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

07/05/2023

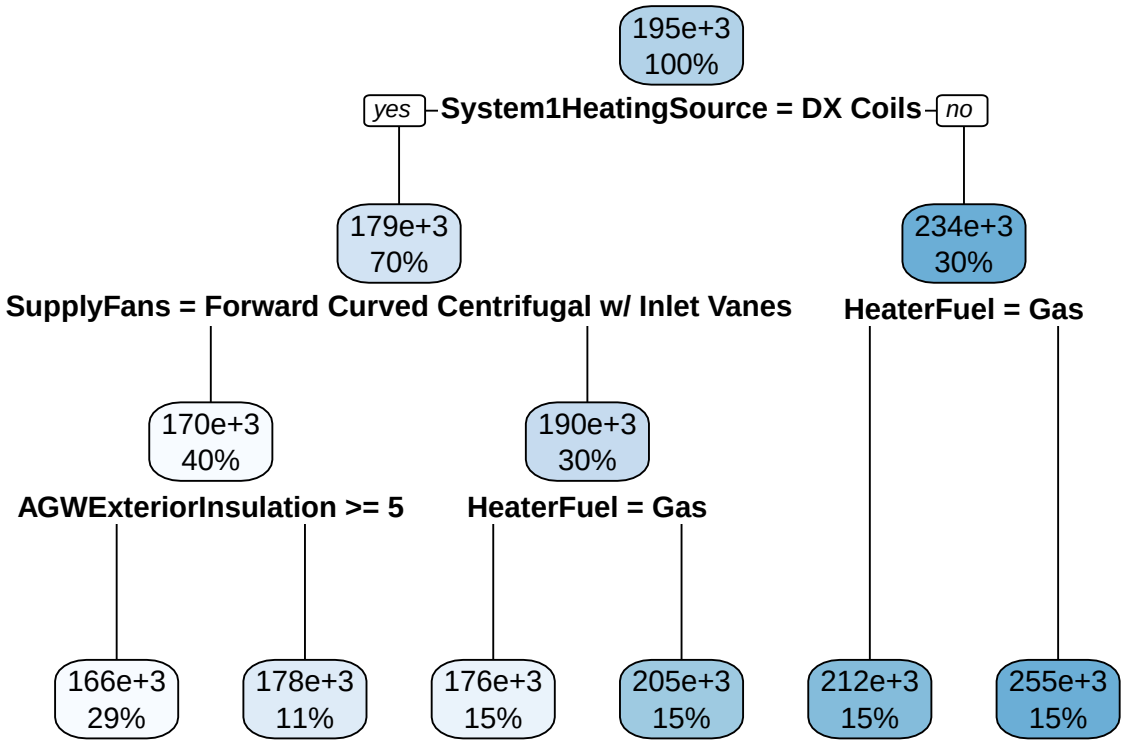
```
#Importing the necessary libraries
library(rpart)
library(rpart.plot)
```

```
##Reading the updated spreadsheet
dfgwp <- readxl::read_xlsx("C:/Users/jaiva/OneDrive/Documents/RET - GWP Output.xlsx")
```

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

```
model2 <- rpart(formula = GWP~.,
                data = dfgwp,
                method = 'anova',)
```

```
##Fitting the initial tree
rpart.plot(model2)
```



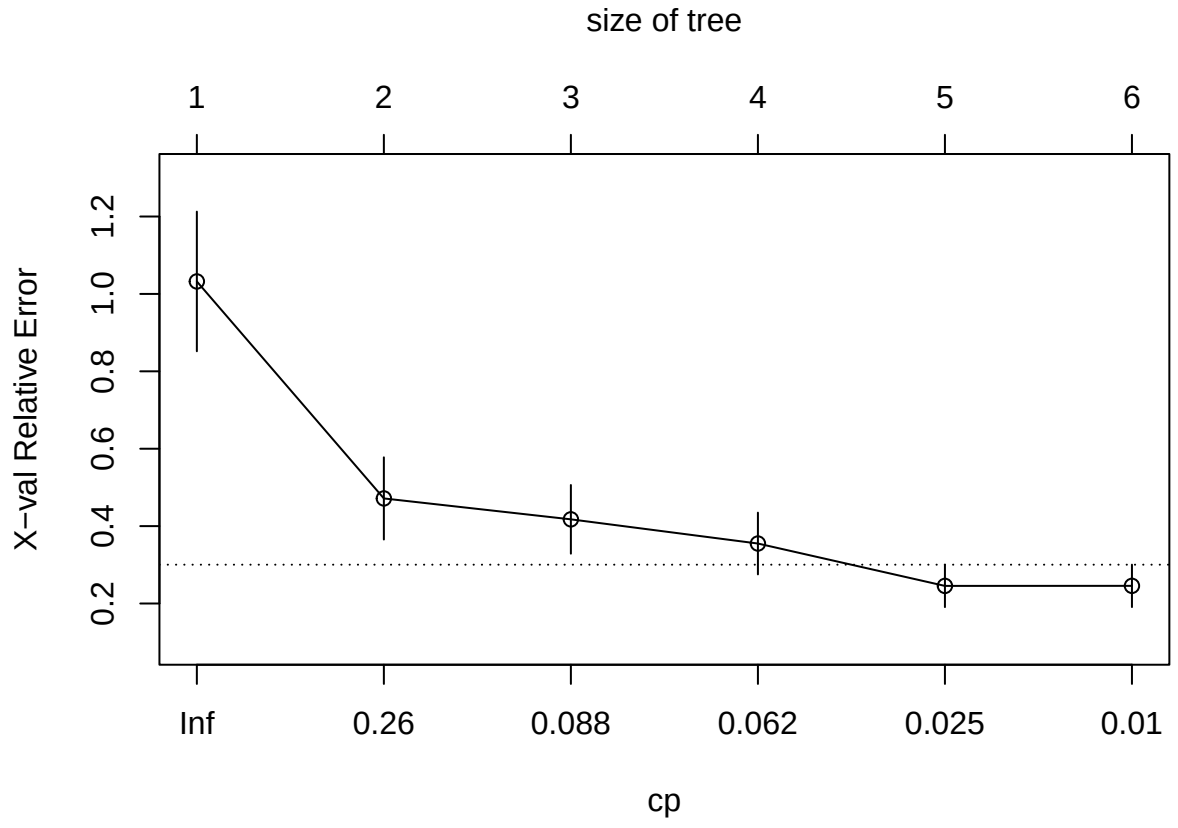

```
summary(model2)
```

```
## Call:
## rpart(formula = GWP ~ ., data = dfgwp, method = "anova")
##   n= 80
##
##           CP nsplit rel error    xerror       xstd
## 1 0.55412275      0 1.0000000 1.0322501 0.18041872
## 2 0.11824492      1 0.4458773 0.4715404 0.10623933
## 3 0.06579178      2 0.3276323 0.4174949 0.08872683
## 4 0.05826506      3 0.2618406 0.3550122 0.07976947
## 5 0.01085387      4 0.2035755 0.2455157 0.05483341
## 6 0.01000000      5 0.1927216 0.2456111 0.05483084
##
## Variable importance
##   System1HeatingSource             HeaterFuel             SupplyFans
##                     55                     21                      7
## RoofExteriorInsulation       AGWAdlInsulation      RoofAdlInsulation
##                      5                      5                      4
##      System1SystemType  AGWExteriorInsulation
##                      3                      1
##
## Node number 1: 80 observations,    complexity param=0.5541227
##   mean=195012.5, MSE=1.145662e+09
##   left son=2 (56 obs) right son=3 (24 obs)
##   Primary splits:
```

```
##       System1HeatingSource splits as  LR,   improve=0.5541227, (0 missing)
##       RoofExtFinish        splits as  LRRL, improve=0.3778275, (0 missing)
##       AGWExtFinish         splits as  RRLL, improve=0.3778275, (0 missing)
##       FrameType            splits as  LRLR, improve=0.3778275, (0 missing)
##       SupplyFans           splits as  LRR,  improve=0.3778275, (0 missing)
##
## Node number 2: 56 observations,    complexity param=0.06579178
##   mean=178517.9, MSE=3.512497e+08
##   left son=4 (32 obs) right son=5 (24 obs)
##   Primary splits:
##       SupplyFans       splits as  LRR,        improve=0.3065592, (0 missing)
##       RoofExtFinish    splits as  LRRL,       improve=0.3065592, (0 missing)
##       AGWExtFinish     splits as  RRLL,       improve=0.3065592, (0 missing)
##       FrameType        splits as  LRLR,       improve=0.3065592, (0 missing)
##       AGWAdlInsulation < 17     to the left,  improve=0.2152182, (0 missing)
##   Surrogate splits:
##       RoofExteriorInsulation < 5      to the right, agree=0.911, adj=0.792, (0 split)
##       AGWAdlInsulation       < 12     to the left,  agree=0.893, adj=0.750, (0 split)
##       RoofAdlInsulation      < 20     to the right, agree=0.804, adj=0.542, (0 split)
##       System1SystemType      splits as  RL,         agree=0.786, adj=0.500, (0 split)
##       HeaterFuel             splits as  LR,         agree=0.786, adj=0.500, (0 split)
##
## Node number 3: 24 observations,    complexity param=0.1182449
##   mean=233500, MSE=8.831667e+08
##   left son=6 (12 obs) right son=7 (12 obs)
##   Primary splits:
##       HeaterFuel             splits as  RL,         improve=0.51129930, (0 missing)
##       GlassTypeThickness     < 0.1875 to the right, improve=0.04662043, (0 missing)
##       GlassTypeSpacing       < 0.375  to the right, improve=0.04662043, (0 missing)
##       GlassTypeEmissivity    splits as  LR,         improve=0.03963798, (0 missing)
##       CeilingsBattInsulation < 10.5   to the left,  improve=0.03635906, (0 missing)
##
## Node number 4: 32 observations,    complexity param=0.01085387
##   mean=169531.2, MSE=6.043652e+07
##   left son=8 (23 obs) right son=9 (9 obs)
##   Primary splits:
##       AGWExteriorInsulation  < 5      to the right, improve=0.51437750, (0 missing)
##       GlassCategory          splits as  LLRL,       improve=0.24897530, (0 missing)
##       GlassTypeSpacing       < 0.125  to the right, improve=0.24897530, (0 missing)
##       AGWAdlInsulation       < 5      to the right, improve=0.18374240, (0 missing)
##       RoofExteriorInsulation < 32.5   to the left,  improve=0.09816278, (0 missing)
##
## Node number 5: 24 observations,    complexity param=0.05826506
##   mean=190500, MSE=4.8775e+08
##   left son=10 (12 obs) right son=11 (12 obs)
##   Primary splits:
##       HeaterFuel             splits as  RL,         improve=0.45619060, (0 missing)
##       CeilingsBattInsulation < 10.5   to the left,  improve=0.11790250, (0 missing)
##       AGWAdlInsulation       < 16     to the left,  improve=0.04464947, (0 missing)
##       RoofExteriorInsulation < 2      to the right, improve=0.03418475, (0 missing)
##       RoofAdlInsulation      < 9.5    to the right, improve=0.03012700, (0 missing)
##
## Node number 6: 12 observations
##   mean=212250, MSE=3.708542e+08
```

```
##
## Node number 7: 12 observations
##   mean=254750, MSE=4.923542e+08
##
## Node number 8: 23 observations
##   mean=166043.5, MSE=1.725898e+07
##
## Node number 9: 9 observations
##   mean=178444.4, MSE=6.024691e+07
##
## Node number 10: 12 observations
##   mean=175583.3, MSE=2.120764e+08
##
## Node number 11: 12 observations
##   mean=205416.7, MSE=3.184097e+08
```

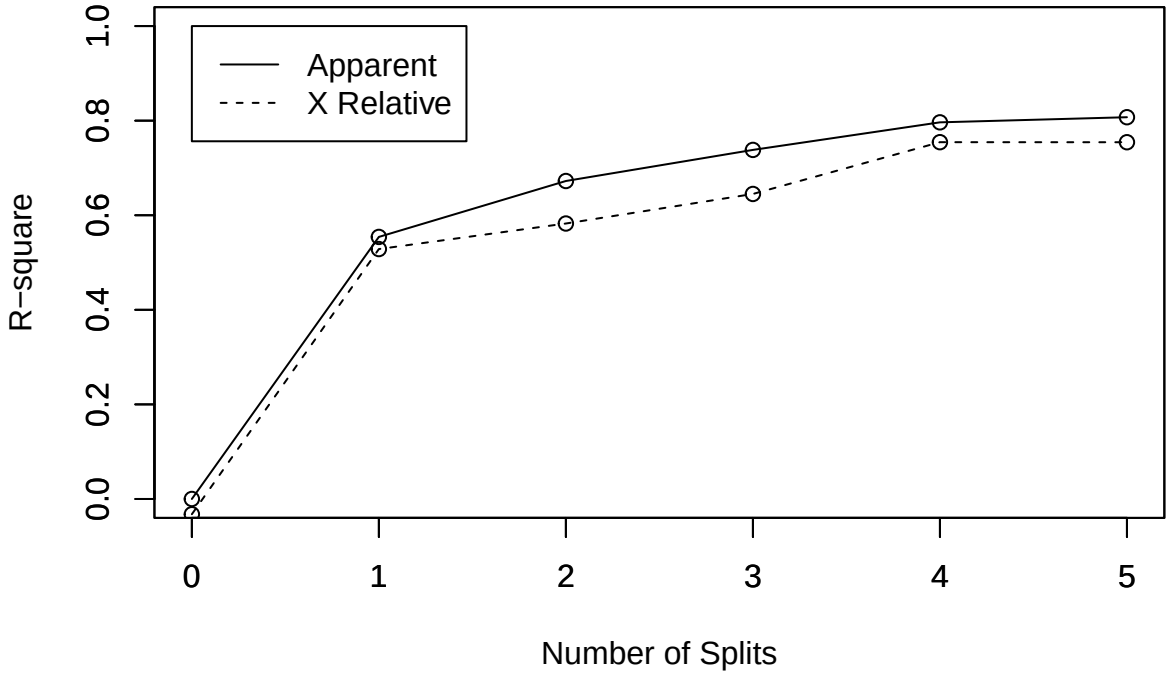
```
plotcp(model2)
```

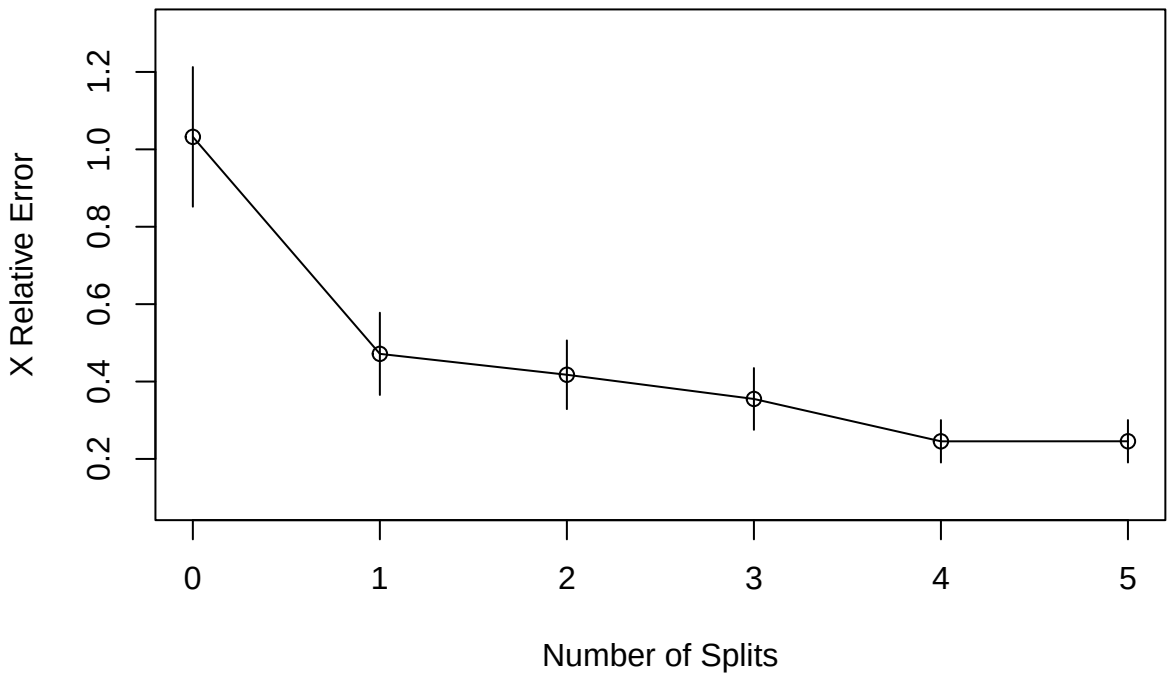



```
#Plotting the r-squared value for initial fitted tree for GWP as response
rsq.rpart(model2)
```

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

```
## [4] System1HeatingSource
##
## Root node error: 9.1653e+10/80 = 1145662344
##
## n= 80
##
##         CP nsplit rel error  xerror     xstd
## 1 0.554123      0   1.00000 1.03225 0.180419
## 2 0.118245      1   0.44588 0.47154 0.106239
## 3 0.065792      2   0.32763 0.41749 0.088727
## 4 0.058265      3   0.26184 0.35501 0.079769
## 5 0.010854      4   0.20358 0.24552 0.054833
## 6 0.010000      5   0.19272 0.24561 0.054831
```

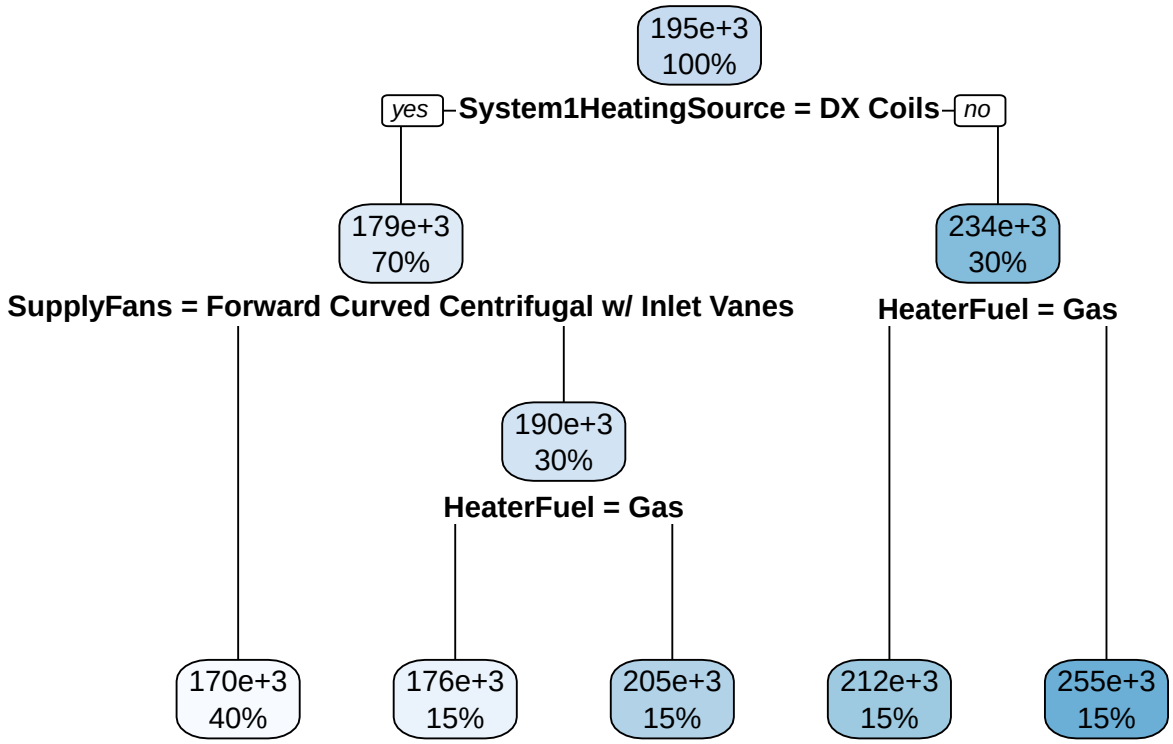





```
#Pruning the initial tree
pruned.tree2 <- prune(model2, cp = model2$cptable[which.min(model2$cptable[,"xerror"]),"CP"])
printcp(pruned.tree2)
```

```
##
## Regression tree:
## rpart(formula = GWP ~ ., data = dfgwp, method = "anova")
##
## Variables actually used in tree construction:
## [1] HeaterFuel           SupplyFans           System1HeatingSource
##
## Root node error: 9.1653e+10/80 = 1145662344
##
## n= 80
##
##         CP nsplit rel error  xerror     xstd
## 1 0.554123      0   1.00000 1.03225 0.180419
## 2 0.118245      1   0.44588 0.47154 0.106239
## 3 0.065792      2   0.32763 0.41749 0.088727
## 4 0.058265      3   0.26184 0.35501 0.079769
## 5 0.010854      4   0.20358 0.24552 0.054833
```

```
#Plotting the pruned tree
rpart.plot(pruned.tree2)
```




```
#The pruned tree is same as the initially fitted tree
printcp(pruned.tree2)
```

```
##
## Regression tree:
## rpart(formula = GWP ~ ., data = dfgwp, method = "anova")
##
## Variables actually used in tree construction:
## [1] HeaterFuel           SupplyFans           System1HeatingSource
##
## Root node error: 9.1653e+10/80 = 1145662344
##
## n= 80
##
##         CP nsplit rel error  xerror     xstd
## 1 0.554123      0   1.00000 1.03225 0.180419
## 2 0.118245      1   0.44588 0.47154 0.106239
## 3 0.065792      2   0.32763 0.41749 0.088727
## 4 0.058265      3   0.26184 0.35501 0.079769
## 5 0.010854      4   0.20358 0.24552 0.054833
```

```
pruned.tree2
```

```
## n= 80
##
## node), split, n, deviance, yval
##       * denotes terminal node
```

```
##
##  1) root 80 91652990000 195012.5
##    2) System1HeatingSource=DX Coils 56 19669980000 178517.9
##      4) SupplyFans=Forward Curved Centrifugal w/ Inlet Vanes 32  1933969000 169531.2 *
##      5) SupplyFans=Two-Speed,Variable 24 11706000000 190500.0
##       10) HeaterFuel=Gas 12  2544917000 175583.3 *
##       11) HeaterFuel=Electricity 12  3820917000 205416.7 *
##    3) System1HeatingSource=Electric Resistance 24 21196000000 233500.0
##      6) HeaterFuel=Gas 12  4450250000 212250.0 *
##      7) HeaterFuel=Electricity 12  5908250000 254750.0 *
```