Executive Summary: NYC Taxi Fare Prediction - Exploratory Data Analysis

Prepared for: New York City Taxi & Limousine Commission
Prepared by: Automatidata

Project Overview

This report presents an exploratory data analysis (EDA) of the 2017 Yellow Taxi Trip Data, focusing on understanding taxi ridership trends and revenue patterns in New York City. The analysis aims to provide actionable insights to the New York City Taxi and Limousine Commission (TLC) by examining trip characteristics, identifying outliers, and visualizing key trends using Python and Tableau.

Key Insights

Outlier Anomalies: Significant outliers in trip distance, fare, and tip amounts necessitate thorough investigation, potentially indicating data errors or unique trip scenarios that impact model accuracy and operational understanding.

Temporal Ridership Peaks: Ridership and revenue exhibit strong temporal patterns, with Thursdays and Fridays consistently showing peak activity, suggesting opportunities for resource optimization.

Location-Based Trip Variances: Notable variations in average trip distances across drop-off locations highlight areas with distinct travel characteristics, enabling targeted service adjustments.

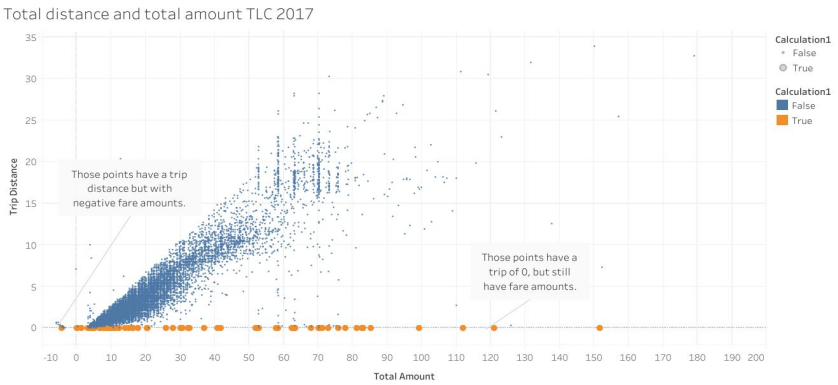
Passenger-Driven Tip Trends: A positive correlation between passenger count and tip amounts indicates a clear customer behavior pattern, informing service and incentive strategies.

Data Skewness and Distribution: The right-skewed distribution of trip distances, fares, and tips emphasizes the prevalence of shorter, lower-cost trips, with occasional outliers significantly impacting overall statistics.

Trip Distance vs Total Amount Correlation: A strong positive correlation exists between trip distance and total amount, with some data anomalies at 0 trip distance and negative total amounts.

Details

- EDA conducted using Python and Tableau for visualization.
- Dataset contains no null values, but significant outliers.
- Scatter plot analysis highlighted the trip distance vs. total amount correlation and anomalies.



Alt Text: Graph displaying New York City TLC data plotting variables for total distance and total amount.

Next Steps

- **Address Outliers:** Implement a strategy to address data anomalies.
- **Integrate External Data:** Add weather and event data for better prediction.
- **Develop Predictive Models:** Create models to forecast trip costs.
- **Optimize Location Strategies:** Improve taxi availability in key areas.
- **Continuous Monitoring:** Track KPIs and refine strategies.
- **Validate Data with TLC:** Confirm data accuracy.