



Data Glacier

Your Deep Learning Partner

Exploratory Data Analysis

Project: Cab Company Investment

13-April-2025

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 : Evaluate and compare the statistics two cab companies, Pink Cab and Yellow Cab, to help XYZ firm in identifying the right company for making investment.
- The analysis will utilize hypothesis results and statistical analysis to identify relationships and trends in the provided data to recommend the better investment company for the XYZ firm.

Data Approach

Data Preprocessing:

- Datasets: Cab_Data.csv, Transaction_ID.csv, Customer_ID.csv, City.csv
- 23 Features (including 9 derived features)
- Joined by shared key identifiers
- Timeframe of the data: 2016-01-31 to 2018-12-31
- Total data points :359,392

Assumptions:

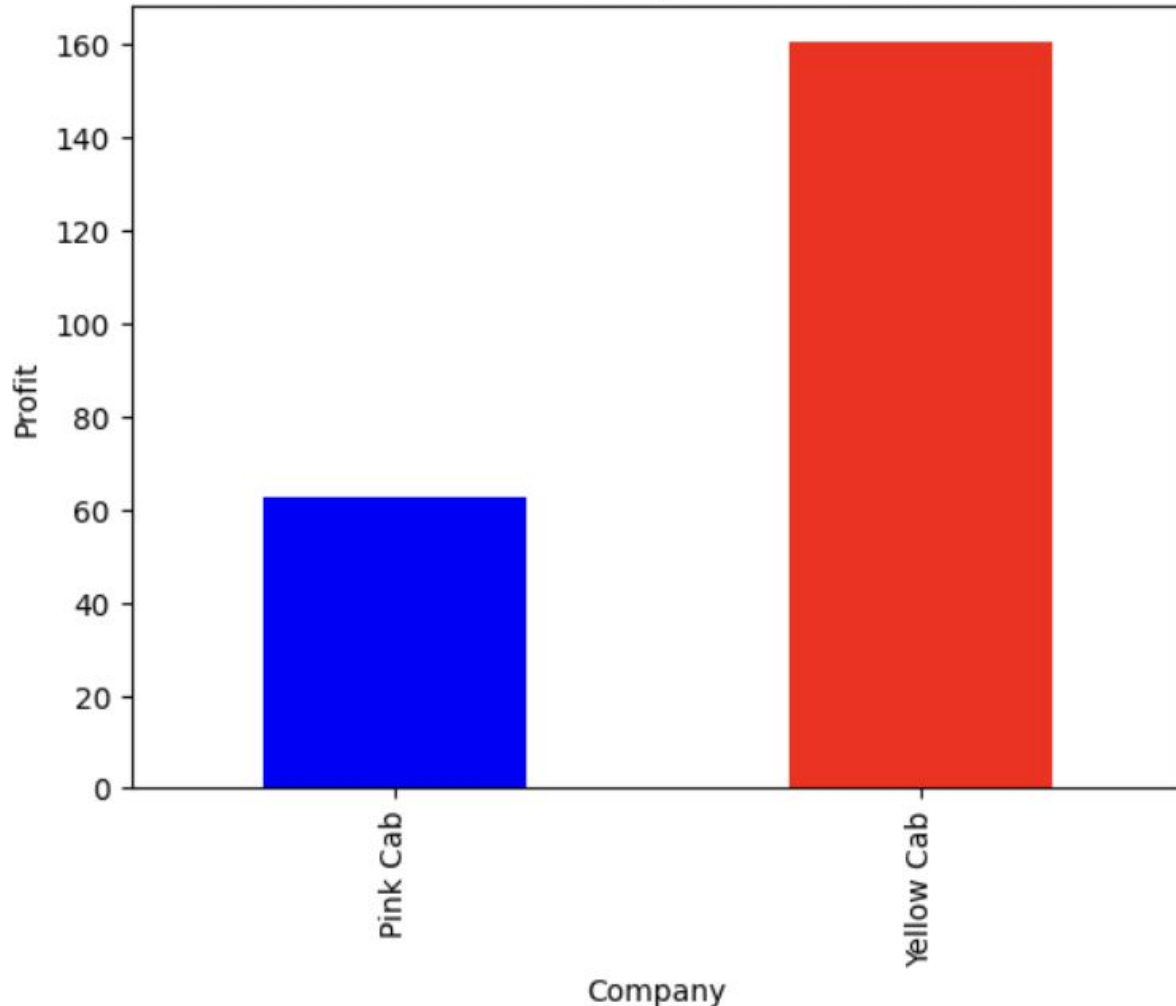
- Outliers and duplicates are omitted during the preprocessing process and is consider as missing values/bad data.
- Profit of rides are calculated keeping other factors constant and only Price_Charged and Cost_of_Trip features used to calculate profit.

Analytical methods:

- Hypothesis testing: 7 in total.
- To reject/keep null hypothesis, I used:
 - Structured analytical approach on key features such as total_rides, profit, features, cities, dates, data income
 - T-test statistical

Profit Analysis

Pink and Yellow Cab Companies vs Average Profits

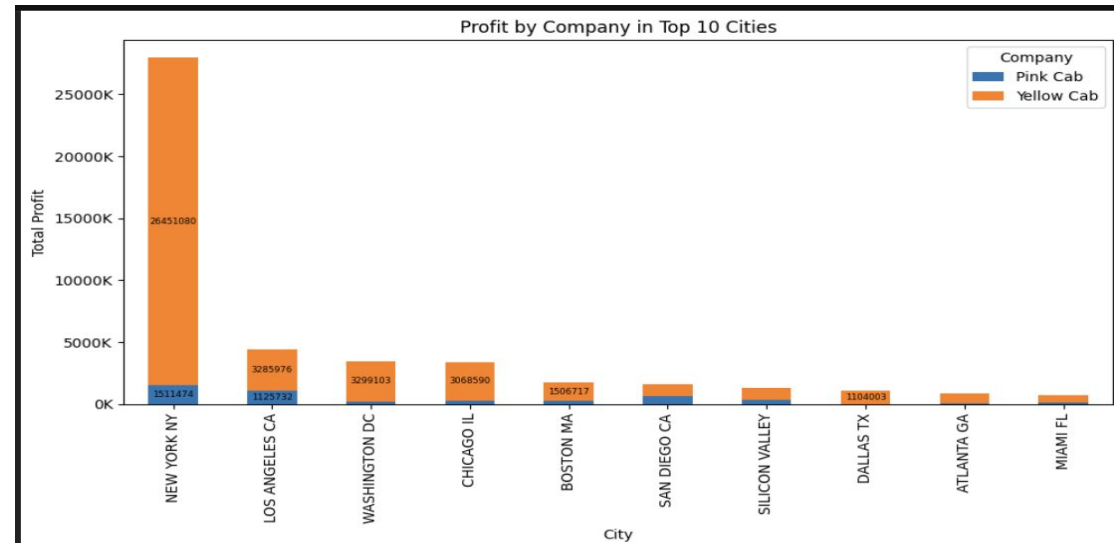
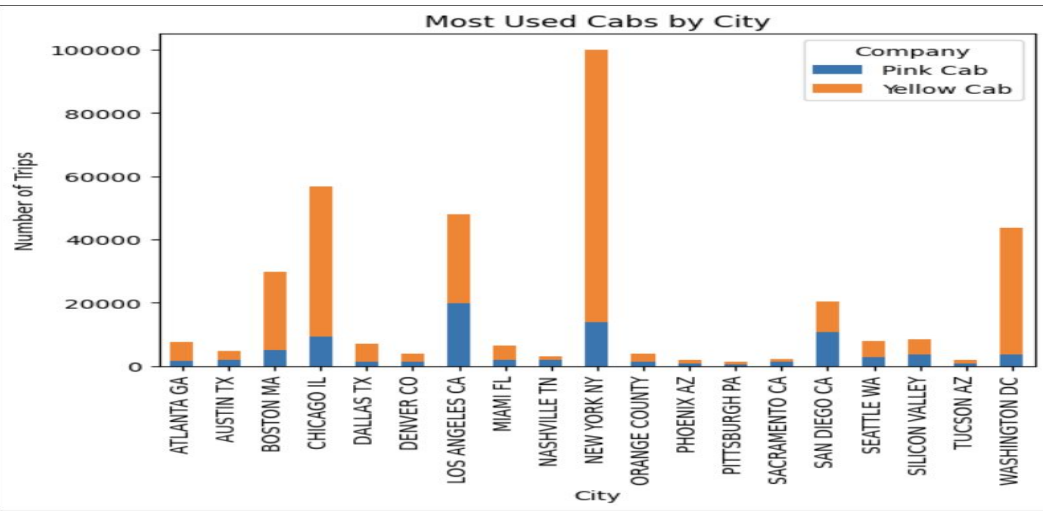


```
Total Rides, KM Travelled, and Profit by Company and Year:
Company  total_rides  total_km_travelled  total_profit
Pink Cab      84711      1911073.11  $5,307,328.32
Yellow Cab   274681      6199417.47  $44,020,373.17
```

Hypothesis 1: It is beneficial to use one cab company over the other.

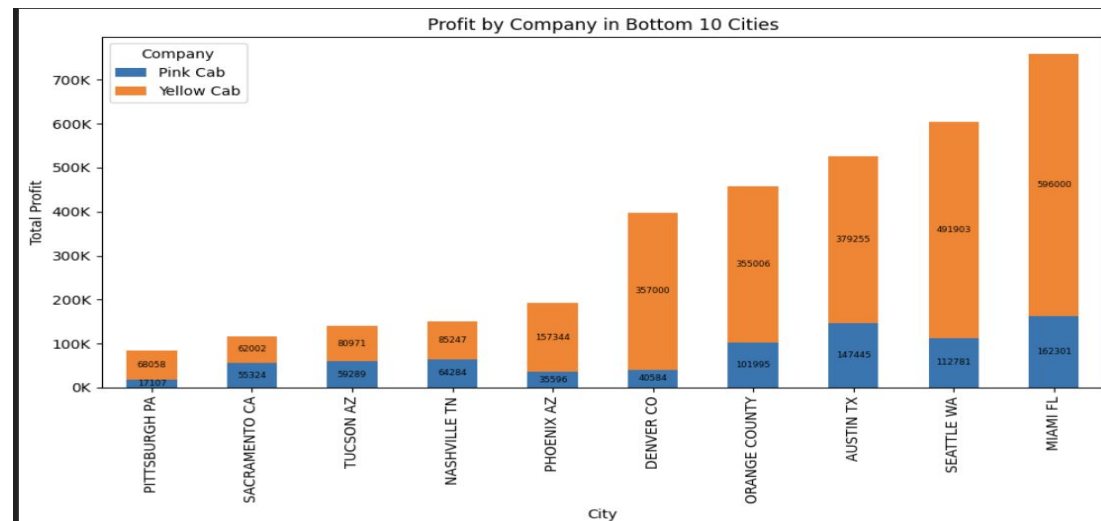
- The bar chart compares the average profit of each ride for two companies. It indicates that Yellow Cab has nearly triple the profit than that of the Pink Cab.
- The table summarizes the total_rides, total distance travelled, and total_profit by each company also follows the same trend that Yellow Cab makes more profit than the other.

Profit by City Analysis

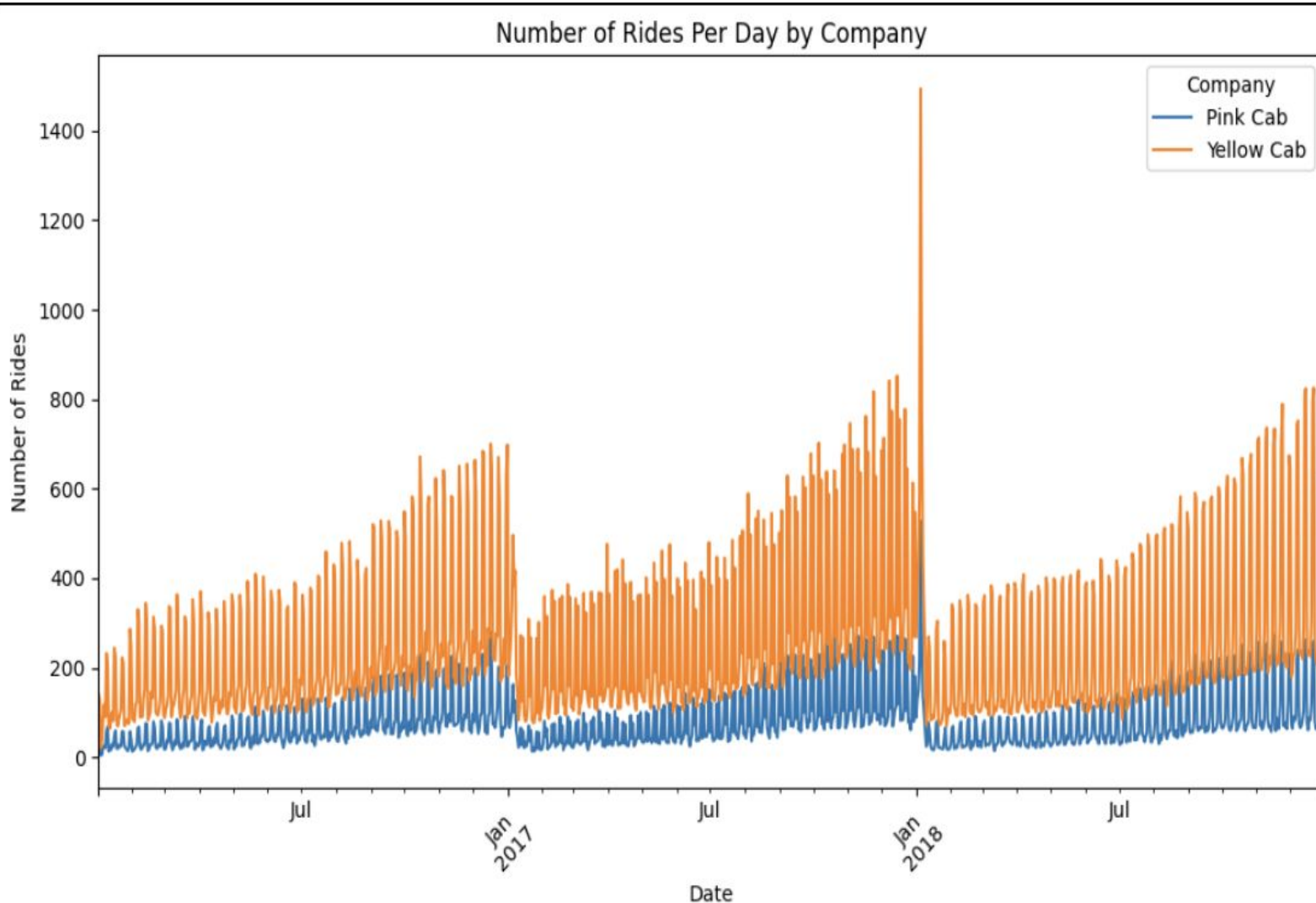


Hypothesis 2: Yellow Cab generates higher total profit than Pink Cab in every city

- The bar chart comparing the most used cars by city shows the dominant figures of Yellow Cab over the Pink one in most cities.
- A closer insight into two following charts compared the profit by each company in cities with most and least profits shows that a very significant proportion of the profit made comes from Yellow Cab.
- Yellow Cab Profit > Pink Cab Profit.



Total Rides Throughout 2016-2018

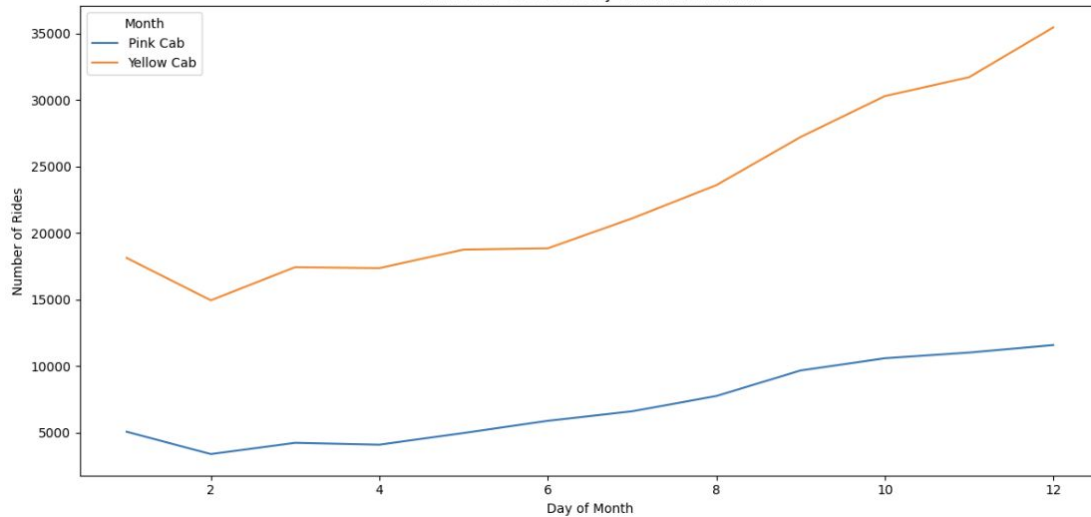


Hypothesis 3: Travelling demands fluctuates seasonally within the year.

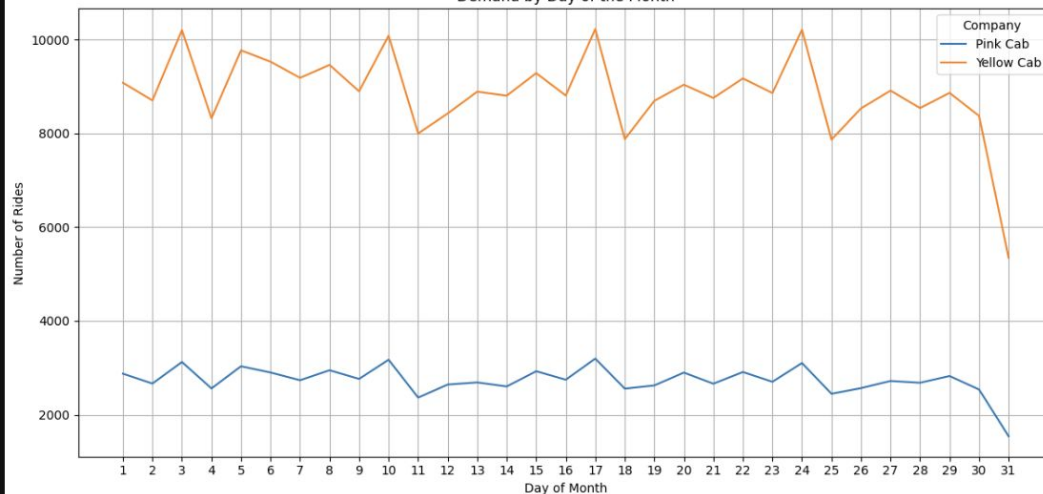
- This line chart compares the rides per day for each company within the time period between 2016-2018 suggests the constant increase trend within the year in both companies.
- The trend remains throughout three observed years.
- Yellow Cab has higher total number of rides than Pink Cab.

Ride Numbers Per Month and Per Day

Number of Rides Per Day within Each Month



Demand by Day of the Month



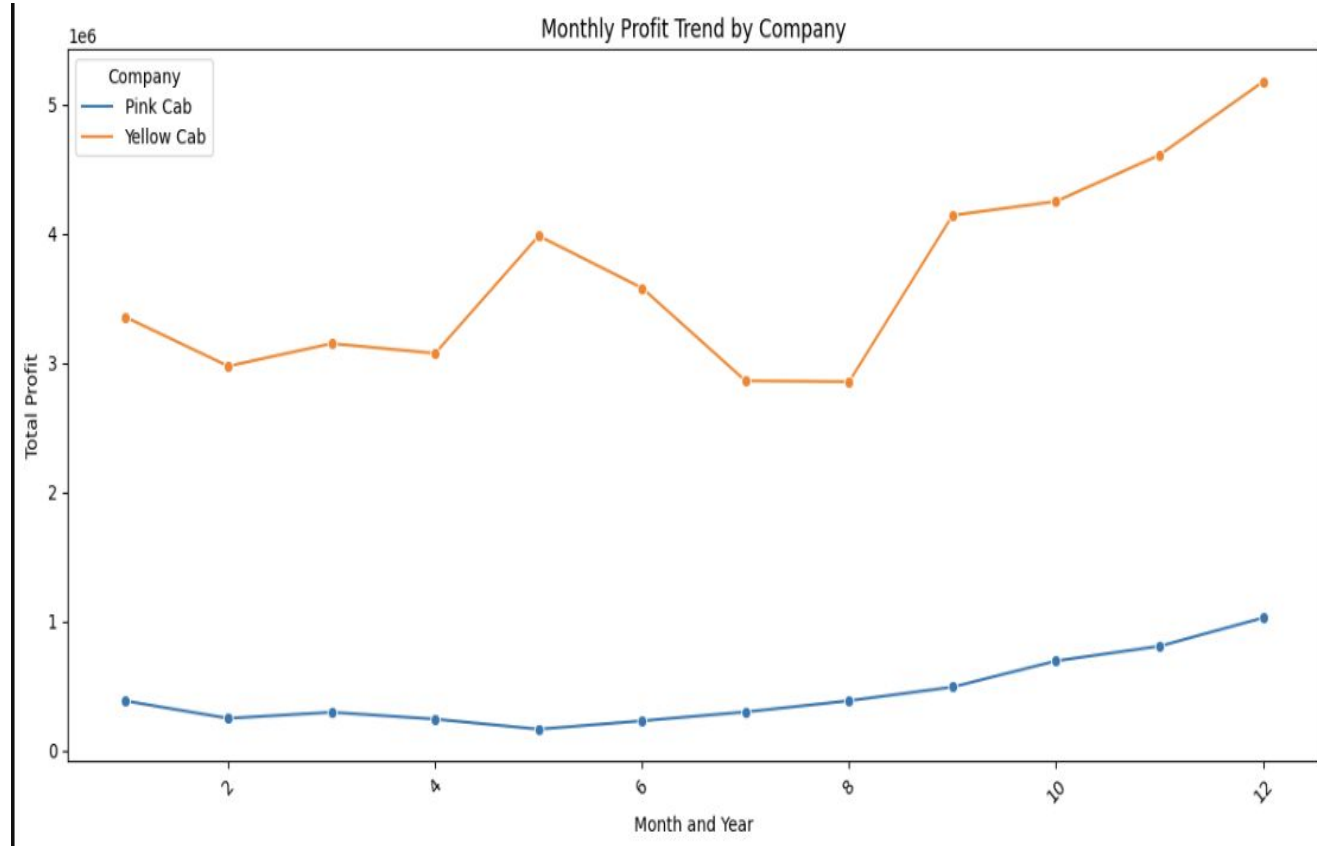
Hypothesis 4: There is a fluctuation in the demand for rides everyday.

- The top line graph showing the trend of the demand of rides by day of the month shows the continuous fluctuation in the demand for both companies to travel.
- The bottom line graph also suggests the repetitive pattern of travelling within the week in which the highest demand falls into the third day of the week, and plummets on the weekends.
- There is a cyclic pattern of the ride demand fluctuation by each day within the month.

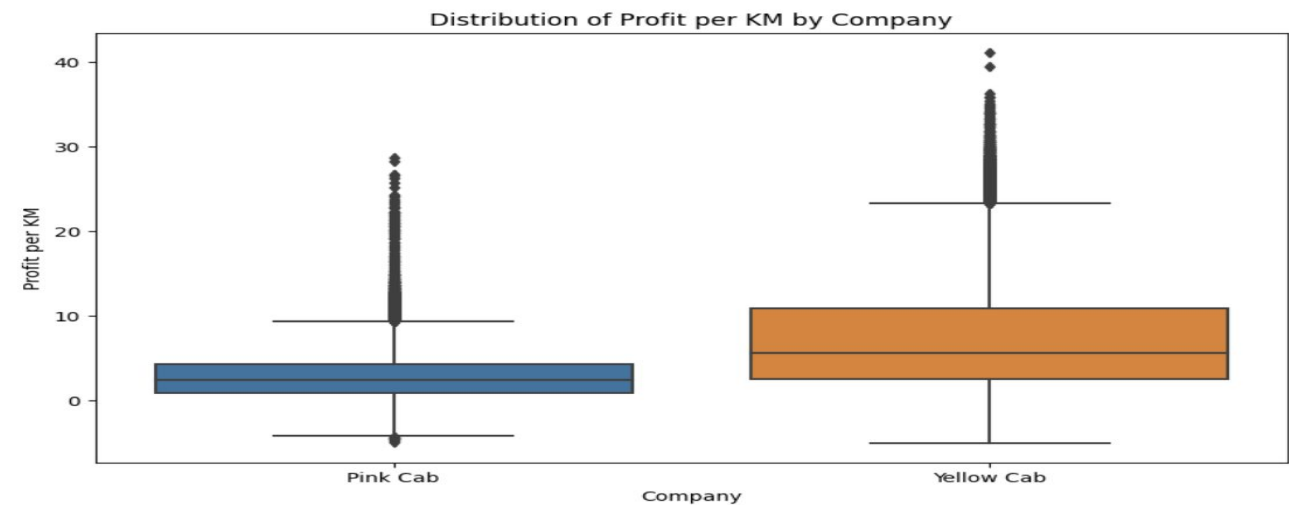
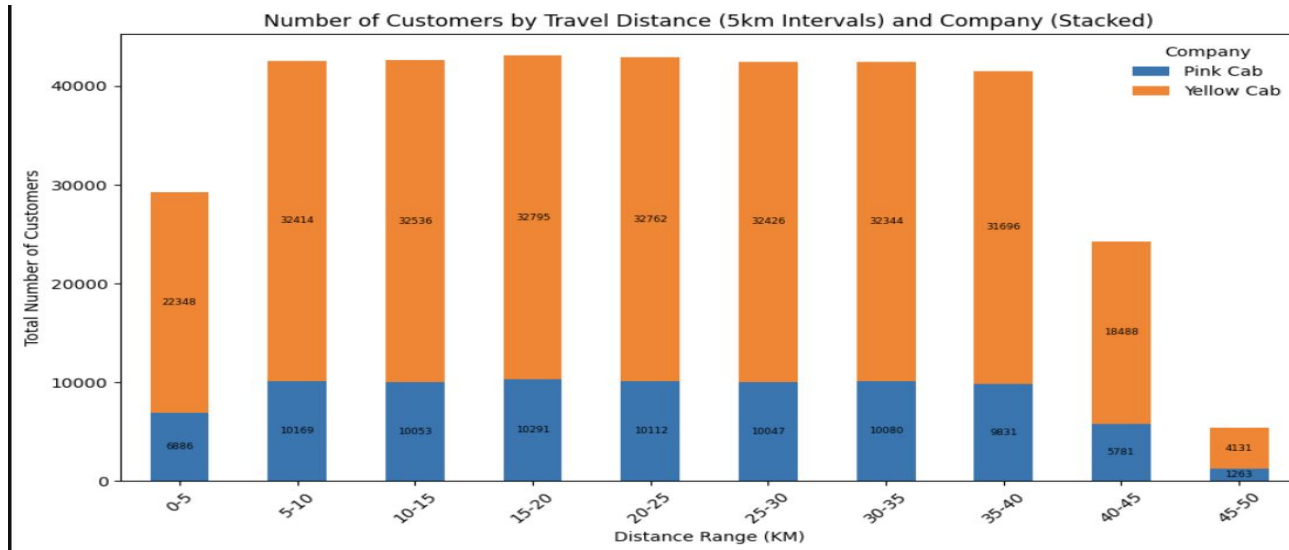
Profit Per Month

Hypothesis 5: There are fluctuations in the profit made in different months.

- The line chart showing the monthly profit trend by company shows that while the pink cab company has a steady increasing profit trend, yellow cab's profit fluctuates throughout the year.
- Both Pink and Yellow companies reach its profit peak at the end months of the year.
- Yellow Cab has a outlier fluctuation in May, with nearly equivalent profit as in September.
- The profit fluctuates throughout the year, with the most profitable months fall between October - December or May.



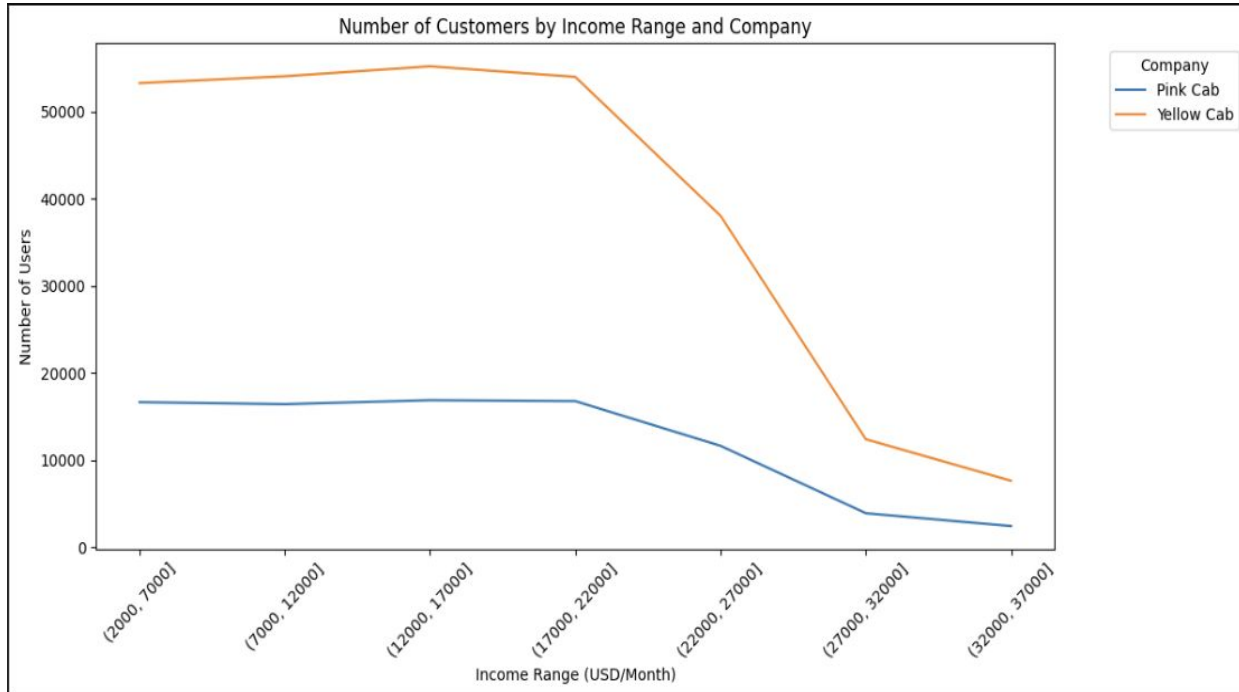
Customer Numbers, Travelled Distance, and Profit



Hypothesis 6: There are changing distributions of customers across the distance range.

- The stacked bar chart shows the steady customer counts (~32,000 for Yellow Cab and ~10,000 for Pink Cab) for customers travelling between 5-40 km.
- For very short distance (<5km) or long distances (40-50 km), customers are around four times likely to use Yellow Cab than Pink Cab.
- Per km travelled, Yellow Cab gains double or triple the profit than the Pink Cab.

City Wise Cab Users Covered By Company



Hypothesis 7: The average customer income is different between the companies

- The t-test analysis to compare customer income between Yellow Cab and Pink Cab shows $p = 0.67 > 0.05$. So we fail to reject the null hypothesis and indicate both companies have the similar income-level customers.
- The line chart shows the similar decreasing trend in the number of users as the income level increases.
- The sharp decline happens when the income surpasses 22,000 USD/ Month, suggesting the target group of customers around 2,000 - 22,000 income for both companies.

Recommendations

The performed analyses recommends the Yellow Cab over the Pink Cab for the following reasons:

1. **Profitable:** Yellow Cab has higher figures in terms of customers and profits than Pink Cab (around 4 times higher regarding number of customers, travelled distance, and average profit; around 8 times higher in terms of the total profit)
2. **Popularity:** Yellow Cab is far more popular in every cities analyzed, especially in big cities with high population and users.
3. **Consistent Customers:** Yellow Cab has better flow of customers throughout the years, months, and days.
4. **Peak Profit Period:** The travelling demand for Yellow Cab doubles that of Pink Cab, with the highest demand between October - December and May.
5. **Better Profit/Km:** Customers are four times likely to use Yellow Cab than Pink Cab, with double to triple profits gained for each km run by the former.
6. **Versatile Customer Group:** Most customers fall between the income range of 2,000-22,000. However, there needs to be more data to conclude the difference in the income level of the customers between two companies.

On the basis of the total profits and customer attraction, though, I will **recommend Yellow Cab** for investment.

Thank You



Data Glacier

Your Deep Learning Partner