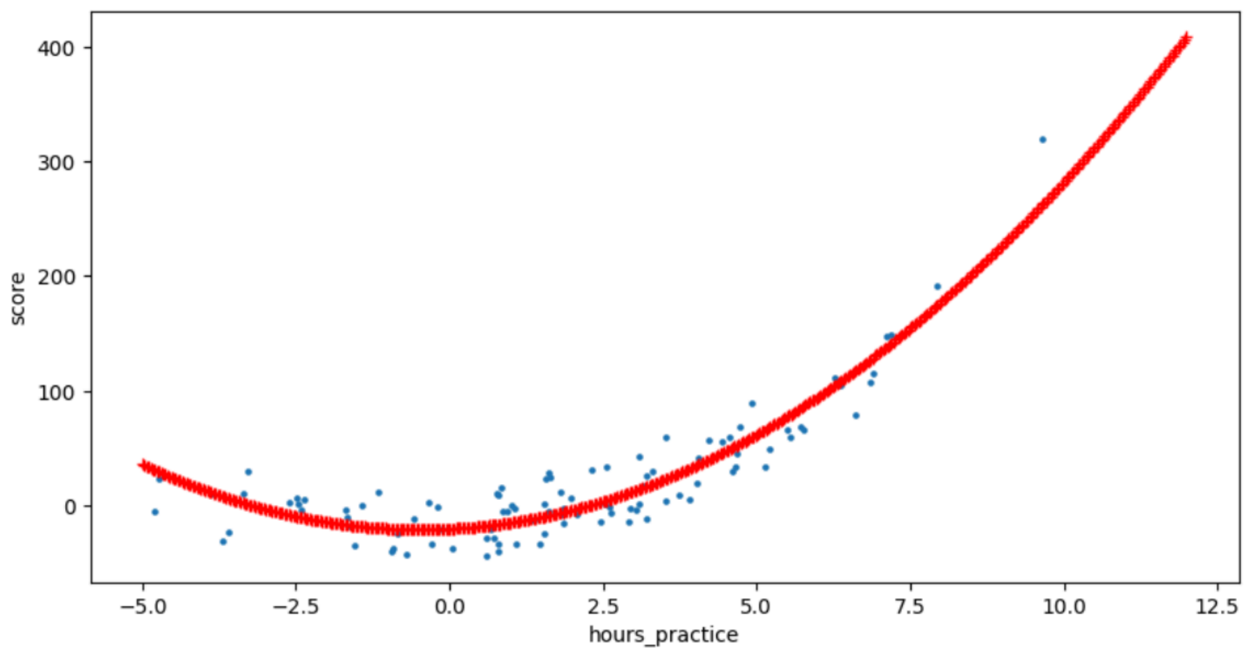
1.3 Linear Model

The shape of the curve is non-linear. The model trained is still considered a linear model because the underlying model identifies linear relationships between the different features and uses these linear relationships to predict outputs. By using feature transformations, we are able to add a new feature to train the linear model on. We are able to train a linear model on the relationship between x and x squared. Since the two features are related to each other via a polynomial function, the resulting model shows a non-linear curve, but the model has learned a linear relationship between the provided features.



2.2 Regularization Effects

1. Regularization greatly improves the model's test performance because it greatly reduces the model's ability to overfit to the training data. The non-regularized model does have a higher training accuracy compared to the model with L1 or L2 regularization, but that is likely due to the model overfitting to the training data (with a very high training score but significantly lower testing score).

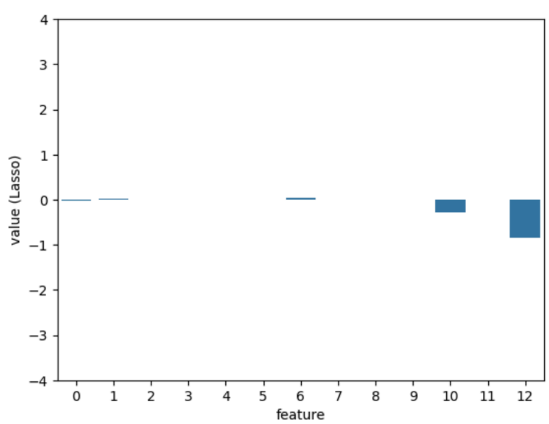
No Regularization Results:

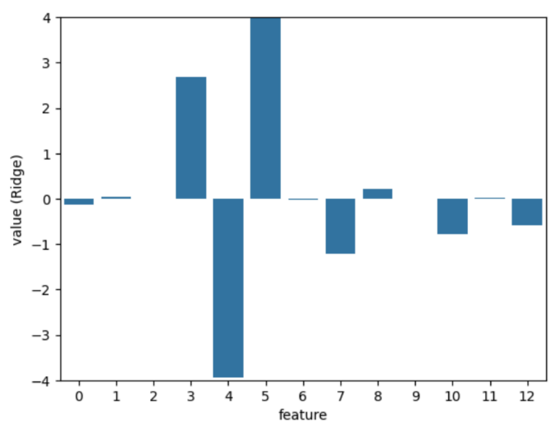
```
Training score: 0.9439354892529316
Test score: 0.6250613759163508
```

With Regularization Results:

```
regularization: L1, train_score: 0.9070657101514069, test_score: 0.8055776105496002
regularization: L2, train_score: 0.9304830234444311, test_score: 0.8079087119535296
```

2. The coefficients of the model with Lasso Regularization are significantly lower than the coefficients of the model with Ridge Regularization. This is because Lasso Regularization can perform feature selection by driving the coefficients of some features to exactly zero. This enables Lasso Regularization to ignore certain unimportant features and can scale the more important features a little bit. Ridge Regularization does not have the same ability to select features, so the coefficients are much larger to enlarge the impact of the most important features.





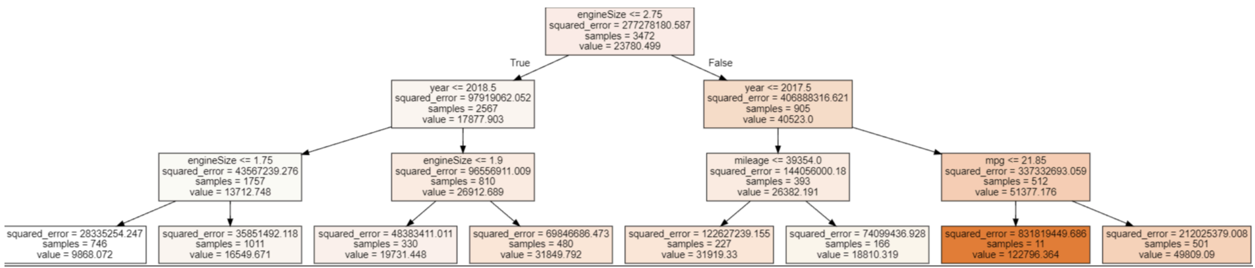
3. Alpha changes the strength of the regularization term on the model. When the regularization term is weaker (alpha is lower), the model coefficients are much larger since the regularization term hasn't been able to penalize the learned coefficients as much. Since the regularization is weaker, the model overfits to the training data more leading to a much worse testing accuracy. When the regularization term is stronger (alpha is higher), the model coefficients are much smaller since the regularization term is able to penalize the learned coefficients a lot more. A strong regularization term also helps models avoid overfitting a lot more which improves the test accuracy greatly.

3.4 Explanation for prediction

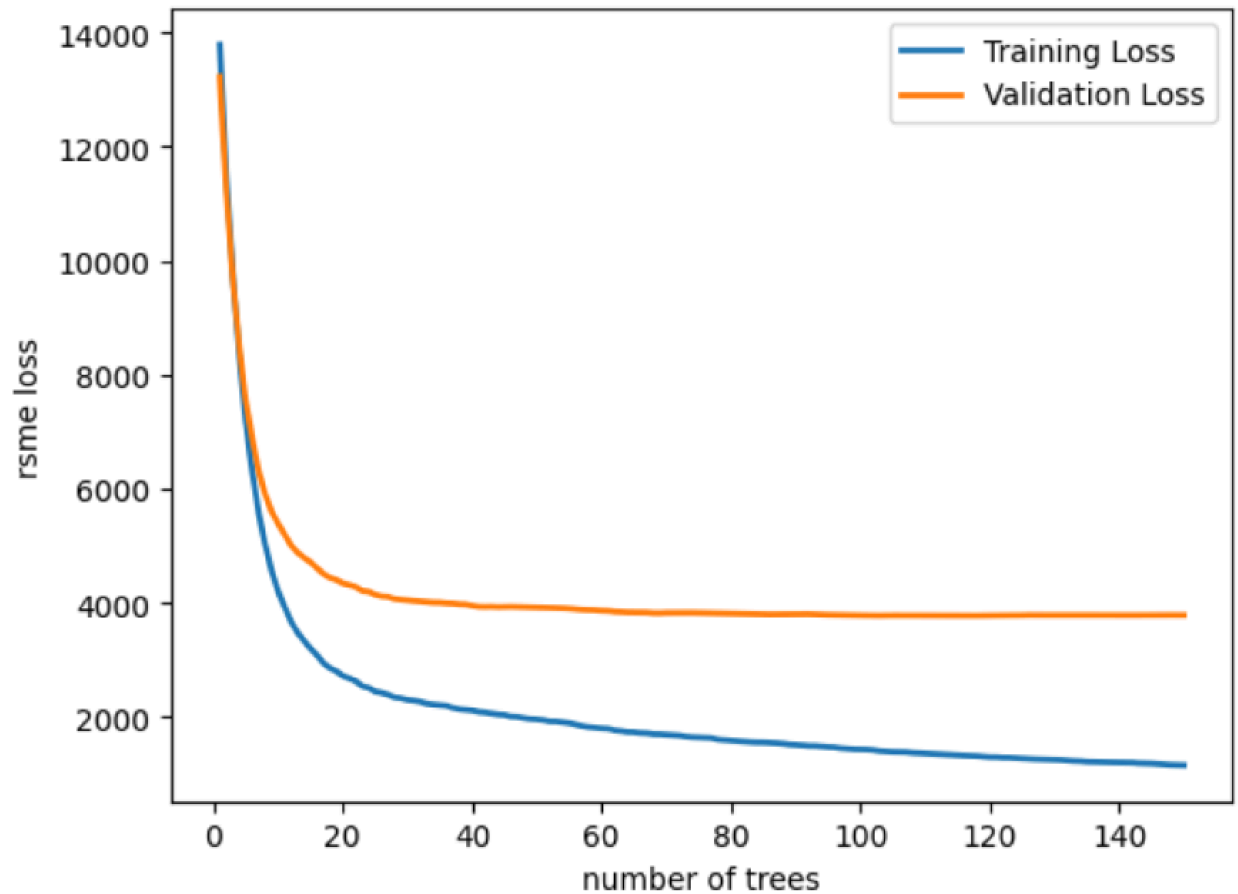
The decision tree is able to reach the prediction for the car price by following this path down the decision tree:

1. engine size = 2.0 <= 2.75 (true, proceed down the left path)
2. year = 2019 <= 2018.5 (false, proceed down the bottom/right path)
3. engine size = 2.0 <= 1.9 (false, proceed down right path)

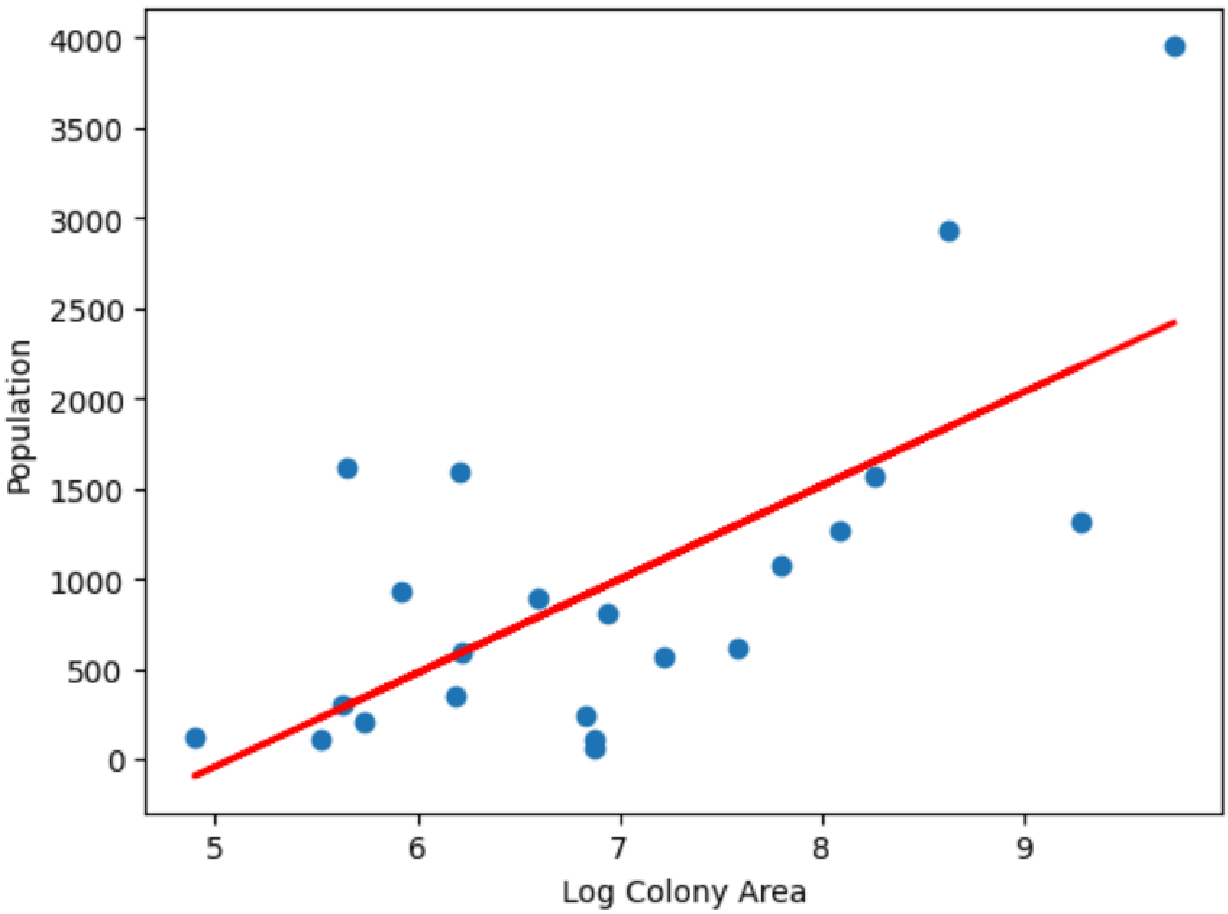
This traversal results in the expected cost for the car to be £31,849.79.



4.2 Plot train and test errors



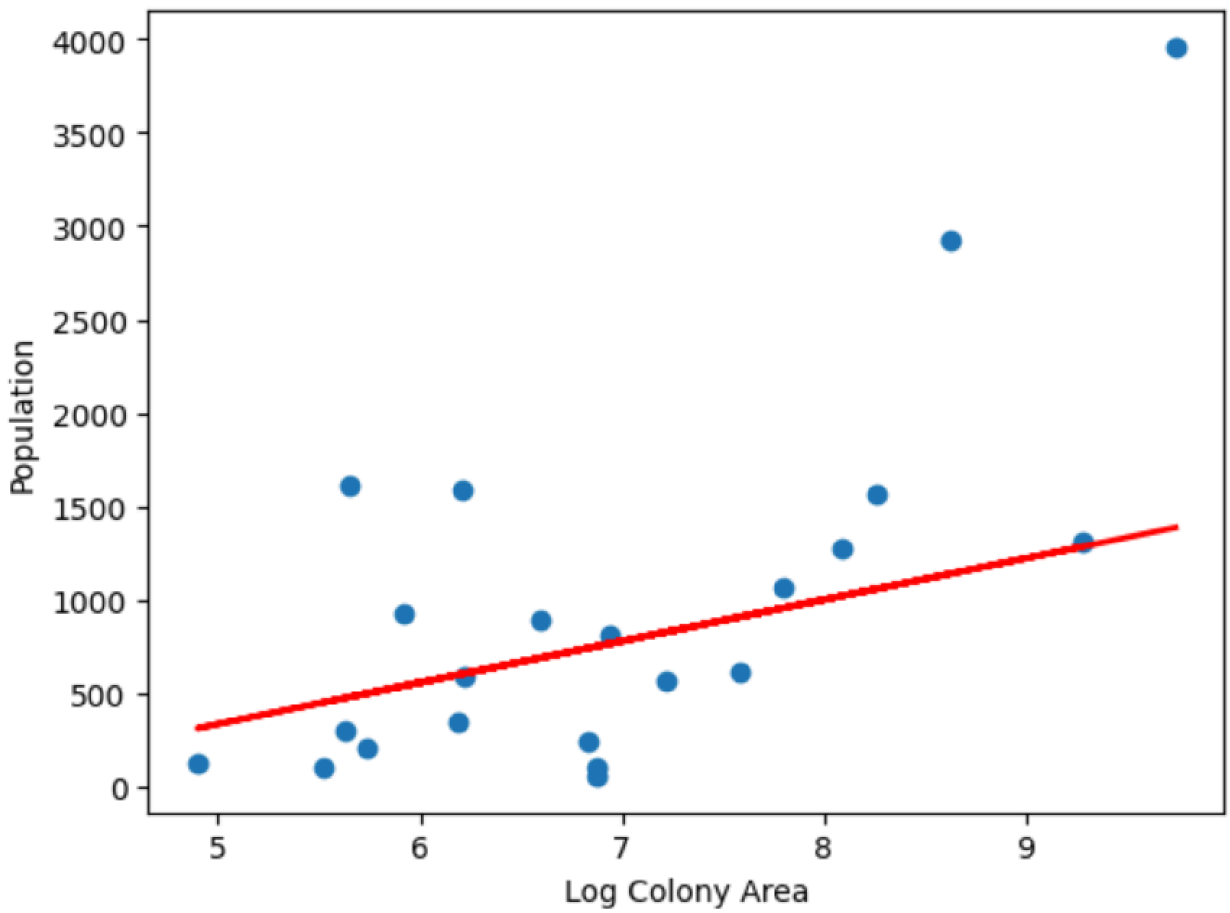
5.1 Fitting regression model



5.2 Performance Analysis

Based on the result from 5.1 (shown below), my regression model explains the data relatively well, especially in smaller log colony areas. If I had a large island, I would not trust the result of this regression as much because there is less data on larger islands, making it more sensitive to outliers. If I had a small island, I would trust the output of this regression model much more because there is significantly more data on smaller islands.

5.3 Plotting regression without outliers



5.4 Comparing Linear Regression Models

After training a linear regression model on the dataset after the two main outliers (the ones with populations much higher than the other islands observed) have been removed, I would trust the predictions of this regression model a little more than the previous one because it seems to be a strong fit for the islands that were included and excludes two clear outliers from the training data. However, I still wouldn’t trust the model very strongly for large islands because of how little data there is on larger islands (the observed outliers may not actually be outliers) and because there may be a non-linear relationship between island log area and population size.