Assignment #6 Predictive Modeling 2 by Joshua Troup

Q1: In terms of number of decision nodes, compare the size of the full tree to the size of the bestpruned tree.

**Full tree (training data) has 970 decision nodes. Best pruned (validation data) has 29 decision nodes.**

Q2: What is the RMSE of the best-pruned tree on the validation data and on the test dataset?

**RMSE of the best-pruned tree on the validation data is 20401. RMSE of the best-pruned tree on the test data is 23169.**

Q3: What is the average error on the validation and the test dataset? What does this suggest?

**The average error on the validation is 384. The average error on the test data is 295. Validation is used to tune the parameters of a classifier. Test is used to assess the performance of a trained classifier. The test is a test of the validation which is a good indicator it is a good model with the average error being relatively low. These small error values in both validation and test suggest the single tree method has created an accurate predictor. Other predictive classifying models may be more or less accurate.**

Q4: Examine the best-pruned tree. What are the critical variables in predicting the price of a home?

**Building Value, Age, Poor Condition, Land Value, Above Space, Acres, and Deck.**

Q5: In terms of number of decision nodes, compare the size of the full tree to the size of the bestpruned tree.

**Full tree (training data) has 817 decision nodes. Best pruned tree (validation data) has 16 decision nodes.**

Q6: What is the RMSE of the best-pruned tree on the validation data and on the test dataset?

**RMSE of the best-pruned tree on the validation data is 22307. RMSE of the best-pruned tree on the test data is 22216.**

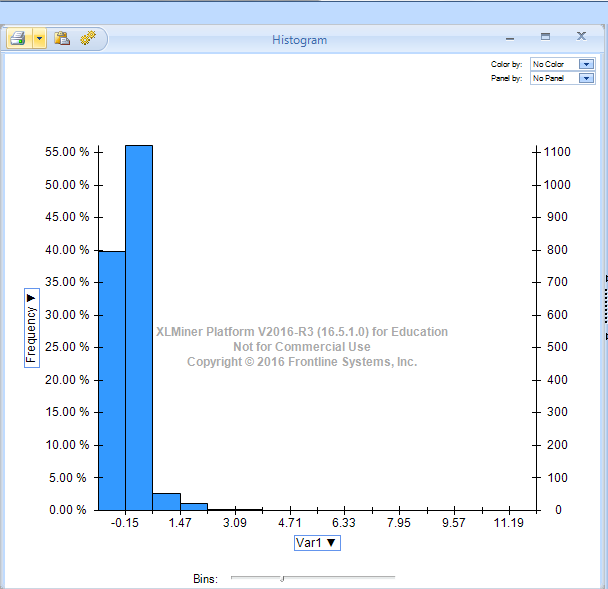
Q7: What is the average error on the validation and the test dataset? What does this suggest?

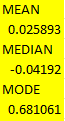
**The average error on the validation is -1017. The average error on the test data is -455. Validation is used to tune the parameters of a classifier. Test is used to assess the performance of a trained classifier. The test is a test of the validation which is a good indicator it is a good model with the average error being relatively low. These small error values in both validation and test suggest the single tree method has created an accurate predictor. Other predictive classifying models may be more or less accurate.**

Q8: Examine the best-pruned tree. What are the critical variables in predicting the price of a home?

**Building Value, Land Value, AC, and Above Space.**

Q9: What is the average percentage change in predicted price between the pre-crisis and post-crisis model? What does this suggest about the impact of the housing bubble?

****

****

**The housing bubble affected the market by a 2.5% average difference. The histogram shows the majority (55%) of the market was affected by the 2.5% difference.**

Q10. Compare the average % change in predicted price using k-NN and Decision Tree. Which is

a better model and why?

**The averages between k-NN and Decision Tree are quite similar. The mean (average) between them both are extremely close (2.4% vs. 2.5%). The Median and Mode were significant however were not as comparable in values as the mean. I believe both models demonstrate a good balance. k-NN may have a slight edge over decision tree as the average values (mean, median, mode) are clustered together better.**