NYC Taxi Fare Prediction— Model Assumption Confirmation Report
Automatidata • NYC Taxi Analytics Project

OVERVIEW

❖ This stage focuses on validating whether the final predictive model meets practical and statistical reliability requirements.

❖ While Ordinary Least Squares (OLS) regression was used earlier as a **diagnostic tool** to understand linear relationships and hypothesis validity, the final production model is a **Random Forest regressor**.

❖ Model assumptions were therefore evaluated using residual behavior and prediction stability rather than strict linear assumptions.

IMPACT

❖ The assumption validation confirms that the final model produces **consistent and reliable fare estimates** without violating practical modelling expectations.

❖ This strengthens confidence in deploying the model for real-world fare prediction, where pricing behavior is inherently non-linear and influenced by multiple interacting factors.

❖ Overall, the validated model supports scalable and trustworthy fare estimation for rider-facing applications.

RESPONSE

❖ Residual diagnostics were performed by comparing predicted fares against residual values for both the baseline and final models.

❖ The final Random Forest model shows residuals that are largely centred around zero with no clear systematic trend.

❖ Compared to the baseline model, residual spread is reduced, and extreme prediction errors are less frequent, indicating improved generalization.

KEY INSIGHTS

❖ Residual plots for the final model indicate minimal bias across low, medium, and high fare ranges.

❖ No strong funnel shape or patterned structure is observed, suggesting stable variance in predictions.

❖ Non-linear pricing behavior present in earlier models is effectively captured by the Random Forest approach.

❖ Residual dispersion remains controlled even for higher predicted fares, indicating robust handling of complex ride scenarios.

Final Model   Model 1