Optimizing Revenue and Driver Retention in the Taxi Industry: A Data-Driven Analysis of Payment Methods, Fare Pricing, and Tip Amounts

Introduction

In the highly competitive rental taxi sector, optimizing revenue and ensuring driver satisfaction are crucial for sustained growth and long-term success. With the increasing reliance on data-driven decision-making, understanding the factors influencing fare pricing and revenue generation has become essential. This research aims to uncover actionable insights by analyzing the impact of different payment methods on fare amounts, and developing predictive models to estimate fare prices and driver tips.

Objective

By leveraging data-driven insights and predictive analytics, this research seeks to provide a comprehensive understanding of revenue dynamics in the taxi industry. The findings will guide strategic decisions to maximize profits, improve driver satisfaction, and ensure a competitive edge in the rapidly evolving market.

- **Payment Methods and Fare Pricing:** Analyze if different payment methods (e.g., cash, credit/debit card) affect fare amounts. This includes conducting A/B testing, using python hypothesis testing and descriptive statistics to extract useful information to generate more cash for drivers. In particular, to validate any observed differences in fare pricing of card versus cash payment.
- **Predictive Modeling:** Develop models to predict fare prices and driver tips, aiding in profit estimation and driver retention. Additionally, suggest tip amounts to customers by predicting tips based on trip distance and duration.

Data Overview

The comprehensive dataset of NYC Taxi Trip records was used for this analysis.
Link - https://www.kaggle.com/datasets/microize/newyork-yellow-taxi-trip-data-2020-2019
Data cleaning and feature engineering procedures were implemented to concentrate solely on the relevant columns to our objective.

Columns relevant for this research:
- passenger_count (1 to 7)
- trip_distance (miles)
- payment_type (Card and Cash)
- fare_amount
- tip_amount
- trip_duration (minutes)

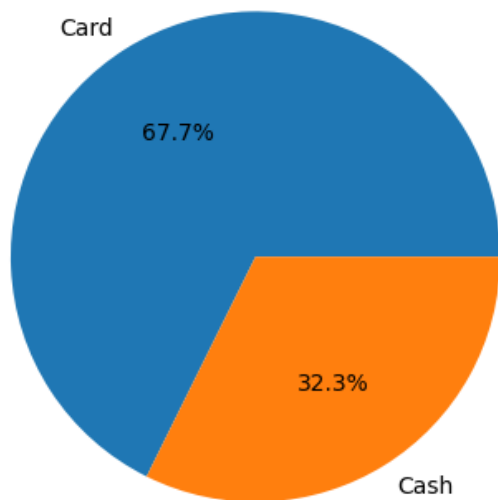| passenger_count	trip_distance	payment_type	fare_amount	tip_amount	trip_duration
1	1.2	Card	6.0	1.47	4.8
1	1.2	Card	7.0	1.5	7.4166
1	0.6	Card	6.0	1.0	6.1833
1	0.8	Card	5.5	1.36	4.85
1	0.0	Cash	3.5	0.0	2.3

Methodology

- **Descriptive Analysis:**
  - Performed statistical analysis to summarize key aspects of the data, focusing on fare amount and payment types.
- **Hypothesis Testing:**
  - Conducted a T-test to evaluate the relationship between payment and fare amount, testing the hypothesis that different payment methods influence fare amounts.
- **Regression Analysis:**
  - Employed linear regression to explore the impact of trip duration (calculated from pickup and dropoff times) and trip distance on fare amount as well as tip amount. Finally resulting in a regression model to help predict both fare amount and tip amount.

Insights

- Customers paying with card are considerably more than cash, clocking at 67.7% for card and rest for cash.

Payment type distribution

- The mean for the card is more than cash in both trip distance and fare amount. This indicates that passengers are more inclined to pay with a card.

	trip_distance		fare_amount	
payment_type	Mean	Standard Deviation	Mean	Standard Deviation
Card	3.134	2.239	13.417	6.324
Cash	2.740	2.164	12.062	6.088

- On any trip distance, card is the preferred mode. However when the fare amount is on a very lower scale, there is less difference between the preferences. But as distance and fare amount increase, the preference to pay with a card increases.
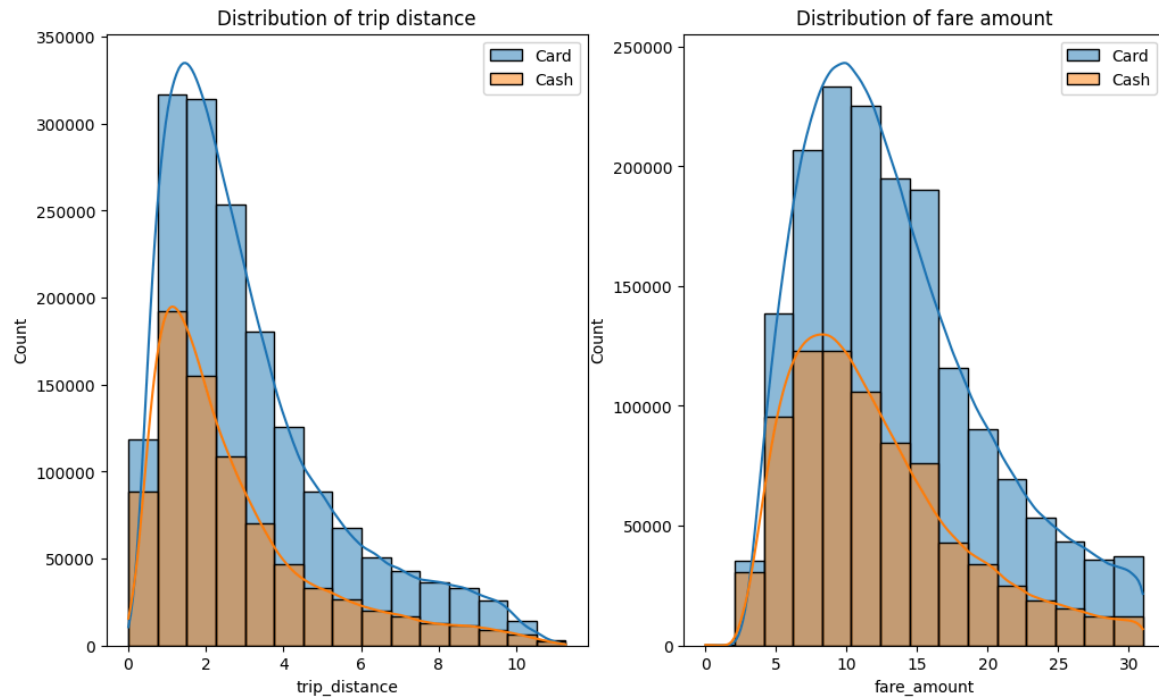
- Among card payments, rides constituting a single passenger comprise the largest portion, valuing at 38% of the total card transactions.
- A similar trend is also visible in cash payments, where rides constituting a single passenger value at 19% of the total cash payments.
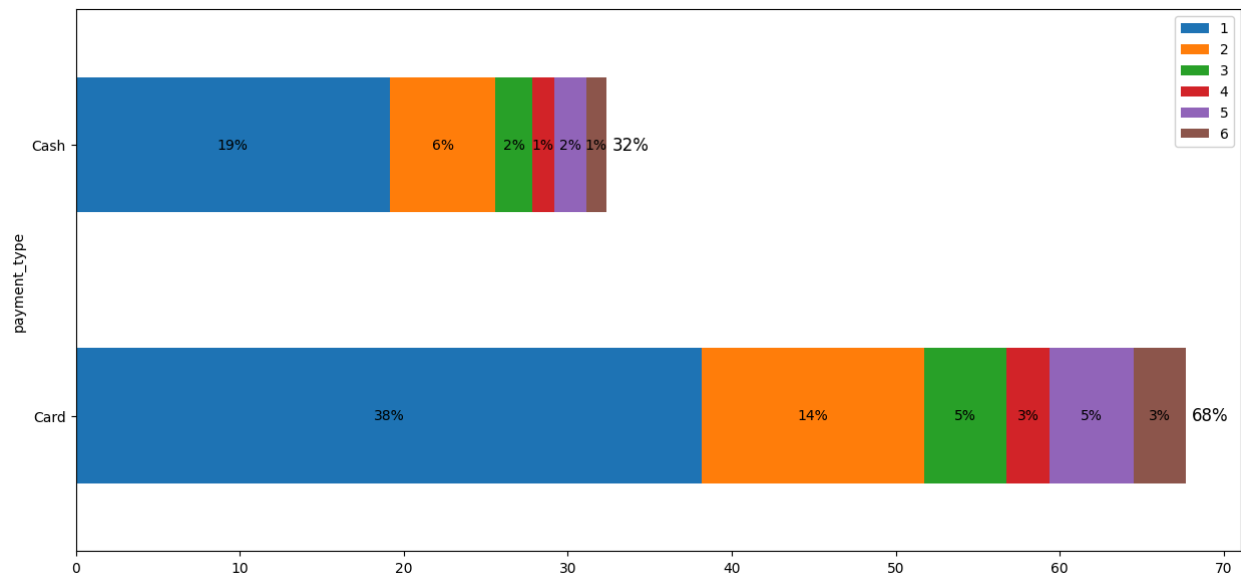
- Furthermore, there is a noticeable decrease in the percentage of transactions suggesting that larger groups are less likely to rent taxis or may avail other payment methods.

- The data also says that card payment is preferred in any number of passenger counts.
- The distribution for tip amount is not normally distributed. There appears to be three peaks with two peaks below the mean value of 2.56, suggesting that max drivers get paid less than average tip by passengers. Businesses can put up ads in the form of suggestions to encourage customers to pay more tip so that driver income and retention can increase.
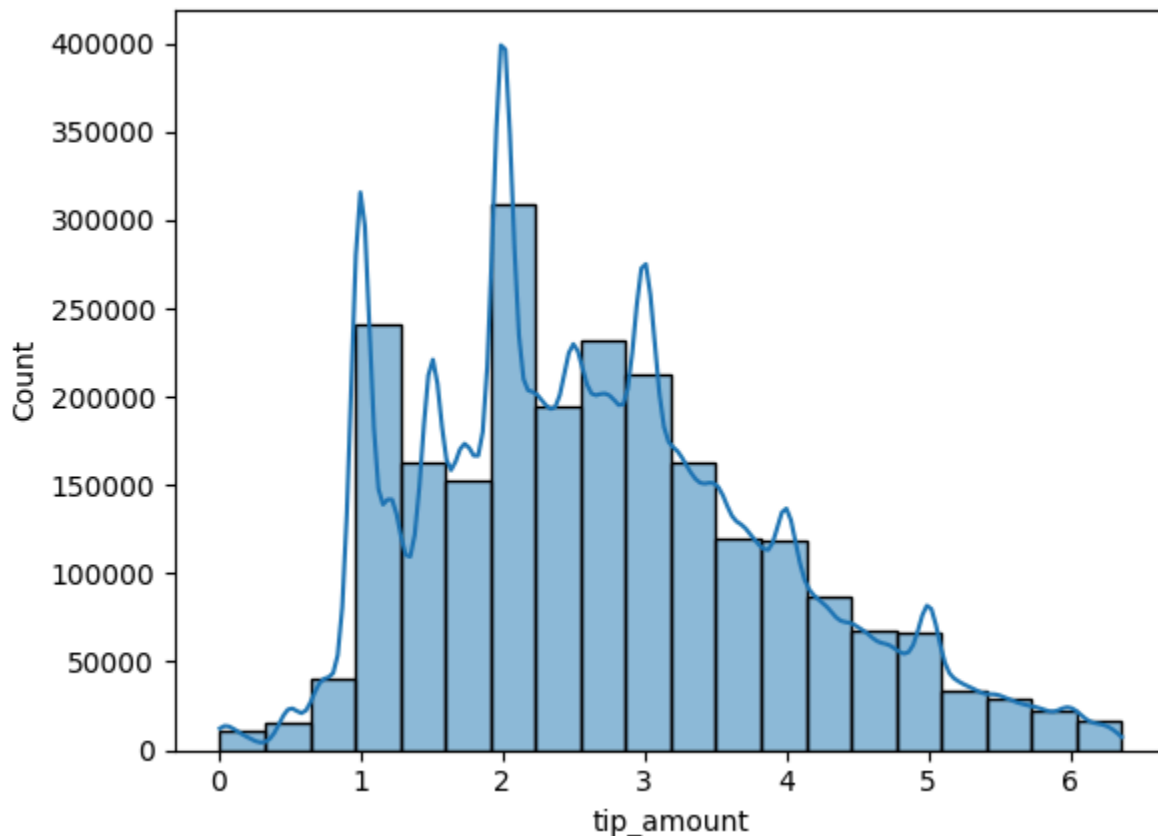
Hypothesis Testing

Null hypothesis: There is no difference in average fare between customers who use credit card and customers who use cash.

Alternate hypothesis: There is a difference in average fare between customers who use credit and customers who use cash.

Based on the fact that p-value is less than the significance level of 5%, the null hypothesis is rejected.
There is a statistically significant difference in average fare amount between passengers who pay with cash and passengers who use a card.

Predictive Modeling

Fare amount based on trip distance and trip duration:
- A linear regression model is built using the two features, trip_distance and trip_duration, to predict the fare amount.
- The initial RMSE value of the model was 0.83, but after implementing gradient boosting, the RMSE value decreased to 0.71 indicating an enhancement of the model.
- Predictions from this model can be used to make decisions in budget allocation and future planning of resources.

Tip amount based on trip distance and trip duration:
- A linear regression model is built using the two features, trip_distance and trip_duration, to predict the tip amount.
- The RMSE value of the model is 0.87
- Predictions from this model can be used to suggest tip amounts automatically to the passengers such that driver income and retention can be increased.

Recommendations

- Company should capitalize on the potential of generating more revenue by encouraging customers to pay using card.
- Strategies such as offering incentives, discounts or value added points for card payments should be implemented for customers to lean more towards paying the fare using a card.
- Seamless transactions and enhanced security options should be introduced to facilitate customer convenience and encourage adoption of this preferred payment method.
- The regression model for predicting fare amount can be used to analyze future profits and planning of resources.
- Predictive model of the tip amount can be implemented for auto suggesting tip amounts to passengers such that driver income and retention can be increased.