

Data Partitioning and Modeling

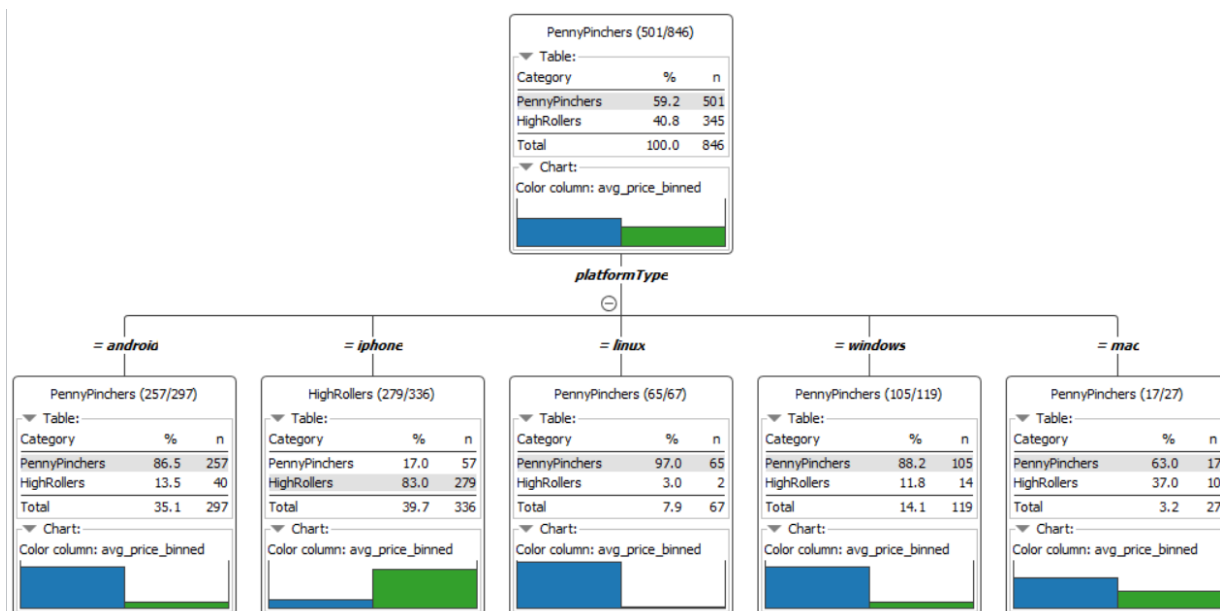
The data was partitioned into train and test datasets.

The training data set without Userid, UsersessionID and Average Price attributes was used to create the decision tree model.

The trained model was then applied to the PennyPinchers and Highrollers dataset.

When partitioning the data using sampling, it is important to set the random seed because it ensures that you will get the same partitions every time you execute this node. This is important for cross checking when multiple users have to work or see the same model. It helps to make the results reproducible.

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



File HiLite

