PotatoPricingInsights

Nirav Shah

3/17/2019

```
library(rpart)
library(rpart.plot)
library(xlsx)
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.5.2
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

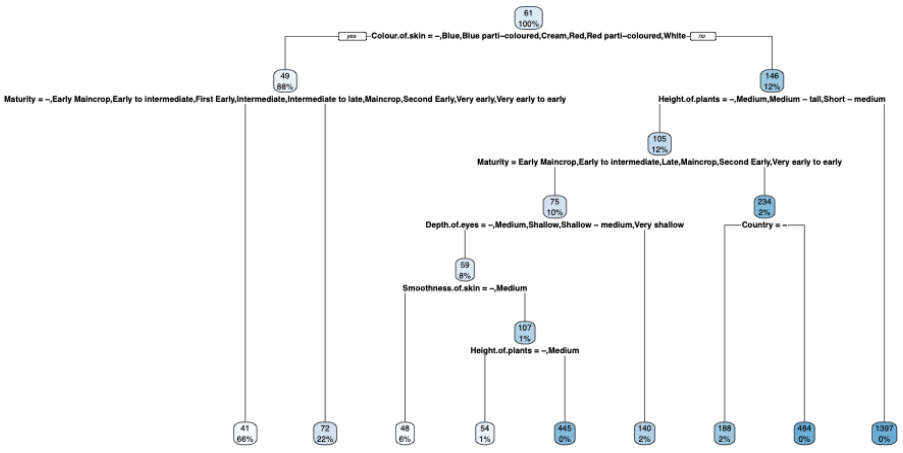
```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
potatoInfo <- read.xlsx("/Users/nvs/Documents/GitHub/PotatoProject/2_Data_Files.xlsx",
                        sheetIndex = 1)
potatoPositions <- read.xlsx("/Users/nvs/Documents/GitHub/PotatoProject/2_Data_Files.xlsx",
                             sheetIndex = 3)

allData <- merge(potatoInfo, potatoPositions, by= 'Product.ID' )

allData <-  allData %>% select( - c(Product.ID, Variety.Name, Quantity, Client.ID))

# basic cart tree with all input variables.
m1 = rpart(formula = 'Price ~.', data = allData,model = )
rpart.plot(m1)
```



```r
# Using the varibale importance parameter to understand which variables are
print(m1$variable.importance)
```

```
##              Height.of.plants                       Country
##                     168648006.1                    48819745.1
##                  Colour.of.skin                      Maturity
##                      26153924.4                    16788491.0
##                   Depth.of.eyes             Smoothness.of.skin
##                       2889530.4                     1106726.0
##                  Shape.of.tuber Colour.of.base.of.lightsprout
##                        879144.1                      212480.3
##                Colour.of.flesh
##                        143446.2
```