Project Title: Automatidata project proposal
Prepared By: Akash Raj
Date: 19-12-2025

Business Problem:

The New York City Taxi and Limousine Commission (TLC) serves millions of passengers daily and regulates a large network of taxi and for-hire vehicles. However, passengers often lack fare transparency before starting a trip, which can lead to uncertainty, dissatisfaction, and disputes.

TLC wants to provide riders with a reliable way to estimate taxi fares before the ride begins. To do this, the agency needs to better understand the factors that influence fare amounts and use historical trip data to build a predictive solution.

Project Objective:

Develop a **regression-based predictive model** that estimates the taxi fare amount before a trip begins using historical TLC trip data, including:

- Trip distance
- Trip duration
- Pickup and drop-off locations
- Passenger count
- Improvement surcharge
- Rate codes and time-based surcharges
- Payment type
- Any additional relevant trip features

The final goals are to:

- Accurately estimate taxi fares before the trip starts
- Identify the key drivers influencing fare amounts
- Improve fare transparency and rider experience
- Support TLC in building customer-facing fare estimation tools

PACE FRAMEWORK:

Stage	Tasks	PACE Stage	Deliverables
1	Understand the business scenario and define the problem	Plan	Project Proposal
2	Data exploration and data cleaning	Plan, Analyze	EDA Summary
3	Determine which models are most appropriate	Analyze, Construct	Model Selection Report
4	Construct the model	Construct	Model Development File
5	Confirm model assumptions	Analyze, Construct	Model Assumptions File
6	Evaluate model results	Analyze	Model Evaluation Report
7	Present actionable insights	Execute	Final Insights Report