Regression Assumptions After Modeling

Executive summary report for the New York City Taxi and Limousine Commission Prepared by **Automatidata**

ISSUE / PROBLEM

The New York City Taxi & Limousine Commission contracted Automatidata to predict customer taxi/rideshare durations to help understand its pricing and fleet viability.

RESPONSE

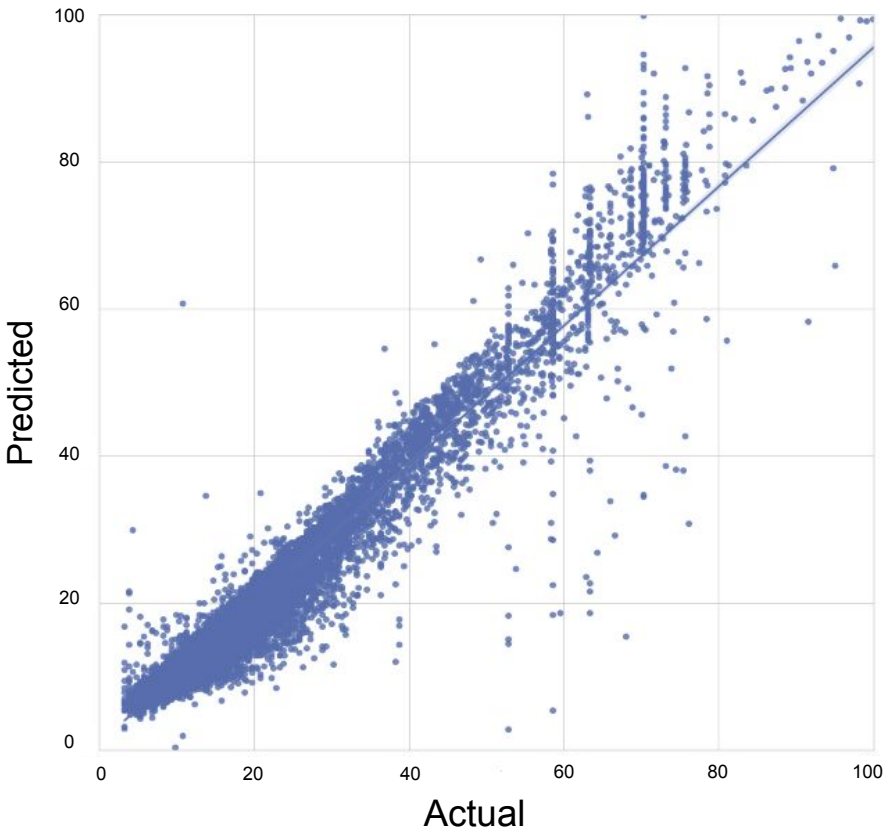
Linear regression model explains 95% price variance based on trip distance, tolls, tips, and payment.

265 pickup and destination codes significantly contributed to prediction performance.

IMPACT

Removing outliers optimized model (amount, passengers, distance).

Model provides a sound framework for predicting price based on pickup and destination distance IDs.



Alt-text: Taxi pricing regression model plot illustrating predicted and actual cost.

Regression assumptions

- Net model tuning resulted in:
  - ✓ R^2 from 0.85 to 0.9
  - ✓ MAE from 0.16 to 0.14
  - ✓ RMSE from 0.39 to 0.23
- Identification of a potentially curious influencer.
  - ✓ Tipping

KEY INSIGHTS

- Outlier removal had significant improvement on the model's effectiveness.
- Request additional data to perform 2nd stage reliability analysis for pricing & duration.
- Consider fine tuning travel time greater than 10 hours (43 pts).

```
Total values > 20 hours==>: 41
total values > 10 hours==>: 43
```