Taxi Industry Analysis

Exploratory Data Analysis

by Xiyuan Wu

Table of contents

Background Information

VenturePath Investments is a private 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 and as per their Go-to-Market(G2M) strategy they want to understand the market before taking final decision.

> *Note: We will use a company name, "VenturePath Investments," instead of "XYZ" for more readability.*

Approach

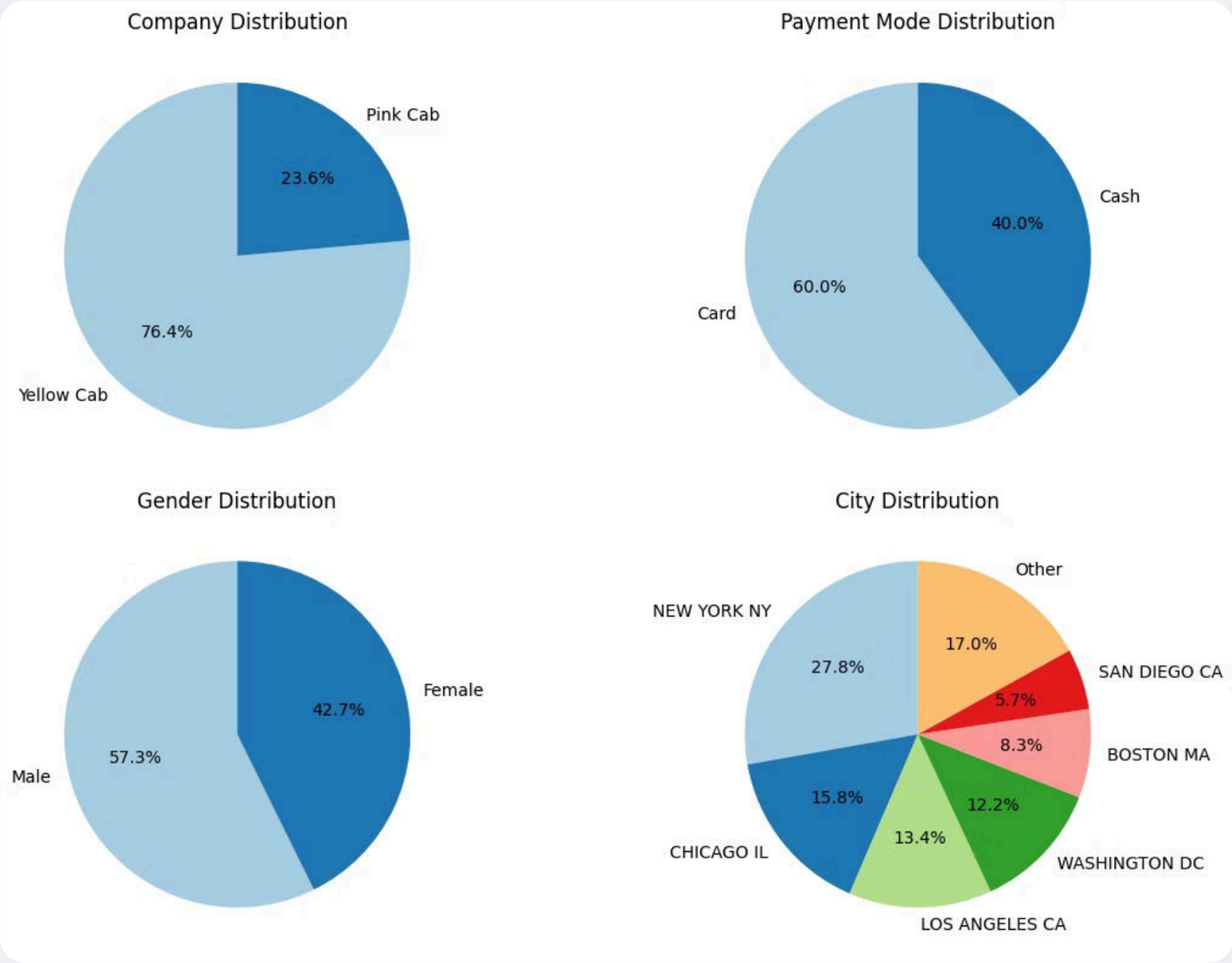
Our strategic approach is to meticulously analyze the provided datasets, encompassing transaction details, customer demographics, transaction-to-customer mappings, and city-specific cab usage statistics. Through this comprehensive analysis, combined with advanced modeling techniques and hypothesis testing, we aim to uncover actionable insights and trends within the cab industry. Hypothesis testing will allow us to validate assumptions about the factors influencing profitability and market dynamics, providing a robust statistical foundation for our analysis. This in-depth investigation will enable us to identify the most viable company for investment, aligning with VenturePath Investments's objectives and market entry strategy. Our findings will be synthesized into a compelling presentation, offering the executive team of VenturePath Investments a clear roadmap for informed decision-making in the competitive landscape of the cab industry.

Analysis Summary

- General Overview
- Profit Overview
- City Overview
- Overall Overview

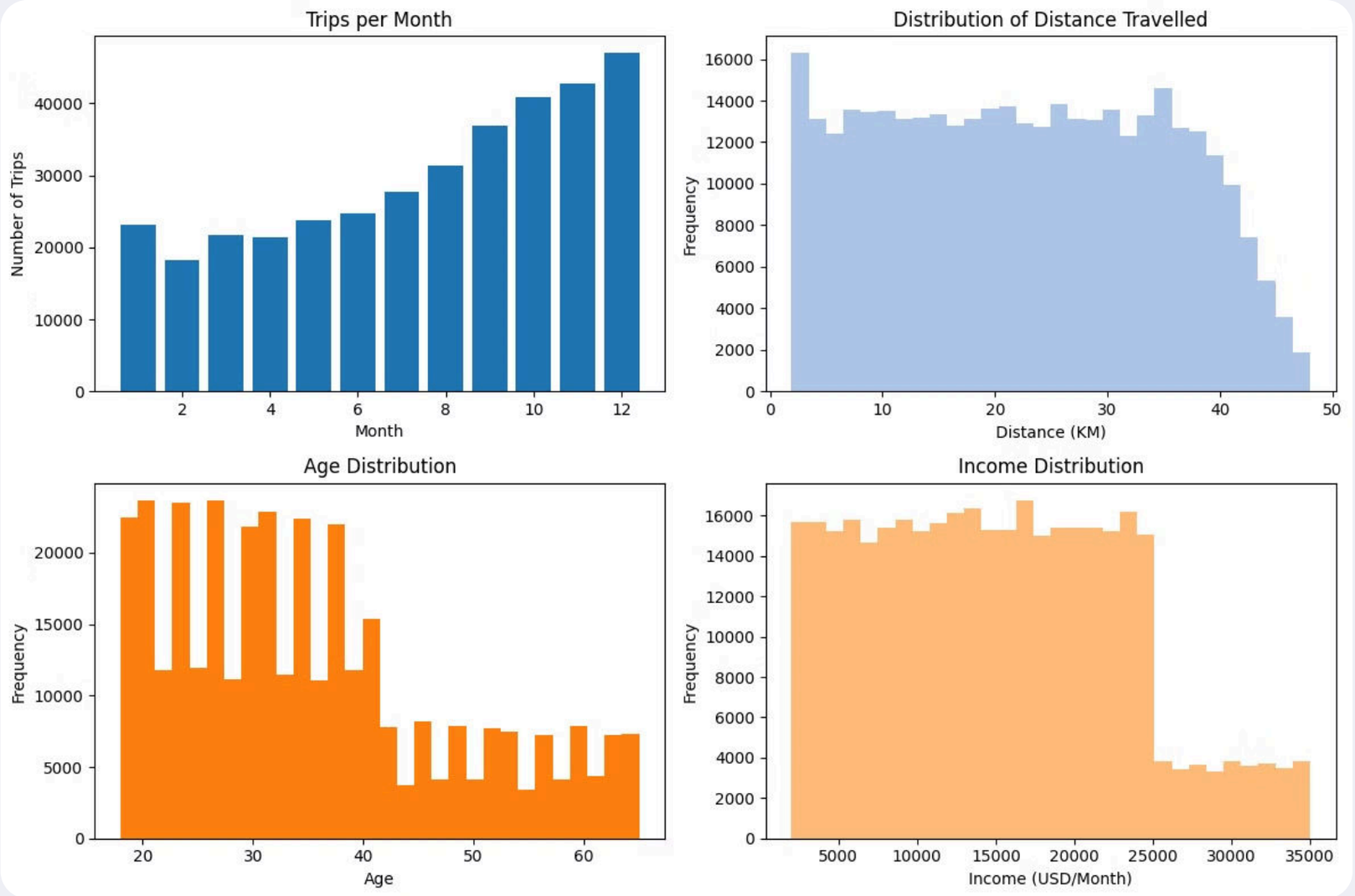
General Overview

Company, Payment Method, Gender and City



Yellow Cab dominates the market share, nearly 80%; although there are more male customers than female, the difference is not significant.

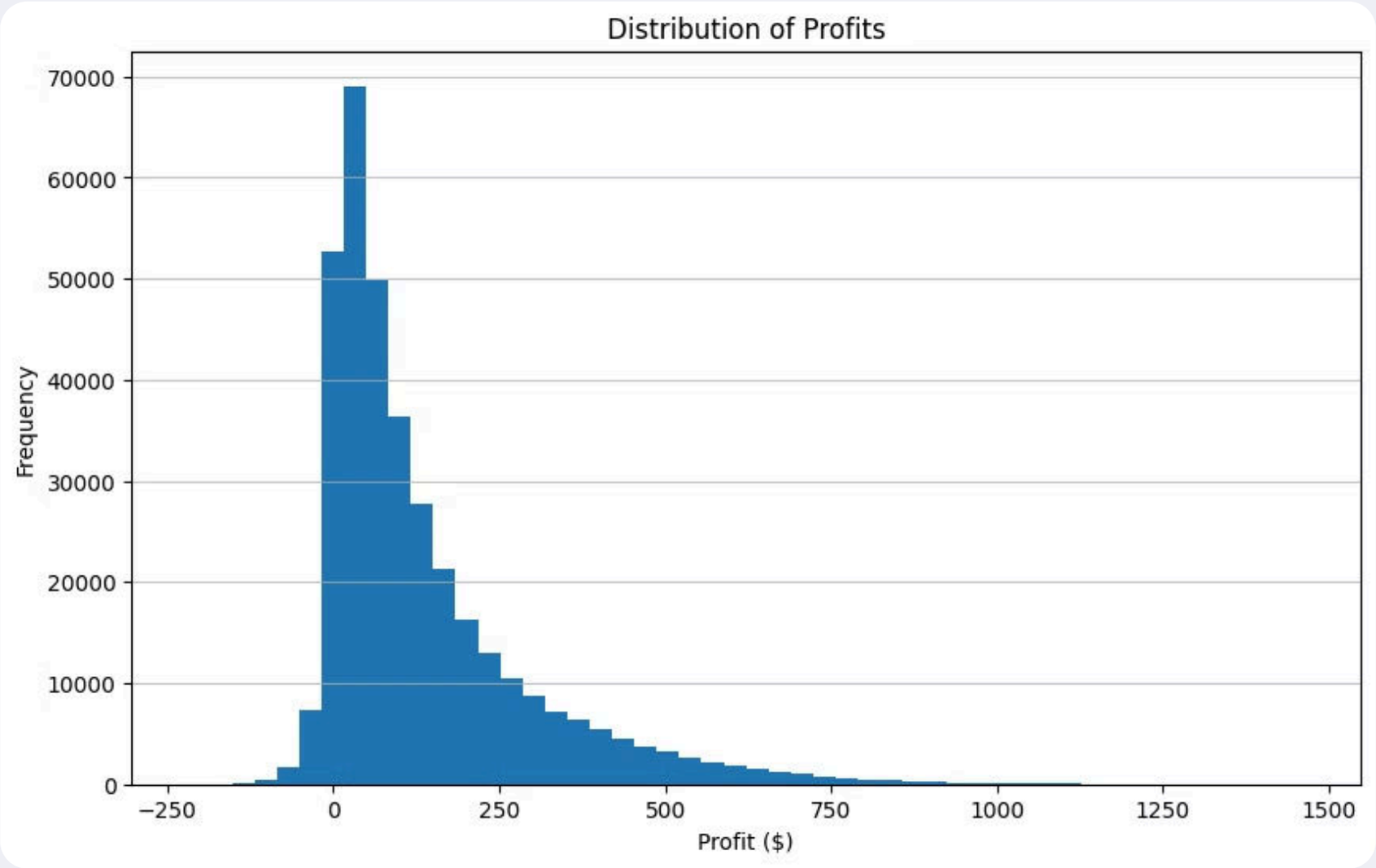
Date of Travel, Distance Travel, Age and Income



- The number of trips per month shows a continuous increase from February to December.
- Regarding distance, the spread is quite even from 2 km to 40 km.
- In terms of income, most people earn an average income between $2,000 and $25,000, with only a few earning above $25,000.

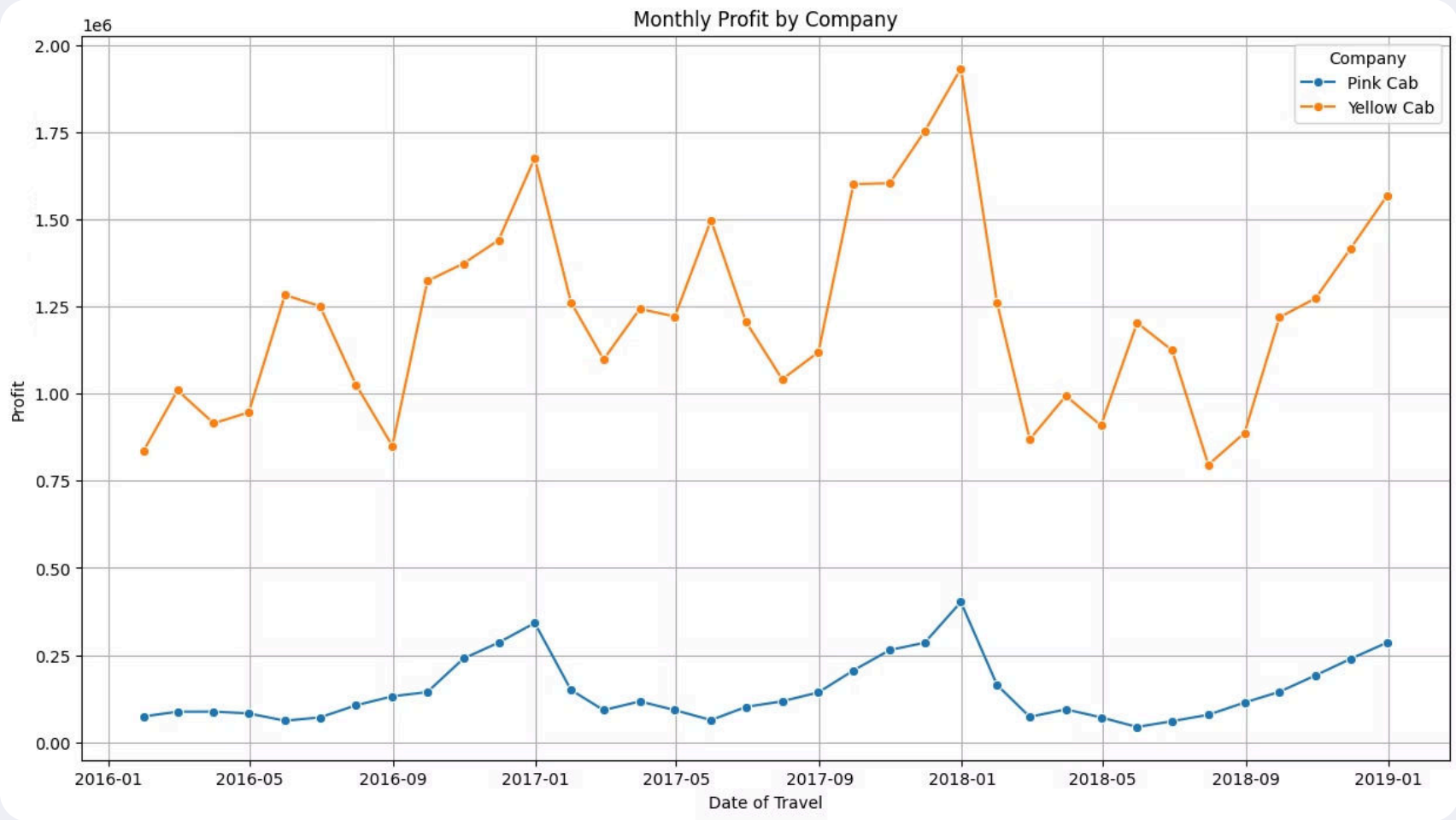
Profit Overview

Distribution of Profits



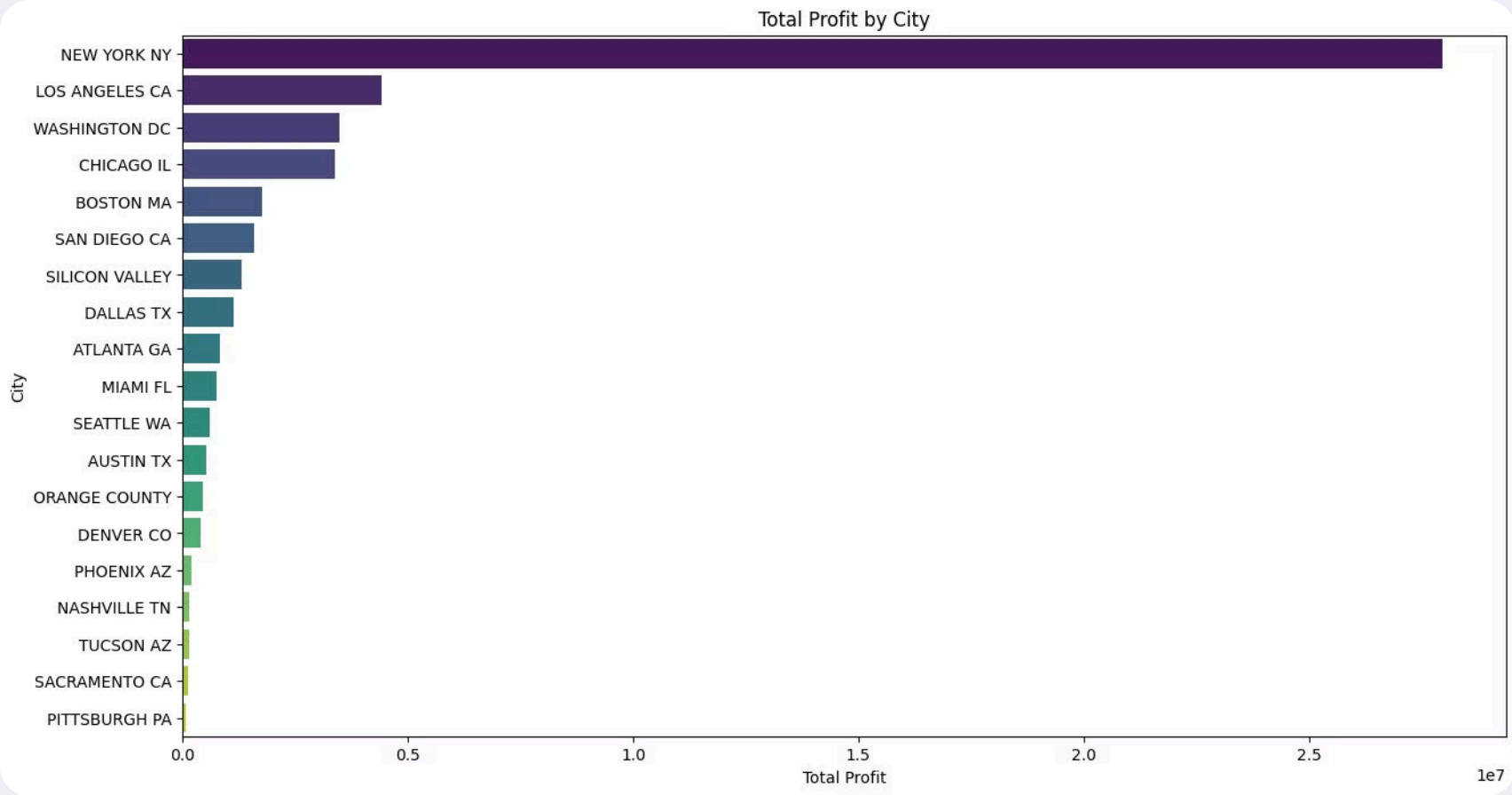
Overall, profits are most widely distributed around $0 to $150.

Monthly Profit by Company



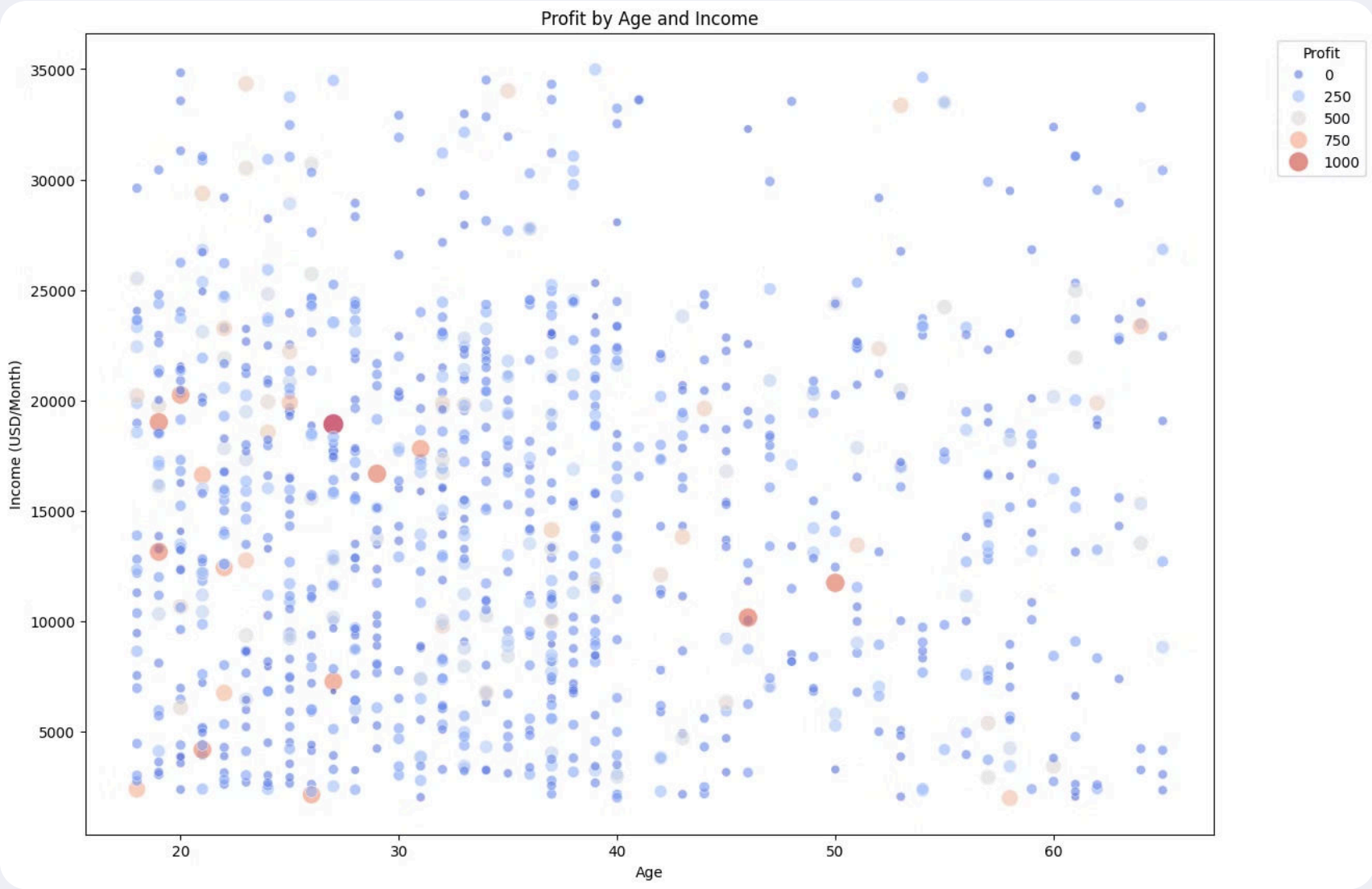
The total profit of Yellow Cab is much more than Pink Cab, which makes sense considering Yellow Cab's market share is about 70%.

Total Profit by City



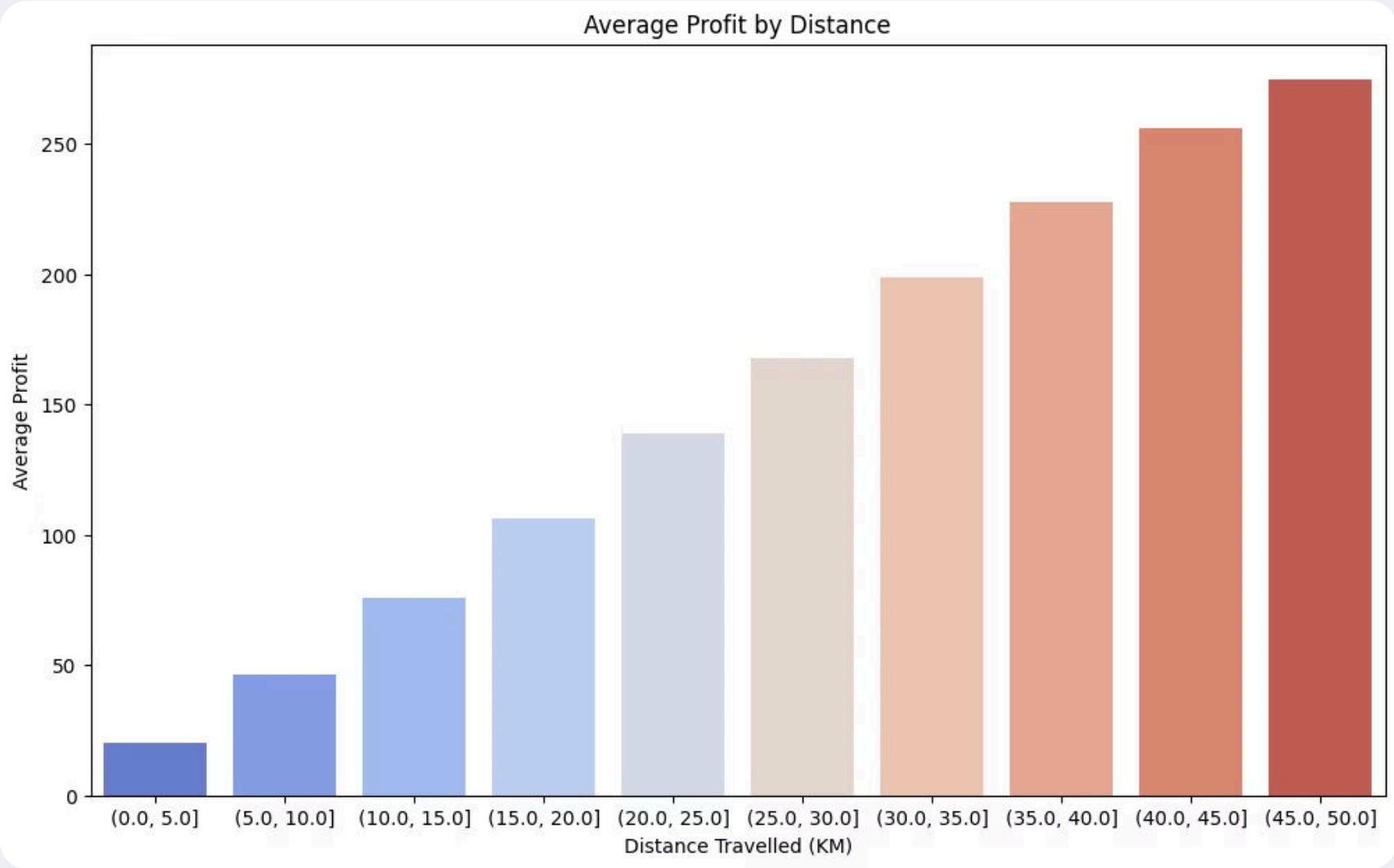
In terms of cities, New York is far ahead, becoming the most profitable city, leading the second place by five times.

Profit and Customer Demographics



There isn’t much correlation found between customer characteristics (age and income) and profit.

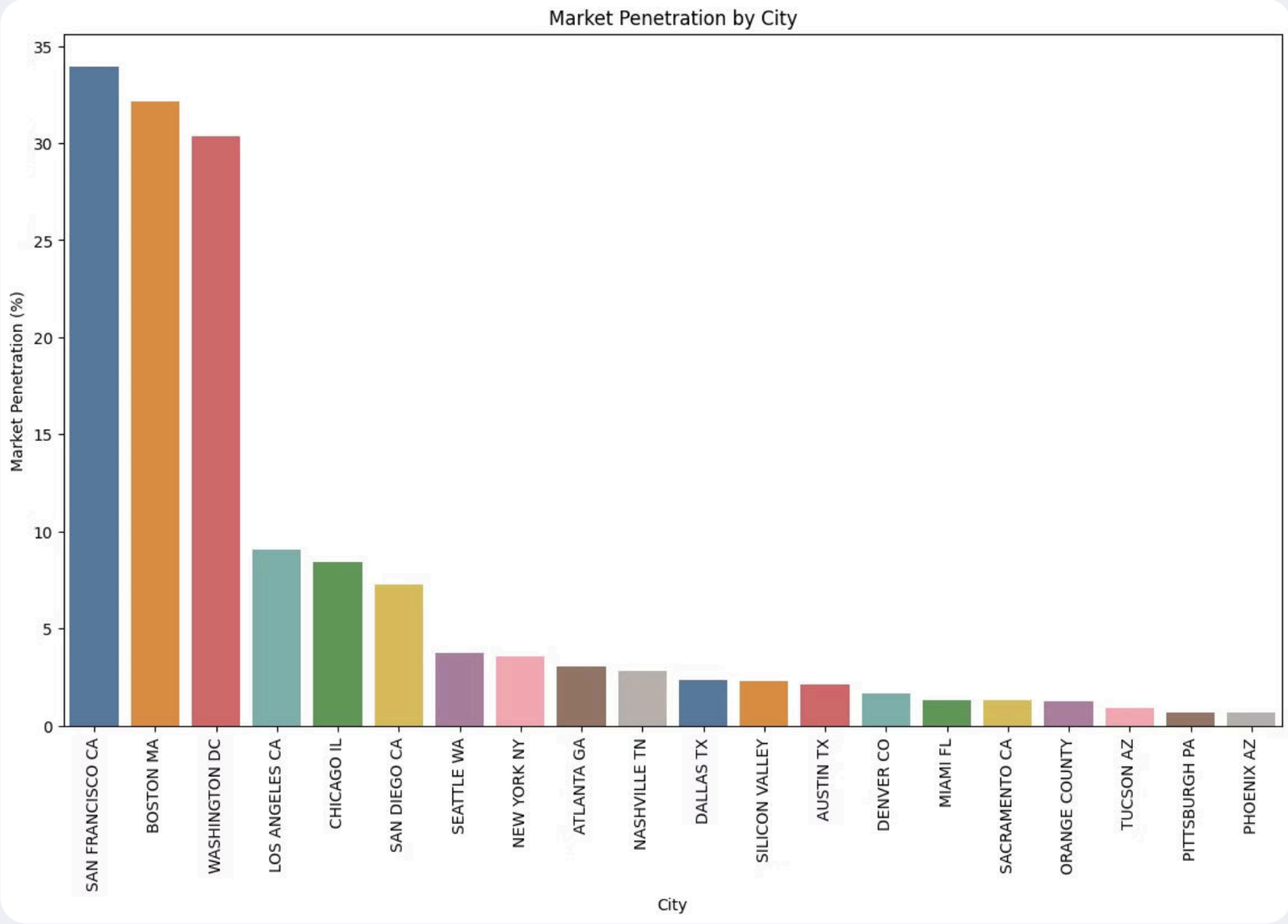
Profit by Distance



The longer the distance, the higher the profit, which is logical.

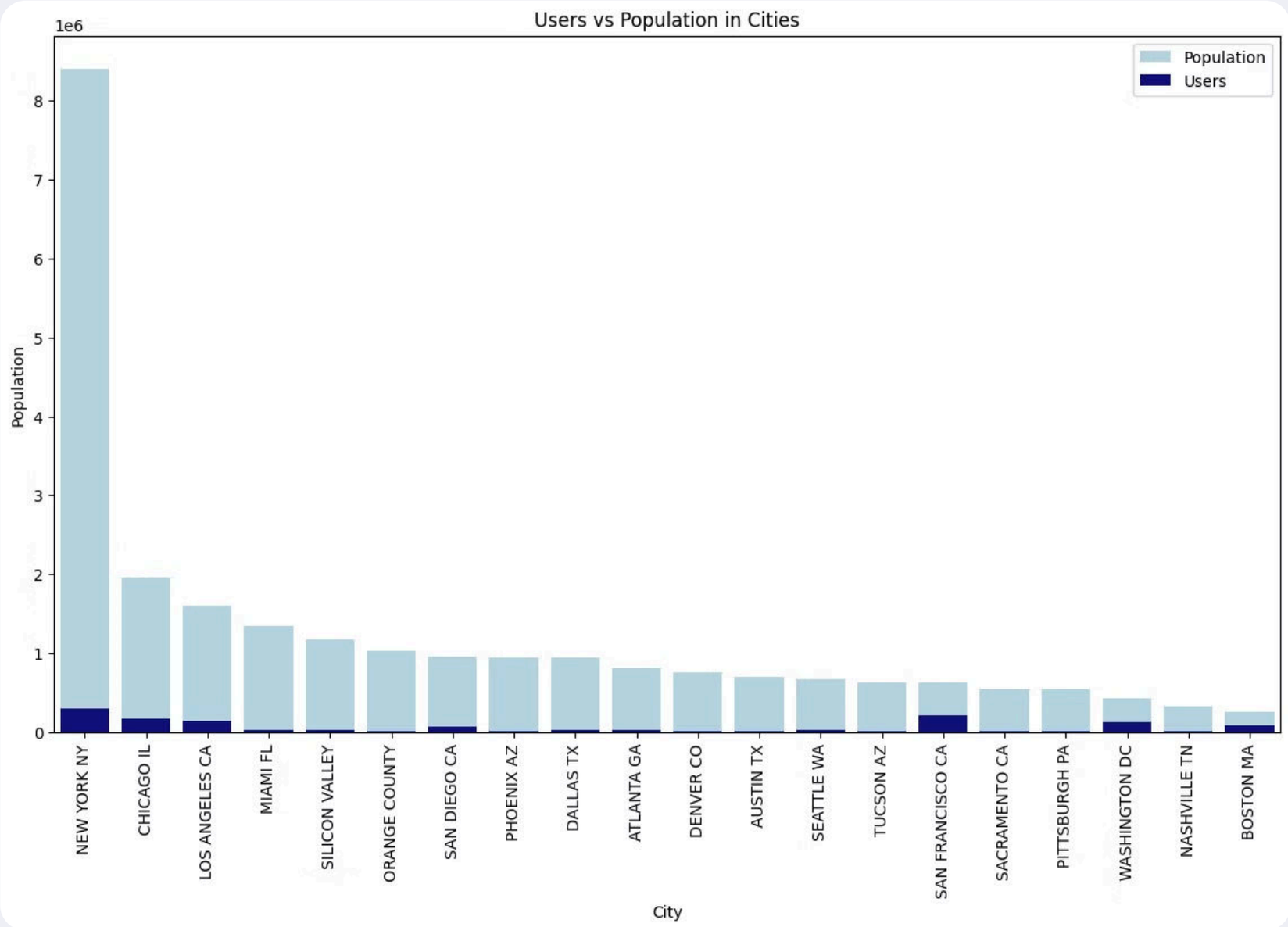
City Overview

Market Penetration



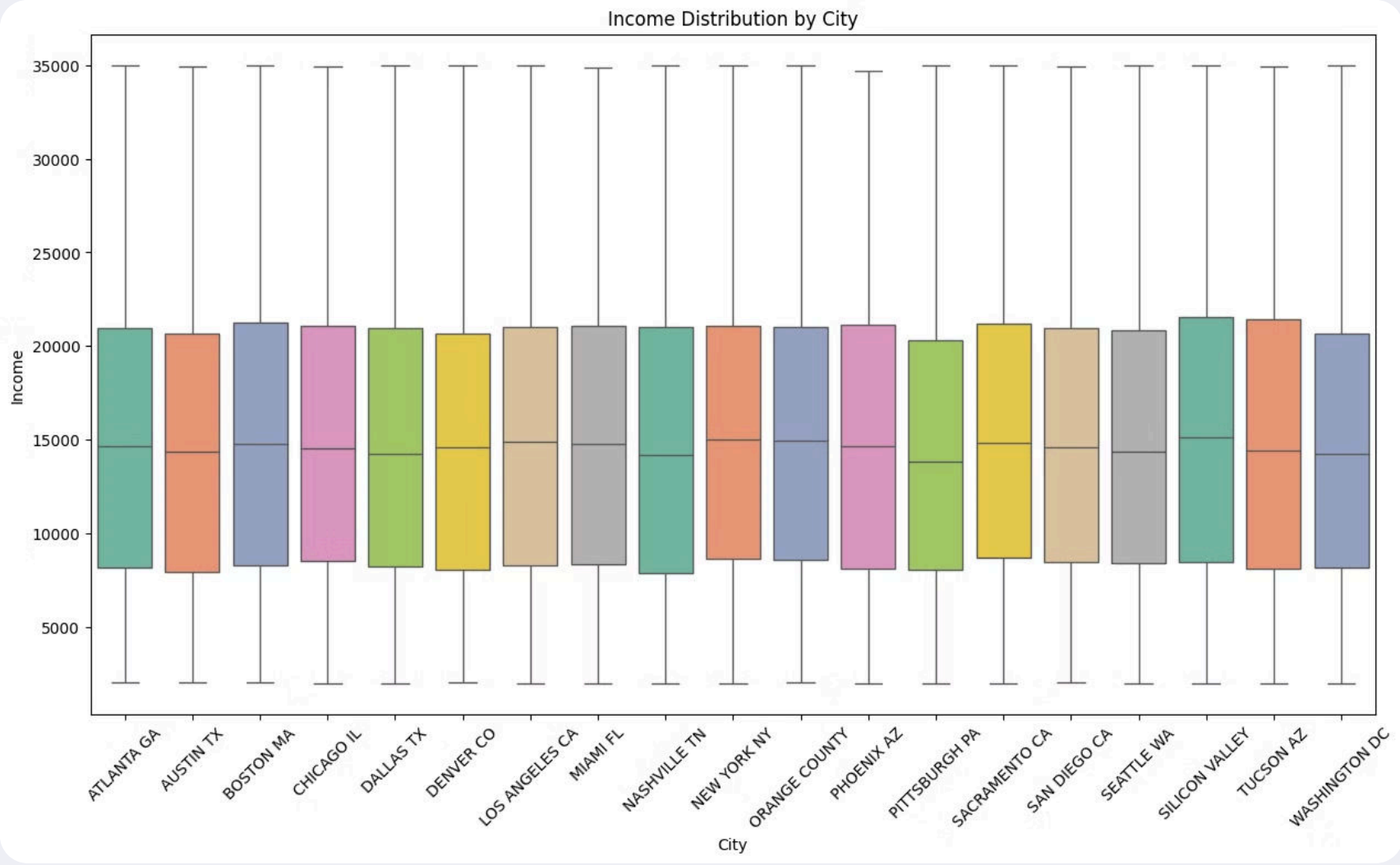
San Francisco, Boston, and Washington D.C. have the highest market penetration rates, far ahead of other cities.

Users vs Population in Cities



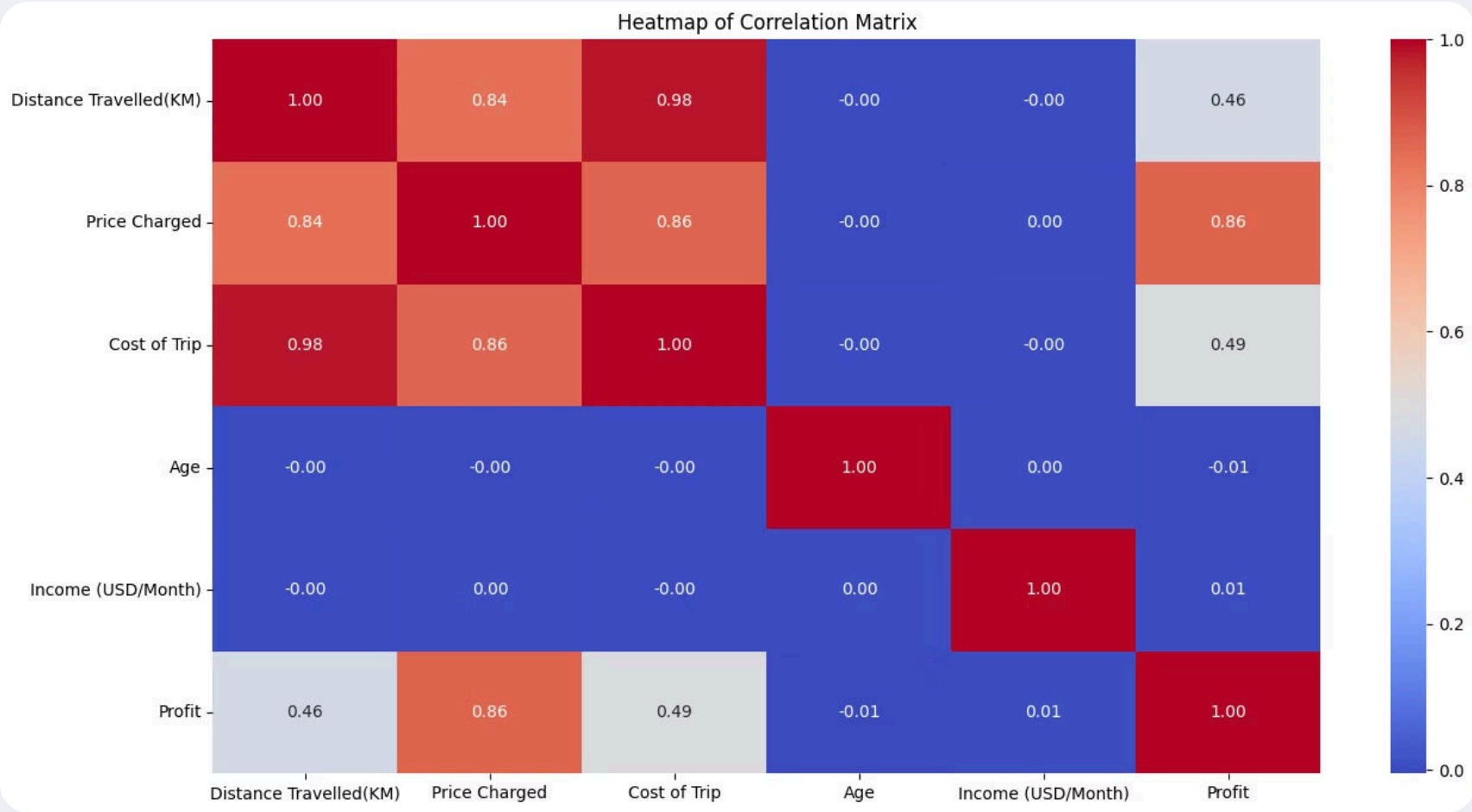
New York has the most significant population and the most users.

City and Income



The income distribution is very uniform across all cities, with almost no difference.

Overall Overview



As expected, the columns with the highest correlation are Price Charged, Cost of Trip, Profits, and Distance.

Hypothesis Testing

- Company Performance Hypothesis
- Market Penetration Hypothesis
- Customer Demographic Hypothesis
- Service Utilization Hypothesis
- Gender Preference Hypothesis

Company Performance Hypothesis

Question

Is there are a significant difference in the average profit per trip between the two cab companies across all cities? This will help in understanding which company has better financial performance.

Hypotheses

- **Null Hypothesis (H0):** There is NO significant difference in the average profit per trip between the two cab companies across all cities. Mathematically, it can be stated as H0: μ1 = μ2, where μ1 is the average profit per trip for Company A, and μ2 is the average profit per trip for Company B.
- **Alternative Hypothesis (H1):** There IS a significant difference in the average profit per trip between the two cab companies across all cities. This can be stated as H1: μ1 ≠ μ2.

Find P Value

```
TtestResult(statistic=-230.99551452746311, pvalue=0.0, df=313532.12464142457)
```

Result

We **reject** the null hypothesis and conclude that there is a statistically significant difference in the average profit per trip between Pink Cab and Yellow Cab across all cities. This finding suggests that one of the companies is more profitable on a per-trip basis, which is a crucial factor for VenturePath Investments to consider in their investment decision.

Market Penetration Hypothesis

Question

Does the company with the highest number of users in cities with the largest populations generate more profit? This aims to analyze the correlation between the market penetration (in terms of users) of the cab companies and their profitability in densely populated cities.

Hypothesis

- **Null Hypothesis (H0):** There is NO significant difference in the profits generated by cab companies in cities with the largest populations, regardless of the number of users.
- **Alternative Hypothesis (H1):** The company with the highest number of users in cities with the largest populations generates significantly more profit than the company with fewer users.

Find P Value

```
TtestResult(statistic=-218.89741242558824, pvalue=0.0, df=187858.54728451738)
```

Result

We **reject** the null hypothesis. This indicates that the difference in the average profits between Pink Cab and Yellow Cab in the largest cities by population is extremely statistically significant. The large magnitude of the t-statistic suggests a very pronounced difference between the profits of the two companies in these cities, highlighting the importance of market penetration in determining profitability.

Customer Demographic Hypothesis

Question

Does the average income of users significantly affect their choice of cab company? This will explore if there's a preference for a specific cab company among users with different income levels.

Hypothesis

- **Null Hypothesis (H0):** The average income of users does NOT significantly affect the choice of cab company. Users of different income levels choose cab companies at similar rates.
- **Alternative Hypothesis (H1):** The average income of users significantly affects the choice of cab company. Users with different income levels show a distinct preference for one cab company over the other.

Find P Value

```
TtestResult(chi2=6.926, pvalue=0.074)
```

Result

We **fail to reject** the null hypothesis. While there may be a trend or a slight association between user income levels and cab company preference, it is not strong enough to be considered statistically significant at the 5% level. However, it is close to the threshold, which could warrant further investigation or consideration of practical significance, suggesting that income may still play a role in consumer choice under certain conditions.

Service Utilization Hypothesis

Question

Is there a significant difference in the distance traveled by customers between the two cab companies, and does this affect overall company profits? This could indicate which company is preferred for longer or shorter trips.

Hypothesis

- **Null Hypothesis (H0):** There is NO significant difference in the average distance traveled by customers between the two cab companies.
- **Alternative Hypothesis (H1):** There IS a significant difference in the average distance traveled by customers between the two cab companies, which may affect overall company profits.

Find P Value

```
TtestResult(statistic=-0.19970297254383812, pvalue=0.8417131770122342, df=140913.77838204397)
```

Result

We **fail to reject** the null hypothesis, indicating that the difference in average distances traveled by customers between Pink Cab and Yellow Cab is not statistically significant. This suggests that customer preferences for Pink Cab versus Yellow Cab are not based on the distance they need to travel, implying that other factors may influence the choice of cab service.

Gender Preference Hypothesis

Question

Does the preference for a cab company differ significantly between male and female users? This examines if gender plays a role in the choice of cab service.

Hypothesis

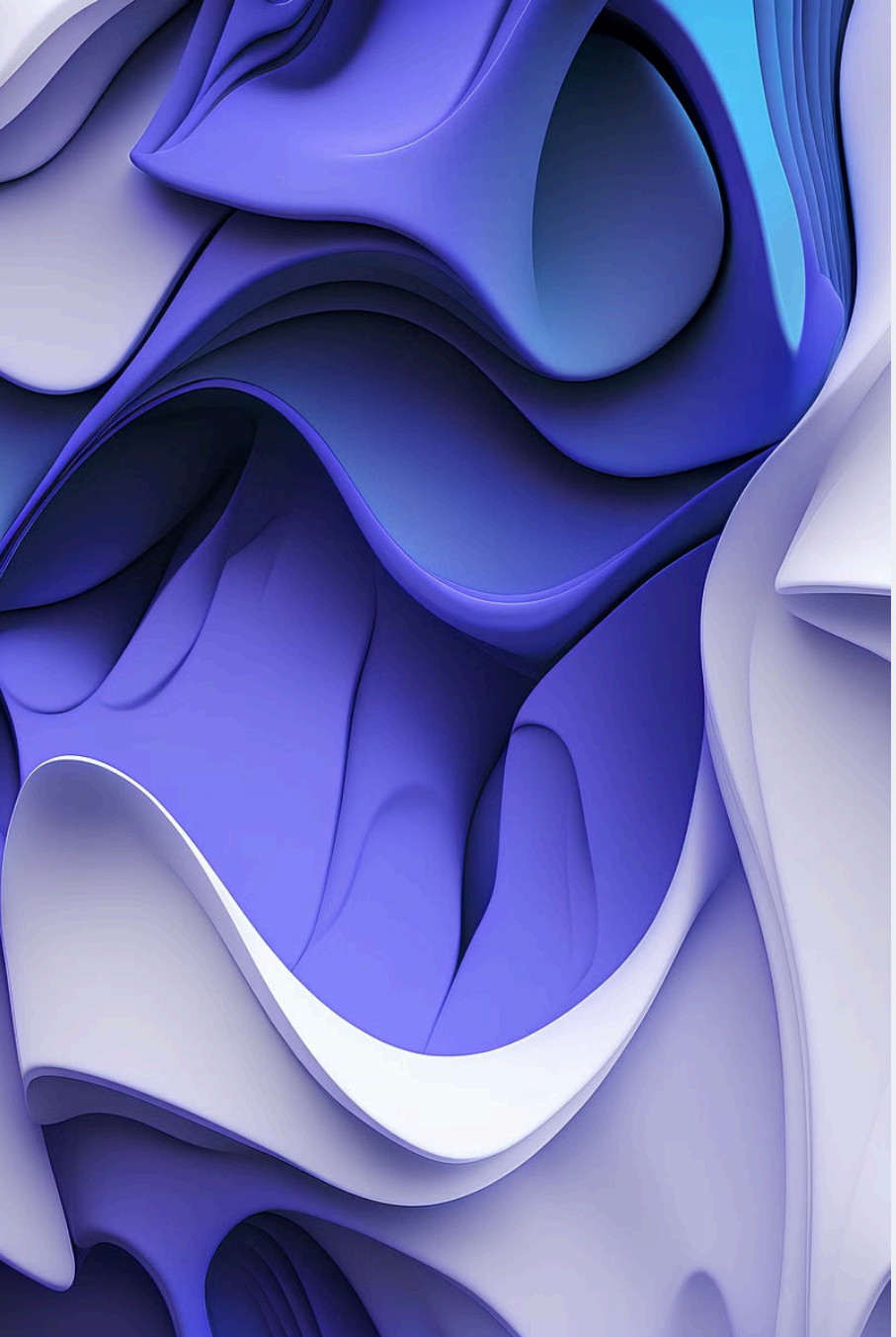
- **Null Hypothesis (H0):** There is NO significant difference in the preference for a cab company between male and female users. Gender does not play a role in the choice of cab service.
- **Alternative Hypothesis (H1):** The preference for a cab company differs significantly between male and female users, indicating that gender plays a role in the choice of cab service.

Find P Value

```
TtestResult(chi2=107.22063897254299, pvalue=3.982674650131372e-25)
```

Result

We **reject** the null hypothesis. This indicates that the observed association between gender and cab company choice is highly statistically significant. There is sufficient evidence to conclude that there is a significant difference in the preference for a cab company between male and female users, highlighting that gender plays a role in the choice of cab service.



Conclusion

- Analysis Summary
- Hypothesis Test Summary
- Insights and Investment Suggestion
- Final Conclusion

Analysis Summary

- **General Analysis** highlights Yellow Cab's dominant market share and the steady increase in trips from February to December, alongside an even spread of trip distances and a wide distribution of customer incomes.
- **Profit Analysis** reveals that Yellow Cab's total profit notably surpasses that of Pink Cab, correlating with its market share. Notably, New York emerges as a significantly profitable city for cab services, with profits increasing alongside trip distances.
- **City Analysis** indicates high market penetration rates in San Francisco, Boston, and Washington D.C., with New York standing out in terms of user base relative to population.

Hypothesis Test Summary

- We observed a **statistically significant difference in average profit per trip** between Pink Cab and Yellow Cab, suggesting varying profitability which could influence investment decisions.
- The **Market Penetration Hypothesis** was confirmed, highlighting the critical role of market penetration in profitability, especially in the largest cities.
- The **Customer Demographic and Service** Utilization Hypotheses indicated no significant effects of income levels or travel distances on cab company preference, suggesting other factors at play.
- The **Gender Preference Hypothesis** showed a significant difference in company preference between genders, pointing to potential targeted marketing strategies.

Insights and Investment Suggestion

Our analysis cover the cab industry, where factors such as market penetration, profitability, and demographic preferences play crucial roles. Yellow Cab's dominance in market share and profitability is clear, yet this alone doesn't provide a full picture of company performance. Critical factors such as operational efficiency, customer satisfaction, service quality, and customer loyalty remain to be thoroughly examined.

While Yellow Cab presents an attractive investment opportunity based on market share and profitability, Pink Cab's position suggests potential untapped opportunities, especially when considering demographic preferences like gender. Before making a definitive investment decision, VenturePath Investments should seek further information on customer satisfaction and service quality. These factors are pivotal in assessing the long-term sustainability and growth potential of both companies.

Final Conclusion

Given the current data, Yellow Cab shows strong performance indicators that align with VenturePath Investments' criteria for profitability and market penetration. However, the investment decision should not overlook the importance of customer-centric metrics, which could significantly impact the long-term viability and competitiveness of the chosen company. Therefore, we recommend a balanced approach, considering both quantitative financial metrics and qualitative measures of customer satisfaction and service quality, to make the most informed investment decision in the cab industry.

Thank you