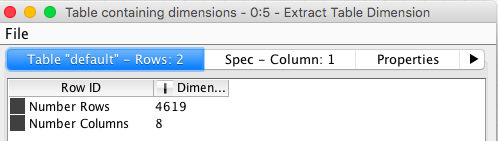
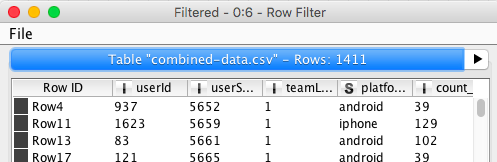
**Data Preparation**

Analysis of combined\_data.csv

Sample Selection

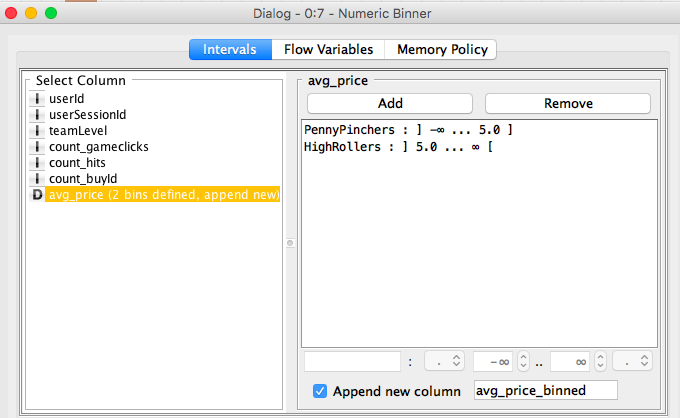
|  |  |
| --- | --- |
| **Item** | **Amount** |
| # of Samples | 4619 |
| # of Samples with Purchases | 1411 |





Attribute Creation

A new categorical attribute was created to enable analysis of players as broken into 2 categories (HighRollers and PennyPinchers). A screenshot of the attribute follows:



Penny Pinchers are users who purchased items costing $5 or less. and HighRollers are users who purchased items costing more than $5. A numeric binner was used to create a new avg\_price\_binned category variable.

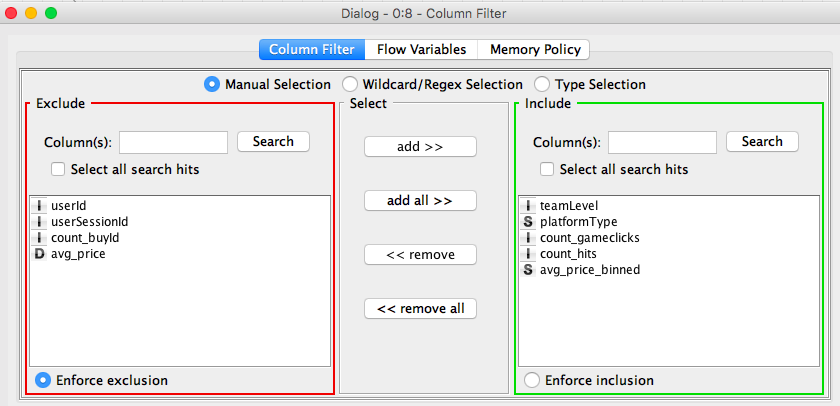
The creation of this new categorical attribute was necessary because:

We need to categorize user in our decision tree algorithm to determine those attributes of a penny pincher and those of high roller.

Attribute Selection

The following attributes were filtered from the dataset for the following reasons:

|  |  |
| --- | --- |
| **Attribute** | **Rationale for Filtering** |
| avg\_price | Replaced by new bins. |
| userid | Don’t need to drill down to this level. |
| userSessionId | Don’t need to keep track of. |
| count\_buyId | # items doesn’t affect penny pincher or high roller status. |



**Data Partitioning and Modeling**

The data was partitioned into train and test datasets.

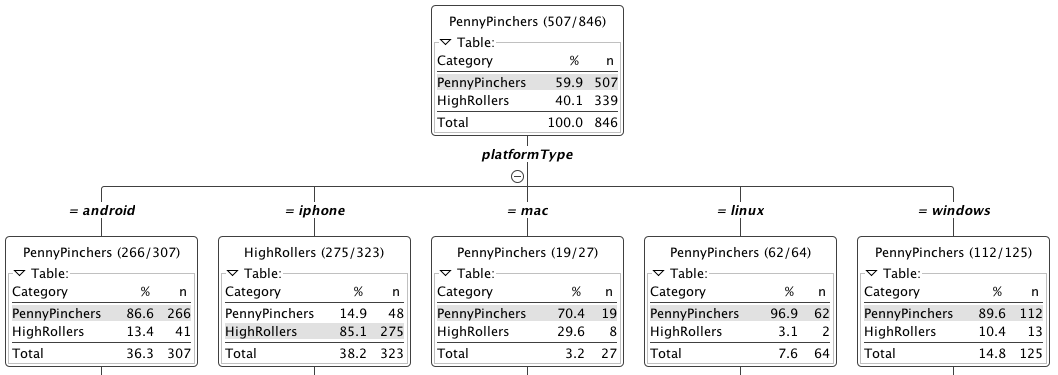
The combined\_data.csv data set was used to create the decision tree model.

The trained model was then applied to 60% of the dataset.

This is important because we want to see if our model represents the data and is not over-fitting

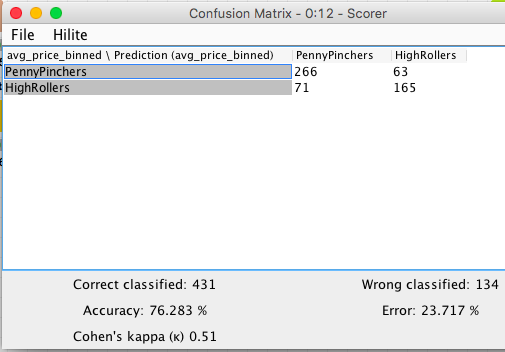
When partitioning the data using sampling, it is important to set the random seed because anyone can reproduce our results.

A screenshot of the resulting decision tree can be seen below:



**Evaluation**

A screenshot of the confusion matrix can be seen below:



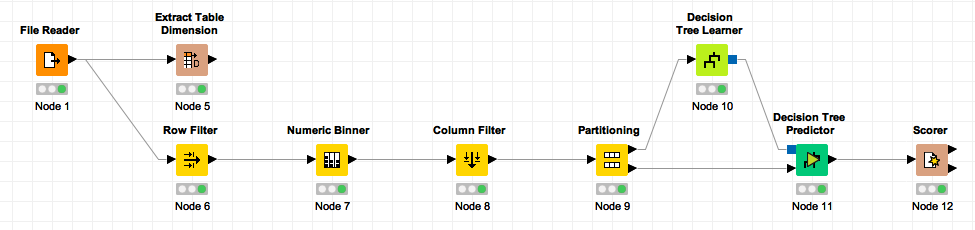
As seen in the screenshot above, the overall accuracy of the model is 76%.

266 Penny Pinchers and 165 High Rollers were accurately classified.

71 Penny Pinchers and 63 High Rollers were mis-classified.

**Analysis Conclusions**

The final KNIME workflow is shown below:



What makes a HighRoller vs. a PennyPincher?

Platform is a good indicator if a user is a High Roller or is a Penny Pincher. iPhone users are by far the highest rollers. Linux users are by far the users who pinch pennies.

|  |
| --- |
| **Specific Recommendations to Increase Revenue** |
| 1. Target expensive items to iPhone users and to Mac users |
| 2. Target inexpensive items to Android users and to some extent Windows users.Skip Linux users. |