Business Analytics: Assignment 2

**Group 1 – Praveen Setty**

# Problem 1

Develop an executive summary of your market basket analysis for upper management at Rite-greens. The summary length should be 1 page (not including the appendix).

* Be sure to justify any settings used in your analysis (e.g., support, confidence, lift ratio, etc.) and address how the results can be used by the company.
* Your summary should reference an appendix containing appropriate output from XLMiner.

## Solution

Ran the Association rules for the data listed in the Binary matrix form. We ran the rules in two iterations

1. Confidence level – 60% and Min Support - 150

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Rule ID** | **A-Support** | **C-Support** | **Support** | **Confidence** | **Lift-Ratio** | **Antecedent** | **Consequent** |
| **Rule 11** | 182 | 357 | 169 | 92.85714286 | 2.60104042 | [Blush,Eye shadow] | [Mascara] |
| **Rule 10** | 184 | 381 | 169 | 91.84782609 | 2.4107041 | [Blush,Mascara] | [Eye shadow] |
| **Rule 6** | 357 | 381 | 321 | 89.91596639 | 2.35999912 | [Mascara] | [Eye shadow] |
| **Rule 13** | 201 | 357 | 179 | 89.05472637 | 2.49453015 | [Concealer,Eye shadow] | [Mascara] |
| **Rule 12** | 204 | 381 | 179 | 87.74509804 | 2.30302095 | [Concealer,Mascara] | [Eye shadow] |
| **Rule 18** | 181 | 381 | 158 | 87.29281768 | 2.29115007 | [Mascara,Lip Gloss] | [Eye shadow] |
| **Rule 16** | 192 | 381 | 166 | 86.45833333 | 2.26924759 | [Mascara,Foundation] | [Eye shadow] |
| **Rule 20** | 175 | 381 | 151 | 86.28571429 | 2.26471691 | [Mascara,Eyeliner] | [Eye shadow] |
| **Rule 7** | 381 | 357 | 321 | 84.2519685 | 2.35999912 | [Eye shadow] | [Mascara] |
| **Rule 21** | 182 | 357 | 151 | 82.96703297 | 2.32400653 | [Eye shadow,Eyeliner] | [Mascara] |
| **Rule 17** | 211 | 357 | 166 | 78.67298578 | 2.20372509 | [Eye shadow,Foundation] | [Mascara] |
| **Rule 19** | 201 | 357 | 158 | 78.60696517 | 2.20187578 | [Eye shadow,Lip Gloss] | [Mascara] |
| **Rule 3** | 234 | 442 | 179 | 76.4957265 | 1.73067255 | [Lip liner] | [Concealer] |
| **Rule 9** | 490 | 536 | 356 | 72.65306122 | 1.35546756 | [Lip Gloss] | [Foundation] |
| **Rule 23** | 227 | 536 | 156 | 68.72246696 | 1.28213558 | [Lip Gloss,Eyeliner] | [Foundation] |
| **Rule 4** | 442 | 457 | 297 | 67.19457014 | 1.4703407 | [Concealer] | [Eyeliner] |
| **Rule 8** | 536 | 490 | 356 | 66.41791045 | 1.35546756 | [Foundation] | [Lip Gloss] |
| **Rule 14** | 231 | 457 | 152 | 65.8008658 | 1.43984389 | [Concealer,Foundation] | [Eyeliner] |
| **Rule 22** | 238 | 490 | 156 | 65.54621849 | 1.33767793 | [Foundation,Eyeliner] | [Lip Gloss] |
| **Rule 5** | 457 | 442 | 297 | 64.98905908 | 1.4703407 | [Eyeliner] | [Concealer] |
| **Rule 15** | 238 | 442 | 152 | 63.86554622 | 1.44492186 | [Foundation,Eyeliner] | [Concealer] |
| **Rule 2** | 279 | 442 | 175 | 62.72401434 | 1.41909535 | [Bronzer] | [Concealer] |
| **Rule 1** | 363 | 442 | 220 | 60.60606061 | 1.37117784 | [Blush] | [Concealer] |

1. Confidence level – 60%, Min Support – 200

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Rule ID** | **A-Support** | **C-Support** | **Support** | **Confidence** | **Lift-Ratio** | **Antecedent** | **Consequent** |
| **Rule 4** | 357 | 381 | 321 | 89.91596639 | 2.35999912 | [Mascara] | [Eye shadow] |
| **Rule 5** | 381 | 357 | 321 | 84.2519685 | 2.35999912 | [Eye shadow] | [Mascara] |
| **Rule 7** | 490 | 536 | 356 | 72.65306122 | 1.35546756 | [Lip Gloss] | [Foundation] |
| **Rule 2** | 442 | 457 | 297 | 67.19457014 | 1.4703407 | [Concealer] | [Eyeliner] |
| **Rule 6** | 536 | 490 | 356 | 66.41791045 | 1.35546756 | [Foundation] | [Lip Gloss] |
| **Rule 3** | 457 | 442 | 297 | 64.98905908 | 1.4703407 | [Eyeliner] | [Concealer] |
| **Rule 1** | 363 | 442 | 220 | 60.60606061 | 1.37117784 | [Blush] | [Concealer] |

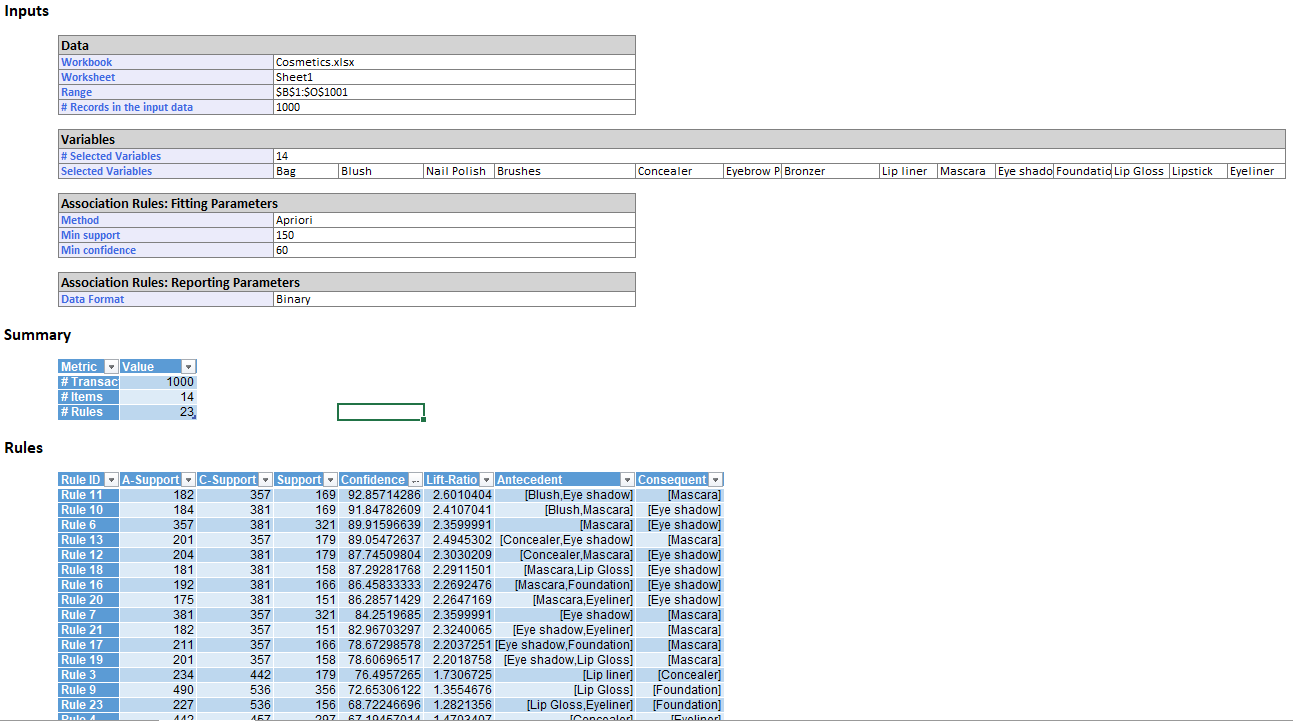
Above are the association rules for the cosmetics based on data provided. The rules above are ordered based on the lift ratio.

Lift Ratio (Definition) – The ratio of a customer probability of buying the Consequent product after the Antecedent product vs just the Consequent product. In other words, by strategically the Antecedent and Consequent products as associated in a way that the sales can be potentially lifted/increased by lift ratio.

### Few Highlights from the Association rules

1. Basic make-up products are highly associated together. First 12 rules in 1st set and first 2 rules in the 2nd set prove the same. Interpretation of some of rules -
   1. Set A, Rule 11 - A Customer buying blush and eye shadow is 2.6% more likely to by Mascara.
   2. Set B, Rule 4 – A Customer buying Mascara is 2.35% more likely to buy Eye shadow.
   3. Set A, Rule 10 - A Customer buying blush and Mascara is 2.41% more likely to buy Eye shadow.
2. Based on above association rules, we need to strategically locate these products in adjacent aisles and run sales promotions on Antecedent products to get the complementary sales on Consequent products.
3. Adding more information like price and profit margin to these rules will provide more insight to specifically target on few products. E.g. We can pick rules associated with products that have high profit margin.

### XLMINER



# Problem 2

Prepare an executive summary of your cluster analysis findings for J&R Equities. The summary length should be 1 page (not including the appendix).

* Be sure to justify any settings used in your analysis (e.g., weights given to different variables, number of clusters formed, etc.) and address how J&R can interpret the clusters (e.g., by naming each cluster, describing patterns, etc.) to better understand the pharmaceutical industry.
* Your summary should reference an appendix containing appropriate output from XLMiner.

## Solution

Iterated the K-means clustering for clusters 2 through 7 and validated the metric to find out the ideal number of clusters. Based on the data and analysis, 3 clusters is an ideal way of grouping the equities with a maximum metric value of 1.38. Below are the 3 clusters and equities in these 3 clusters.

## Summary

* Within the pharmaceutical industry, cluster 1 are growth companies with a Revenue Growth (Cluster Center) at 0.91 (normalized data).
* Cluster 2 are the companies with best Return on Asset
* Cluster 3 are the companies with worst Return on Asset
* From an investing point of view, it is a good idea to have a healthy split between Cluster 1 and 2, and avoid cluster 3 companies.

### Growth Companies

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Record ID** | **Cluster** | **Cluster 1** | **Cluster 2** | **Cluster 3** | **Cluster 4** | **Cluster 5** |
| **Record 5** | 1 | 1.99284699 | 2.84364606 | 3.0532988 | AVE | Aventis |
| **Record 8** | 1 | 3.60563887 | 5.59998543 | 4.56394399 | CHTT | Chattem, Inc |
| **Record 9** | 1 | 2.35576764 | 5.24668086 | 4.81280823 | ELN | Elan Corporation, plc |
| **Record 12** | 1 | 1.16304345 | 3.39983321 | 2.62837772 | IVX | IVAX Corporation |
| **Record 14** | 1 | 1.82880495 | 4.42384131 | 4.01903949 | MRX | Medicis Pharmaceutical Corporation |
| **Record 20** | 1 | 2.21910926 | 3.72350989 | 3.49353847 | WPI | Watson Pharmaceuticals, Inc. |

### Best ROA Companies

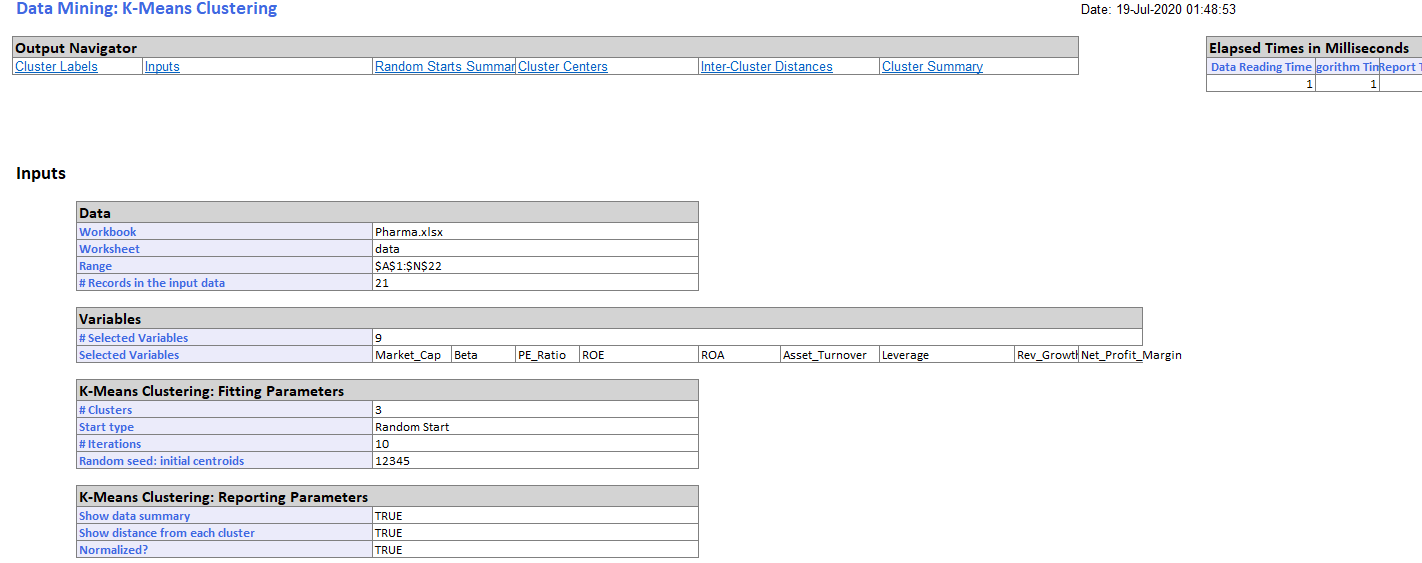
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Record ID** | **Cluster** | **Cluster 1** | **Cluster 2** | **Cluster 3** | **Cluster 4** | **Cluster 5** |
| **Record 1** | 2 | 2.94712765 | 1.37425702 | 2.89358244 | ABT | Abbott Laboratories |
| **Record 4** | 2 | 3.4855838 | 1.16815931 | 3.48635479 | AZN | AstraZeneca PLC |
| **Record 7** | 2 | 3.79515504 | 1.28062169 | 3.96342322 | BMY | Bristol-Myers Squibb Company |
| **Record 10** | 2 | 3.6723083 | 1.66625372 | 4.00492502 | LLY | Eli Lilly and Company |
| **Record 11** | 2 | 5.46232904 | 2.56418355 | 5.90188407 | GSK | GlaxoSmithKline plc |
| **Record 13** | 2 | 4.5823579 | 1.62595174 | 4.19325046 | JNJ | Johnson & Johnson |
| **Record 15** | 2 | 4.49150555 | 1.92237492 | 4.33643105 | MRK | Merck & Co., Inc. |
| **Record 16** | 2 | 4.02063027 | 2.50829483 | 3.99912832 | NVS | Novartis AG |
| **Record 17** | 2 | 5.25122404 | 2.72448543 | 5.86220574 | PFE | Pfizer Inc |
| **Record 19** | 2 | 3.06461039 | 1.56833045 | 3.00774257 | SGP | Schering-Plough Corporation |
| **Record 21** | 2 | 4.09732163 | 2.55948776 | 5.04816344 | WYE | Wyeth |

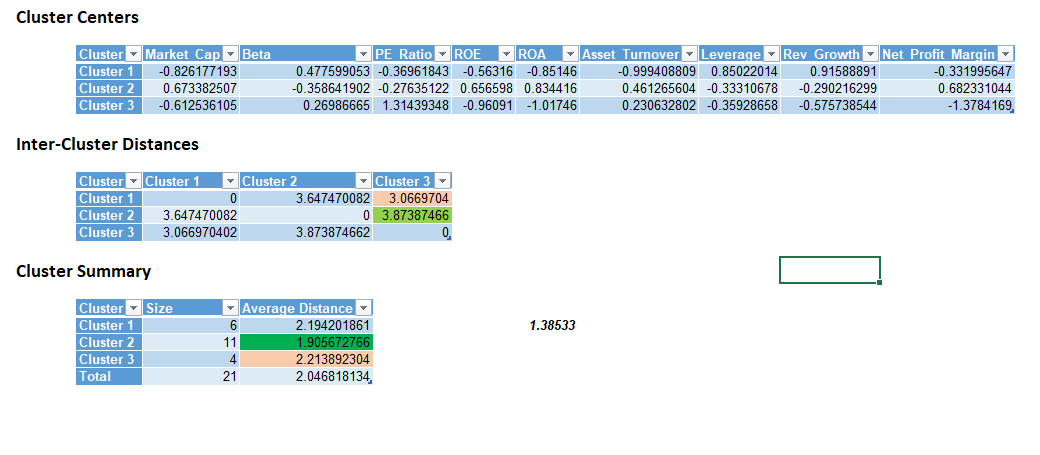
### Worst ROA Companies

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Record ID** | **Cluster** | **Cluster 1** | **Cluster 2** | **Cluster 3** | **Cluster 4** | **Cluster 5** |
| **Record 2** | 3 | 4.84436576 | 5.2313158 | 2.45938375 | AGN | Allergan, Inc. |
| **Record 3** | 3 | 2.88000583 | 2.87889366 | 2.04995541 | AHM | Amersham plc |
| **Record 6** | 3 | 4.06084691 | 5.43871688 | 2.82359572 | BAY | Bayer AG |
| **Record 18** | 3 | 3.14871921 | 3.91602839 | 1.52263433 | PHA | Pharmacia Corporation |

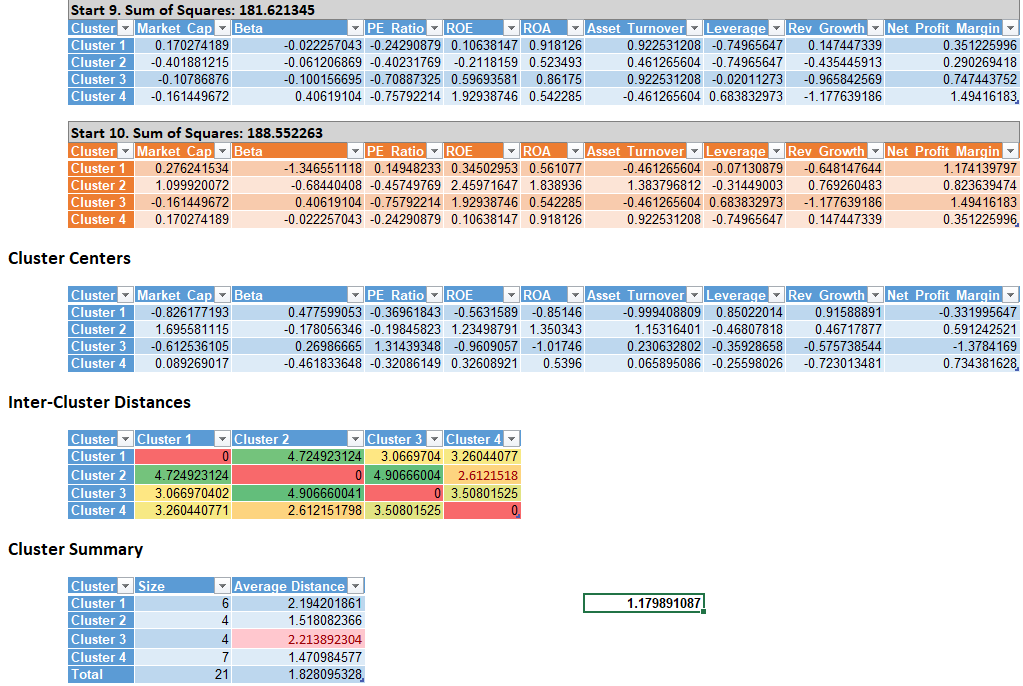
## XLMINER

### 3 Clusters





### 4 Clusters

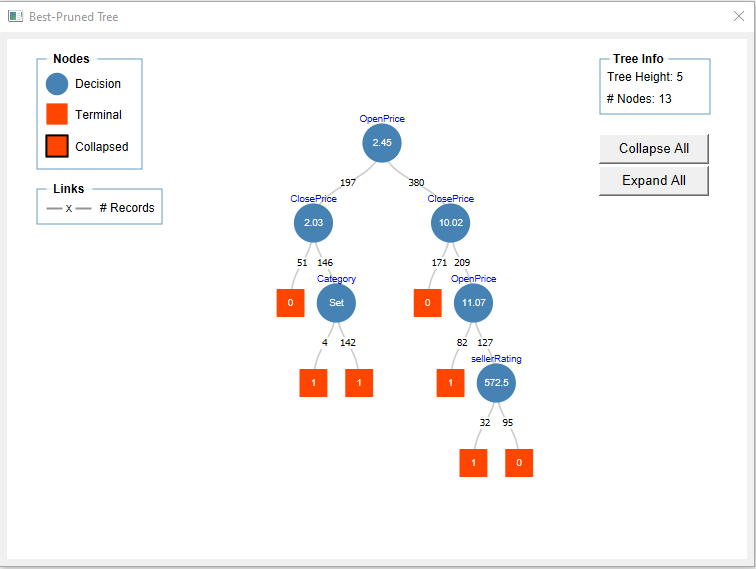


# Problem 3

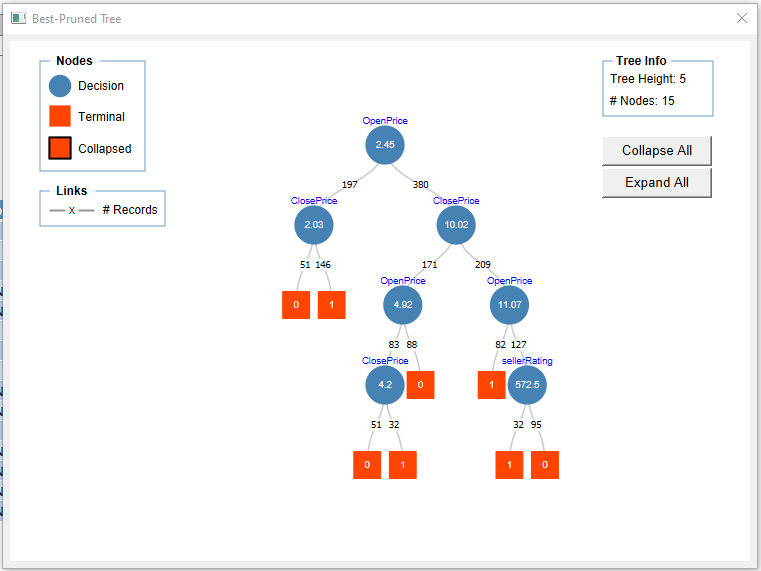
## a1.

Fit a classification tree with all predictors, using the best-pruned tree. To avoid overfitting set the minimum number of records in a leaf node to 50. Also, set the maximum number of levels displayed at seven (the max. allowed in XLMiner). Is this model practical for predicting the outcome of a new auction? Why or why not?

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Limit min # records in leaves** | | | | 191 | |
| **Overall** | 385 | 67 | **17.4026** | |



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Limit min # records in leaves** | | | | 50 | |
| **Overall** | 577 | 89 | **15.42461** | |



Yes, this model is practical to predict the outcome of an auction but at the same time does not consider other parameters like end day and duration. The classification tree predictability is based on Category, OpenPrice, ClosePrice and SellerRating.

By limiting the min # records in leaves from 191 to 50, we reduced the overall error % from 17.4% to 15.42% but an important parameter category is being removed from the classification tree.

## a2.

Describe the interesting and uninteresting information that your tree's rules provide.

Tree provides very interesting insights based on the OpenPrice, ClosePrice and SellerRating that can predict if the auction is competitive or not.

At the same time, the tree must also include other parameters like Category, Duration and End Date as these might impact if the auction is competitive or not because

* Most of the customers do more shopping in weekends and so this can make an auction more competitive.
* Duration – The Auctions that are active for a smaller number of days might be more competitive depending on the category.

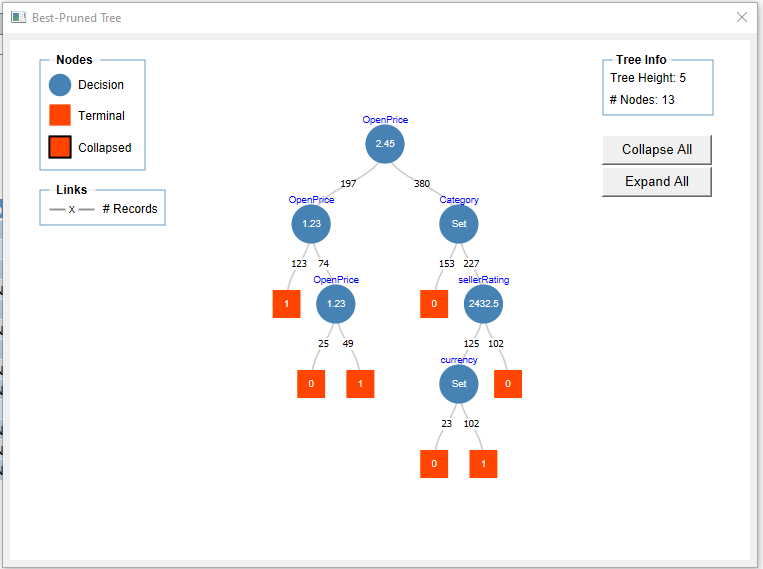
Analysis based on the Rules

* Interesting – Seller Rating is important only when the price of the category is more than $11.07 (and other currencies too)
* Interesting – If the price of the category is less than < 10.02, the seller rating is not considered.
* Uninteresting – Open Price is evaluated again based on the close price. This model will not efficiently work for future auctions, where close price is not known.

## b1.

Fit another classification tree (using the best-pruned tree, with a minimum number of records per leaf node =50 and seven displayed levels) this time only with predictors that can be used for predicting the outcome of a new auction. Describe the resulting tree in terms of rules.

* The tree breaks at an OpenPrice of 2.44987 and if the price is less than this no other parameters are being considered to evaluate the competitiveness.
* If the OpenPrice is greater than 2.44987, the other parameters Catetory, Seller Rating and Current are evaluated is the same order.
* In UK and US, the auction is competitive when the seller rating is between 0 and 2432.5 and the price of the category is greater than 2.45 (Not for all the categories though).



## b2.

Based on the tree you created in b1, what can you conclude from these data about how the chances of a competitive auction relate to the auction settings set by the seller (duration, opening price, ending day, currency)? What would you recommend for a seller as the strategy that will most likely lead to a competitive auction?

To make the auction competitive, the seller must

1. The price of the category must be less than 1.22999
2. The categories must not be Automotive, Clothing/Accessories, Coins/Stamps, Health/Beauty, Jewelry, Music/Movie/Game, Pottery/Glass.
3. Seller must wait to gain a rating of 2432.5 before selling categories expensive than 2.45.
4. Seller need to focus his sales in US and UK.