Exploratory Data Analysis

G2M insight for Cab Investment firm

21-Apr-2024

Problem Statement

- 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** : Provide actionable insights to help XYZ firm in identifying the right company for making investment.

The analysis has been divided into four parts:

- Data Understanding and Exploration
- Profitability Analysis
- Market Positioning and Competitive Analysis
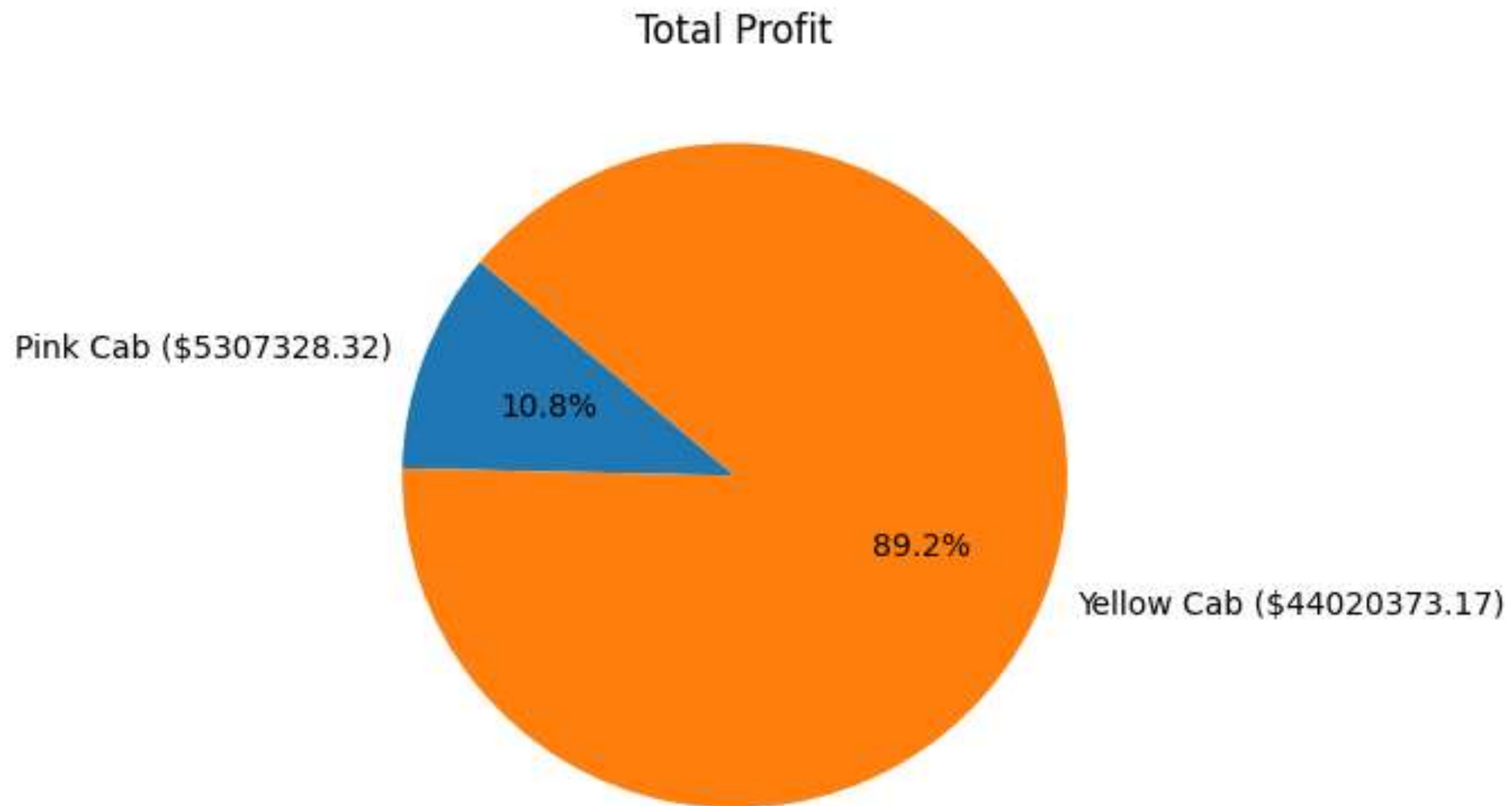- Recommendations for investment

Data Exploration

- 18 Features(including 1 derived feature)
- Timeframe of the data: 2016-01-31 to 2018-12-31
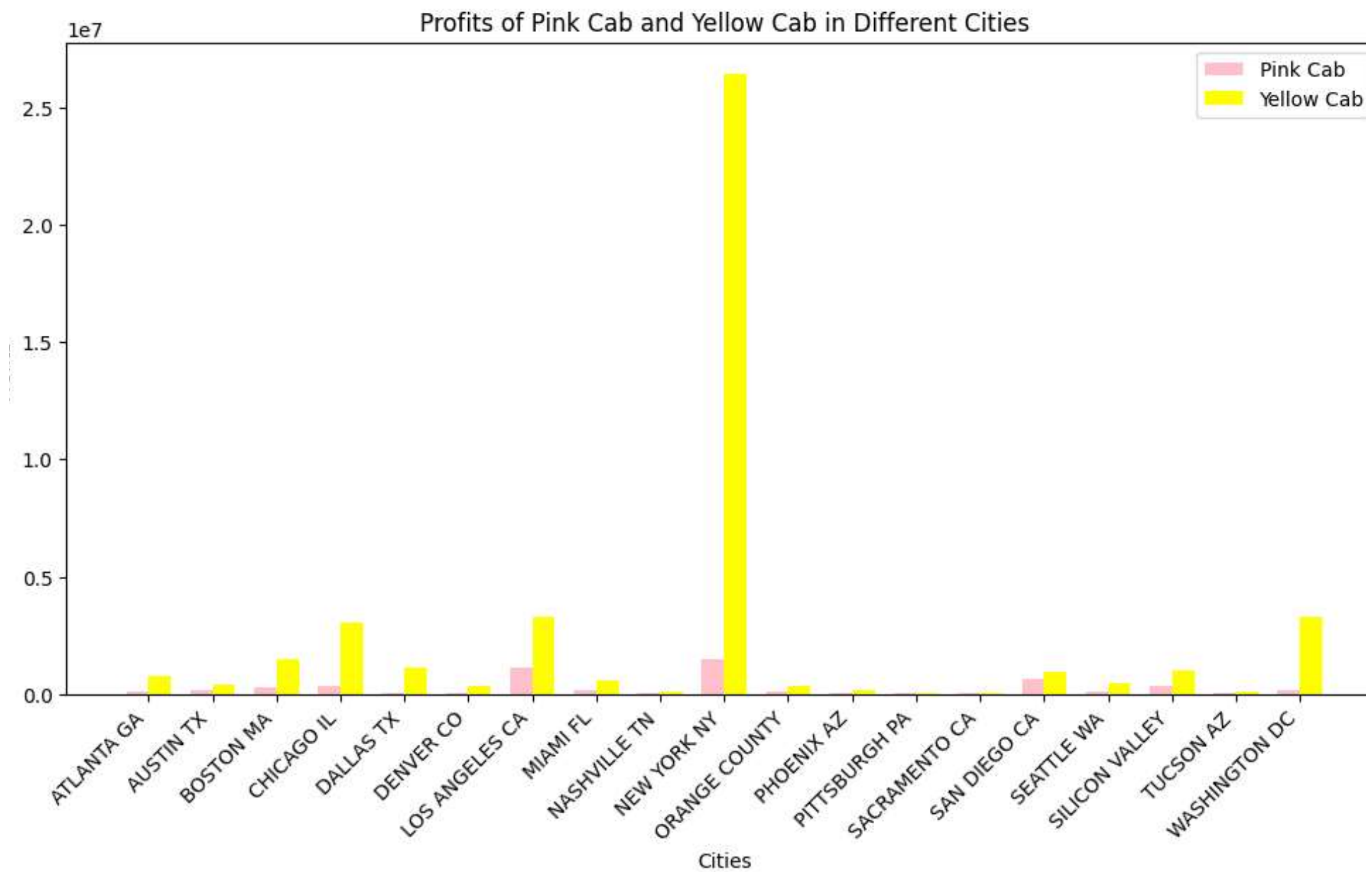- Total data points :359,393

Assumptions:

- Outliers are present in Price_Charged feature but due to unavailability of trip duration details ,we are not treating this as outlier.

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

- Users feature of city dataset is treated as number of cab users in the city.We have assumed that this can be other cab users as well(including Yellow and Pink cab)
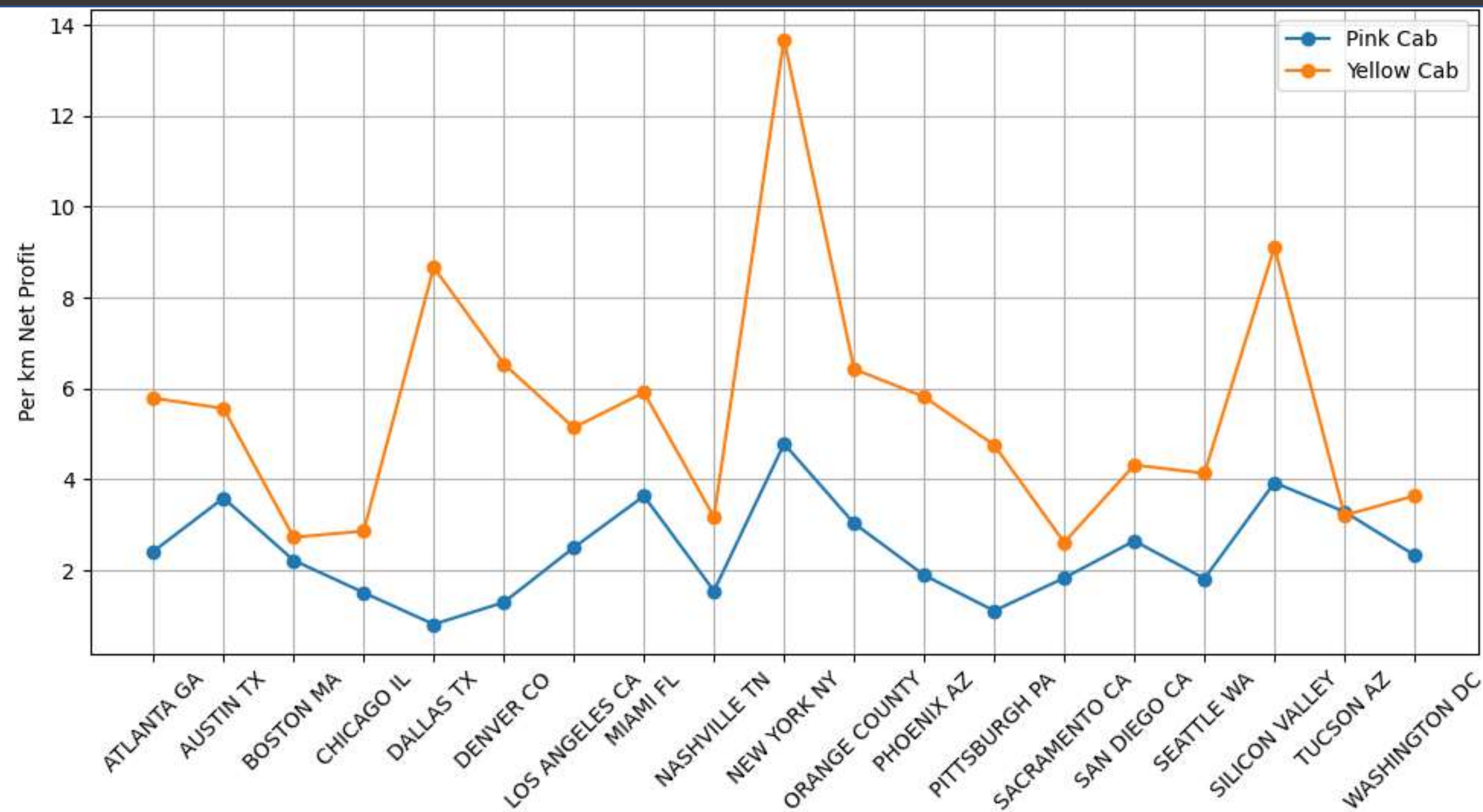
Total Profit Analysis

Total Profit

Pink Cab ($5307328.32)

10.8%

89.2%

Yellow Cab ($44020373.17)

City Wise Profit Analysis

Profits of Pink Cab and Yellow Cab in Different Cities

- Yellow Cab outperforms Pink Cab across multiple cities, with significantly higher profits.

- For instance, in New York, Yellow Cab's profits ($26,451,080.77) dwarf Pink Cab's ($1,511,474.39).

- Similarly, in Chicago, Yellow Cab's profits ($3,068,590.93) far exceed Pink Cab's ($318,722.49).

- Given these substantial differences in profitability, investing in Yellow Cab presents a more lucrative opportunity for XYZ firm.

Per Kilometer Average Net Profit Analysis

City wise Ride Analysis: Average Metrics and Earnings Breakdown

1. Pink Cab Data							
Cities	Avg KM	Avg Price	Avg Cost	Per KM Price	Per KM Cost	Total Num of Rides	Per KM Avg Net Profit
ATLANTA GA	22.539824	302.131555	247.619594	13.404344	10.985871	1762	2.418473
AUSTIN TX	21.969176	320.435070	241.502901	14.585667	10.992807	1868	3.592860
BOSTON MA	22.740746	300.551489	250.030529	13.216430	10.994825	5186	2.221605
CHICAGO IL	22.511070	281.613731	247.565821	12.510011	10.997514	9361	1.512496
DALLAS TX	22.021203	260.025652	242.105281	11.807968	10.994190	1380	0.813778
DENVER CO	22.459218	276.202453	247.088437	12.297955	11.001649	1394	1.296306
LOS ANGELES CA	22.635629	305.624252	248.955132	13.501911	10.998375	19865	2.503536
MIAMI FL	22.299925	326.006309	244.936783	14.619166	10.983749	2002	3.635417

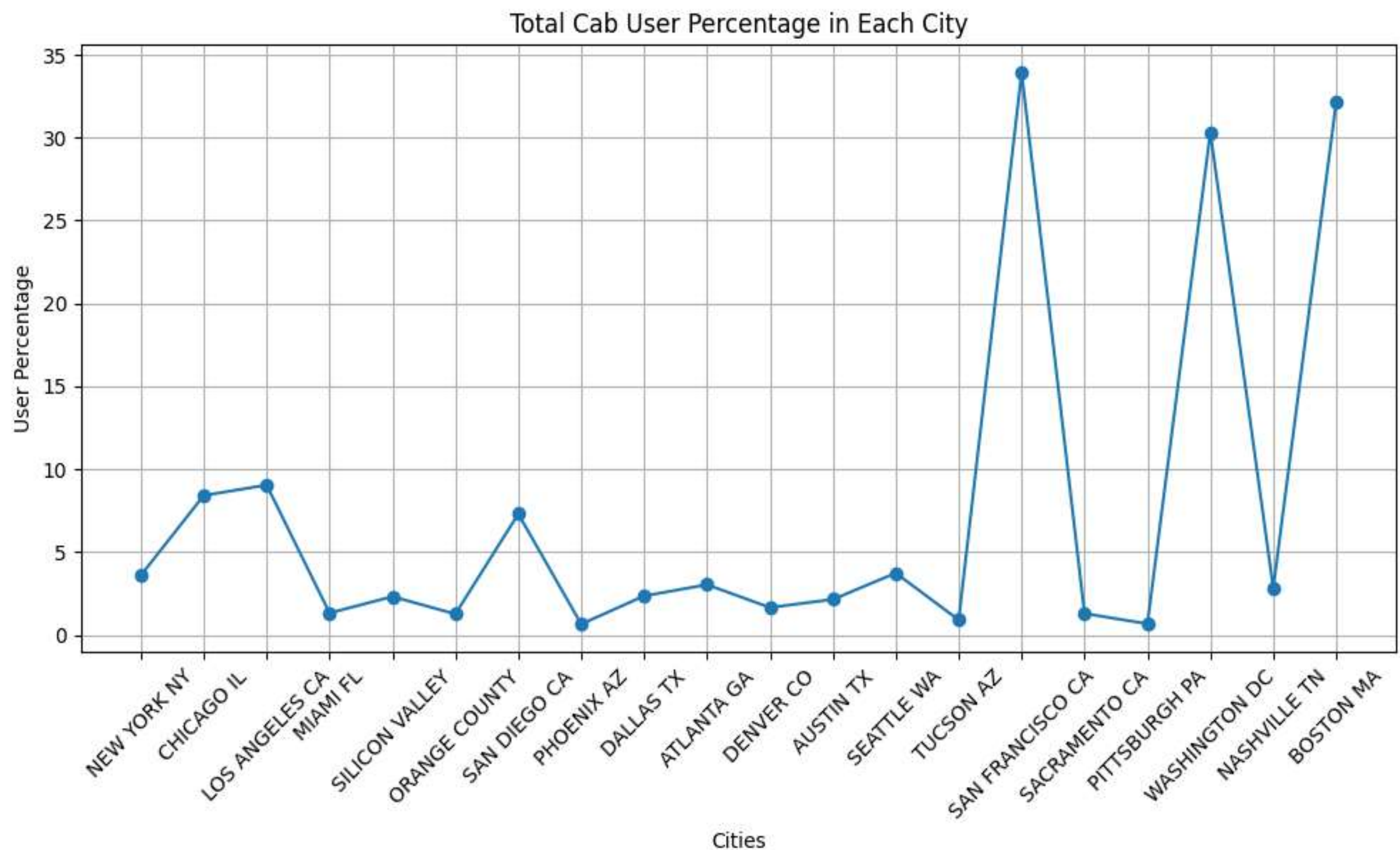| Cities	Avg KM	Avg Price	Avg Cost	Per KM Price	Per KM Cost	Total Num of Rides	Per KM Avg Net Profit
NEW YORK NY	22.622415	357.074129	248.856589	15.784085	11.000443	13967	4.783642
ORANGE COUNTY	22.213153	312.026874	244.613898	14.046942	11.012120	1513	3.034823
PHOENIX AZ	21.747384	280.321458	239.121928	12.889893	10.995434	864	1.894459
PITTSBURGH PA	22.581613	273.879355	248.794695	12.128423	11.017579	682	1.110844
SACRAMENTO CA	22.604303	290.110810	248.638537	12.834318	10.999611	1334	1.834707
SAN DIEGO CA	22.554710	307.938193	248.165914	13.652944	11.002842	10672	2.650102
SEATTLE WA	22.694133	291.223898	249.942297	12.832564	11.013521	2732	1.819043
SILICON VALLEY	22.791838	340.377008	250.766650	14.934162	11.002476	3797	3.931686
TUCSON AZ	22.556909	323.213554	249.008830	14.328805	11.039138	799	3.289667

City wise Ride Analysis: Average Metrics and Earnings Breakdown

2. Yellow Cab Data

Cities	Avg KM	Avg Price	Avg Cost	Per KM Price	Per KM Cost	Total Num of Rides	Per KM Avg Net Profit
ATLANTA GA	22.248247	422.413446	293.615723	18.986370	13.197252	5795	5.789118
AUSTIN TX	22.512840	422.248940	296.999358	18.755916	13.192443	3028	5.563473
BOSTON MA	22.543649	359.140431	297.656812	15.930891	13.203577	24506	2.727314
CHICAGO IL	22.659493	364.021927	299.097441	16.064875	13.199653	47264	2.865222
DALLAS TX	22.599395	493.807790	297.958374	21.850487	13.184352	5637	8.666135
DENVER CO	22.444097	443.235413	296.382238	19.748418	13.205354	2431	6.543065
LOS ANGELES CA	22.688117	416.231136	299.574768	18.345777	13.204038	28168	5.141739
MIAMI FL	22.635606	432.944872	299.072443	19.126718	13.212478	4452	5.914241

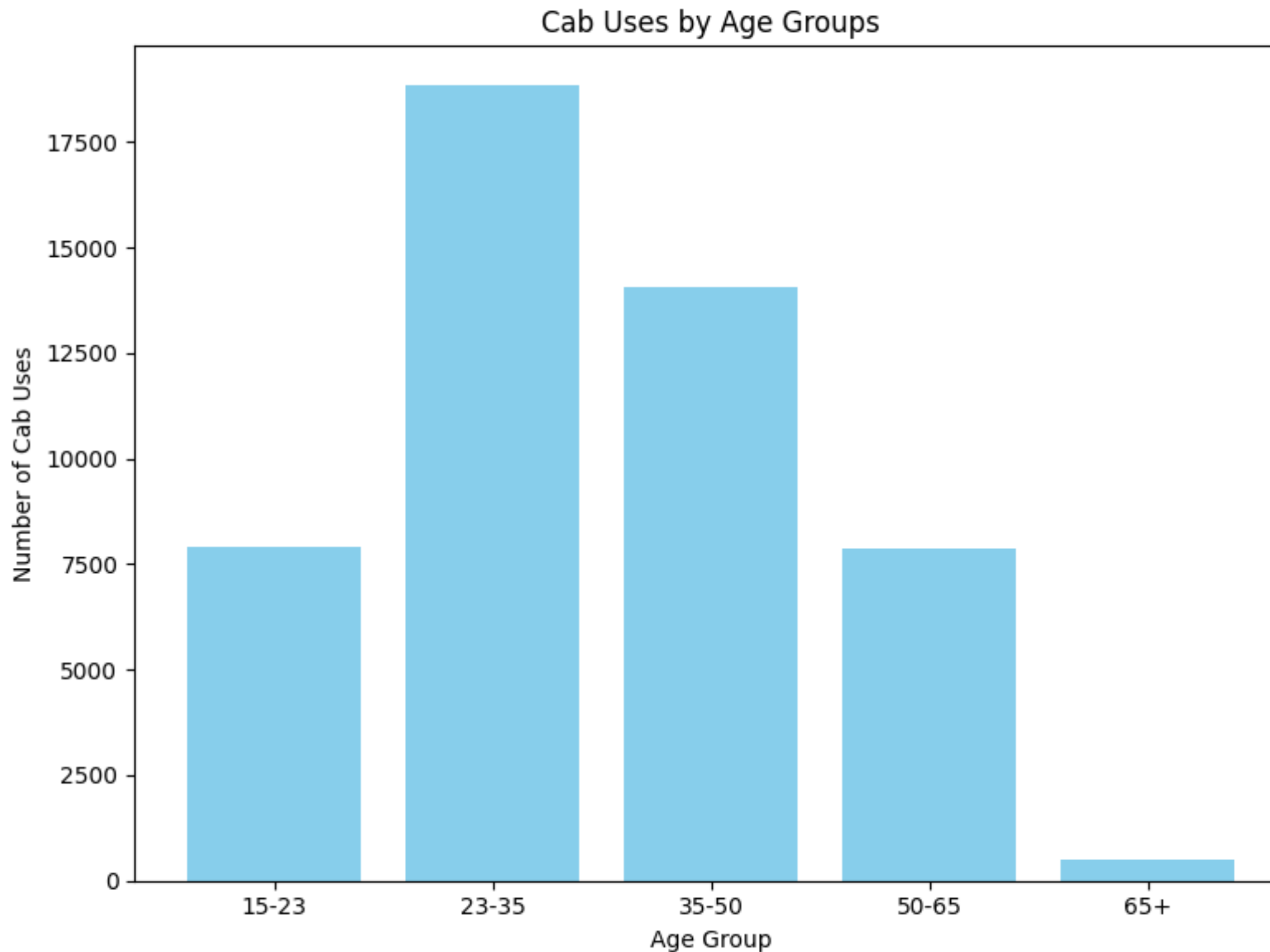| Cities	Avg KM	Avg Price	Avg Cost	Per KM Price	Per KM Cost	Total Num of Rides	Per KM Avg Net Profit
NEW YORK NY	22.496362	604.841911	296.977659	26.886209	13.201141	85918	13.685068
ORANGE COUNTY	22.364597	438.548623	294.763268	19.609055	13.179905	2469	6.429150
PHOENIX AZ	22.529450	428.737417	297.617010	19.030088	13.210132	1200	5.819956
PITTSBURGH PA	22.670523	405.313661	297.455945	17.878443	13.120824	631	4.757619
SACRAMENTO CA	22.949719	362.637106	302.615634	15.801374	13.186028	1033	2.615347
SAN DIEGO CA	22.391995	392.223328	295.516553	17.516230	13.197420	9816	4.318810
SEATTLE WA	22.583783	391.014733	297.585765	17.313960	13.176967	5265	4.136994
SILICON VALLEY	22.696391	506.499502	299.711120	22.316301	13.205232	4722	9.111069
TUCSON AZ	22.267527	366.018224	294.488966	16.437309	13.225042	1132	3.212268

- Based on the provided data for Pink Cab and Yellow Cab across various cities, several factors can be considered for making a decision:

  1. **Average Price and Cost:** Pink Cab generally has lower average prices and costs compared to Yellow Cab across most cities.

  2. **Per KM Price and Cost:** Yellow Cab tends to have higher per kilometer prices and costs compared to Pink Cab.

  3. **Total Number of Rides:** Yellow Cab serves a larger number of rides in many cities compared to Pink Cab.

  4. **Per KM Net Profit:** Pink Cab often has a higher per kilometer net profit compared to Yellow Cab.

- Considering these factors, the decision might vary based on the specific goals and priorities of XYZ firm. If XYZ values higher profitability per ride, Pink Cab could be a suitable option. However, if XYZ prioritizes higher market share and total revenue, Yellow Cab might be preferred due to its larger customer base and total number of rides served.
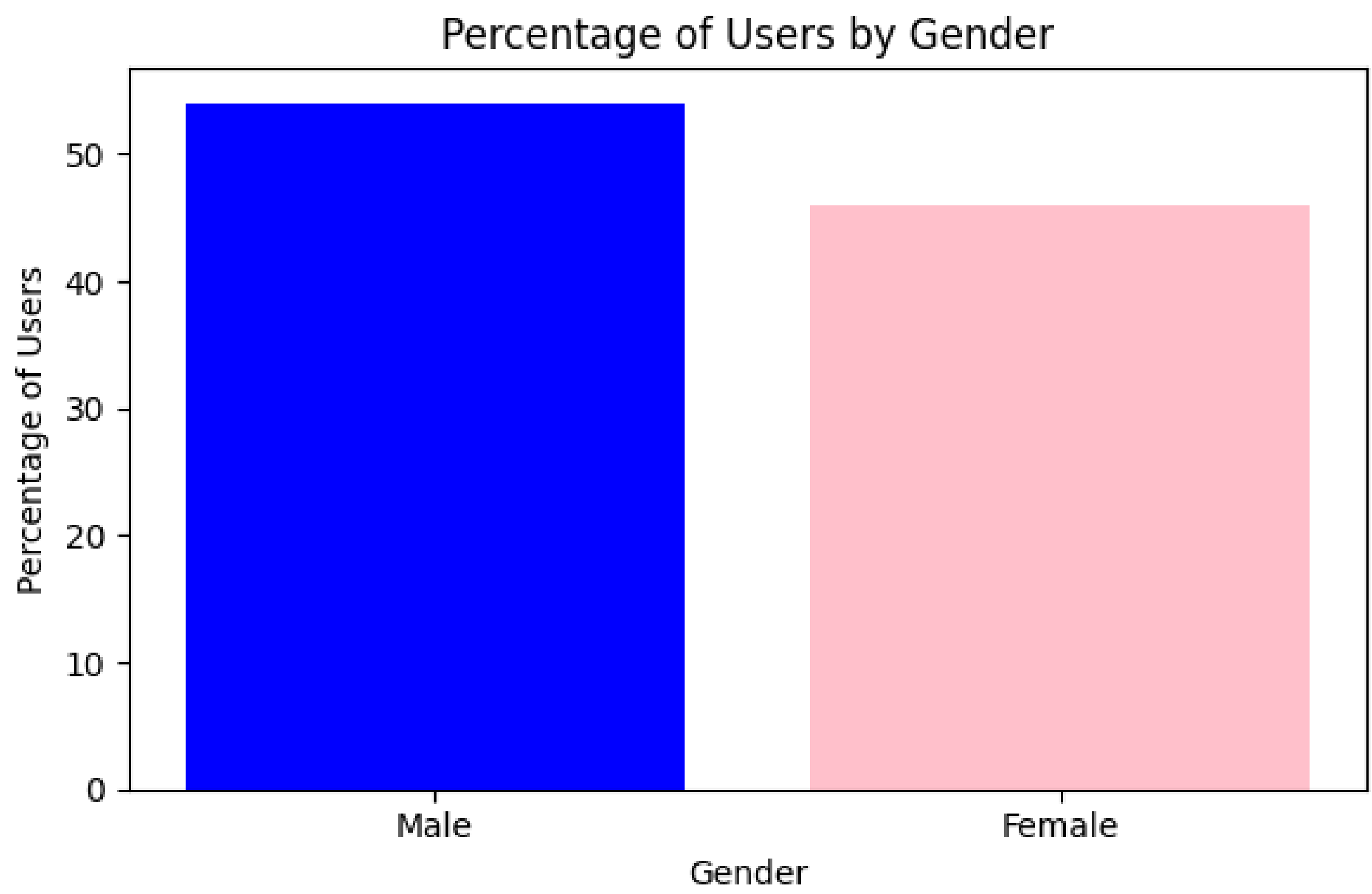
City wise Total Cab User Analysis

Total Cab User Percentage in Each City

- Based on user percentage, investing in cab services in cities like **New York**, **San Francisco**, and **Washington DC** offers the most promising market potential and customer base for XYZ firm.

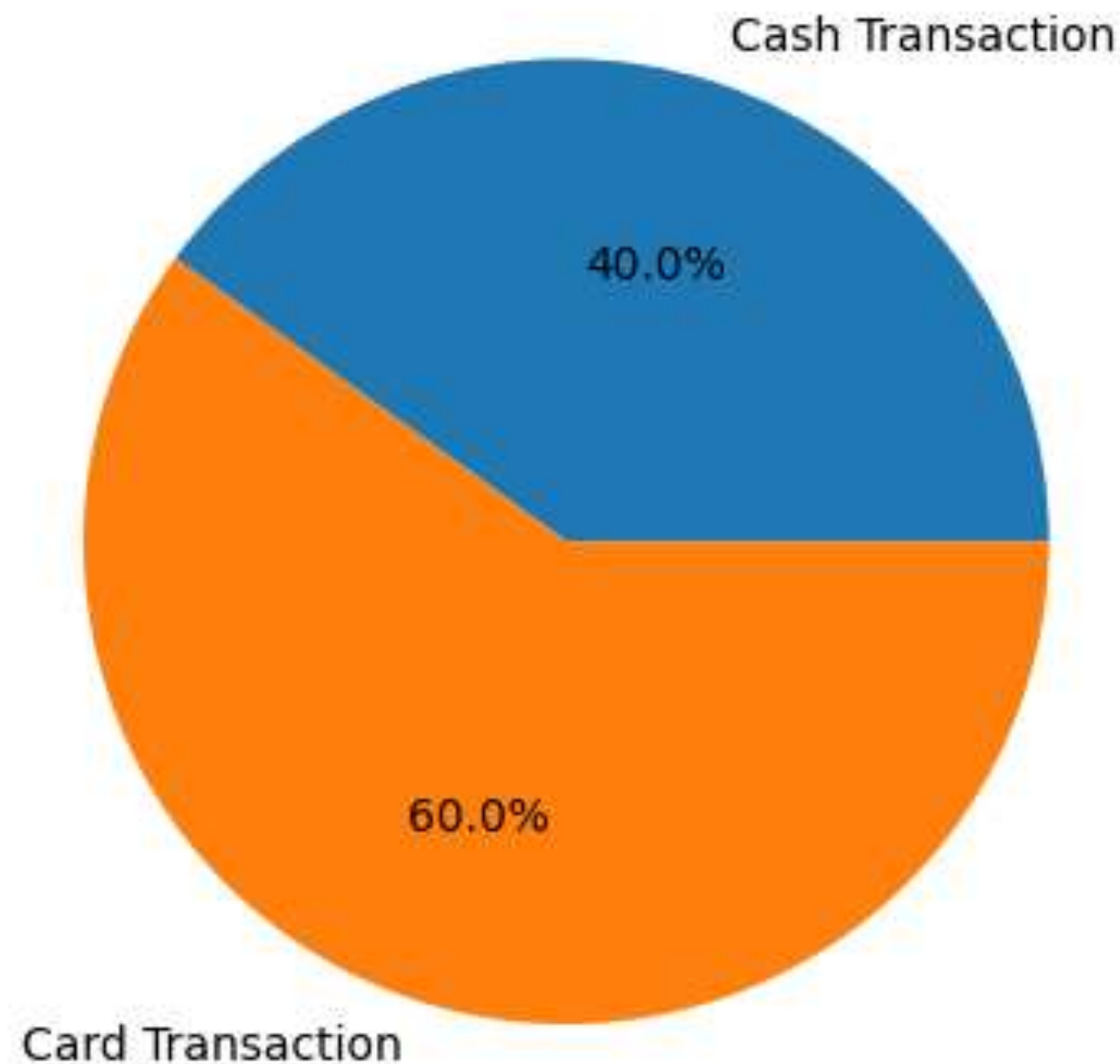# Cab Usage by Age Groups

Cab Uses by Age Groups

- The age group of 23-35 accounts for the highest cab usage with **18,837** rides.

- Followed by the 35-50 age group with **14,084** rides.

- Age groups 15-23 and 50-65 also contribute significantly with **7,903** and **7,855** rides respectively.

- The 65+ age group shows the lowest cab usage with 492 rides.

- Tailoring marketing strategies to target the age groups with the highest cab usage could optimize XYZ's investment impact.

Gender-Based Cab User Percentage Analysis

Percentage of Users by Gender

Transaction Mode Analysis

- **60%** of cab transactions are via cards, signaling a preference for electronic payments.

- Cash transactions make up **40%** of total transactions, highlighting the need to cater to both payment methods.

- XYZ can invest in fintech for improved digital payment experiences and target marketing to boost card usage.

- Understanding cash transaction patterns can guide service enhancements.

Recommendations

1. Total Profitability: Yellow Cab's total profit is **$44,020,373.17** and Pink Cab's total profit is **$5,307,328.32**. This reflects superior cost management and market dominance, making Yellow Cab the preferred investment with a potential **8x** higher return.

2. City-wise Profit: In key cities like New York and Chicago, Yellow Cab's profits exceed Pink Cab's by over **$25 million** and **$2.75 million** respectively. Such substantial differences highlight Yellow Cab's stronger revenue potential and market share.

3. Cost-Benefit Analysis: Despite higher per kilometer prices and costs, Yellow Cab's larger ride volume ensures competitive profitability. For instance, while Pink Cab may have lower average prices and costs, Yellow Cab's higher profit per ride translates to better returns.

4. Market Potential: Cities like **New York**, **San Francisco**, and **Washington DC** boast high user percentages, indicating strong market demand. Investing in these cities offers significant growth opportunities and access to a large customer base.

5. Targeted Marketing: Focusing on age groups with the highest cab usage, such as the **23-35** and **35-50** demographics, can optimize ROI. For instance, tailoring marketing strategies to these age groups, which contribute over **18,000** and **14,000** rides respectively, can maximize engagement and conversion rates.

Investing in **Yellow Cab** promises higher profitability and market dominance, particularly in cities like New York and Chicago. Its strong revenue potential, despite higher costs, positions it as the optimal choice for XYZ's investment.

Thank You