

New York TLC Project Regression Model

Executive Summary Report

➤ ISSUE / PROBLEM

The Automatidata team asks the client to build a regression model to estimate which variables affect fare prices.

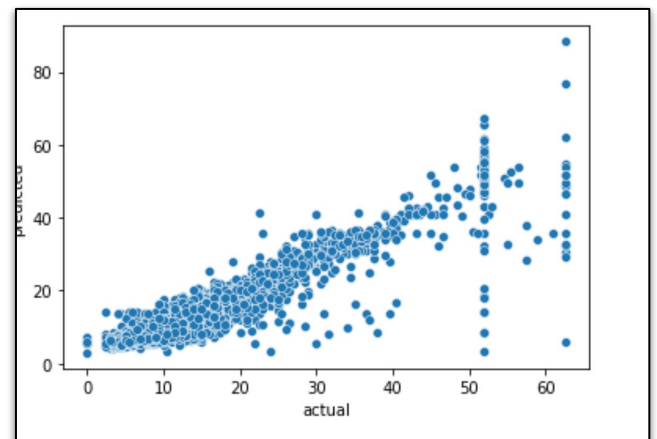
➤ RESPONSE

The Automatidata team choose multiple linear regression as the regression model, and the model successfully estimate cab fares. The model performance was great on both train and test datasets, suggesting that there are no overfitting occurred and the model is not biased.

➤ KEY INSIGHTS

➤ IMPACT

Extensive exploratory data analysis is a requirement to be done before modelling, removing the outliers would significantly lessen the impact.



Actual vs predicted scatter plot for the multiple linear regression model built by the Automatidata team.