

Regression Trees for RET Project Rev.1 (Response variable - HHP)

07/05/2023

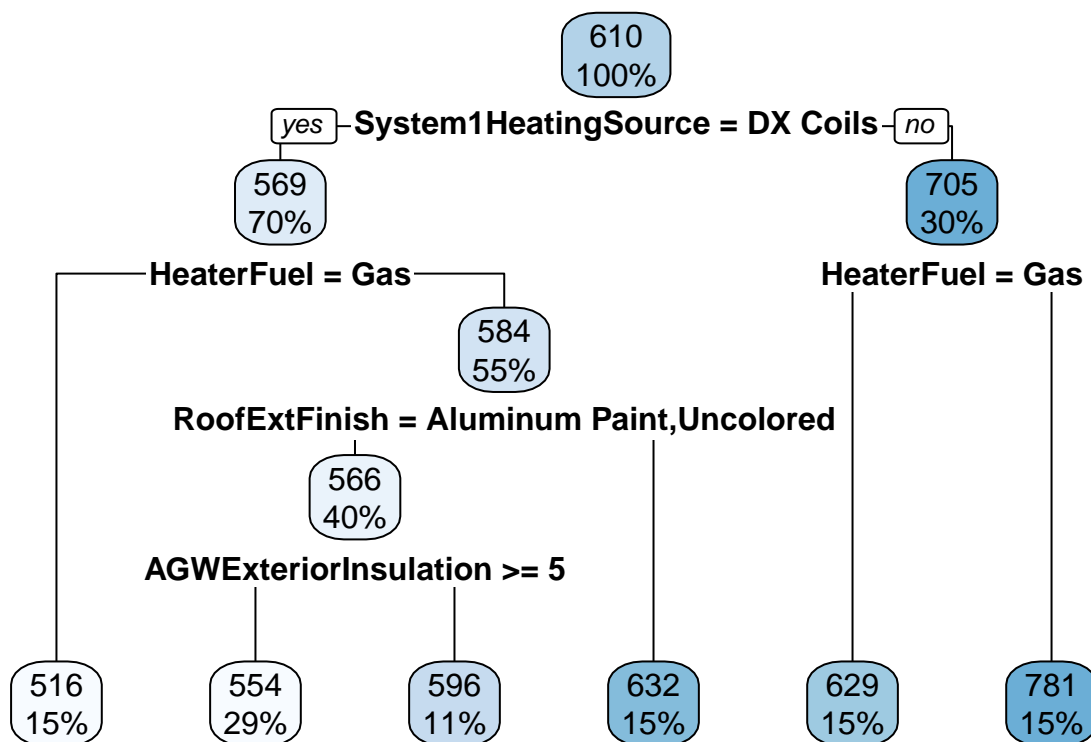
```
#Importing the necessary libraries
library(rpart)
library(rpart.plot)

#Reading the update spreadsheet for HHP response
dfhhp <- readxl::read_xlsx("C:/Users/jaiva/OneDrive/Documents/RET - HHP Output.xlsx")

#Defining the categorical variables
##Defining the categorical variables
RoofExtFinish <- as.factor(dfhhp$RoofExtFinish)
AGWExtFinish <- as.factor(dfhhp$AGWExtFinish)
GlassCategory <- as.factor(dfhhp$GlassCategory)
GlassTypeEmissivity <- as.factor(dfhhp$GlassTypeEmissivity)
FrameType <- as.factor(dfhhp$FrameType)
System1HeatingSource <- as.factor(dfhhp$System1HeatingSource)
System1SystemType <- as.factor(dfhhp$System1SystemType)
SupplyFans <- as.factor(dfhhp$SupplyFans)
HeaterFuel <- as.factor(dfhhp$HeaterFuel)
HeaterType <- as.factor(dfhhp$HeaterType)

#Fitting Initial tree
model3 <- rpart(formula = HHP~.,
                 data = dfhhp,
                 method='anova')

#Plotting the initial fitted tree
rpart.plot(model3)
```



#Summary of the initial fitted tree
`summary(model3)`

```
## Call:
## rpart(formula = HHP ~ ., data = dfhhp, method = "anova")
##   n= 80
##
##           CP nsplit rel error   xerror   xstd
## 1 0.43294731      0 1.0000000 1.0355754 0.20121501
## 2 0.19318814      1 0.5670527 0.5981380 0.12560043
## 3 0.05996071      2 0.3738646 0.4277231 0.08661288
## 4 0.05374320      3 0.3139038 0.4328515 0.08916096
## 5 0.01594910      4 0.2601606 0.3251504 0.06651323
## 6 0.01000000      5 0.2442115 0.3280538 0.06639104
##
## Variable importance
##   System1HeatingSource      HeaterFuel      RoofExtFinish
##              49              28              6
## RoofExteriorInsulation    AGWAdlInsulation      HeaterType
##              4              3              3
##   System1SystemType    RoofAdlInsulation    AGWExteriorInsulation
##              3              3              2
##
## Node number 1: 80 observations,   complexity param=0.4329473
##   mean=609.9875, MSE=8959.662
##   left son=2 (56 obs) right son=3 (24 obs)
```

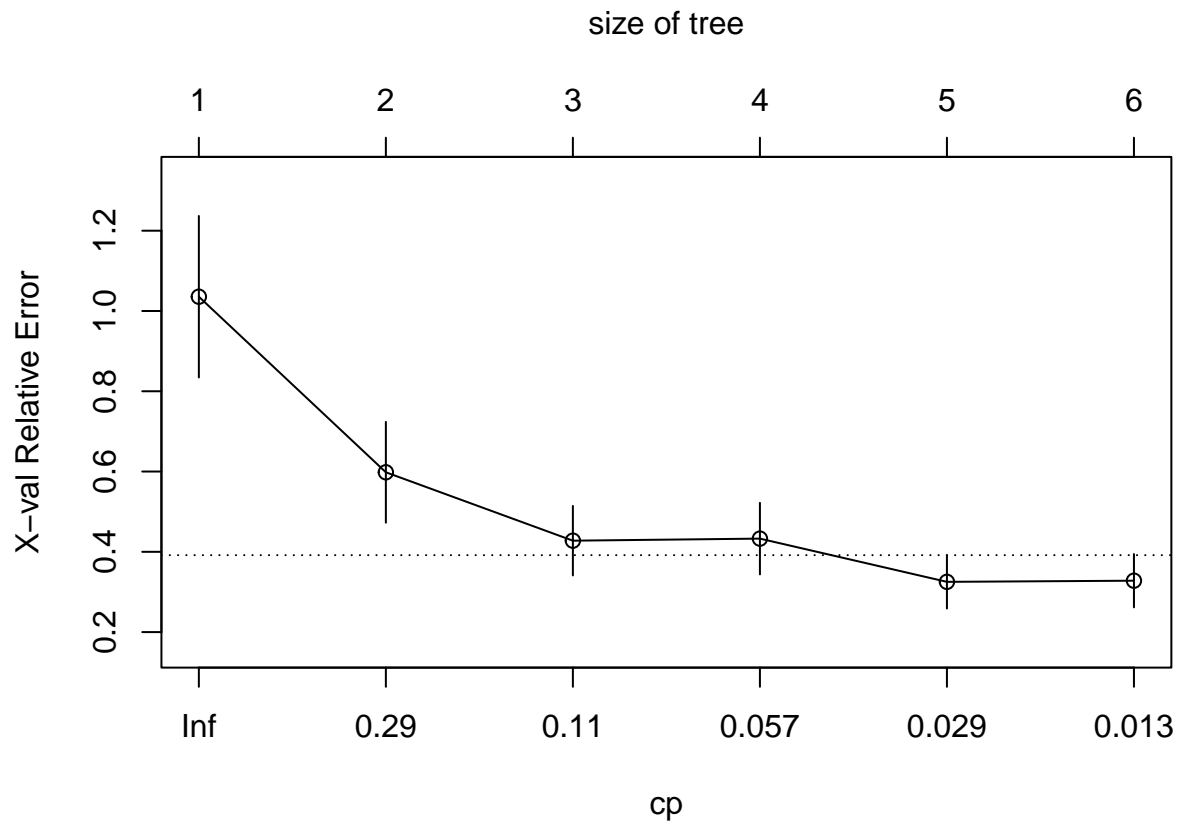
```

## Primary splits:
##   System1HeatingSource splits as LR,   improve=0.4329473, (0 missing)
##   RoofExtFinish       splits as LRRL,  improve=0.1468493, (0 missing)
##   AGWExtFinish        splits as RRLL,  improve=0.1468493, (0 missing)
##   FrameType           splits as LRLR,  improve=0.1468493, (0 missing)
##   SupplyFans          splits as LRR,   improve=0.1468493, (0 missing)
##
## Node number 2: 56 observations,   complexity param=0.05996071
##   mean=569.2143, MSE=3132.454
##   left son=4 (12 obs) right son=5 (44 obs)
##   Primary splits:
##     HeaterFuel          splits as RL,      improve=0.24500540, (0 missing)
##     CeilingsBattInsulation < 1.5   to the left, improve=0.06828378, (0 missing)
##     AGWExteriorInsulation < 5      to the right, improve=0.04313505, (0 missing)
##     AGWAdlInsulation      < 17     to the left, improve=0.03761610, (0 missing)
##     RoofAdlInsulation     < 1.5     to the right, improve=0.02870979, (0 missing)
##
## Node number 3: 24 observations,   complexity param=0.1931881
##   mean=705.125, MSE=9626.276
##   left son=6 (12 obs) right son=7 (12 obs)
##   Primary splits:
##     HeaterFuel          splits as RL,      improve=0.59936660, (0 missing)
##     CeilingsBattInsulation < 10.5  to the left, improve=0.06955084, (0 missing)
##     GlassTypeThickness   < 0.1875 to the right, improve=0.04034981, (0 missing)
##     GlassTypeSpacing     < 0.375  to the right, improve=0.04034981, (0 missing)
##     AGWExteriorInsulation < 15.75 to the right, improve=0.02929067, (0 missing)
##
## Node number 4: 12 observations
##   mean=516.1667, MSE=2768.972
##
## Node number 5: 44 observations,   complexity param=0.0537432
##   mean=583.6818, MSE=2254.808
##   left son=10 (32 obs) right son=11 (12 obs)
##   Primary splits:
##     RoofExtFinish       splits as LRRL,    improve=0.3882781, (0 missing)
##     AGWExtFinish        splits as RRLL,    improve=0.3882781, (0 missing)
##     FrameType           splits as LRLR,    improve=0.3882781, (0 missing)
##     SupplyFans          splits as LRR,     improve=0.3882781, (0 missing)
##     RoofExteriorInsulation < 2          to the right, improve=0.2881640, (0 missing)
##   Surrogate splits:
##     RoofExteriorInsulation < 5          to the right, agree=0.886, adj=0.583, (0 split)
##     AGWAdlInsulation      < 12          to the left, agree=0.864, adj=0.500, (0 split)
##     System1SystemType     splits as RL,    agree=0.864, adj=0.500, (0 split)
##     HeaterType            splits as RL,    agree=0.864, adj=0.500, (0 split)
##     RoofAdlInsulation     < 1.5         to the right, agree=0.841, adj=0.417, (0 split)
##
## Node number 6: 12 observations
##   mean=629.1667, MSE=3433.806
##
## Node number 7: 12 observations
##   mean=781.0833, MSE=4279.41
##
## Node number 10: 32 observations,   complexity param=0.0159491
##   mean=565.5625, MSE=632.0586

```

```
## left son=20 (23 obs) right son=21 (9 obs)
## Primary splits:
## AGWExteriorInsulation < 5 to the right, improve=0.5652109, (0 missing)
## GlassCategory splits as LLRL, improve=0.2655036, (0 missing)
## GlassTypeSpacing < 0.125 to the right, improve=0.2655036, (0 missing)
## AGWAdlInsulation < 5 to the right, improve=0.1448841, (0 missing)
## RoofExteriorInsulation < 32.5 to the left, improve=0.1216449, (0 missing)
##
## Node number 11: 12 observations
## mean=632, MSE=3372
##
## Node number 20: 23 observations
## mean=553.7391, MSE=160.2798
##
## Node number 21: 9 observations
## mean=595.7778, MSE=567.5062
```

```
plotcp(model3)
```



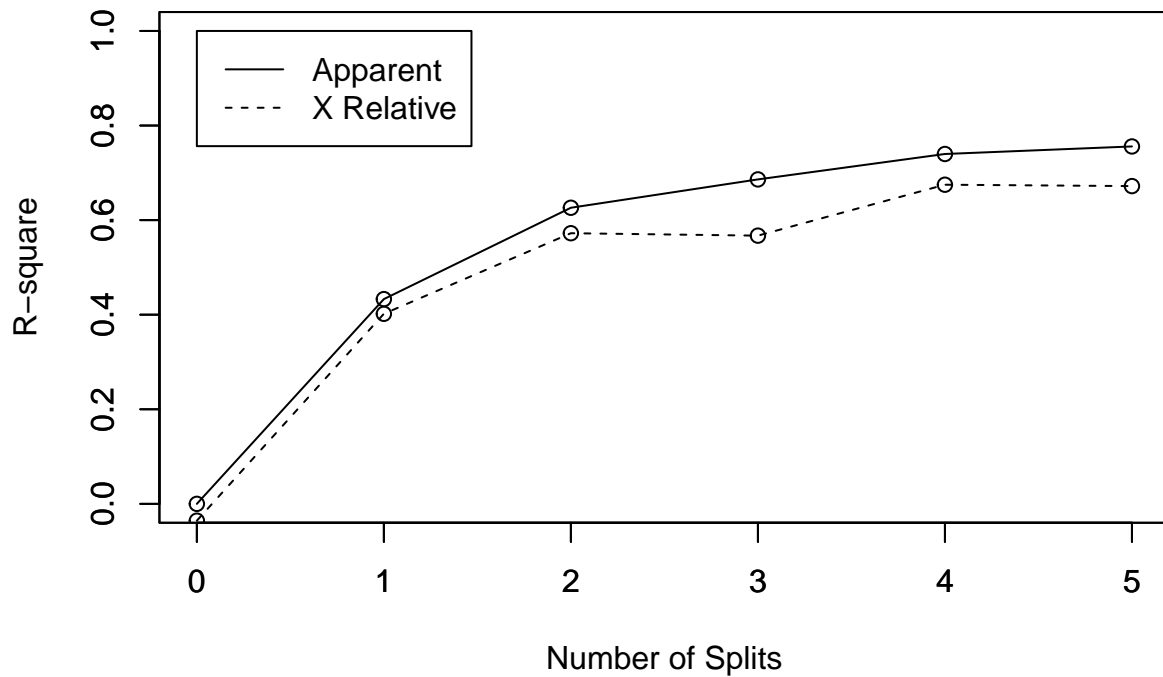
```
#Plotting the r-square value of intial fitted tree
rsq.rpart(model3)
```

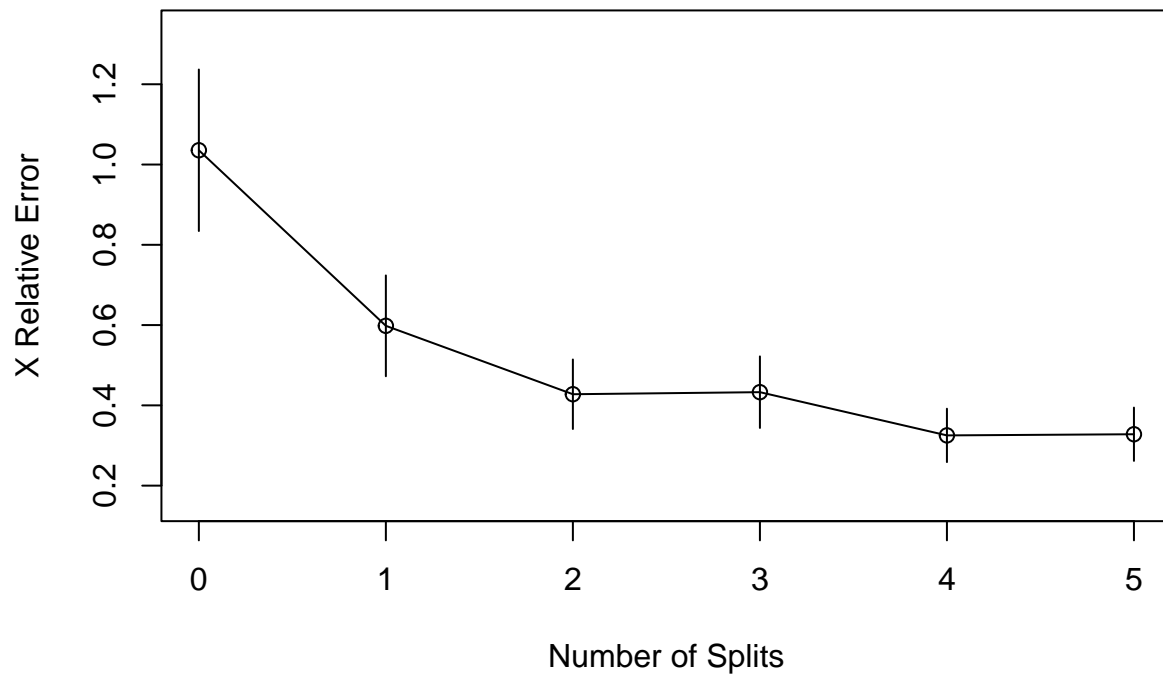
```
##
## Regression tree:
## rpart(formula = HHP ~ ., data = dfhhp, method = "anova")
##
## Variables actually used in tree construction:
```

```

## [1] AGWExteriorInsulation HeaterFuel      RoofExtFinish
## [4] System1HeatingSource
##
## Root node error: 716773/80 = 8959.7
##
## n= 80
##
##      CP nsplit rel error  xerror   xstd
## 1 0.432947    0  1.00000 1.03558 0.201215
## 2 0.193188    1  0.56705 0.59814 0.125600
## 3 0.059961    2  0.37386 0.42772 0.086613
## 4 0.053743    3  0.31390 0.43285 0.089161
## 5 0.015949    4  0.26016 0.32515 0.066513
## 6 0.010000    5  0.24421 0.32805 0.066391

```

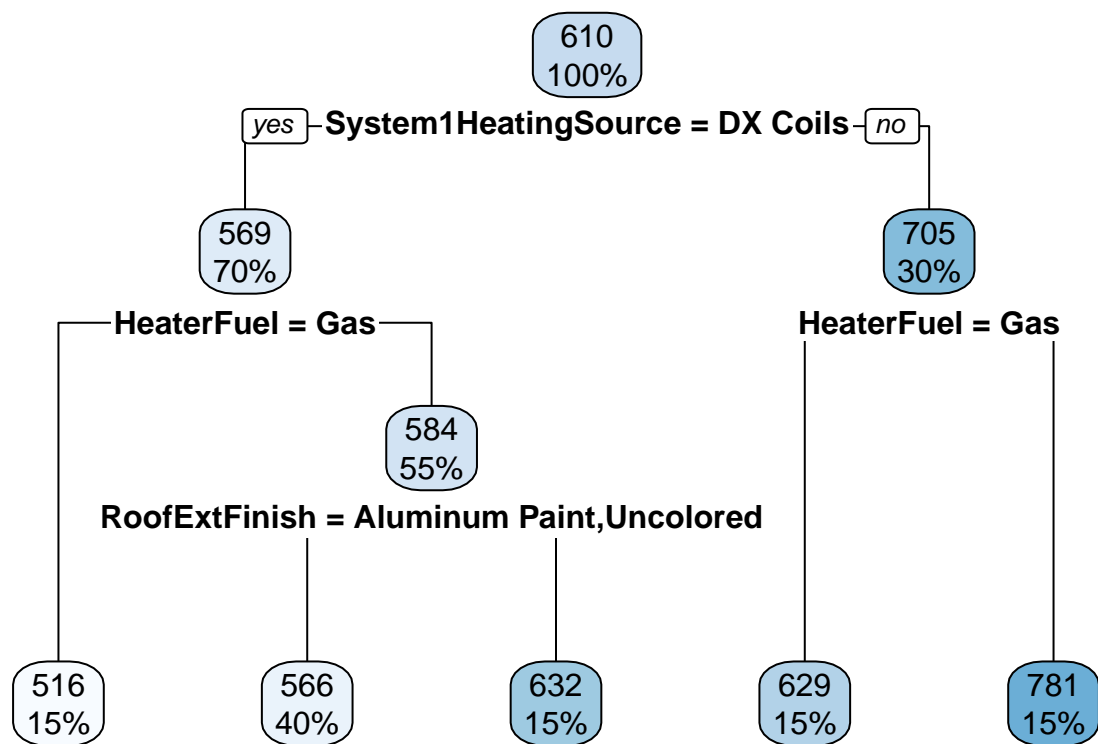




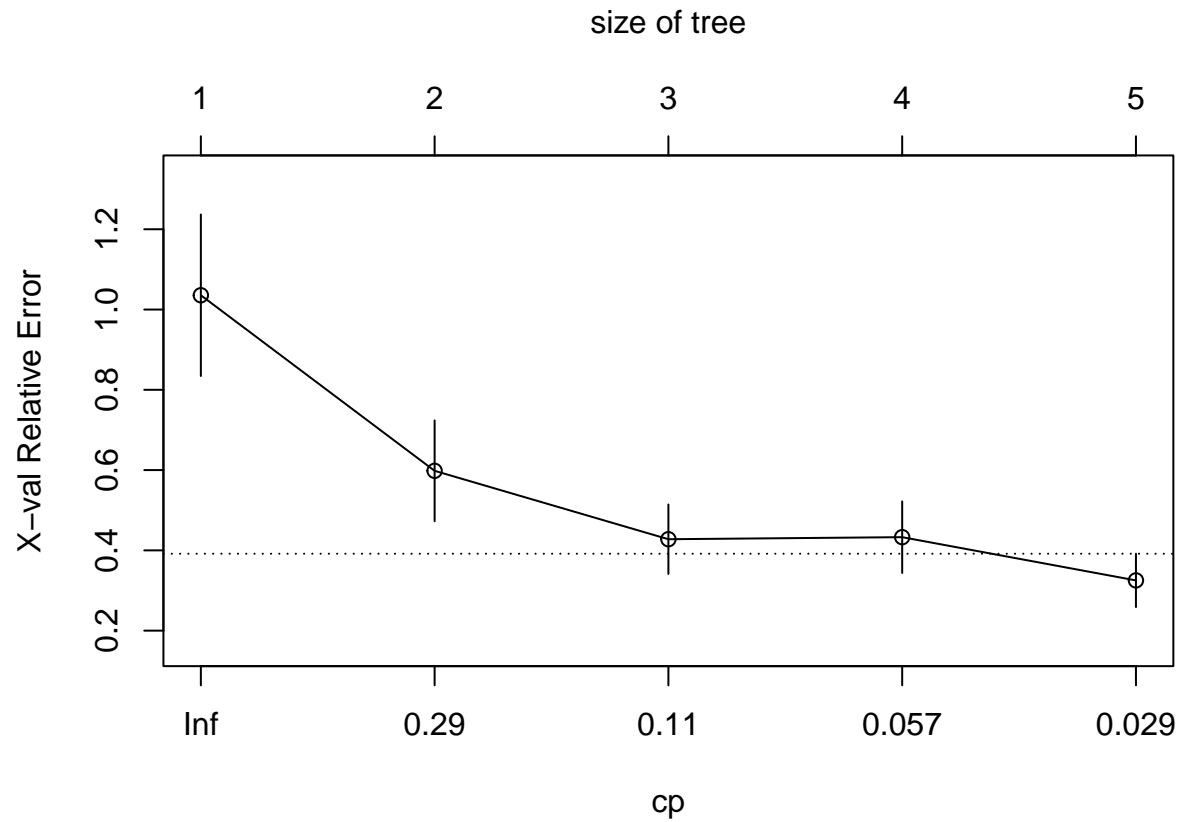
```
#Pruning the initial tree
pruned.tree3 <- prune(model3, cp = model3$cptable[which.min(model3$cptable[, "xerror"]), "CP"])
printcp(pruned.tree3)
```

```
##
## Regression tree:
## rpart(formula = HHP ~ ., data = dfhhp, method = "anova")
##
## Variables actually used in tree construction:
## [1] HeaterFuel      RoofExtFinish    System1HeatingSource
##
## Root node error: 716773/80 = 8959.7
##
## n= 80
##
##      CP nsplit rel error  xerror   xstd
## 1 0.432947     0  1.00000 1.03558 0.201215
## 2 0.193188     1  0.56705 0.59814 0.125600
## 3 0.059961     2  0.37386 0.42772 0.086613
## 4 0.053743     3  0.31390 0.43285 0.089161
## 5 0.015949     4  0.26016 0.32515 0.066513
```

```
#Plotting the pruned tree
rpart.plot(pruned.tree3)
```



```
plotcp(pruned.tree3)
```



```
pruned.tree3
```

```
## n= 80
##
## node), split, n, deviance, yval
##      * denotes terminal node
##
## 1) root 80 716773.00 609.9875
##    2) System1HeatingSource=DX Coils 56 175417.40 569.2143
##      4) HeaterFuel=Gas 12 33227.67 516.1667 *
##      5) HeaterFuel=Electricity 44 99211.55 583.6818
##        10) RoofExtFinish=Aluminum Paint,Uncolored 32 20225.88 565.5625 *
##        11) RoofExtFinish=Dark,Light 12 40464.00 632.0000 *
##    3) System1HeatingSource=Electric Resistance 24 231030.60 705.1250
##      6) HeaterFuel=Gas 12 41205.67 629.1667 *
##      7) HeaterFuel=Electricity 12 51352.92 781.0833 *
```