Energy Usage Analysis - Feature Engineering + Modelling + Insight Generation

2023-12-01

### Group : IST687 M004 Group 1

## Contents

### 1.Install Dependancies

### 2.Data Cleaning + Feature Engineering

### 3.Exploratory Data Analysis

### 4.Modelling

### 5.Insight Generation

### 6.Prediction - Future Energy Consumption

##—————————————————————————————————————————————————————–>

## 1.Load Required Dependancies

# Load required Libraries  
  
options(warn=-1)  
# install.packages("corrplot")  
# install.packages("catboost")  
# install.packages('devtools')  
# devtools::install\_url('BINARY\_URL'[, INSTALL\_opts = c("--no-multiarch", "--no-test-load")])  
# devtools::install\_url('https://github.com/catboost/catboost/releases/download/v1.2.2/catboost-R-Windows-1.2.2.tgz', INSTALL\_opts = c("--no-multiarch", "--no-test-load"))  
# install.packages("xgboost")  
# install.packages("shapviz")  
library(arrow)

##   
## Attaching package: 'arrow'

## The following object is masked from 'package:utils':  
##   
## timestamp

library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.2 ✔ readr 2.1.4  
## ✔ forcats 1.0.0 ✔ stringr 1.5.0  
## ✔ ggplot2 3.4.3 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.2 ✔ tidyr 1.3.0  
## ✔ purrr 1.0.2

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ lubridate::duration() masks arrow::duration()  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(lobstr)  
library(imputeTS)

## Registered S3 method overwritten by 'quantmod':  
## method from  
## as.zoo.data.frame zoo

library(curl)

## Using libcurl 8.3.0 with Schannel  
##   
## Attaching package: 'curl'  
##   
## The following object is masked from 'package:readr':  
##   
## parse\_date

library(httr)

##   
## Attaching package: 'httr'  
##   
## The following object is masked from 'package:curl':  
##   
## handle\_reset

library(xml2)  
library(aws.s3)  
library(corrplot)

## corrplot 0.92 loaded

library(xgboost)

##   
## Attaching package: 'xgboost'  
##   
## The following object is masked from 'package:dplyr':  
##   
## slice

library(readr)  
library(stringr)  
library(dplyr)  
library(caret)

## Loading required package: lattice  
##   
## Attaching package: 'caret'  
##   
## The following object is masked from 'package:httr':  
##   
## progress  
##   
## The following object is masked from 'package:purrr':  
##   
## lift

library(car)

## Loading required package: carData  
##   
## Attaching package: 'car'  
##   
## The following object is masked from 'package:dplyr':  
##   
## recode  
##   
## The following object is masked from 'package:purrr':  
##   
## some

library(catboost)  
library(recipes)

##   
## Attaching package: 'recipes'  
##   
## The following object is masked from 'package:stringr':  
##   
## fixed  
##   
## The following object is masked from 'package:stats':  
##   
## step

library(ggplot2)  
library(shapviz)  
  
  
  
# library(knitr)  
# hook\_output = knit\_hooks$get('output')  
# knit\_hooks$set(output = function(x, options) {  
# # this hook is used only when the linewidth option is not NULL  
# if (!is.null(n <- options$linewidth)) {  
# x = xfun::split\_lines(x)  
# # any lines wider than n should be wrapped  
# if (any(nchar(x) > n)) x = strwrap(x, width = n)  
# x = paste(x, collapse = '\n')  
# }  
# hook\_output(x, options)  
# })  
#   
# knitr::opts\_chunk$set(tidy.opts = list(width.cutoff = 60), tidy = TRUE)

##—————————————————————————————————————————————————————–>

## 2.Data Cleaning + Feature Engineering

### 2.1. Load Dataset

static\_house\_energy\_weather\_df <- read\_csv("C:/Users/Himanshu/OneDrive/Desktop/Syracuse Sem 1/IST-687 Intro to DS Lab/Final Project/static\_house\_energy\_weather\_df\_17130\_rows\_with\_time\_of\_day.csv",show\_col\_types = FALSE)

### 2.2. Study Unique Values in Dataset

# Unique Values  
unique\_value\_in\_each\_column <- lapply(static\_house\_energy\_weather\_df, unique)  
  
dim(static\_house\_energy\_weather\_df)

## [1] 17130 180

### 2.3.

### Removing columns based on info in metadata file and columns with only 1 value.

### Ordinal Encode Required Columns

### Remove Negative energy rows

#Define Required Columns  
  
cols\_req <- c('bldg\_id','in.county', 'in.sqft', 'in.bedrooms',  
'in.building\_america\_climate\_zone', 'in.ceiling\_fan', 'in.city', 'in.clothes\_dryer', 'in.clothes\_washer', 'in.cooking\_range', 'in.cooling\_setpoint', 'in.cooling\_setpoint\_has\_offset', 'in.cooling\_setpoint\_offset\_magnitude', 'in.cooling\_setpoint\_offset\_period',   
'in.dishwasher', 'in.ducts',   
 'in.federal\_poverty\_level', 'in.geometry\_attic\_type', 'in.geometry\_floor\_area', 'in.geometry\_floor\_area\_bin', 'in.geometry\_foundation\_type', 'in.geometry\_garage', 'in.geometry\_stories',  
 'in.geometry\_wall\_exterior\_finish', 'in.geometry\_wall\_type', 'in.has\_pv', 'in.heating\_fuel', 'in.heating\_setpoint', 'in.heating\_setpoint\_has\_offset',  
'in.heating\_setpoint\_offset\_magnitude', 'in.heating\_setpoint\_offset\_period', 'in.hot\_water\_fixtures', 'in.hvac\_cooling\_efficiency', 'in.hvac\_cooling\_partial\_space\_conditioning',  
'in.hvac\_cooling\_type', 'in.hvac\_has\_ducts', 'in.hvac\_has\_zonal\_electric\_heating', 'in.hvac\_heating\_efficiency', 'in.hvac\_heating\_type', 'in.hvac\_heating\_type\_and\_fuel',  
 'in.income', 'in.infiltration', 'in.insulation\_ceiling',   
'in.insulation\_floor', 'in.insulation\_foundation\_wall', 'in.insulation\_rim\_joist', 'in.insulation\_roof', 'in.insulation\_slab', 'in.insulation\_wall',   
'in.misc\_extra\_refrigerator', 'in.misc\_freezer', 'in.misc\_gas\_fireplace',  
'in.misc\_gas\_grill', 'in.misc\_gas\_lighting', 'in.misc\_hot\_tub\_spa', 'in.misc\_pool', 'in.misc\_pool\_heater', 'in.misc\_pool\_pump', 'in.misc\_well\_pump', 'in.natural\_ventilation', 'in.neighbors',  
'in.occupants', 'in.orientation', 'in.plug\_load\_diversity', 'in.pv\_orientation', 'in.pv\_system\_size',   
'in.refrigerator', 'in.roof\_material', 'in.tenure', 'in.units\_represented', 'in.usage\_level', 'in.vacancy\_status', 'in.vintage', 'in.vintage\_acs',  
'in.water\_heater\_efficiency', 'in.water\_heater\_fuel', 'in.window\_areas',  
'in.windows', 'upgrade.insulation\_roof', 'upgrade.water\_heater\_efficiency', 'upgrade.hvac\_cooling\_efficiency', 'upgrade.infiltration\_reduction', 'upgrade.geometry\_foundation\_type',  
'upgrade.clothes\_dryer', 'upgrade.insulation\_ceiling', 'upgrade.ducts', 'upgrade.hvac\_heating\_type', 'upgrade.insulation\_wall', 'upgrade.insulation\_foundation\_wall',  
'upgrade.hvac\_heating\_efficiency', 'upgrade.cooking\_range', 'time\_of\_day', 'total\_energy\_consumption', 'dry\_bulb\_temperature\_[°c]', 'relative\_humidity\_[%]', 'wind\_speed\_[m/s]', 'wind\_direction\_[deg]', 'global\_horizontal\_radiation\_[w/m2]',  
 'direct\_normal\_radiation\_[w/m2]', 'diffuse\_horizontal\_radiation\_[w/m2]')  
  
static\_house\_energy\_weather\_df2 <- static\_house\_energy\_weather\_df %>% select(all\_of(cols\_req))  
  
# Convert 'None' to NA  
static\_house\_energy\_weather\_df2[static\_house\_energy\_weather\_df2 == 'None'] <- NA  
  
  
  
# Function to drop columns with only one unique value  
drop\_single\_unique\_columns <- function(data) {  
 single\_unique\_cols <- sapply(data, function(col) length(unique(col)) == 1)  
 return(data[, !single\_unique\_cols, drop = FALSE])  
}  
  
# Use the function to drop columns  
static\_house\_energy\_weather\_df2 <- drop\_single\_unique\_columns(static\_house\_energy\_weather\_df2)  
  
# Convert time of day and other column columns to numbers - Ordinal Encoding  
time\_of\_day\_mapping <- c("morning"=1,"afternoon-evening"=2,"night"=3)  
in\_vacancy\_status\_mapping <- c("Occupied"=1, "Vacant"=0 )  
in\_geometry\_floor\_area\_mapping <- c("0-499"=0 ,"500-749"=1,"750-999"=2,"1000-1499"=3,"1500-1999"=4,"2000-2499"=5,"2500-2999"=6,"3000-3999"=7,"4000+"=8)   
in\_hot\_water\_fixtures\_mapping <- c("100% Usage"=1, "50% Usage"=0, "200% Usage"=2)  
upgrade\_cooking\_range\_mapping <- c("Electric, Induction, 100% Usage"=1, "Electric, Induction, 80% Usage"=0, "Electric, Induction, 120% Usage"=3)  
in\_occupants\_mapping <- c("1"=1 , "2"=2,"3"=3,"4"=4,"5"=5,"8"=8,"6"=6,"7"=7,"10+"=10,"9"=9)  
income\_mapping <- c("<10000"=1, "10000-14999"=2, "15000-19999"=3, "20000-24999"=4, "25000-29999"=5, "30000-34999"=6, "35000-39999"=7, "40000-44999"=8, "45000-49999"=9, "50000-59999"=10, "60000-69999"=11, "70000-79999"=12, "80000-99999"=13, "100000-119999"=14, "120000-139999"=15, "140000-159999"=16, "160000-179999"=17, "180000-199999"=18, "200000+"=19)  
  
static\_house\_energy\_weather\_df2$time\_of\_day <- as.numeric(time\_of\_day\_mapping[static\_house\_energy\_weather\_df2$time\_of\_day])  
static\_house\_energy\_weather\_df2$in.vacancy\_status <- as.numeric(in\_vacancy\_status\_mapping[static\_house\_energy\_weather\_df2$in.vacancy\_status])  
static\_house\_energy\_weather\_df2$in.geometry\_floor\_area <- as.numeric(in\_geometry\_floor\_area\_mapping[static\_house\_energy\_weather\_df2$in.geometry\_floor\_area])  
static\_house\_energy\_weather\_df2$in.hot\_water\_fixtures <- as.numeric(in\_hot\_water\_fixtures\_mapping[static\_house\_energy\_weather\_df2$in.hot\_water\_fixtures])  
static\_house\_energy\_weather\_df2$upgrade.cooking\_range <- as.numeric(upgrade\_cooking\_range\_mapping[static\_house\_energy\_weather\_df2$upgrade.cooking\_range])  
static\_house\_energy\_weather\_df2$in.occupants <- as.numeric(in\_occupants\_mapping[static\_house\_energy\_weather\_df2$in.occupants])  
str(static\_house\_energy\_weather\_df2$time\_of\_day)

## num [1:17130] 3 1 2 3 1 2 3 1 2 3 ...

static\_house\_energy\_weather\_df2$in.income <- as.numeric(income\_mapping[static\_house\_energy\_weather\_df2$in.occupants])  
str(static\_house\_energy\_weather\_df2$in.income)

## num [1:17130] 3 3 3 1 1 1 2 2 2 2 ...

# Remove negative Target  
static\_house\_energy\_weather\_df2 <- static\_house\_energy\_weather\_df2 %>% filter(total\_energy\_consumption>=0)  
dim(static\_house\_energy\_weather\_df2)

## [1] 17085 96

###2.4. #### Calculate Percentage of nulls in each row - Remove those greater than 80%. #### Fill Categorical Nulls with mode at Country - Income Level. #### Remove any remaining Null rows, if any.

static\_house\_energy\_weather\_df3 <- static\_house\_energy\_weather\_df2  
  
# Function to calculate percentage of nulls in each column  
percentage\_nulls <- function(column) {  
 sum(is.na(column)) / length(column) \* 100  
}  
  
# Apply the function to each column  
column\_null\_percentage <- sapply(static\_house\_energy\_weather\_df3, percentage\_nulls)  
  
# Extract columns with more than 80% nulls  
columns\_above\_threshold <- names(column\_null\_percentage[column\_null\_percentage < 80])  
  
# Extract only those columns from the original data frame  
static\_house\_energy\_weather\_df3 <- static\_house\_energy\_weather\_df3[, columns\_above\_threshold]  
  
# dim(static\_house\_energy\_weather\_df3)  
  
  
# Function to fill missing values with the mode within each group  
fill\_mode <- function(x) {  
 if (is.character(x)) {  
 mode\_table <- table(x)  
 if (length(mode\_table) > 0) {  
 mode\_val <- names(sort(mode\_table, decreasing = TRUE))[1]  
 x[is.na(x)] <- mode\_val  
 }  
 }  
 x  
}  
  
# Apply the fill\_mode function to each character column within each group  
static\_house\_energy\_weather\_df3 <- static\_house\_energy\_weather\_df3 %>%  
 group\_by(in.city,in.income) %>%  
 mutate(across(where(is.character), fill\_mode))  
  
# Apply the fill\_mode function again to those that do not have a mode inside a group  
static\_house\_energy\_weather\_df3 <- static\_house\_energy\_weather\_df3 %>%  
 mutate(across(where(is.character), fill\_mode))  
  
  
  
# Drop NA Rows - 90% of rows remaining  
static\_house\_energy\_weather\_df3 <- na.omit(static\_house\_energy\_weather\_df3)  
dim(static\_house\_energy\_weather\_df3)

## [1] 16340 85

column\_null\_percentage <- sapply(static\_house\_energy\_weather\_df3, percentage\_nulls)

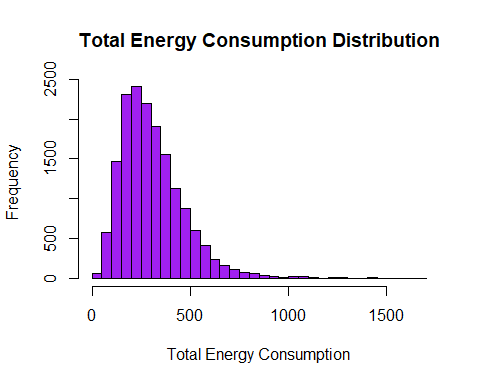
##—————————————————————————————————————————————————————–>

### 3.3.Exploratory Data Analysis

#### 3.1.Distribution of Target Variable - Total Energy Consumption

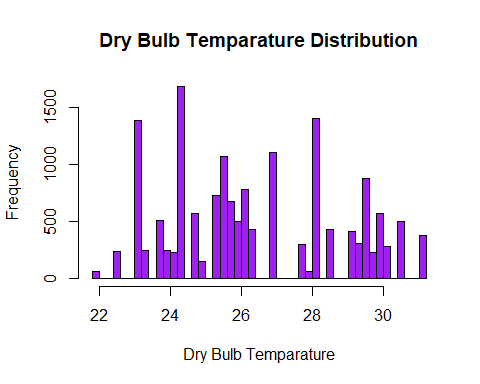
#### -> It is observed that target variable is slightly right-skewed.

# Histogram plot  
hist(static\_house\_energy\_weather\_df3$total\_energy\_consumption, col='purple', border='black', breaks=50,  
 main='Total Energy Consumption Distribution', xlab='Total Energy Consumption', ylab='Frequency')

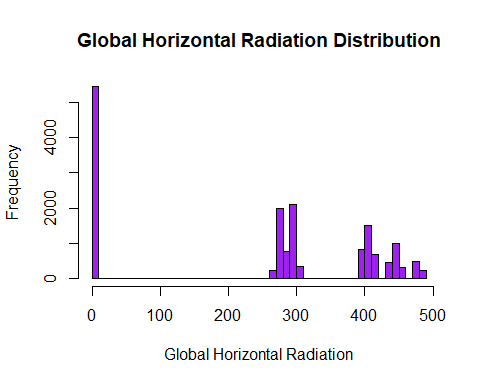


#### 3.2 Distribution of Continuous Independanct Features

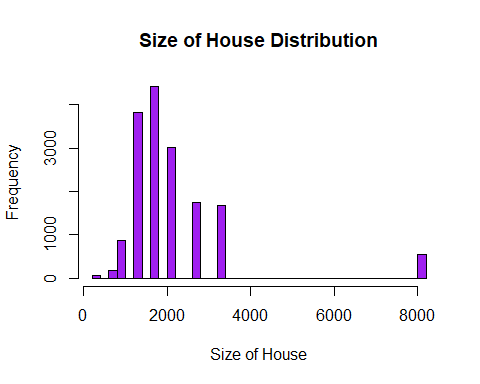
# Histogram plot -> dry\_bulb\_temperature\_  
hist(static\_house\_energy\_weather\_df3$`dry\_bulb\_temperature\_[°c]`, col='purple', border='black', breaks=50,  
 main='Dry Bulb Temparature Distribution', xlab='Dry Bulb Temparature', ylab='Frequency')



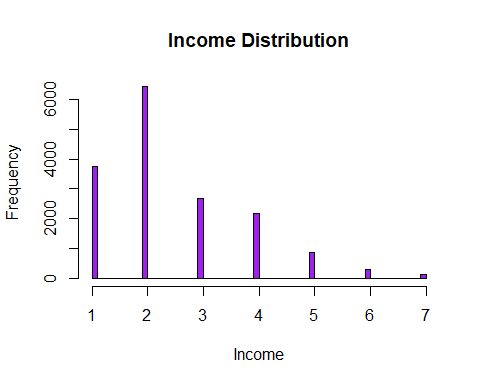
# Histogram plot -> global\_horizontal\_radiation\_  
hist(static\_house\_energy\_weather\_df3$`global\_horizontal\_radiation\_[w/m2]`, col='purple', border='black', breaks=50,  
 main='Global Horizontal Radiation Distribution', xlab='Global Horizontal Radiation', ylab='Frequency')



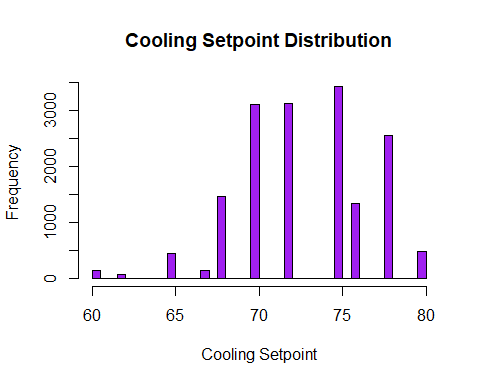
# Histogram plot -> in.sqft  
hist(static\_house\_energy\_weather\_df3$in.sqft, col='purple', border='black', breaks=50,  
 main='Size of House Distribution', xlab='Size of House', ylab='Frequency')



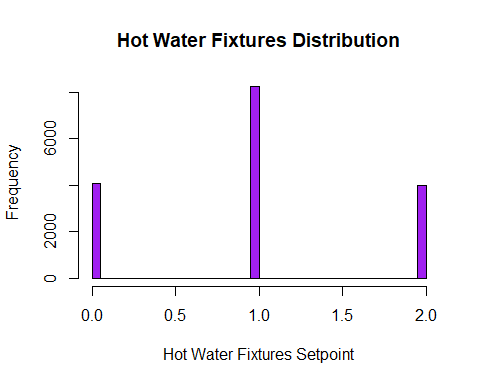
# Histogram plot -> Income  
hist(static\_house\_energy\_weather\_df3$in.income, col='purple', border='black', breaks=50,  
 main='Income Distribution', xlab='Income', ylab='Frequency')



# Histogram plot -> Colling Setpoint  
hist(static\_house\_energy\_weather\_df3$in.cooling\_setpoint, col='purple', border='black', breaks=50,  
 main='Cooling Setpoint Distribution', xlab='Cooling Setpoint', ylab='Frequency')



# Histogram plot -> Colling Setpoint  
hist(static\_house\_energy\_weather\_df3$in.hot\_water\_fixtures, col='purple', border='black', breaks=50,  
 main='Hot Water Fixtures Distribution', xlab='Hot Water Fixtures Setpoint', ylab='Frequency')

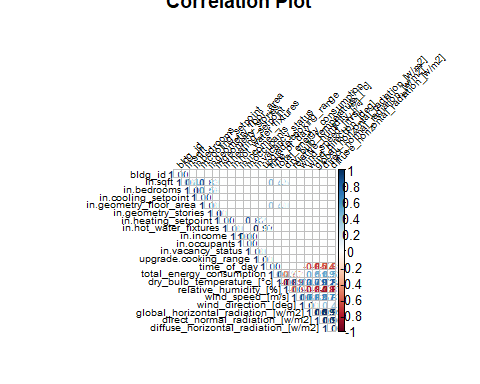


#### 3.3. Study Correlation between Continuous Features

#### Raditation and Humidity and Highly Negatively Correlated

#### Wind Speed and Radiation have a High Positive Correlation

# Select only the numerical variables from the dataset  
numerical\_vars <- static\_house\_energy\_weather\_df3[sapply(static\_house\_energy\_weather\_df3, is.numeric)]  
  
# Calculate the correlation matrix  
cor\_matrix <- cor(numerical\_vars)  
# Set a threshold for correlation values  
threshold <- 0.4  
  
# Filter correlation matrix for values above the threshold  
filtered\_cor\_matrix <- cor\_matrix \* (abs(cor\_matrix) > threshold)  
  
corrplot(filtered\_cor\_matrix, method = "number", type = "upper", tl.col = "black", tl.srt = 45, tl.cex = 0.6, number.cex = 0.6,title = "Correlation Plot",number.digits = 2)



par(mfrow = c(1, 1), mar = c(1,1,1,1) + 5)

##—————————————————————————————————————————————————————–>

### 4. Modelling

#### Linear,Xgboost and Catboost were the three models that were used.

#### XGBoost and Catboost gave thebest accuracy with a MAPE of 17% (83% accruracy).

#### 4.1.Run Linear Regression - Run Linear Regression algorithm on Dataset

# Drop a few columns based on correlation  
static\_house\_energy\_weather\_df4 <- static\_house\_energy\_weather\_df3  
  
static\_house\_energy\_weather\_df4 <- static\_house\_energy\_weather\_df4 %>%  
 select(where(~n\_distinct(.) > 1))  
  
  
# Create dataset with just building\_id and county - To be used in prediction section  
static\_house\_energy\_weather\_df\_for\_prediction <- static\_house\_energy\_weather\_df4   
static\_house\_energy\_weather\_df\_building\_and\_county <- static\_house\_energy\_weather\_df4[,c('bldg\_id','in.county','time\_of\_day')]  
static\_house\_energy\_weather\_df4 <- static\_house\_energy\_weather\_df4 %>% select(-c( 'bldg\_id','in.county','global\_horizontal\_radiation\_[w/m2]',  
 'direct\_normal\_radiation\_[w/m2]'))  
  
# lapply(static\_house\_energy\_weather\_df4, unique)  
  
rm(static\_house\_energy\_weather\_df)  
rm(static\_house\_energy\_weather\_df2)  
  
set.seed(123)  
  
# Split the data into training (80%) and testing (20%) sets  
index <- createDataPartition(static\_house\_energy\_weather\_df4$total\_energy\_consumption, p = 0.8, list = FALSE)  
train\_data <- static\_house\_energy\_weather\_df4[index, ]  
test\_data <- static\_house\_energy\_weather\_df4[-index, ]  
  
# Get character columns  
character\_columns <- names(train\_data)[sapply(train\_data, is.character)]  
  
# Loop through character columns and apply the steps  
for (col in character\_columns) {  
 # Get unique values in the current column  
  
 unique\_values <- unique(train\_data[[col]])  
  
 # Filter rows in other columns based on unique values  
 test\_data <- test\_data[test\_data[[col]] %in% unique\_values, ]  
}  
  
  
# Run linear regression with both categorical and continuous columns  
model <- lm(total\_energy\_consumption ~ ., data = train\_data)  
  
# Print the summary of the model  
# summary(model)  
  
# Make predictions on the test data  
predictions <- predict(model, newdata = test\_data)  
# test\_data$predictions <- predictions  
# Evaluate the model on the test data (e.g., calculate RMSE)  
rmse <- sqrt(mean((test\_data$total\_energy\_consumption - predictions)^2))  
print(paste("Root Mean Squared Error on test data:", rmse))

## [1] "Root Mean Squared Error on test data: 83.3659103617599"

cat("Minimum:", min(test\_data$total\_energy\_consumption), "\n")

## Minimum: 3.707

cat("Maximum:", max(test\_data$total\_energy\_consumption), "\n")

## Maximum: 1670.53

cat("Mean:", mean(test\_data$total\_energy\_consumption), "\n")

## Mean: 310.9359

# Calculate MAPE  
mape <- mean(abs((test\_data$total\_energy\_consumption - predictions) / test\_data$total\_energy\_consumption )) \* 100  
  
# Print the result  
print(paste("MAPE:", mape))

## [1] "MAPE: 25.8964937733538"

#R2  
r\_squared <- R2(predictions, test\_data$total\_energy\_consumption)  
print(paste("R-squared Linear Regression:", r\_squared))

## [1] "R-squared Linear Regression: 0.756463521318137"

#### 4.2. Run CatBoost - Catboost is a Boosting Regression Model,

#### which specializes in handling categorical variables - which we have a lot of.

#### Boosting models are good at handling overfitting and non-linear data as they as equipped with

#### regularization parameters and do not require prior coolumn scaling.

# Create Train-Test Split  
train\_data\_ctb <- train\_data  
test\_data\_ctb <- test\_data  
# # Convert all character columns to factors - Need for Catboost  
# train\_data\_ctb[sapply(train\_data\_ctb, is.character)] <- lapply(train\_data\_ctb[sapply(train\_data\_ctb, is.character)],   
# as.factor)  
# test\_data\_ctb[sapply(test\_data\_ctb, is.character)] <- lapply(test\_data\_ctb[sapply(test\_data\_ctb, is.character)],   
# as.factor)  
  
# Identify character columns  
char\_columns <- names(train\_data\_ctb)[sapply(train\_data\_ctb, is.character)]  
  
# Create a recipe for label encoding  
label\_recipe <- recipe(~ ., data = train\_data\_ctb) %>%  
 step\_dummy(all\_nominal\_predictors(), one\_hot = FALSE)  
  
# Apply the label encoding recipe to the data  
train\_data\_ctb <- prep(label\_recipe) %>% bake(new\_data = train\_data\_ctb)

## New names:  
## New names:  
## New names:  
## New names:  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..1h...116`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..2h...117`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..3h...118`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..4h...119`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..5h...120`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..1h...121`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..2h...122`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..3h...123`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..4h...124`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..5h...125`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..1h...127`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..2h...128`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..3h...129`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..4h...130`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..5h...131`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..1h...132`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..2h...133`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..3h...134`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..4h...135`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..5h...136`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..1h...138`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..2h...139`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..3h...140`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..4h...141`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..5h...142`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..1h...143`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..2h...144`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..3h...145`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..4h...146`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..5h...147`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..1h...149`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..2h...150`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..3h...151`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..4h...152`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..5h...153`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..1h...154`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..2h...155`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..3h...156`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..4h...157`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..5h...158`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..1h...160`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..2h...161`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..3h...162`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..4h...163`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..5h...164`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..1h...165`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..2h...166`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..3h...167`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..4h...168`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..5h...169`

# Apply the saved encoding recipe to the new data  
test\_data\_ctb <-prep(label\_recipe) %>% bake(new\_data = test\_data\_ctb)

## New names:  
## New names:  
## New names:  
## New names:  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..1h...116`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..2h...117`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..3h...118`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..4h...119`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..5h...120`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..1h...121`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..2h...122`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..3h...123`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..4h...124`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..5h...125`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..1h...127`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..2h...128`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..3h...129`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..4h...130`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..5h...131`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..1h...132`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..2h...133`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..3h...134`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..4h...135`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..5h...136`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..1h...138`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..2h...139`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..3h...140`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..4h...141`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..5h...142`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..1h...143`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..2h...144`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..3h...145`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..4h...146`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..5h...147`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..1h...149`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..2h...150`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..3h...151`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..4h...152`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..5h...153`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..1h...154`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..2h...155`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..3h...156`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..4h...157`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..5h...158`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..1h...160`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..2h...161`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..3h...162`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..4h...163`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..5h...164`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..1h...165`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..2h...166`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..3h...167`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..4h...168`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..5h...169`

# Specify the columns for features and the response variable  
response <- "total\_energy\_consumption"  
  
# Select all columns except the excluded one  
selected\_columns <- train\_data\_ctb[, !colnames(train\_data\_ctb) %in% response]  
  
# Convert the selected columns to a vector  
features <- colnames(selected\_columns)  
  
# Convert response variable to numeric (CatBoost requires numeric target)  
train\_data\_ctb[[response]] <- as.numeric(train\_data\_ctb[[response]])  
  
# Convert the data to the CatBoost dataset format  
train\_data\_ctb2 <- catboost.load\_pool(data = train\_data\_ctb[features], label = train\_data\_ctb[[response]])  
  
# Specify CatBoost parameters  
catboost\_params <- list(  
 iterations = 100, # Number of boosting iterations  
 learning\_rate = 0.1, # Learning rate  
 depth = 6, # Depth of the trees  
 loss\_function = "RMSE", # Loss function for regression  
 logging\_level = 'Silent'  
)  
  
# Train the CatBoost model  
# model <- catboost.train(train\_data\_ctb2, params = catboost\_params)  
model <- catboost.train(train\_data\_ctb2,NULL, params = catboost\_params)  
  
  
#Test Data  
new\_data\_pool <- catboost.load\_pool(data = test\_data\_ctb)  
predictions\_ctb <- catboost.predict(model, new\_data\_pool)  
  
  
#Catboost  
rmse <- sqrt(mean((test\_data\_ctb$total\_energy\_consumption - predictions\_ctb)^2))  
print(paste("Root Mean Squared Error on test data:", rmse))

## [1] "Root Mean Squared Error on test data: 70.1861792978264"

cat("Minimum:", min(test\_data\_ctb$total\_energy\_consumption), "\n")

## Minimum: 3.707

cat("Maximum:", max(test\_data\_ctb$total\_energy\_consumption), "\n")

## Maximum: 1670.53

cat("Mean:", mean(test\_data\_ctb$total\_energy\_consumption), "\n")

## Mean: 310.9359

# Calculate MAPE  
mape <- mean(abs((test\_data\_ctb$total\_energy\_consumption - predictions\_ctb) / test\_data\_ctb$total\_energy\_consumption )) \* 100  
  
# Print the result  
print(paste("MAPE:", mape))

## [1] "MAPE: 18.9428563237314"

#R2  
r\_squared <- R2(predictions\_ctb, test\_data\_ctb$total\_energy\_consumption)  
print(paste("R-squared CTB:", r\_squared))

## [1] "R-squared CTB: 0.829868885578707"

ctb\_feature\_imp\_matrix <- catboost.get\_feature\_importance(model,  
 pool = NULL,  
 type = 'FeatureImportance',  
 thread\_count = -1)  
  
# convert the matrix into dataframe   
dataframe\_ctb\_feature\_imp=as.data.frame(ctb\_feature\_imp\_matrix)  
colnames(dataframe\_ctb\_feature\_imp) <- c("Feature\_Importance")  
dataframe\_ctb\_feature\_imp <- arrange(dataframe\_ctb\_feature\_imp, desc(Feature\_Importance))  
  
# Top Features from Catboost  
head(dataframe\_ctb\_feature\_imp)

## Feature\_Importance  
## dry\_bulb\_temperature\_[°c] 18.895609  
## in.sqft 12.038119  
## time\_of\_day 9.316663  
## in.heating\_setpoint 9.055751  
## in.geometry\_floor\_area 8.063708  
## in.vacancy\_status 6.898157

#### 4.3, Run XGboost : Extreme Gradient Boosting model that is great at handling non-linear data.

#### Boosting models are good at handling overfitting and non-linear data as they as equipped with

#### regularization parameters and do not require prior coolumn scaling.

# https://tilburgsciencehub.com/building-blocks/analyze-data/machine-learning/xgboost/  
  
#Use train-test split created before  
train\_data\_xgb <- train\_data  
test\_data\_xgb <- test\_data  
# Convert all character columns to factors   
# train\_data\_xgb[sapply(train\_data\_xgb, is.character)] <- lapply(train\_data\_xgb[sapply(train\_data\_xgb, is.character)],   
# as.factor)  
# test\_data\_xgb[sapply(test\_data\_xgb, is.character)] <- lapply(test\_data\_xgb[sapply(test\_data\_xgb, is.character)],   
# as.factor)  
  
# Identify character columns  
char\_columns <- names(train\_data\_xgb)[sapply(train\_data\_xgb, is.character)]  
  
# Create a recipe for label encoding  
label\_recipe <- recipe(~ ., data = train\_data\_xgb) %>%  
 step\_dummy(all\_nominal\_predictors(), one\_hot = FALSE)  
  
# Apply the label encoding recipe to the data  
train\_data\_xgb <- prep(label\_recipe) %>% bake(new\_data = train\_data\_xgb)

## New names:  
## New names:  
## New names:  
## New names:  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..1h...116`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..2h...117`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..3h...118`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..4h...119`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..5h...120`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..1h...121`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..2h...122`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..3h...123`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..4h...124`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..5h...125`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..1h...127`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..2h...128`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..3h...129`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..4h...130`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..5h...131`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..1h...132`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..2h...133`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..3h...134`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..4h...135`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..5h...136`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..1h...138`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..2h...139`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..3h...140`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..4h...141`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..5h...142`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..1h...143`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..2h...144`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..3h...145`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..4h...146`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..5h...147`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..1h...149`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..2h...150`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..3h...151`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..4h...152`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..5h...153`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..1h...154`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..2h...155`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..3h...156`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..4h...157`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..5h...158`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..1h...160`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..2h...161`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..3h...162`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..4h...163`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..5h...164`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..1h...165`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..2h...166`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..3h...167`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..4h...168`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..5h...169`

# Apply the saved encoding recipe to the new data  
test\_data\_xgb <-prep(label\_recipe) %>% bake(new\_data = test\_data\_xgb)

## New names:  
## New names:  
## New names:  
## New names:  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..1h...116`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..2h...117`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..3h...118`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..4h...119`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..5h...120`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..1h...121`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..2h...122`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..3h...123`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..4h...124`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..5h...125`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..1h...127`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..2h...128`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..3h...129`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..4h...130`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..5h...131`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..1h...132`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..2h...133`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..3h...134`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..4h...135`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..5h...136`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..1h...138`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..2h...139`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..3h...140`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..4h...141`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..5h...142`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..1h...143`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..2h...144`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..3h...145`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..4h...146`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..5h...147`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..1h...149`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..2h...150`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..3h...151`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..4h...152`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..5h...153`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..1h...154`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..2h...155`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..3h...156`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..4h...157`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..5h...158`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..1h...160`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..2h...161`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..3h...162`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..4h...163`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..5h...164`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..1h...165`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..2h...166`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..3h...167`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..4h...168`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..5h...169`

# Define the parameter grid for hyperparameter tuning  
param\_grid <- expand.grid(  
 nrounds = c(50, 100, 150),  
 max\_depth = c(3, 6, 9),  
 eta = c(0.01, 0.05, 0.1),  
 gamma = c(0, 1, 2)  
)  
  
params <- list(  
 max\_depth = 10,  
 min\_child\_weight = 1,   
 gamma = 1,  
 eta = 0.1,   
 colsample\_bytree = 0.8,   
 objective = "reg:squarederror", # Set the objective for regression  
 eval\_metric = "rmse", # Use RMSE for evaluation  
 nrounds = 100   
)  
response <- "total\_energy\_consumption"  
  
X\_train <- train\_data\_xgb %>% select(-c("total\_energy\_consumption"))  
y\_train <- train\_data\_xgb$total\_energy\_consumption  
X\_test <- test\_data\_xgb %>% select(-c("total\_energy\_consumption"))  
y\_test <- test\_data\_xgb$total\_energy\_consumption  
  
#convert both sets to a DMatrix format, in order for xgboost to work with the data   
dtrain <- xgb.DMatrix(data = as.matrix(X\_train), label = y\_train)  
dtest <- xgb.DMatrix(data = as.matrix(X\_test), label = y\_test)  
  
  
xgb\_model <- xgb.train(  
 params = params,  
 data = dtrain,  
 nrounds = params$nrounds,   
 early\_stopping\_rounds = 20, # stop iteration when there test set does not improve for 20 rounds  
 watchlist = list(train = dtrain, test = dtest),  
 verbose = 0   
)

## [23:35:11] WARNING: src/learner.cc:767:   
## Parameters: { "nrounds" } are not used.

predictions\_xgb <- predict(xgb\_model, newdata = dtest)  
  
#XGBoost  
rmse <- sqrt(mean((test\_data\_xgb$total\_energy\_consumption - predictions\_xgb)^2))  
print(paste("Root Mean Squared Error on test data:", rmse))

## [1] "Root Mean Squared Error on test data: 66.7868768273033"

cat("Minimum:", min(test\_data\_xgb$total\_energy\_consumption), "\n")

## Minimum: 3.707

cat("Maximum:", max(test\_data\_xgb$total\_energy\_consumption), "\n")

## Maximum: 1670.53

cat("Mean:", mean(test\_data\_xgb$total\_energy\_consumption), "\n")

## Mean: 310.9359

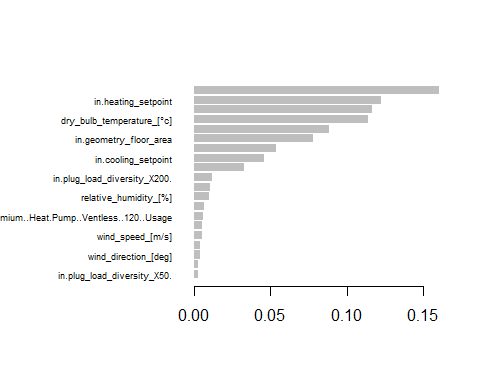
# Calculate MAPE  
mape <- mean(abs((test\_data\_xgb$total\_energy\_consumption - predictions\_xgb) / test\_data\_xgb$total\_energy\_consumption )) \* 100  
  
# Print the result  
print(paste("MAPE:", mape))

## [1] "MAPE: 18.8225847915731"

#R2  
r\_squared <- R2(predictions\_xgb, test\_data\_xgb$total\_energy\_consumption)  
print(paste("R-squared XGB:", r\_squared))

## [1] "R-squared XGB: 0.844847534851268"

#Plot XGB Feature IMportance  
importance\_matrix <- xgb.importance(feature\_names = colnames(X\_train), model = xgb\_model)  
xgb.plot.importance(importance\_matrix,top\_n = 20)



importance\_matrix <- importance\_matrix[order(importance\_matrix$Importance, decreasing = TRUE), ]  
top\_15\_features <- importance\_matrix$Feature[0:15]

##—————————————————————————————————————————————————————–>

## 5. Insight Generation

#### 1.Using Shapley to generate insights from the model.

#### 2.Shapley calculates the indivdial effect of the features towards total energy consumption.

#### 3.Analysing impact on Total energy consumption by studying Partial Dependance Plots and Feature Importance

#### 4.Instructions on reading the graph:

#### Understanding the y-axis:

#### The y-axis represents the predicted outcome (total energy consumption) of the model.

#### The 0 mark on the y-axis indicates average total energy consumption in the dataset

#### Understanding the x-axis:

#### The x-axis represents the values of the feature for which you are creating the Shapley PDP.

#### 5.Some notable insights:

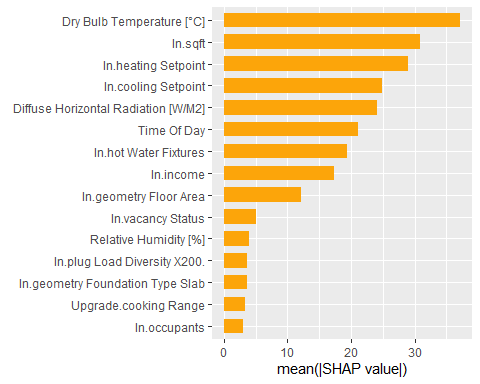
#### a.Dry Bulb Temperature, House Size (sqft), Cooling and Heating Setpoint, Humidity are some of the top drivers of energy consumption.

#### b. Income : People who earn more than $30,000, consume the same amount of energy (50kWh in an 8 hour time span).

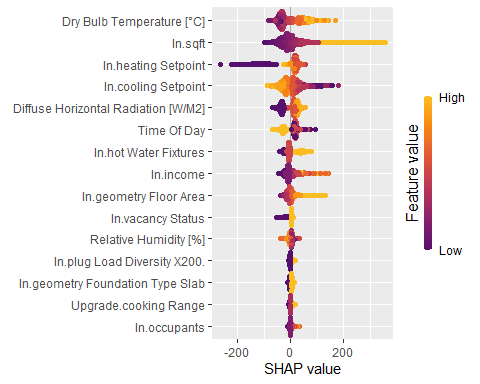
#### Below this income level, there is a steady decline in energy consumption as income reduces.

#### c.Temperature : After the weather outside crosses a temperature of 26 C, there is a spike in energy consumption across household.

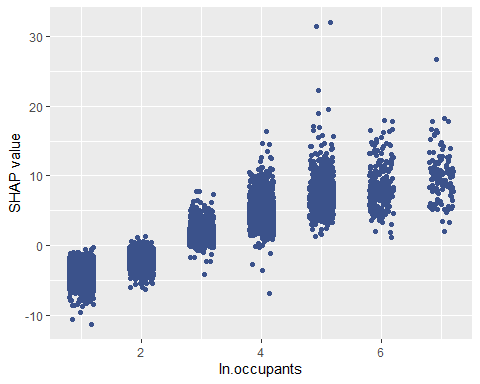
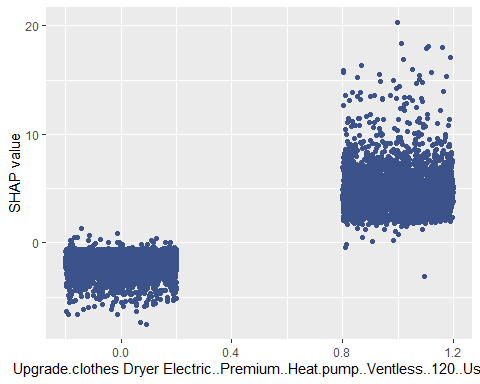
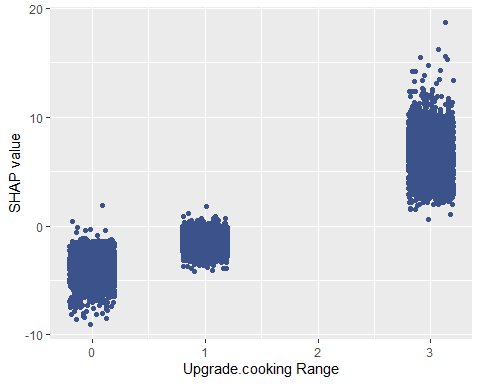
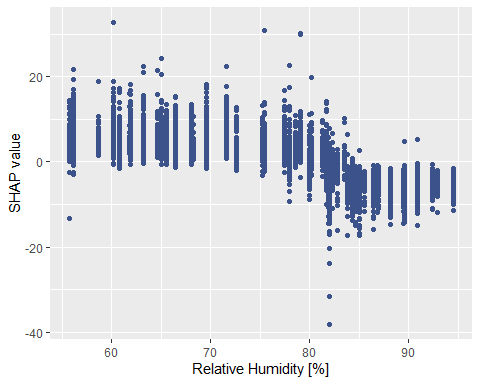
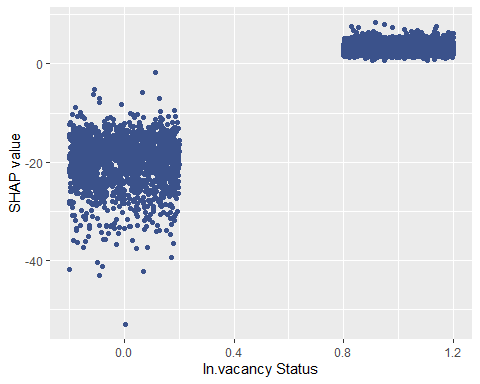
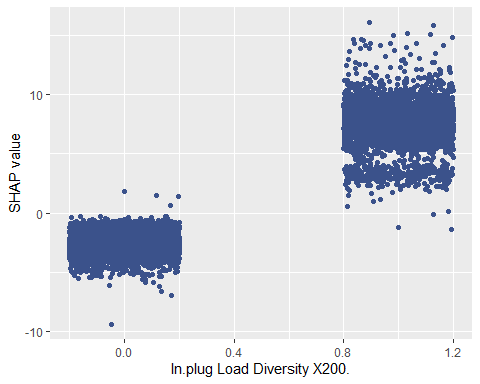
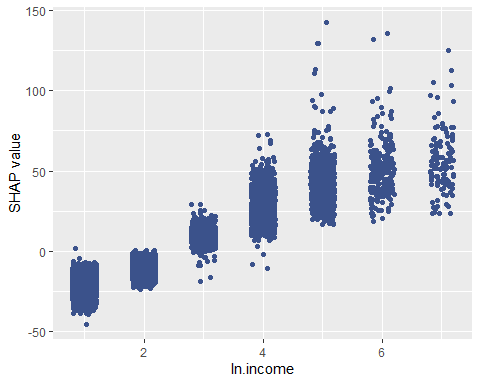
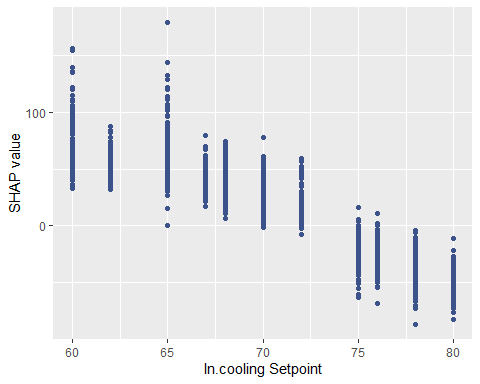
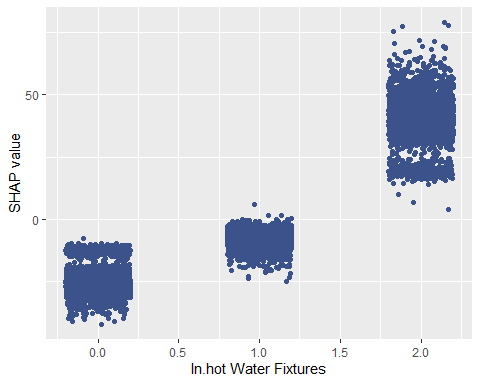
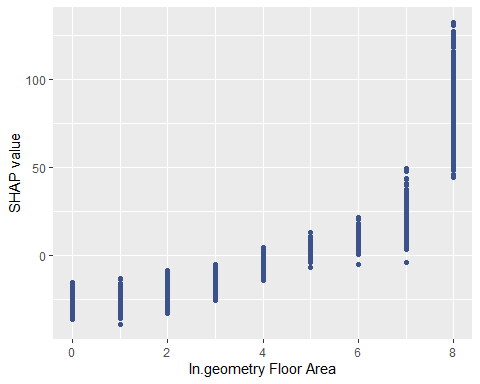
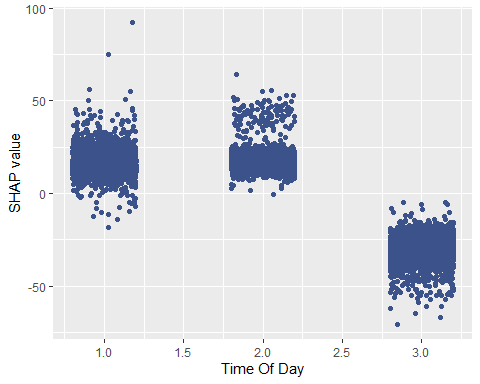
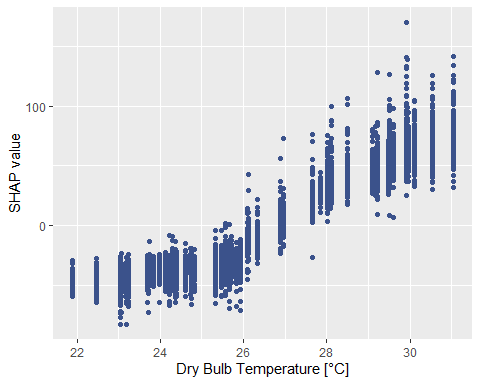
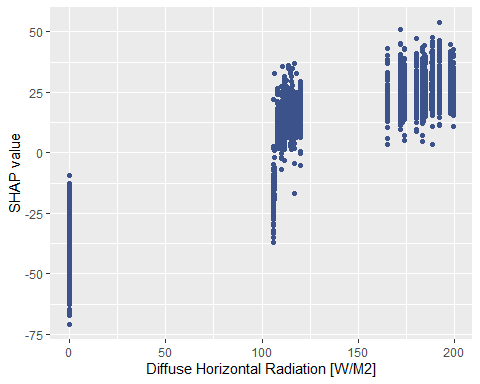
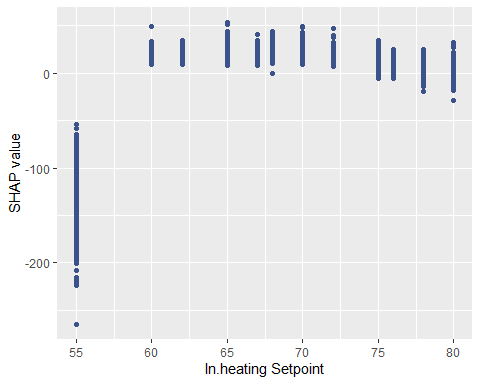
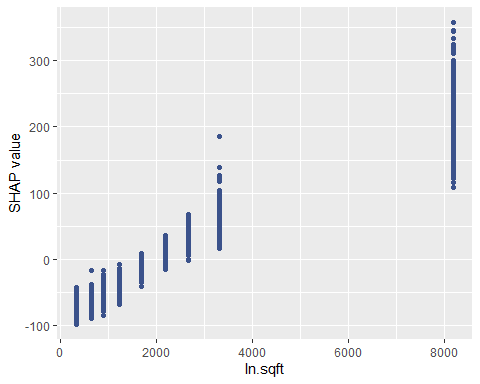
#Create Shap object   
  
shp <- shapviz(xgb\_model, X\_pred = dtrain, X = X\_train)  
colnames(shp) <- gsub("\_", " ", colnames(shp))  
# Function to capitalize each word before and after a full stop  
capitalize\_words <- function(x) {  
 words <- unlist(strsplit(x, "\\.", fixed = TRUE))  
 capitalized\_words <- sapply(words, str\_to\_title)  
 return(paste(capitalized\_words, collapse = "."))  
}  
# Capitalize each word in column names  
colnames(shp) <- sapply(colnames(shp), capitalize\_words)  
  
  
# Display Shapley Feature Importance  
sv\_importance(shp)



# Feature Importance with Direction of impact   
sv\_importance(shp, kind = "beeswarm")



# Main effect of carat and its interactions  
for (col in top\_15\_features){  
 col <- gsub("\_", " ", col)  
 col <- sapply(col, capitalize\_words)  
 impact\_of\_column\_on\_energy <- sv\_dependence(shp,v = col,color\_var = col)  
 print(impact\_of\_column\_on\_energy)}



# Key to map the x axis of below graphs  
time\_of\_day\_mapping <- c("morning"=1,"afternoon-evening"=2,"night"=3)  
in\_vacancy\_status\_mapping <- c("Occupied"=1, "Vacant"=0 )  
in\_geometry\_floor\_area\_mapping <- c("0-499"=0 ,"500-749"=1,"750-999"=2,"1000-1499"=3,"1500-1999"=4,"2000-2499"=5,"2500-2999"=6,"3000-3999"=7,"4000+"=8)   
in\_hot\_water\_fixtures\_mapping <- c("100% Usage"=1, "50% Usage"=0, "200% Usage"=2)  
upgrade\_cooking\_range\_mapping <- c("Electric, Induction, 100% Usage"=1, "Electric, Induction, 80% Usage"=0, "Electric, Induction, 120% Usage"=3)  
in\_occupants\_mapping <- c("1"=1 , "2"=2,"3"=3,"4"=4,"5"=5,"8"=8,"6"=6,"7"=7,"10+"=10,"9"=9)  
income\_mapping <- c("<10000"=1, "10000-14999"=2, "15000-19999"=3, "20000-24999"=4, "25000-29999"=5, "30000-34999"=6, "35000-39999"=7, "40000-44999"=8, "45000-49999"=9, "50000-59999"=10, "60000-69999"=11, "70000-79999"=12, "80000-99999"=13, "100000-119999"=14, "120000-139999"=15, "140000-159999"=16, "160000-179999"=17, "180000-199999"=18, "200000+"=19)

##—————————————————————————————————————————————————————–>

## 6. Predictions - By increasing temp by 5

#### Add 5 degrees C to the Original Dataset, and Predict total Energy Consumed.

#### Calculate peak Future Energy Demand for Geographies,House size and Time of the Day.

#### Total energy predicted next july after increasing temparature by 5 degrees is: 5983238 kWh

#### and increase in total energy consumed next july is : 927594.8 kWh Percentage increase in Energy Consumed is : 18.34771 %

# Use prediction DF created in Linear Regression Section  
  
# ----------------->   
static\_house\_energy\_weather\_df\_for\_prediction2 <- static\_house\_energy\_weather\_df\_for\_prediction  
static\_house\_energy\_weather\_df\_building\_and\_county <- static\_house\_energy\_weather\_df\_for\_prediction2[,c('bldg\_id','in.county','time\_of\_day')]  
static\_house\_energy\_weather\_df\_for\_prediction2\_processed\_for\_loaded\_xgb\_model <- static\_house\_energy\_weather\_df\_for\_prediction2 %>% select(-c( 'bldg\_id','in.county','global\_horizontal\_radiation\_[w/m2]',  
 'direct\_normal\_radiation\_[w/m2]'))  
  
# Encode it using the label\_recipe encoder created in previous steps (xgb cell)  
static\_house\_energy\_weather\_df4\_prediction <- prep(label\_recipe) %>% bake(new\_data = static\_house\_energy\_weather\_df\_for\_prediction2\_processed\_for\_loaded\_xgb\_model)

## New names:  
## New names:  
## New names:  
## New names:  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..1h...116`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..2h...117`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..3h...118`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..4h...119`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..5h...120`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..1h...121`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..2h...122`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..3h...123`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..4h...124`  
## • `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.and.Night.Setup..5h...125`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..1h...127`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..2h...128`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..3h...129`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..4h...130`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..5h...131`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..1h...132`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..2h...133`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..3h...134`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..4h...135`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup..5h...136`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..1h...138`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..2h...139`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..3h...140`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..4h...141`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..5h...142`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..1h...143`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..2h...144`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..3h...145`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..4h...146`  
## • `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Day.Setup.and.Night.Setback..5h...147`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..1h...149`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..2h...150`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..3h...151`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..4h...152`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..5h...153`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..1h...154`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..2h...155`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..3h...156`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..4h...157`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setback..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setback..5h...158`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..1h...160`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..2h...161`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..3h...162`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..4h...163`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..5h...164`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..1h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..1h...165`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..2h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..2h...166`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..3h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..3h...167`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..4h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..4h...168`  
## • `in.cooling\_setpoint\_offset\_period\_Night.Setup..5h` ->  
## `in.cooling\_setpoint\_offset\_period\_Night.Setup..5h...169`

#-----------------> Increase Temp by 5  
static\_house\_energy\_weather\_df4\_prediction$`dry\_bulb\_temperature\_[°c]` <- static\_house\_energy\_weather\_df4\_prediction$`dry\_bulb\_temperature\_[°c]`+5  
  
# Increase Temp by 5-> Create dataset to be converted to a martix for xgb prediction  
static\_house\_energy\_weather\_df4\_prediction\_independant <- static\_house\_energy\_weather\_df4\_prediction %>% select (-c('total\_energy\_consumption'))  
static\_house\_energy\_weather\_df4\_prediction\_dependant <- static\_house\_energy\_weather\_df4\_prediction$total\_energy\_consumption  
  
  
#Create matrix which is a pre-requisite to XGboost prediction  
static\_house\_energy\_weather\_df4\_matrix <- xgb.DMatrix(data = as.matrix(static\_house\_energy\_weather\_df4\_prediction\_independant ), label = static\_house\_energy\_weather\_df4\_prediction\_dependant)  
predictions\_plus\_5\_temp <- predict(xgb\_model, newdata = static\_house\_energy\_weather\_df4\_matrix)  
  
#Add prediction columns to original DF  
static\_house\_energy\_weather\_df\_for\_prediction2$total\_energy\_consumption\_prediction\_after\_increasing\_temp <- predictions\_plus\_5\_temp  
  
#Predicted - Current = Increase in energy   
static\_house\_energy\_weather\_df\_for\_prediction2$change\_in\_total\_energy <- static\_house\_energy\_weather\_df\_for\_prediction2$total\_energy\_consumption\_prediction\_after\_increasing\_temp - static\_house\_energy\_weather\_df\_for\_prediction2$total\_energy\_consumption   
  
#TotalEnergy  
total\_energy\_next\_july\_after\_increasing\_temp <- sum(static\_house\_energy\_weather\_df\_for\_prediction2$total\_energy\_consumption\_prediction\_after\_increasing\_temp)  
  
total\_change\_in\_energy\_this\_july\_to\_last\_july <- sum(static\_house\_energy\_weather\_df\_for\_prediction2$change\_in\_total\_energy)  
  
percentage\_increase\_in\_energy\_consumed <- total\_change\_in\_energy\_this\_july\_to\_last\_july\*100/sum(static\_house\_energy\_weather\_df\_for\_prediction2$total\_energy\_consumption)  
  
cat("Total energy predicted next july after increasing temparature by 5 degrees is: ",total\_energy\_next\_july\_after\_increasing\_temp," kWh", " and increase in total energy consumed next july is :",total\_change\_in\_energy\_this\_july\_to\_last\_july," kWh")

## Total energy predicted next july after increasing temparature by 5 degrees is: 5983238 kWh and increase in total energy consumed next july is : 927594.8 kWh

cat(" Percentage increase in Energy Consumed is :",percentage\_increase\_in\_energy\_consumed, " %")

## Percentage increase in Energy Consumed is : 18.34771 %

# #calculate peak energy demand for geographies and regions  
options(dplyr.summarise.inform = FALSE)  
time\_of\_day\_reverse\_mapping <- c("1"="morning","2"="afternoon-evening","3"="night")  
  
geography\_energy\_demand\_df <- static\_house\_energy\_weather\_df\_for\_prediction2 %>%   
 group\_by(in.county) %>%   
 dplyr::summarise(total\_energy\_prediction\_next\_year\_per\_hour = sum(total\_energy\_consumption\_prediction\_after\_increasing\_temp)/8, total\_energy\_\_consumption\_this\_year\_per\_hour = sum(total\_energy\_consumption)/8)  
  
geography\_energy\_demand\_df$percentage\_change\_in\_energy\_consumption\_per\_hour <- (geography\_energy\_demand\_df$total\_energy\_prediction\_next\_year\_per\_hour-geography\_energy\_demand\_df$total\_energy\_\_consumption\_this\_year\_per\_hour)\*100/geography\_energy\_demand\_df$total\_energy\_\_consumption\_this\_year\_per\_hour   
print(geography\_energy\_demand\_df)

## # A tibble: 46 × 4  
## in.county total\_energy\_prediction\_next\_year\_per\_hour total\_energy\_\_consumpt…¹  
## <chr> <dbl> <dbl>  
## 1 G4500010 3769. 3014.  
## 2 G4500030 26575. 22183.  
## 3 G4500050 1331. 1100.  
## 4 G4500070 33435. 28071.  
## 5 G4500090 2636. 2278.  
## 6 G4500110 2550. 2155.  
## 7 G4500130 33108. 29216.  
## 8 G4500150 27237. 22716.  
## 9 G4500170 2119. 1753.  
## 10 G4500190 62152. 52936.  
## # ℹ 36 more rows  
## # ℹ abbreviated name: ¹​total\_energy\_\_consumption\_this\_year\_per\_hour  
## # ℹ 1 more variable: percentage\_change\_in\_energy\_consumption\_per\_hour <dbl>

## Calculating peak energy demand for sqft  
  
sqft\_wise\_energy\_demand\_df <- static\_house\_energy\_weather\_df\_for\_prediction2 %>%   
 group\_by(in.sqft) %>%   
 dplyr::summarise(total\_energy\_prediction\_next\_year\_per\_hour = sum(total\_energy\_consumption\_prediction\_after\_increasing\_temp)/8, total\_energy\_\_consumption\_this\_year\_per\_hour = sum(total\_energy\_consumption)/8)  
sqft\_wise\_energy\_demand\_df$percentage\_change\_in\_energy\_consumption <- (sqft\_wise\_energy\_demand\_df$total\_energy\_prediction\_next\_year\_per\_hour-sqft\_wise\_energy\_demand\_df$total\_energy\_\_consumption\_this\_year\_per\_hour)\*100/sqft\_wise\_energy\_demand\_df$total\_energy\_\_consumption\_this\_year\_per\_hour  
print(sqft\_wise\_energy\_demand\_df)

## # A tibble: 9 × 4  
## in.sqft total\_energy\_predictio…¹ total\_energy\_\_consum…² percentage\_change\_in…³  
## <dbl> <dbl> <dbl> <dbl>  
## 1 328 2067. 1664. 24.2  
## 2 633 6235. 5164. 20.7  
## 3 885 28427. 23969. 18.6  
## 4 1220 143299. 121350. 18.1  
## 5 1690 189916. 160208. 18.5  
## 6 2176 143894. 120327. 19.6  
## 7 2663 89270. 75202. 18.7  
## 8 3301 95438. 80732. 18.2  
## 9 8194 49358. 43339. 13.9  
## # ℹ abbreviated names: ¹​total\_energy\_prediction\_next\_year\_per\_hour,  
## # ²​total\_energy\_\_consumption\_this\_year\_per\_hour,  
## # ³​percentage\_change\_in\_energy\_consumption

## Calculating peak energy demand for time of the day  
time\_of\_day\_wise\_energy\_demand\_df <- static\_house\_energy\_weather\_df\_for\_prediction2 %>%   
 group\_by(time\_of\_day) %>%   
 dplyr::summarise(total\_energy\_prediction\_next\_year\_per\_hour = sum(total\_energy\_consumption\_prediction\_after\_increasing\_temp)/8, total\_energy\_\_consumption\_this\_year\_per\_hour = sum(total\_energy\_consumption)/8)  
time\_of\_day\_wise\_energy\_demand\_df$time\_of\_day <- as.character(time\_of\_day\_reverse\_mapping[time\_of\_day\_wise\_energy\_demand\_df$time\_of\_day])  
  
time\_of\_day\_wise\_energy\_demand\_df$percentage\_change\_in\_energy\_consumption\_per\_hour <- (time\_of\_day\_wise\_energy\_demand\_df$total\_energy\_prediction\_next\_year\_per\_hour-time\_of\_day\_wise\_energy\_demand\_df$total\_energy\_\_consumption\_this\_year\_per\_hour)\*100/time\_of\_day\_wise\_energy\_demand\_df$total\_energy\_\_consumption\_this\_year\_per\_hour  
  
  
  
print(time\_of\_day\_wise\_energy\_demand\_df)

## # A tibble: 3 × 4  
## time\_of\_day total\_energy\_prediction\_next\_year\_p…¹ total\_energy\_\_consum…²  
## <chr> <dbl> <dbl>  
## 1 morning 264819. 215282.  
## 2 afternoon-evening 291670. 278974.  
## 3 night 191416. 137700.  
## # ℹ abbreviated names: ¹​total\_energy\_prediction\_next\_year\_per\_hour,  
## # ²​total\_energy\_\_consumption\_this\_year\_per\_hour  
## # ℹ 1 more variable: percentage\_change\_in\_energy\_consumption\_per\_hour <dbl>

# write.csv(sqft\_wise\_energy\_demand\_df, file = "C:/Users/Himanshu/OneDrive/Desktop/Syracuse Sem 1/IST-687 Intro to DS Lab/Final Project/size\_of\_house\_surge.csv",row.names = FALSE)