Energy Efficiency Report

Study of Dataset:

If we notice the two dependant variables, we see that they have a correlation and almost form a linear relationship.

Step 1: We load the data set into R

We read the file in R with below command :

```
setwd("C:/Users/Trupti/Desktop")

require(XLConnect)

EnergyEfficiency = loadWorkbook("EnergyEfficiency.xlsx")

EnergyEfficiencyanalysis = readWorksheet(EnergyEfficiency, sheet = "EnergyEfficiency", header = TRUE)

EnergyEfficiencyanalysis

cor(EnergyEfficiencyanalysis)
```

```
                X1            X2         X3            X4         X5
   X6
X1  1.000000e+00 -9.919015e-01 -0.2037817 -8.688234e-01  0.8277473  0.0000
00000
X2 -9.919015e-01  1.000000e+00  0.1955016  8.807195e-01 -0.8581477  0.0000
00000
X3 -2.037817e-01  1.955016e-01  1.0000000 -2.923165e-01  0.2809757  0.0000
00000
X4 -8.688234e-01  8.807195e-01 -0.2923165  1.000000e+00 -0.9725122  0.0000
00000
X5  8.277473e-01 -8.581477e-01  0.2809757 -9.725122e-01  1.0000000  0.0000
00000
X6  0.000000e+00  0.000000e+00  0.0000000  0.000000e+00  0.0000000  1.0000
00000
X7  7.617400e-20  4.664140e-20  0.0000000 -1.197187e-19  0.0000000  0.0000
00000
X8  0.000000e+00  0.000000e+00  0.0000000  0.000000e+00  0.0000000  0.0000
00000
Y1  6.222719e-01 -6.581199e-01  0.4556714 -8.618281e-01  0.8894305 -0.0025
86763
Y2  6.343391e-01 -6.729989e-01  0.4271170 -8.625466e-01  0.8957852  0.0142
89598
               X7          X8           Y1          Y2
X1  7.617400e-20 0.00000000  0.622271936  0.63433907
X2  4.664140e-20 0.00000000 -0.658119917 -0.67299893
X3  0.000000e+00 0.00000000  0.455671365  0.42711700
X4 -1.197187e-19 0.00000000 -0.861828052 -0.86254660
X5  0.000000e+00 0.00000000  0.889430464  0.89578517
X6  0.000000e+00 0.00000000 -0.002586763  0.01428960
X7  1.000000e+00 0.21296422  0.269841685  0.20750499
X8  2.129642e-01 1.00000000  0.087368460  0.05052512
Y1  2.698417e-01 0.08736846  1.000000000  0.97586174
Y2  2.075050e-01 0.05052512  0.975861739  1.00000000
```

Step 2

a.) For Yield 1 :

We take decision tree regression using rpart for dependant variable Y1 as below :

DescTreeRegressionY1<-rpart(Y1~X1+X2+X3+X4+X5+X6+X7+X8, data=EnergyEfficiencyanalysis, method = "anova")

summary(DescTreeRegressionY1)

#Summary Results in Console

```
Call:
rpart(formula = Y1 ~ X1 + X2 + X3 + X4 + X5 + X6 + X7 + X8, data = EnergyE
fficiencyanalysis,
    method = "anova")
  n= 768

          CP nsplit  rel error     xerror        xstd
1 0.79108655      0 1.00000000 1.00217463 0.031454805
2 0.08443686      1 0.20891345 0.21060171 0.012659713
3 0.02895920      2 0.12447659 0.12571678 0.008238705
4 0.01474893      3 0.09551738 0.09680288 0.006122283
5 0.01270084      4 0.08076845 0.08705066 0.005172301
6 0.01014178      5 0.06806761 0.07174207 0.004366187
7 0.01000000      6 0.05792583 0.06715091 0.004087295

Variable importance
X1 X2 X4 X5 X3 X7 X8
23 23 21 21  9  2  1

Node number 1: 768 observations,    complexity param=0.7910866
  mean=22.3072, MSE=101.6796
  left son=2 (384 obs) right son=3 (384 obs)
  Primary splits:
      X4 < 183.75  to the right, improve=0.7910866, (0 missing)
      X5 < 5.25    to the left,  improve=0.7910866, (0 missing)
      X1 < 0.75    to the left,  improve=0.7910866, (0 missing)
      X2 < 673.75  to the right, improve=0.7910866, (0 missing)
      X3 < 281.75  to the left,  improve=0.2104531, (0 missing)
  Surrogate splits:
      X1 < 0.75    to the left,  agree=1.000, adj=1.000, (0 split)
      X2 < 673.75  to the right, agree=1.000, adj=1.000, (0 split)
      X5 < 5.25    to the left,  agree=1.000, adj=1.000, (0 split)
      X3 < 281.75  to the left,  agree=0.667, adj=0.333, (0 split)

Node number 2: 384 observations,    complexity param=0.01270084
  mean=13.33851, MSE=7.119698
  left son=4 (144 obs) right son=5 (240 obs)
  Primary splits:
      X7 < 0.175   to the left,  improve=0.3627730, (0 missing)
      X1 < 0.65    to the right, improve=0.3128954, (0 missing)
      X3 < 330.75  to the left,  improve=0.3128954, (0 missing)
      X2 < 771.75  to the left,  improve=0.3128954, (0 missing)
      X8 < 0.5     to the left,  improve=0.3093992, (0 missing)
  Surrogate splits:
```

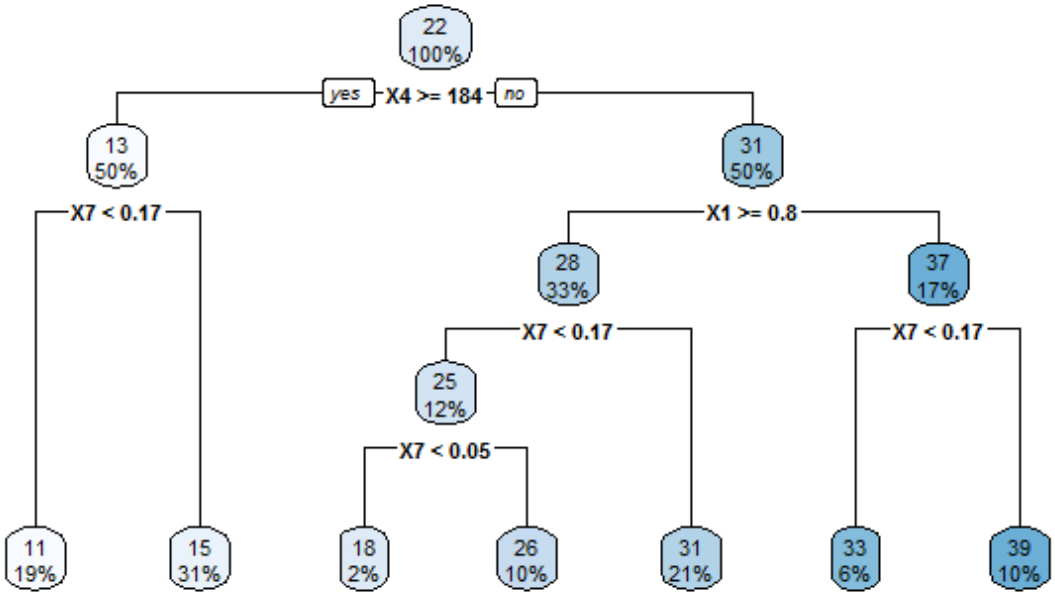
```
      X8 < 0.5     to the left,  agree=0.688, adj=0.167, (0 split)

Node number 3: 384 observations,    complexity param=0.08443686
  mean=31.27589, MSE=35.36479
  left son=6 (256 obs) right son=7 (128 obs)
  Primary splits:
      X1 < 0.805   to the right, improve=0.4855400, (0 missing)
      X2 < 624.75  to the left,  improve=0.4855400, (0 missing)
      X3 < 330.75  to the left,  improve=0.4855400, (0 missing)
      X7 < 0.175   to the left,  improve=0.2513315, (0 missing)
      X8 < 0.5     to the left,  improve=0.1997350, (0 missing)
  Surrogate splits:
      X2 < 624.75  to the left,  agree=1, adj=1, (0 split)
      X3 < 330.75  to the left,  agree=1, adj=1, (0 split)

Node number 4: 144 observations
  mean=11.26372, MSE=5.476865

Node number 5: 240 observations
  mean=14.58337, MSE=3.972863

Node number 6: 256 observations,    complexity param=0.0289592
  mean=28.34578, MSE=19.12815
  left son=12 (96 obs) right son=13 (160 obs)
  Primary splits:
      X7 < 0.175   to the left,  improve=0.46181610, (0 missing)
      X8 < 0.5     to the left,  improve=0.36685200, (0 missing)
      X1 < 0.84    to the left,  improve=0.13569390, (0 missing)
      X2 < 600.25  to the right, improve=0.13569390, (0 missing)
      X4 < 134.75  to the right, improve=0.08723999, (0 missing)
  Surrogate splits:
      X8 < 0.5     to the left,  agree=0.688, adj=0.167, (0 split)

Node number 7: 128 observations,    complexity param=0.01474893
  mean=37.13609, MSE=16.32503
  left son=14 (48 obs) right son=15 (80 obs)
  Primary splits:
      X7 < 0.175   to the left,  improve=0.5511780, (0 missing)
      X8 < 0.5     to the left,  improve=0.4383955, (0 missing)
      X3 < 379.75  to the right, improve=0.1327052, (0 missing)
      X4 < 134.75  to the left,  improve=0.1327052, (0 missing)
      X2 < 649.25  to the right, improve=0.1327052, (0 missing)
  Surrogate splits:
      X8 < 0.5     to the left,  agree=0.688, adj=0.167, (0 split)

Node number 12: 96 observations,    complexity param=0.01014178
  mean=24.50875, MSE=12.73365
  left son=24 (16 obs) right son=25 (80 obs)
  Primary splits:
      X7 < 0.05    to the left,  improve=0.64786610, (0 missing)
      X8 < 0.5     to the left,  improve=0.64786610, (0 missing)
      X2 < 600.25  to the right, improve=0.10668190, (0 missing)
      X1 < 0.84    to the left,  improve=0.10668190, (0 missing)
      X4 < 116.375 to the left,  improve=0.07098019, (0 missing)
  Surrogate splits:
      X8 < 0.5     to the left,  agree=1, adj=1, (0 split)

Node number 13: 160 observations
  mean=30.648, MSE=8.830945

Node number 14: 48 observations
  mean=33.26354, MSE=12.453
```

```
Node number 15: 80 observations
  mean=39.45963, MSE=4.251449

Node number 24: 16 observations
  mean=18.08625, MSE=4.279523

Node number 25: 80 observations
  mean=25.79325, MSE=4.524837
```

Upon plotting the graph in R, we get plot as below :

prp(DescTreeRegressionY1)

rpart.plot(DescTreeRegressionY1)



b.) For Yield 2:

We take decision tree regression using rpart for dependant variable Y2 as below :

DescTreeRegressionY2<-rpart(Y2~X1+X2+X3+X4+X5+X6+X7+X8, data=EnergyEfficiencyanalysis, method = "anova")

summary(DescTreeRegressionY2)summary(DescTreeRegressionY2)

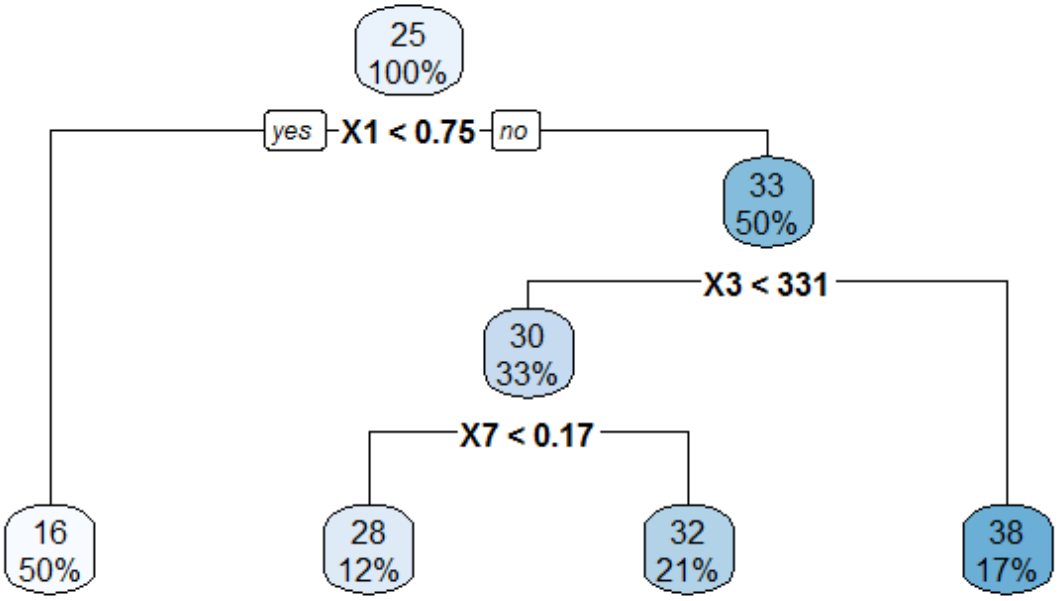
#Summary Results in Console

```
Call:
rpart(formula = Y2 ~ X1 + X2 + X3 + X4 + X5 + X6 + X7 + X8, data = EnergyE
fficiencyanalysis,
    method = "anova")
```

```
  n= 768

          CP nsplit rel error    xerror        xstd
1 0.80243107      0 1.0000000 1.0008901 0.033330398
2 0.07534417      1 0.1975689 0.1991311 0.012655022
3 0.01885084      2 0.1222248 0.1236709 0.006822588
4 0.01000000      3 0.1033739 0.1053744 0.005705307

Variable importance
X1 X2 X4 X5 X3 X7
24 24 22 22  9  1

Node number 1: 768 observations,    complexity param=0.8024311
  mean=24.58776, MSE=90.38514
  left son=2 (384 obs) right son=3 (384 obs)
  Primary splits:
      X1 < 0.75    to the left,  improve=0.8024311, (0 missing)
      X5 < 5.25    to the left,  improve=0.8024311, (0 missing)
      X2 < 673.75  to the right, improve=0.8024311, (0 missing)
      X4 < 183.75  to the right, improve=0.8024311, (0 missing)
      X3 < 281.75  to the left,  improve=0.2067026, (0 missing)
  Surrogate splits:
      X2 < 673.75  to the right, agree=1.000, adj=1.000, (0 split)
      X4 < 183.75  to the right, agree=1.000, adj=1.000, (0 split)
      X5 < 5.25    to the left,  agree=1.000, adj=1.000, (0 split)
      X3 < 281.75  to the left,  agree=0.667, adj=0.333, (0 split)

Node number 2: 384 observations
  mean=16.07143, MSE=5.844996

Node number 3: 384 observations,    complexity param=0.07534417
  mean=33.10409, MSE=29.8696
  left son=6 (256 obs) right son=7 (128 obs)
  Primary splits:
      X3 < 330.75  to the left,  improve=0.4559816, (0 missing)
      X2 < 624.75  to the left,  improve=0.4559816, (0 missing)
      X1 < 0.805   to the right, improve=0.4559816, (0 missing)
      X7 < 0.175   to the left,  improve=0.1645447, (0 missing)
      X4 < 116.375 to the left,  improve=0.1011679, (0 missing)
  Surrogate splits:
      X1 < 0.805   to the right, agree=1, adj=1, (0 split)
      X2 < 624.75  to the left,  agree=1, adj=1, (0 split)

Node number 6: 256 observations,    complexity param=0.01885084
  mean=30.49449, MSE=15.61716
  left son=12 (96 obs) right son=13 (160 obs)
  Primary splits:
      X7 < 0.175   to the left,  improve=0.32730070, (0 missing)
      X8 < 0.5     to the left,  improve=0.17424270, (0 missing)
      X1 < 0.84    to the left,  improve=0.12983510, (0 missing)
      X2 < 600.25  to the right, improve=0.12983510, (0 missing)
      X4 < 134.75  to the right, improve=0.06719687, (0 missing)
  Surrogate splits:
      X8 < 0.5     to the left,  agree=0.688, adj=0.167, (0 split)

Node number 7: 128 observations
  mean=38.32328, MSE=17.51451

Node number 12: 96 observations
  mean=27.57573, MSE=10.59244

Node number 13: 160 observations
```

mean=32.24575, MSE=10.45358

Upon plotting the graph in R, we get plot as below :



Conclusion :

It is observed from the correlation that the Dependant variables Y1 and Y2 bear a linear relationship . The independent variables X3,X4,X5 and X7 also bear a relationship between each other(as per correlation). Hence the behaviour of the dependant variables would change as per changes in these independent variables.