

Data Intake Report

Introduction

The provided dataset offers an in-depth examination of taxi services, specifically focusing on two primary cab companies: the Yellow Cab and the Pink Cab. It covers a timespan of several years, capturing hundreds of thousands of transactions across numerous cities in the United States. The wealth of data available presents ample opportunities for robust analysis and strategic decision-making.

The primary data fields include:

- Transaction ID: the transaction ID of each taxi trip
- Date of Travel: the date of the taxi trip
- Company: two taxi companies, namely Pink Cab and Yellow Cab
- City: different cities in United States
- KM Travelled: the total distance (measured in kilometers) for each taxi trip
- Price Charged: the price charged of each taxi trip
- Cost of Trip: the cost of each taxi trip
- Customer ID: the ID of each customer of the taxi trip
- Payment_Mode: the method of payments(cash or card)
- Gender: the gender of each customer
- Age: the age of each customer
- Income (USD/Month): the income(in USD) of each customer

Detailed Findings

1. Seasonality in Cab Usage:

Quarterly seasonality: The 4th quarter (October–December) of each year appears to have more cab rides compared to the other quarters, showing an increase in the number of cab rides in this quarter. This could suggest that people use cab services more frequently in the last quarter of the year.

Monthly seasonality: If we look at the monthly data, it seems like there's a similar trend where the months from October to January have more cab rides, and then there's a decrease in the following months. The pattern seems to repeat every year.

Yearly trend: It seems there is an overall increase in cab usage from 2016 to 2018. However, in 2019, there is a sudden drop in cab usage in February. Please note that this might be due to incomplete data for 2019. The data demonstrated evidence of seasonality in the number of customers utilizing cab services. Additional statistical analysis is needed to fully understand and quantify the

nature and extent of this seasonality.

2. Company Popularity: During the observation period, the Pink Cab company demonstrated a lower customer count than the Yellow Cab company. Further investigation could explore why this discrepancy exists and what strategic steps Pink Cab might take to boost its customer base.

3. Pricing Factors: A predictive model was developed to assess the impact of various factors like income, age, cost of the trip, and distance travelled on the price charged. While all these variables showed a statistically significant relationship, 'Cost of Trip' and 'KM Travelled' exhibited the highest influence.

4. Company Revenue: In terms of total payments generated, the Yellow Cab company outperformed the Pink Cab company.

5. Geographic Variation: The city of New York emerged as the most profitable, leading the tally in total payments. This identifies New York as a high-traffic region for cab services, presenting opportunities for targeted business growth.

6. Correlation between Income and KM travelled

An analysis of customer income against the average fare showed a clear positive correlation. This suggests that as the income of a customer increases, the amount they spend on cab rides also tends to increase. This insight may indicate a tendency of higher-income customers to choose more expensive or longer rides.

7. Customer Demographics: The average income of customers was approximately 15048 USD per month. The age group that predominantly utilizes the service falls between 25 and 42 years, representing the working-age population. This information can guide demographic targeting in marketing campaigns.

Opportunities for Further Investigation

While the findings generated valuable insights, there are several areas where further, more detailed investigation could be beneficial:

- Seasonality Analysis: A detailed examination of seasonality could reveal peak demand periods, allowing companies to adjust their resources and strategies accordingly.
- Pricing Dynamics: Building a more sophisticated predictive model could provide insights into the factors with the most significant impact on pricing.
- Customer Segmentation: Differentiating customers by income and age can help

companies tailor their marketing and pricing strategies, ensuring they resonate with the intended audience.

Data Quality Assessment

The dataset provided was free of missing or null values, indicating a high level of data cleanliness and integrity. However, future data collection could consider modifications to enhance data utility. For instance, the 'Age' field, currently captured as continuous numerical data, could be better represented as segmented age groups, enabling easier analysis of customer demographics.

Adding additional demographic information, such as customer occupation or trip purpose, would augment the richness of the dataset and support a more comprehensive customer understanding.

Concluding Remarks

Overall, the dataset represents a thorough snapshot of the operations and customer demographics of the Yellow Cab and Pink Cab companies. Preliminary analysis has surfaced numerous valuable insights, but the depth and breadth of the data offer ample opportunities for further, more granular analysis. The dataset has the potential to become a potent tool driving strategic decision-making, informing everything from resource allocation to marketing strategy and pricing models.