Q1→*Of the features available for each data point, choose three that you feel are significant and give a brief description for each of what they measure.*

Answer: CRIM(value:11.95): per capita crime rate by town -->>Because more the crime,less livable is the neighbourhod hence the house prices will be lower.INDUS(value:18.1): proportion of non-retail business acres per town-->>More INDUS,more developed the neighbourhood,more the house price,consequently. LSTAT(value:12.13):% lower status of the population-->>More this %,more less income people live in the area poiting to the fact that the area is cheaper enough where the low income people can afford to buy house.

Q2→ *Using your client's feature set `CLIENT_FEATURES`, which values correspond with the features you've chosen above?*

Answer: CRIM(value:11.95): per capita crime rate by town INDUS(value:18.1): proportion of non-retail business acres per town LSTAT(value:12.13):% lower status of the population

Q3→Why do we split the data into training and testing subsets for our model?
Answer: The train partition helps build our model and the test partition helps to provide us honest assessments of the performance of our predictive models.Hence to avoid overfitting,have possibly optimum performance of our model and simulate real life data that our model may encounter.

Q4→ *Which performance metric below did you find was most appropriate for predicting housing prices and analyzing the total error. Why?*

- *Accuracy*
- *Precision*
- *Recall*
- *F1 Score*
- *Mean Squared Error (MSE)*
- *Mean Absolute Error (MAE)*

Answer: In many circumstances it makes sense to give more weight to points further away from the mean--that is, being off by 10 is more than twice as bad as being off by 5. In such cases MSE is a more appropriate measure of error and hence I chose MSE.Also only MSE and MAE can be used for our model as only they support continuos multi output data.The rest of the metrics that are:-accuracy,precision,recall and F1 score are not suitable for continuos multioutput data

Q5→ *What is the grid search algorithm and when is it applicable?*
Answer: In the context of machine learning, hyperparameter optimization or model selection is the problem of choosing a set of hyperparameters for a learning algorithm, usually with the goal of optimizing a measure of the algorithm's performance on an independent data set.The

traditional way of performing hyperparameter optimization has been grid search, or a parameter sweep, which is simply an exhaustive searching through a manually specified subset of the hyperparameter space of a learning algorithm.
It unfortunately suffers from the curse of dimesionality so,it can only be used when we have not very high dimensional data. With a gridsearch you just try a set of values for your parameters and look at which value the objective function is largest (or smallest). Lets say you have one parameter and want to try for the values $1,2,3,4,5,6,7,8,9,10$. (this is not realistic, normally you try many more possible values). In this example you need to compute the objective function 10 times.

Lets say there is a second parameter. Now you need to try the values:
$(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(1,7),(1,8),(1,9),(1,10),(2,1),(2,2),(2,3),(2,4),(2,5),(2,6),(2,7),(2,8),(2,9),(2,10),\cdots,(10,1),(10,2),(10,3),(10,4),(10,5),(10,6),(10,7),(10,8),(10,9),(10,10)$.
So now you need to compute the objective function $10\times10=100$ times. If you have a third parameter you would end up with $10^3=1000$ evaluations, etc.
If we use a more realist number of tries per parameter, say 1000, then we end up with a $1000^1=1000$ evaluations for one parameter, $1000^2=1,000,000$ evaluations for two parameters, $1000^3=1,000,000,000$ evaluations for three parameters, etc. You can see that you can quickly end up with an unmanageable number of evaluations.

Q6→ *What is cross-validation, and how is it performed on a model? Why would cross-validation be helpful when using grid search?*
Answer: Cross-validation is a model validation technique for assessing how the results of a statistical analysis will generalize to an independent data set.

It is performed as given->>One round of cross-validation involves partitioning a sample of data into complementary subsets, performing the analysis on one subset (called the training set), and validating the analysis on the other subset (called the validation set or testing set)-->>"validating the analysis on the other subset" means that the model is valid if it has error less then a certain chosen threshold and thus when it is below that threshold,it means that it is able to predict within an acceptable margin of errors. To reduce variability, multiple rounds of cross-validation are performed using different partitions, and the validation results are averaged over the rounds.

Cross-validation helps to estimate the generalization performance of a model built using the grid search. Cross-validation is when you reserve part of your data to use in evaluating your model. There are different cross-validation methods. The simplest conceptually is to just take 70% (just making up a number here, it doesn't have to be 70%) of your data and use that for training, and then use the remaining 30% of the data to evaluate the model's performance. The reason you need different data for training and evaluating the model is to protect against overfitting. There are other (slightly more involved) cross-validation techniques, of course, like k-fold cross-validation, which often used in practice.

Grid search means you have a set of models (which differ from each other in their parameter values, which lie on a grid). What you do is you then train each of the models and evaluate it using cross-validation. You then select the one that performed best.

To give a concrete example, if you're using a support vector machine, you could use different values for gamma and C. So, for example you could have a grid with the following values for (gamma, C): (1, 1), (0.1, 1), (1, 10), (0.1, 10). It's a grid because it's like a product of [1, 0.1] for gamma and [1, 10] for C. Grid-search would basically train a SVM for each of these four pair of (gamma, C) values, then evaluate it using cross-validation, and select the one that did best.

Q7→ *Choose one of the learning curve graphs that are created above. What is the max depth for the chosen model? As the size of the training set increases, what happens to the training error? What happens to the testing error?*
Answer: Max Depth is 10. As the size of the training set increases, the training error decreases non linearly,although initially its very low. The testing error,also decreases and increases in a constant pattern,one after the another with the increase in data points i the training set.When the training set is small, the trained model can essentially “memorize” all of the training data. As the training set gets larger, the model won’t be able to fit all of the training data exactly. The opposite is happening with the test set. When the training set is small, then it’s more likely the model hasn’t seen similar data before. As the training set gets larger, it becomes more likely that the model has seen similar data before.

Q8→ *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?*
Answer: Model with max-depth 1 suffers from high bias as whereas the model with max-depth 10 suffers from high variance.The reason for this phenomenon is that a model with max depth 1 is a simpler model that don't tend to overfit, but may underfit their training data, failing to capture important regularities.It oversimplifies any existing correlation between atrributes and the target variable.However,a model with max-depth 10 is more complex enabling it to represent the training set more accurately. In the process, however, it may also represent a large noise component in the training set, making their predictions less accurate - despite their added complexity and as a consequence suffering from overfitting.

Q9→ *From the model complexity graph above, 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?*
Answer: With the increase in max depth,training error decreases because a more complex model is much more able to repreent the relationships among the attributes thus developing a model that better reprents the data.However as the model complexity increases beyond a certain point it starts overfitting.While the test error increases and decreases in a patterns when the training set is small, then it’s more likely the model hasn’t seen similar data before. As the training set gets larger, it becomes more likely that the model has seen similar data before.Based on my interpretation,I think the model with max depth equal to 4 OR 5 generalizes the best as it has a low training error as well a low test error.This trade off between training and test error seems to be best for the model with this max depth.

Q10→ *Using grid search on the entire dataset, what is the optimal `max_depth` parameter for your model? How does this result compare to your intial intuition?*
Answer: Final model has an optimal max_depth parameter of 4 as 13 is highly overfitting here, as the test error is well above our training error and we are basically just memorizing our data.

Q11→ *With your parameter-tuned model, what is the best selling price for your client's home? How does this selling price compare to the basic statistics you calculated on the dataset?*
Answer: $21630.This seems to be good enough as it is very close to the median and mean house price calculated earlier.

Q12→*In a few sentences, discuss whether you would use this model or not to predict the selling price of future clients' homes in the Greater Boston area.*
Answer: Though the accuracy is good,I would not use this model.I think it needs more helpful features like median household income,some metric to assess nearby facilities on their qauntity,quaility and availability,distance to nearest hospitals,entertainment venues such as movie theaters etc.