Regression Modelling and Insights

Executive summary report for the New York City Taxi and Limousine Commission

Prepared by **Automatidata**

ISSUE / PROBLEM

The New York City Taxi & Limousine Commission enlisted Automatidata to forecast taxi cab fares. The data team at Automatidata developed the requested deliverable for their client: a regression model.

RESPONSE

Opting for a multiple linear regression (MLR) model, the Automatidata data team analyzed the provided data's type and distribution. The MLR model demonstrated effectiveness in estimating taxi fares, preemptively.

With strong performance observed across both training and test sets, the model indicates minimal bias and avoids overfitting.

IMPACT

Imputing outliers optimized the model, specifically in regards to the variables of: fare amount and duration.

The linear regression model provides a sound framework for predicting the estimated fare amount for taxi rides.

In order to showcase the efficacy of the linear regression model, the Automatidata data team included a scatter plot comparing the predicted and actual fare amount. This model can be used to predict the fare amount of taxi cab rides with reasonable confidence. The provided notebook exhibits further analysis on the model residuals.
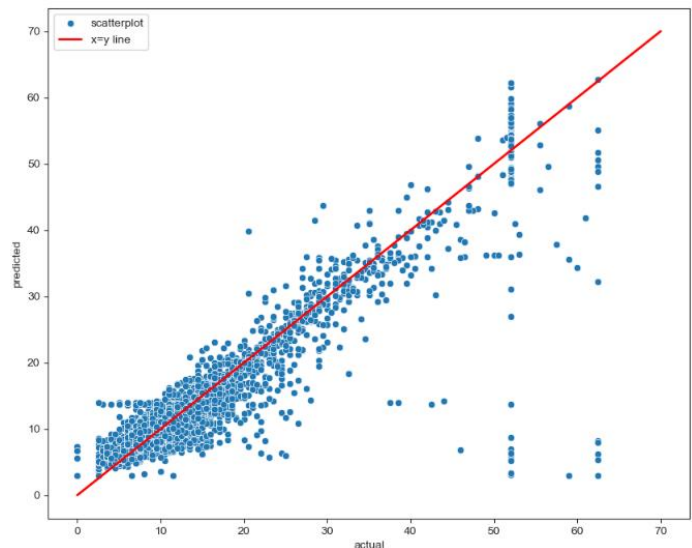
Alt-text: The scatter plot shows a linear regression model plot illustrating predicted and actual fare amount for taxi cab rides. The two vertical lines in scatterplot represent special cases/scenarious.

Model metrics:

- Net model tuning resulted in:
  - ✓ R^2 0.83, meaning that 83% of the variance is described by the model.
  - ✓ MAE 2.1
  - ✓ MSE: 17.5
  - ✓ RMSE 4.2

KEY INSIGHTS

- The feature with the greatest effect on fare amount was ride duration, which was not unexpected. The model revealed a mean increase of $2 for each additional mile travelled. This is not very reliable, however, due to high correlation between some features, especially mean_duration and mean_distance.

- Additional data for unrepresented or underrepresented cases could help.

- This regression model can be used to by NYC Taxi and Limousine commission to create an app that predicts fares before starting rides.

- The model provides a reliable and decent prediction for the fares.