


1. **Question 1: How many variables are in this data set?**
 - a. 11 variables
2. **Question 2: Are there any variables that are categorical? If so, write down the names.**
 - a. Yes, the categorical variables are shelveLoc, us, urban
3. **Question 3: How many nodes are produced?**
 - a. 27 Nodes
4. **Question 4: What is the training error rate?**
 - a. 9%
5. **Question 5: What is the variable of the first branch? How is that branch split?**
 - a. First name is ShelveLoc: Bad, Medium
 - b. The bad or medium go to the left and the good and high go to the right
6. **Question 6: What is the test error rate?**

	High.test	
tree.pred	No	Yes
No	104	33
Yes	13	50

- a. 
- b. $46/200 = 0.23$
7. **Question 7: What is the lowest cross-validation error, dev**
 - a. 74
8. **Question 8: What is the number of terminal nodes, size that corresponds to the lowest cross-validation error?**
 - a. 8
9. **Question 9: What is the test error rate for the pruned tree?**
 - a. 24.5
10. **Question 10: What variables were used to construct this tree?**
 - a. "rm" "lstat" "crim" "age"
11. **Question 11: How many nodes are used to construct this tree?**
 - a. 7
12. **Question 12: What is the predicted median house price for medium sized homes ($6.9595 \leq rm < 7.553$)?**
 - a. \$33,420
13. **Question 13: How many nodes would be best to use?**

- a. We should still use 7 Nodes

14.Question 14: What is the test set MSE associated with the regression tree?

- a. 35.28688