Project Proposal & PACE Strategy Document

# Project Overview

Project Title: Predictive Fare Model for NYC Taxi and Limousine Commission  
Project Goal: Build a regression model to estimate taxi fares based on trip characteristics such as distance, pickup/dropoff time, and location, using NYC TLC trip data.

# PACE Strategy

## PLAN

Goal: Develop a regression model that estimates taxi fares before the ride begins, using historical data from the NYC TLC.  
  
Key Questions:  
- What variables (e.g., distance, time, pickup/dropoff location) are available in the dataset?  
- Are there missing, duplicate, or inconsistent entries in the data?  
- What metrics will determine a successful model (e.g., MAE, RMSE, R²)?  
- Who are the primary and secondary stakeholders?  
- What level of technical detail should be communicated to each stakeholder?  
  
Stakeholders:  
Internal: Deshawn Washington, Luana Rodriquez, Udo Bankole, Uli King  
External: Juliana Soto, Titus Nelson  
  
Constraints: Timeline (TBD), non-technical communication for TLC, tools: Python

## ANALYZE

Tasks:  
- Load and explore the dataset  
- Perform data cleaning (remove nulls, fix data types, remove duplicates)  
- Conduct EDA to uncover trends and correlations  
- Identify and engineer relevant features  
- Analyze correlation between fare and key variables  
  
Tools: Python (pandas, matplotlib, seaborn), Correlation matrices, Descriptive statistics

## CONSTRUCT

Tasks:  
- Choose and build regression model (e.g., linear regression, decision tree)  
- Use A/B testing to compare model approaches  
- Optimize model with parameter tuning  
- Validate model (cross-validation, RMSE, R²)  
  
Considerations: Use explainable models, ensure fairness

## EXECUTE

Tasks:  
- Finalize and document regression model  
- Prepare non-technical presentation for TLC  
- Create visuals showing insights and predictions  
- Present recommendations and highlight performance  
  
Deliverables: Predictive model, dashboard/visuals, project summary, presentation

# Project Milestones

* M1 - Project Planning: Define goals, assign team roles, communication (Plan)
* M2 - Data Acquisition & Inspection: Load and inspect dataset (Analyze)
* M3 - Data Cleaning: Clean and format data (Analyze)
* M4 - EDA: Discover patterns and trends (Analyze)
* M5 - Feature Engineering: Create useful variables (Analyze)
* M6 - Model Development: Build regression model (Construct)
* M7 - Model Testing & Validation: Evaluate model (Construct)
* M8 - Insight Generation: Translate output to insights (Execute)
* M9 - Visualization Creation: Develop visuals for TLC (Execute)
* M10 - Final Presentation: Present findings to TLC (Execute)

# Stakeholder Communication Strategy

Automatidata Team (Technical):  
- Share model performance, EDA findings, and code-level updates  
  
TLC Team (Non-technical):  
- Share visual summaries, model benefits, and key insights in simple terms

# Key Questions to Explore Before Modeling

1. 1. What are the most predictive variables for fare amount?
2. 2. Are there seasonal, hourly, or locational pricing trends?
3. 3. Are there anomalies or outliers (e.g., $0 fares)?
4. 4. Should external datasets (e.g., weather, traffic) be considered?

# Final Deliverables

- Regression model for fare prediction  
- Executive-friendly visual report  
- Summary of evaluation metrics  
- Presentation for TLC team