Some statistics calculated using NumPy:

```
        Size of data set is  506
        Number of features  13
        Minimum price  5.0
        Maximum price  50.0
        Mean price is  22.5328063241
        Median is  21.2
        Standard deviation is  9.18801154528
```

1) Of the available features for a given home, choose three you feel are significant and give a brief description for each of what they measure.

   a. CRIM: Crime rate in the particular town (per capita)
   b. NOX: Atmospheric Toxicity (Chemical: Nitric Oxide concentration)
   c. RM: Average number of rooms per dwelling

   I think these should be the most important factors (in decreasing order of importance) that any person would look at first if they wish to buy a house in a city. The value of CRIM and NOX should be the least and at the same time RM should be as high as possible.

2) Using your client's feature set in the template code, which values correspond to the chosen features?

   CRIM = 11.95
   NOX = 0.659
   RM = 5.609

3) Why do we split the data into training and testing subsets?

   The predictor (algorithm) first needs to learn the trends in the given dataset to make a model for making prediction. Then based on what patterns have been recognised, its knowledge needs to be tested. Thus the dataset is divided into 2 parts, namely training set (to learn patterns) and testing set (to test its knowledge). The results of applying the model on the test set shows how the model would perform in a real environment in the future.

4) Which performance metric below is most appropriate for predicting housing prices and analyzing error? Why?

   Accuracy, Precision, Recall and F1 score are not appropriate as they are used as metrics for classification models. Mean Squared Error (MSE) and Mean Absolute Error (MAE) are regression metrics and any one can be used.
   I preferred MSE as it is more efficient than MAE in improving model performance.

5) What is the grid search algorithm and when is it applicable?

   Grid Search Algorithm is used to fine tune the model and 'search' for the best parameters that help to generalise the predicting model. Generally, it is applicable when we have multiple ways to reach a goal and the optimal one needs to be chosen. In Boston Housing problem, grid search was used in the decision tree regressor to find the best max depth which helps to generalise the data.

6) What is cross-validation and how is it performed on a model? Why would cross-validation be helpful when using grid search?

Cross validation is a process of shuffling the entire dataset to create new training and testing sets.

When we divide the entire dataset into the 2 parts, the values of the testing set are never used for training and vice-versa. This would reduce the efficiency of the model as it would not get to learn from the testing set or be able to test the training set. To prevent this, the dataset is shuffled multiple times and the training and test sets are reassigned each time. This way, almost always the data once used in testing occurs in training and the training data is also used in testing in future iterations. This reduces error, making the algorithm more accurate.

Analyzing Model Performance

7) Choose one of the learning curve graphs your code creates. What is the max depth for the model? As the size of the training set increases, what happens to the training error? Describe what happens to the testing error.

At max depth 8, as the size of the training set is increasing, the training error is increasing very slightly. The testing error reduces greatly when the training size increases from 0 to 30. The testing error is minimum when the entire training set is used to make predicitons.

8) Look at the learning curve graphs for the model with a max depth of 1 and a max depth of 10. When the model is using the full training set, does it suffer from high bias or high variance when the max depth is 1? What about when the max depth is 10?

When depth is 1, the model suffers from high bias because even when the full training set is being utilised, the error in training and testing remains almost unaffected. The model is too simple and the data is said to be underfit.

As the training size increases, the model suffers from high variance when depth is 10 because at maximum training size, the test and train error varies a lot more than at max depth as 1. Thus at depth as 10, the model is becoming increasingly complicated; model is unable to predict more accurate values for testing data.

9) From the model complexity graph, describe the training and testing errors as the max depth increases. Based on your interpretation of the graph, which max depth results in a model that best generalizes the dataset? Why?

Using the model complexity graph, as max depth increases from 0 to 5, there is a sharp reduction in the testing and training error. At max depth 15 the training error is becoming 0 but the variance in the testing error is high. The testing error does not continue to reduce after max depth 4. For the number of parameters considered at this stage, the model requires more data. The inability of the model to make better predictions after max depth 4 is due to overfitting (model is getting too complicated).

Analysing the graph, max depth 10 seems to best generalise the dataset because at this point the testing and training error seems to be minimum in the graph.

Model Prediction

10) Using grid search, what is the optimal max depth for your model? How does this result compare to your initial intuition?

> The grid search algorithm shows the optimal max depth to be 6 for this model. This result is not much different from my initial intuition as I expected the optimal depth to be 10.

11) With your parameter-tuned model, what is the best selling price for your client's home? How does this selling price compare to the statistics you calculated on the dataset?

> The predicted selling price for the client is 20.77. This price is approximately 2 points below average price of a house being sold in Boston, which is well within the standard deviation so it seems good.

12) In a few sentences, discuss whether you would use this model or not to predict the selling price of future clients' homes in the Boston area.

> I would use this model with max depth as 6 to predict selling prices in Boston, as the error in testing set is still low and the data is being generalised very well.