Regression Tress for RET Project Rev.1 (Response Variable - kWh/year)

07/05/2023

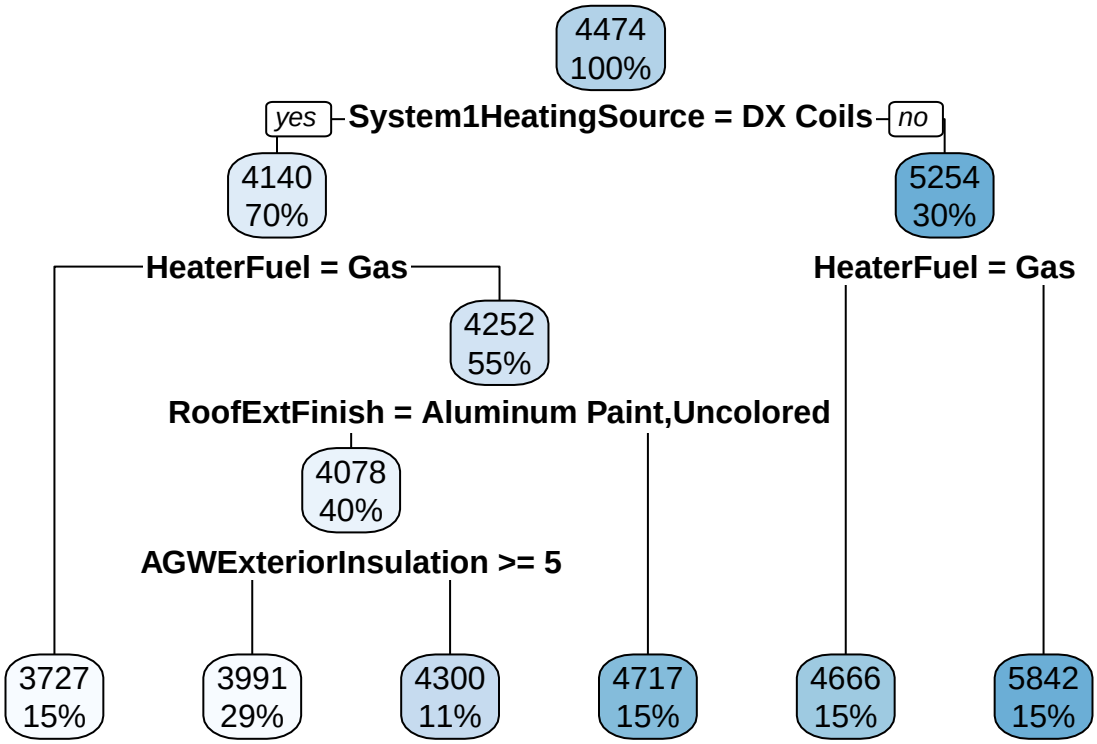
```
#Importing necessary libraries
library(rpart)
library(rpart.plot)
library(ggplot2)
library(readxl)
```

```
#Importing the updated spreadsheet
df <- read_xlsx("C:/Users/jaiva/OneDrive/Documents/RET - Cost Output.xlsx")
```

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

```
model1 <- rpart(formula = kWhperyear ~.,
                data = df,
                method ='anova')
```

```
##Plotting the initial fitted tree
rpart.plot(model1)
```



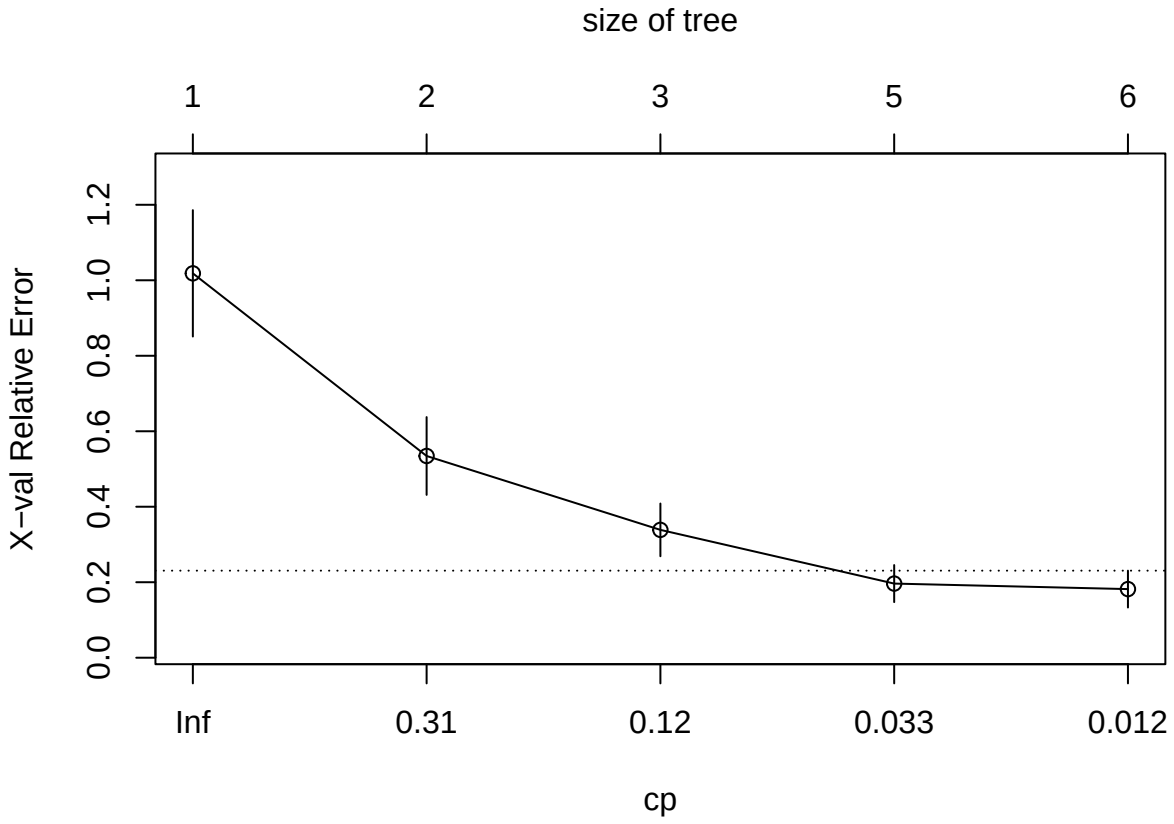

```
summary(model1)
```

```
## Call:
## rpart(formula = kWhperyear ~ ., data = df, method = "anova")
##   n= 80
##
##           CP nsplit rel error    xerror       xstd
## 1 0.49440119      0 1.0000000 1.0183826 0.16747418
## 2 0.19665421      1 0.5055988 0.5344511 0.10308917
## 3 0.07309843      2 0.3089446 0.3386701 0.06985099
## 4 0.01464606      4 0.1627477 0.1964131 0.04910081
## 5 0.01000000      5 0.1481017 0.1817688 0.04885274
##
## Variable importance
##   System1HeatingSource             HeaterFuel          RoofExtFinish
##                     47                     24                         8
## RoofExteriorInsulation       AGWAdlInsulation             HeaterType
##                      5                      4                         4
##      System1SystemType      RoofAdlInsulation  AGWExteriorInsulation
##                      4                      3                         1
##
## Node number 1: 80 observations,    complexity param=0.4944012
##   mean=4474.154, MSE=527589
##   left son=2 (56 obs) right son=3 (24 obs)
##   Primary splits:
##       System1HeatingSource splits as  LR,   improve=0.4944012, (0 missing)
```

```
##       RoofExtFinish        splits as  LRRL, improve=0.1982518, (0 missing)
##       AGWExtFinish         splits as  RRLL, improve=0.1982518, (0 missing)
##       FrameType            splits as  LRLR, improve=0.1982518, (0 missing)
##       SupplyFans           splits as  LRR,  improve=0.1982518, (0 missing)
##
## Node number 2: 56 observations,    complexity param=0.07309843
##   mean=4139.805, MSE=168782.1
##   left son=4 (12 obs) right son=5 (44 obs)
##   Primary splits:
##       HeaterFuel             splits as  RL,         improve=0.27546600, (0 missing)
##       RoofExteriorInsulation < 2      to the right, improve=0.06898060, (0 missing)
##       HeaterType             splits as  RL,         improve=0.04764771, (0 missing)
##       AGWAdlInsulation       < 17     to the left,  improve=0.03616048, (0 missing)
##       RoofExtFinish          splits as  LRRL,       improve=0.03012129, (0 missing)
##
## Node number 3: 24 observations,    complexity param=0.1966542
##   mean=5254.3, MSE=495336.3
##   left son=6 (12 obs) right son=7 (12 obs)
##   Primary splits:
##       HeaterFuel            splits as  RL,         improve=0.69819640, (0 missing)
##       GlassTypeThickness    < 0.1875 to the right, improve=0.04019332, (0 missing)
##       GlassTypeSpacing      < 0.375  to the right, improve=0.04019332, (0 missing)
##       AGWExteriorInsulation < 15.75  to the right, improve=0.03341331, (0 missing)
##       AGWExtFinish          splits as  RL--,       improve=0.02313160, (0 missing)
##
## Node number 4: 12 observations
##   mean=3726.917, MSE=63431.46
##
## Node number 5: 44 observations,    complexity param=0.07309843
##   mean=4252.411, MSE=138340.3
##   left son=10 (32 obs) right son=11 (12 obs)
##   Primary splits:
##       RoofExtFinish          splits as  LRRL,       improve=0.5859889, (0 missing)
##       AGWExtFinish           splits as  RRLL,       improve=0.5859889, (0 missing)
##       FrameType              splits as  LRLR,       improve=0.5859889, (0 missing)
##       SupplyFans             splits as  LRR,        improve=0.5859889, (0 missing)
##       RoofExteriorInsulation < 2      to the right, improve=0.3689811, (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=4666.217, MSE=184247.1
##
## Node number 7: 12 observations
##   mean=5842.383, MSE=114741.4
##
## Node number 10: 32 observations,    complexity param=0.01464606
##   mean=4078.056, MSE=37632.89
##   left son=20 (23 obs) right son=21 (9 obs)
##   Primary splits:
```

```
##       AGWExteriorInsulation  < 5      to the right, improve=0.5133211, (0 missing)
##       GlassTypeSpacing       < 0.125  to the right, improve=0.2290074, (0 missing)
##       GlassCategory          splits as  LLRL,       improve=0.2290074, (0 missing)
##       AGWAdlInsulation       < 5      to the right, improve=0.1784729, (0 missing)
##       CeilingsBattInsulation < 20     to the right, improve=0.1022451, (0 missing)
##
## Node number 11: 12 observations
##   mean=4717.358, MSE=109651.8
##
## Node number 20: 23 observations
##   mean=3991.113, MSE=10308.62
##
## Node number 21: 9 observations
##   mean=4300.244, MSE=38776.21
```

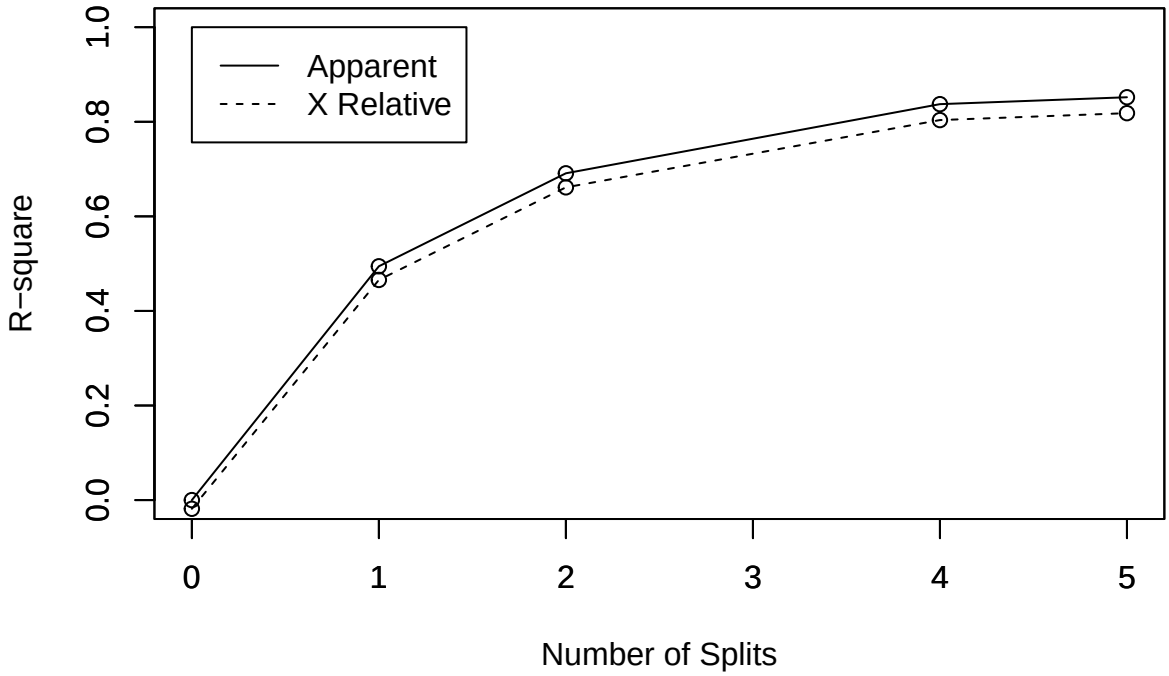
```
#Plotting the cp for the initial fitted tree
plotcp(model1)
```

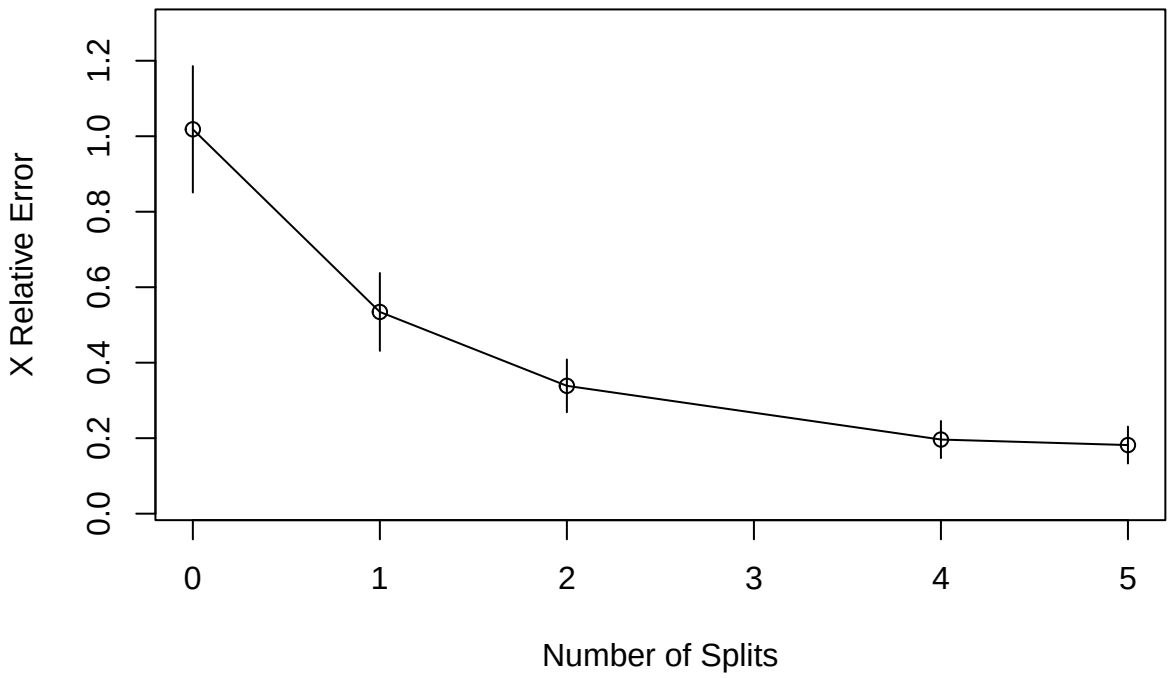



```
##Plotting the r-square value of the initial fitted tree
rsq.rpart(model1)
```

```
##
## Regression tree:
## rpart(formula = kWhperyear ~ ., data = df, method = "anova")
##
## Variables actually used in tree construction:
## [1] AGWExteriorInsulation HeaterFuel            RoofExtFinish
```

```
## [4] System1HeatingSource
##
## Root node error: 42207122/80 = 527589
##
## n= 80
##
##         CP nsplit rel error  xerror     xstd
## 1 0.494401      0   1.00000 1.01838 0.167474
## 2 0.196654      1   0.50560 0.53445 0.103089
## 3 0.073098      2   0.30894 0.33867 0.069851
## 4 0.014646      4   0.16275 0.19641 0.049101
## 5 0.010000      5   0.14810 0.18177 0.048853
```

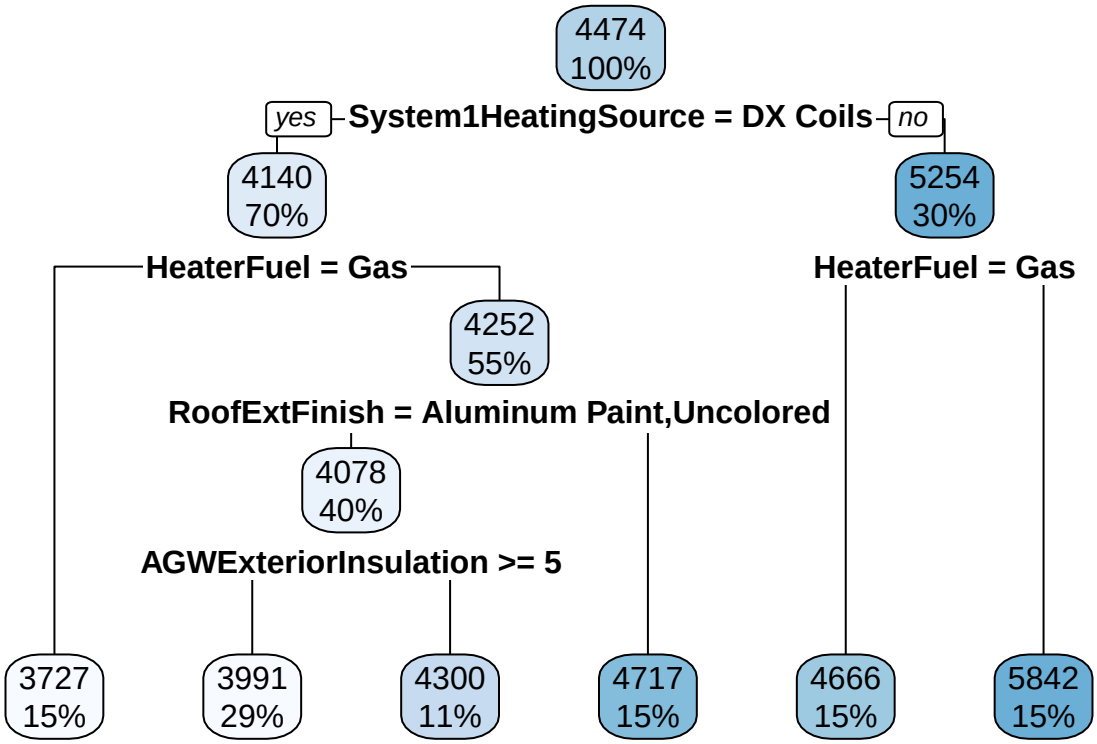




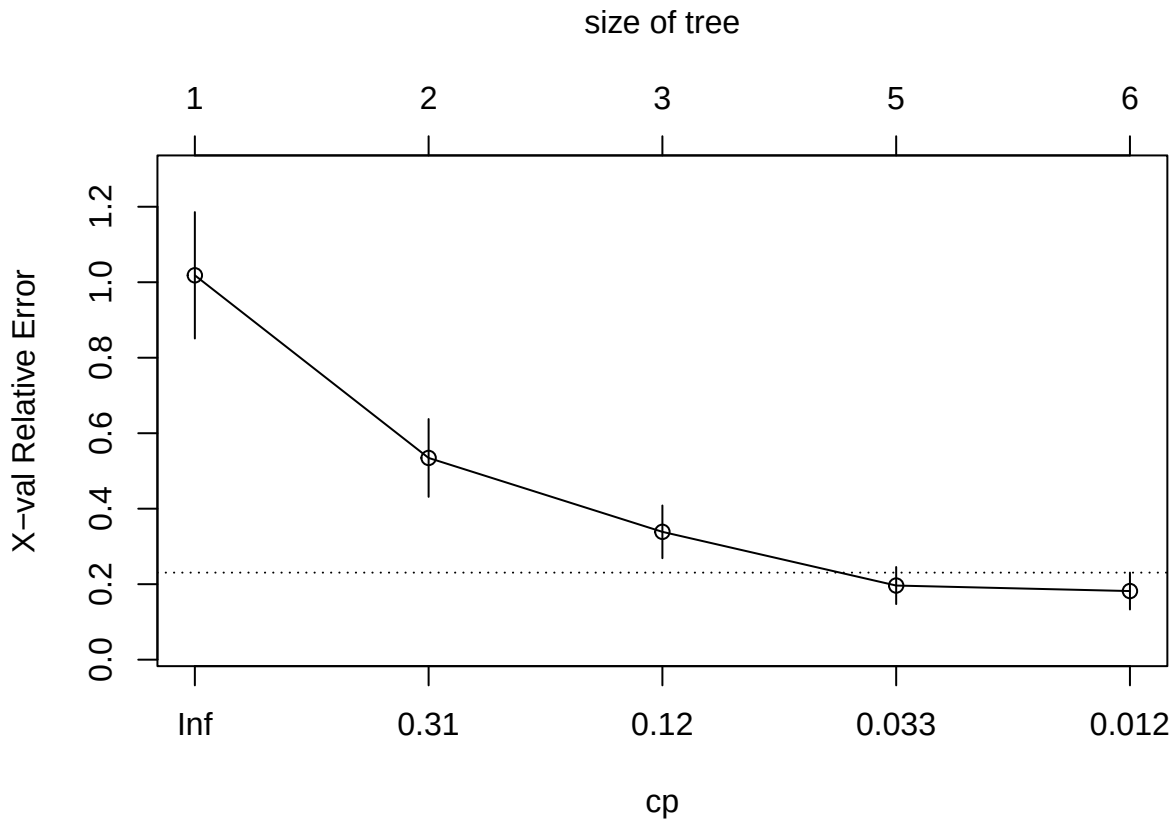

```
#Pruning the tree using minimum cp parameter
pruned.tree <- prune(model1, cp = model1$cptable[which.min(model1$cptable[,"xerror"]),"CP"])
printcp(pruned.tree)
```

```
##
## Regression tree:
## rpart(formula = kWhperyear ~ ., data = df, method = "anova")
##
## Variables actually used in tree construction:
## [1] AGWExteriorInsulation HeaterFuel            RoofExtFinish
## [4] System1HeatingSource
##
## Root node error: 42207122/80 = 527589
##
## n= 80
##
##         CP nsplit rel error  xerror     xstd
## 1 0.494401      0   1.00000 1.01838 0.167474
## 2 0.196654      1   0.50560 0.53445 0.103089
## 3 0.073098      2   0.30894 0.33867 0.069851
## 4 0.014646      4   0.16275 0.19641 0.049101
## 5 0.010000      5   0.14810 0.18177 0.048853
```

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




```
#The pruned tree is same as the initially fitted tree
plotcp(pruned.tree)
```




```
pruned.tree
```

```
## n= 80
##
## node), split, n, deviance, yval
##       * denotes terminal node
##
##  1) root 80 42207120.0 4474.154
##    2) System1HeatingSource=DX Coils 56  9451800.0 4139.805
##      4) HeaterFuel=Gas 12   761177.5 3726.917 *
##      5) HeaterFuel=Electricity 44  6086973.0 4252.411
##       10) RoofExtFinish=Aluminum Paint,Uncolored 32  1204252.0 4078.056
##         20) AGWExteriorInsulation>=5 23   237098.3 3991.113 *
##         21) AGWExteriorInsulation< 5 9   348985.9 4300.244 *
##       11) RoofExtFinish=Dark,Light 12  1315822.0 4717.358 *
##    3) System1HeatingSource=Electric Resistance 24 11888070.0 5254.300
##      6) HeaterFuel=Gas 12  2210965.0 4666.217 *
##      7) HeaterFuel=Electricity 12  1376897.0 5842.383 *
```