1.) Statistics Analysis and Data Exploration
Number of data points: 506
Number of features: 14
Minimum and Maximum house prices: 5.0, 50.0
Mean and Median house prices: 22.532806324110698, 21.199999999999999
Standard Deviation: 9.1880115452782061

2.) Evaluating Model Performance
The **mean squared error** is an appropriate measure for measuring the performance of the model selected. It is appropriate because of the following

- It uses all data to check for error.
- It is not affected by outliers
- The values can be seen across a single plain

Other rubrics for measuring model performance are
1. Mean Absolute Error
2. The Interquartile range
3. The Range
4. Variance

It is important to split the Boston housing data in order to get an idea of how well the model is able to predict data it has not been trained on.

Grid Search is a technique optimize parameters for an estimator. The goal of grid search is to find the best parameters that minimize errors across all the data specified.

Cross Validation is a technique used to test how well a model performs on a given set of data. It involves splitting a dataset in to n folds then iteratively using each of the folds as test data. Cross Validation can be combined with grid search to test the selected parameters in grid search.

3.) Analyzing Model Performance
After observing the learning curve graphs, the general trend observed is that as training size increases, the distance between training and testing errors also increases.

When the model is fully trained, it suffers from high variance / overfitting. That explains the low training error and high testing error.

The model with Max Depth = 2 best generalizes the dataset, Because at that time, both Training and testing errors were the closest.

4.) Model Prediction
The most reasonable price reported by grid search is 21.4
According to the earlier statistic, this appears to be a reasonable model since the predicted value falls within the standard deviation. However, We are not entirely sure if the model is biased to pick values closer to the mean.