



Data Glacier

Your Deep Learning Partner

G2M Case Study

Virtual Internship - 19 May 2023

Zohra Bouchamaoui

Agenda

- US Cab Industry Overview
- Client Overview and Objective
- EDA (1, 2, 3)
 - Datasets
 - Approach
 - Assumptions
- Profit Analysis
 - Income Classes
 - Gender
 - Age
 - KM Travelled
- Customer Loyalty
- Cab ride analysis – Holidays
- 2019 Forecasts - Profits
- Recommendations (1, 2)



US Cab Industry Overview

Market Overview:

The US cab industry plays a crucial role in providing transportation services to millions of people across cities and urban areas. Although the industry has faced disruption from ride-hailing services, traditional cab companies still hold a significant market share. The market is characterized by established players, regional operators, and emerging start-ups, each with its own strengths and challenges.

Key Drivers and Challenges:

Investors should consider several factors when evaluating investment opportunities in the US cab industry. These include:

- 1. Consumer Demand:** Analysing consumer demand for cab services is crucial, considering changing preferences and the competition from ride-hailing platforms. Factors such as population density, urbanization, and commuting patterns should be considered.
- 2. Competitive Landscape:** Assessing the competitive landscape is crucial, considering the presence of established cab companies and the growing dominance of ride-hailing platforms. Differentiating factors such as service quality, pricing, and customer loyalty programs should be evaluated.
- 3. Technological Advancements:** The integration of technology has become increasingly important for cab companies to remain competitive. Evaluating investments in companies that embrace technological innovations, such as mobile apps for booking and ride tracking, can be advantageous.
- 4. Regulatory Environment:** The regulatory landscape surrounding the cab industry varies across states and cities. Understanding local regulations and their potential impact on business operations is essential.

Client Overview and Objective

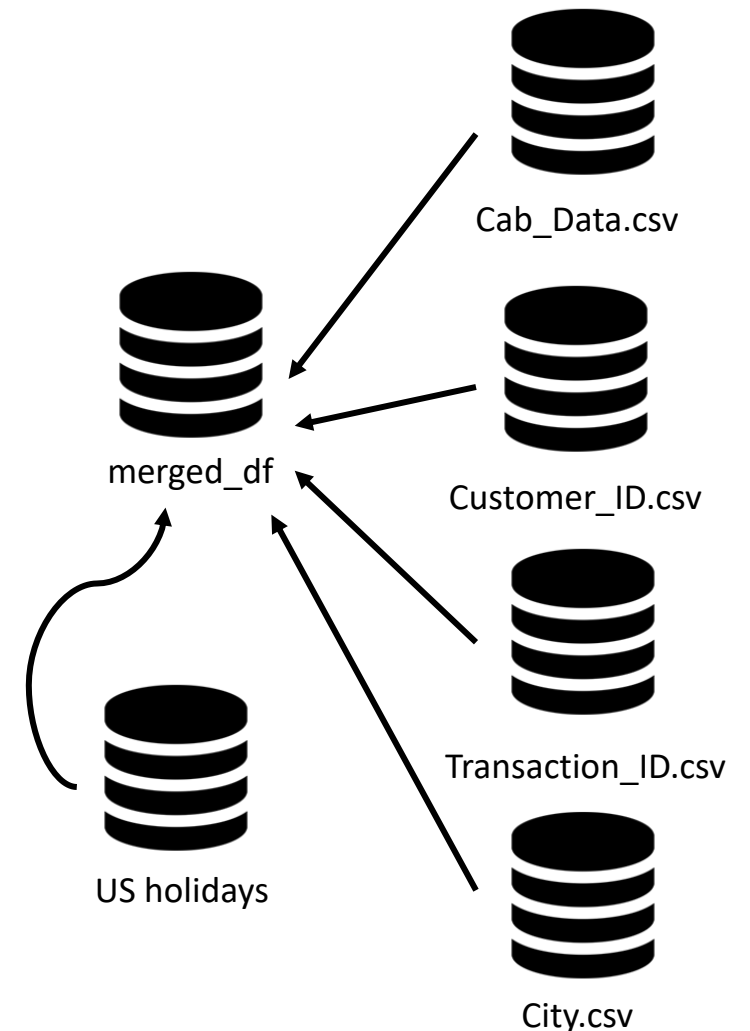
- XYZ is a private firm in the US seeking to invest in the cab industry. With remarkable growth and multiple key players in the market, XYZ aims to understand the market landscape and customer profiles before making a final investment decision. They have engaged a team of experts to conduct a comprehensive analysis using extensive data sets and advanced analytical techniques.
- The objective is to provide actionable insights and strategic recommendations to help XYZ evaluate risks, identify opportunities, and develop a well-defined investment strategy. The findings will be presented to XYZ's Executive team to guide their decision-making process and ensure alignment with their long-term growth objectives.

Explorative Data Analysis (1)

Datasets:

We have been provided 4 individual data sets. Time period of data is from *31/01/2016 to 31/12/2018*.

- **Cab_Data.csv** – this file includes details of transaction for 2 cab companies.
- **Customer_ID.csv** – this is a mapping table that contains a unique identifier which links the customer's demographic details.
- **Transaction_ID.csv** – this is a mapping table that contains transaction to customer mapping and payment mode.
- **City.csv** – this file contains list of US cities, their population and number of cab users.
- The **holidays** dataset in the holidays package provides information about various holidays observed in different countries, including the United States.
- The merged dataset (**merged_df**) has 359,392 rows and 21 columns (features) – including 6 derived features.



Explorative Data Analysis (2)

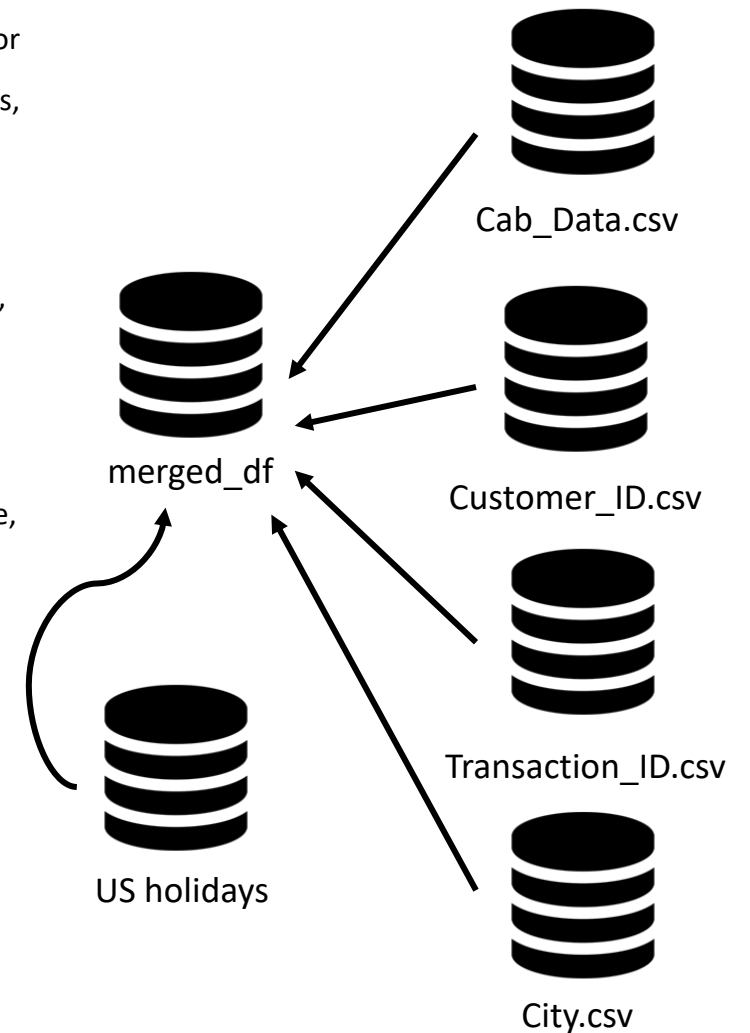
Approach:

The case study focuses on providing actionable insights for a cab investment firm, XYZ, to help them identify the right company for investment in the cab industry. The project involves analysing multiple datasets related to cab companies, customer profiles, transactions, and city information.

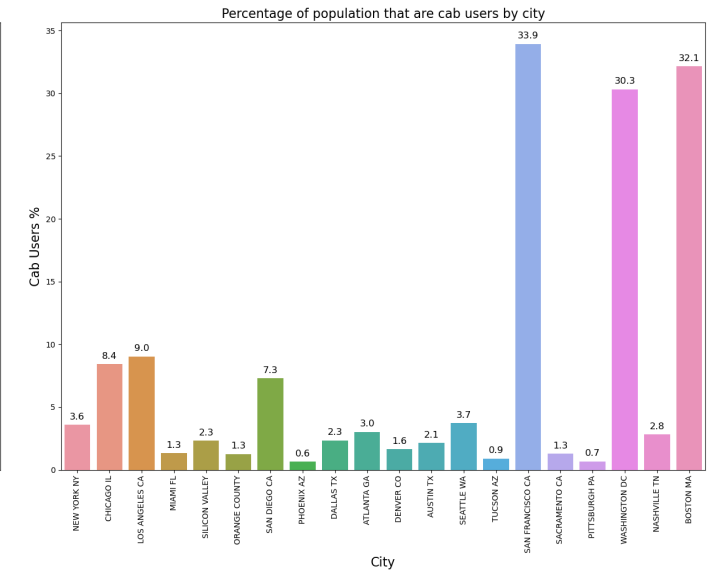
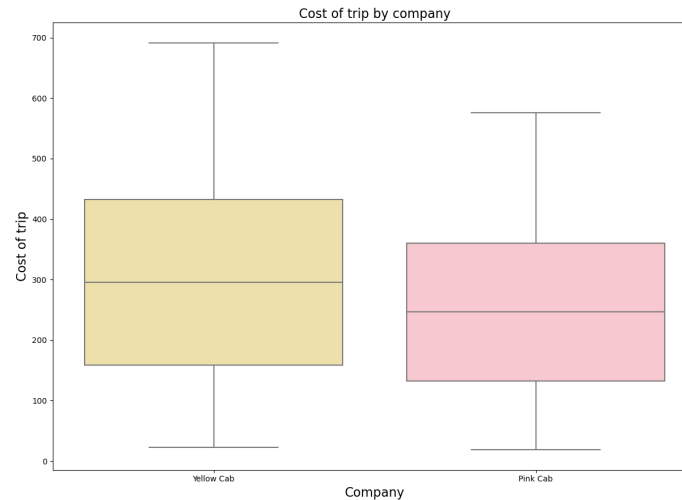
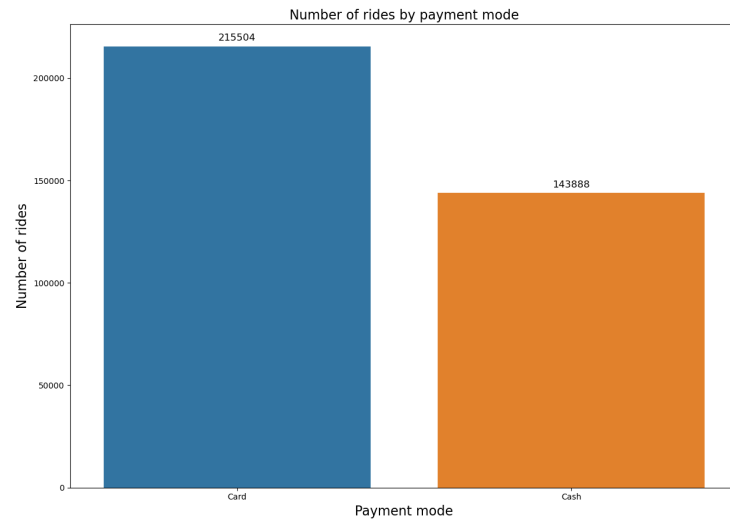
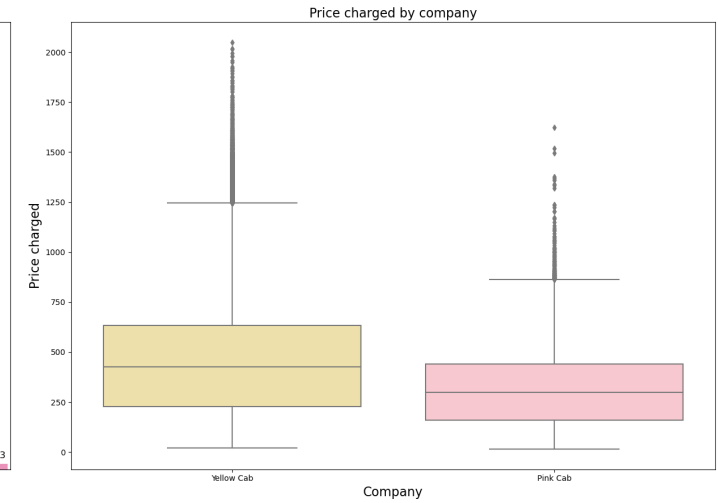
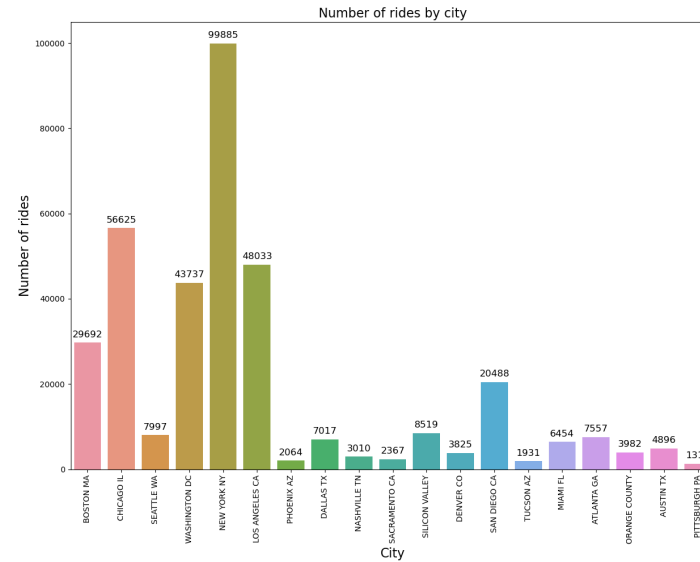
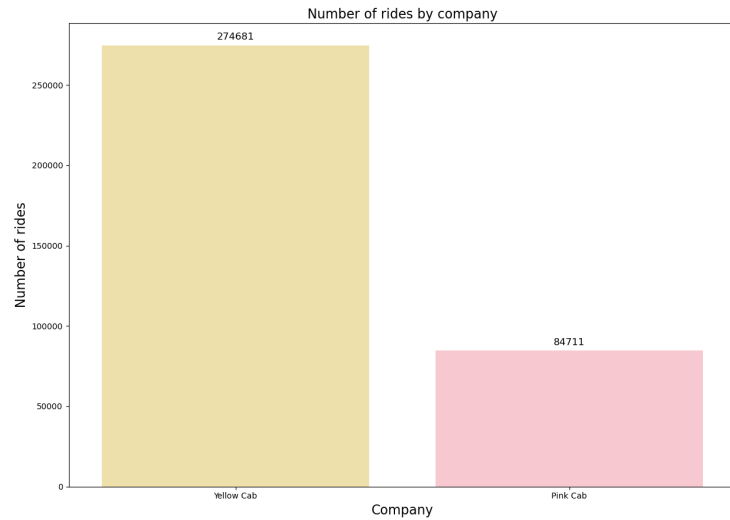
- **Data exploration phase** involves examining the datasets to understand their structure, check for missing values, and obtain statistical summaries. Data cleaning and reformatting are performed, including modifying date formats, filling missing values, and merging datasets based on common keys.
- After cleaning the individual datasets, they are merged into one dataset (merged_df).
- **Visualizations** are then created to gain insights. These visualizations include the number of rides by company, payment mode, and city, as well as the average price charged and cost of the trip by company. The profits made per trip are also calculated and added as a new column in the dataset.

Assumptions:

- Price charged has 237 outliers but for the purpose of this case study we will not be treating them as outliers.
- Profits are calculated using the 'Cost of trip' and the 'Price Charged'.

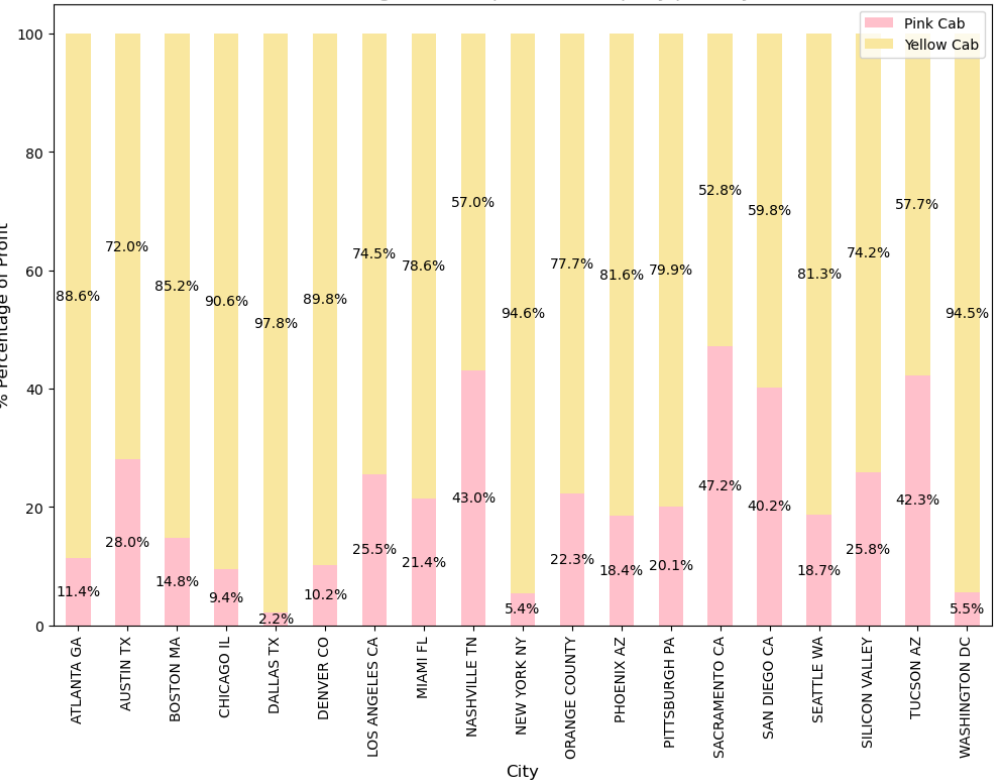


Exploratory Analysis (3)



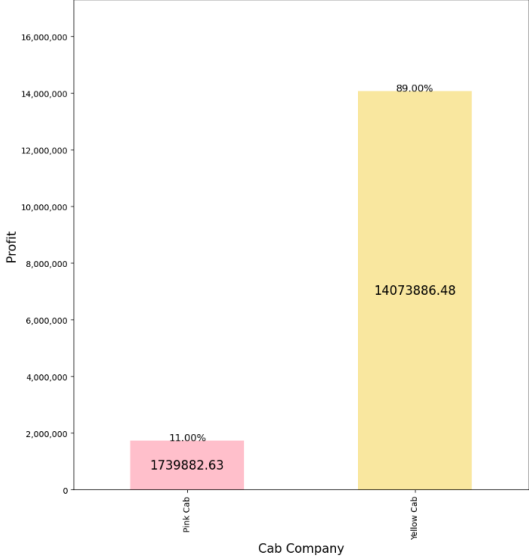
Profit Analysis

Percentage of Profit per Cab Company per City

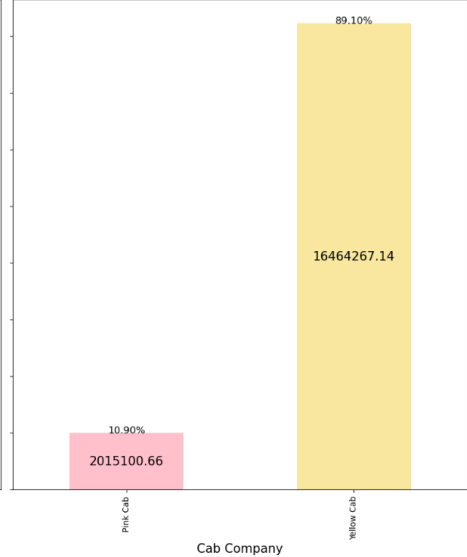


Company	Total Profit	Number of rides	Average profit per ride
Yellow	44,020,373.17	2,746,81	160.26
Pink	5,307,328.32	84,711	62.65

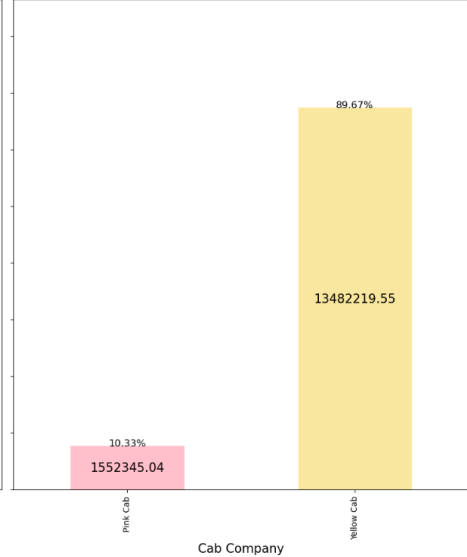
Profit Analysis by Cab Company - 2016



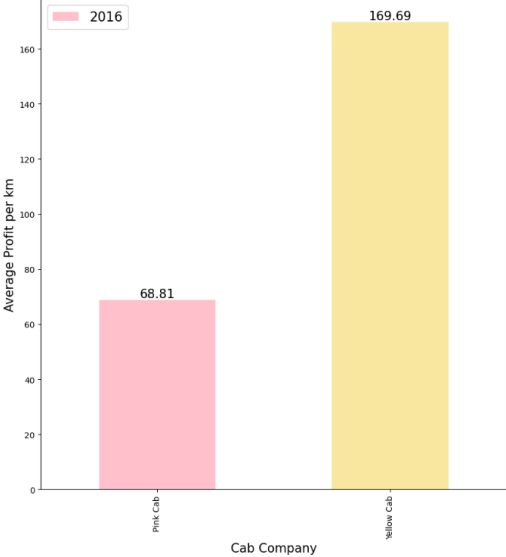
Profit Analysis by Cab Company - 2017



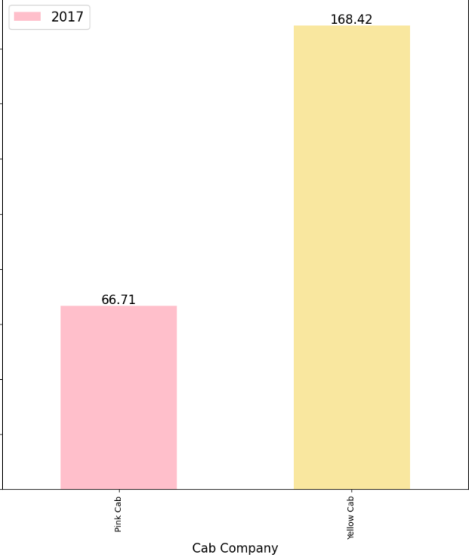
Profit Analysis by Cab Company - 2018



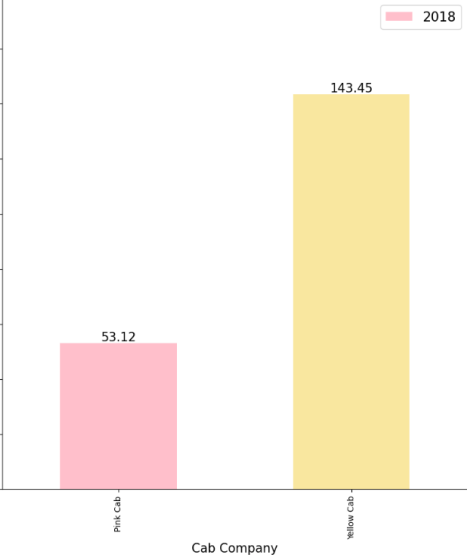
Average Profit per km by Cab Company - 2016



Average Profit per km by Cab Company - 2017



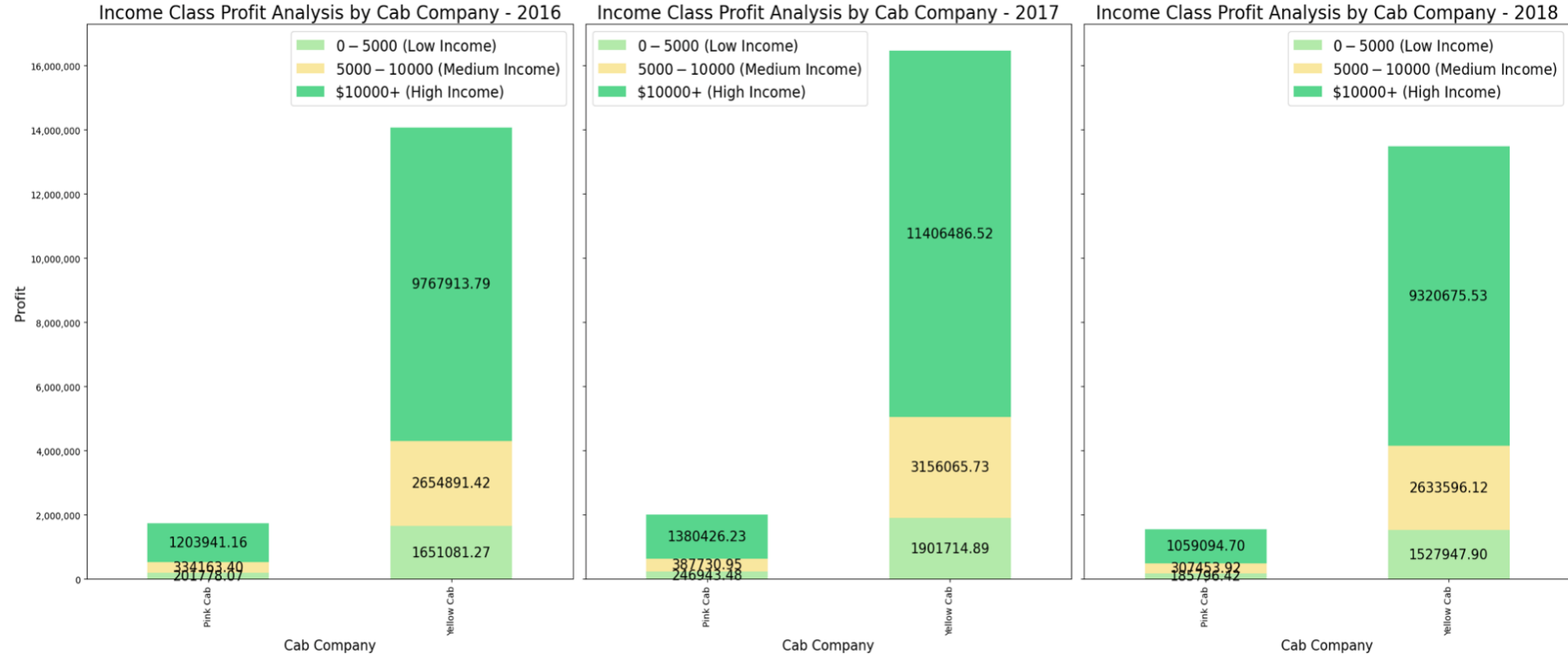
Average Profit per km by Cab Company - 2018



Profit Analysis – Income Classes

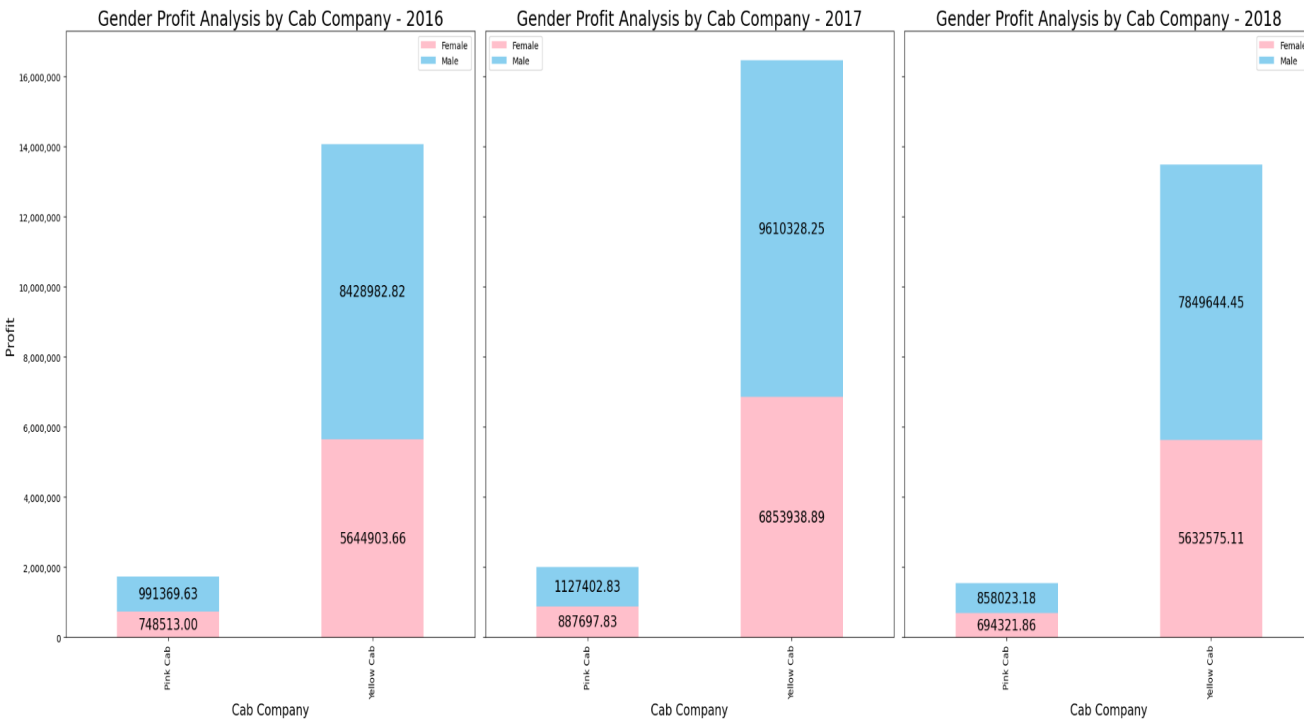
We have 3 income classes: Low: 0 - \$5,000, Medium: \$5,000 - \$10,000, and High: \$10,000+. For both the yellow and pink companies, most profit comes from customers with high income. The profits coming from the high-income customers is almost triple the amount coming from medium income customers.

For both cab companies, the high-income users contribute most to their profits.

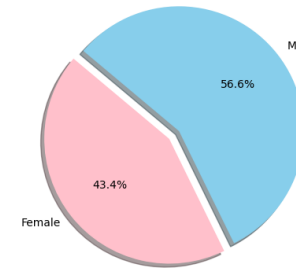


Profit Analysis - Gender

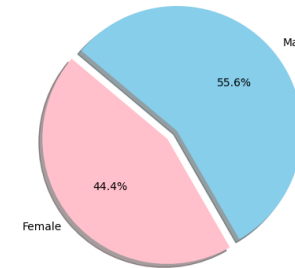
The gender distribution for each company is almost equal for both the yellow and pink cab companies, with a slightly higher number of male users. Both companies made their highest yearly profits in 2017, and their lowest profits in 2018. We can see that despite the number of male and female being almost equal, there is a slight difference in the percentage of male users being higher.



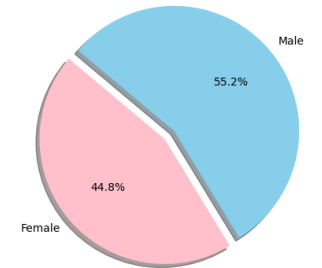
Gender Distribution for Pink Cab in 2016



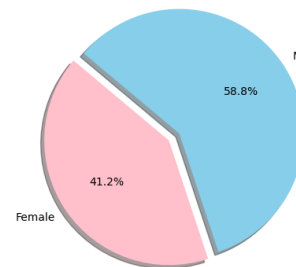
Gender Distribution for Pink Cab in 2017



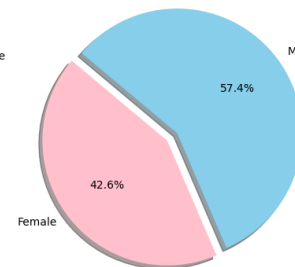
Gender Distribution for Pink Cab in 2018



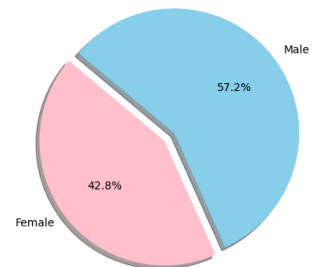
Gender Distribution for Yellow Cab in 2016



Gender Distribution for Yellow Cab in 2017

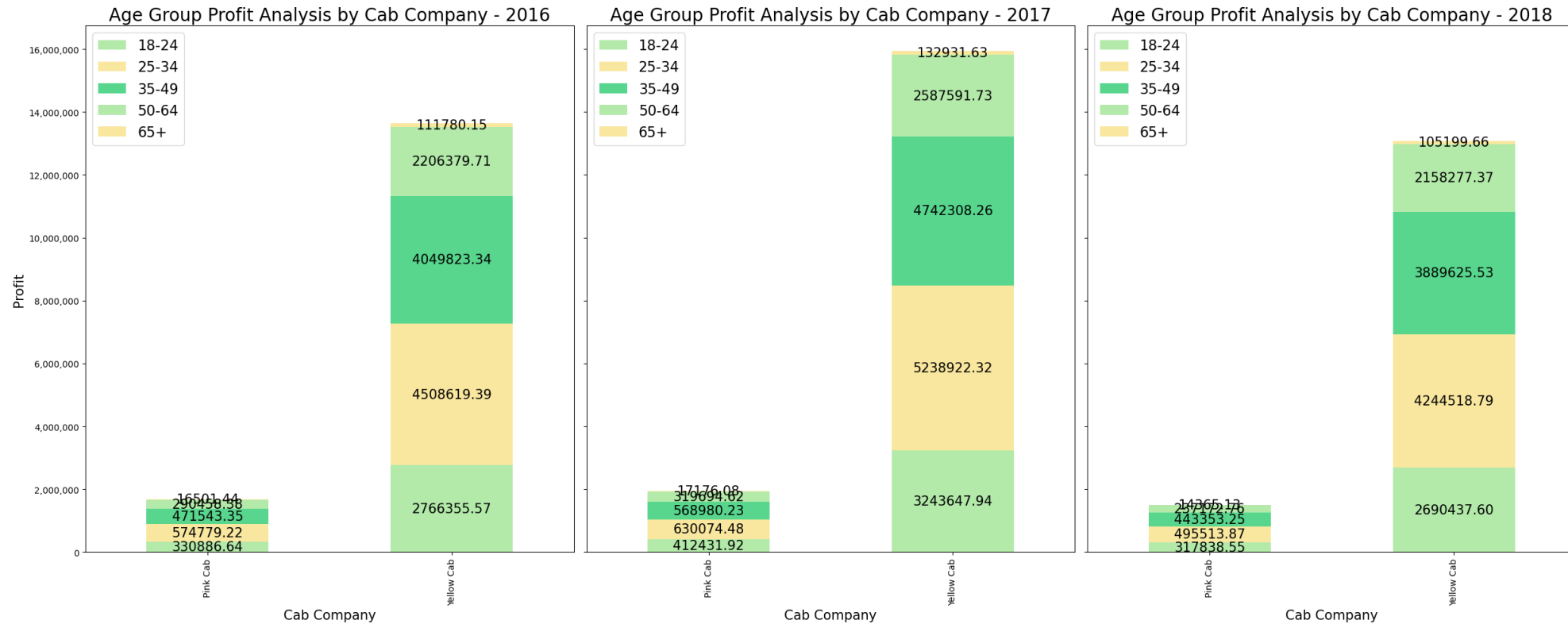


Gender Distribution for Yellow Cab in 2018



Profit Analysis - Age

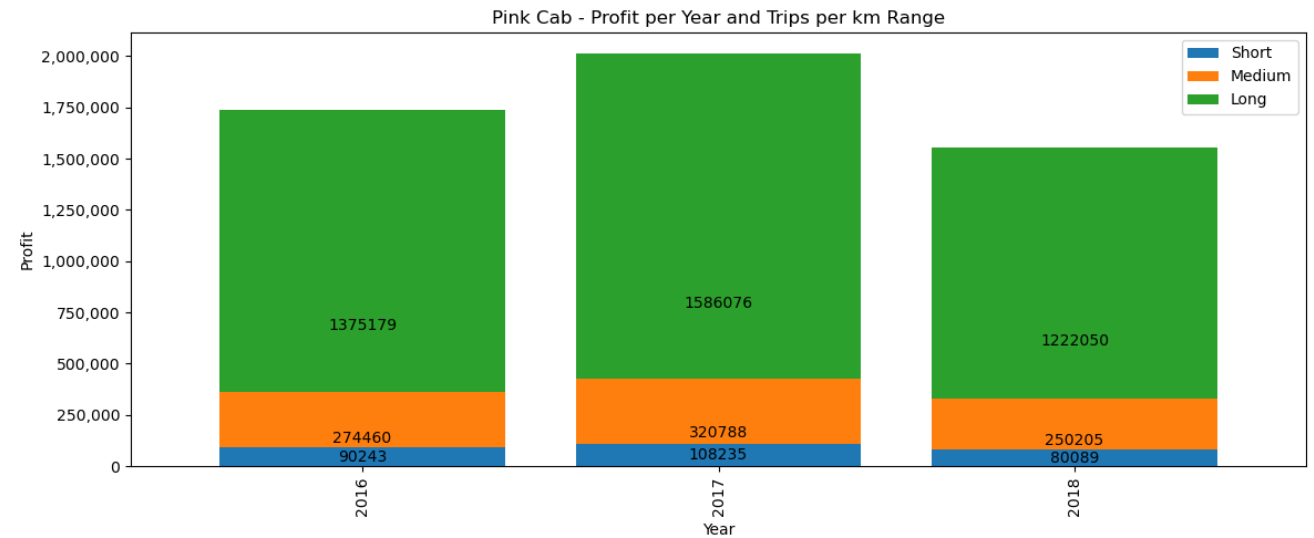
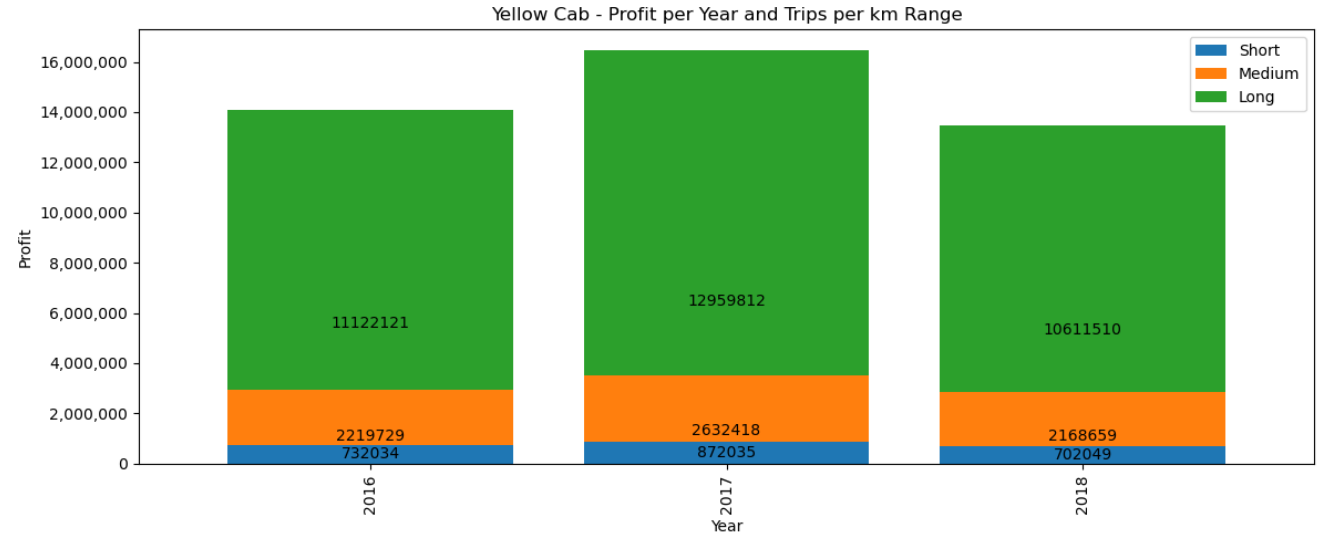
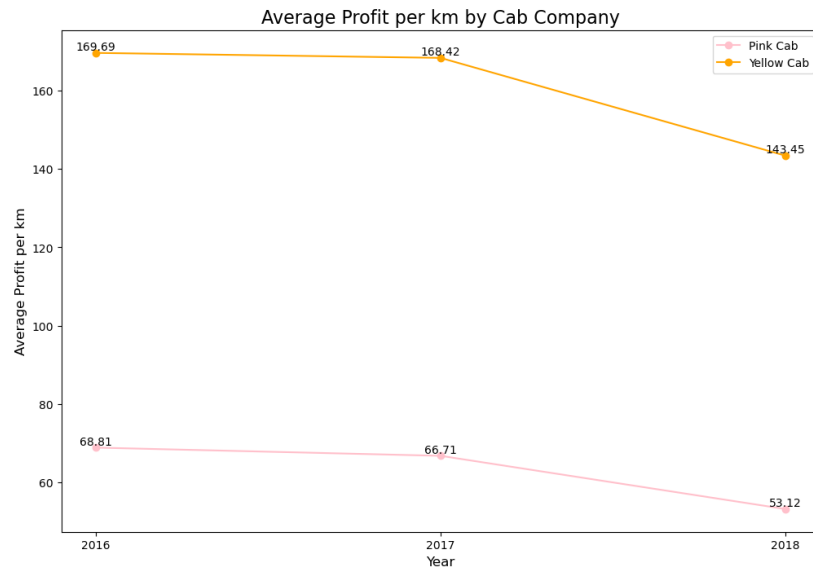
We have created five age groups: 18-24, 25-34, 35-49, 50-64, and 65+. For both companies, the highest profits come from both the 25-34 and 35-49 age groups. The 65+ age group shows the least profit for both companies. The distribution of profit for each age group has been almost constant over the years.



Profit Analysis – KM Travelled

We have created three KM travelled ranges: short, medium, and long. We can see that the average profit per KM travelled for each company has decrease slightly between 2016 and 2017, to a drastic drop from 2017 to 2018.

It is also clear that the longer the trips, the more profits are made by the cab companies.

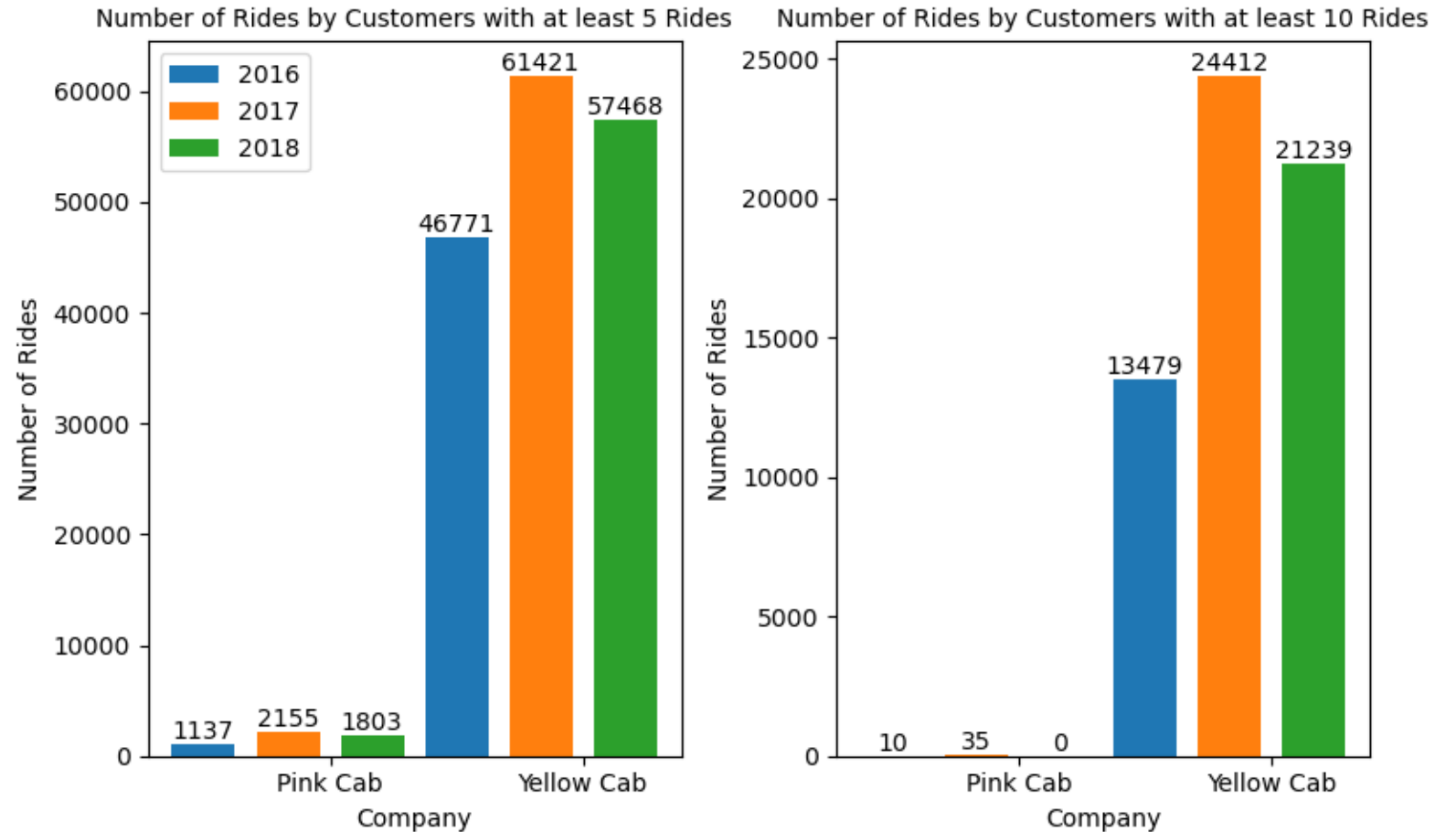


Customer loyalty

These two plots show the the number of customers that have taken at least 5 and 10 rides with each company for each year.

It appears that the yellow company has a higher number of customers with at least 5 and 10 rides than the pink company. The company shows small numbers of customers with at least 5 rides, and almost no customer with at least 10 rides – with 10 in 2016, 35 in 2017, and 0 in 2018.

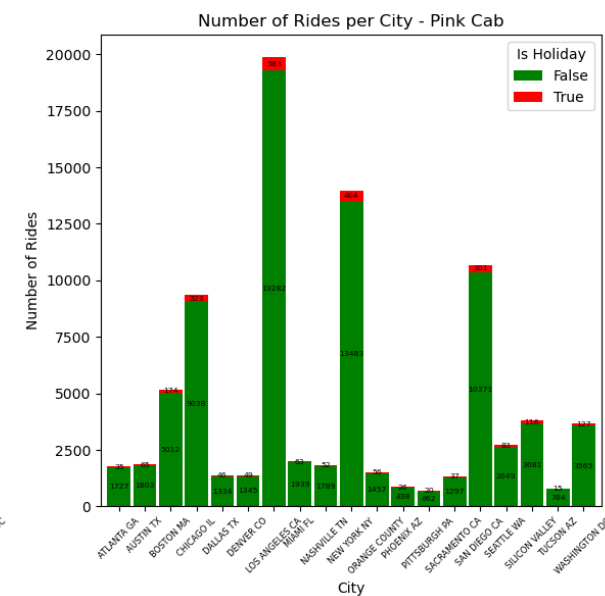
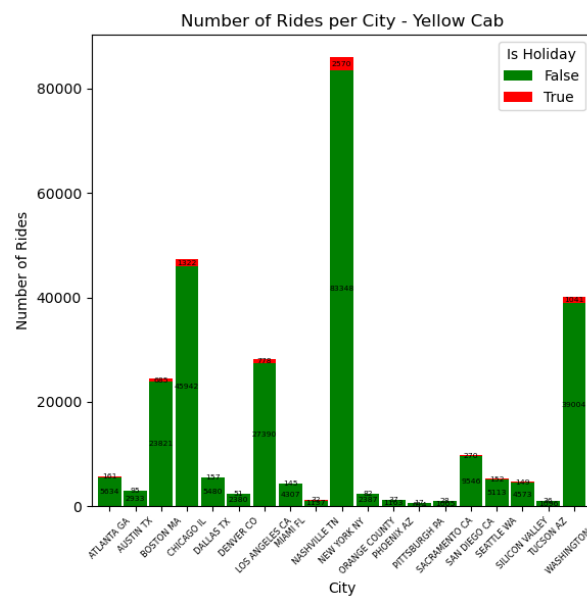
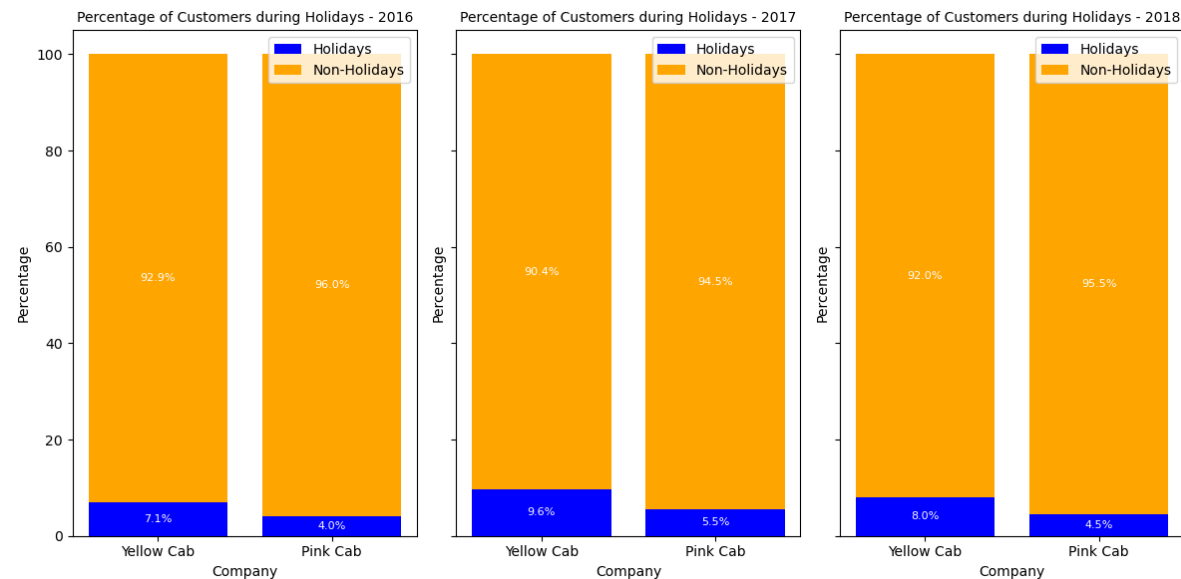
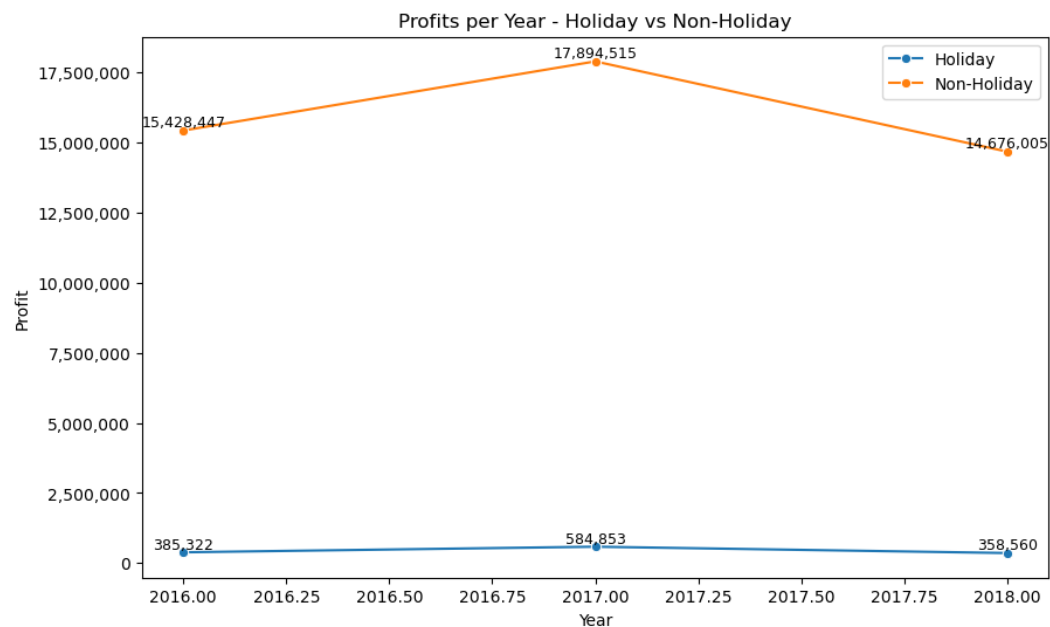
This shows that the Yellow Company has a better customer loyalty than the pink company and that they might have a higher customer satisfaction.



Cab ride analysis - Holidays

The percentage of cab riders during the holidays for both companies is relatively low, meaning that their analysis might not be as pertinent as the one of the previous features.

As we can see, the profits for non-holidays cab rides are significantly higher than the ones on holiday, which is not usually what would be expected.

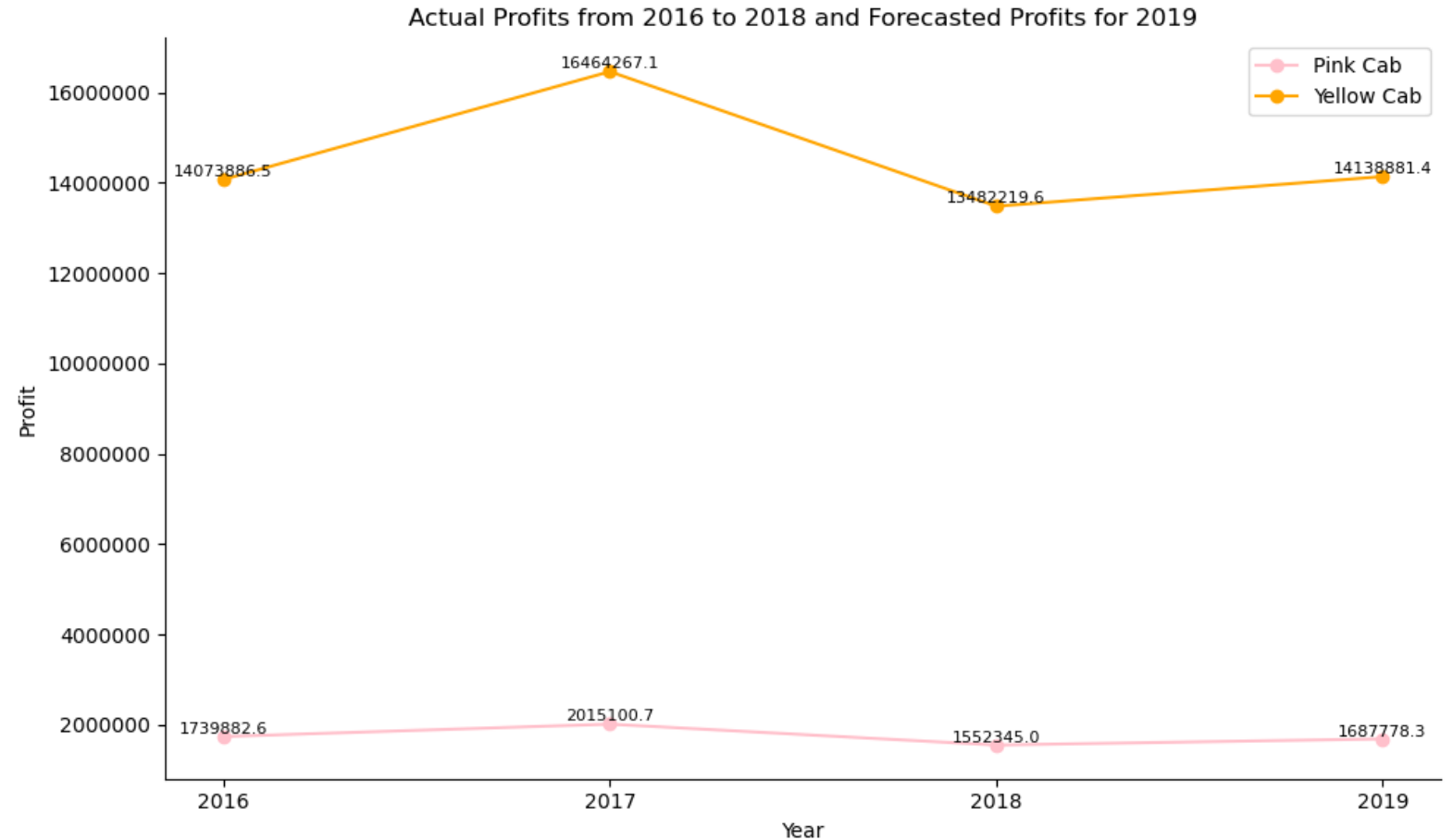


2019 Forecasts - Profits

Using the ARIMA model, we have forecasted the profits for both the yellow and pink companies in 2019.

It is predicted that there will be an increase of profits for both companies, 8.72% for the Pink Company and 4.87% for the Yellow Company. Despite an increase in profit from 2018 to 2019, both companies are doing worse than they did during the 2016 – 2018 time period.

As previously seen in other plots, the yellow company is performed significantly better than the pink company.



Recommendations (1)

- **Focus on Yellow Cab's Profitability:** The Yellow Cab company has consistently shown higher profits compared to the Pink Cab company. It is recommended to allocate more resources and investments towards the Yellow Cab company to capitalize on its success and maximize profitability.
- **Exploit Growth Opportunities in Top Cities:** Identify the top-performing cities such as Boston, Chicago, Washington DC, New York, and Los Angeles, which have the highest number of rides. Focus on expanding operations in these cities and capture a larger market share by improving service quality, optimizing driver availability, and investing in marketing and promotions.
- **Target Cities with High Cab User Percentages:** In addition to the cities with the highest ride numbers, consider cities like San Francisco, Washington DC, and Boston that have a high percentage of the population as cab users. These cities show potential for a larger customer base and increased revenue generation.
- **Optimize Marketing and Pricing Strategies:** Conduct market research and analysis to identify customer preferences, price sensitivity, and competitor strategies. Utilize this information to fine-tune marketing campaigns, target specific customer segments effectively, and optimize pricing strategies to maximize profits.
- **Address Declining Profit per Kilometre:** The analysis indicates a decline in average profit per kilometer traveled between 2016 and 2018. Investigate the factors contributing to this decline and implement measures to improve profitability. This may involve optimizing operational efficiency, reducing costs, or exploring alternative revenue streams.
- **Re-evaluate Holiday Strategy:** As the analysis suggests, the percentage of cab riders during holidays is relatively low, and profits tend to be lower compared to non-holiday periods. Re-evaluate the holiday strategy and consider reallocating resources or adjusting pricing strategies during these periods. Explore potential opportunities to increase demand and profitability during holidays through targeted promotions, special offers, or unique services.

Recommendations (2)

- **Target High-Income Customers:** Both Yellow Cab and Pink Cab companies generate a significant portion of their profits from high-income customers. To further increase profits, it is advisable to develop strategies to attract and retain high-income customers. This can include offering premium services, loyalty programs, and personalized experiences that cater to their needs.
- **Focus on Profitable Customer Segments:** Analyse the profitability of different customer segments, such as age groups and income classes. Allocate resources and marketing efforts towards the age groups and income classes that generate the highest profits for both Yellow Cab and Pink Cab. Tailor services and promotional activities to cater to their specific needs and preferences.
- **Enhance Customer Loyalty Programs and Capitalize on Customer Satisfaction and Loyalty:** Leverage the higher customer loyalty observed in Yellow Cab compared to Pink Cab. Further strengthen customer loyalty by introducing or improving loyalty programs that reward frequent riders, such as discounts, exclusive offers, and personalized perks. This can help increase customer retention and foster long-term relationships with riders. Invest in driver training programs, improve service quality, and ensure a seamless and enjoyable customer experience. Positive word-of-mouth and customer recommendations can help attract new customers and strengthen the brand reputation.

Thank You



Data Glacier

Your Deep Learning Partner