

Revenue Optimization for NYC Yellow Taxi Drivers

Introduction

In the dynamic landscape of urban transportation, taxi services face the dual challenge of maintaining operational efficiency and ensuring driver satisfaction. One critical factor that influences driver earnings is the payment method used by passengers. This project investigates the relationship between payment type and fare amount using publicly available data from the New York City Taxi & Limousine Commission (TLC). The overarching goal is to derive actionable insights that can help drivers maximize revenue.

Objective

The primary objective of this project is to analyze historical taxi trip data to answer the following research question:

Does the payment method (Credit Card vs. Cash) have a significant impact on the fare amount, and how can this insight be leveraged to increase driver revenue?

Dataset Overview

Dataset Source: NYC Yellow Taxi Trip Data - Kaggle

Time Period: January 2015 & January-March 2016

Key Attributes Used:

- tpep_pickup_datetime: Start time of the trip
- tpep_dropoff_datetime: End time of the trip
- passenger_count: Number of passengers
- trip_distance: Distance covered during the trip (in miles)
- payment_type: Mode of payment (coded as integers)
- fare_amount: Fare charged for the trip

Additional computed columns:

Revenue Optimization for NYC Yellow Taxi Drivers

- duration (in minutes): Derived from pickup and dropoff times.

Data Preprocessing

The data preprocessing steps include:

- Parsing date fields into datetime objects.
- Removing rows with null/missing values, non-sensible values (e.g., zero fare), or outliers using IQR.
- Filtering only Credit Card (1) and Cash (2) payments, and valid passenger counts (1 to 5).

After cleaning, the dataset was reduced to approximately 386,000 meaningful trip records.

Exploratory Data Analysis

Fare amounts and trip durations are right-skewed. Boxplots showed significant outliers, removed using the Interquartile Range (IQR) method.

Credit Card fares tend to have higher average values. Cash fares are more concentrated in the lower range (under \$10). Presence of tips in Credit Card payments may contribute to this.

Statistical Analysis

To validate the observed difference, a statistical test such as independent two-sample t-test can be conducted to determine if the difference in fare distributions by payment type is statistically significant.

Key Insights

- Credit Card payments are associated with higher fare amounts.
- Cash payments are generally for shorter, lower-value trips and may underreport revenue due to missing tip data.

Revenue Optimization for NYC Yellow Taxi Drivers

Business Implications

For Drivers:

- Target areas/times with high card usage.
- Promote card payments for convenience and potential tips.

For Taxi Operators:

- Invest in card infrastructure.
- Use data to guide driver deployment and incentive programs.

Recommendations

1. Incentivize Credit Card Usage.
2. Time-Slot Analysis for payment patterns.
3. Zone-Based Strategy using geospatial data.
4. Predictive modeling for revenue optimization.
5. Include Tip Prediction Model.

Limitations

- Tip data missing for cash transactions.
- Payment preference bias may affect analysis.
- Results are based on a subset and may not generalize to all periods.

Future Work

- Incorporate spatial data for geospatial analysis.
- Analyze trip frequency and driver performance.
- Use regression or forecasting for fare prediction.

Revenue Optimization for NYC Yellow Taxi Drivers

Conclusion

The project reveals that credit card payments yield higher fare values, largely due to longer trips and inclusion of tips. These insights support more profitable strategies for drivers and taxi operators to enhance revenue outcomes.