Data Glacier

Your Deep Learning Partner

G2M Case Study

Virtual Internship

Priyanjali Patel

27-May-2025

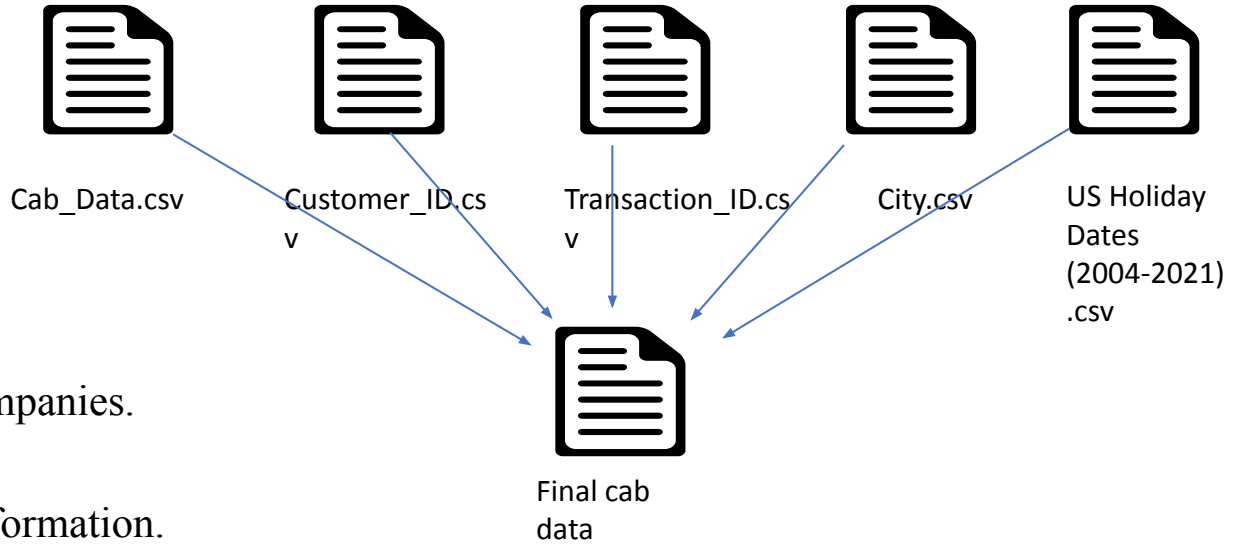
Background –G2M(cab industry) case study

- **Business Problem**

- **Client:** XYZ (Private Firm, USA)

Objective:
Due to rapid growth in the cab industry and the presence of multiple major players, Client XYZ is exploring potential investment opportunities in this sector. As part of their Go-to-Market (G2M) strategy, they aim to gain a comprehensive understanding of the market landscape before making an informed investment decision.

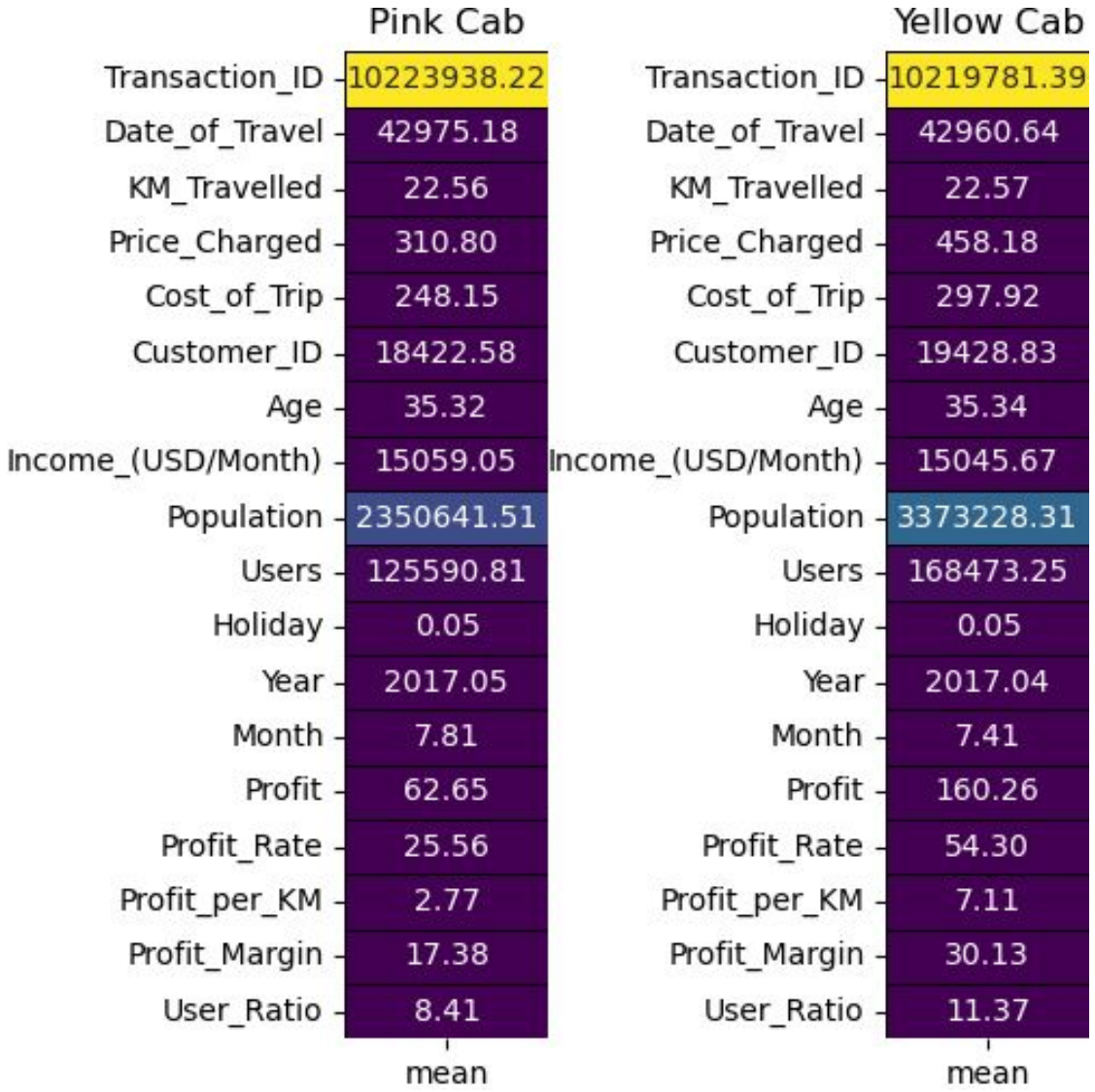
Data Exploration



Provided Datasets (Time Period: Jan 31, 2016 – Dec 31, 2018)

1. **Cab_Data.csv** – Contains transaction details from two cab companies.
2. **Customer_ID.csv** – Links customer IDs with demographic information.
3. **Transaction_ID.csv** – Maps transactions to customer IDs and payment methods.
4. **City.csv** – Includes U.S. city-level data: population and cab user count.
5. **US Holiday Dates (2004-2021).csv** – This dataset contains the official holiday dates in the United States from 2004 to 2021.

Average Feature Comparison Between Pink Cab and Yellow Cab



Feature Engineering: New Metrics Introduced

- **Profit**
  = Price_Charged - Cost_of_Trip
  Measures the absolute earnings per trip, but may be misleading alone if both cost and revenue are high.

- **Profit Rate (% Return on Cost)**
  = (Profit / Cost_of_Trip) * 100
  Indicates how much profit is earned for every dollar spent on cost.

- **Profit per KM**
  = Profit / KM_Travelled
  Normalizes profit based on distance, revealing which routes are more profitable

Feature Engineering: New Metrics Introduced

- **Profit Margin (% Return on Revenue)**
  = (Profit / Price_Charged) * 100
  Shows how efficiently each company converts revenue into profit.

- **User Ratio (% of Population Using Cabs)**
  = (Users / Population) * 100
  Helps understand market penetration in different cities.

2. Exploratory Data Analysis (EDA)

Overall Analysis: Company-Wise Comparison

Hypothesis 1:

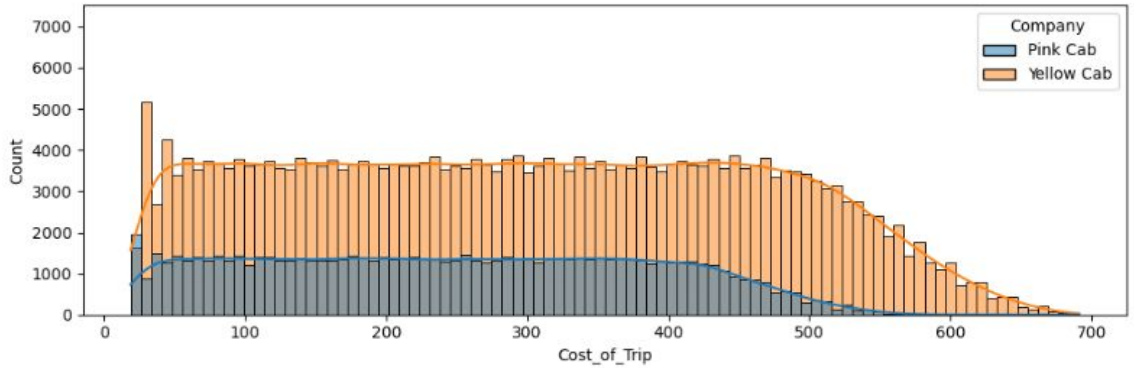
"The distribution of profit-related metrics (e.g., profit, profit rate, profit per km) significantly differs between Pink Cab and Yellow Cab."

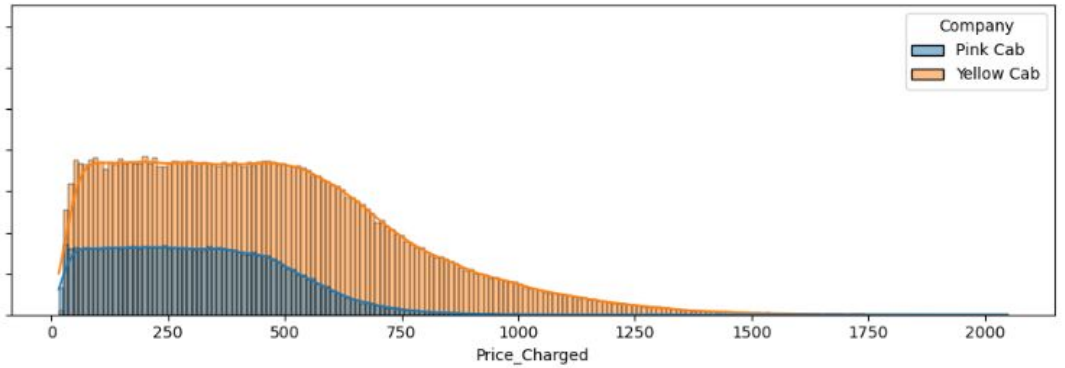
Key Questions:

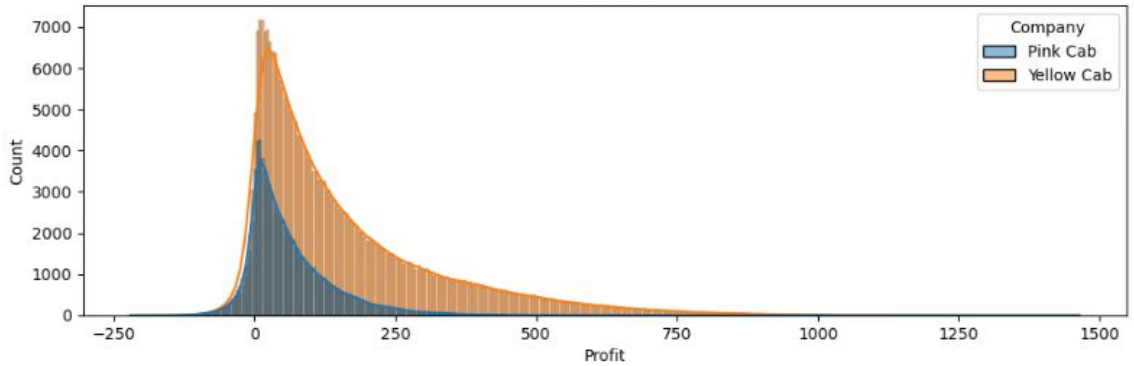
- Which company has a higher average profit rate?
- Is the cost structure (Cost of Trip) similar for both?
- Does Yellow Cab serve more long-distance/high-cost trips compared to Pink Cab?

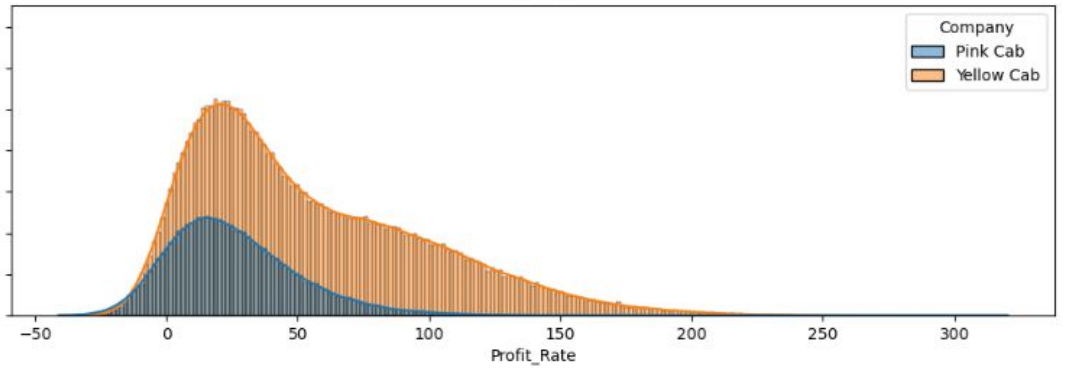
2. Exploratory Data Analysis (EDA)

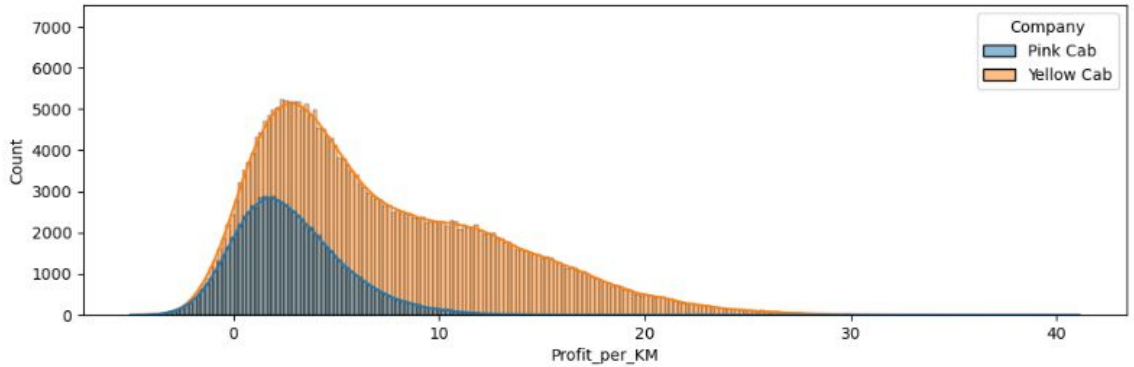
Overall Analysis: Company-Wise Comparison

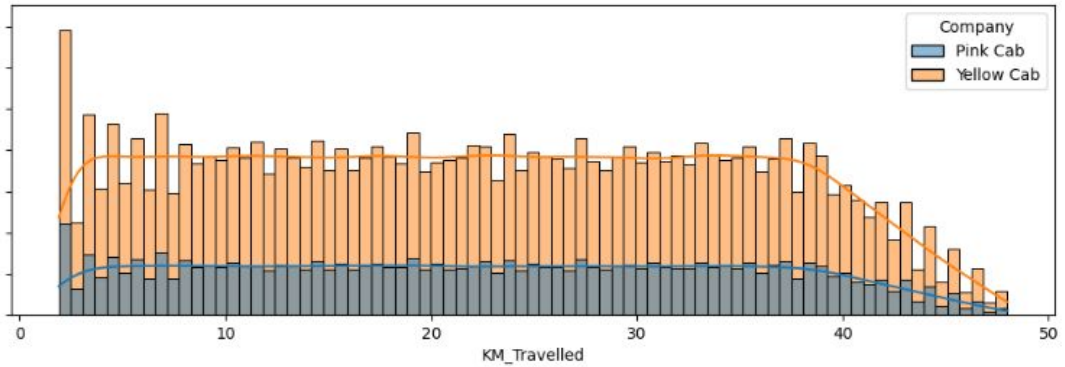












2. Exploratory Data Analysis (EDA)

Overall Analysis: Gender-Based Analysis

Hypothesis 2:

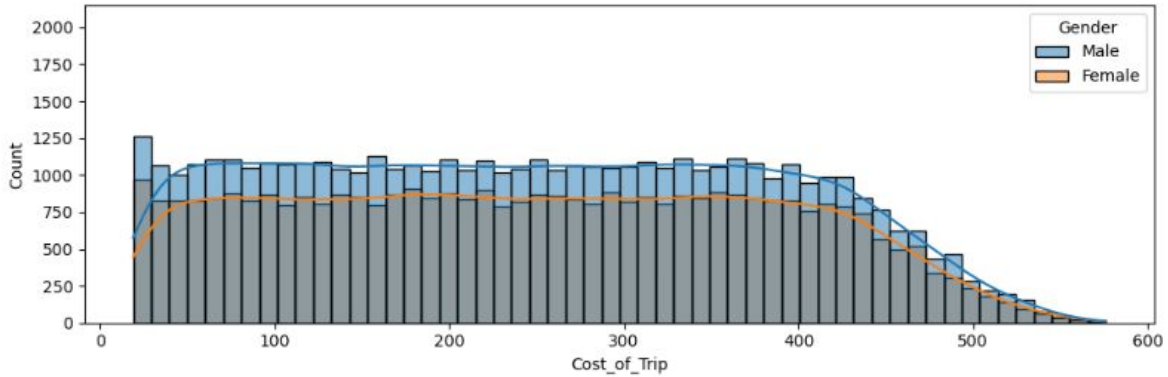
"Profit metrics vary by customer gender within each company."

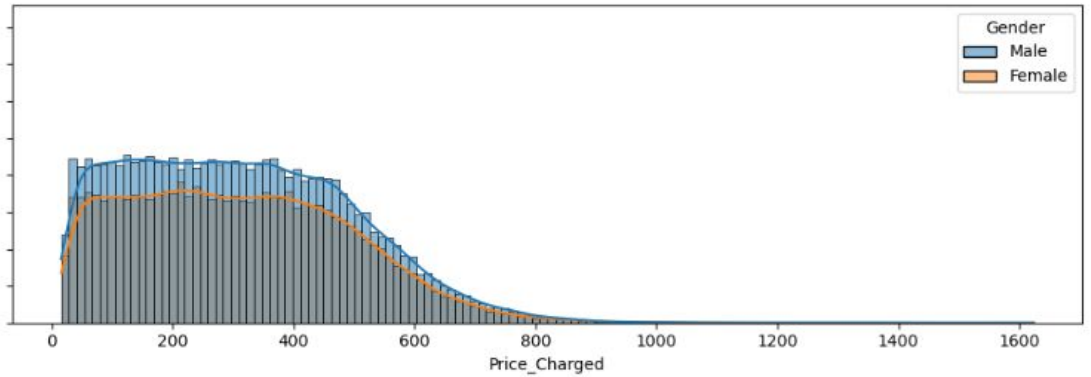
Key Questions:

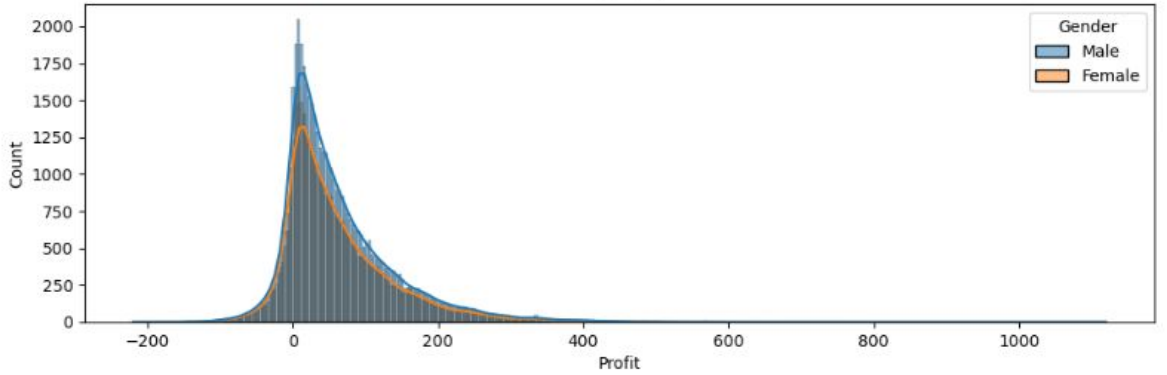
- Do male or female customers lead to higher profitability?
- Are there pricing or trip length differences based on gender?

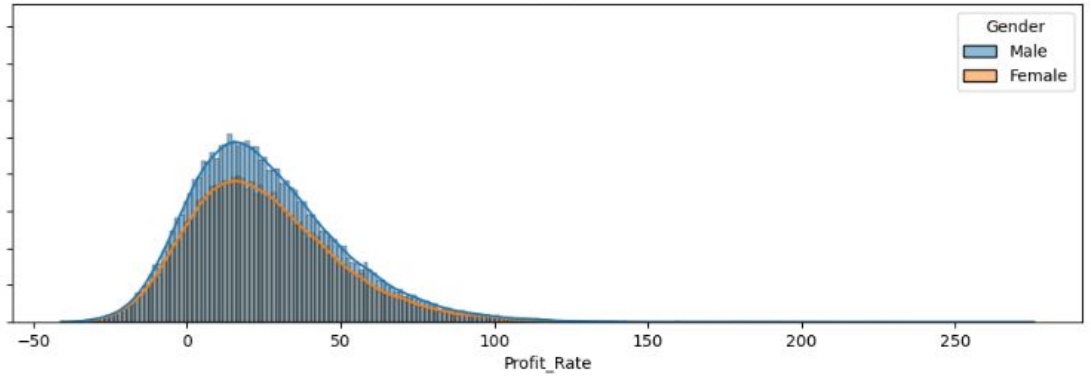
2. Exploratory Data Analysis (EDA)

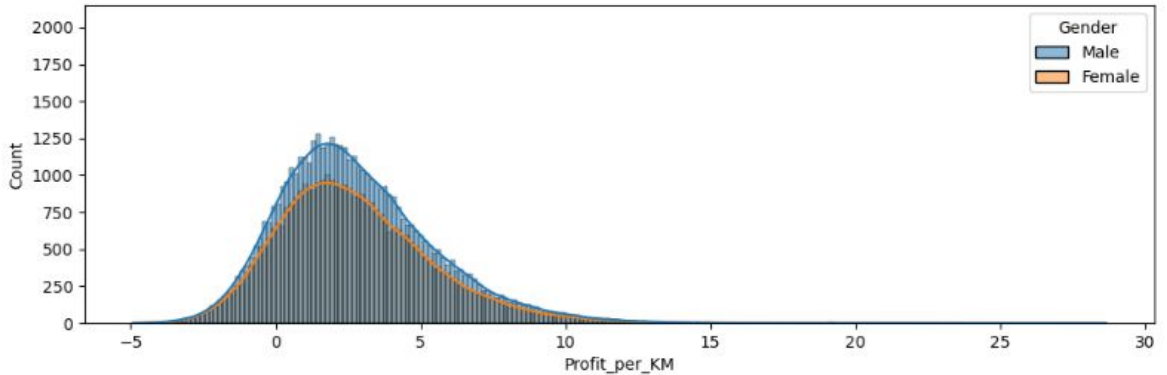
Overall Analysis: Gender-Based Analysis for Pink Cab

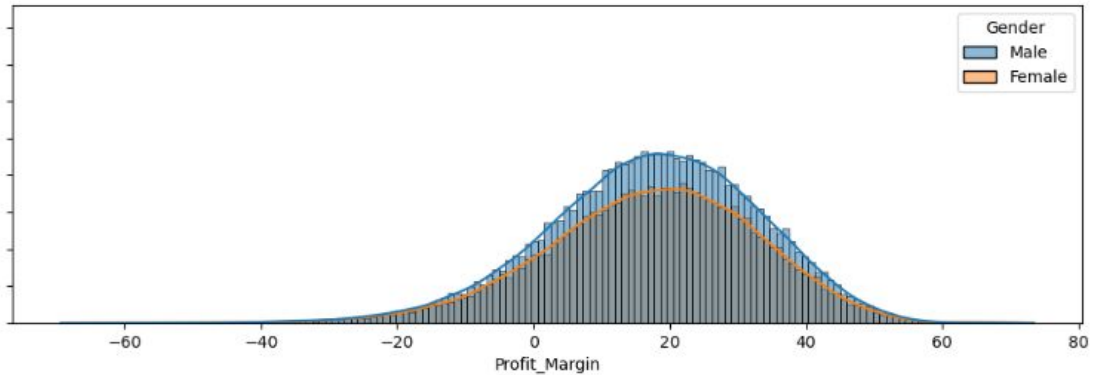






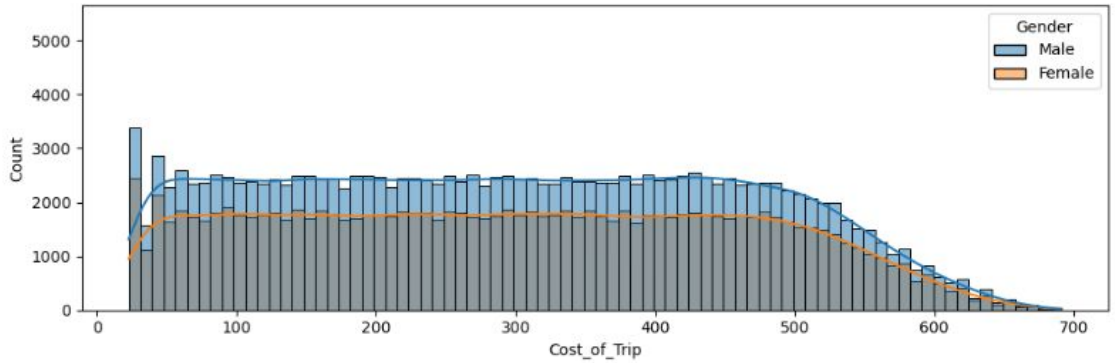


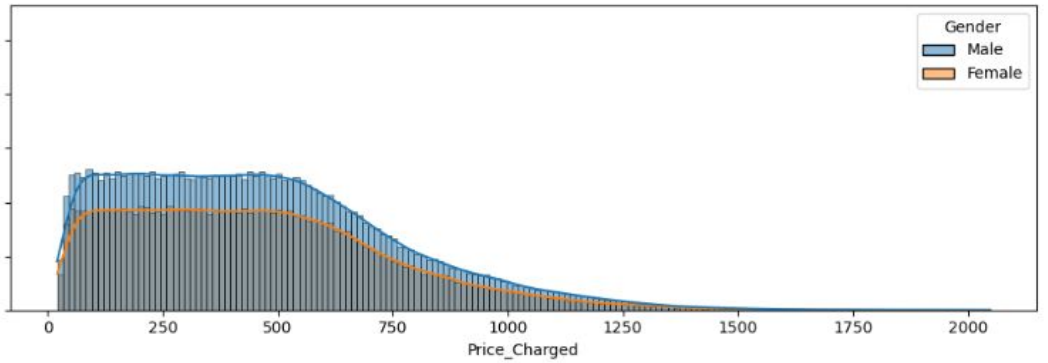


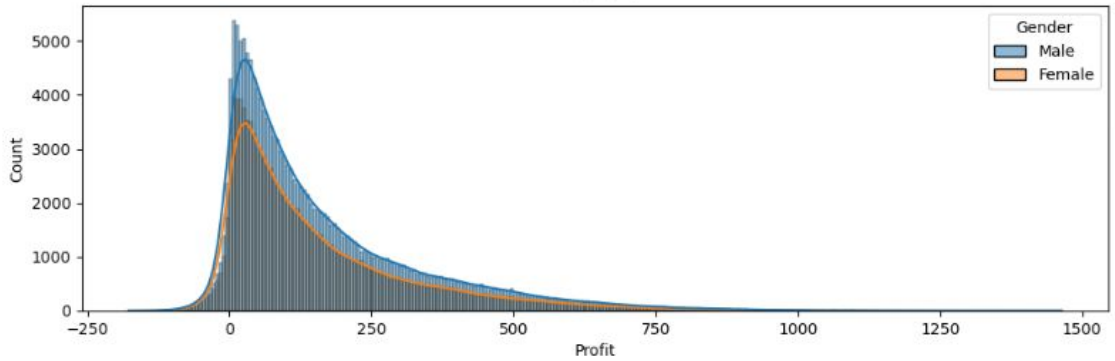


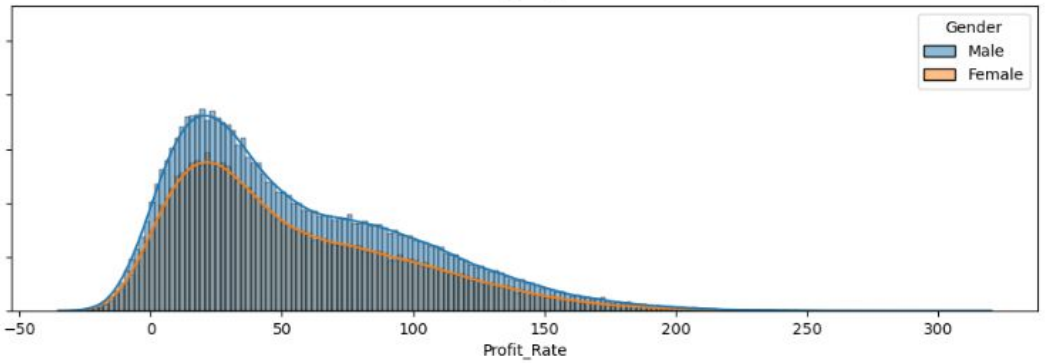
2. Exploratory Data Analysis (EDA)

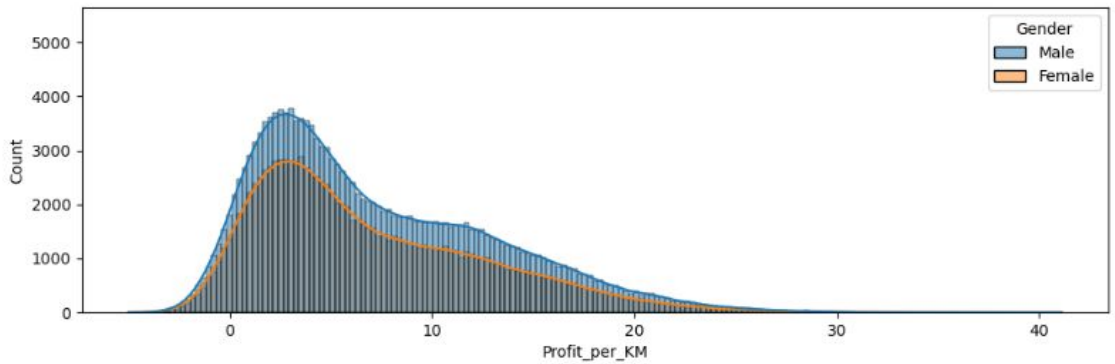
Overall Analysis: Gender-Based Analysis for Yellow Cab

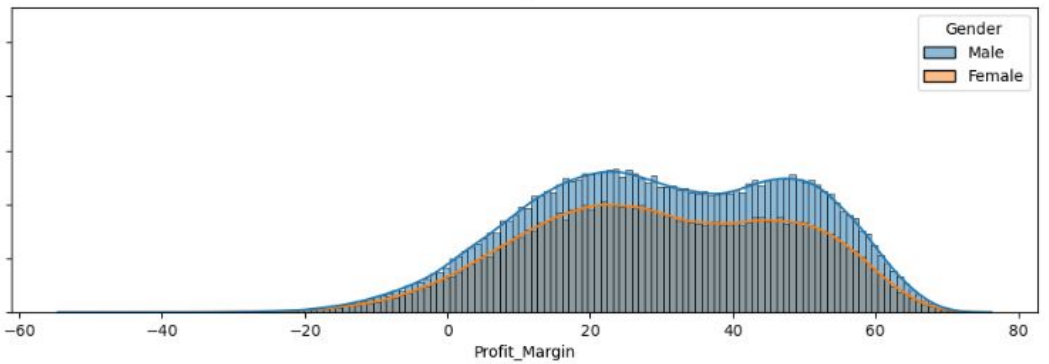












2. Exploratory Data Analysis (EDA)

Overall Analysis: Holiday Impact

Hypothesis 3:

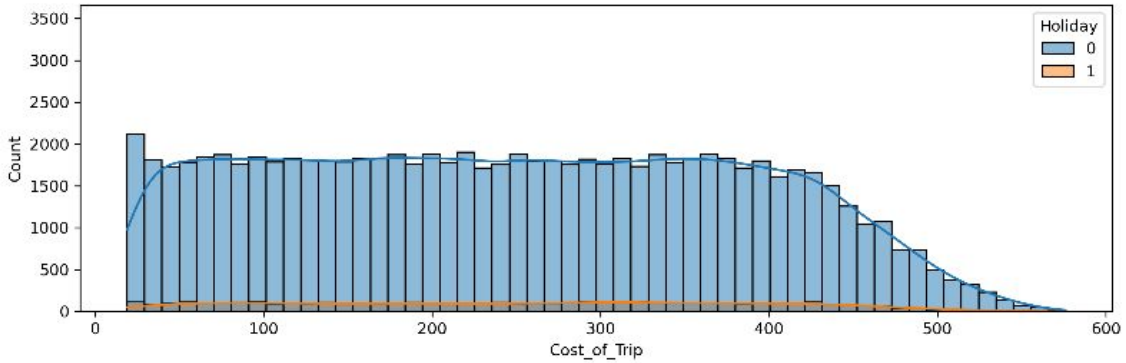
"Cab profitability increases on holidays due to higher demand."

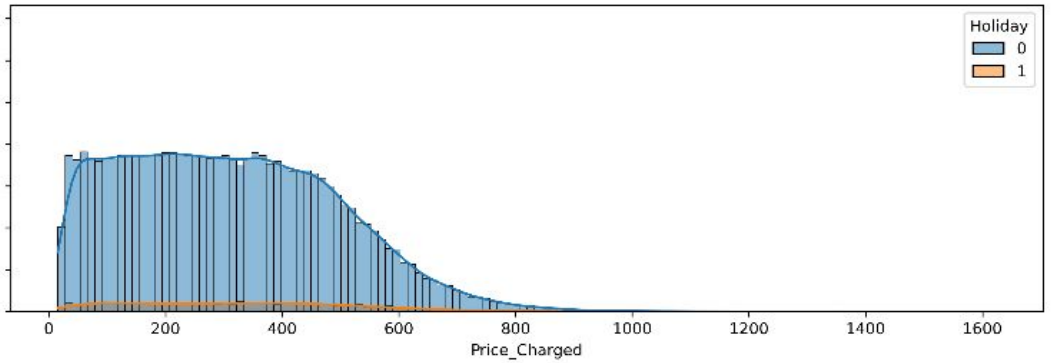
Key Questions:

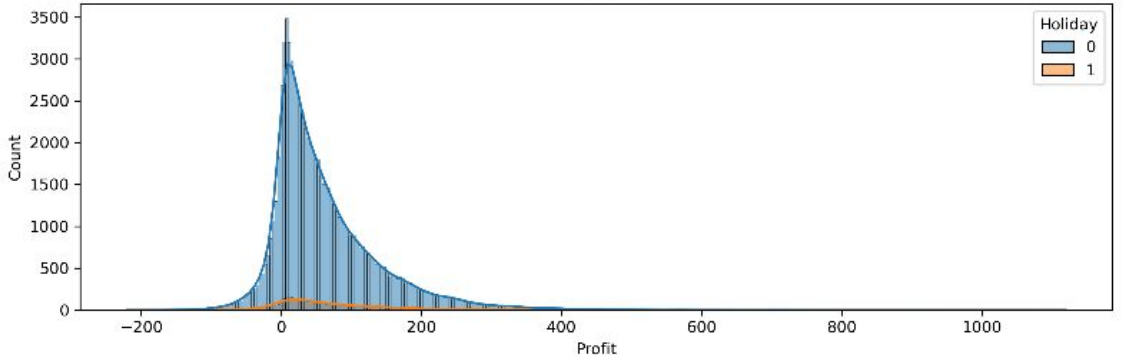
- Is there an increase in cost or price during holidays?
- Does profit or profit per km rise significantly on holidays?

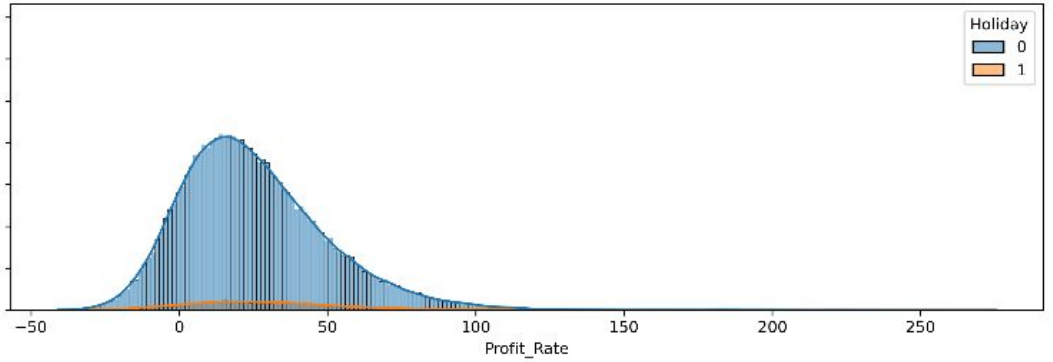
2. Exploratory Data Analysis (EDA)

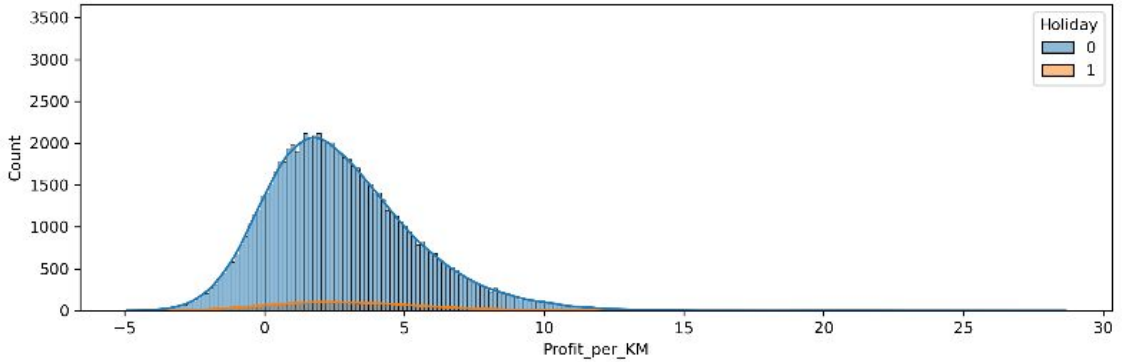
Overall Analysis: Holiday Impact for Pink Cab

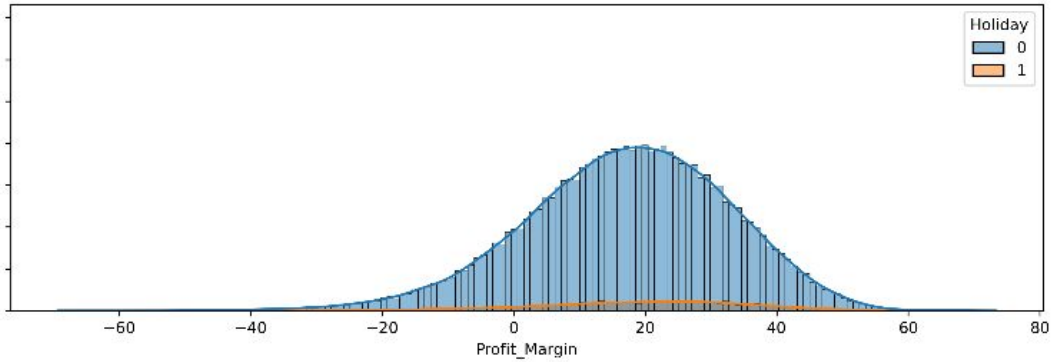






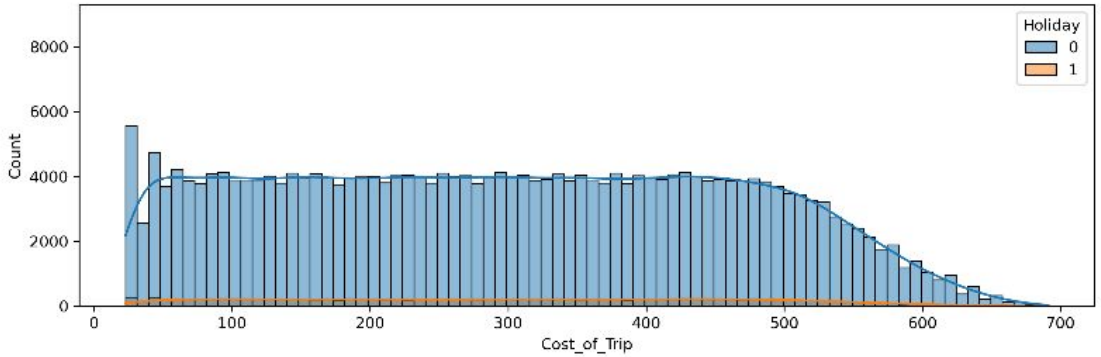


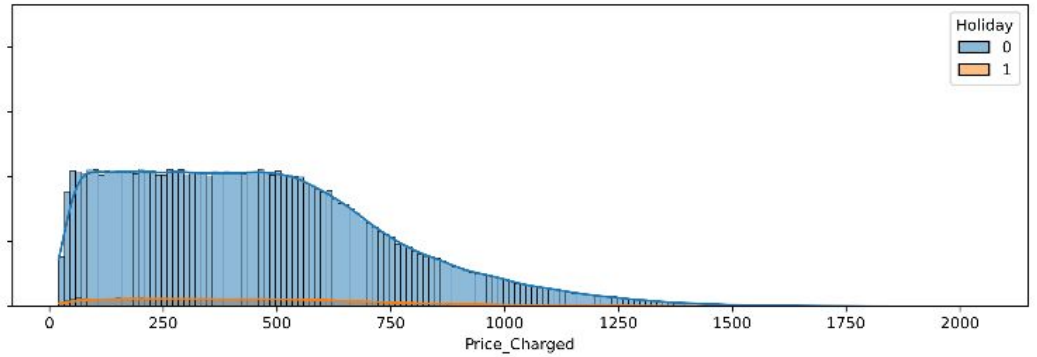


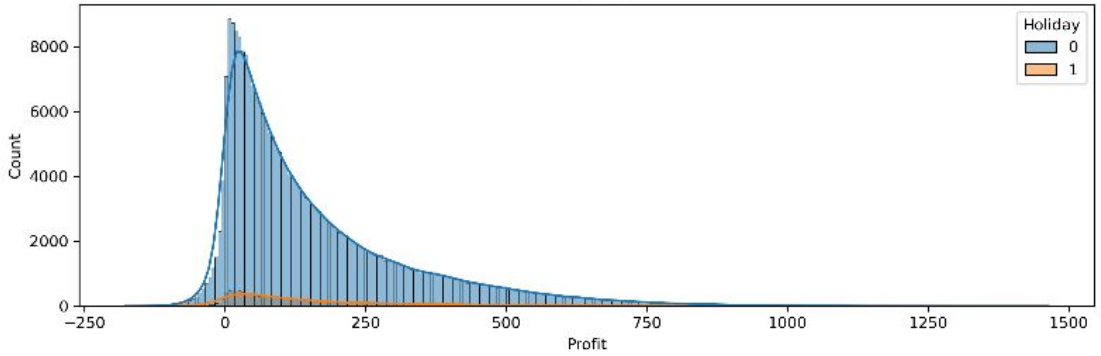


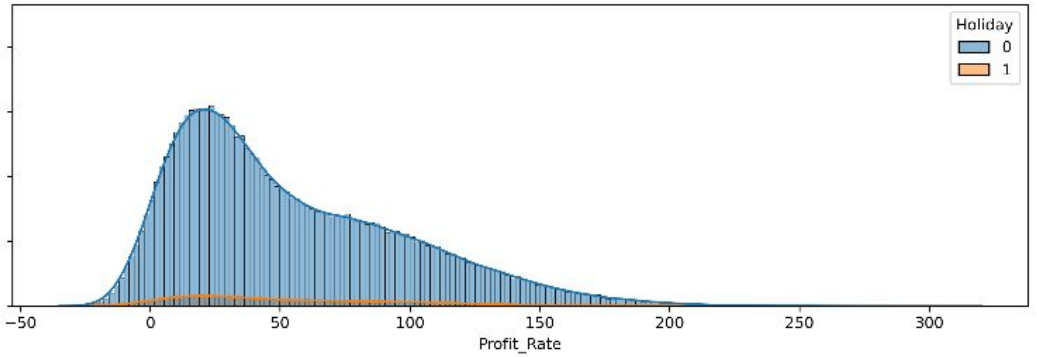
2. Exploratory Data Analysis (EDA)

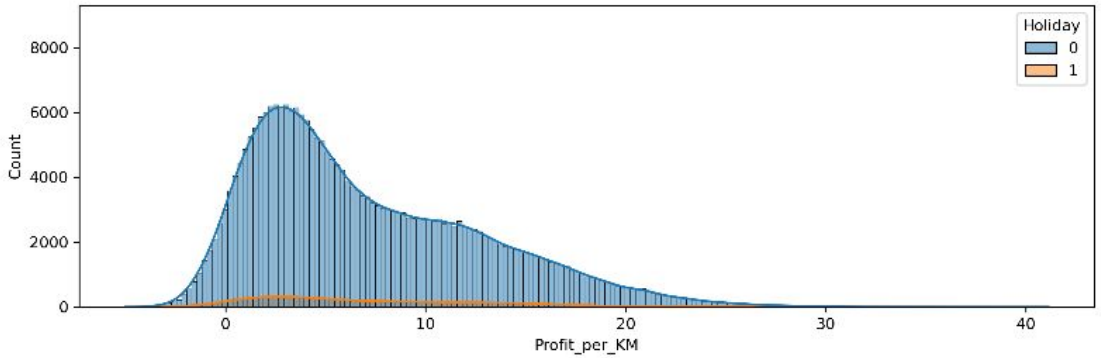
Overall Analysis: Holiday Impact for Yellow Cab

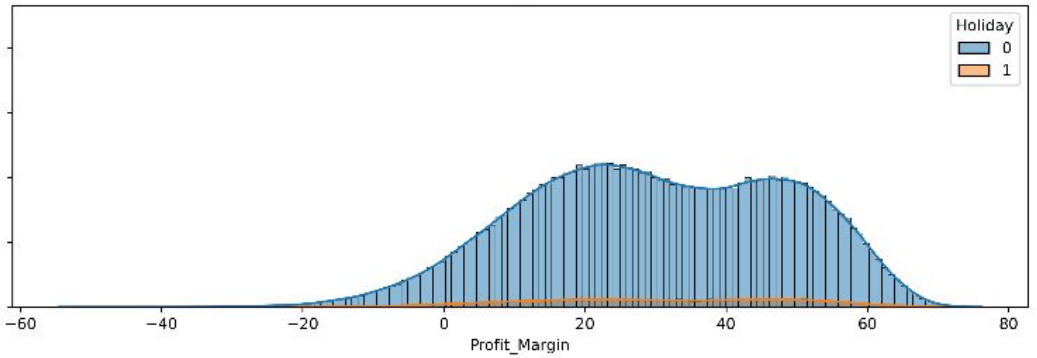












2. Exploratory Data Analysis (EDA)

Overall Analysis : City-Wise Analysis

Hypothesis 4:

"Certain cities contribute more to overall profit due to better margins or higher pricing."

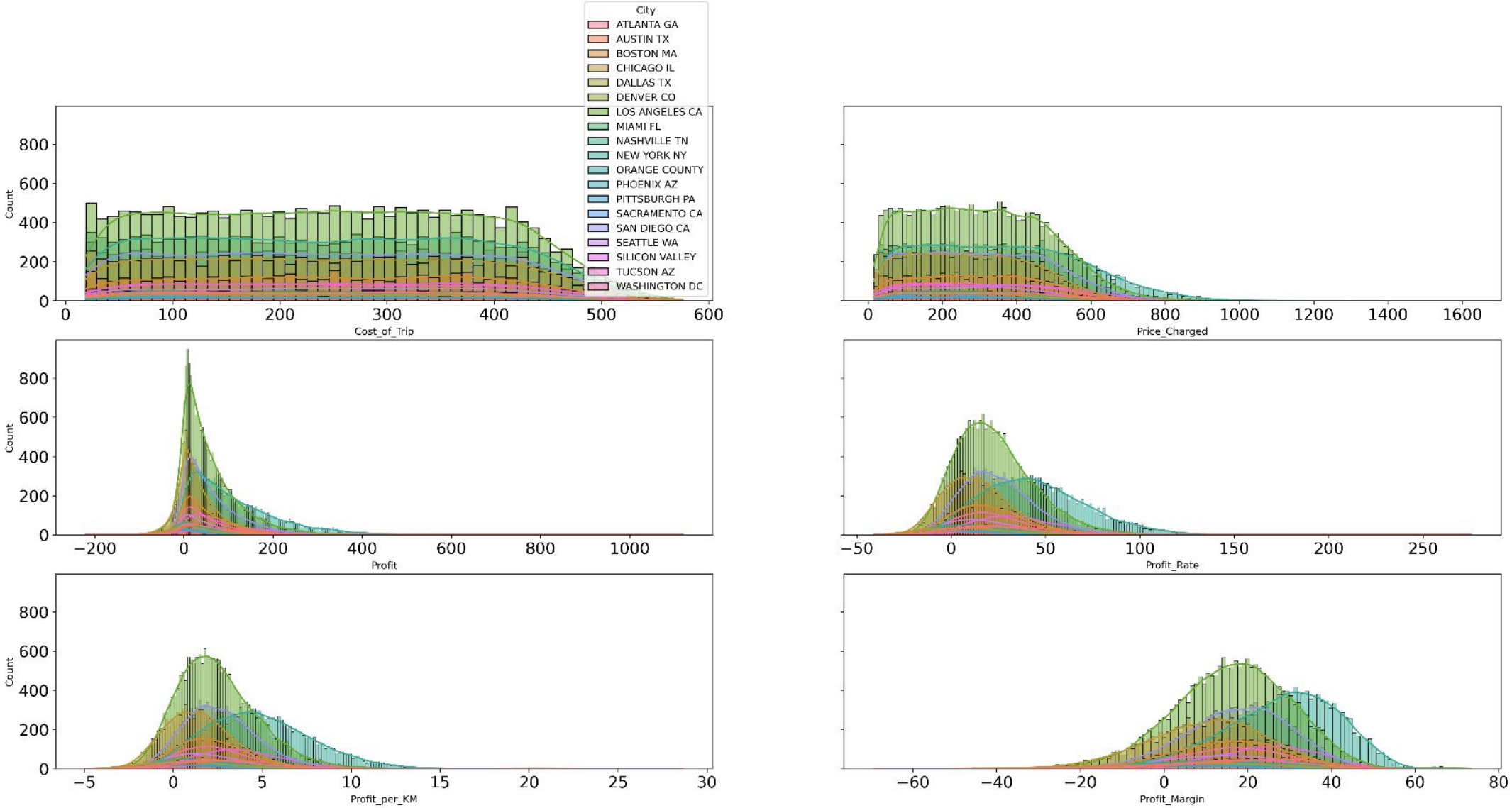
Key Questions:

- Are profit metrics consistent across cities?

- Which cities are the most or least profitable?

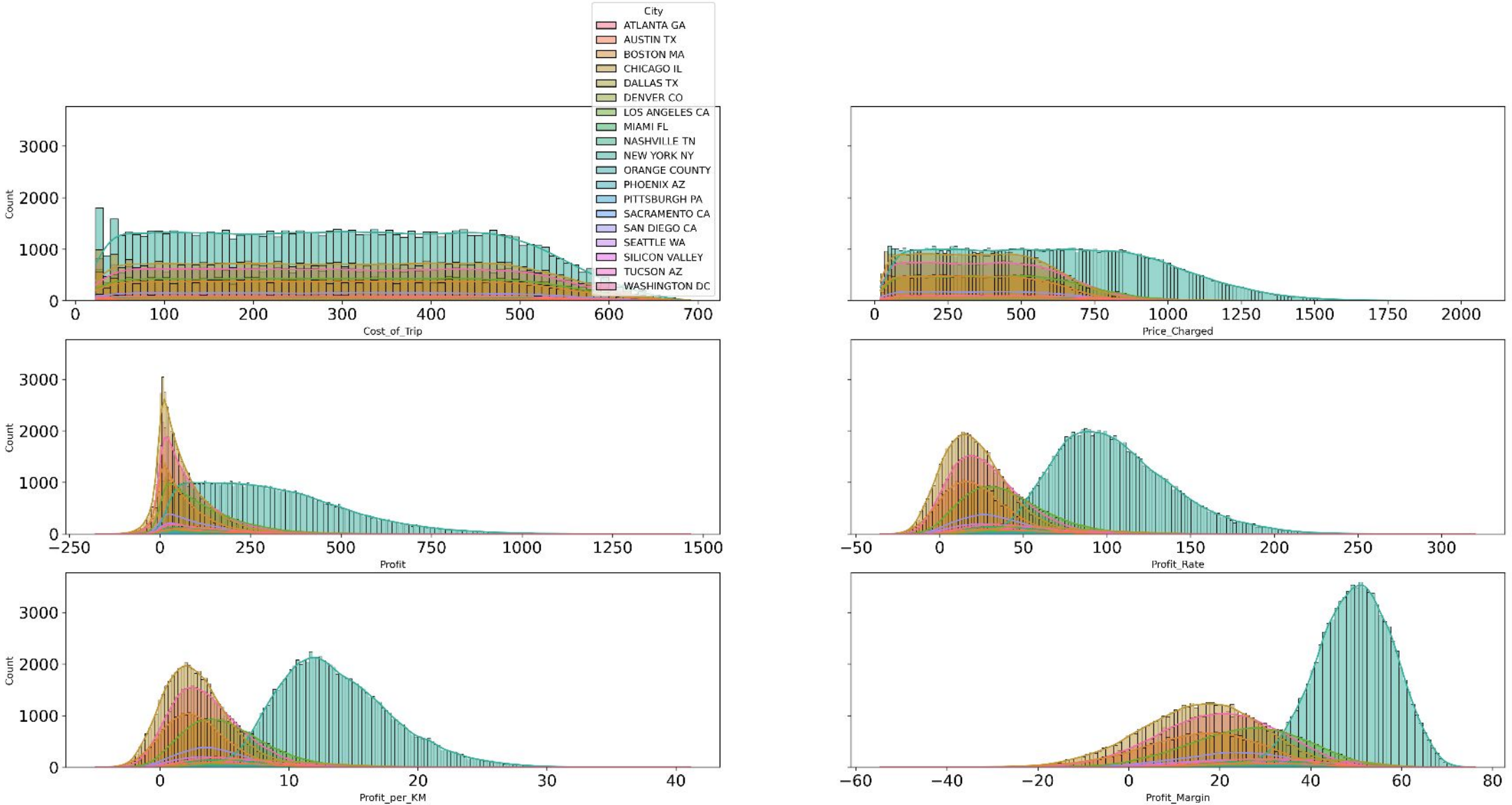
2. Exploratory Data Analysis (EDA)

Overall Analysis: City-Wise Analysis for Pink Cab



2. Exploratory Data Analysis (EDA)

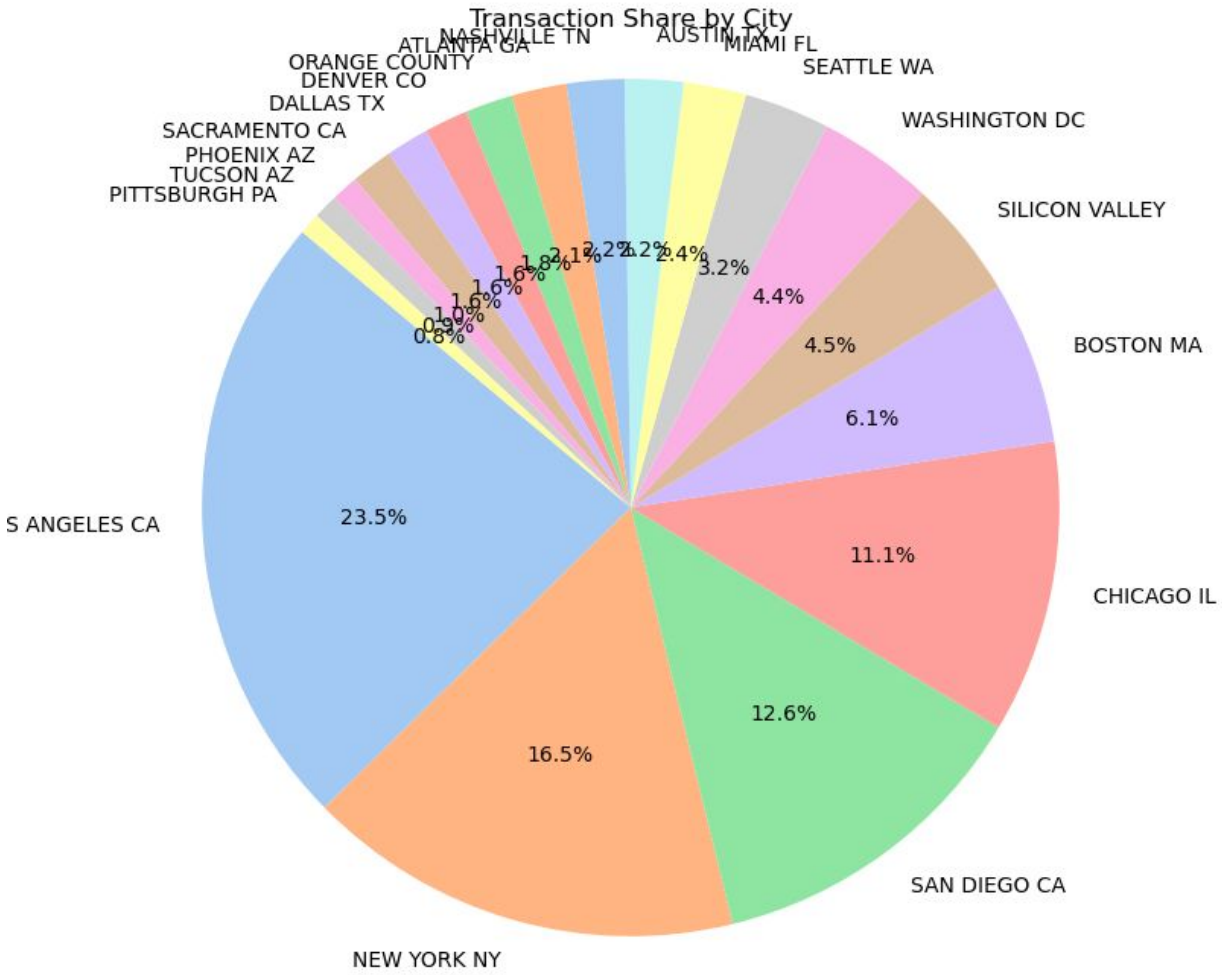
Overall Analysis: City-Wise Analysis for Yellow Cab



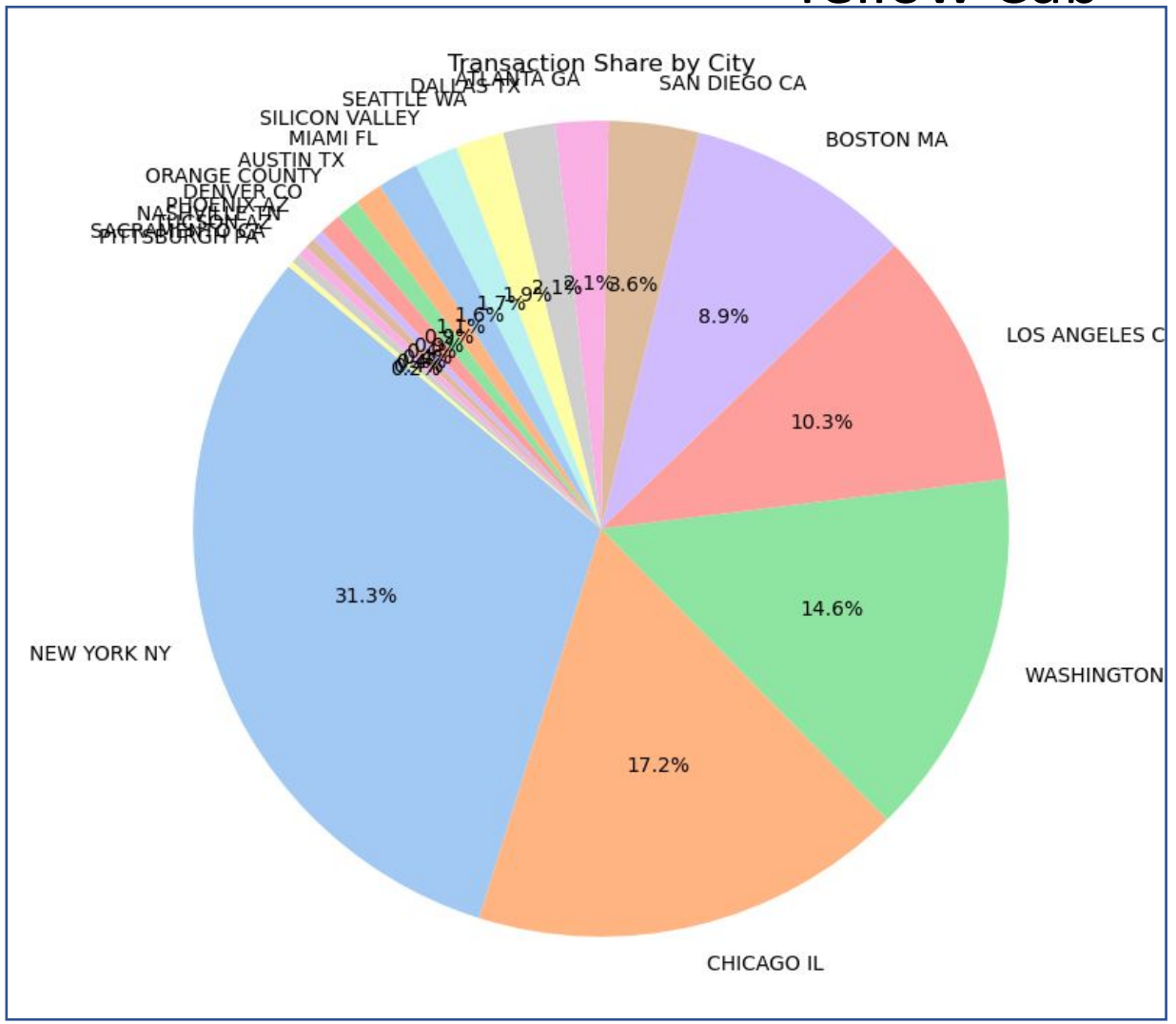
2. Exploratory Data Analysis (EDA)

Overall Analysis: City-Wise Analysis

Pink Cab



Yellow Cab



2. Exploratory Data Analysis (EDA)

Overall Analysis: Temporal/Yearly Trends

Hypothesis 5:

"Cab company performance changes over time, with evolving profit rates or cost structures."

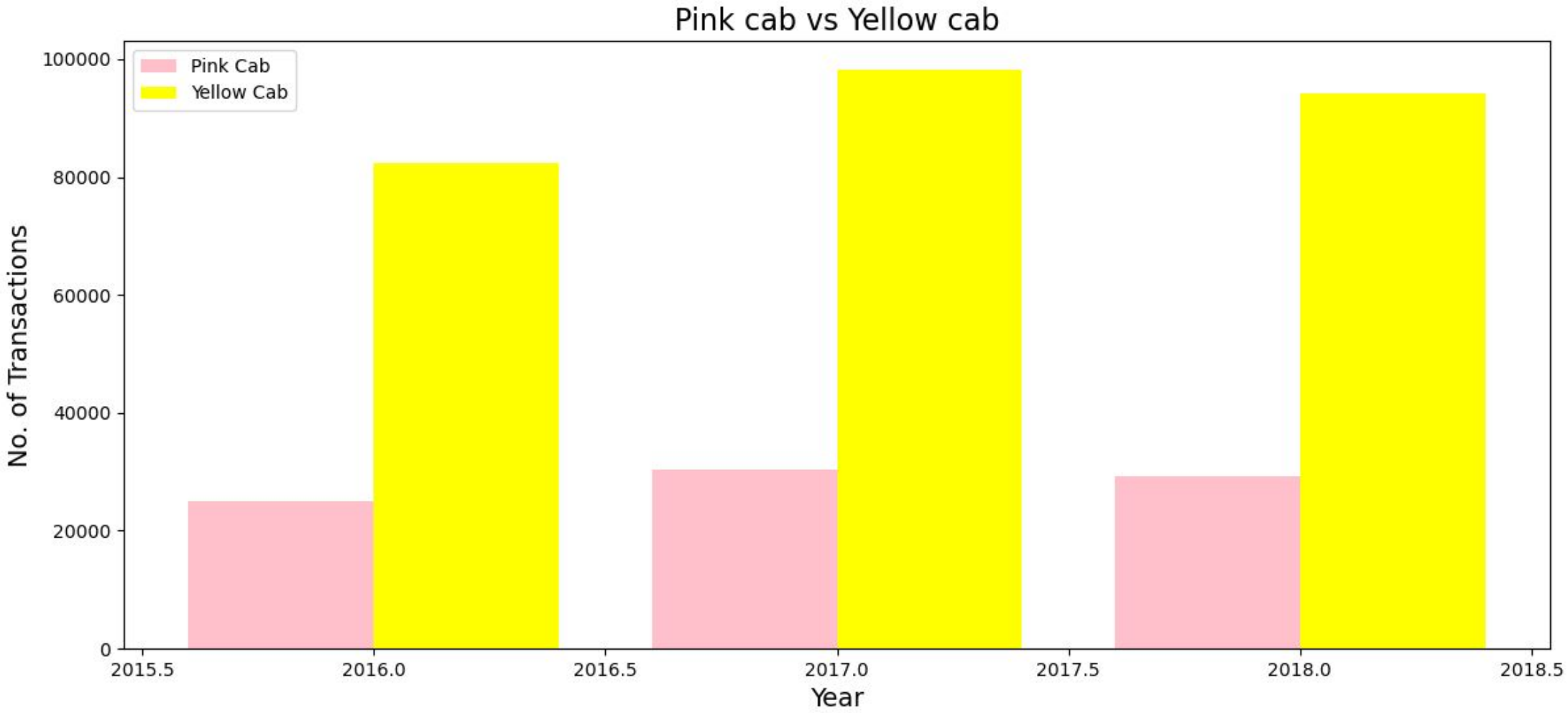
Key Questions:

- Is there a trend in profitability over the years (2016–2018)?

- Do Cost of Trip or Price Charged show seasonal variation?

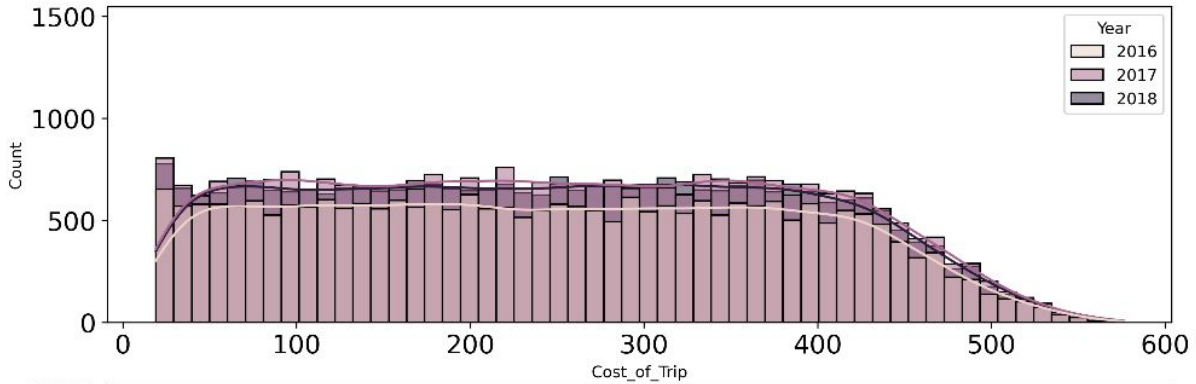
2. Exploratory Data Analysis (EDA)

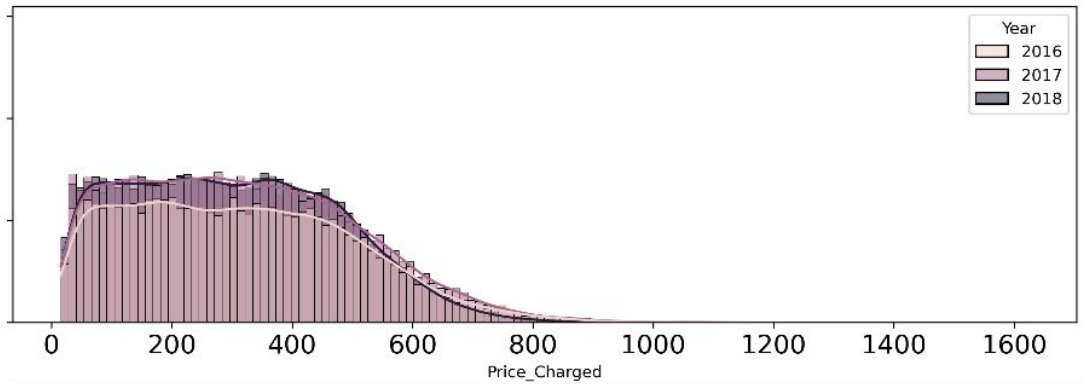
Overall Analysis: Yearly Trend

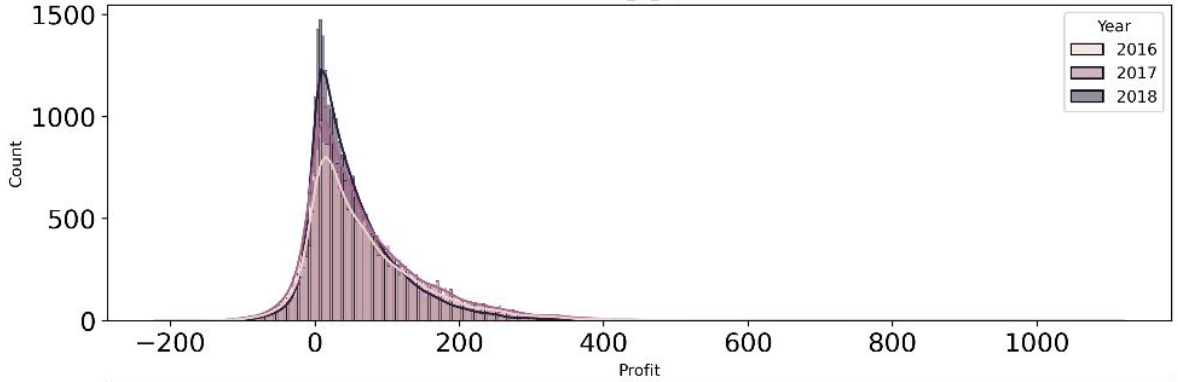


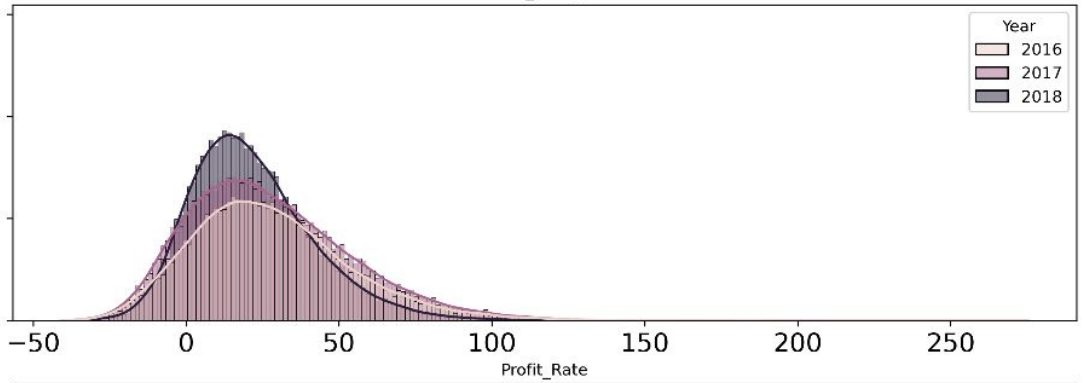
2. Exploratory Data Analysis (EDA)

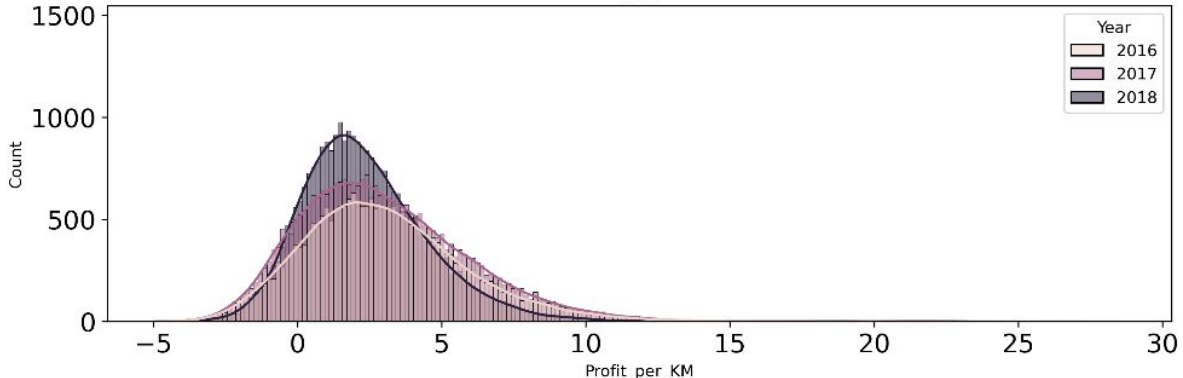
Overall Analysis: Yearly Trends for Pink Cab

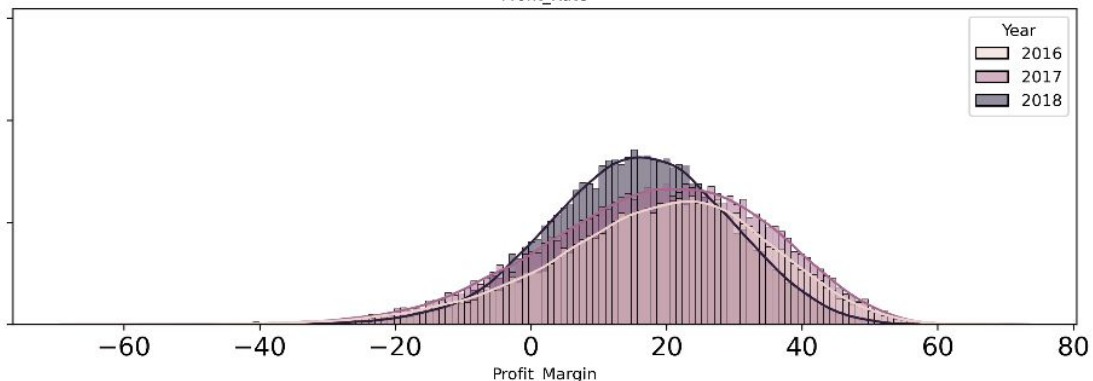






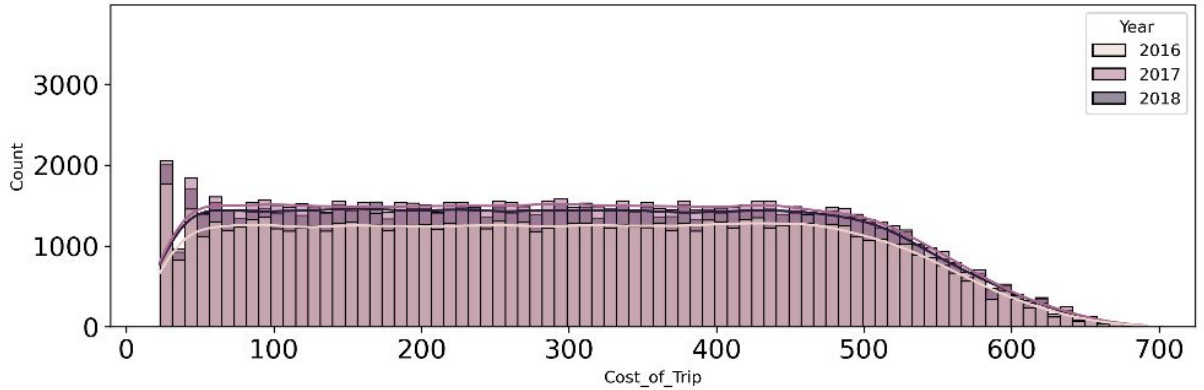


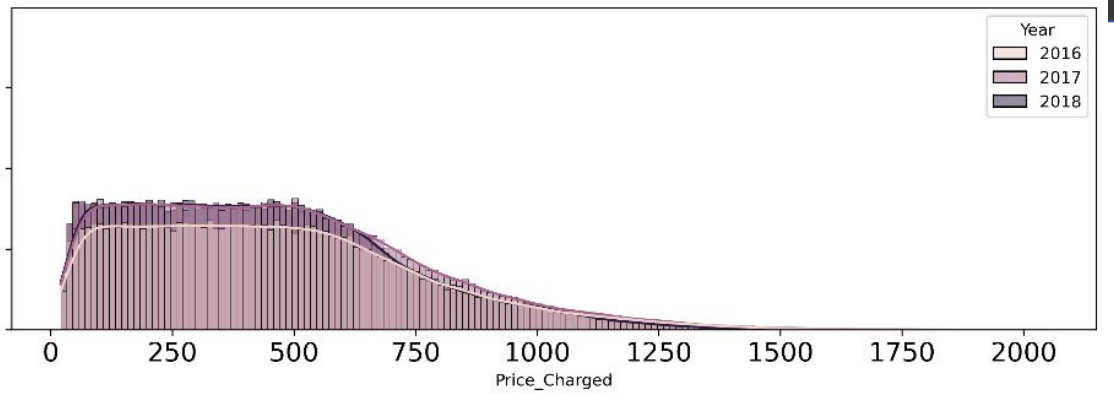


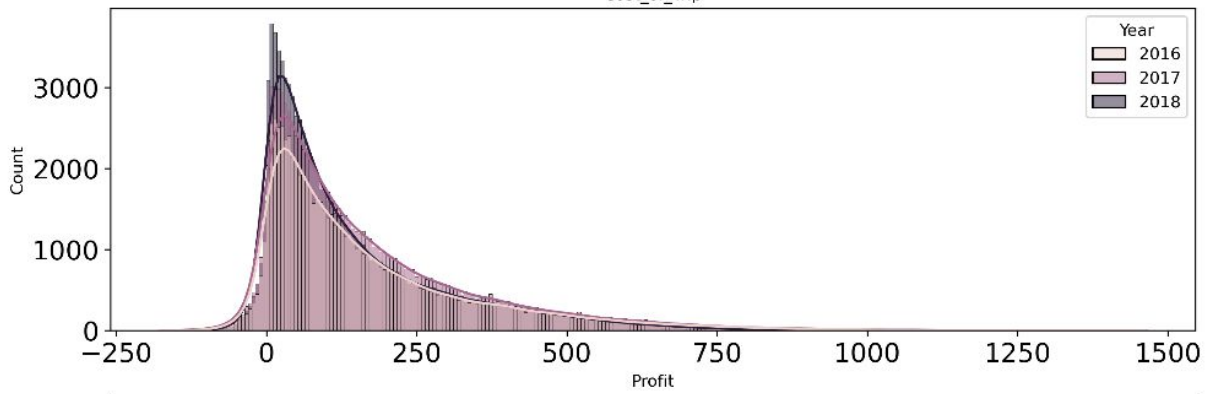


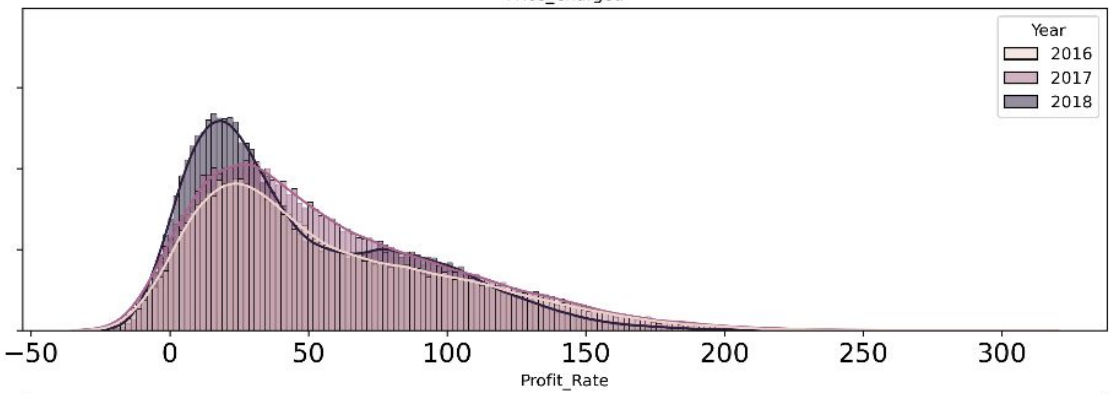
2. Exploratory Data Analysis (EDA)

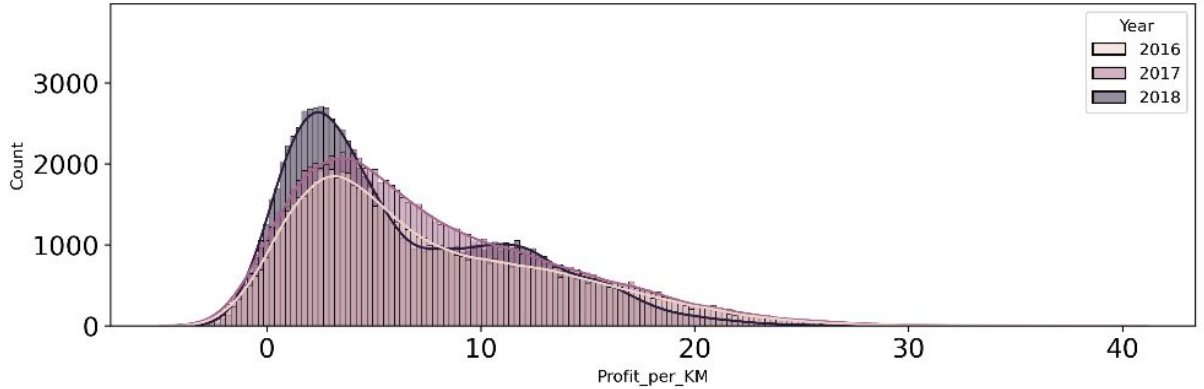
Overall Analysis: Yearly Trends for Yellow Cab

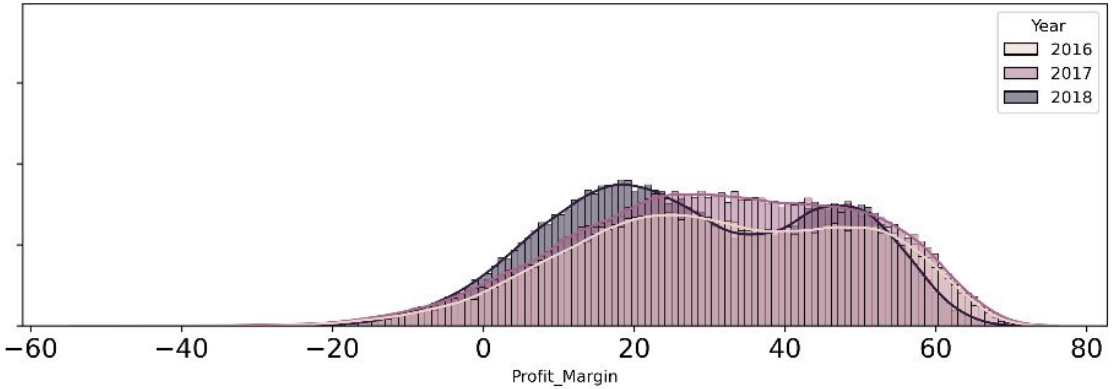






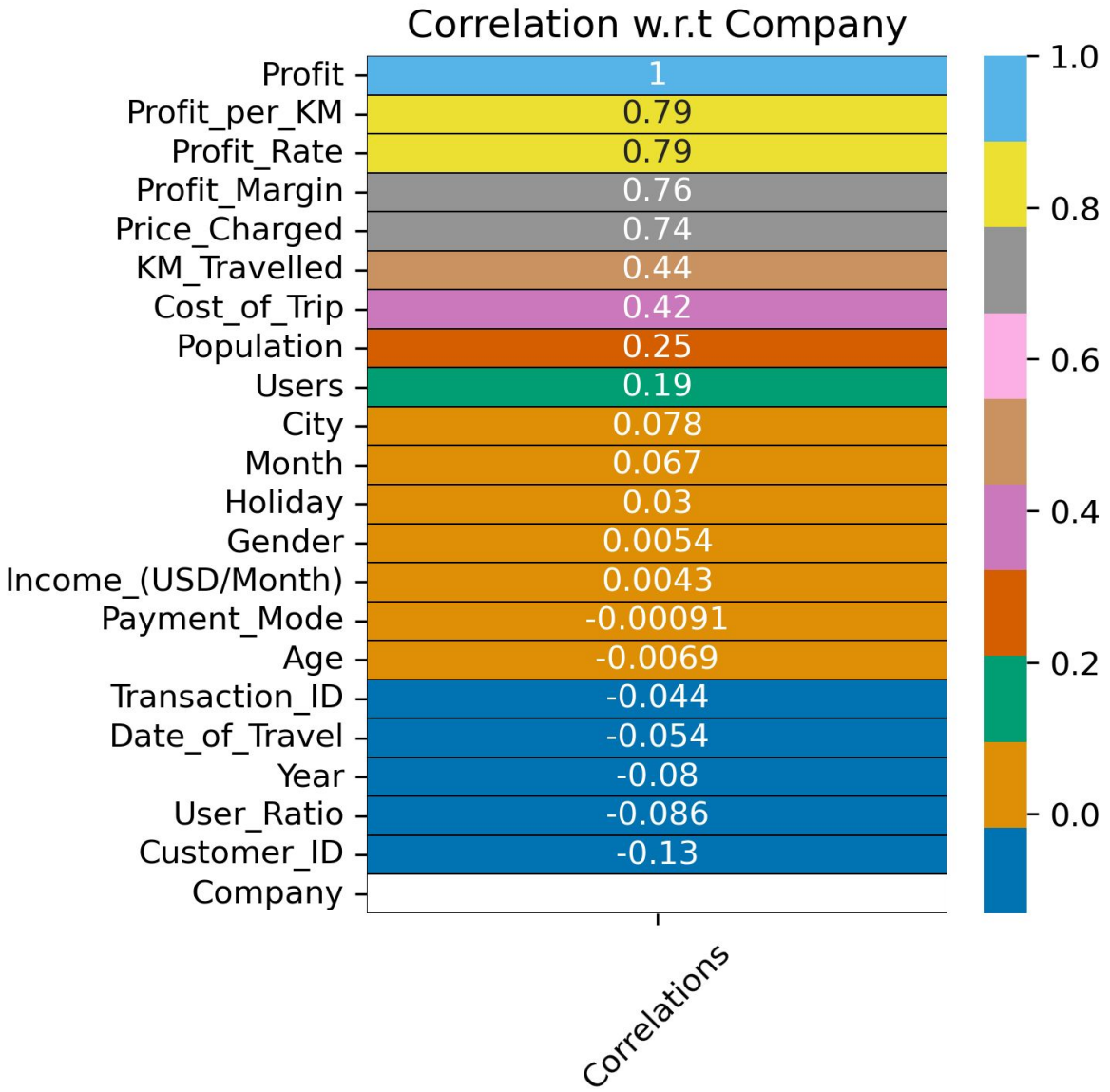






Profit and customer base Analysis

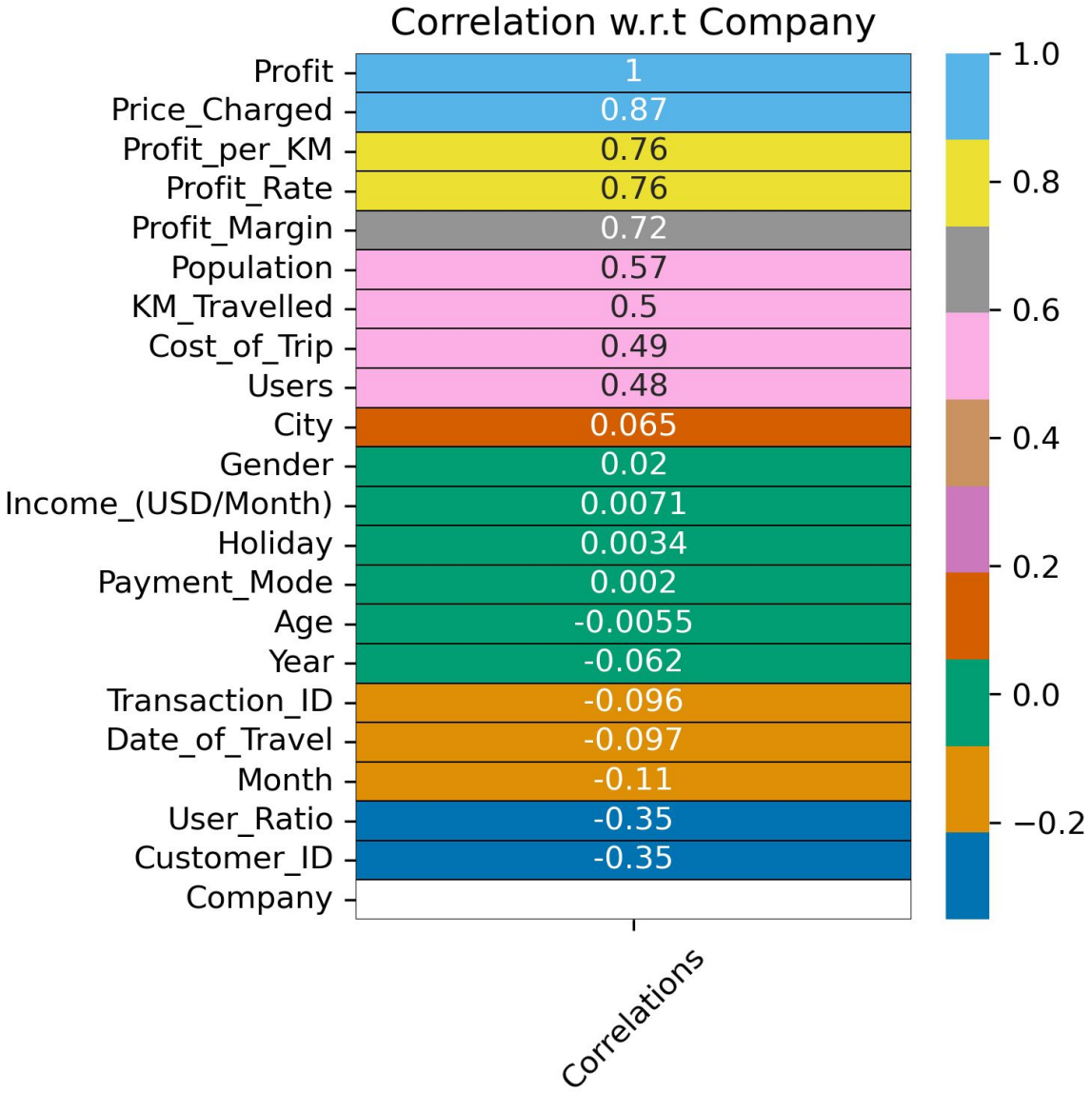
- **"Which features have the most influence (positive or negative) on Profit for Pink Cab ?"**



In the pink Cab, profit is strongly correlated to Price_Charged, moderately correlated to KM travelled and Cost_of_Trip.

Profit and customer base Analysis

- **"Which features have the most influence (positive or negative) on Profit for Yellow Cab ?"**

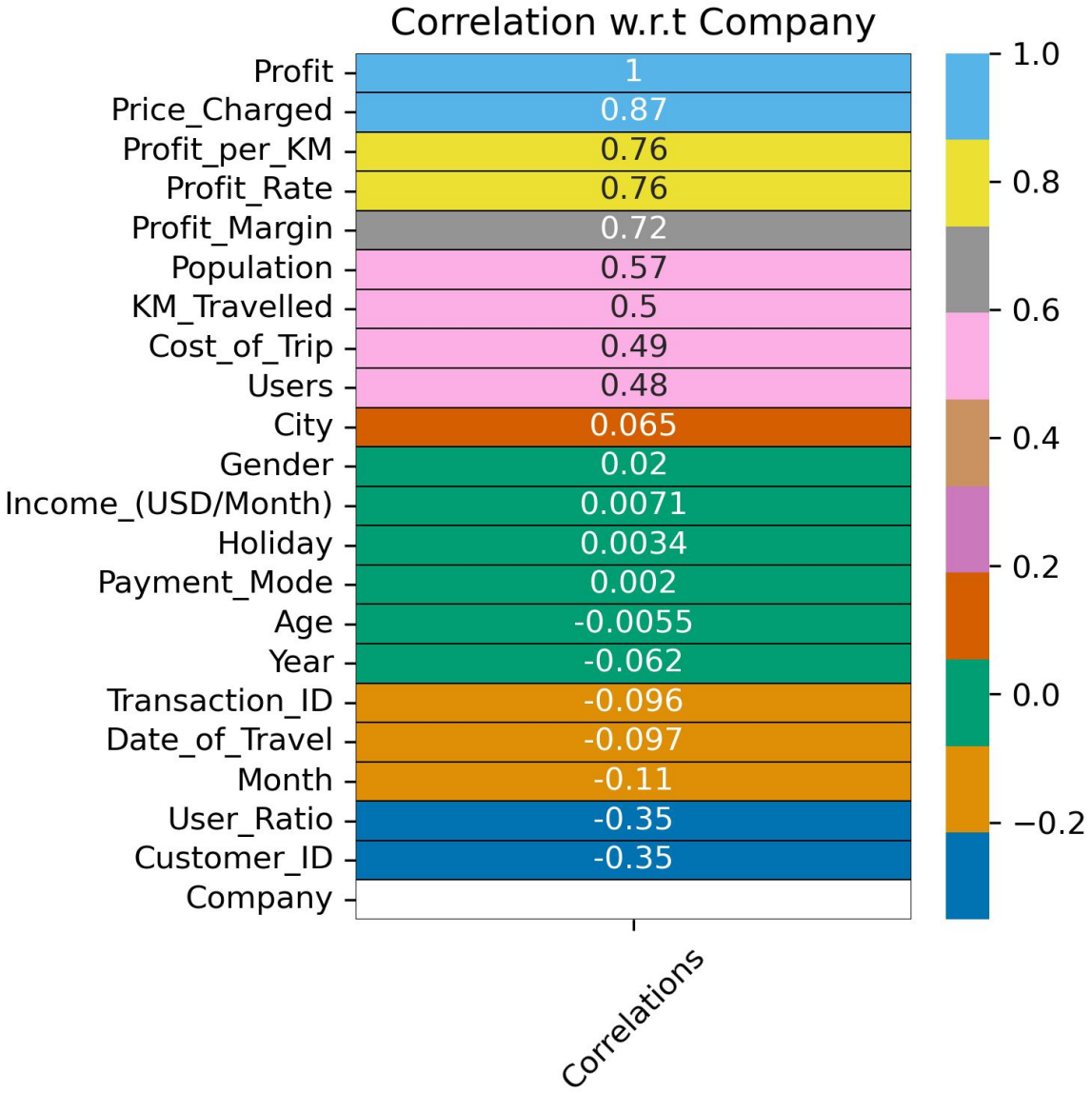


In the yellow cab, profit is strongly correlated to Price_Charged, moderately correlated to Population and KM travelled.

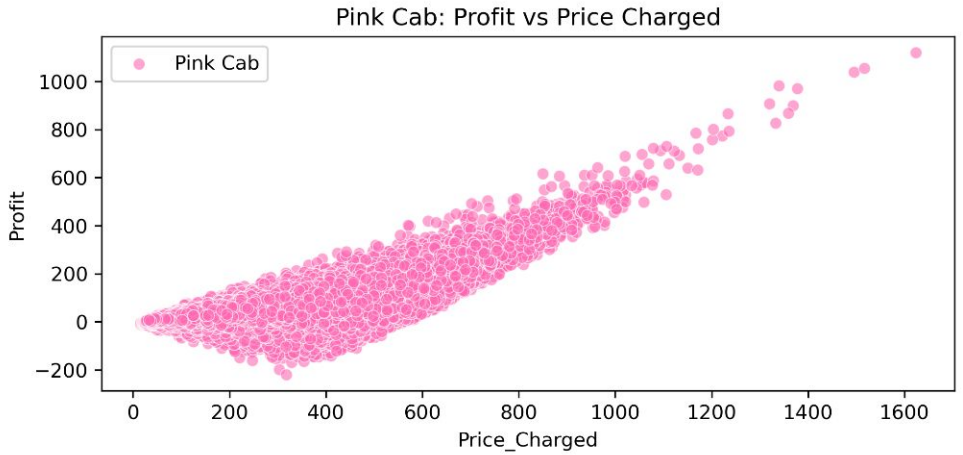
Since in both cabs, profit is strongly correlated to Price charged and moderately correlated to KM travelled. Let's check how does the scatter plot looks like.

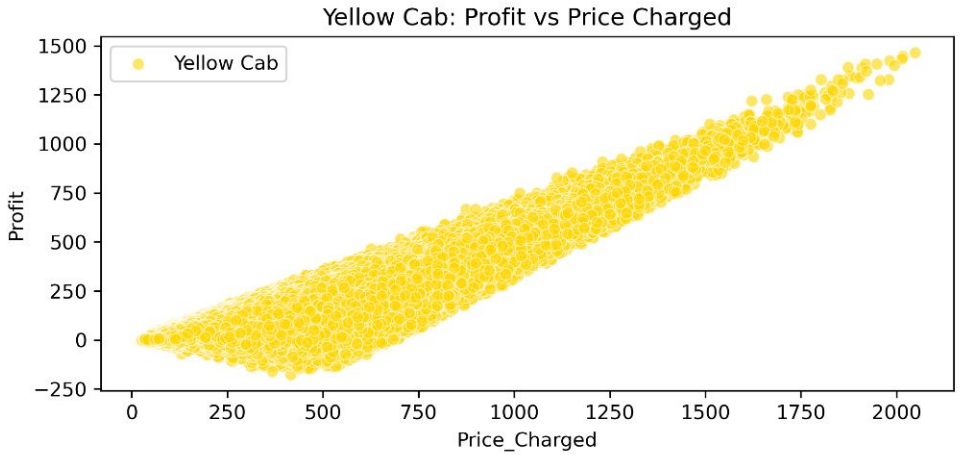
Profit and customer base Analysis

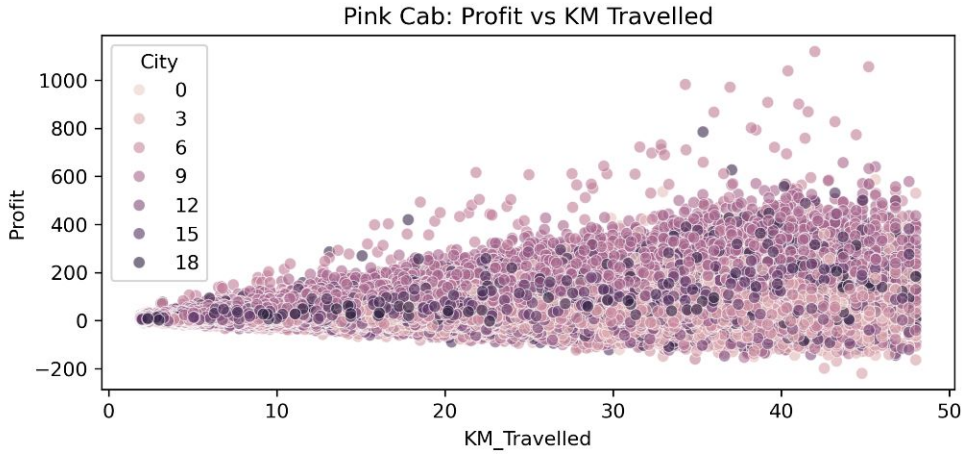
- **"Which features have the most influence (positive or negative) on Profit for Yellow Cab ?"**

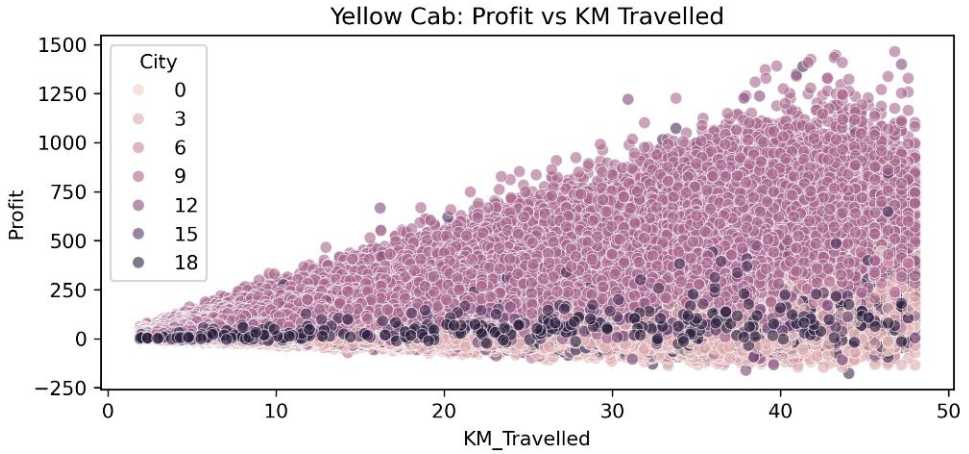


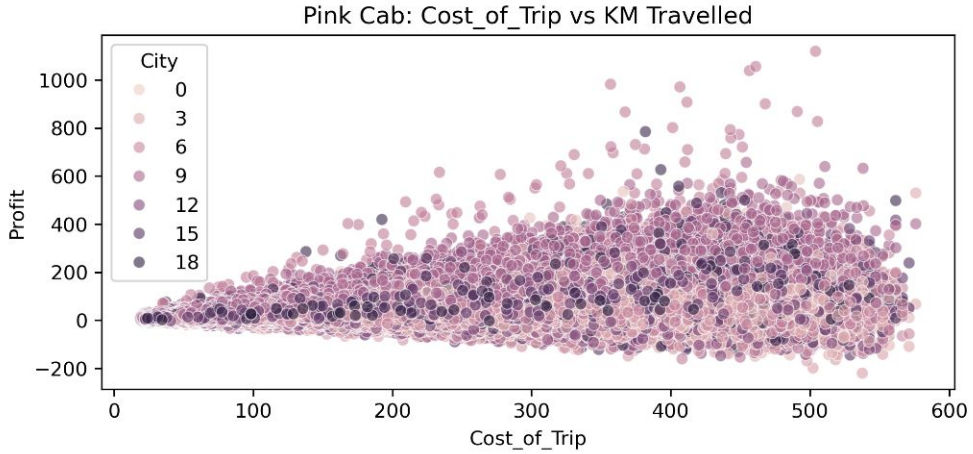
In the yellow cab, profit is strongly correlated to Price_Charged, moderately correlated to Population and KM travelled.

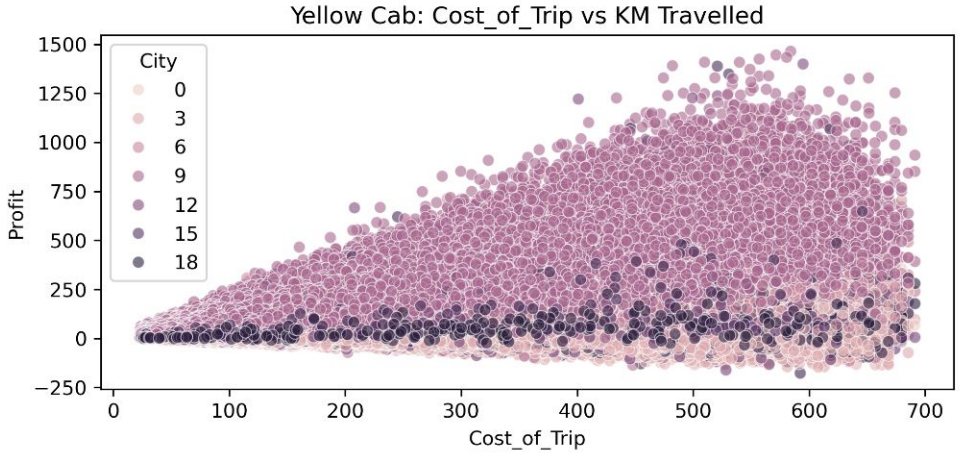

Pink Cab: Profit vs Price Charged
Pink Cab
Profit
Price_Charged


Yellow Cab: Profit vs Price Charged
Yellow Cab
Profit
Price_Charged


Pink Cab: Profit vs KM Travelled
City
0
3
6
9
12
15
18
Profit
KM_Travelled


Yellow Cab: Profit vs KM Travelled
City
0
3
6
9
12
15
18
Profit
KM_Travelled


Pink Cab: Cost_of_Trip vs KM Travelled
City
0
3
6
9
12
15
18
Profit
Cost_of_Trip


Yellow Cab: Cost_of_Trip vs KM Travelled
City
0
3
6
9
12
15
18
Profit
Cost_of_Trip

Recommendations

- Profit is strongly correlated with **kilometers traveled**, **price charged**, and **cost of trip**, which is expected since:
- Profit=Price Charged−Cost of Trip
- As the **kilometers traveled** increase, both the **price charged** and **cost of trip** generally increase, resulting in higher profits.
- However, when comparing the two companies, we observe that **Yellow Cab shows more consistent (less variable) profits** for similar values of distance, price, and cost. This suggests a more **stable and predictable profit model**, which is advantageous for investors.
- Therefore, based on the analysis, **investing in Yellow Cab is the better option** due to its **more uniform and reliable profitability** in similar operating conditions.

Thank You