Regression Assumptions after Modelling

Executive Summary for the New York Taxi and Limousine Commission
Prepared by **Automatidata**

ISSUE / PROBLEM

The New York Taxi and Limousine Commission contacted Automatidata to predict the cab fares. In this part of the project, the team developed a regression model to predict those fare, since it was part of the original project.

RESPONSE

The team created a Multiple Linear Regression (MLR) model to predict the cab fare prior to the ride with the data provided.

The results from our model showed an excellent performance on both training and test data. Indicating that the model is not over-biased nor over-fit.

IMPACT

Our model uses the independent variable as fare commission and the dependent variables were passenger count, mean distance, mean hour, and rush hour.
As mentioned above, these were used for our model, creating a reliable model to predict the cab fare.

Below, it shows a graph that evaluates the actual value of the fare ride with the expected one. Due to the results obtained, it can be concluded that this model can be used for the organization. In the notebook, there's further details about the process, including the residual.
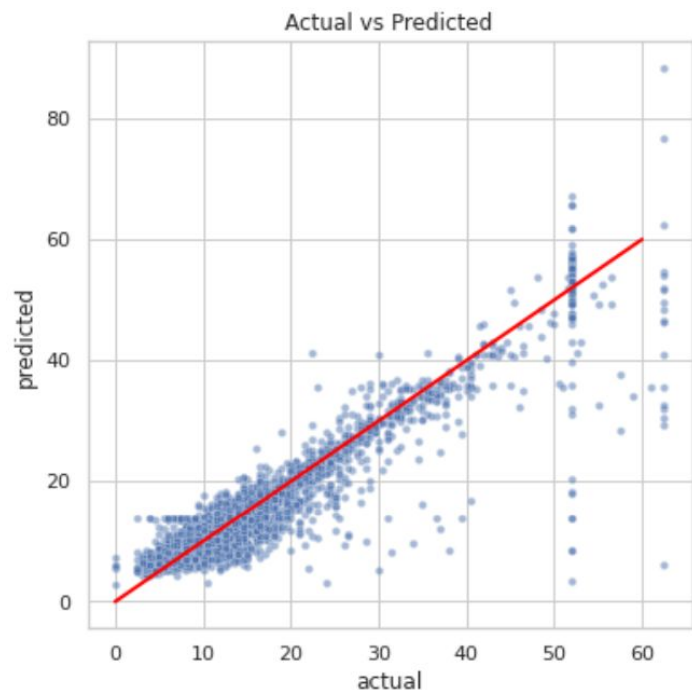
Alt -text: The graph shows a linear regression between the the predicted and actual values for the cab fire.

Model Metrics:

- R^2: 86.82%
- MAE: 2.13 %
- MSE: 14.33%
- RMSE: 3.78&

KEY INSIGHTS

- It was found that the mean distance was the variable with the greatest weight, indicating that for every 3.57 miles, the mean fare increase by $7.13. Although, further investigation is required to consider external factors..

- The New York Taxi and Limousine can use this model to predict the cab fare before the client request a ride.

- The model metrics presents a good variance and results, indicating an excellent viability.