

**NYC Taxi Fare Prediction-Regression Model Proposal**

Leveraging NYC TLC Trip Data to Identify Key Variables and Build Predictive Models for Fare Optimization

**Project Overview:** The NYC Taxi and Limousine Commission (TLC) has collected detailed taxi and rideshare trip data. Our objective is to create a regression model to predict taxi fares before each ride using distance, time of day, and other relevant factors. This project will follow the **PACE strategy**: **Plan, Analyze, Construct, Execute** to ensure a structured and transparent workflow.

# Details

Details

Key Insights

Key Insights

* Credit card is the most common payment method (≈ 67% of trips), followed by cash.
* Average tip for credit card payments is **$2.73**, while cash trips have **no recorded tips**.
* Vendor 1 and Vendor 2 have almost identical mean total fares (~$16.30).
* Most credit card trips have **1 passenger**, with 2 passengers as the second most common.
* Average tip amounts are relatively consistent across passenger counts for credit card trips.
* Passenger counts of **0** still appear in the data, which may indicate data entry errors.

***Linear Relationship between ‘Fare Amount’ and ‘Trip Distance’***

# Next Steps

Next Steps

* Clean data and remove outliers.
* Train regression & tree-based models.
* Evaluate using RMSE, MAE, and R².
* Create an interactive dashboard.
* Align insights with NYC TLC objectives.