Assignment #5 Predictive Modeling 1 by Joshua Troup

Q1: What value of k minimizes the RMSE on the validation data?

**K value=7. Training Error=1103. Validation Error=23389 which is the lowest value validation error based on K=1-20**

Q2: What is the RMSE on the validation data and test set? (Scoring)

**RMSE is 23389 for validation. RMSE is 27268 for test.**

Q3: What is the average error on the validation data and test set? What does this suggest? **Average Error is 3119 for validation. Average Error is 3987 for test. 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. 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.**

Q4: What value of k minimizes the RMSE on the validation data?

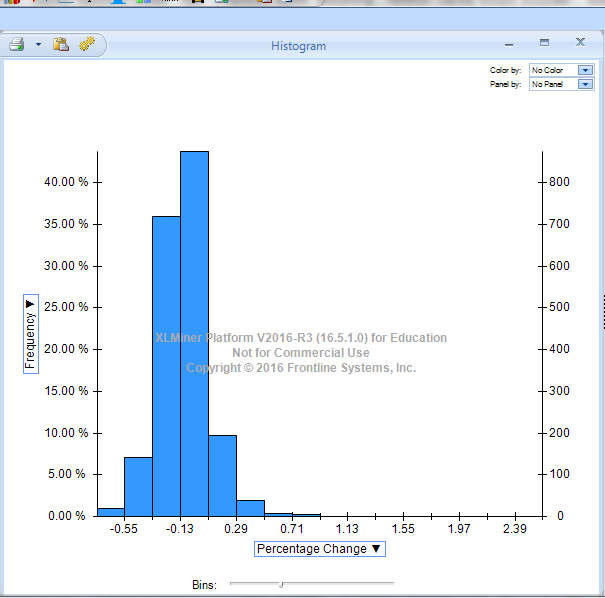
**K value=4. Training Error=0 Validation Error=24152 which is lowest validation value error based on K=1-20**

Q5: What is the RMSE on the validation data and test set? (Scoring)

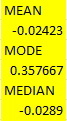
**RMSE is 24152 for validation. RMSE is 24395 for test.**

Q6: What is the average error on the validation data and test set? What does this suggest?

**Average Error is 1289 for validation. Average Error is 2831 for test. 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. 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.**

****

Q7: What is the average % change in predicted price between the pre-crisis and post-crisis model?

****

**The housing bubble affected the market by a 2.4% average difference. The histogram shows nearly half (45%) of the market was affected by the 2.4% difference.**