**Data Partitioning and Modeling**

The data was partitioned into train and test datasets.

The **Train** data set was used to create the decision tree model.

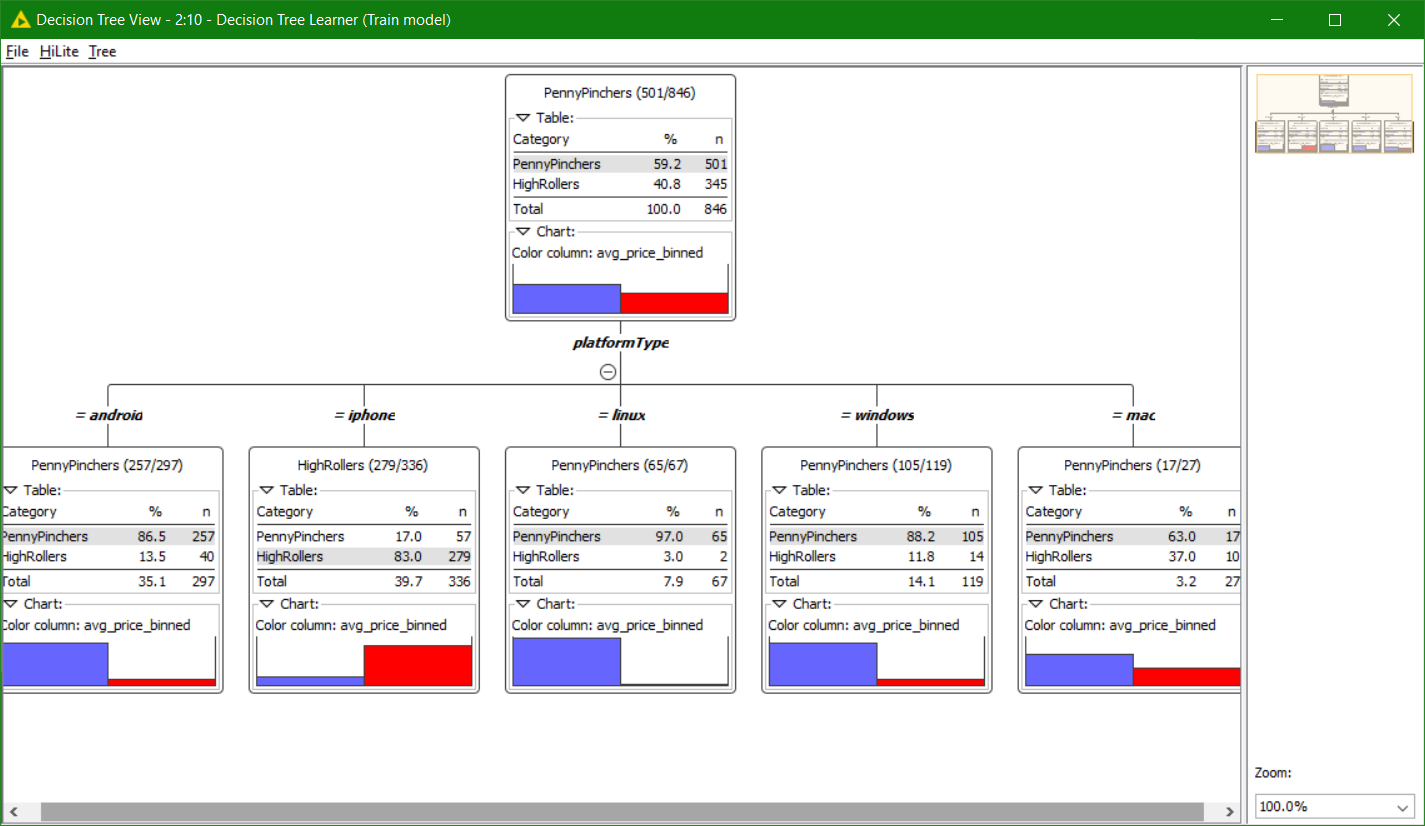
The trained model was then applied to the **Test** dataset.

This is important because the train data set consist of the records whose labels are already known and this will facilitate us to build the classification model(a decision tree) while the test data set would contain records with known labels, thus serving as an unbiased means of evaluating the performance of the trained model. Later on we will be comparing the results of the trained model(i.e accuracy, etc)

When partitioning the data using sampling, it is important to set the random seed because it ensures that I will get the same partitions every time I execute this node(i.e to replicate the same partitioned datasets for repeated executions of the training and scoring process). It is important to get reproducible results. Also, it is not set by default, so we will need to set it when we use this node.

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

Zoom-in the image to view more clearly



The following is the Decision Tree View of the Test Model

