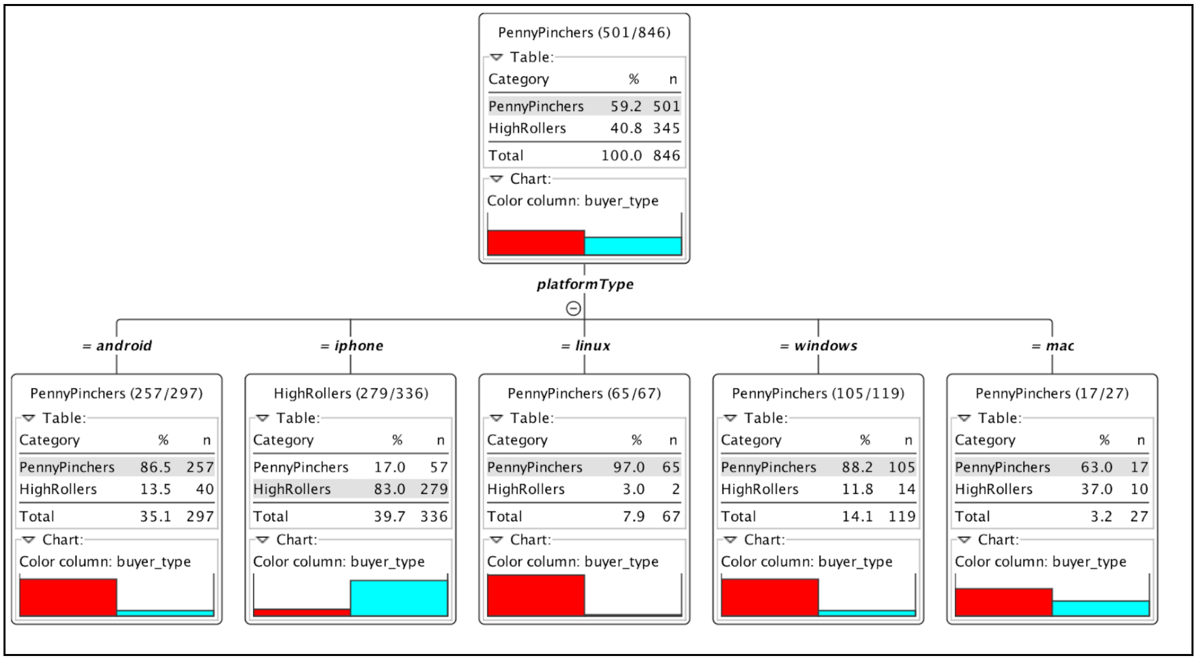
Data Partitioning and Modeling

The data was partitioned into train and test datasets.
The 60% training data set was used to create the decision tree model.
The decision tree model was then applied to the remaining 40% test dataset.
To split the data into a training and test datasets is important because it allows us to build a model based on the training data and then apply this model to the remaining rows (test data set) to measure the accuracy of classification of our model.

When partitioning the data using sampling, it is important to set the same random seed because we can have consistent (same) data partitions when partition node is executed.

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



PennyPinchers (501/846)

Table:

Category	%	n
PennyPinchers	59.2	501
HighRollers	40.8	345
Total	100.0	846

Chart:
Color column: buyer_type

platformType

= android

PennyPinchers (257/297)

Table:

Category	%	n
PennyPinchers	86.5	257
HighRollers	13.5	40
Total	35.1	297

Chart:
Color column: buyer_type

= iphone

HighRollers (279/336)

Table:

Category	%	n
PennyPinchers	17.0	57
HighRollers	83.0	279
Total	39.7	336

Chart:
Color column: buyer_type

= linux

PennyPinchers (65/67)

Table:

Category	%	n
PennyPinchers	97.0	65
HighRollers	3.0	2
Total	7.9	67

Chart:
Color column: buyer_type

= windows

PennyPinchers (105/119)

Table:

Category	%	n
PennyPinchers	88.2	105
HighRollers	11.8	14
Total	14.1	119

Chart:
Color column: buyer_type

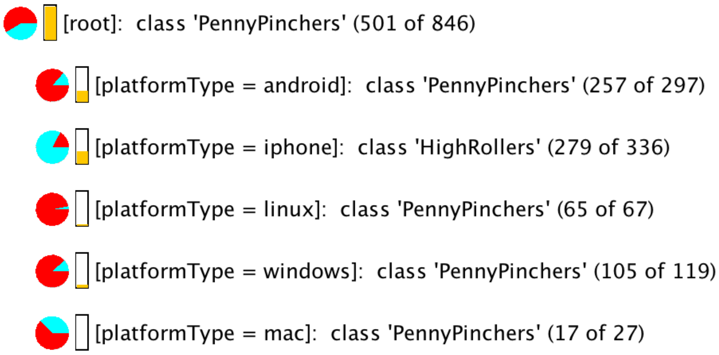
= mac

PennyPinchers (17/27)

Table:

Category	%	n
PennyPinchers	63.0	17
HighRollers	37.0	10
Total	3.2	27

Chart:
Color column: buyer_type



[root]: class 'PennyPinchers' (501 of 846)

[platformType = android]: class 'PennyPinchers' (257 of 297)

[platformType = iphone]: class 'HighRollers' (279 of 336)

[platformType = linux]: class 'PennyPinchers' (65 of 67)

[platformType = windows]: class 'PennyPinchers' (105 of 119)

[platformType = mac]: class 'PennyPinchers' (17 of 27)