Data Glacier
Your Deep Learning Partner

Exploratory Data Analysis

G2m Insight For Cab Investment

05/19/2024

Agenda

Problem Statement

Approach

EDA: Based on Analysis Aspects.

EDA Summary

Recommendations

Problem Statement.

Title:

Investment Analysis in the U.S. Cab Industry

Background:

XYZ, a private firm in the U.S., is considering an investment in the cab industry due to its remarkable growth in recent years. With multiple key players in the market, XYZ aims to make an informed decision by understanding the market dynamics, customer demographics, and financial performance of potential cab companies.

Objective:

The primary objective is to analyze and compare two cab companies based on various datasets provided. The goal is to generate actionable insights that will assist XYZ in identifying the more promising investment opportunity.

Scope:

The analysis will cover the period from January 31, 2021, to December 31, 2023. The datasets include transaction details, customer demographics, transaction-to-customer mapping, and city-specific data on population and cab usage.

Problem Statement.

Datasets Provided:

Cab_Data.csv: Contains transaction details for two cab companies.

Customer_ID.csv: Maps customer demographic details using unique identifiers.

Transaction_ID.csv: Maps transactions to customers and includes payment modes.

City.csv: Lists U.S. cities with their population and number of cab users.

Hypotheses to Investigate:

1. Is there any seasonality in the number of customers using the cab service?

2. Which company has the maximum cab users at a particular time period?

3. Does the margin proportionally increase with the number of customers?

4. What are the key attributes of the customer segments?

5. How do demographic factors influence cab usage?

Problem Statement.

Outcome:

The analysis will help XYZ make a data-driven decision on which cab company to invest in by providing a comprehensive understanding of the market and identifying key factors influencing the performance of the two companies.

Approach

Client:

XYZ is a US-based private company. As a result of the last few years' exceptional expansion in the taxi business and the presence of several major companies, the industry is planning investments, and in line with their go-to-market (G2M) strategy, they want to fully understand the market before making a choice.

Objective:

The company is interested in using my approaches to help them identify the best company to invest in.

Analysis Aspects:

Profit Margins, Location, Age Group, Payment Method(Credit vs Cash), Customer Retainment.

Data Exploration:

There are 359,352 total rows in the dataset. There are 20 total features that was given and an additional 4 that were derived from the dataset. The time period is between 2020 and 2021.

Assumptions:

1. Both companies' data analyses were conducted with the assumption that there was external noise in addition to the supplied data.

2. Data analysis was done for both businesses with the understanding that the datasets were

Approach

3. limited time frame from 2021 to 2023.

4. Fifth, the datasets were chosen at random.

5. The only payment options that are taken into consideration are cash and cards.

6. Price Charged: The only variables used to determine profit were the trip costs.

1. Profit Analysis.

Motivation:

Calculating the profit margins of the two companies is the first step in differentiating their values. A corporation with a higher profit margin will be worth more money.

Tests:

The first test looked at the combined profit margin for the two businesses over the course of the four quarters in 2020–2023.

The purpose of the second test was to display the profit margin frequency for a single transaction.
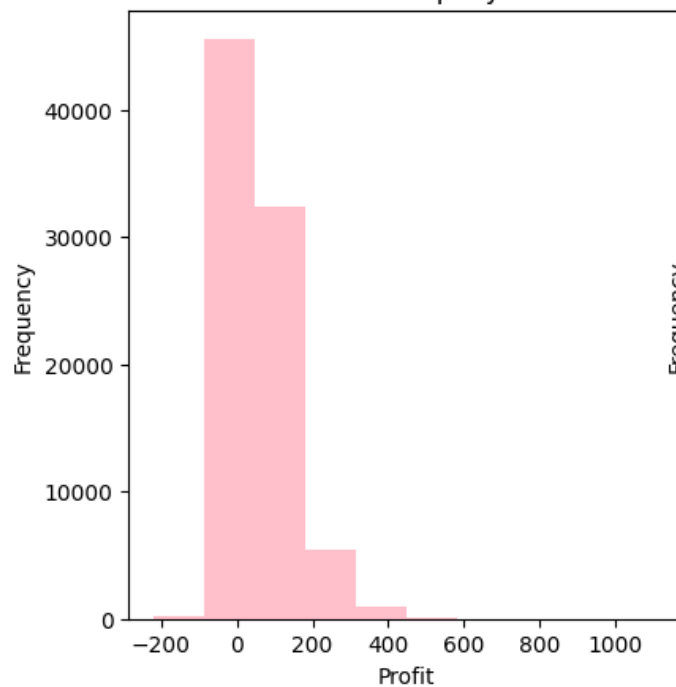
Conclusion:

Yellow Cab has higher profit margins than Pink Cab.
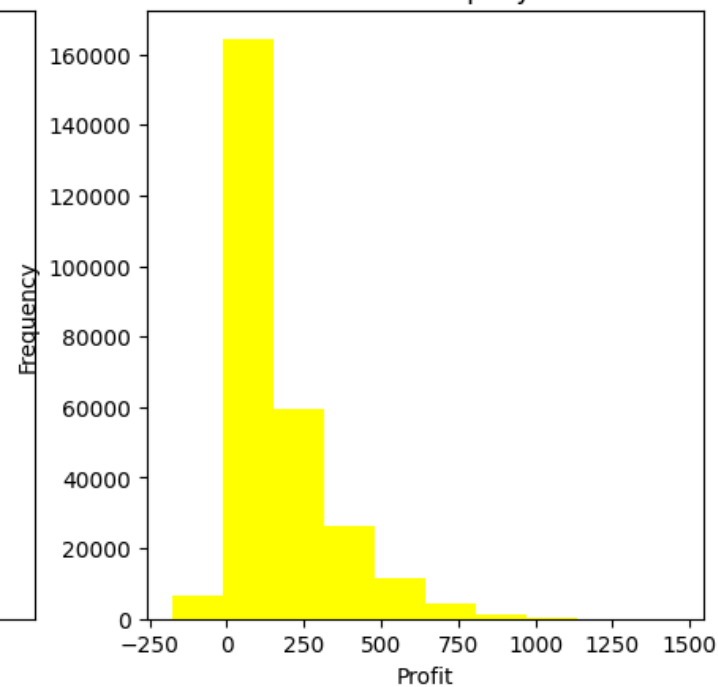
Profit Analysis Results.
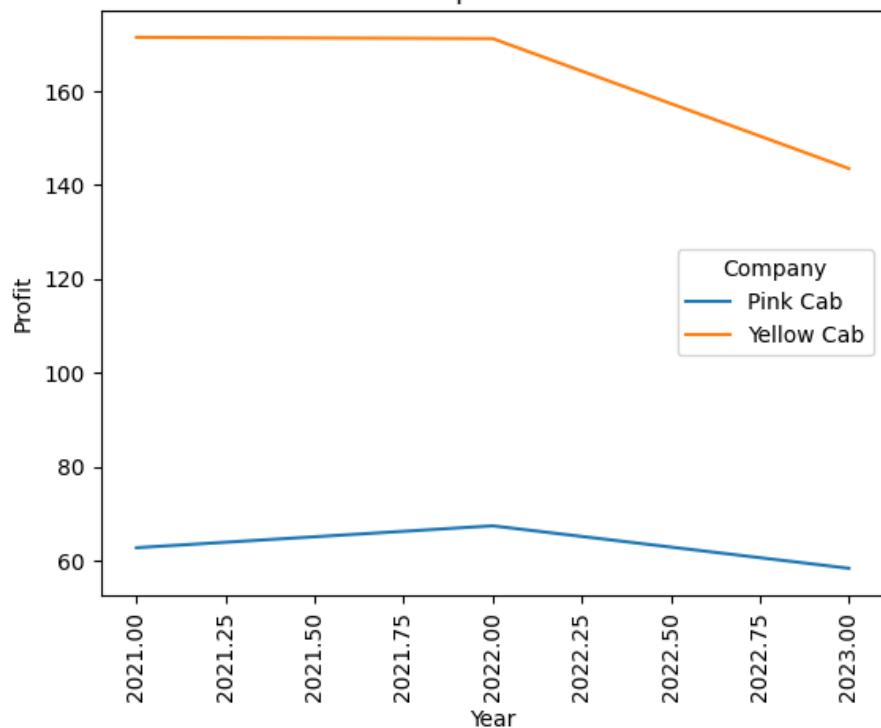
2. Location Analysis.

Motivation:

My objective was to ascertain the allocation of earnings among the top three cities that generated the highest profits for both companies. It may be safer and less probable for them to see a decline in revenues during local economic downturns if there is a more equitable distribution of profits from various places.

Tests:

Creating a histogram of the total number of transactions reported for each city was the first test. The second exam involved creating a pie chart showing how each company's top three cities' profits were distributed.

Conclusion:

Based on locations, profit distribution is higher in New York City for both companies, but Yellow Cab has a higher profit margin in New York City compared to Pink Cab. Pink Cab has a higher profit margin in Los Angeles, CA compared to Yellow Cab.
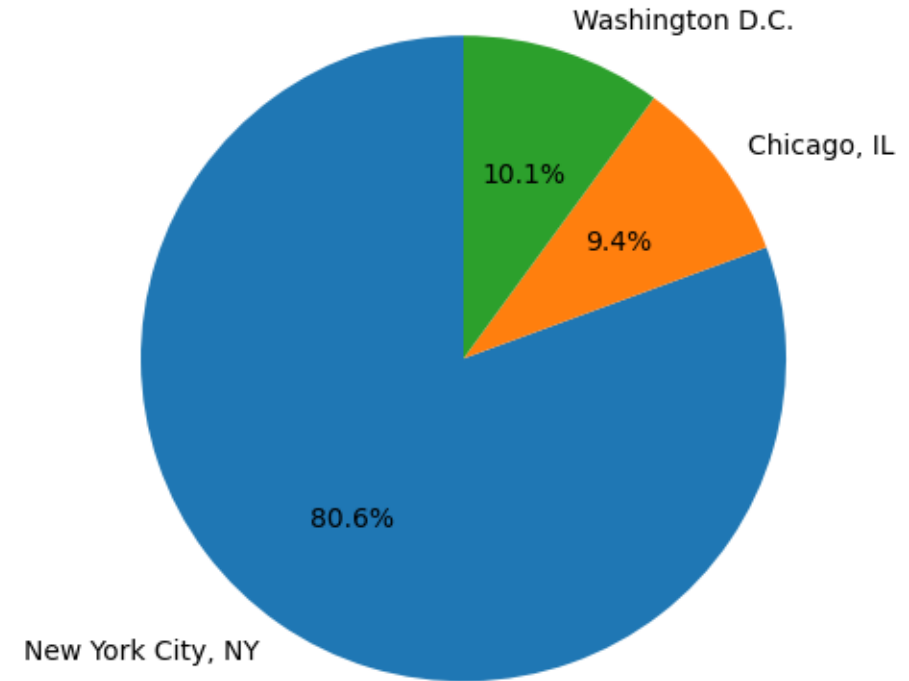
Location Analysis Results.
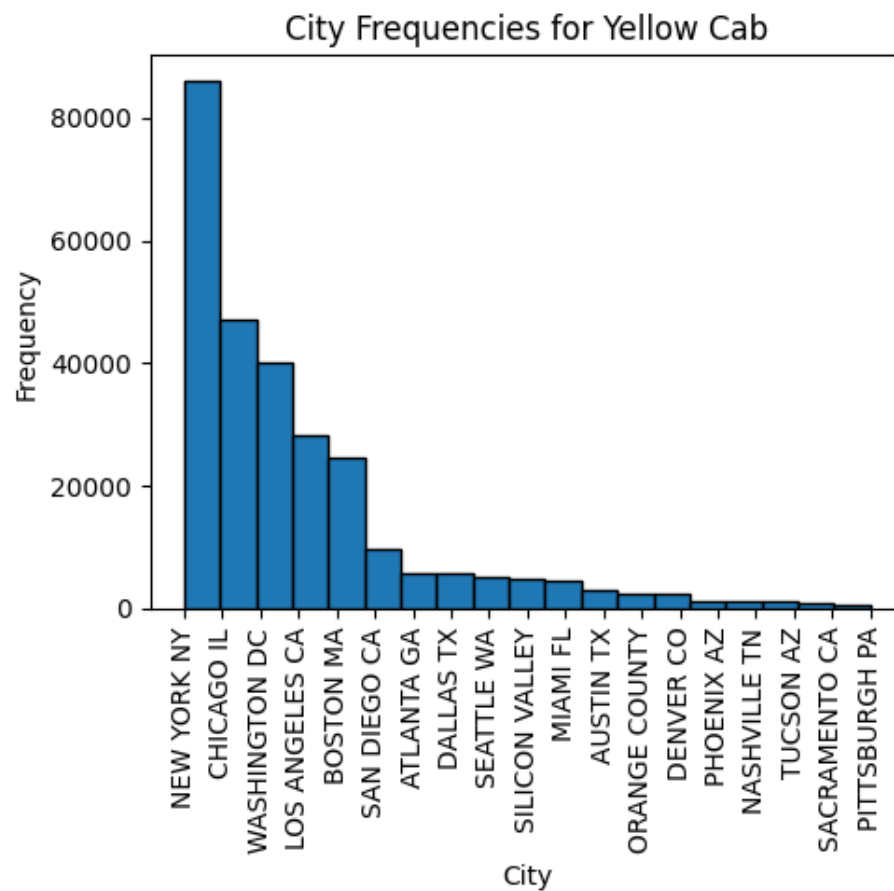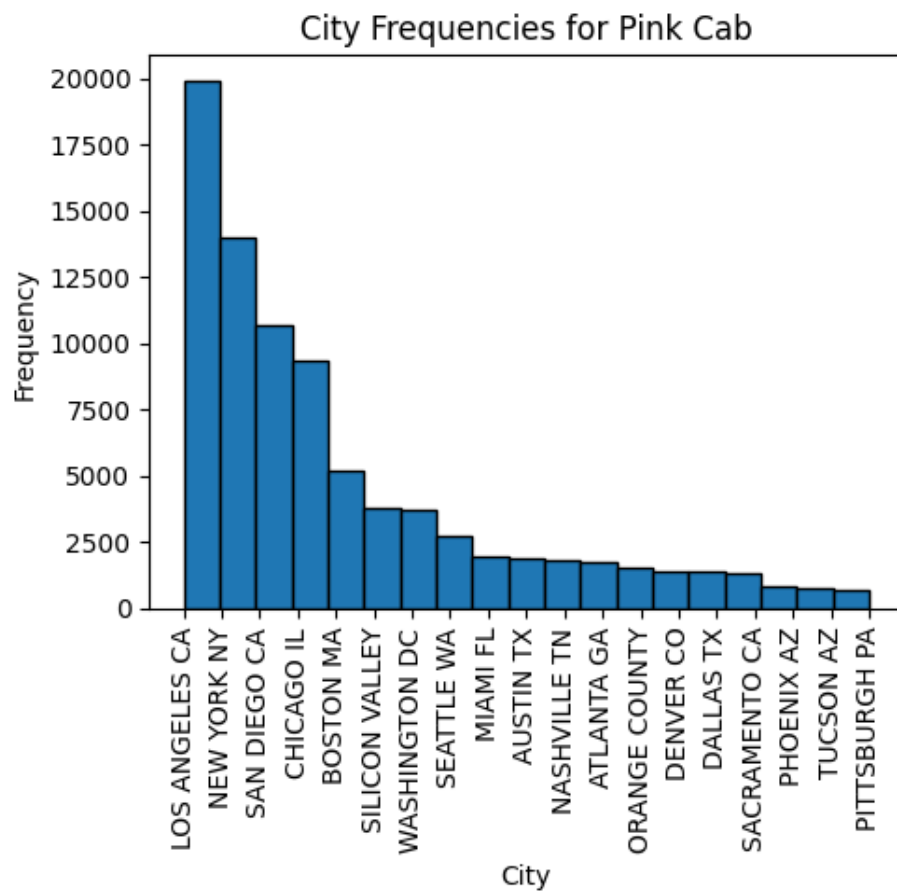
Location Analysis Results.

Distribution of Profits per City for Pink Cab

San Diego, CA — 19.5%
Los Angeles, CA — 34.4%
New York City, NY — 46.2%
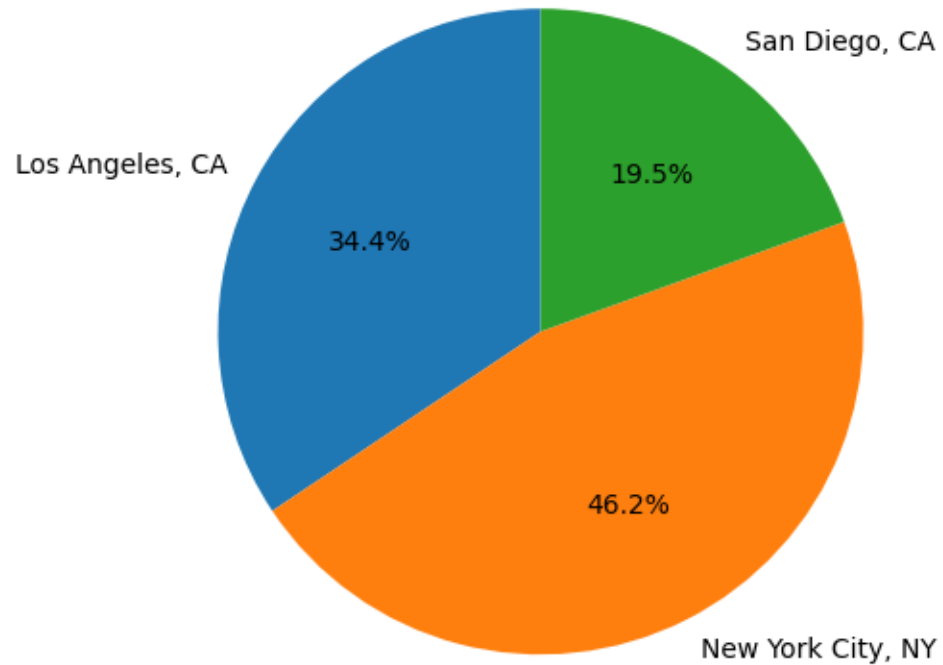
Distribution of Profits per City for Yellow Cab

Washington D.C. — 10.1%
Chicago, IL — 9.4%
New York City, NY — 80.6%

3. Age Group Analysis

Motivation:

I was interested in examining the distribution of age groups across the two businesses according to their clientele. Since younger customers are more likely to stick with a service for a longer amount of time, I was interested in finding out whether company had a higher proportion of younger clients in its distribution

Tests:

The first test was making a pie chart of four different age groups: young adults, middle adults, older adults, and elderly. The second test was making a pie chart of the total profit distribution in terms of these four age groups
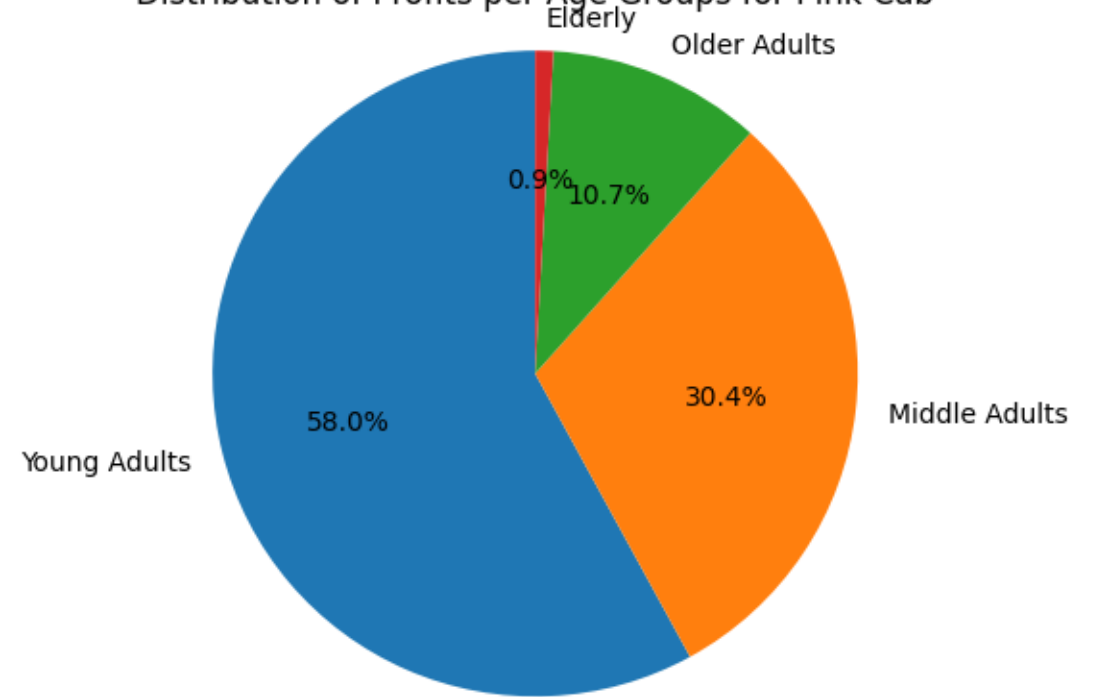
Conclusion:

Both companies have the same age group distribution, but Yellow Cab has a higher profit margin in all age groups compared to Pink Cab.

Age Group Results
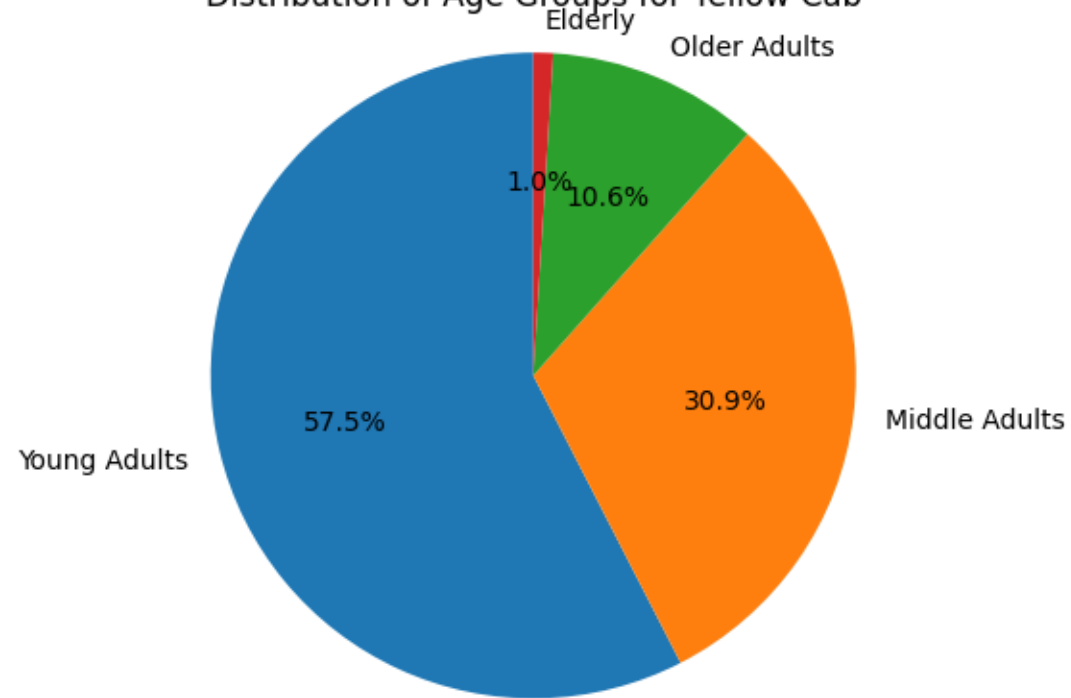

Distribution of Age Groups for Pink Cab
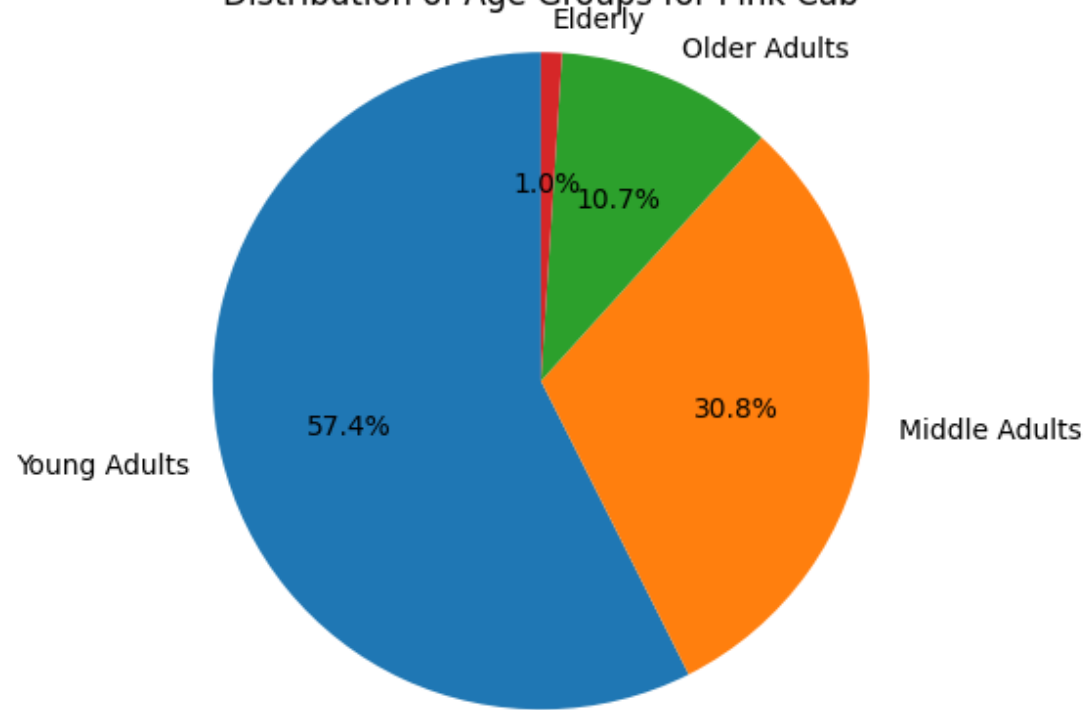
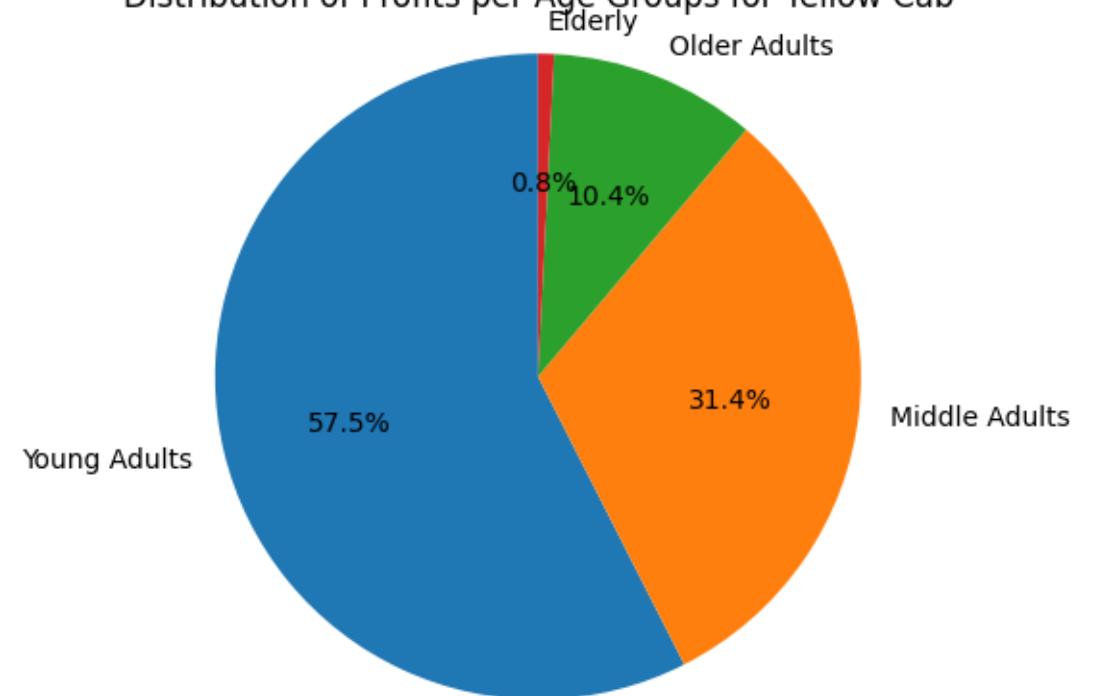Distribution of Profits per Age Groups for Pink Cab

Age Group Results

Distribution of Age Groups for Yellow Cab

Distribution of Profits per Age Groups for Yellow Cab

4. Payment method Analysis

Motivation:

I was interested in determining whether the two companies' cash and card user bases differed. An increase in card transactions would indicate that the business is adjusting to the rapidly evolving technology, and eventually, this payment option will take the lead.

Tests:

Creating a pie chart showing the two companies' cash and card users was the first test. The second test involved creating a line graph showing the two organizations' respective changes in the percentage of card transactions between 2021 and 2023.

Conclusion:

The two companies' results are strikingly comparable. About 40% of customers at both organizations report using cash, while about 60% use cards. This indicates that there aren't any notable distinctions between the two businesses. The results are also indicated by the line graph, which shows that there will be very little growth in card transactions for either company between 2021 and 2023. Therefore, it is not possible to conclude which of the two companies fared better based on this analysis's lack of differentiation.

5. Customer Retainment Analysis

Motivation:

How frequently the same clients wanted to keep using each company's taxi services was what I was trying to find out. The percentage of clients who used the taxi service in 2021 and 2023 were the two primary components that I examined. The other component shows the variation in the average number of taxi service calls made by the same customer between 2021 and 2023. This will demonstrate how each business is performing in terms of keeping customers, which is a key measure of a company's value.

Tests:

The first test was making a pie chart of how many customers were retained between 2021 and 2023. The second test was making a line graph that showed how the average number of cab service calls from 2021 to 2023 changed for the same customer.

Conclusion:

Pink Cab ratio of lost customers to retained customers is higher as compared to yellow cab company, meaning that Yellow Cab provided a cab service that a higher percentage of their customers wanted to continue using their service. Thus, this means that Yellow Cab has better quality of sercvces.

Customer Retainment Results.

Pink Cab % of customers Reatined/Lost (2021-2023)

Retained 17.9%

Lost 82.1%
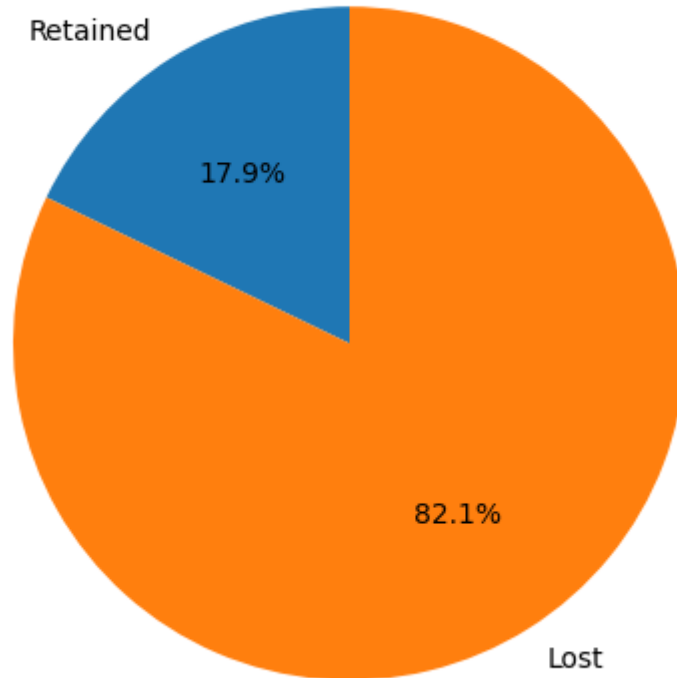
Yellow Cab % of Customers Retained/Lost (2021-2022)

Retained 39.8%

Lost 60.2%
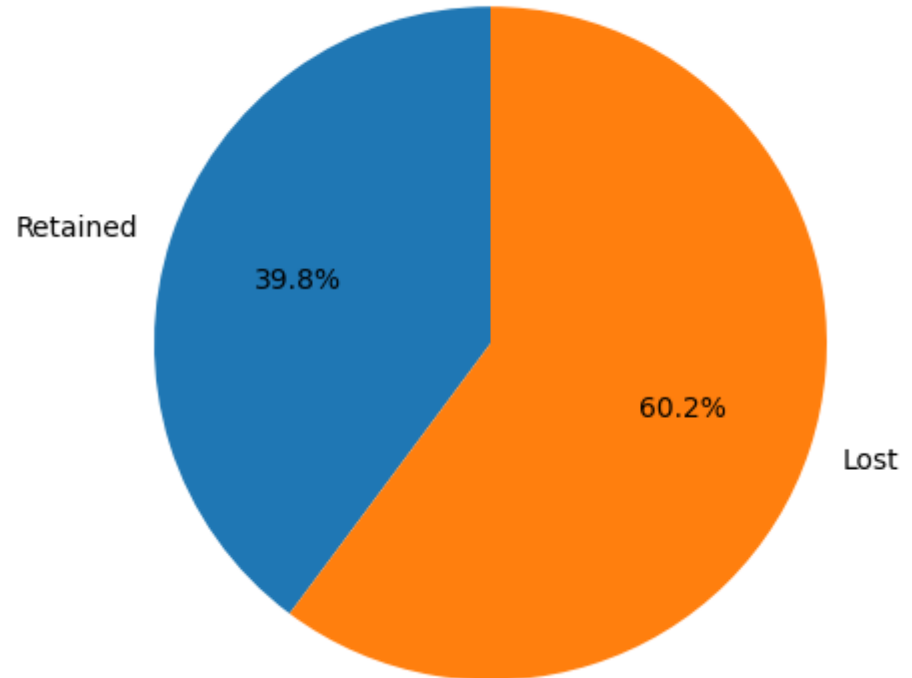
Recommendations

Categories Analyzed:

Profits, Locations, Age Groups. Payment Methods, Customer Retention.

General Overview:

Age groups and payment methods were similar between the two companies. However, locations, profits, and customer retention showed significant differences.

Location Analysis:

Yellow Cab's profits are highly concentrated, with 80% coming from New York City. In contrast, Pink Cab's profits are more evenly distributed across cities. A more even distribution is generally safer, favoring Pink Cab slightly in this category.

Profit Analysis:

Yellow Cab outperformed Pink Cab with higher total profits over three years and higher profit margins per transaction. This category clearly favors Yellow Cab.

Conclusion:

Yellow Cab excels in profits and customer retention, which are more critical for future performance. Therefore, Yellow Cab is the better investment choice.

Final Recommendation:

Invest in Yellow Cab.

Thank You