Model Comparison: Gradient Boosting vs XGBoost

Evaluation Metrics

- **MAE (Mean Absolute Error)**:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

Measures the average magnitude of errors in a set of predictions, without considering their direction.

- **RMSE (Root Mean Squared Error)**:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Penalizes larger errors more heavily than MAE.

- **R^2 Score (Coefficient of Determination)**:

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

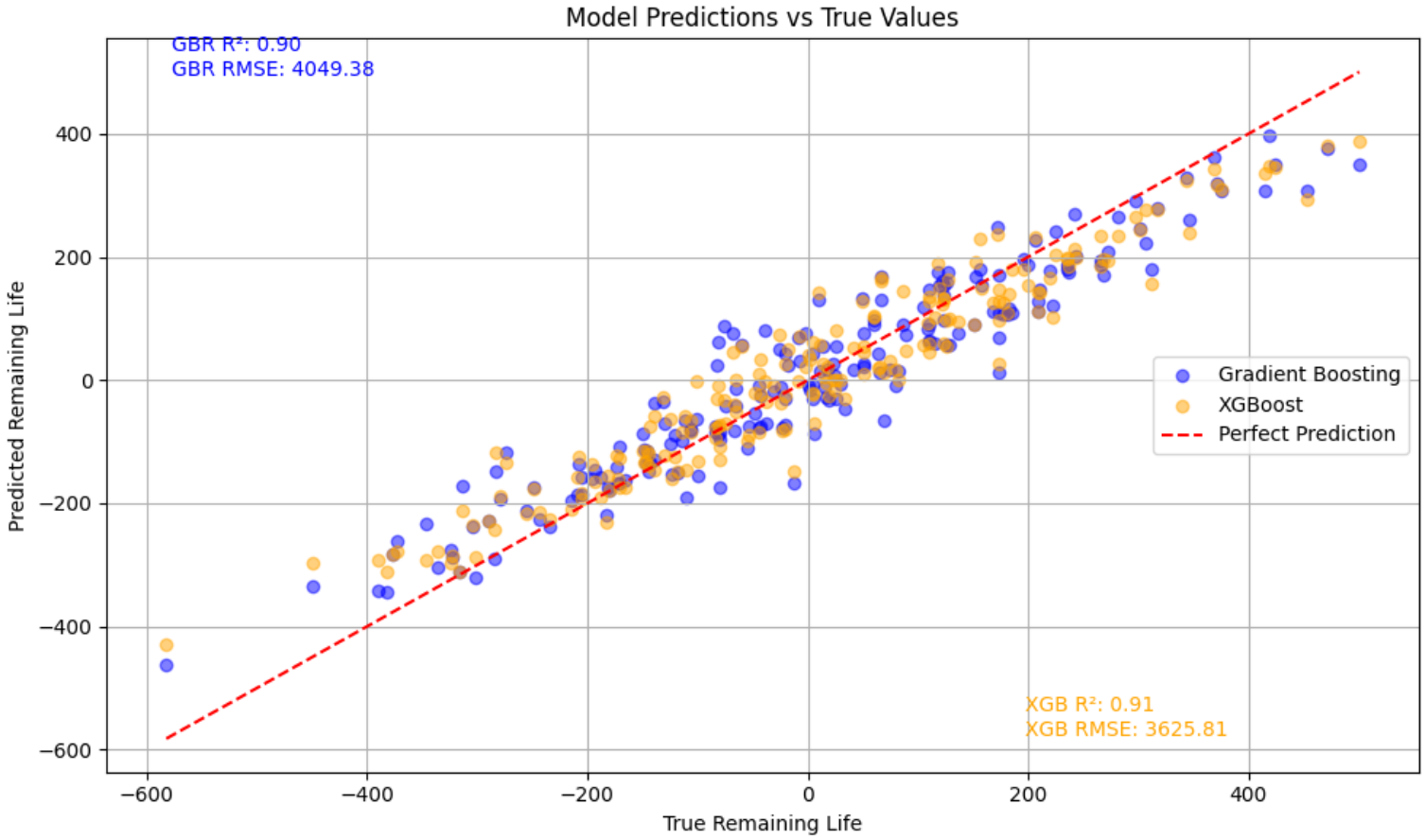
Indicates the proportion of variance explained by the model (1 is perfect).

Performance Comparison

Model	**MAE**	**RMSE**	**R^2**
Gradient Boosting	–	3541.33	0.99
XGBoost	45.37	3407.65	–

Table: Comparison of regression models on test set

Predictions vs True Values



Closer points to the red dashed line indicate more accurate predictions.

Interpretation

- **Gradient Boosting Regressor**:
  - Typically performs well on smaller datasets
  - Easy to interpret and tune
- **XGBoost Regressor**:
  - Often achieves higher accuracy, especially on large and complex datasets
  - Efficient, handles regularization and missing values well
- **Overall:** XGBoost is generally preferred for its speed and performance, though both methods can perform similarly depending on tuning and data.