**Analysis of Regression Models Performance**

The models have been evaluated on a dataset, and the results show how well each model performs in predicting the target variable (likely the price

of used cars) based on the selected features (year, odometer, drive\_fwd, fuel\_diesel, cylinders\_8). The key metrics evaluated are Mean Absolute Error

(MAE) and R-Squared (R²).

**1. Linear Regression**

Mean Absolute Error (MAE): 7012.50

R-Squared (R²): 0.3190

Interpretation:

Linear Regression captures about 31.9% of the variance in the data, meaning it's not very effective at explaining the relationship between the

features and the target variable. The high MAE indicates that the predictions are, on average, off by approximately $7,012.50, which is quite

significant and suggests that a linear relationship may not be the best fit for this data.

**2. Ridge Regression**

Mean Absolute Error (MAE): 7012.64

R-Squared (R²): 0.3190

Interpretation:

Ridge Regression, which adds a penalty to the size of coefficients, performs almost identically to Linear Regression. This suggests that

regularization (which Ridge provides) does not significantly improve model performance, indicating that the model might not be overfitting and that the linear assumption remains insufficient.

**3. Lasso Regression**

Mean Absolute Error (MAE): 7012.81

R-Squared (R²): 0.3190

Interpretation:

Lasso Regression, which can shrink some coefficients to zero (effectively performing feature selection), also provides similar results. This implies that no single feature dominates the predictions, and the linear model's limitations are likely due to the linear assumption rather than overfitting or irrelevant features.

**4. Decision Tree**

Mean Absolute Error (MAE): 3661.09

R-Squared (R²): 0.6410

Interpretation:

The Decision Tree model significantly outperforms the linear models, capturing about 64.1% of the variance in the data. The lower MAE of approximately $3,661.09 indicates that the model is much better at making accurate predictions. This improvement suggests that the relationships between the features and the target variable are likely non-linear, which the Decision Tree can capture.

**5. Random Forest**

Mean Absolute Error (MAE): 3577.77

R-Squared (R²): 0.7323

Interpretation:

The Random Forest model performs the best among all the models, capturing 73.23% of the variance and achieving the lowest MAE of approximately $3,577.77. Random Forest, which is an ensemble of multiple Decision Trees, further reduces errors by averaging out the predictions, making it the most reliable model in this scenario.

**Key Takeaways:**

Non-Linear Relationships: The significant improvement in performance from Linear Regression models to Decision Trees and Random Forests indicates that the relationship between the features and the target variable is non-linear.

**Model Suitability:**

Linear Models: Linear, Ridge, and Lasso Regression are not well-suited for this data, as they fail to capture the complexity of the relationships. Tree-Based Models: Decision Trees and, more effectively, Random Forests are much better suited for this dataset, as they capture non-linear interactions and provide better predictive accuracy.

**Recommendation:**

Given the results, Random Forest should be the preferred model for making predictions on this dataset. It offers the best balance between accuracy (as measured by R-Squared) and prediction error (as measured by MAE). Further optimization of the Random Forest model, such as tuning hyperparameters, might yield even better results.

**Insights for the Client:**

The Random Forest model indicates that certain non-linear factors (possibly interactions between features) are crucial in predicting the target variable. The dealership can trust the Random Forest model more for pricing predictions, but they should be aware that simple linear models will likely provide less accurate estimates due to the complexity of the relationships in the data.

