Data Glacier
Your Deep Learning Partner

Exploratory Data Analysis
G2M Insights for Cab Investment Case Study

July 28, 2024

Agenda

- Executive Summary
- Problem Statement
- Approach
- EDA
- EDA Summary
- Recommendations

Executive Summary

XYZ, a private US firm, is exploring investment opportunities in the cab industry. To support their Go-to-Market strategy, an in-depth market analysis has been conducted using data from two prominent cab companies. The objective is to identify which company presents the most promising investment opportunity based on various customer and transaction metrics.

Problem Statement

XYZ needs a comprehensive analysis of the market and performance of two potential cab companies. The goal is to provide actionable insights to help XYZ identify the most promising investment opportunity.

Key Areas of Focus:
1. Customer Volume and Trends
2. Financial Performance
3. Customer Demographics and Segmentation
4. Geographical Analysis

Approach

1. Data Understanding and Preparation:
- Review and understand field names, data types, and relationships across datasets.
- Merge datasets to create a comprehensive master dataset.
- Rename columns and format data types where necessary.
- Perform feature engineering.

2. Exploratory Data Analysis (EDA):
- Conduct initial EDA to uncover patterns, trends, and outliers.
- Formulate and test hypotheses to investigate various aspects of the data.

Approach

3. Analysis and Visualization:
- Use statistical analysis and data visualization techniques to derive insights.
- Compare and contrast the performance of the two cab companies.

4. Findings and Recommendations:
- Summarize findings and provide data-driven recommendations to XYZ's executive team.
- Identify which company is the better investment opportunity based on the analysis.

Case Study Flowchart
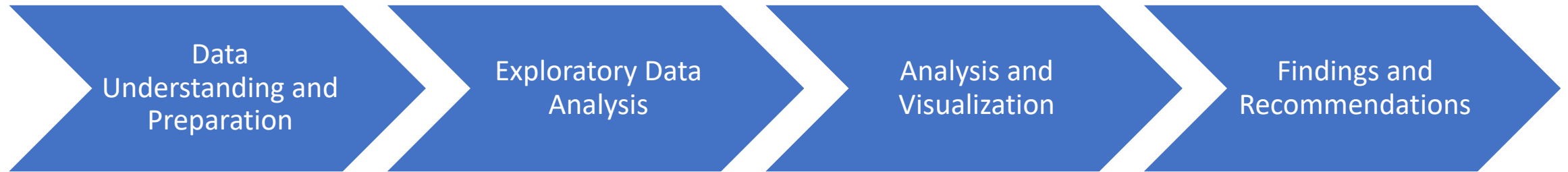
EDA with Data Visualization

1. Relationship between trip profit and price charged
2. Number of transactions per company
3. Gross revenue per company
4. Gross profit per company
5. Number of transactions per gender
6. Company profitability per city
7. Gender profitability per city
8. Profitability per year
9. Profitability trend
10. Total number of customers per year per company

EDA with Data Visualization

11. Customer number trend
12. Number of customers per age group
13. Number of customer per income group
14. Gender profitability per age group
15. Gender profitability per income group

Insights Drawn from EDA

Correlation of Numeric Features

|	Date_of_Travel	Transaction_ID	KM_Travelled	Price_Charged	Cost_of_Trip	Trip_Profit	Customer_ID	
Date_of_Travel	1.000000	0.993030	-0.001621	-0.055559	-0.004484	-0.090672	-0.017653	-0.0(
Transaction_ID	0.993030	1.000000	-0.001429	-0.052902	-0.003462	-0.087130	-0.016912	-0.0
KM_Travelled	-0.001621	-0.001429	1.000000	0.835753	0.981848	0.462768	0.000389	-0.0(
Price_Charged	-0.055559	-0.052902	0.835753	1.000000	0.859812	0.864154	-0.177324	-0.0(
Cost_of_Trip	-0.004484	-0.003462	0.981848	0.859812	1.000000	0.486056	0.003077	-0.0(
Trip_Profit	-0.090672	-0.087130	0.462768	0.864154	0.486056	1.000000	-0.306527	-0.0(
Customer_ID	-0.017653	-0.016912	0.000389	-0.177324	0.003077	-0.306527	1.000000	-0.0(
Age	-0.001346	-0.001267	-0.000369	-0.003084	-0.000189	-0.005093	-0.004735	1.0(
Income_(USD/Month)	-0.001368	-0.001570	-0.000544	0.003228	-0.000633	0.006148	-0.013608	0.0(

Noteworthy correlations:
Distance travelled is strongly correlated with price charged and trip cost but weakly correlated with trip profit.

Trip Profit vs. Price Charged

Relationship Between Price and Profit per Trip
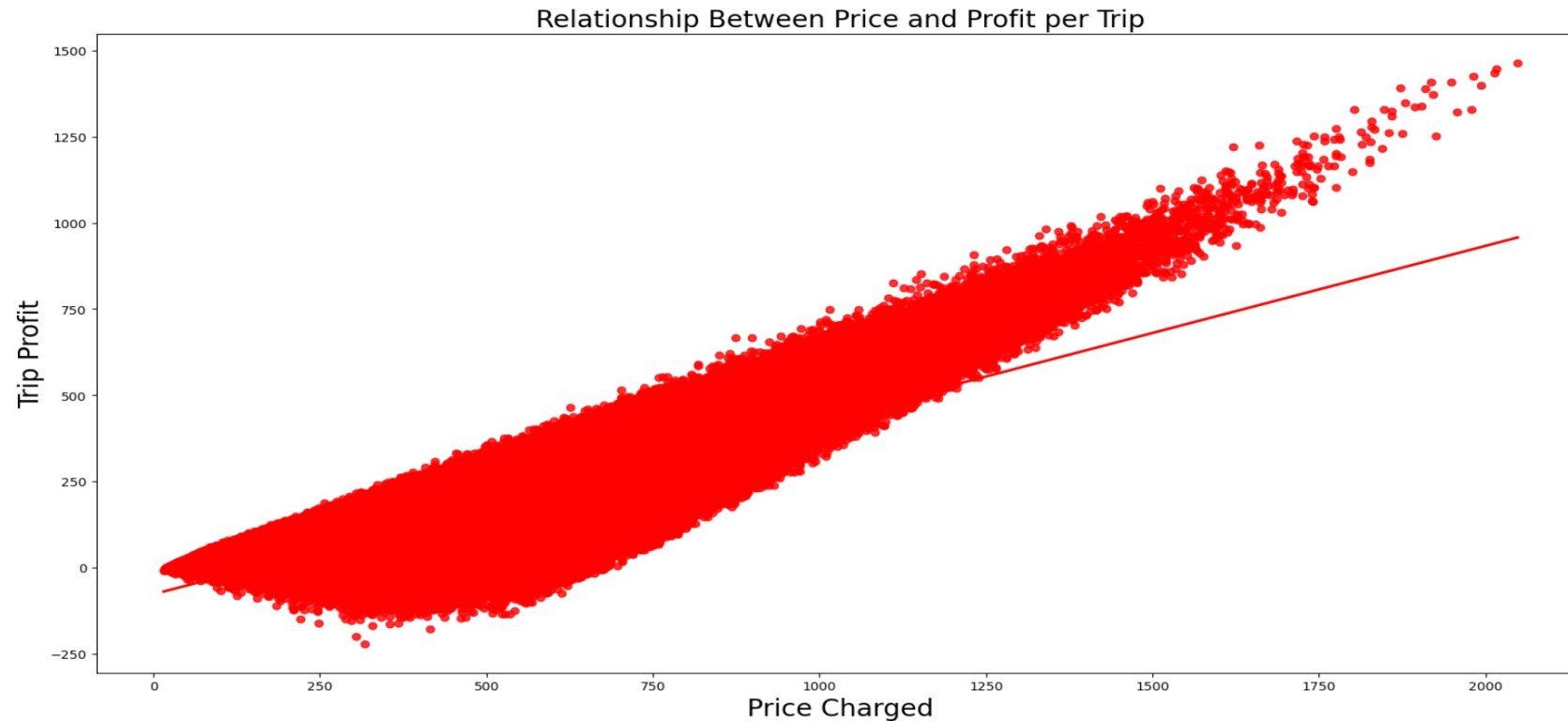
Weak correlation which might be due to price discounts or deeper issue with pricing model.

Data Glacier

# Transactions, Gross Revenue and Gross Profit per Company
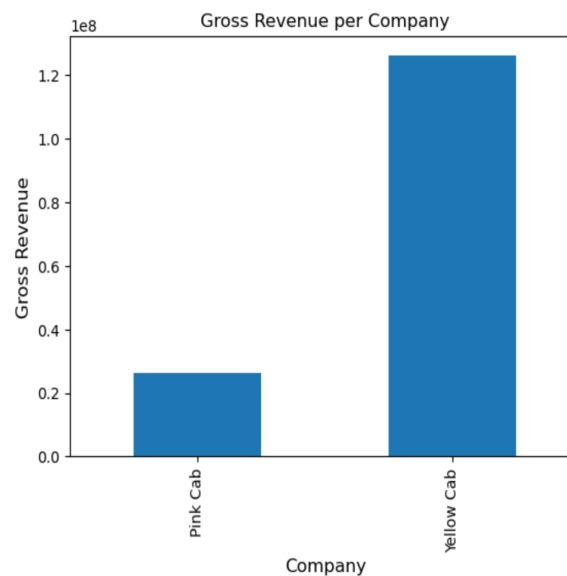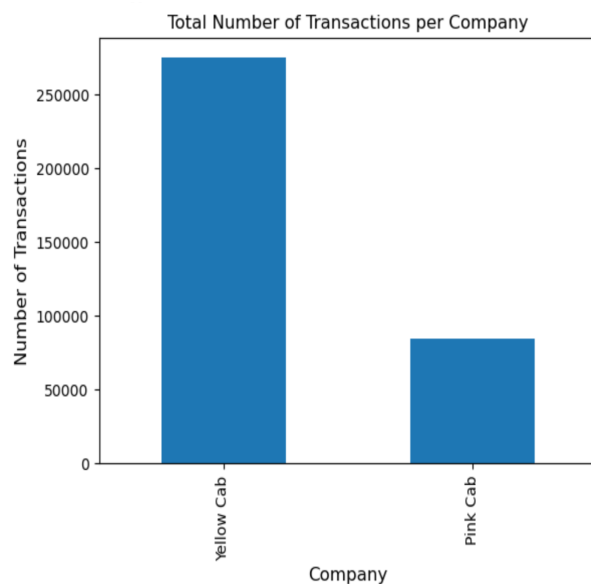
Yellow Cab outperforms Pink Cab in the areas of number of transactions, revenue and profit. Additionally, Yellow Cab does just about 3 times as many transactions but is about 7.5 times as profitable.

Number of Transactions by Gender

Total Number of Transactions per Gender
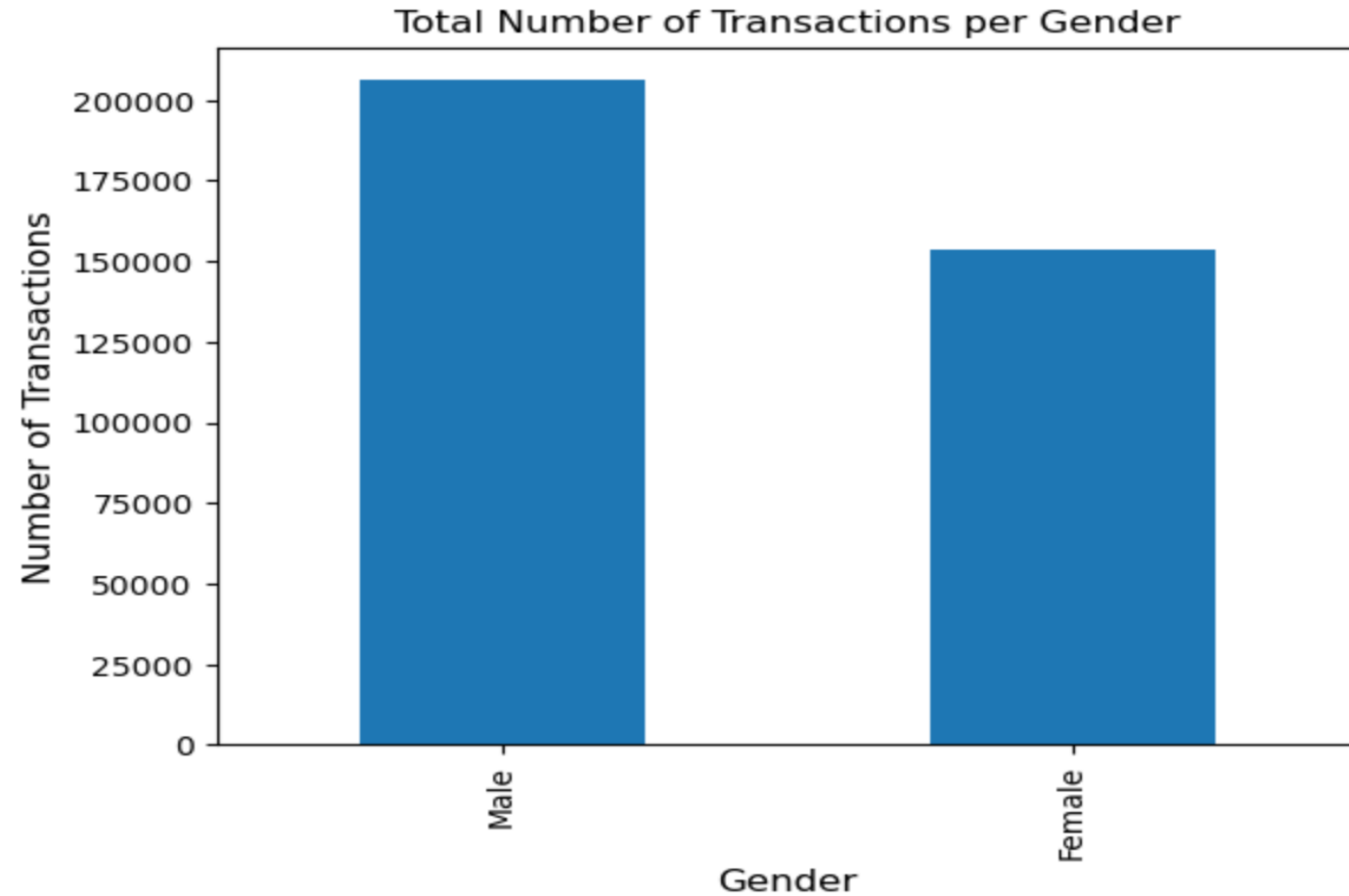
There are more male users across both platforms relative to females.

Company Profitability per City
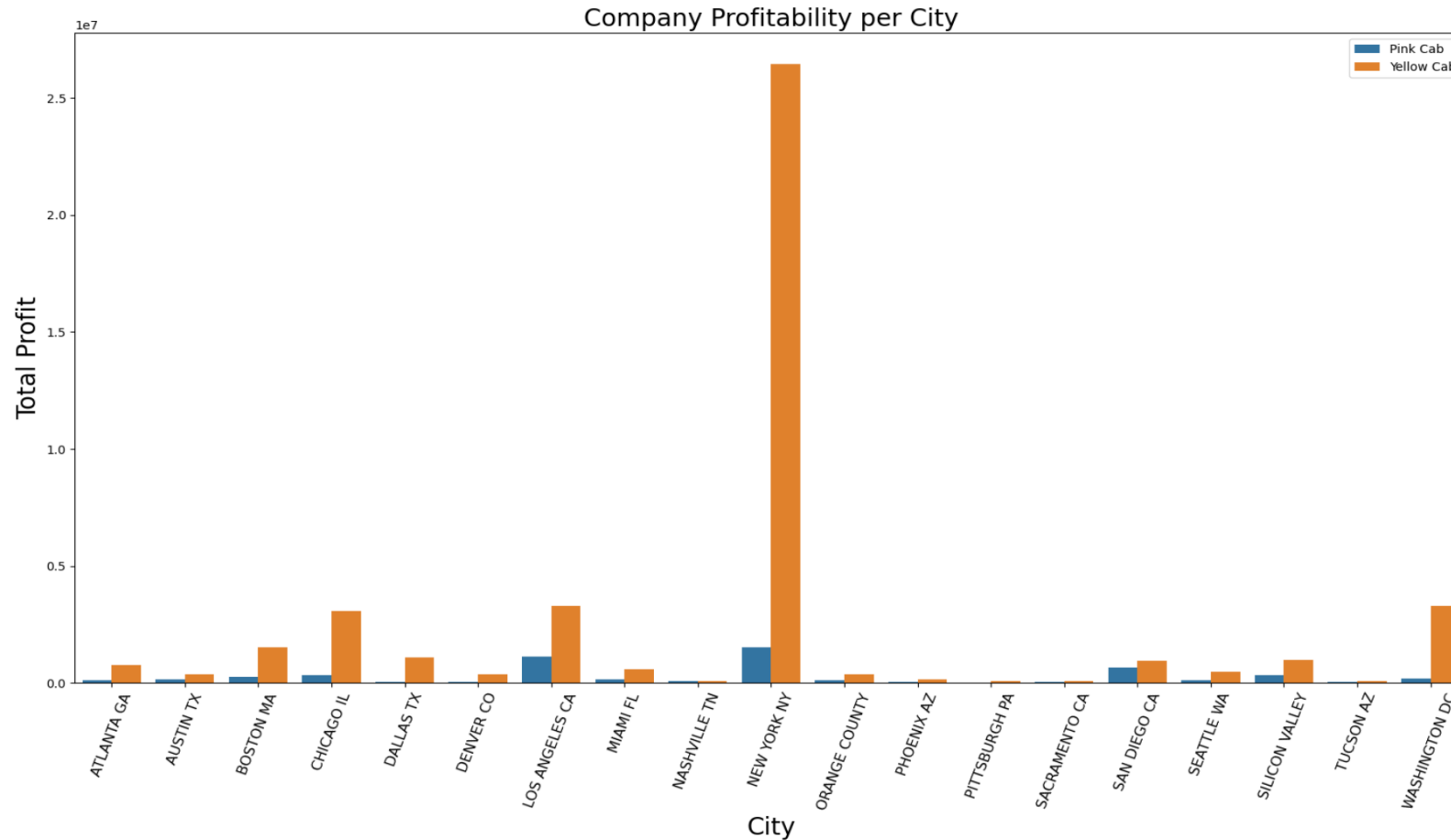
Yellow Cab outperforms its competition in all 20 studied cities.
Also, New York was about 8 times as profitable as the next most profitable city, which is Los Angeles.
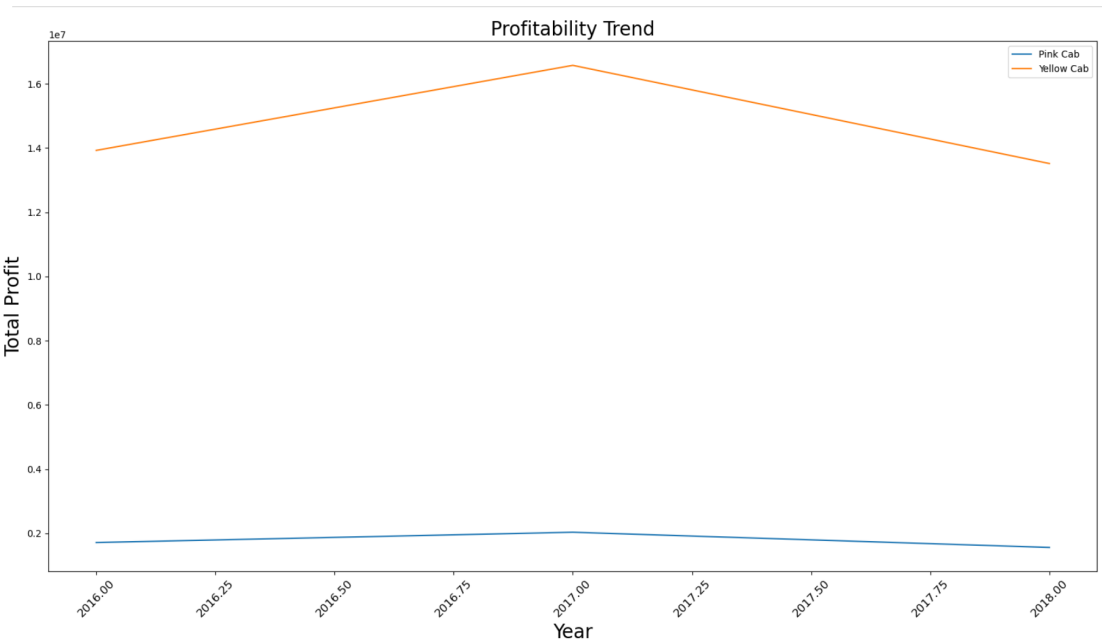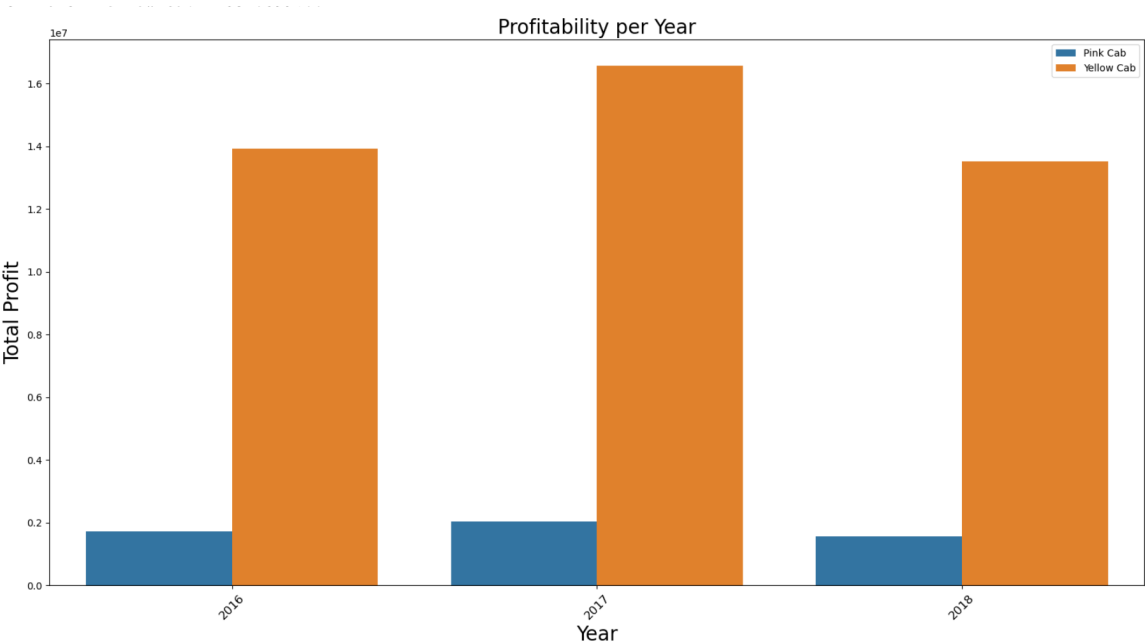
Gender Profitability per City

Gender Profitability per City
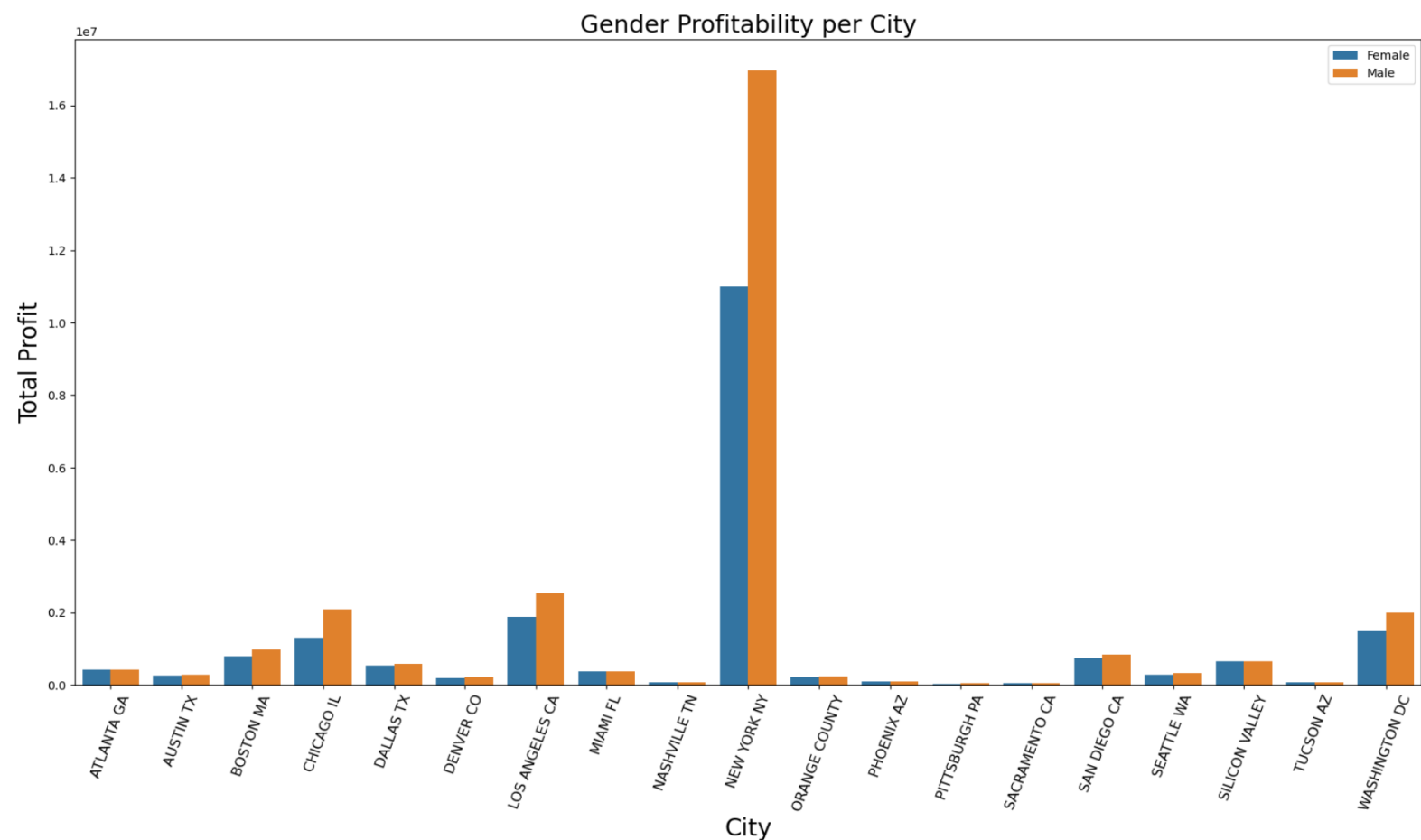
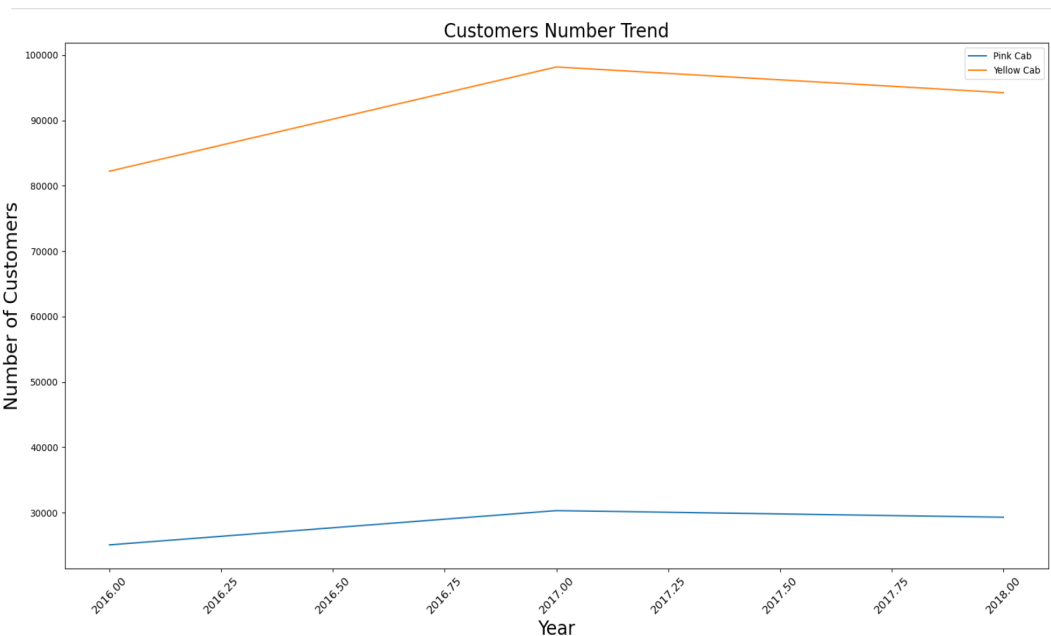There are more male users in more profitable cities in the study, with the widest gender gap in New York.

Yellow Cab is about 7 times as profitable as Pink Cab.

Number of Customers per Year | Customer Numbers Trend

Both companies have stable customer bases, but Yellow Cab has about 3 times more customers.
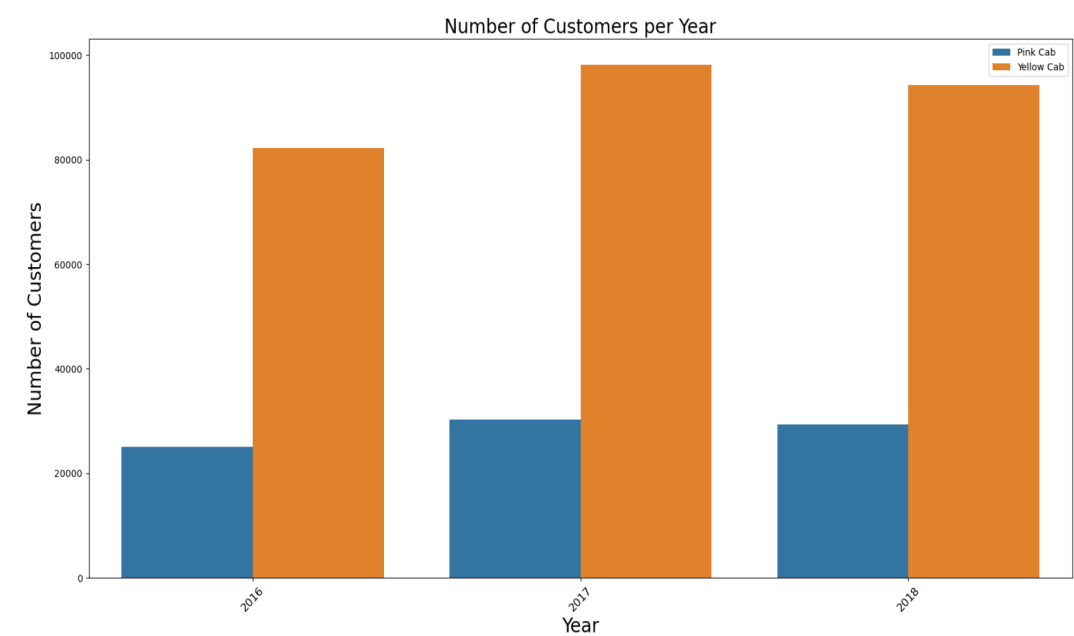
Customer Distribution by Age and Income

The 26–35 age group is the most frequent user group, followed by 19–25 and 36–45 age groups. Higher income groups use the services more frequently. Both observations are true for both genders.
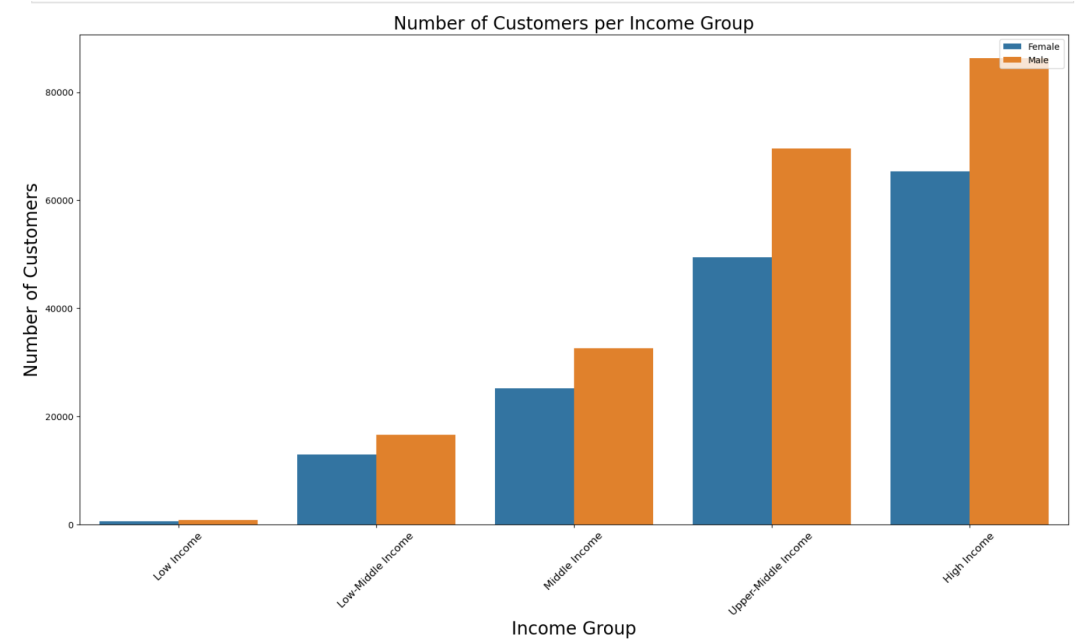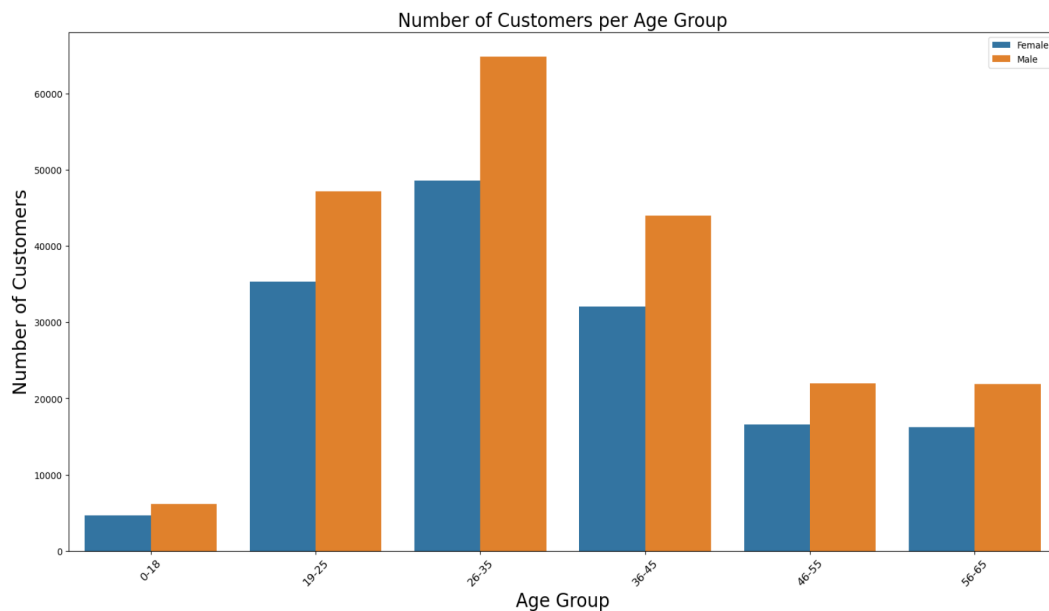
Profitability per Age and Income Groups

Profitability per age and income groups mirrors the customer distribution: the 26–35 age group is the most profitable, and higher income groups more profitable.
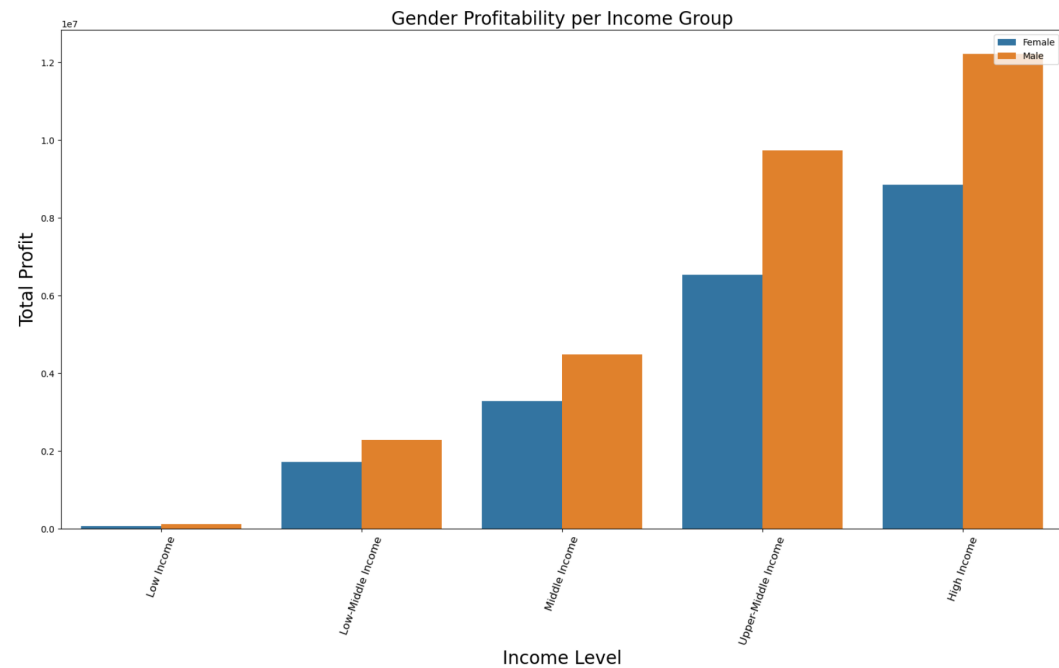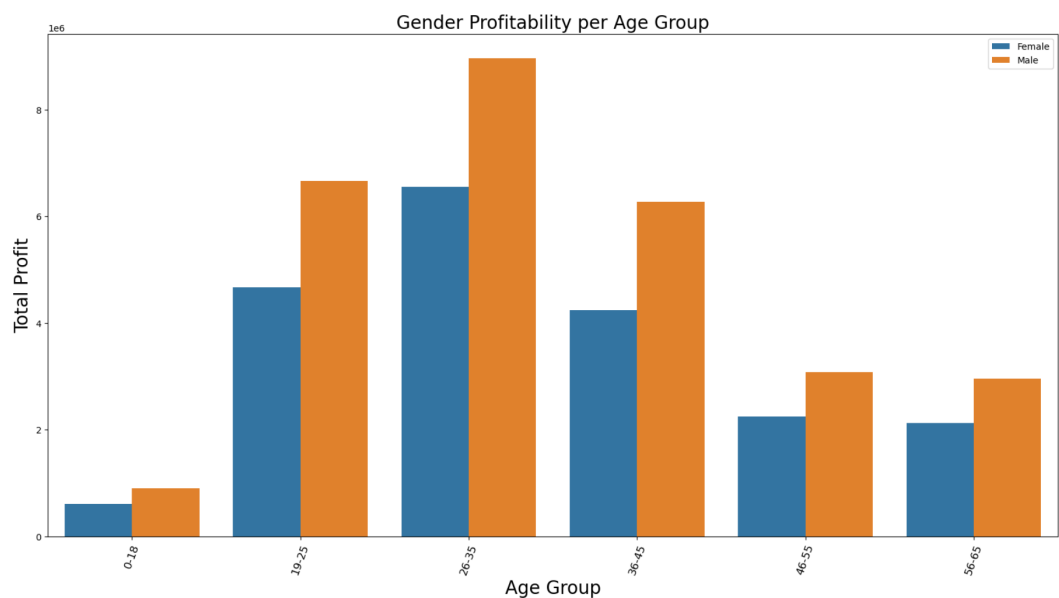
EDA Summary

1. Distance traveled is strongly correlated with trip cost but weakly correlated with trip profit, possibly due to price discounts or issues with the pricing model.
2. Yellow Cab outperforms Pink Cab in all 20 cities studied, as well as in transactions, revenue, and profit. Yellow Cab has about 3 times more transactions and is 7.5 times more profitable.
3. New York is 8 times more profitable than Los Angeles, the second most profitable city.
4. Both companies have stable customer bases, but Yellow Cab has about 3 times more customers.
5. There are more male users than female users on both platforms, especially in more profitable cities, with the widest gender gap in New York.
6. The 26–35 age group is the most frequent user group, followed by 19–25 and 36–45 age groups. Higher income groups use the services more frequently, true for both genders.
7. Profitability per age and income groups mirrors the customer distribution: the 26–35 age group is the most profitable, and higher income groups are more profitable.

Recommendations

Based on Analysis

1. Yellow Cab is the superior investment option, outperforming Pink Cab in all metrics and cities studied.

2. Focus Areas:
- Establish major urban hubs: New York, Los Angeles, Washington DC, Chicago, and Boston, as these are the most profitable.
- Launch a special campaign for New York, the most profitable city by a significant margin.

3. Target Demographics:
- Focus on male users, particularly those aged 19-45, and higher income earners.

4. Pricing Review:
- Review the pricing model to ensure it is optimized and free of issues.

Thank You