**Executive Summary: NYC Taxi Trends and Revenue Optimization Project Proposal**

**Overview**

This project explores data from the New York City Taxi and Limousine Commission (TLC) to identify fare trends, uncover revenue opportunities, and support data-driven operational strategies. The analysis leverages Python, statistical testing, regression modeling, and machine learning to provide actionable insights for optimizing fare structures and improving financial performance.

**Problem**

The NYC TLC needs to understand how various factors—such as trip distance, time, and payment method—influence fare amounts. Inconsistent revenue patterns and potential inefficiencies in payment systems limit the agency’s ability to forecast demand and optimize pricing strategies.

**Solution**

We performed an end-to-end data analysis using a large TLC trip dataset. The workflow included data cleaning, exploratory data analysis (EDA), hypothesis testing on payment methods, multiple linear regression modeling, and machine learning-based fare prediction. The goal was to identify the primary drivers of fare variation and recommend operational strategies for revenue optimization.

**Key Insights**

* Payment Method Impact: Credit card transactions consistently yielded higher average fares and more frequent tipping, contributing significantly to total revenue. Hypothesis testing confirmed a statistically significant difference in mean fares between cash and credit card payments.
* Trip Distance as Primary Driver: Regression analysis showed that trip distance had the strongest influence on fare, followed by trip duration.
* Fare Prediction: A machine learning model (linear regression with regularization) achieved an R² of ~0.70, indicating strong predictive power based on tripcharacteristics. However, performance declined for short trips due to overestimation bias.
* Data Anomalies: Zero or negative fare values and durations were identified as outliers and excluded or imputed to improve model integrity.
* Time of Day and Vendor: These factors had weaker influence and were not significant predictors in the final models.

**Next Steps**

* Encourage Card Payments: Implement incentives for digital payments to drive higher average fares and tipping.
* Target High-Value Routes: Focus on promoting longer-distance rides and possibly reevaluate flat-rate zones.
* Refine Predictive Models: Incorporate external data sources like weather or traffic to improve prediction accuracy, especially for short-distance trips.
* Monitor Seasonal Trends: Use time-series dashboards to visualize and plan around monthly demand fluctuations.
* Evaluate Fare Adjustment Pilots: Run A/B tests on pricing strategies in different boroughs or time slots.

**Impact**

The findings provide a data-driven foundation for strategic decisions at the NYC TLC. By shifting customer behavior toward more profitable payment methods and optimizing services around distance-based revenue trends, the agency can significantly improve fare collection and operational efficiency. The predictive models also enable better forecasting for budgeting and service planning.