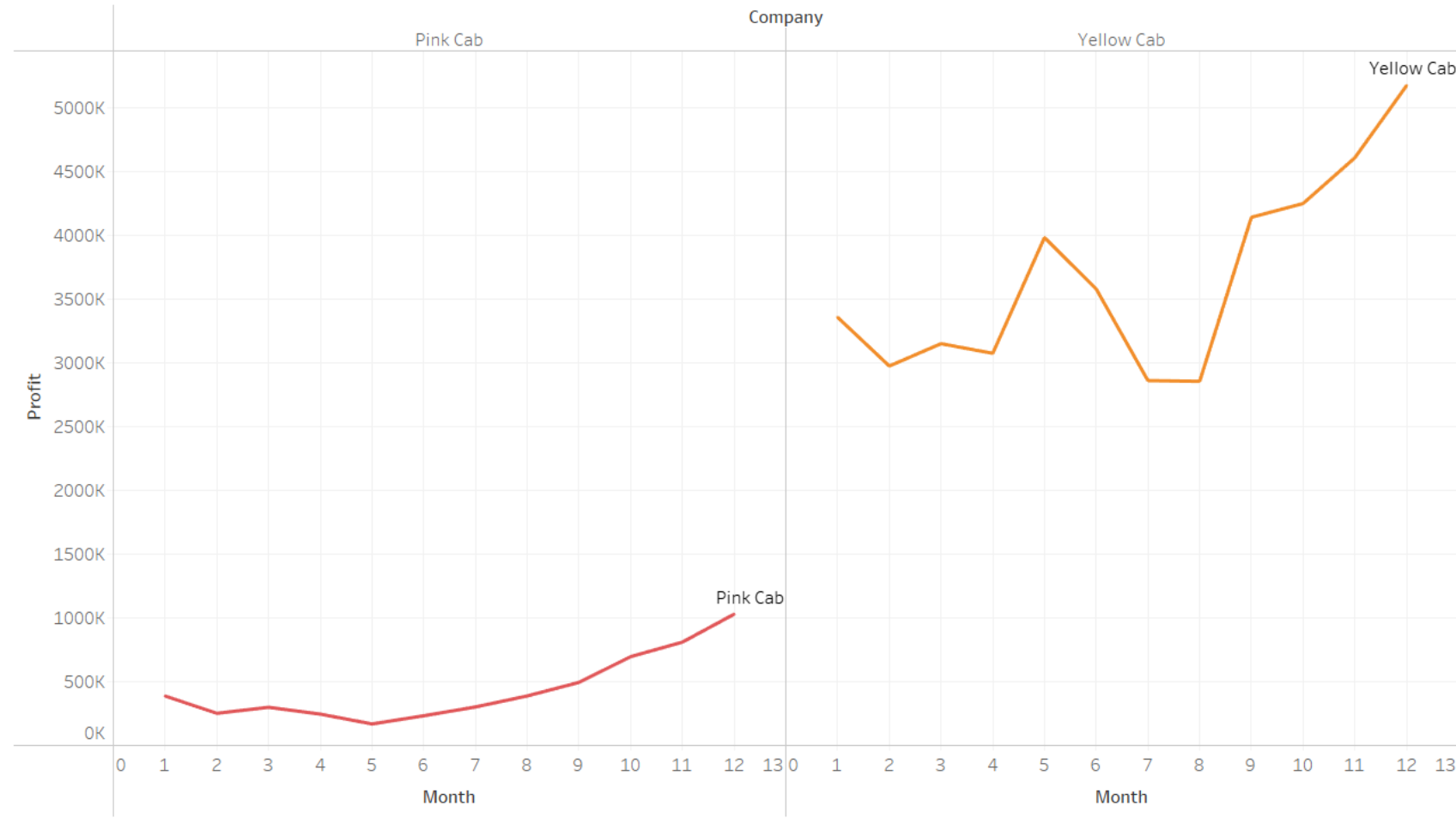- Users feature of city dataset is treated as number of cab users in the city.
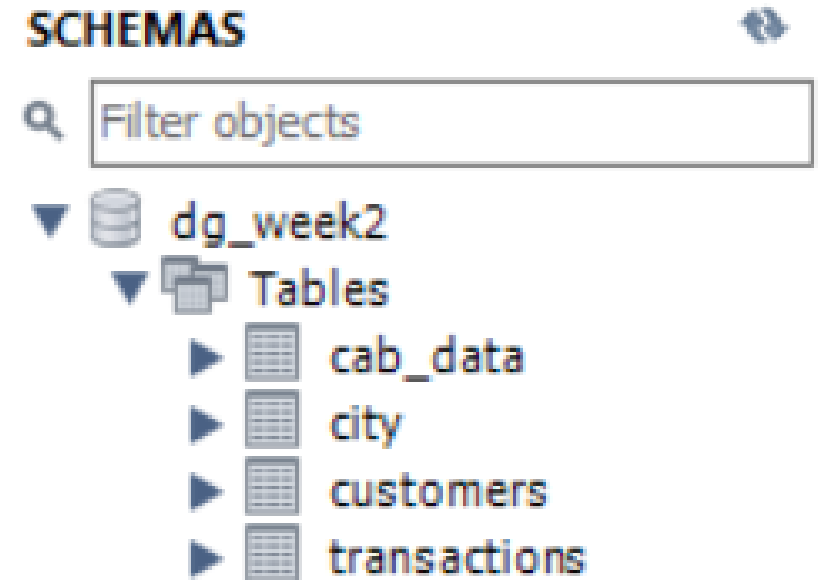
Profit Analysis
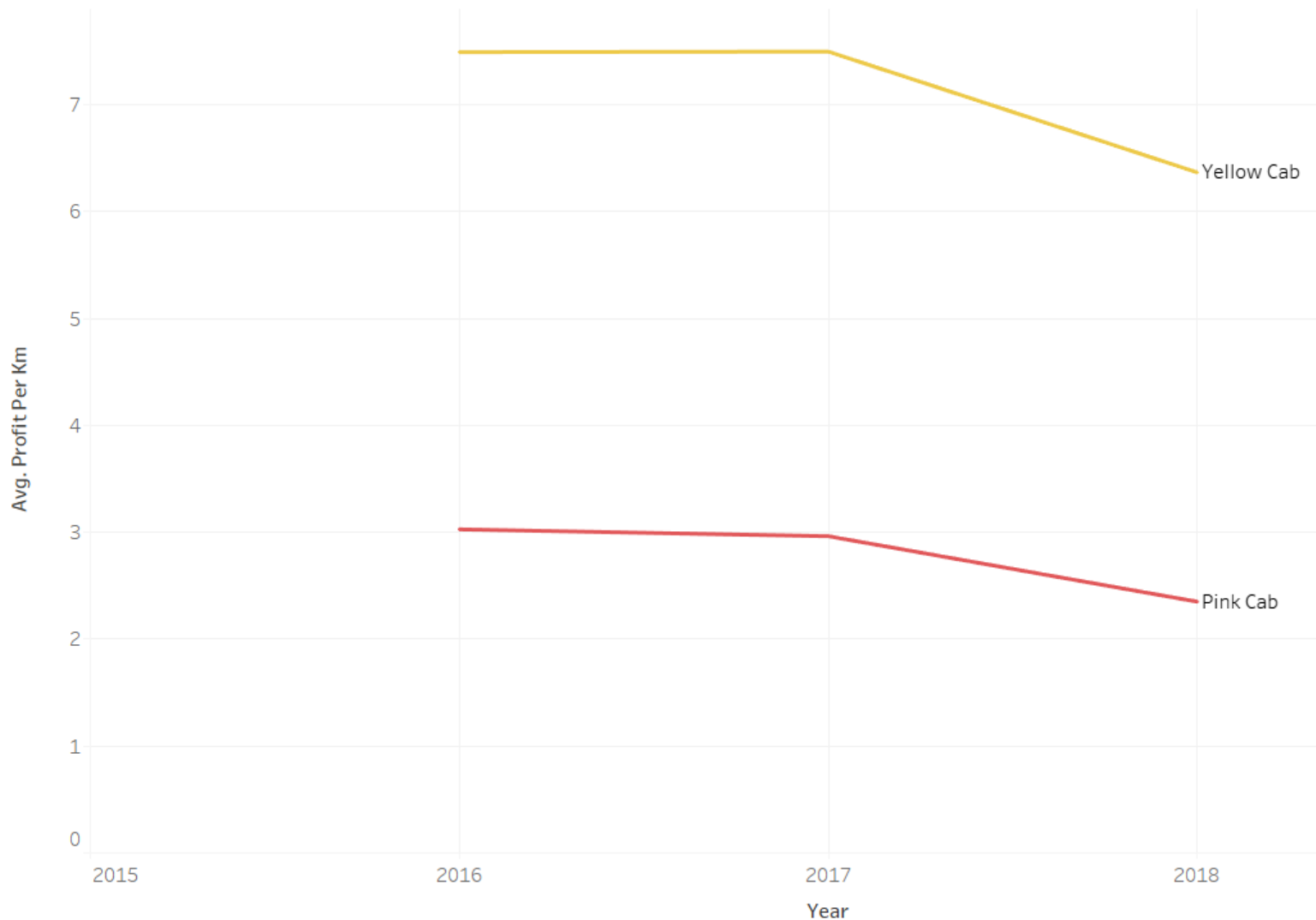
Yellow Cab seems to be much more profitable per km. this indicates good efficiency and management. The decline is higher for yellow cab though, suggesting that the reason is possibly due to market volatility.

Monthly Profit Analysis

Yellow cab seems to be more volatile throughout the three years as illustrated by the graph. On the other hand, pink cab seems to be more consistent with low variations in the profit throughout the year. This suggests that pink cab is less responsive to market demand than yellow cab

Market Share analysis city wise
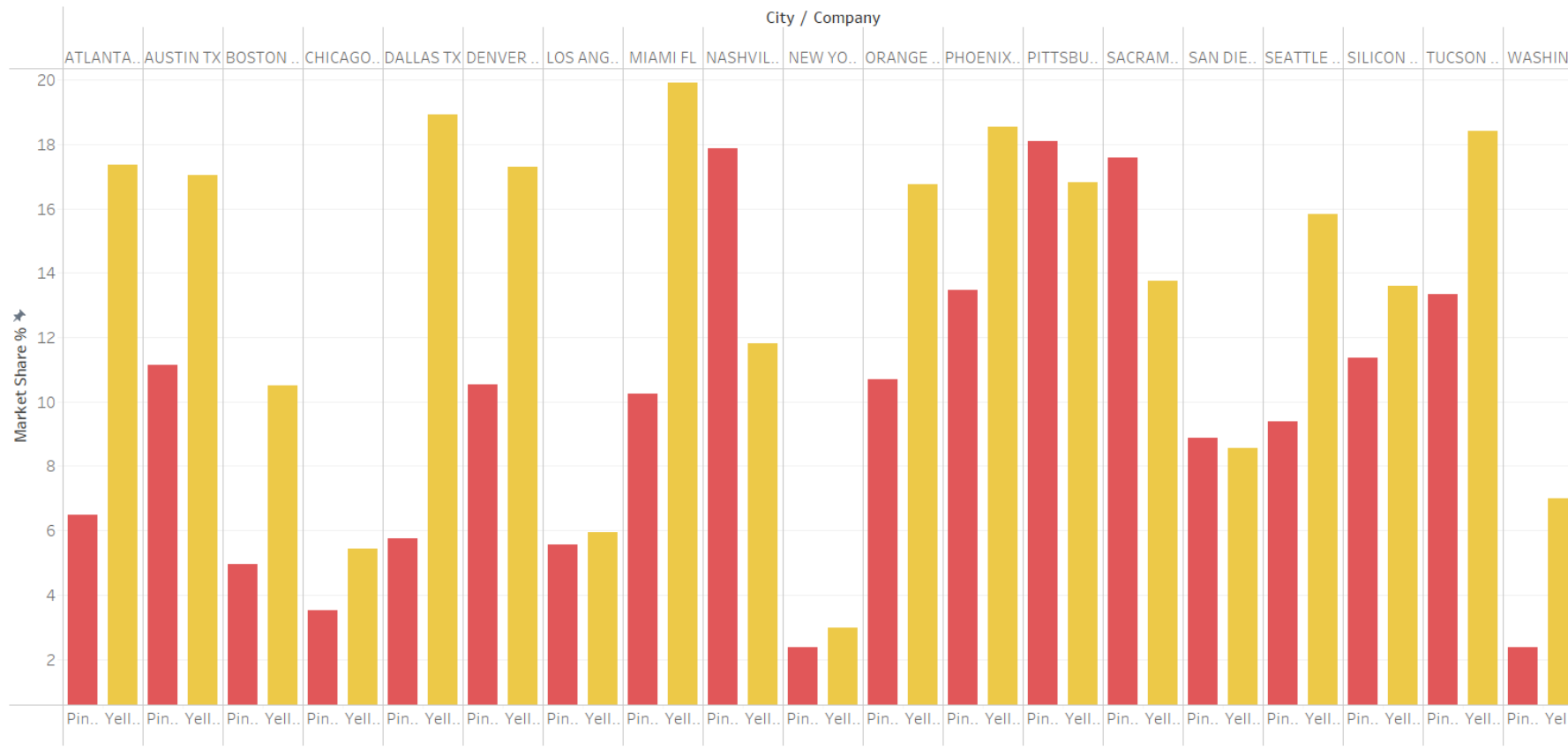
Pink cab has more market share in NASHVILLE, PITTSBURGH and SACRAMENTO. In Every other city in the dataset, Yellow cab has more market share. It is also worth mentioning that in the cities where yellow cab has less, pink cab is only marginally advantaged
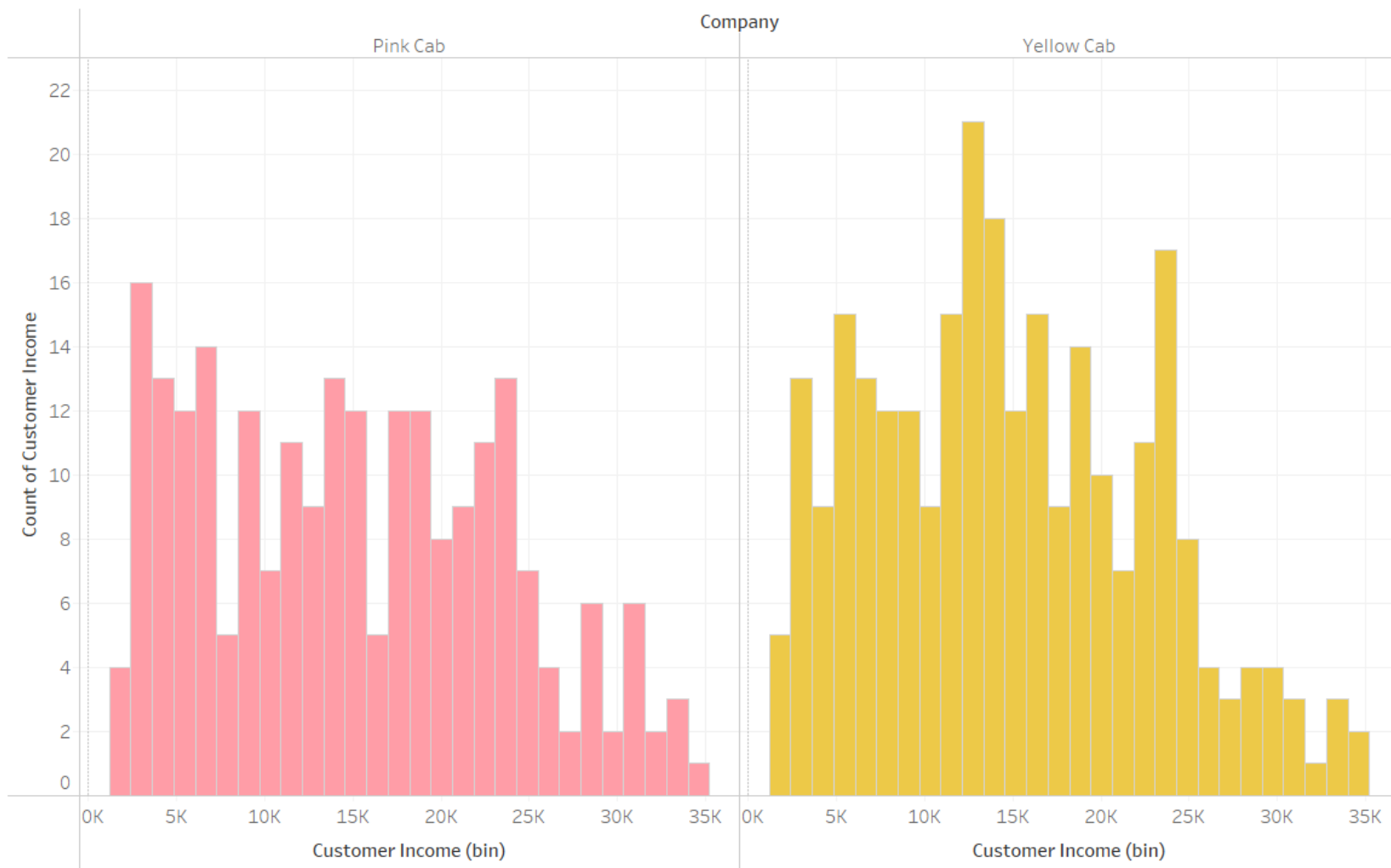
Gender Customer base analysis
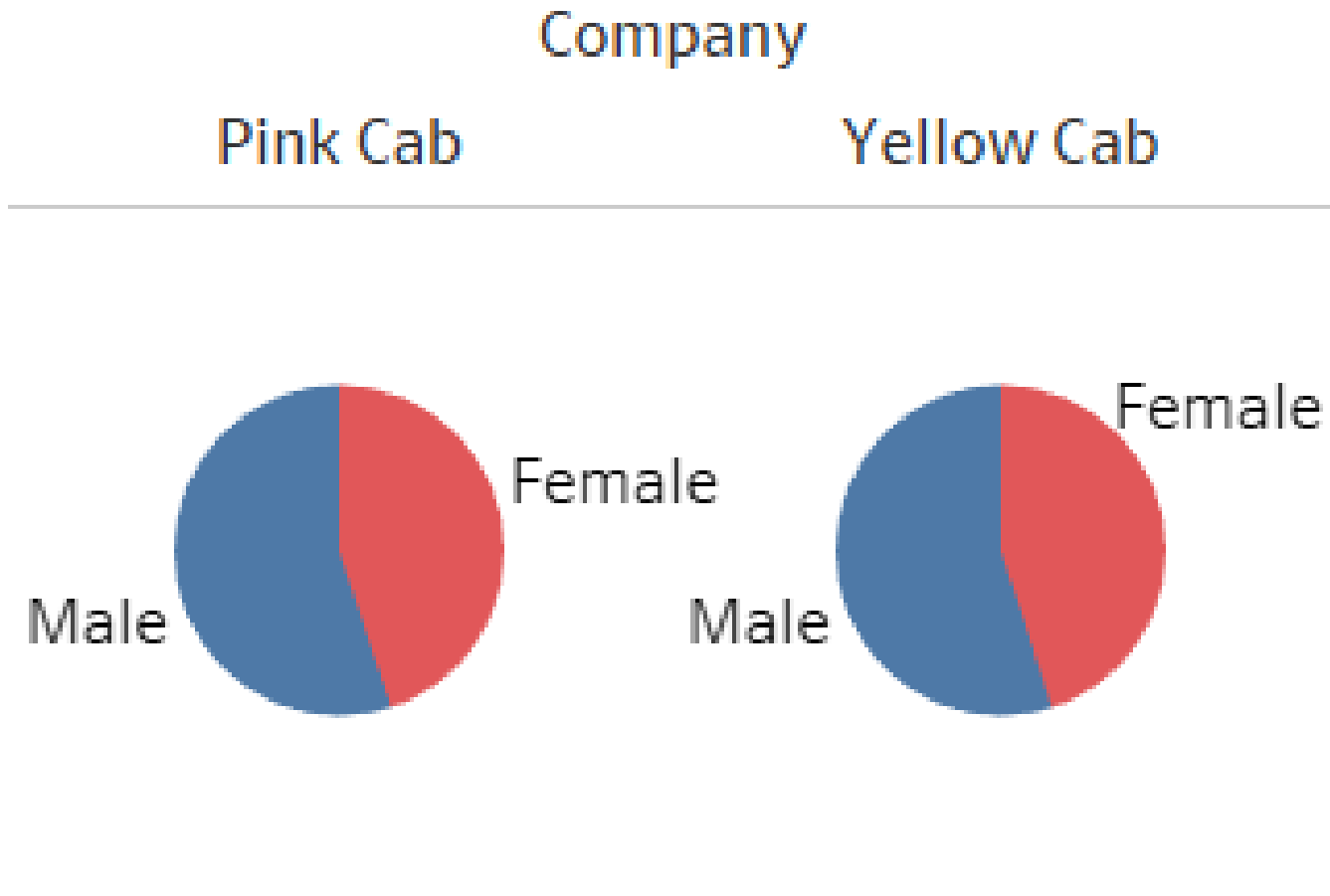
The companies are very close in their gender demographic. Having slightly more men. Which is expected given the fact that there are slightly more men in the customer base of cab companies than women. We can infer from this that neither company have a bias for a certain gender
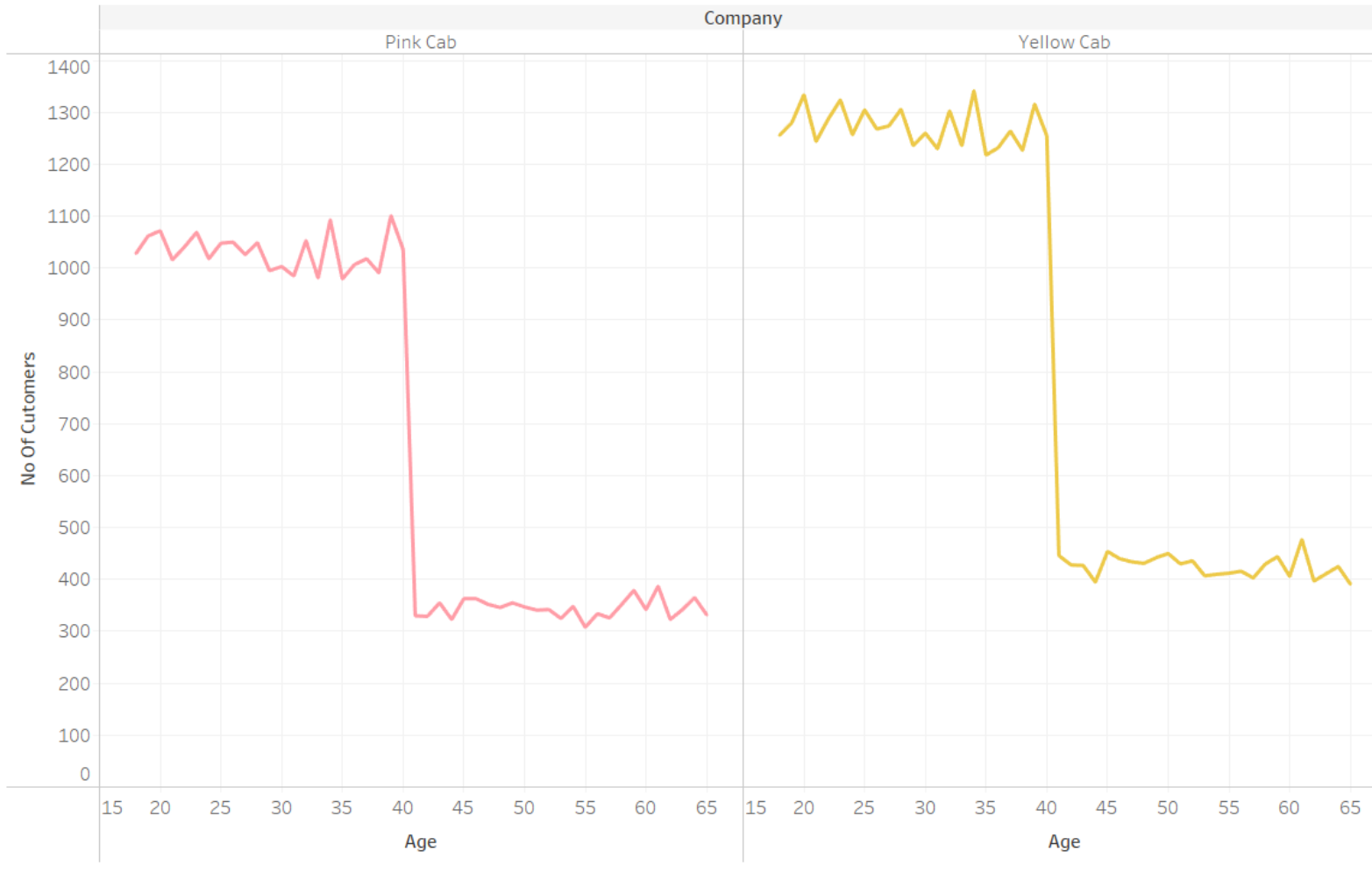
Income Customer base analysis

The distributions tell us that yellow cab has more customers from the middle class of people. Pink Cab has more customers from the lower class and upper class of people.

Customer base age-wise analysis

The distributions suggest that both companies have identical age demographics, indicating a lack of bias for the companies among age groups

Disease spreading potential
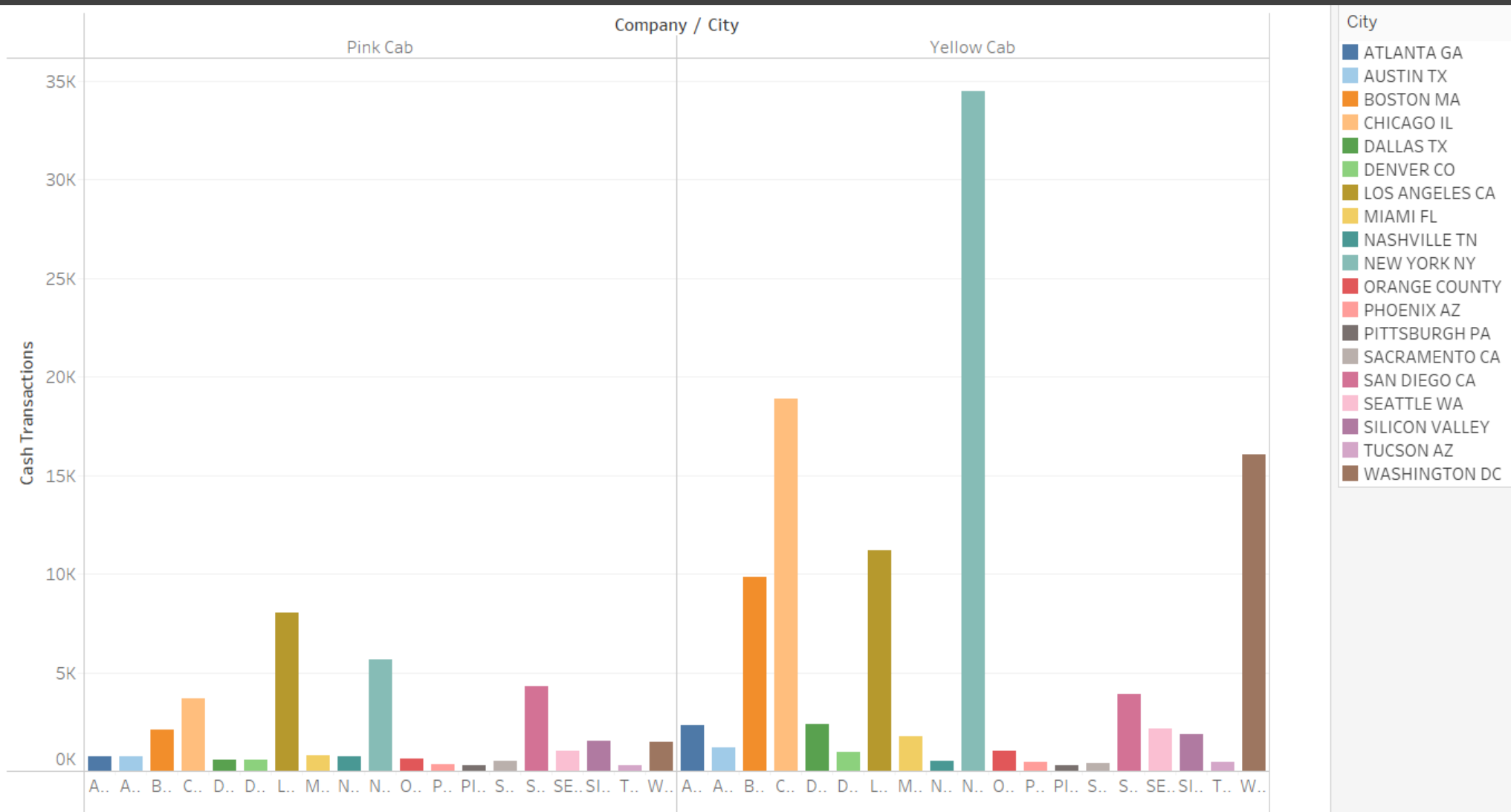
Yellow cab has substantially more cash transactions than pink cab due to the difference in market share. This adds responsibility to yellow cab for decreasing cash transactions for public health. Yellow cab has too many cash transactions in LA despite the similarity in market share between the two companies.

Recommendation

- Customer Reach : Yellow cab has higher customer reach in 25 cities while Pink cab has higher customer reach in 4 cities. We have also observed that Yellow cab is doing good in covering other cab users as compared to Pink cab.

- Age wise Reach : Yellow cab has customers in all age groups.

- Average Profit per KM: Yellow cab's average profit per KM is almost three times the average profit per KM of the Pink cab.

- Income wise Reach :Yellow cab is more popular among middle class groups but offers its services to all class groups.

- Yellow cab responds to market volatility, profits increase during high seasons while pink cab is static by comparison

Based on the points made, we will recommend Yellow cab for investment.

Thank You