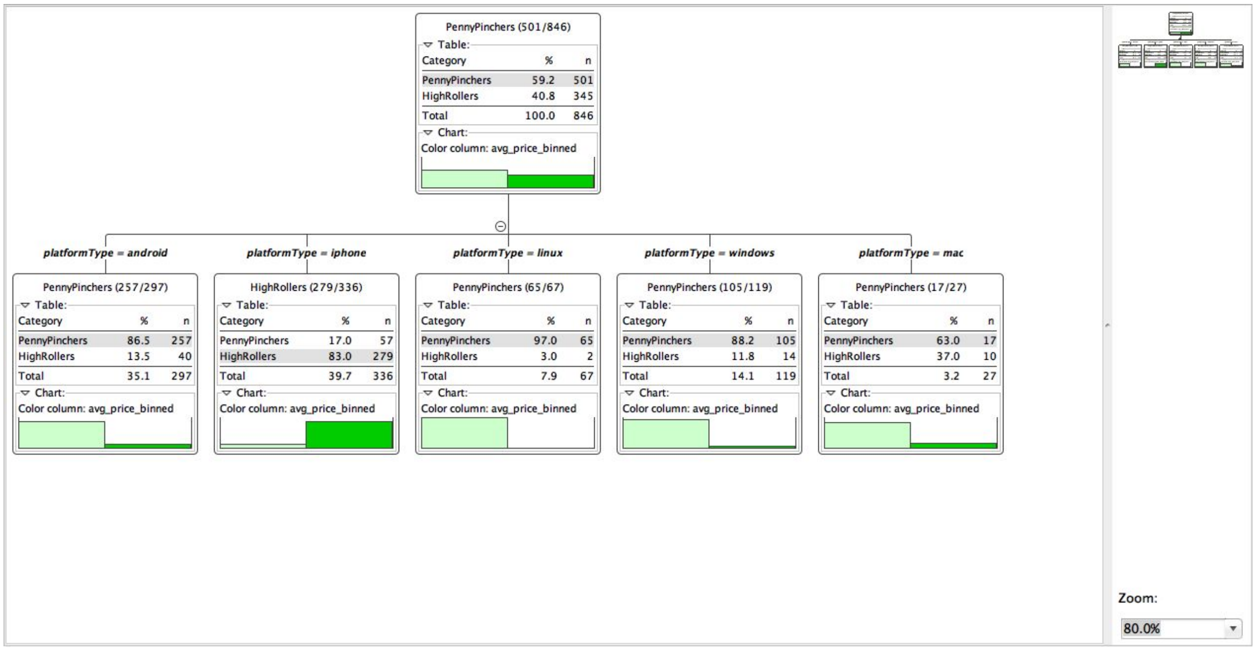
## Data Partitioning and Modeling

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 we want to make sure our classifier is not overfitting or underfitting.

When partitioning the data using sampling, it is important to set the random seed because we want to obtain the same training and test datasets to train and test the decision tree algorithm throughout the participants in the course.

The data was partitioned into 60% train and 40% test datasets. The training data is used to build the classification algorithm, whereas the test data is used to evaluate the classifier on new, unseen data to make sure our classification is not overfitting or underfitting. For this project, we selected the stratified sampling method which divides the datasets into separate groups referred to as strata and a probability sample is drawn from each group. When partitioning the data, it is important to set the random seed to a certain number which in this case is 1466016757670 so that we obtain the same training and test datasets to train and test the decision tree algorithm.

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: avg_price_binned

platformType = android

PennyPinchers (257/297)

Table:

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

Chart:
Color column: avg_price_binned

platformType = iphone

HighRollers (279/336)

Table:

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

Chart:
Color column: avg_price_binned

platformType = linux

PennyPinchers (65/67)

Table:

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

Chart:
Color column: avg_price_binned

platformType = windows

PennyPinchers (105/119)

Table:

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

Chart:
Color column: avg_price_binned

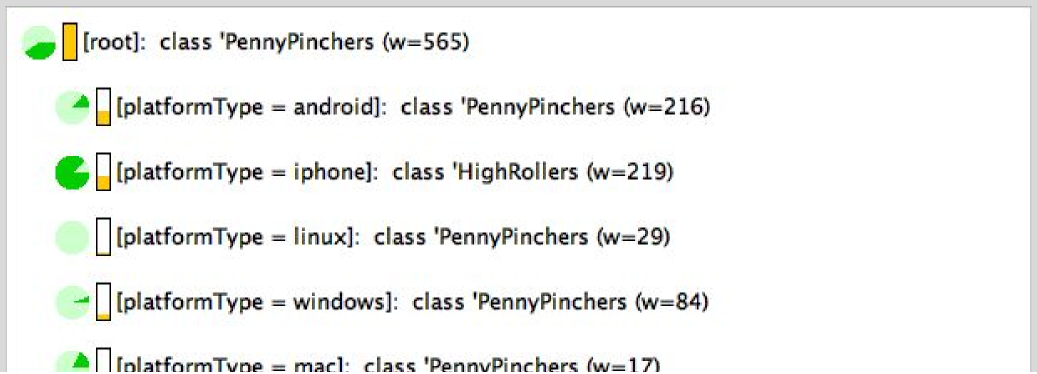
platformType = mac

PennyPinchers (17/27)

Table:

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

Chart:
Color column: avg_price_binned



```
[root]:  class 'PennyPinchers (w=565)
    [platformType = android]:  class 'PennyPinchers (w=216)
    [platformType = iphone]:  class 'HighRollers (w=219)
    [platformType = linux]:  class 'PennyPinchers (w=29)
    [platformType = windows]:  class 'PennyPinchers (w=84)
    [platformType = mac]:  class 'PennyPinchers (w=17)
```