Automatidata – Taxi Fare Prediction Model

New York City Taxi and Limousine Commision Executive Report

ISSUE / PROBLEM

The New York City Taxi and Limousine Commission (TLC) has commissioned Automatidata to construct a predictive model. This model, built on a diverse set of variables, aims to accurately forecast taxi fare prices with a high degree of confidence.

RESPONSE

The Automatidata team engineered a multiple linear regression model, utilizing variables with the highest correlation with the dependent variable.

This model proficiently predicted taxi fares before the start of the ride.

The model demonstrated robust performance across both training and testing datasets.

IMPACT

During the EDA process, the treatment of outliers, specifically those related to duration and fare amount, enhanced the model's optimization.

The constructed model is capable of reliably forecasting the estimated fare amounts for taxi rides before the journey begins.

The scatterplot presented below offers a graphical illustration of the predicted versus actual fare amounts. The plot clearly demonstrates that the model possesses a high degree of reliability in forecasting the fare amount for taxi rides.
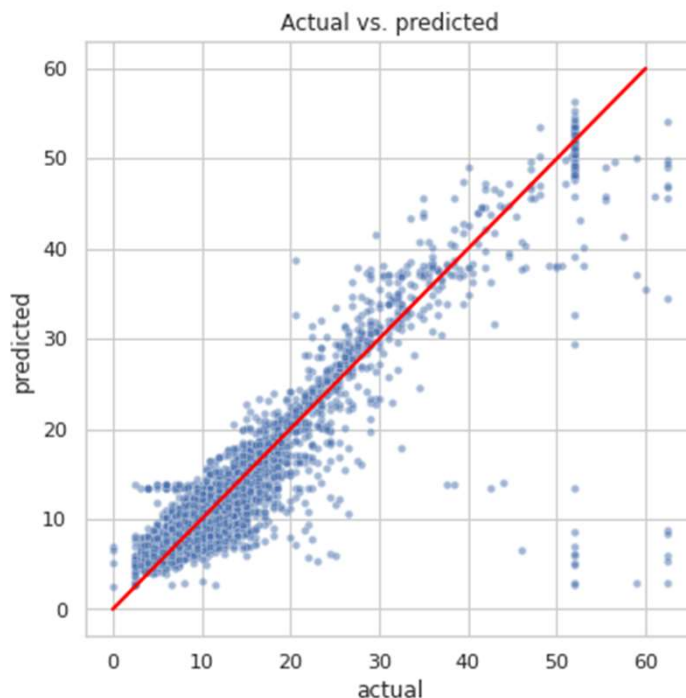
The plot shows a linear regression model of predicted vs actual fare amount values.

Model Metrics

R-squared (R2): 0.83 *(83% of variance described by the model)*
Mean Absolute Error (MAE): 2.15
Mean Squared Error (MSE): 17.52
Root Mean Squared Error (RMSE): 4.16

KEY INSIGHTS

- The feature that held the most significant predictive value was ride_duration.

- For each increment of 3.2 miles traversed, there is an average fare increase of $7.53. However, due to the presence of multi-collinearity among some features, this prediction may not be entirely dependable.

- In general, the model offers a robust prediction of fares.

- The New York Taxi and Limousine Commission (TLC) could leverage this model to develop an application that would enable riders to view estimated fares prior to the start of their journey.