NYC Taxi Fare Price Prediction Using Regression Modeling

Overview: The goal is to develop a regression model that accurately predicts taxi fare prices in NYC using real-world TLC ride data. This model will consider variables such as trip distance, pickup/dropoff time, and other ride-related metrics. The outcome will improve fare estimate accuracy for stakeholders and support future fare policy decisions.

Milestone	Tasks	Deliverables/Reports	Relevant Stakeholder (Optional Activity)
1	Write a project proposal Plan ▾	• Global-level project document	Automatidata Lead
1a	Establish structure for project workflow (PACE) Plan ▾	Timeline + Task map	Internal team
2	Compile summary information about the data Analyze ▾	• Data files ready for EDA	N/A

Course 1: Foundations of Data Science

2a	**Begin exploring the data** ˅ **Analyze** ˅	Cleaning data (removing outliers, handling nulls) is part of data construction.	**N/A**
3	**Data exploration and cleaning** ˅ **Analyze** ˅ and **Plan** ˅	● EDA report	N/A
3a	**Visualization building** ˅ **Construct** ˅ and **Analyze** ˅	● Tableau dashboard/visualizations	Business stakeholders
4	**Compute descriptive statistics** **Analyze** ˅	● Analysis of testing results between two important variables	N/A
4a	**Conduct hypothesis testing** ˅ **Construct** ˅ and **Analyze** ˅	Regression model	Data team
5	**Build a regression model** ˅ **Analyze** ˅ and **Construct** ˅	RMSE/R² results	N/A
5a	**Evaluate the model** ˅ **Execute** ˅	● Determine the success of the model	Automatidata Lead

6	Build a machine learning model Construct ▾	• Final model	N/A
6a	Communicate final insights with stakeholders Execute ▾	• Report to all stakeholders	NYC TLC