Data Glacier
Your Deep Learning Partner

Exploratory Data Analysis
G2M insight for Cab Investment firm

Gideon Osei Bonsu

Agenda

Data Glacier
Your Deep Learning Partner

Executive Summary

- The two cab companies have a lot of potential for growth.

- The companies have a much larger presence in Washington DC, San Francisco, and Boston, serving over 30% of the population in these cities.

- The Yellow Cab company has secured greater profits.

- The Pink Cab company has had a better average profit trend since 2018.

- The Yellow Cab company has better customer retention.

Data Glacier
Your Deep Learning Partner

Problem Statement

- XYZ is a private firm in the US. Due to remarkable growth in the cab industry in the last few years, and multiple key players in the market, it is planning for an investment in the cab industry. As per their Go-to-Market (G2M) strategy, they want to understand the market before taking any further steps.

- Having been provided with multiple datasets that contain information on 2 cab companies from 31/01/2016 to 31/12/2018, XYZ is interested in using actionable insights derived from my analysis to help them identify the right company to commit to investing in.

Approach Taken

The steps below were taken to examine the data:

- **Ask**: In this phase, I posed crucial questions regarding the purpose of the analysis and the desired outcomes from the customer.

- **Prepare**: This is when I imported the necessary libraries and loaded the datasets.

- **Process**: At this stage, I performed data cleaning, which involved identifying and handling null values and empty cells, standardizing column names, and consolidating datasets to create a unified and comprehensive dataset. Additionally, numeric date values (e.g., 54690) were transformed into a date format using January 1, 1900 as the reference date.

- **Analyze**: During this phase, I examined the different columns, employing various analytical techniques to extract meaningful insights from the data.

- **Share**: This stage ran in tandem with the "Analyze" stage, I disseminated visual representations of different segments of the data analysis.

- **Act**: In this phase, I offered recommendations based on the insights gleaned from my data analysis.
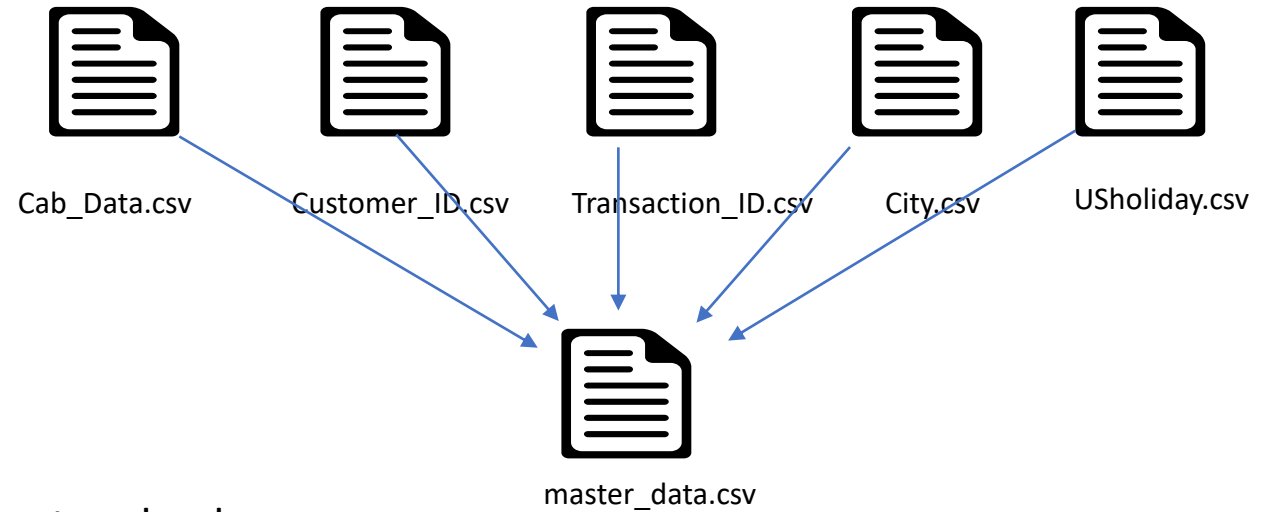
Exploratory Data Analysis

Master Data contains:

- 18 Features (including 4 derived features).
- Timeframe of the data: 2016-01-31 to 01/29/2019.
- Total data points: 359,392.

Cab_Data.csv   Customer_ID.csv   Transaction_ID.csv   City.csv   USholiday.csv

master_data.csv

Assumptions:

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

- The initially provided end date and start date were not in alignment. To rectify this, the end date was amended to 01/29/2019, as opposed to the originally specified 2018-12-31. This adjustment was necessary due to discrepancies in the numerical date values provided in the dataset, which did not correspond with 2018-12-31 as the end date. This disparity arises from the indicated start date of 31/01/2016 when converted to Date format. Additional elaboration can be found in the Jupyter notebook.

Customer Preference Analysis

Cab Company Share of All Journeys

Pink Cab — 23.57%

Yellow Cab — 76.43%

2016-01-31 to 2019-01-29
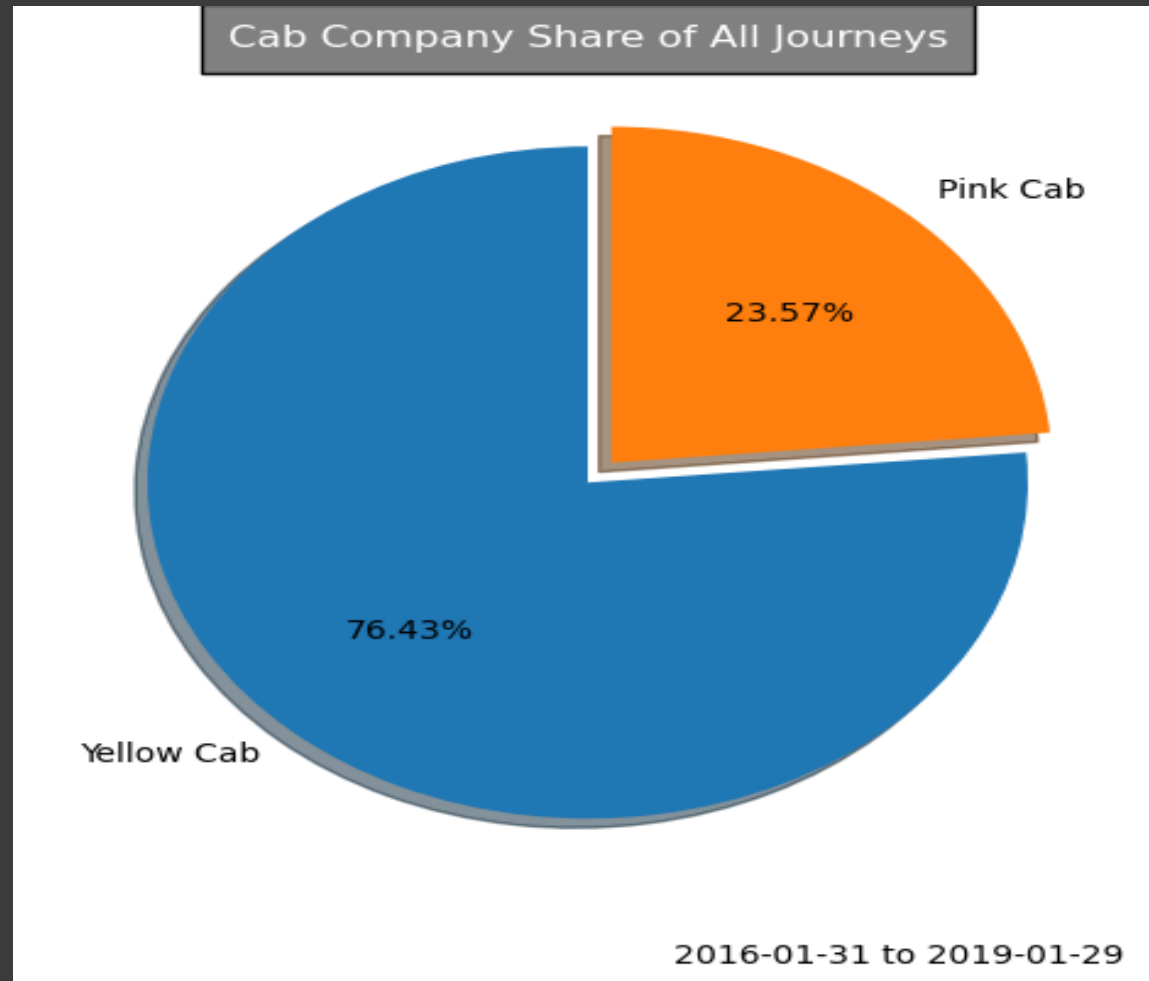
- The Yellow Cab is significantly more popular among customers compared to the Pink Cab.

- Out of the total 359,392 transactions in the dataset, the Yellow Cab accounted for 274,681, constituting a dominant 76.43% of all cab journeys.

- In contrast, the Pink Cab only had 84,711 journeys, making up a much smaller 23.57% of the total.
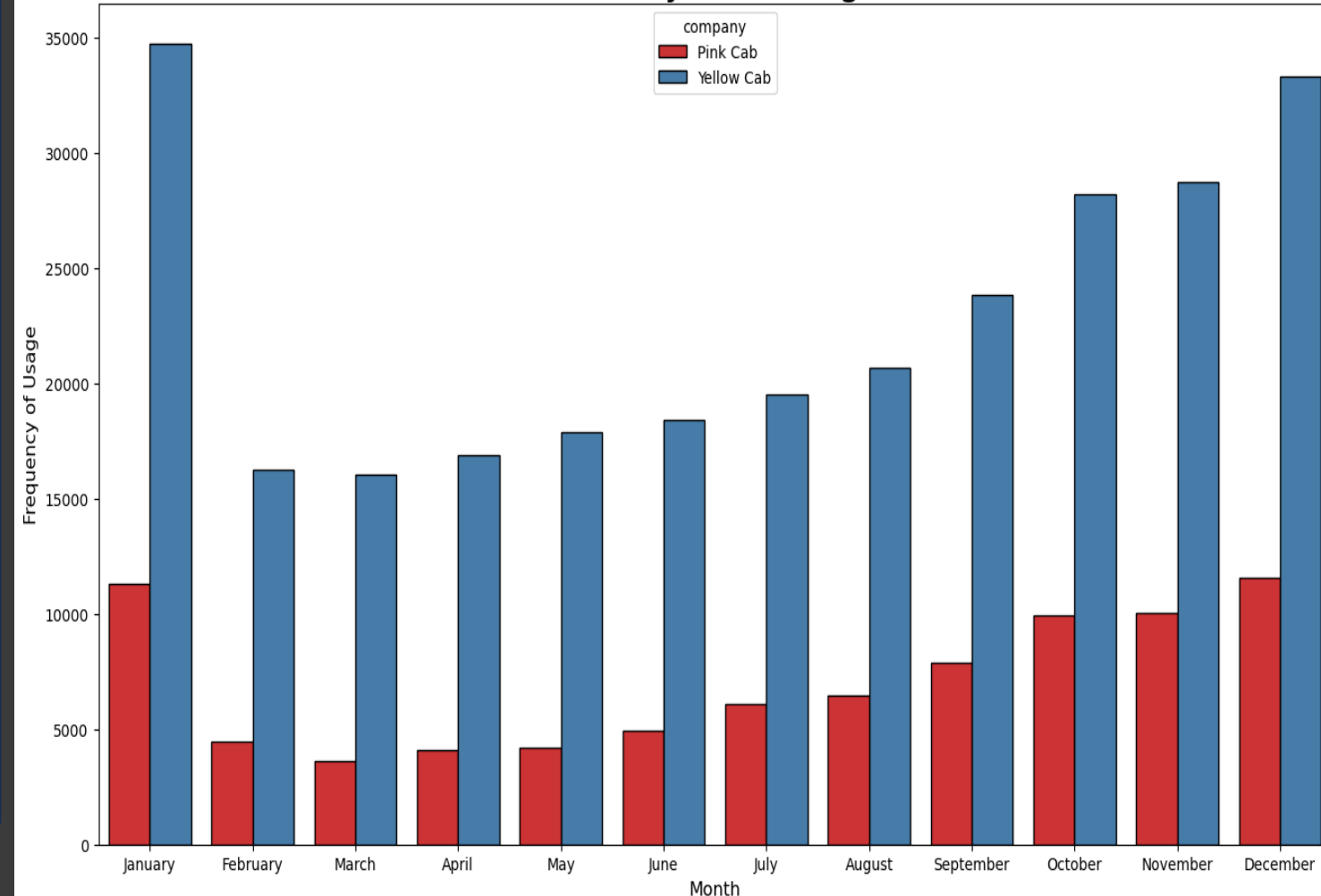
Seasonality Analysis

Yearly Cab Usage
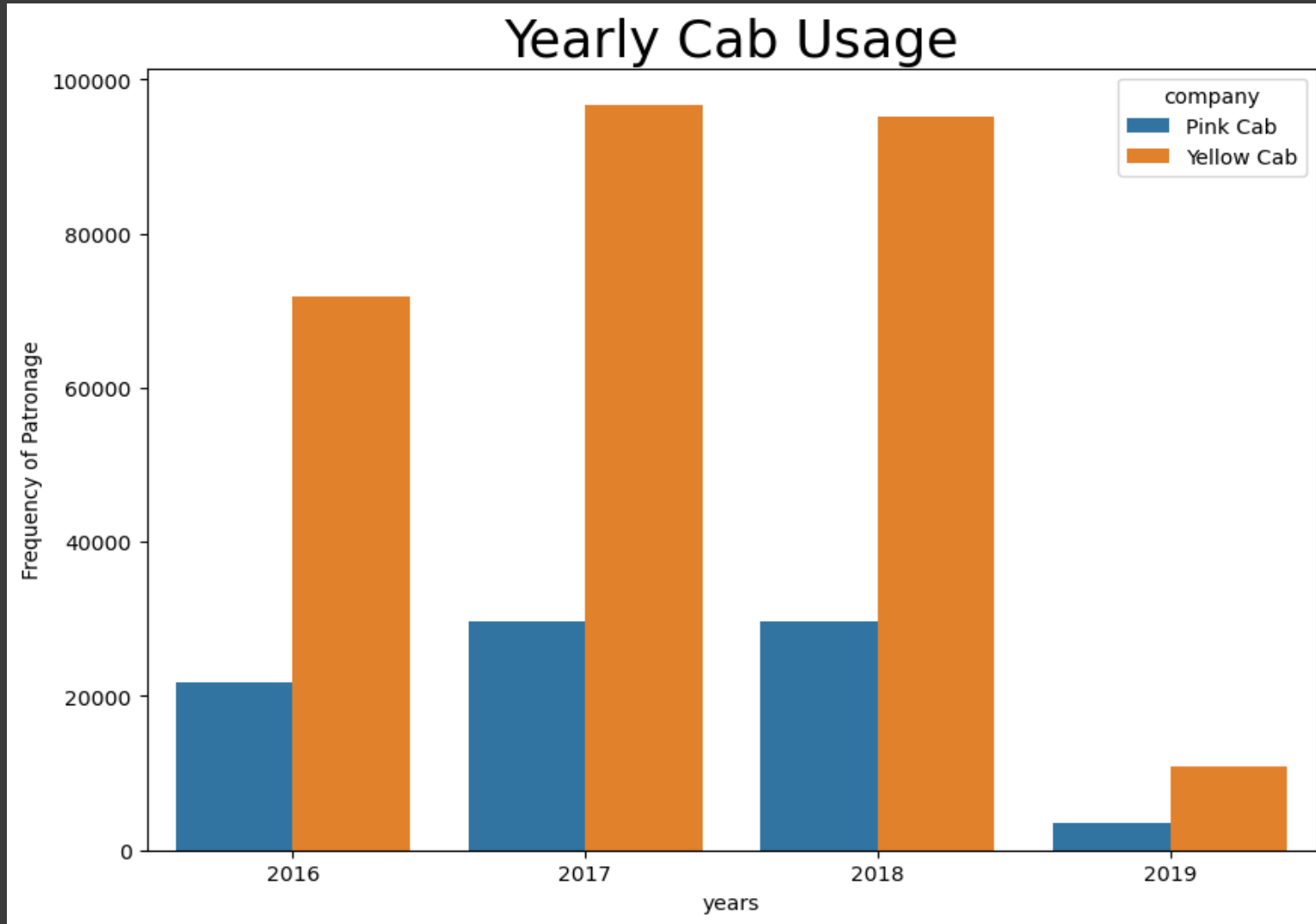
- Despite the Yellow Cab having a lot more transactions, its number of transactions saw a 1.49% decrease from 96,626 in 2017 to 95186 in 2018.

- The Pink Cab only had a 0.07% decrease from 29,750 in 2017 to 29,730 in 2018.
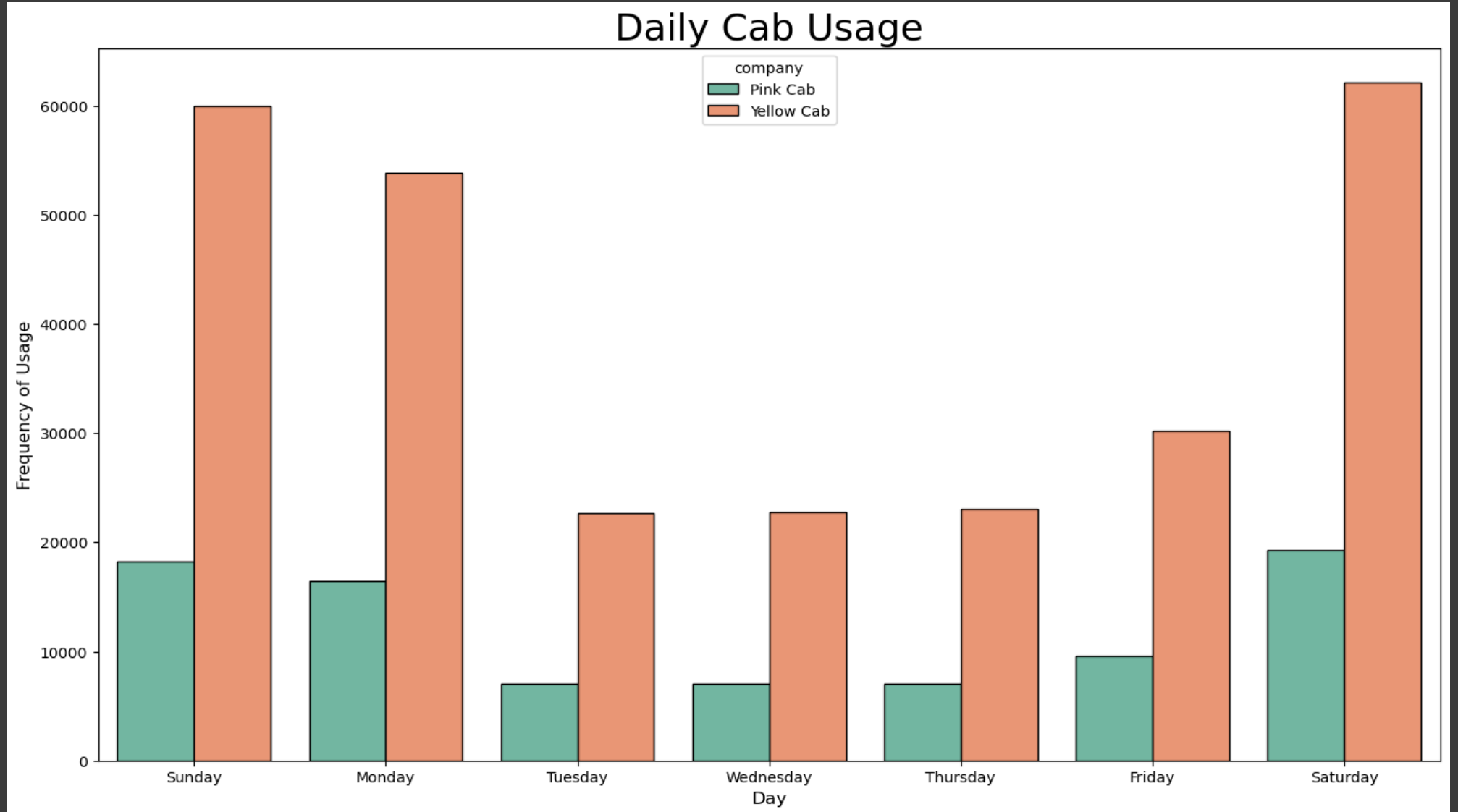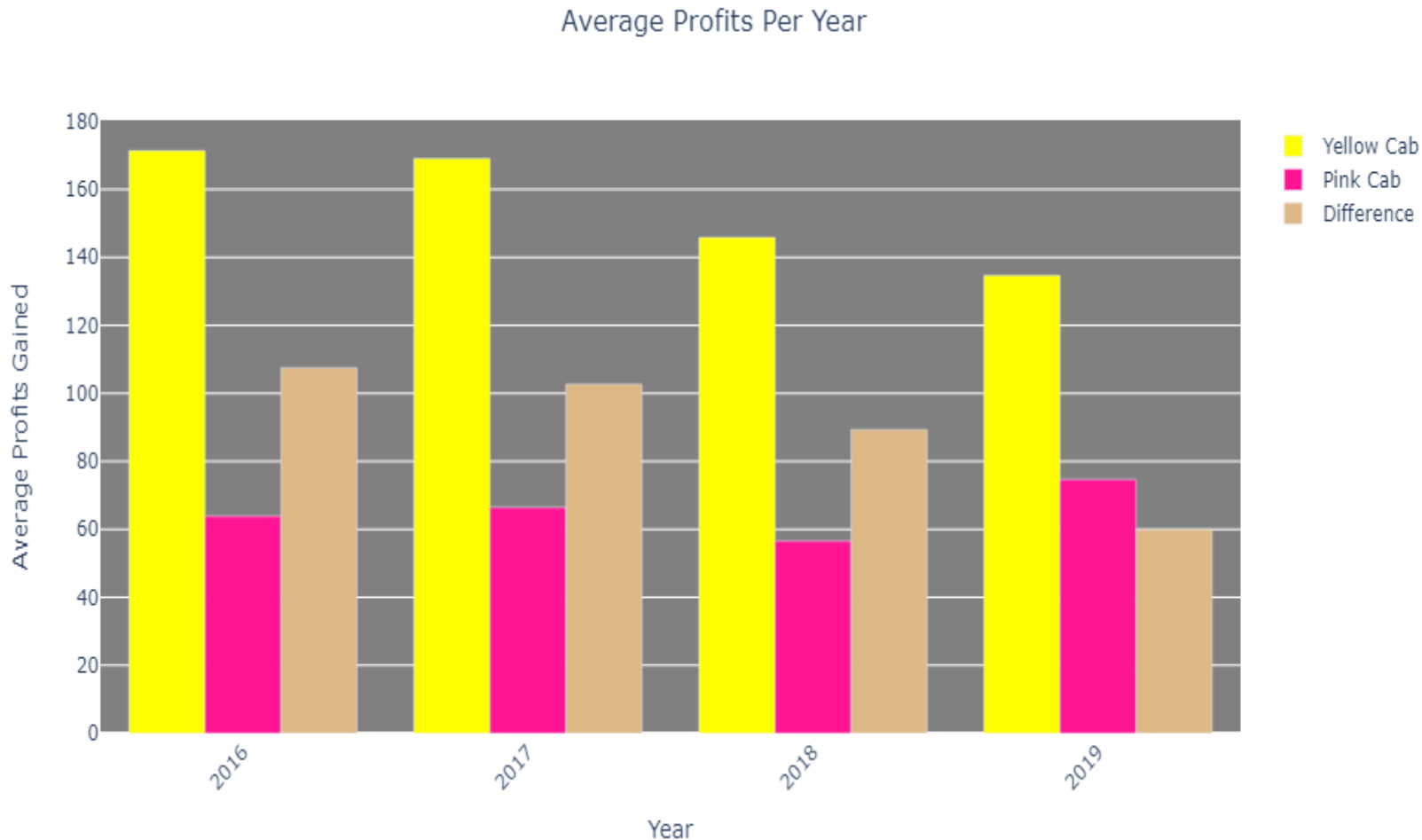
Seasonality Analysis

Monthly Cab Usage

- There is considerable seasonality in cab usage.

- Cab usage rises almost steadily from February to January of the next year.

- Cabs are busiest in January and December, which are generally holiday months.
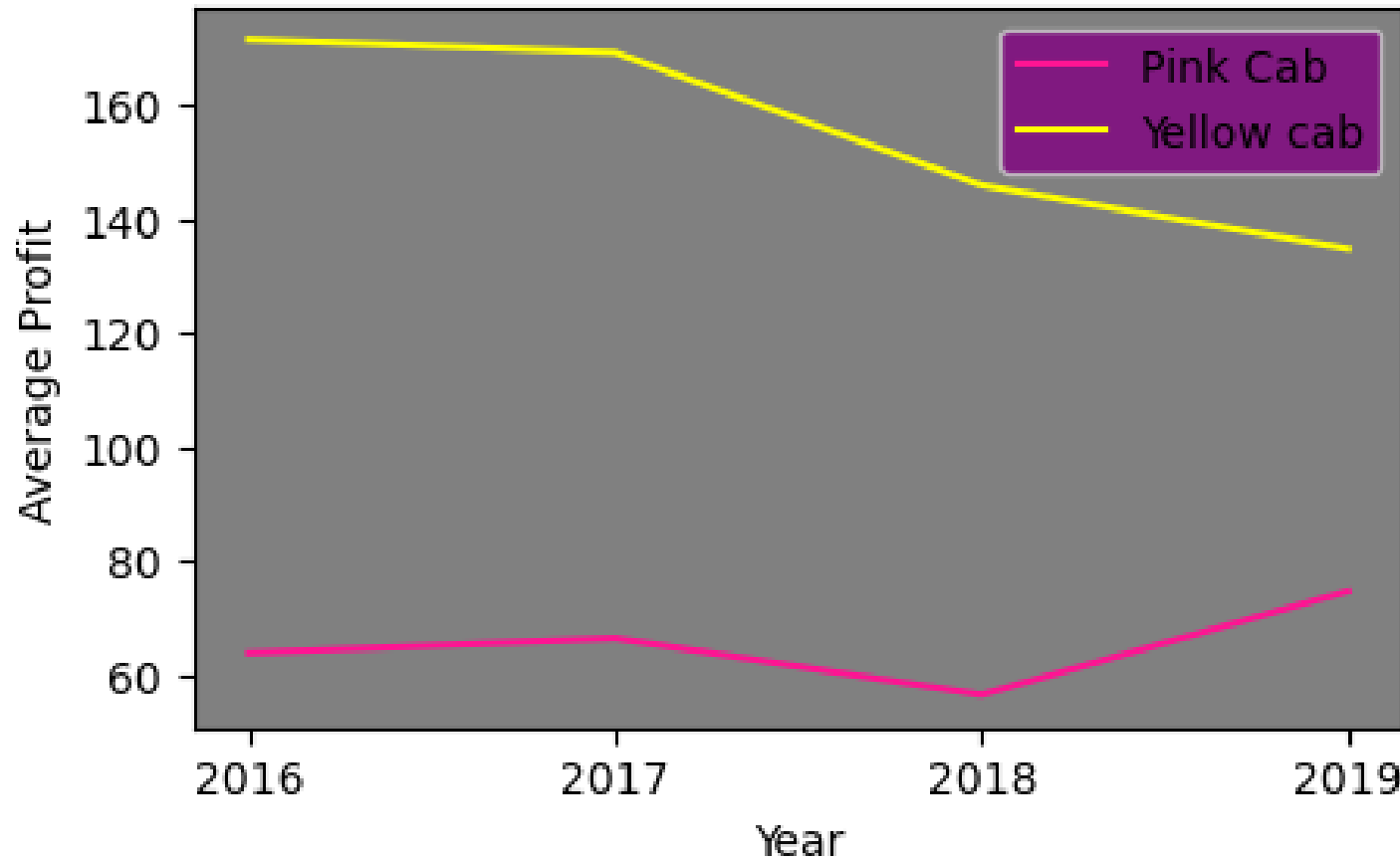
Seasonality Analysis

Daily Cab Usage

Profit Analysis

Average Profits Per Year

- The Yellow Cab has its average profits being significantly higher than the Pink Cab's.

- From 2016 to 2019, there exists an average profit gap of $89.86 in favor of the Yellow Cab.
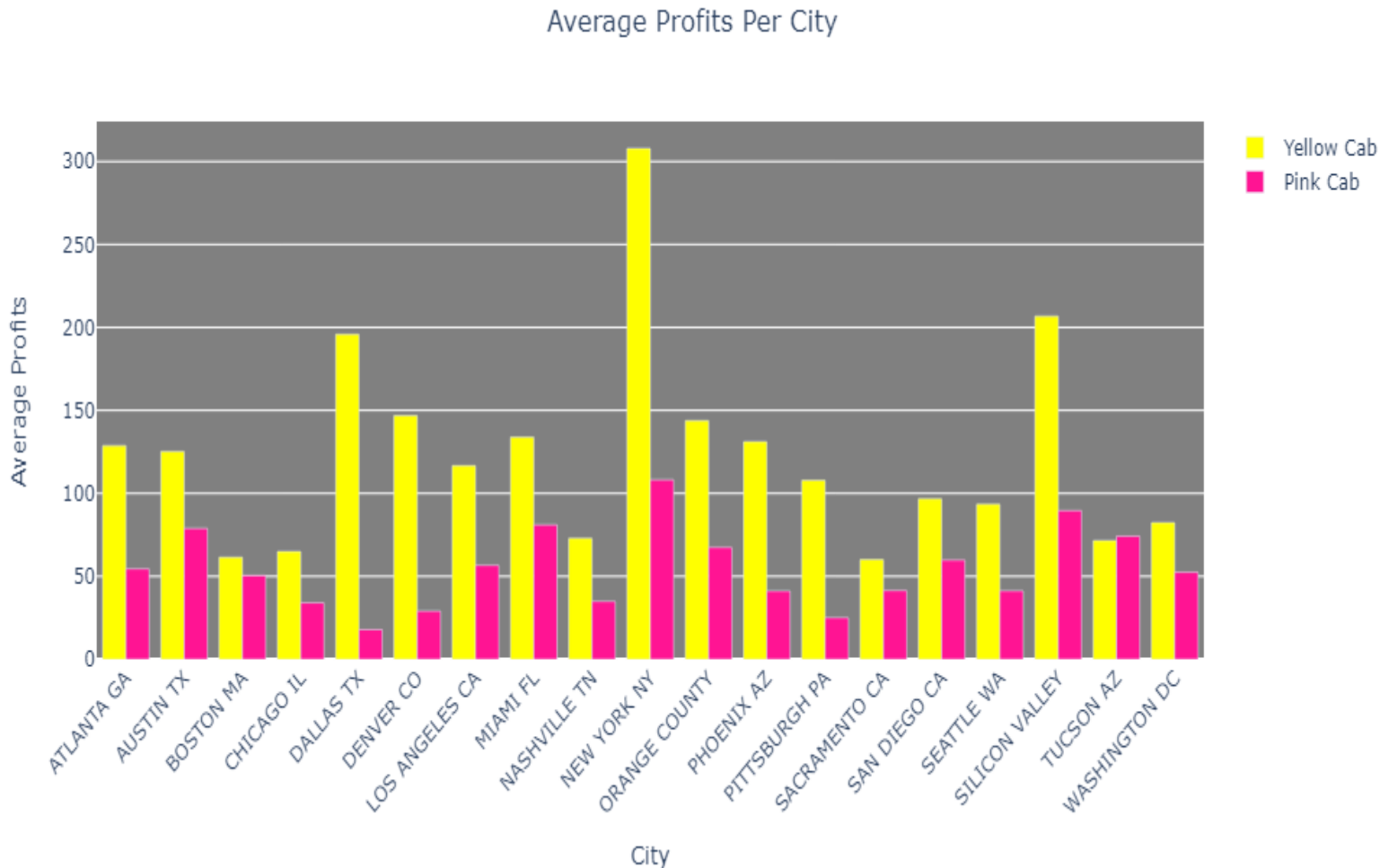
Profit Analysis

Yearly Average Profit Trend

- Despite the Yellow Cab's higher annual profits, it experienced a consistent decline in earnings from 2017 to 2019. Conversely, the Pink Cab saw a surge in profits between 2018 and 2019.

Profit Analysis

Average Profits Per City

- Both cab services excel in New York and Silicon Valley in terms of average profits. The Yellow Cab stands out further in Dallas, while the Pink Cab thrives in Austin and Miami.

- We will see in the next slides that cities with most users are not necessarily the cities where the companies make the most average profits.
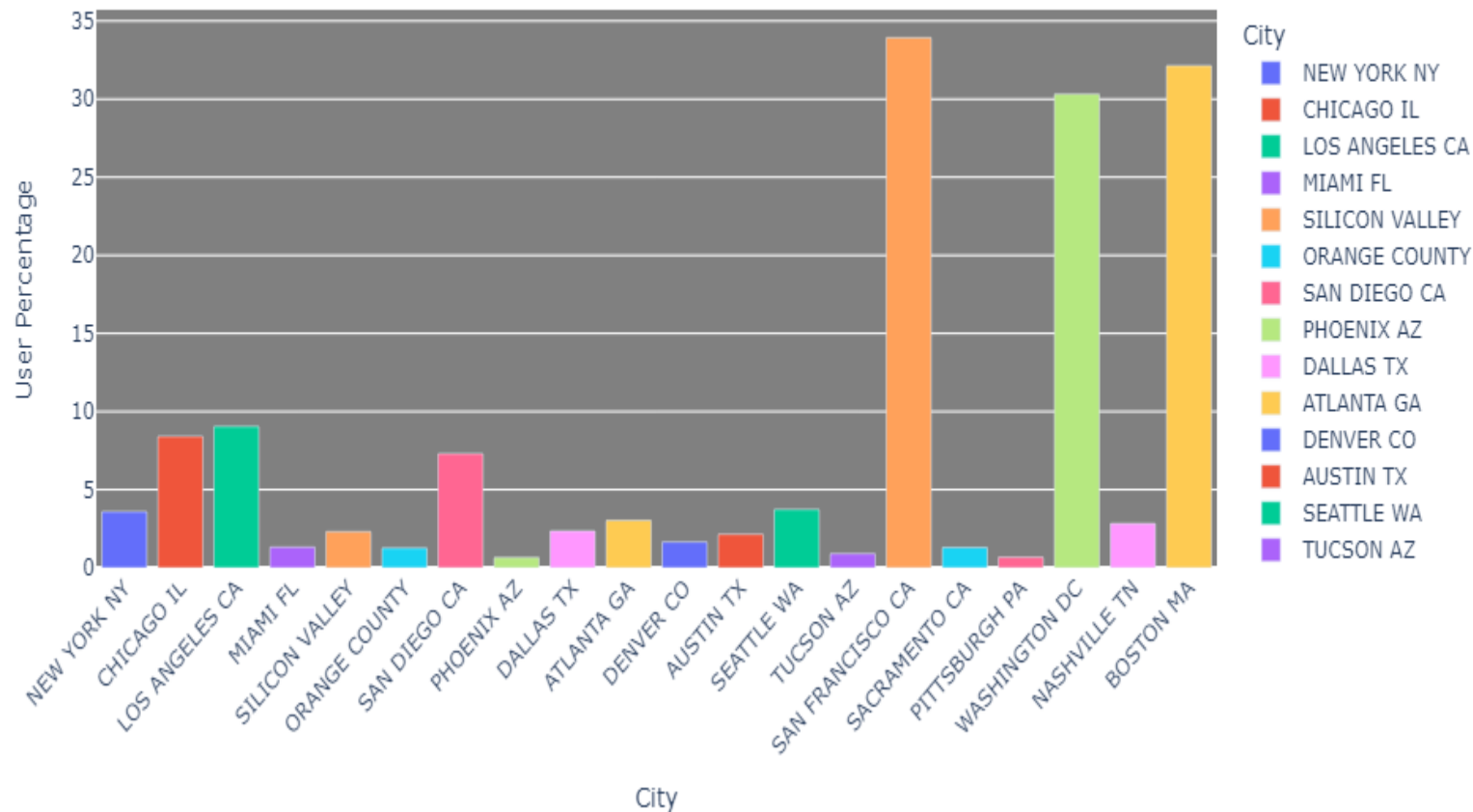
City Data Analysis

Cab Companies' Cities of Operation
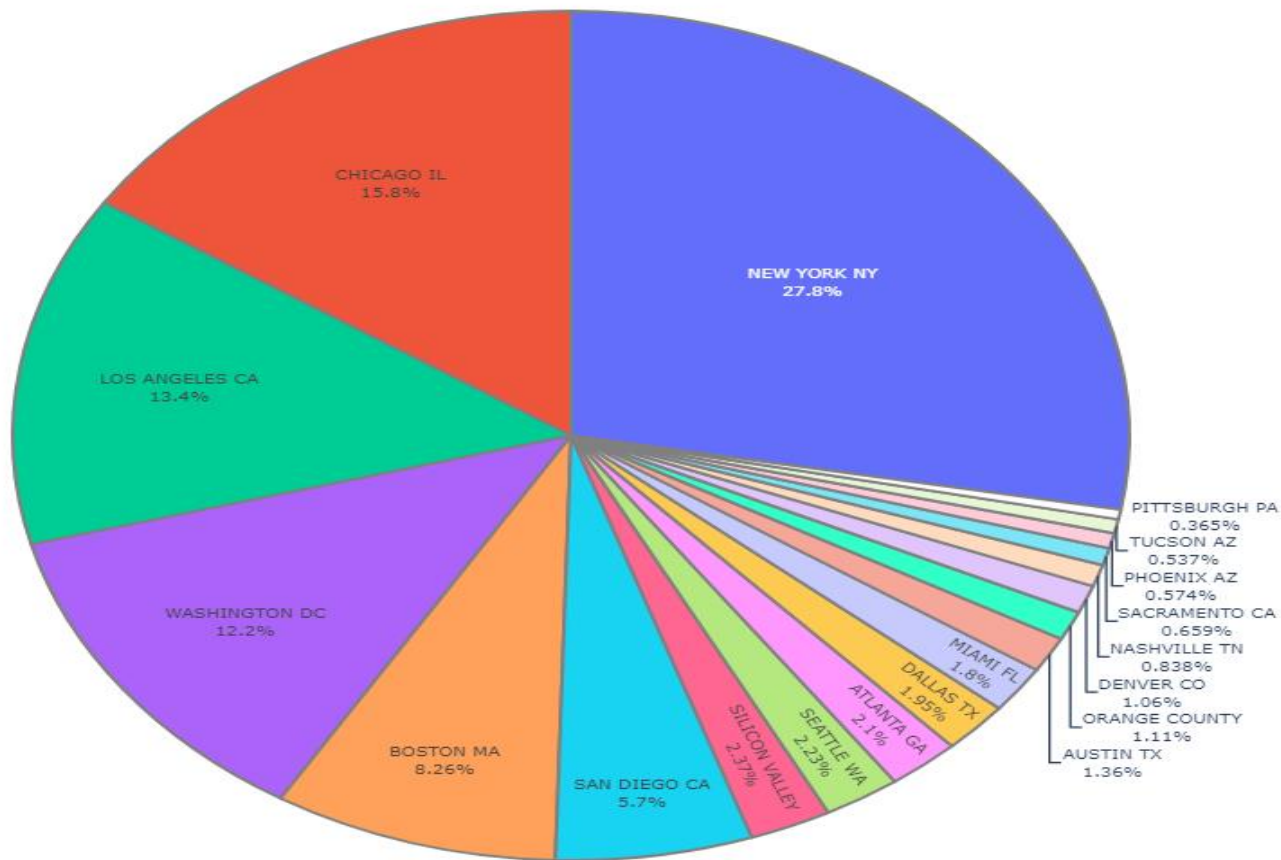
- New York, Chicago, Los Angeles, Washington, DC and Boston are the top 5 cities of operation for these two cab companies, with respect to number of users.

City Data Analysis

- The two Cab companies have over 30% of the entire population of Washington DC, San Francisco and Boston using their services.

- There is a lot of room for growth, especially considering that they have less than 10% of the people in each of the other cities using their services.

Customer Retention Analysis

Cab Company	5+ Transaction Customers	5+ Retention Rate	10+ Transaction Customers	10+ Retention Rate	15+ Transaction Customers	15+ Retention Rate
Pink	3690	4.36	240	0.28	7	0.01
Yellow	14631	5.33	9709	3.53	5474	1.99

❖ It's evident that the Yellow Cab demonstrates superior customer retention.

- 5.33% of the Yellow Cab company's customers returned at least 5 times, compared to 4.36% for the Pink Cab.

- 3.53% of Yellow Cab customers used the service at least 10 times, while only 0.28% did so for Pink Cab.

- Yellow Cab leads with a 1.99% retention rate for 15+ trips, compared to Pink Cab's 0.01%.

Mode of Payment Analysis

Payment Modes for Transactions

30.6%

45.9%

14.1%

9.46%

- Yellow Cab Card
- Yellow Cab Cash
- Pink Cab Card
- Pink Cab Cash
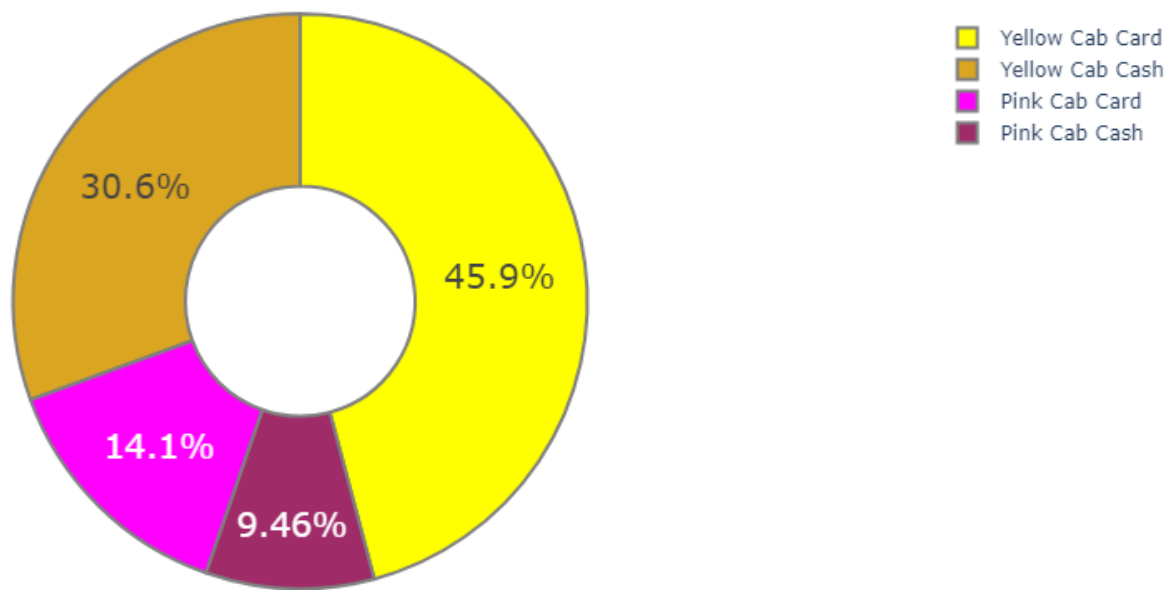
- Cards are the preferred payment mode for customers.

- Cards are used for payment 60% of the time, cash 40% - this is true for both cab companies.

- I suggest maintaining both card and cash payment alternatives since a considerable portion of customers utilize each of these methods.

Price and Distance Correlation

Correlation Between Price Charged and Trip Distance
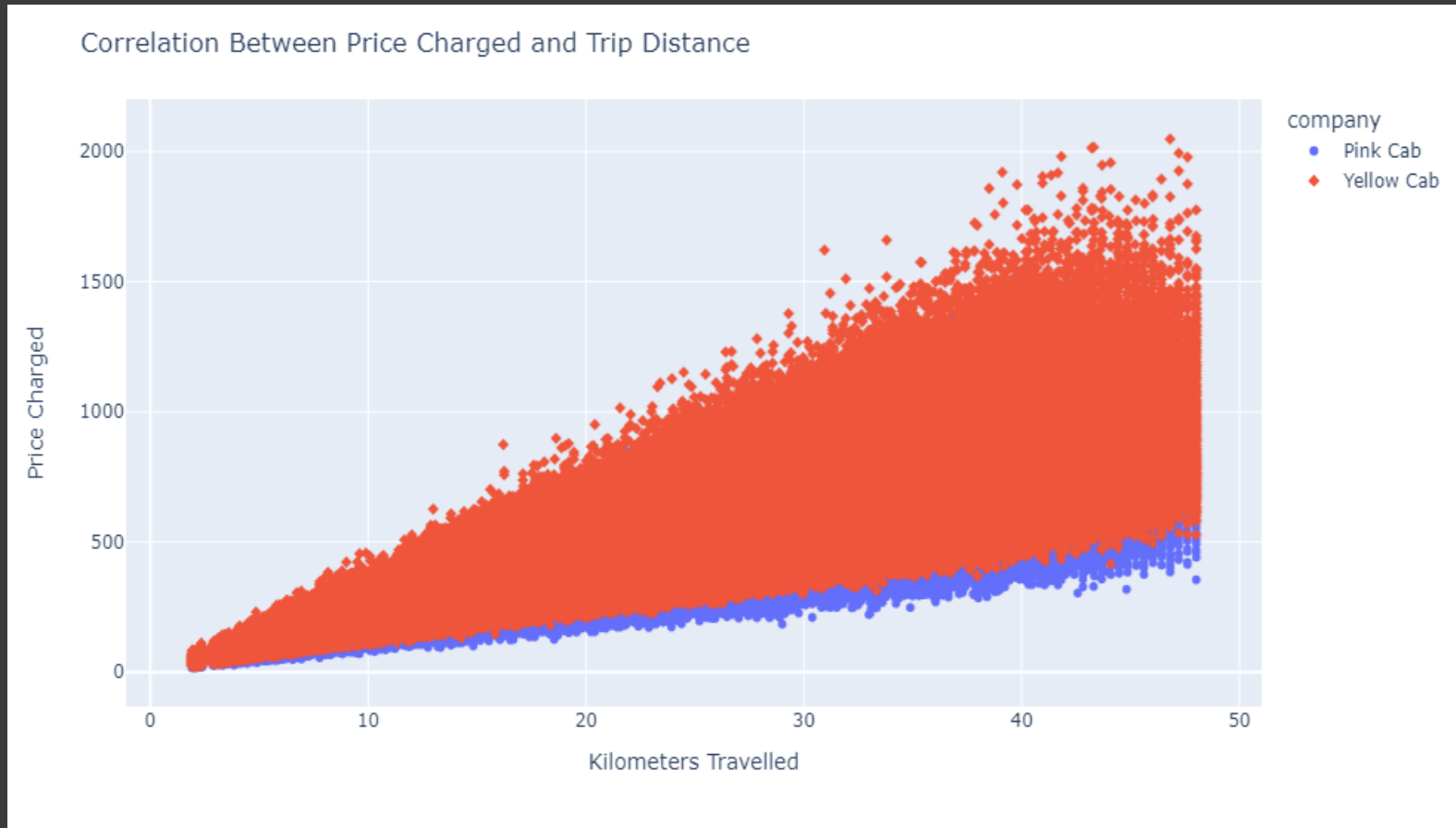
- As expected, there is a positive correlation between Price Charged and Distance travelled.

Data Glacier
Your Deep Learning Partner

Feature Correlation Heatmap

Feature Correlation Heatmap
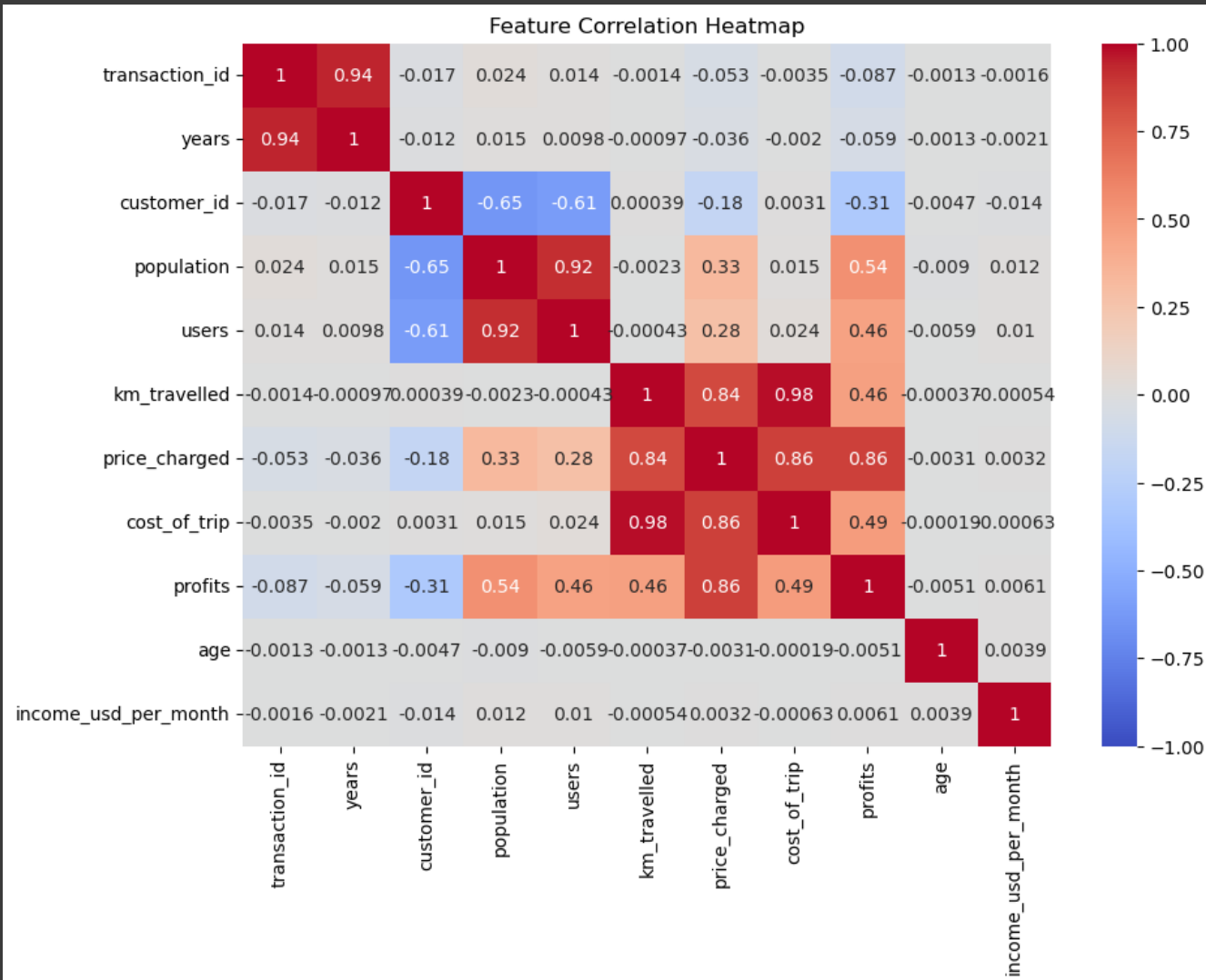
- As expected, there is a strong correlation between Population and Users, as well as between Distance Travelled, Cost of Trip and Price Charged.

- Again, there is a strong correlation between Transaction ID and the Years of the transactions.
    - - A transaction at a later date would comprise of a greater number for its Transaction ID.

Recommendations

- The Cab companies collectively serve more than 30% of the total population in Washington DC, San Francisco, and Boston. Nevertheless, there exists significant potential for expansion, particularly given that their service utilization remains below 10% of the population in each of the other cities.

- I recommend that XYZ consider investing in the Yellow Cab company for the following reasons:
    - ❑ The Yellow Cab company demonstrates a significantly higher customer retention rate compared to the Pink Cab. This suggests that even in challenging times, the Yellow Cab is more likely to thrive due to its strong customer loyalty.
    - ❑ The Yellow Cab is the preferred choice for the majority of customers, accounting for 76.43% of all cab journeys, surpassing the Pink Cab's share of 23.57%.
    - ❑ The Yellow Cab achieves notably higher average profits. Over the course of four years,there is an average profit difference of $89.86 in favor of the Yellow Cab.
- Notable Drawback:
    - ❑ While the Yellow Cab has significantly higher annual profits, its earnings have exhibited a consistent decline from 2017 to 2019. In contrast, the Pink Cab saw a notable upswing in profits between 2018 and 2019.

Thank You