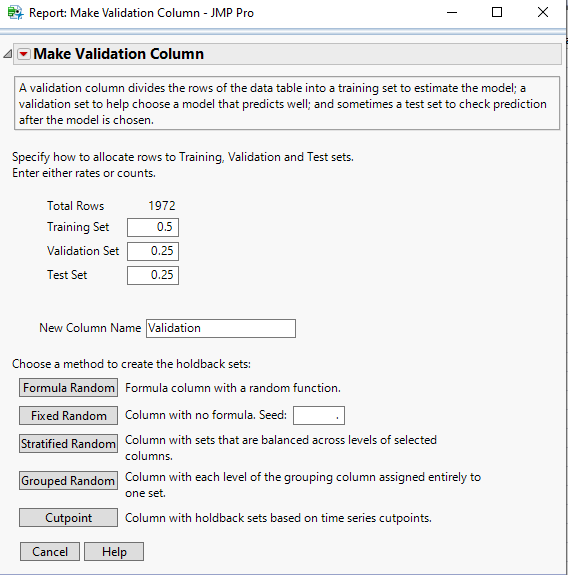
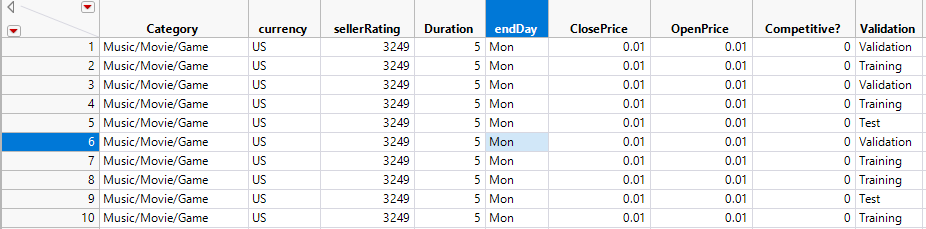
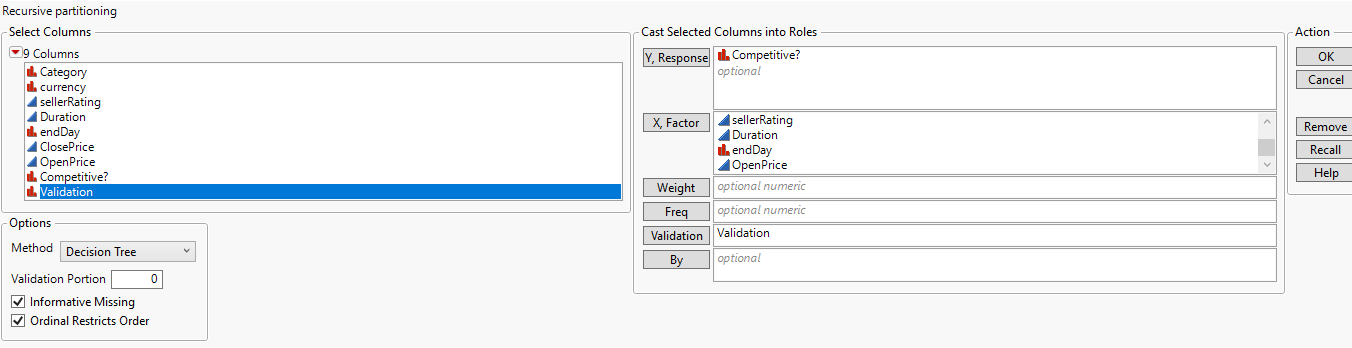
1. **Split the data into training (50%), validation (25%), and test (25%).  Include a screen shot of the first 10 rows of data.**

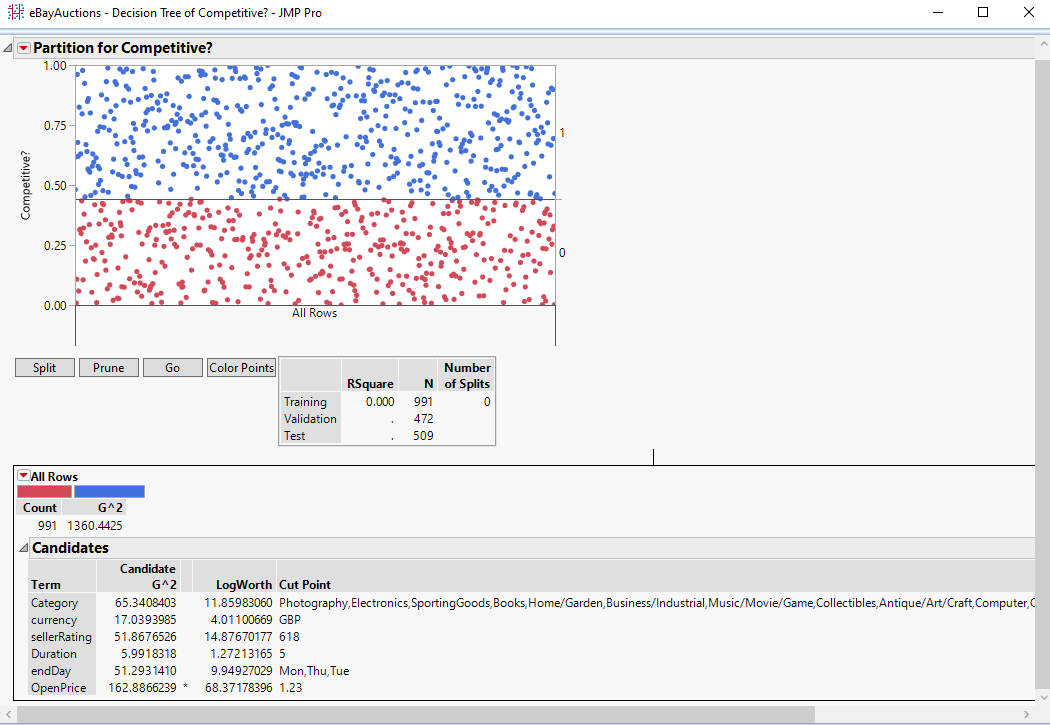




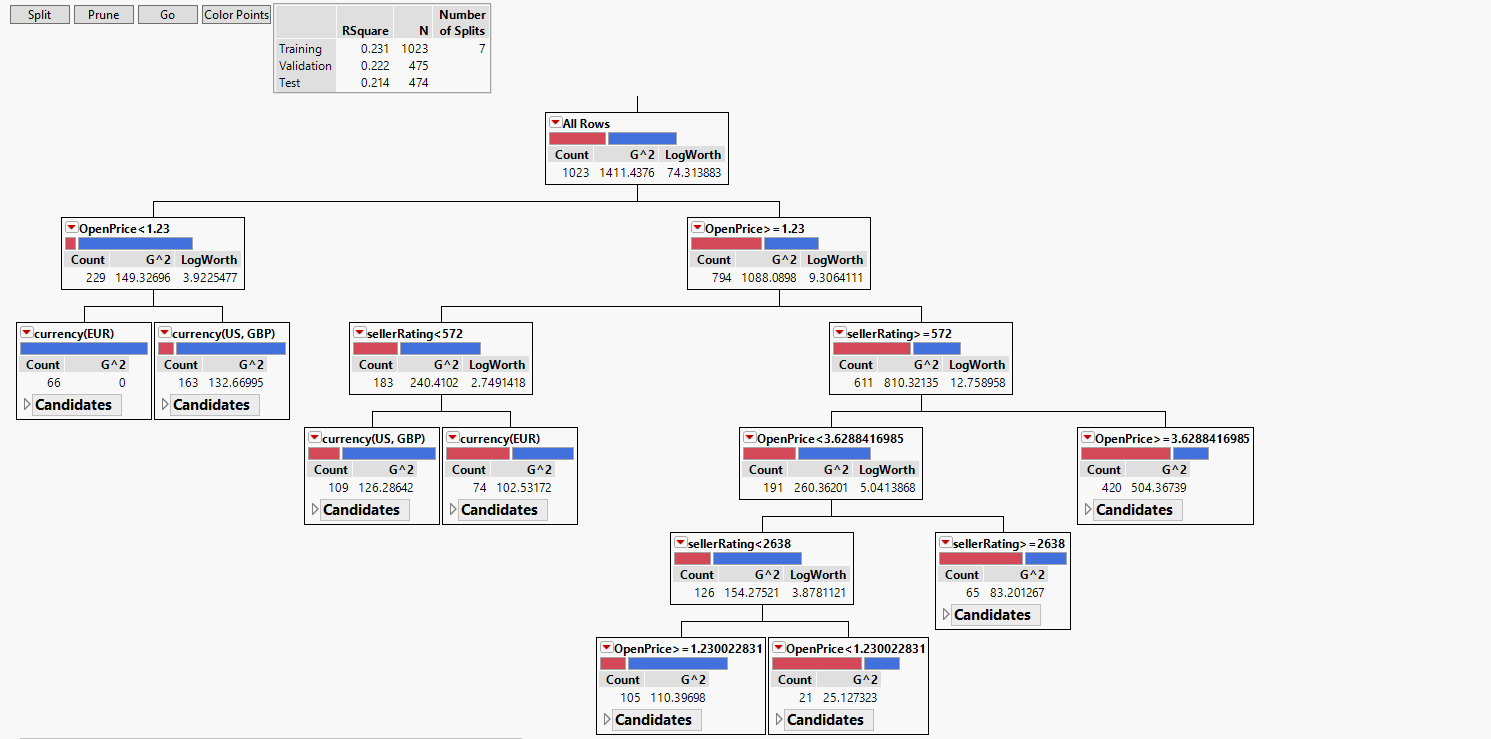
1. **Fit a classification tree using all the predictors except Close Price.  How many splits are optimal?  Include a screen shot of the Split History and the Leaf Report.**

**Analyze -> Predictive Modeling -> Partition**

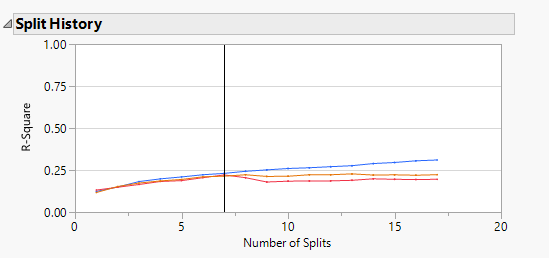




If the **logworth** is greater than 2, then the variable that is used in the branch is significant and should be included in the tree. From the partition we can see that, opening price is the main predictor to find out whether or not the auction will be competitive because of the high LogWorth value.

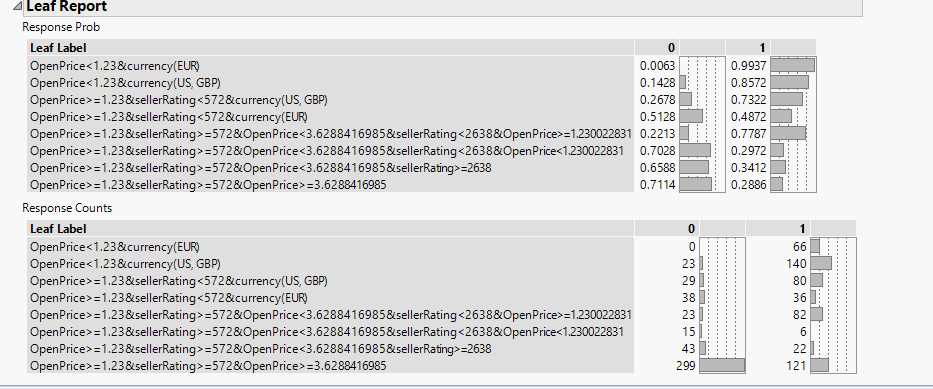


**Split History:**

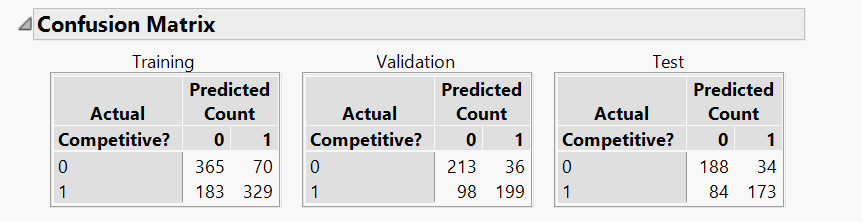


According to the Split History table, the optimal number of splits is 7.

**Leaf Report:**



1. **What is the total accuracy and the accuracy of a 1 prediction as measured by the test data for the tree created in question 2? Include a screen shot of the Confusion Matrix for all three partitions.**



**Accuracy of test data** = Total number of correct predictions/ Total predictions

= (188+173) / 479

= 361/479

= 75.3%

**Accuracy of 1’s in the test data** = Number of correct 1’s prediction/ Total no. of 1’s prediction

= 173 / (173+34)

= 173/207

= 83.5%

1. **Describe one benefit of the Bootstrap Forest and one benefit of the Boosted Tree.**

**Bootstrap Forest:**

The Bootstrap Forest platform fits an ensemble model by averaging many decision trees each of which is fit to a bootstrap sample of the training data. Each split in each tree considers a random subset of the predictors. In this way, many weak models are combined to produce a more powerful model. The final prediction for an observation is the average of the predicted values for that observation over all the decision trees.

Benefits of Bootstrap Forest:

* **Reduction in overfitting** by averaging several trees.
* **Less variance**. Averaging a set of observations reduces variance. We reduce the chance of stumbling across a classifier that doesn’t perform well because of the relationship between the training and test data.

**Boosted Tree:**

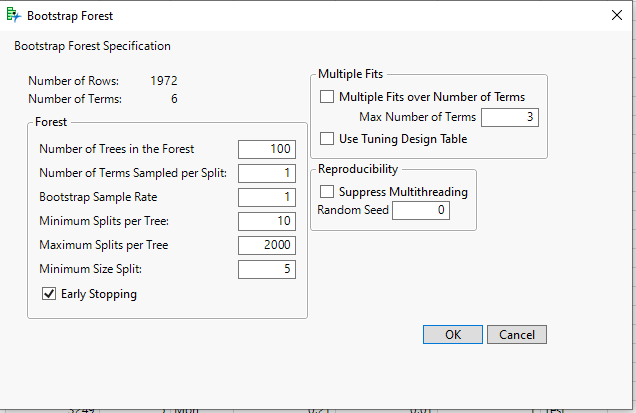
Boosting is the process of building a large, additive decision tree by fitting a sequence of smaller decision trees, called *layers*. The tree at each layer consists of a small number of splits. The tree is fit based on the residuals of the previous layers, which allows each layer to correct the fit for bad fitting data from the previous layers. The final prediction for an observation is the sum of the predictions for that observation over all of the layers.

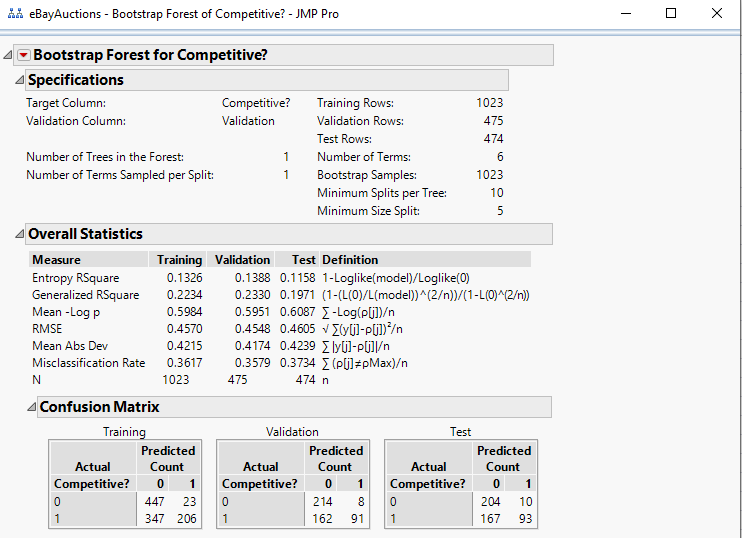
Benefit of Boosted Trees:

* **Less variance**. By using multiple trees, you reduce the chance of stumbling across a classifier that doesn’t perform well because of the relationship between the train and test data.

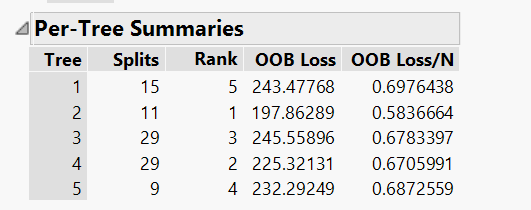
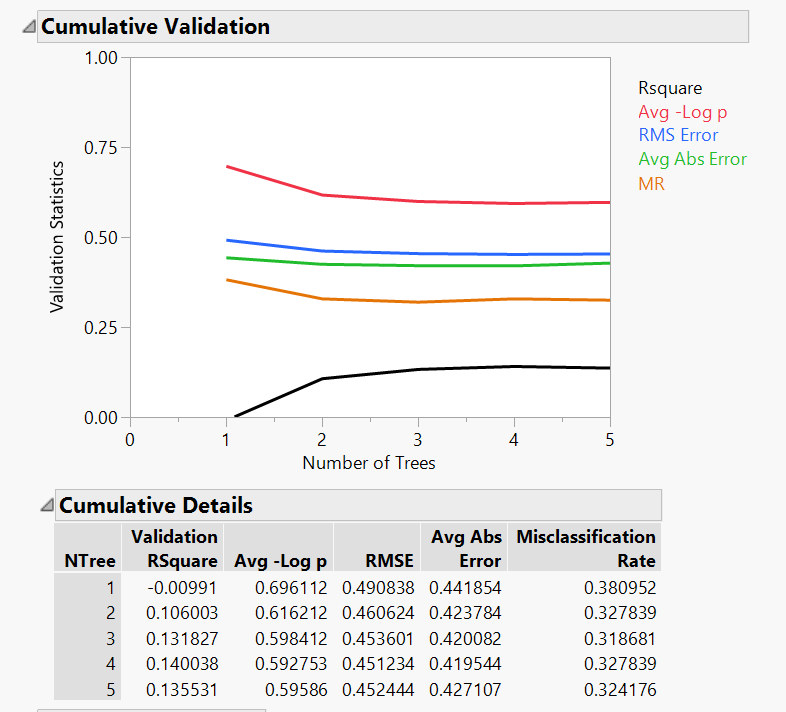
1. **Build either a Bootstrap Forest or Boosted Tree model to try to improve performance.  Include a screen shot of the Specifications and Overall Statistics.**

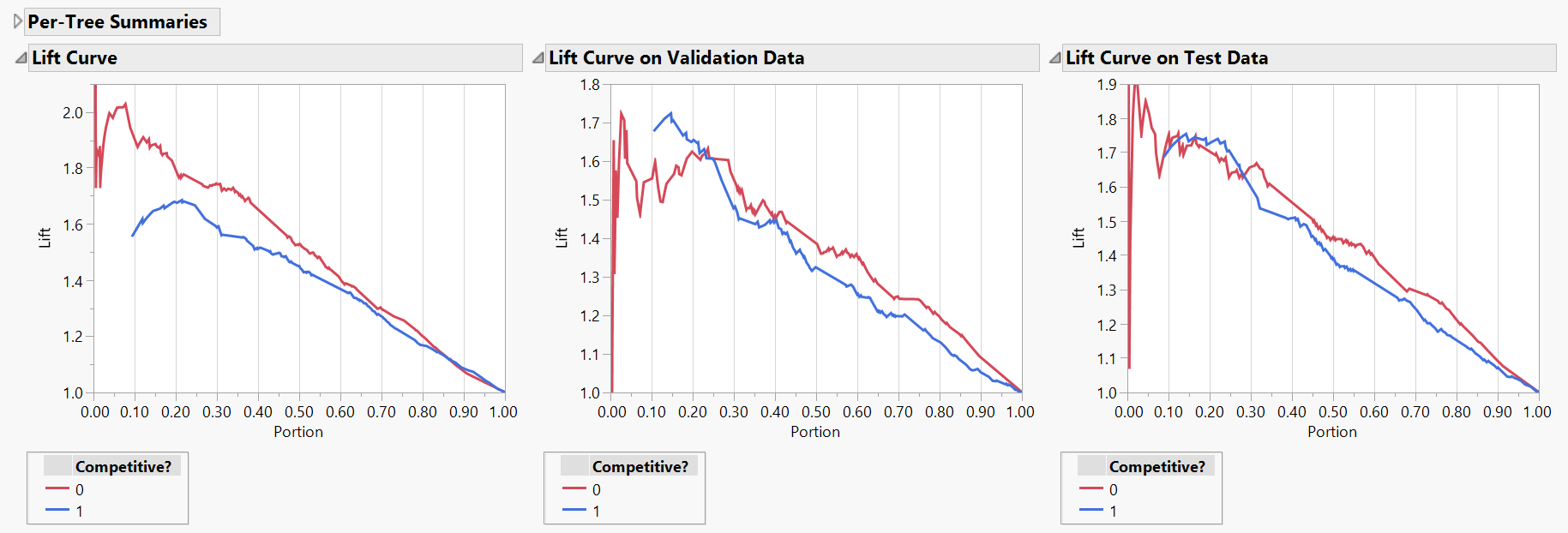
**Bootstrap Forest:**



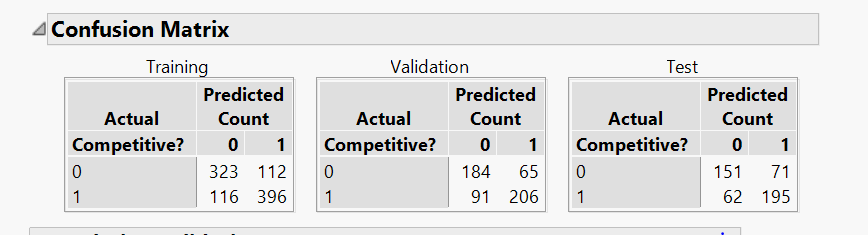


From the confusion matrix, we can see that the accuracy of the Training model is 75.9%, the validation model is 71.4% and the test model is 72.2%. Therefore, we can clearly see that there is less overfitting, thus improving the overall predictive performance.





1. **What is the total accuracy and the accuracy of a 1 prediction as measured by the test data for the tree created in question 5? Include a screen shot of the Confusion Matrix for all three partitions.**



Total Accuracy = Number of correct predictions/ Total predictions

= 151 + 195 / 479

= 346/479

= 72.2%

The Accuracy of 1’s prediction = Number of correct 1’s predictions / Total 1’s prediction

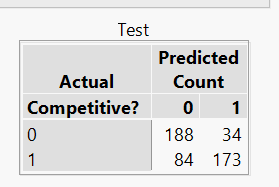
= 195 / 71+195

= 195 / 266

= 73.3%

1. **Compare and contrast the model created in question 2 with the model created in question 5.**

The test matrix for the partitioning classification –



The Accuracy of the test data = Total number of correct predictions/ Total predictions

= 188+173 / 479

= 361/479

= 75.3%

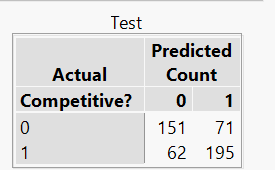
The Accuracy of 1’s in the Test data = Number of correct 1’s prediction/ Total no. of 1’s prediction

= 173 / (173+34)

= 173/207

= 83.5%

The test matrix for the bootstrap classification –



Total Accuracy = Number of correct predictions/ Total predictions

= 151 + 195 / 479

= 346/479

= 72.2%

The Accuracy of 1’s prediction = Number of correct 1’s predictions / Total 1’s prediction

= 195 / 71+195

= 195 / 266

= 73.3%

**The partitioning method performs better than the Bootstrap forest as the dataset is very small.**