DS_Kaggle_BikeShare_OLS

Divya

6/10/2017

```
# cat("\014")
setwd("~/Desktop/MIDS/DivyaGitHub/TpT-BikeShareKaggle/")

#libraries
library(car)
library(lmtest) #regression with heteroskadasticity
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
library(sandwich) #regression with heteroskadasticity
library(stargazer)
```

```
##
## Please cite as:
```

```
##  Hlavac, Marek (2015). stargazer: Well-Formatted Regression and Summary Statistics Tables.
```

```
##  R package version 5.2. http://CRAN.R-project.org/package=stargazer
```

```
library(effsize) #for cohen's d (practical significance)
library(rpart) #for tree
library(Metrics) #for rmsle
```

```
## Warning: package 'Metrics' was built under R version 3.3.2
```

```
library(party)
```

```
## Warning: package 'party' was built under R version 3.3.2
```

```
## Loading required package: grid
```

```
## Loading required package: mvtnorm
```

```
## Loading required package: modeltools
```

```
## Loading required package: stats4
```

```
## Loading required package: strucchange
```

```
train = read.csv("train.csv", sep = ',')
train_data = read.csv("train_data.csv", sep = ',')
dev_data = read.csv("dev_data.csv", sep = ',')
test_data = read.csv("test_data.csv", sep = ',')
```

RPART MODEL

Using rpart (recursive partitioning and regression trees)

RPART Train Data

Let's try use the rpart model to train with our train_data set.

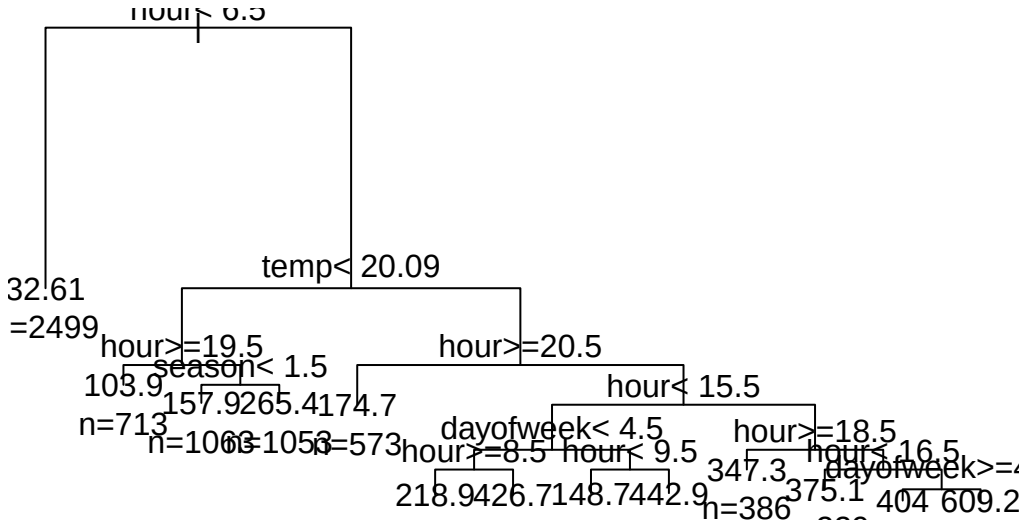
```
# choosing the variables to include in the model
formula_rpart = count ~ hour + temp + humidity + season + weather + dayofweek

# fitting forumula to the model
fit_rpart = rpart(formula_rpart, data=train_data)

# tells us the importance of each variable in the model
fit_rpart
```

```
## n= 8708
##
## node), split, n, deviance, yval
##       * denotes terminal node
##
##   1) root 8708 284496500.0 191.19990
##     2) hour< 6.5 2499   3773467.0  32.60624 *
##     3) hour>=6.5 6209 192570500.0 255.03080
##       6) temp< 20.09 2829  55691080.0 184.31810
##        12) hour>=19.5 713   3039984.0 103.92990 *
##        13) hour< 19.5 2116  46490930.0 211.40550
##          26) season< 1.5 1063  13997170.0 157.89930 *
##          27) season>=1.5 1053  26378320.0 265.41980 *
##       7) temp>=20.09 3380 110893800.0 314.21600
##        14) hour>=20.5 573   3637662.0 174.69110 *
##        15) hour< 20.5 2807  93824440.0 342.69750
##          30) hour< 15.5 1753  41616670.0 285.56420
##            60) dayofweek< 4.5 1290  23404730.0 258.05040
##             120) hour>=8.5 1047   7353893.0 218.89880 *
##             121) hour< 8.5 243   7531033.0 426.74070 *
##            61) dayofweek>=4.5 463  14514570.0 362.22250
##             122) hour< 9.5 127    967599.4 148.68500 *
##             123) hour>=9.5 336   5567149.0 442.93450 *
##          31) hour>=15.5 1054  36968530.0 437.72110
##            62) hour>=18.5 386   6036184.0 347.33160 *
##            63) hour< 18.5 668  25956270.0 489.95210
##             126) hour< 16.5 229   3876739.0 375.12230 *
##             127) hour>=16.5 439  17484840.0 549.85190
##               254) dayofweek>=4.5 127   1954156.0 404.01570 *
##               255) dayofweek< 4.5 312  11730150.0 609.21470 *
```

```
plot(fit_rpart)
text(fit_rpart, use.n=TRUE)
```



According to this model, the most important factor is hour (biggest split).

RPART Predict With Dev Data Set

Let's try use the rpart model to predict with our dev_data set. And then we can calculate rmsle to evaluate our model.

```
#dev_data
predict_rpart_dev = predict(fit_rpart, dev_data)

# putting our predictions + hours into dataframe
submit_rpart_dev = data.frame(datetime = dev_data$datetime, count=predict_rpart_dev)

# writing the dataframe to a csv file --> submit to kaggle
write.csv(submit_rpart_dev, file="submit_rpart_dev_v1.csv",row.names=FALSE)

#checking root mean squared log error (like the evaluation in kaggle)
rmsle(dev_data$count, abs(predict_rpart_dev))
```

```
## [1] 0.8729415
```

RPART Predict With Test Data

Let's try use the rpart model to predict with our test_data set. We'll save the predictions for the test_data set along with the datetime column as a dataframe and convert and save that into a csv file to upload to kaggle.

```
#test_data
predict_rpart_test = predict(fit_rpart, test_data)

# putting our predictions + hours into dataframe
```

```
submit_rpart_test = data.frame(datetime = test_data$datetime, count=predict_rpart_test)

# writing the dataframe to a csv file --> submit to kaggle
write.csv(submit_rpart_test, file="submit_rpart_test_v1.csv",row.names=FALSE)
```

PARTY MODEL

PARTY Train Data

Let's try use the party model to train with our train_data set.

Using party (recursive partitioning and regression trees)

```
formula_ctree = count ~ hour + temp + humidity + season + weather + dayofweek

#fitting forumula to the model
fit_ctree = ctree(formula_ctree, data=train_data)

#tells us the importance of each variable in the model
fit_ctree
```

```
##
##   Conditional inference tree with 165 terminal nodes
##
## Response:  count
## Inputs:  hour, temp, humidity, season, weather, dayofweek
## Number of observations:  8708
##
## 1) temp <= 22.14; criterion = 1, statistic = 1360.418
##   2) hour <= 6; criterion = 1, statistic = 567.471
##     3) temp <= 10.66; criterion = 1, statistic = 63.641
##       4) season <= 2; criterion = 1, statistic = 33.725
##         5) hour <= 5; criterion = 1, statistic = 18.877
##           6) dayofweek <= 4; criterion = 1, statistic = 45.446
##             7) hour <= 0; criterion = 0.986, statistic = 9.277
##               8)*  weights = 34
##             7) hour > 0
##               9)*  weights = 174
##           6) dayofweek > 4
##             10) hour <= 2; criterion = 1, statistic = 50.052
##               11)*  weights = 38
##             10) hour > 2
##               12) hour <= 3; criterion = 0.997, statistic = 12.295
##                 13)*  weights = 15
##               12) hour > 3
##                 14)*  weights = 37
##         5) hour > 5
##           15) dayofweek <= 4; criterion = 1, statistic = 16.914
##             16)*  weights = 50
##           15) dayofweek > 4
##             17)*  weights = 16
##       4) season > 2
```

```
##         18)*  weights = 156
##     3) temp > 10.66
##       19) weather <= 2; criterion = 1, statistic = 20.167
##         20) dayofweek <= 4; criterion = 0.998, statistic = 12.964
##           21) hour <= 5; criterion = 1, statistic = 147.078
##             22) hour <= 0; criterion = 1, statistic = 40.769
##               23)*  weights = 98
##             22) hour > 0
##               24) hour <= 4; criterion = 1, statistic = 17.883
##                 25) hour <= 1; criterion = 1, statistic = 77.355
##                   26)*  weights = 103
##                 25) hour > 1
##                   27) hour <= 2; criterion = 1, statistic = 20.967
##                     28)*  weights = 106
##                   27) hour > 2
##                     29) season <= 1; criterion = 1, statistic = 23.034
##                       30)*  weights = 36
##                     29) season > 1
##                       31)*  weights = 169
##               24) hour > 4
##                 32) season <= 2; criterion = 0.998, statistic = 13.183
##                   33)*  weights = 67
##                 32) season > 2
##                   34)*  weights = 51
##           21) hour > 5
##             35) temp <= 18.04; criterion = 1, statistic = 16.011
##               36) season <= 3; criterion = 0.996, statistic = 11.798
##                 37)*  weights = 42
##               36) season > 3
##                 38)*  weights = 40
##             35) temp > 18.04
##               39) weather <= 1; criterion = 0.963, statistic = 7.455
##                 40)*  weights = 28
##               39) weather > 1
##                 41)*  weights = 8
##         20) dayofweek > 4
##           42) hour <= 2; criterion = 1, statistic = 177.241
##             43) hour <= 0; criterion = 1, statistic = 24.732
##               44) temp <= 19.68; criterion = 0.991, statistic = 10.131
##                 45) season <= 2; criterion = 0.998, statistic = 13.292
##                   46)*  weights = 21
##                 45) season > 2
##                   47)*  weights = 20
##               44) temp > 19.68
##                 48)*  weights = 8
##             43) hour > 0
##               49) temp <= 19.68; criterion = 0.999, statistic = 14.324
##                 50) season <= 2; criterion = 1, statistic = 17.033
##                   51)*  weights = 43
##                 50) season > 2
##                   52) hour <= 1; criterion = 0.977, statistic = 8.328
##                     53)*  weights = 18
##                   52) hour > 1
##                     54)*  weights = 20
```

```
##               49) temp > 19.68
##                 55)*  weights = 16
##           42) hour > 2
##             56)*  weights = 177
##       19) weather > 2
##         57) hour <= 5; criterion = 0.999, statistic = 14.369
##           58) dayofweek <= 4; criterion = 1, statistic = 16.64
##             59)*  weights = 77
##           58) dayofweek > 4
##             60)*  weights = 10
##         57) hour > 5
##           61)*  weights = 19
##   2) hour > 6
##     62) temp <= 13.94; criterion = 1, statistic = 404.885
##       63) season <= 2; criterion = 1, statistic = 184.529
##         64) hour <= 19; criterion = 1, statistic = 106.236
##           65) temp <= 9.84; criterion = 1, statistic = 29.2
##             66) dayofweek <= 4; criterion = 0.996, statistic = 11.704
##               67) hour <= 9; criterion = 1, statistic = 16.951
##                 68)*  weights = 123
##               67) hour > 9
##                 69) hour <= 16; criterion = 1, statistic = 56.314
##                   70)*  weights = 134
##                 69) hour > 16
##                   71) hour <= 18; criterion = 0.998, statistic = 12.621
##                     72)*  weights = 40
##                   71) hour > 18
##                     73)*  weights = 19
##             66) dayofweek > 4
##               74) hour <= 9; criterion = 1, statistic = 26.484
##                 75) hour <= 7; criterion = 0.999, statistic = 14.401
##                   76)*  weights = 19
##                 75) hour > 7
##                   77)*  weights = 30
##               74) hour > 9
##                 78)*  weights = 61
##           65) temp > 9.84
##             79) weather <= 2; criterion = 0.998, statistic = 12.976
##               80) hour <= 8; criterion = 0.964, statistic = 7.499
##                 81) dayofweek <= 4; criterion = 1, statistic = 15.823
##                   82) hour <= 7; criterion = 1, statistic = 16.093
##                     83)*  weights = 23
##                   82) hour > 7
##                     84)*  weights = 18
##                 81) dayofweek > 4
##                   85)*  weights = 14
##               80) hour > 8
##                 86)*  weights = 279
##             79) weather > 2
##               87)*  weights = 38
##         64) hour > 19
##           88) hour <= 21; criterion = 1, statistic = 69.831
##             89) temp <= 11.48; criterion = 1, statistic = 18.168
##               90) dayofweek <= 4; criterion = 0.97, statistic = 7.883
```

```
##                 91) temp <= 9.84; criterion = 0.989, statistic = 9.635
##                   92) hour <= 20; criterion = 0.955, statistic = 7.107
##                     93)*  weights = 25
##                   92) hour > 20
##                     94)*  weights = 22
##                 91) temp > 9.84
##                   95)*  weights = 9
##               90) dayofweek > 4
##                 96)*  weights = 23
##             89) temp > 11.48
##               97)*  weights = 49
##           88) hour > 21
##             98) temp <= 9.02; criterion = 1, statistic = 22.461
##               99)*  weights = 65
##             98) temp > 9.02
##               100) hour <= 22; criterion = 1, statistic = 17.253
##                 101)*  weights = 45
##               100) hour > 22
##                 102)*  weights = 42
##       63) season > 2
##         103) dayofweek <= 4; criterion = 1, statistic = 40.02
##           104) hour <= 8; criterion = 1, statistic = 38.767
##             105) hour <= 7; criterion = 1, statistic = 28.241
##               106)*  weights = 28
##             105) hour > 7
##               107)*  weights = 25
##           104) hour > 8
##             108) humidity <= 61; criterion = 0.999, statistic = 13.932
##               109) dayofweek <= 3; criterion = 0.965, statistic = 7.565
##                 110)*  weights = 141
##               109) dayofweek > 3
##                 111)*  weights = 47
##             108) humidity > 61
##               112) hour <= 9; criterion = 0.999, statistic = 14.343
##                 113)*  weights = 10
##               112) hour > 9
##                 114) dayofweek <= 2; criterion = 0.99, statistic = 9.841
##                   115)*  weights = 19
##                 114) dayofweek > 2
##                   116) temp <= 12.3; criterion = 0.988, statistic = 9.547
##                     117)*  weights = 10
##                   116) temp > 12.3
##                     118)*  weights = 13
##         103) dayofweek > 4
##           119) temp <= 10.66; criterion = 1, statistic = 31.259
##             120) temp <= 9.02; criterion = 0.979, statistic = 8.486
##               121)*  weights = 17
##             120) temp > 9.02
##               122)*  weights = 34
##           119) temp > 10.66
##             123)*  weights = 86
##     62) temp > 13.94
##       124) weather <= 2; criterion = 1, statistic = 98.318
##         125) hour <= 19; criterion = 1, statistic = 88.66
```

```
##           126) season <= 2; criterion = 1, statistic = 76.966
##             127) temp <= 16.4; criterion = 1, statistic = 40.556
##               128) weather <= 1; criterion = 0.995, statistic = 11.204
##                 129) hour <= 15; criterion = 0.954, statistic = 7.092
##                   130)*  weights = 83
##                 129) hour > 15
##                   131)*  weights = 44
##               128) weather > 1
##                 132)*  weights = 70
##             127) temp > 16.4
##               133) weather <= 1; criterion = 0.998, statistic = 12.977
##                 134) season <= 1; criterion = 0.974, statistic = 8.099
##                   135) humidity <= 72; criterion = 0.997, statistic = 12.369
##                     136) hour <= 16; criterion = 0.999, statistic = 14.915
##                       137) dayofweek <= 4; criterion = 1, statistic = 18.868
##                         138) humidity <= 63; criterion = 0.967, statistic = 7.659
##                           139) hour <= 15; criterion = 0.975, statistic = 8.216
##                             140)*  weights = 39
##                           139) hour > 15
##                             141)*  weights = 10
##                         138) humidity > 63
##                           142)*  weights = 10
##                       137) dayofweek > 4
##                         143) temp <= 19.68; criterion = 0.998, statistic = 12.581
##                           144)*  weights = 26
##                         143) temp > 19.68
##                           145)*  weights = 9
##                     136) hour > 16
##                       146)*  weights = 41
##                   135) humidity > 72
##                     147)*  weights = 14
##                 134) season > 1
##                   148) hour <= 16; criterion = 0.986, statistic = 9.304
##                     149)*  weights = 139
##                   148) hour > 16
##                     150) humidity <= 35; criterion = 0.996, statistic = 11.521
##                       151)*  weights = 21
##                     150) humidity > 35
##                       152)*  weights = 14
##               133) weather > 1
##                 153)*  weights = 148
##           126) season > 2
##             154) hour <= 15; criterion = 1, statistic = 20.421
##               155)*  weights = 411
##             154) hour > 15
##               156) dayofweek <= 4; criterion = 1, statistic = 15.8
##                 157)*  weights = 126
##               156) dayofweek > 4
##                 158) hour <= 17; criterion = 1, statistic = 23.611
##                   159) temp <= 18.86; criterion = 0.951, statistic = 6.959
##                     160)*  weights = 19
##                   159) temp > 18.86
##                     161)*  weights = 11
##                 158) hour > 17
```

```
##                   162) temp <= 17.22; criterion = 0.963, statistic = 7.478
##                     163)*  weights = 22
##                   162) temp > 17.22
##                     164)*  weights = 15
##         125) hour > 19
##           165) hour <= 21; criterion = 1, statistic = 164.577
##             166) dayofweek <= 4; criterion = 1, statistic = 30.937
##               167) temp <= 18.04; criterion = 1, statistic = 20.694
##                 168) season <= 1; criterion = 0.987, statistic = 9.369
##                   169)*  weights = 21
##                 168) season > 1
##                   170) humidity <= 66; criterion = 0.998, statistic = 12.982
##                     171)*  weights = 29
##                   170) humidity > 66
##                     172)*  weights = 24
##               167) temp > 18.04
##                 173) hour <= 20; criterion = 0.999, statistic = 15.059
##                   174)*  weights = 35
##                 173) hour > 20
##                   175)*  weights = 42
##             166) dayofweek > 4
##               176)*  weights = 66
##           165) hour > 21
##             177) hour <= 22; criterion = 1, statistic = 35.143
##               178) temp <= 18.04; criterion = 0.985, statistic = 9.081
##                 179)*  weights = 50
##               178) temp > 18.04
##                 180)*  weights = 56
##             177) hour > 22
##               181) temp <= 18.86; criterion = 0.955, statistic = 7.122
##                 182) season <= 2; criterion = 0.986, statistic = 9.274
##                   183)*  weights = 44
##                 182) season > 2
##                   184)*  weights = 30
##               181) temp > 18.86
##                 185)*  weights = 40
##       124) weather > 2
##         186)*  weights = 205
## 1) temp > 22.14
##   187) hour <= 6; criterion = 1, statistic = 743.715
##     188) dayofweek <= 4; criterion = 1, statistic = 44.625
##       189) hour <= 5; criterion = 1, statistic = 46.692
##         190) hour <= 0; criterion = 1, statistic = 94.519
##           191) dayofweek <= 3; criterion = 0.999, statistic = 15.371
##             192)*  weights = 79
##           191) dayofweek > 3
##             193)*  weights = 23
##         190) hour > 0
##           194)*  weights = 408
##       189) hour > 5
##         195)*  weights = 68
##     188) dayofweek > 4
##       196) hour <= 2; criterion = 1, statistic = 156.598
##         197) hour <= 0; criterion = 1, statistic = 50.139
```

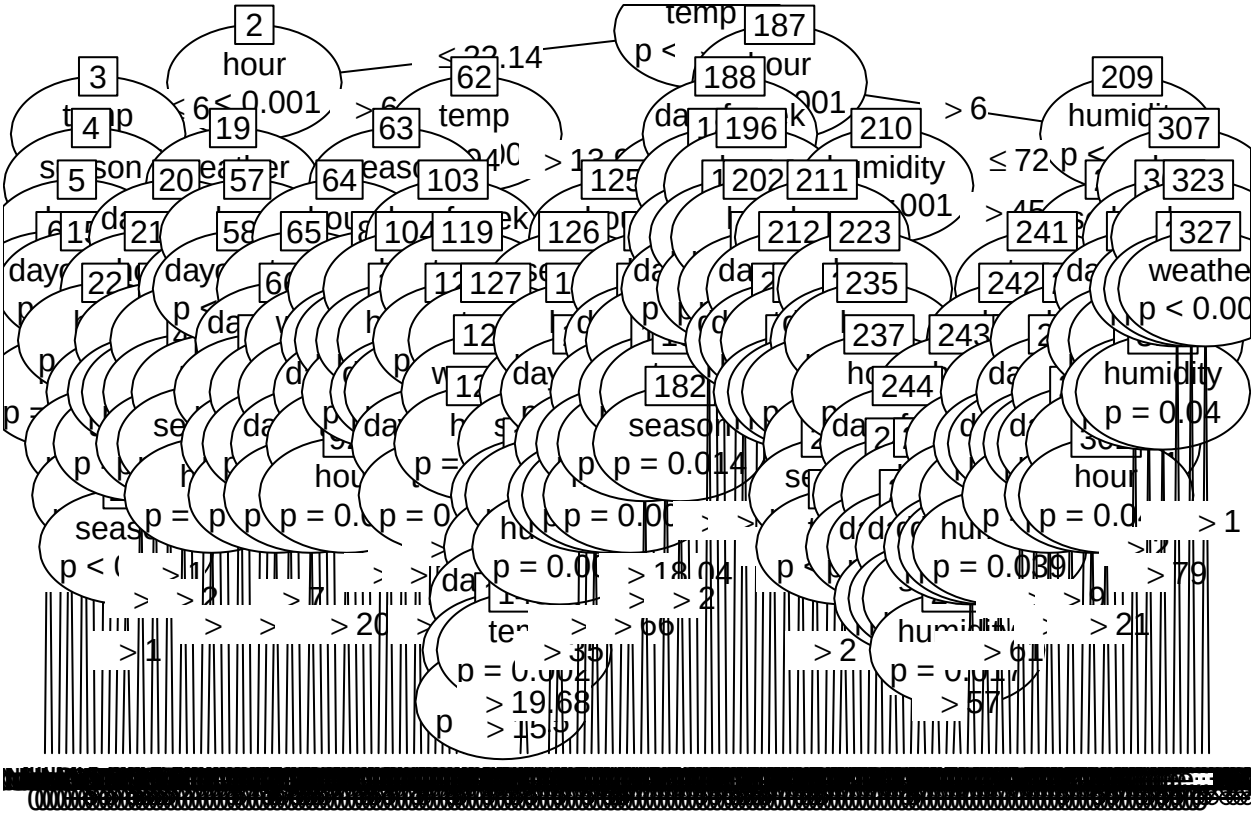
```
##           198)*  weights = 36
##         197) hour > 0
##           199) hour <= 1; criterion = 1, statistic = 17.082
##             200)*  weights = 38
##           199) hour > 1
##             201)*  weights = 34
##       196) hour > 2
##         202) hour <= 3; criterion = 0.985, statistic = 9.131
##           203)*  weights = 33
##         202) hour > 3
##           204) hour <= 5; criterion = 1, statistic = 24.635
##             205)*  weights = 56
##           204) hour > 5
##             206) dayofweek <= 5; criterion = 0.987, statistic = 9.353
##               207)*  weights = 13
##             206) dayofweek > 5
##               208)*  weights = 14
##   187) hour > 6
##     209) humidity <= 72; criterion = 1, statistic = 221.214
##       210) humidity <= 45; criterion = 1, statistic = 39.556
##         211) hour <= 15; criterion = 1, statistic = 57.047
##           212) dayofweek <= 4; criterion = 1, statistic = 186.633
##             213) dayofweek <= 3; criterion = 1, statistic = 16.861
##               214)*  weights = 271
##             213) dayofweek > 3
##               215)*  weights = 67
##           212) dayofweek > 4
##             216) temp <= 31.98; criterion = 1, statistic = 16.335
##               217) hour <= 11; criterion = 0.998, statistic = 12.693
##                 218)*  weights = 14
##               217) hour > 11
##                 219)*  weights = 59
##             216) temp > 31.98
##               220) season <= 2; criterion = 0.999, statistic = 14.154
##                 221)*  weights = 8
##               220) season > 2
##                 222)*  weights = 37
##         211) hour > 15
##           223) hour <= 19; criterion = 1, statistic = 54.948
##             224) dayofweek <= 3; criterion = 0.992, statistic = 10.31
##               225) hour <= 16; criterion = 1, statistic = 21.388
##                 226)*  weights = 55
##               225) hour > 16
##                 227) hour <= 18; criterion = 1, statistic = 18.684
##                   228) season <= 2; criterion = 0.975, statistic = 8.171
##                     229)*  weights = 50
##                   228) season > 2
##                     230) temp <= 28.7; criterion = 0.999, statistic = 14.273
##                       231)*  weights = 9
##                     230) temp > 28.7
##                       232)*  weights = 33
##                 227) hour > 18
##                   233)*  weights = 25
##             224) dayofweek > 3
```

```
##               234)*  weights = 131
##           223) hour > 19
##             235) hour <= 20; criterion = 1, statistic = 30.338
##               236)*  weights = 36
##             235) hour > 20
##               237) hour <= 21; criterion = 0.99, statistic = 9.8
##                 238)*  weights = 23
##               237) hour > 21
##                 239)*  weights = 16
##       210) humidity > 45
##         240) season <= 3; criterion = 0.978, statistic = 8.454
##           241) temp <= 29.52; criterion = 0.997, statistic = 12.063
##             242) hour <= 20; criterion = 0.963, statistic = 7.477
##               243) hour <= 15; criterion = 1, statistic = 48.333
##                 244) dayofweek <= 3; criterion = 1, statistic = 23.704
##                   245) hour <= 8; criterion = 1, statistic = 43.305
##                     246) hour <= 7; criterion = 0.983, statistic = 8.885
##                       247)*  weights = 16
##                     246) hour > 7
##                       248)*  weights = 18
##                   245) hour > 8
##                     249)*  weights = 160
##                 244) dayofweek > 3
##                   250) hour <= 9; criterion = 1, statistic = 24.036
##                     251) dayofweek <= 4; criterion = 1, statistic = 17.44
##                       252)*  weights = 17
##                     251) dayofweek > 4
##                       253) hour <= 8; criterion = 1, statistic = 21.765
##                         254)*  weights = 19
##                       253) hour > 8
##                         255)*  weights = 14
##                   250) hour > 9
##                     256) dayofweek <= 4; criterion = 1, statistic = 27.918
##                       257) hour <= 12; criterion = 0.996, statistic = 11.62
##                         258)*  weights = 15
##                       257) hour > 12
##                         259)*  weights = 9
##                     256) dayofweek > 4
##                       260) season <= 2; criterion = 0.998, statistic = 13.052
##                         261)*  weights = 36
##                       260) season > 2
##                         262) humidity <= 57; criterion = 0.983, statistic = 8.919
##                           263)*  weights = 8
##                         262) humidity > 57
##                           264)*  weights = 16
##               243) hour > 15
##                 265) hour <= 18; criterion = 0.996, statistic = 11.727
##                   266) hour <= 16; criterion = 0.994, statistic = 10.909
##                     267) dayofweek <= 3; criterion = 0.988, statistic = 9.587
##                       268)*  weights = 16
##                     267) dayofweek > 3
##                       269)*  weights = 12
##                   266) hour > 16
##                     270) dayofweek <= 4; criterion = 0.997, statistic = 12.231
```

```
##                     271)*  weights = 44
##                   270) dayofweek > 4
##                     272)*  weights = 19
##                 265) hour > 18
##                   273) hour <= 19; criterion = 0.998, statistic = 13.346
##                     274) humidity <= 61; criterion = 0.961, statistic = 7.39
##                       275)*  weights = 18
##                     274) humidity > 61
##                       276)*  weights = 17
##                   273) hour > 19
##                     277)*  weights = 35
##               242) hour > 20
##                 278) hour <= 22; criterion = 1, statistic = 66.156
##                   279) hour <= 21; criterion = 1, statistic = 16.017
##                     280)*  weights = 48
##                   279) hour > 21
##                     281)*  weights = 53
##                 278) hour > 22
##                   282) dayofweek <= 2; criterion = 0.955, statistic = 7.095
##                     283)*  weights = 22
##                   282) dayofweek > 2
##                     284)*  weights = 32
##             241) temp > 29.52
##               285) hour <= 15; criterion = 0.999, statistic = 15.122
##                 286) dayofweek <= 4; criterion = 1, statistic = 23.613
##                   287) hour <= 8; criterion = 1, statistic = 27.072
##                     288) hour <= 7; criterion = 0.981, statistic = 8.729
##                       289)*  weights = 11
##                     288) hour > 7
##                       290)*  weights = 18
##                   287) hour > 8
##                     291)*  weights = 175
##                 286) dayofweek > 4
##                   292) hour <= 9; criterion = 1, statistic = 44.85
##                     293)*  weights = 19
##                   292) hour > 9
##                     294)*  weights = 75
##               285) hour > 15
##                 295) hour <= 20; criterion = 1, statistic = 65.236
##                   296) dayofweek <= 4; criterion = 1, statistic = 15.636
##                     297)*  weights = 132
##                   296) dayofweek > 4
##                     298) hour <= 17; criterion = 0.992, statistic = 10.318
##                       299)*  weights = 23
##                     298) hour > 17
##                       300)*  weights = 22
##                 295) hour > 20
##                   301) hour <= 22; criterion = 1, statistic = 22.339
##                     302) hour <= 21; criterion = 0.955, statistic = 7.095
##                       303)*  weights = 27
##                     302) hour > 21
##                       304)*  weights = 24
##                   301) hour > 22
##                     305)*  weights = 12
```

```
##         240) season > 3
##           306)*  weights = 136
##     209) humidity > 72
##       307) hour <= 20; criterion = 1, statistic = 42.79
##         308) humidity <= 89; criterion = 1, statistic = 27.106
##           309) dayofweek <= 4; criterion = 0.999, statistic = 13.5
##             310) weather <= 2; criterion = 1, statistic = 22.966
##               311)*  weights = 238
##             310) weather > 2
##               312)*  weights = 59
##           309) dayofweek > 4
##             313) hour <= 8; criterion = 1, statistic = 28.62
##               314) hour <= 7; criterion = 1, statistic = 18.549
##                 315)*  weights = 21
##               314) hour > 7
##                 316)*  weights = 14
##             313) hour > 8
##               317) humidity <= 79; criterion = 0.96, statistic = 7.316
##                 318)*  weights = 56
##               317) humidity > 79
##                 319)*  weights = 28
##         308) humidity > 89
##           320) season <= 3; criterion = 0.991, statistic = 10.123
##             321)*  weights = 26
##           320) season > 3
##             322)*  weights = 12
##       307) hour > 20
##         323) hour <= 22; criterion = 1, statistic = 49.608
##           324) humidity <= 79; criterion = 1, statistic = 18.989
##             325)*  weights = 65
##           324) humidity > 79
##             326)*  weights = 53
##         323) hour > 22
##           327) weather <= 1; criterion = 0.999, statistic = 15.435
##             328)*  weights = 31
##           327) weather > 1
##             329)*  weights = 32
```

```
plot(fit_ctree)
```



According to this model, the most important factor is temp (biggest split).

PARTY Predict With Dev Data Set

Let's try use the party model to predict with our dev_data set. And then we can calculate rmsle to evaluate our model.

```
#dev_data
predict_ctree_dev = predict(fit_ctree, dev_data)

# putting our predictions + hours into dataframe
submit_ctree_dev = data.frame(datetime = dev_data$datetime, count=predict_ctree_dev)

# writing the dataframe to a csv file --> submit to kaggle
write.csv(submit_ctree_dev, file="submit_ctree_dev_v1.csv",row.names=FALSE)

#checking root mean squared log error (like the evaluation in kaggle)
rmsle(dev_data$count, abs(predict_ctree_dev))
```

```
## [1] 0.5906517
```

PARTY Predict With Test Data

Let's try use the party model to predict with our test_data set. We'll save the predictions for the test_data set along with the datetime column as a dataframe and convert and save that into a csv file to upload to kaggle.

```r
#test_data
predict_ctree_test = predict(fit_ctree, test_data)

# putting our predictions + hours into dataframe
submit_ctree_test = data.frame(datetime = test_data$datetime, count=predict_ctree_test)

# writing the dataframe to a csv file --> submit to kaggle
write.csv(submit_ctree_test, file="submit_ctree_test_v1.csv",row.names=FALSE)
```