

Main_data_Models

2023-12-03

Version-1 Model

```
in.sqft in.bedrooms in.building_america_climate_zone in.ceiling_fan in.cooling_setpoint
in.cooling_setpoint_has_offset in.cooling_setpoint_offset_magnitude in.cooling_setpoint_offset_period in.ducts
in.geometry_foundation_type in.geometry_wall_type in.has_pv in.heating_fuel in.hot_water_fixtures
in.hvac_cooling_partial_space_conditioning in.hvac_cooling_type in.hvac_heating_type
in.hvac_heating_type_and_fuel in.insulation_ceiling in.insulation_wall in.lighting in.misc_extra_refrigerator
in.misc_freezer in.misc_pool_pump in.occupants in.pv_system_size in.refrigerator in.roof_material
in.usage_level in.vacancy_status in.water_heater_efficiency in.water_heater_fuel
```

```
library (arrow)
```

```
## Warning: package 'arrow' was built under R version 4.3.2
```

```
##
## Attaching package: 'arrow'
```

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

```
library(arrow)
library(tidyverse)
```

```
## — Attaching core tidyverse packages ————— tidyverse 2.0.0 —
## ✓ dplyr      1.1.3      ✓ readr      2.1.4
## ✓ forcats    1.0.0      ✓ stringr    1.5.0
## ✓ ggplot2     3.4.4      ✓ 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 be
come errors
```

```
Merged_Final<-read_parquet("Aggregate_Final_Dataset.parquet")
```

```
str(Merged_Final)
```

```
## tibble [137,040 × 102] (S3: tbl_df/tbl/data.frame)
## $ in.county                : chr [1:137040] "G4500010" "G4500010" "G4500
010" "G4500010" ...
## $ hour                     : num [1:137040] 0 0 0 0 0 0 0 0 0 ...
## $ Dry Bulb Temperature [°C] : num [1:137040] 22.4 22.4 22.4 22.4 22.4 ...
## $ Relative Humidity [%]    : num [1:137040] 95.2 95.2 95.2 95.2 95.2 ...
## $ Wind Speed [m/s]         : num [1:137040] 1.09 1.09 1.09 1.09 1.09 ...
## $ Wind Direction [Deg]     : num [1:137040] 126 126 126 126 126 ...
## $ Global Horizontal Radiation [W/m2] : num [1:137040] 0 0 0 0 0 0 0 0 0 ...
## $ Direct Normal Radiation [W/m2]    : num [1:137040] 0 0 0 0 0 0 0 0 0 ...
## $ Diffuse Horizontal Radiation [W/m2] : num [1:137040] 0 0 0 0 0 0 0 0 0 ...
## $ bldg_id                   : num [1:137040] 410602 465218 473719 29915 1
02598 ...
## $ in.sqft                  : num [1:137040] 1220 2176 3301 2663 1690 ...
## $ in.bathroom_spot_vent_hour : chr [1:137040] "Hour20" "Hour11" "Hour4" "H
our19" ...
## $ in.bedrooms              : num [1:137040] 4 4 5 3 3 4 3 4 3 2 ...
## $ in.building_america_climate_zone : chr [1:137040] "Mixed-Humid" "Mixed-Humid"
"Mixed-Humid" "Mixed-Humid" ...
## $ in.ceiling_fan           : chr [1:137040] "Standard Efficiency" "Stand
ard Efficiency" "Standard Efficiency" "Standard Efficiency, No usage" ...
## $ in.city                  : chr [1:137040] "In another census Place" "N
ot in a census Place" "Not in a census Place" "Not in a census Place" ...
## $ in.clothes_dryer         : chr [1:137040] "Electric, 120% Usage" "Gas,
100% Usage" "Electric, 80% Usage" "Propane, 100% Usage" ...
## $ in.clothes_washer        : chr [1:137040] "EnergyStar, 120% Usage" "En
ergyStar, 100% Usage" "Standard, 80% Usage" "EnergyStar, 100% Usage" ...
## $ in.clothes_washer_presence : chr [1:137040] "Yes" "Yes" "Yes" "Yes" ...
## $ in.cooking_range          : chr [1:137040] "Electric, 120% Usage" "Elec
tric, 100% Usage" "Electric, 80% Usage" "Electric, 100% Usage" ...
## $ in.cooling_setpoint       : chr [1:137040] "75F" "70F" "75F" "75F" ...
## $ in.cooling_setpoint_has_offset : chr [1:137040] "No" "No" "No" "No" ...
## $ in.cooling_setpoint_offset_magnitude : chr [1:137040] "0F" "0F" "0F" "0F" ...
## $ in.cooling_setpoint_offset_period : chr [1:137040] "None" "None" "None" "None"
...
## $ in.county_and_puma       : chr [1:137040] "G4500010, G45001600" "G4500
010, G45001600" "G4500010, G45001600" "G4500010, G45001600" ...
## $ in.dishwasher            : chr [1:137040] "290 Rated kWh, 120% Usage"
"318 Rated kWh, 100% Usage" "290 Rated kWh, 80% Usage" "None" ...
## $ in.ducts                 : chr [1:137040] "20% Leakage, R-4" "20% Leak
age, R-8" "20% Leakage, R-4" "20% Leakage, R-4" ...
## $ in.federal_poverty_level : chr [1:137040] "300-400%" "150-200%" "400%
+" "400%+" ...
## $ in.geometry_attic_type   : chr [1:137040] "Vented Attic" "Vented Atti
c" "Vented Attic" "Vented Attic" ...
## $ in.geometry_floor_area   : chr [1:137040] "1000-1499" "2000-2499" "300
0-3999" "2500-2999" ...
## $ in.geometry_floor_area_bin : chr [1:137040] "0-1499" "1500-2499" "2500-3
999" "2500-3999" ...
## $ in.geometry_foundation_type : chr [1:137040] "Slab" "Slab" "Slab" "Slab"
...
## $ in.geometry_garage       : chr [1:137040] "None" "2 Car" "2 Car" "Non
e" ...
## $ in.geometry_stories      : num [1:137040] 1 1 2 1 2 2 1 2 1 1 ...
## $ in.geometry_stories_low_rise : num [1:137040] 1 1 2 1 2 2 1 2 1 1 ...
```

```

## $ in.geometry_wall_exterior_finish      : chr [1:137040] "Wood, Medium/Dark" "Brick,
Medium/Dark" "Vinyl, Light" "Aluminum, Light" ...
## $ in.geometry_wall_type                  : chr [1:137040] "Wood Frame" "Wood Frame" "W
ood Frame" "Steel Frame" ...
## $ in.has_pv                              : chr [1:137040] "No" "No" "No" "No" ...
## $ in.heating_fuel                        : chr [1:137040] "Electricity" "Electricity"
"Propane" "Electricity" ...
## $ in.heating_setpoint                    : chr [1:137040] "70F" "72F" "65F" "55F" ...
## $ in.heating_setpoint_has_offset          : chr [1:137040] "Yes" "Yes" "No" "No" ...
## $ in.heating_setpoint_offset_magnitude   : chr [1:137040] "3F" "3F" "0F" "0F" ...
## $ in.heating_setpoint_offset_period      : chr [1:137040] "Night" "Day and Night -4h"
"None" "None" ...
## $ in.hot_water_fixtures                  : chr [1:137040] "200% Usage" "100% Usage" "5
0% Usage" "100% Usage" ...
## $ in.hvac_cooling_efficiency              : chr [1:137040] "AC, SEER 15" "Heat Pump" "A
C, SEER 13" "Heat Pump" ...
## $ in.hvac_cooling_partial_space_conditioning: chr [1:137040] "100% Conditioned" "100% Con
ditioned" "100% Conditioned" "100% Conditioned" ...
## $ in.hvac_cooling_type                   : chr [1:137040] "Central AC" "Heat Pump" "Ce
ntral AC" "Heat Pump" ...
## $ in.hvac_has_ducts                      : chr [1:137040] "Yes" "Yes" "Yes" "Yes" ...
## $ in.hvac_has_zonal_electric_heating     : chr [1:137040] "No" "No" "No" "No" ...
## $ in.hvac_heating_efficiency              : chr [1:137040] "Electric Furnace, 100% AFU
E" "ASHP, SEER 13, 7.7 HSPF" "Fuel Furnace, 80% AFUE" "ASHP, SEER 13, 7.7 HSPF" ...
## $ in.hvac_heating_type                   : chr [1:137040] "Ducted Heating" "Ducted Hea
t Pump" "Ducted Heating" "Ducted Heat Pump" ...
## $ in.hvac_heating_type_and_fuel          : chr [1:137040] "Electricity Electric Furnac
e" "Electricity ASHP" "Propane Fuel Furnace" "Electricity ASHP" ...
## $ in.income                             : chr [1:137040] "45000-49999" "50000-59999"
"160000-179999" "80000-99999" ...
## $ in.income_recs_2015                   : chr [1:137040] "40000-59999" "40000-59999"
"140000+" "80000-99999" ...
## $ in.income_recs_2020                   : chr [1:137040] "40000-59999" "40000-59999"
"150000+" "60000-99999" ...
## $ in.infiltration                       : chr [1:137040] "15 ACH50" "25 ACH50" "4 ACH
50" "15 ACH50" ...
## $ in.insulation_ceiling                  : chr [1:137040] "R-30" "R-30" "R-7" "R-30"
...
## $ in.insulation_floor                    : chr [1:137040] "None" "None" "None" "None"
...
## $ in.insulation_foundation_wall          : chr [1:137040] "None" "None" "None" "None"
...
## $ in.insulation_rim_joist                : chr [1:137040] "None" "None" "None" "None"
...
## $ in.insulation_roof                    : chr [1:137040] "Unfinished, Uninsulated" "U
nfinished, Uninsulated" "Unfinished, Uninsulated" "Unfinished, Uninsulated" ...
## $ in.insulation_slab                     : chr [1:137040] "Uninsulated" "2ft R10 Unde
r, Horizontal" "Uninsulated" "Uninsulated" ...
## $ in.insulation_wall                     : chr [1:137040] "Wood Stud, Uninsulated" "Wo
od Stud, R-15" "Wood Stud, Uninsulated" "Wood Stud, R-11" ...
## $ in.lighting                           : chr [1:137040] "100% Incandescent" "100% In
candescent" "100% LED" "100% CFL" ...
## $ in.misc_extra_refrigerator             : chr [1:137040] "EF 15.9" "None" "None" "Non
e" ...
## $ in.misc_freezer                        : chr [1:137040] "None" "EF 12, National Aver
age" "None" "EF 12, National Average" ...

```

```

## $ in.misc_gas_fireplace : chr [1:137040] "None" "None" "None" "None"
...
## $ in.misc_gas_grill : chr [1:137040] "Gas Grill" "None" "None" "N
one" ...
## $ in.misc_gas_lighting : chr [1:137040] "None" "None" "None" "None"
...
## $ in.misc_hot_tub_spa : chr [1:137040] "None" "None" "None" "Electr
ic" ...
## $ in.misc_pool : chr [1:137040] "None" "None" "None" "None"
...
## $ in.misc_pool_heater : chr [1:137040] "None" "None" "None" "None"
...
## $ in.misc_pool_pump : chr [1:137040] "None" "None" "None" "None"
...
## $ in.misc_well_pump : chr [1:137040] "None" "None" "None" "None"
...
## $ in.occupants : chr [1:137040] "1" "5" "4" "2" ...
## $ in.orientation : chr [1:137040] "West" "South" "East" "Nort
h" ...
## $ in.plug_load_diversity : chr [1:137040] "200%" "100%" "50%" "100%"
...
## $ in.puma : chr [1:137040] "G45001600" "G45001600" "G45
001600" "G45001600" ...
## $ in.puma_metro_status : chr [1:137040] "Not/partially in metro are
a" "Not/partially in metro area" "Not/partially in metro area" "Not/partially in metro area"
...
## $ in.pv_orientation : chr [1:137040] "None" "None" "None" "None"
...
## $ in.pv_system_size : chr [1:137040] "None" "None" "None" "None"
...
## $ in.range_spot_vent_hour : chr [1:137040] "Hour9" "Hour19" "Hour2" "Ho
ur16" ...
## $ in.reeds_balancing_area : num [1:137040] 95 95 95 95 95 95 95 95 95 9
5 ...
## $ in.refrigerator : chr [1:137040] "EF 17.6, 100% Usage" "EF 1
7.6, 100% Usage" "EF 17.6, 100% Usage" "EF 17.6, 100% Usage" ...
## $ in.roof_material : chr [1:137040] "Composition Shingles" "Wood
Shingles" "Composition Shingles" "Composition Shingles" ...
## $ in.tenure : chr [1:137040] "Owner" "Renter" "Owner" "Ow
ner" ...
## $ in.usage_level : chr [1:137040] "High" "Medium" "Low" "Mediu
m" ...
## $ in.vacancy_status : chr [1:137040] "Occupied" "Occupied" "Occup
ied" "Vacant" ...
## $ in.vintage : chr [1:137040] "1960s" "2000s" "1970s" "199
0s" ...
## $ in.vintage_acs : chr [1:137040] "1960-79" "2000-09" "1960-7
9" "1980-99" ...
## $ in.water_heater_efficiency : chr [1:137040] "Electric Standard" "Electri
c Standard" "Electric Standard" "Electric Standard" ...
## $ in.water_heater_fuel : chr [1:137040] "Electricity" "Electricity"
"Electricity" "Electricity" ...
## $ in.weather_file_city : chr [1:137040] "Greenwood Co" "Greenwood C
o" "Greenwood Co" "Greenwood Co" ...
## $ in.weather_file_latitude : num [1:137040] 34.2 34.2 34.2 34.2 34.2 ...
## $ in.weather_file_longitude : num [1:137040] -82.2 -82.2 -82.2 -82.2 -82.

```

```

2 ...
## $ in.window_areas : chr [1:137040] "F18 B18 L18 R18" "F12 B12 L
12 R12" "F12 B12 L12 R12" "F30 B30 L30 R30" ...
## $ in.windows : chr [1:137040] "Single, Clear, Metal" "Double, Clear, Metal, Air" "Double, Low-E, Non-metal, Air, M-Gain" "Double, Clear, Non-metal, Air" ...
## $ upgrade.water_heater_efficiency : chr [1:137040] "Electric Heat Pump, 66 gal, 3.35 UEF" "Electric Heat Pump, 66 gal, 3.35 UEF" "Electric Heat Pump, 80 gal, 3.45 UEF" "Electric Heat Pump, 50 gal, 3.45 UEF" ...
## $ upgrade.clothes_dryer : chr [1:137040] "Electric, Premium, Heat Pump, Ventless, 120% Usage" "Electric, Premium, Heat Pump, Ventless, 100% Usage" "Electric, Premium, Heat Pump, Ventless, 80% Usage" "Electric, Premium, Heat Pump, Ventless, 100% Usage" ...
## [list output truncated]

```

```

# cols_1<-c('in.sqft',
# 'in.bedrooms',
# 'in.building_america_climate_zone',
# 'in.ceiling_fan',
# 'in.cooling_setpoint',
# 'in.cooling_setpoint_has_offset',
# 'in.cooling_setpoint_offset_magnitude',
# 'in.cooling_setpoint_offset_period',
# 'in.ducts',
# 'in.geometry_foundation_type',
# 'in.geometry_wall_type',
# 'in.has_pv',
# 'in.heating_fuel',
# 'in.hot_water_fixtures',
# 'in.hvac_cooling_partial_space_conditioning',
# 'in.hvac_cooling_type',
# 'in.hvac_heating_type',
# 'in.hvac_heating_type_and_fuel',
# 'in.insulation_ceiling',
# 'in.insulation_wall',
# 'in.lighting',
# 'in.misc_extra_refrigerator',
# 'in.misc_freezer',
# 'in.misc_pool_pump',
# 'in.occupants',
# 'in.pv_system_size',
# 'in.refrigerator',
# 'in.roof_material',
# 'in.usage_level',
# 'in.vacancy_status',
# 'in.water_heater_efficiency',
# 'in.water_heater_fuel',
# 'Final_Energy_KWH'
# )
#
# Subset_V1<-Merged_Final[,cols_1]

```

```
# str(Subset_V1)
# non_numeric_cols <- sapply(Subset_V1, function(x) !is.numeric(x))
# Subset_V1[non_numeric_cols] <- lapply(Subset_V1[non_numeric_cols], as.factor)
# str(Subset_V1)
#
#
#
# # Example assuming 'energy_consumption' is the target variable
# model_lm <- lm( Final_Energy_KWH~ ., data = Subset_V1)
# summary(model_lm)
```

```
cols_2<-c(
'Dry Bulb Temperature [°C]',
'Relative Humidity [%]',
'Global Horizontal Radiation [W/m2]',
'in.sqft',
'in.bedrooms',
'in.building_america_climate_zone',
'in.ceiling_fan',
'in.cooling_setpoint',
'in.cooling_setpoint_has_offset',
'in.cooling_setpoint_offset_magnitude',
'in.clothes_dryer',
'in.clothes_washer',
'in.ducts',
'in.geometry_foundation_type',
'in.geometry_wall_type',
'in.has_pv',
'in.heating_fuel',
'in.hot_water_fixtures',
'in.hvac_cooling_partial_space_conditioning',
'in.hvac_cooling_type',
'in.hvac_heating_type',
'in.insulation_ceiling',
'in.insulation_wall',
'in.lighting',
'in.misc_extra_refrigerator',
'in.misc_freezer',
'in.misc_pool_pump',
'in.occupants',
'in.pv_system_size',
'in.refrigerator',
'in.roof_material',
'in.usage_level',
'in.vacancy_status',
'in.water_heater_efficiency',
'in.water_heater_fuel',
'Final_Energy_KWH'
)

Subset_V2<-Merged_Final[,cols_2]
```

```
str(Subset_V2)
```

```
## tibble [137,040 × 36] (S3: tbl_df/tbl/data.frame)
## $ Dry Bulb Temperature [°C] : num [1:137040] 22.4 22.4 22.4 22.4 22.4 ...
## $ Relative Humidity [%] : num [1:137040] 95.2 95.2 95.2 95.2 95.2 ...
## $ Global Horizontal Radiation [W/m2] : num [1:137040] 0 0 0 0 0 0 0 0 0 ...
## $ in.sqft : num [1:137040] 1220 2176 3301 2663 1690 ...
## $ in.bedrooms : num [1:137040] 4 4 5 3 3 4 3 4 3 2 ...
## $ in.building_america_climate_zone : chr [1:137040] "Mixed-Humid" "Mixed-Humid"
"Mixed-Humid" "Mixed-Humid" ...
## $ in.ceiling_fan : chr [1:137040] "Standard Efficiency" "Stand
ard Efficiency" "Standard Efficiency" "Standard Efficiency, No usage" ...
## $ in.cooling_setpoint : chr [1:137040] "75F" "70F" "75F" "75F" ...
## $ in.cooling_setpoint_has_offset : chr [1:137040] "No" "No" "No" "No" ...
## $ in.cooling_setpoint_offset_magnitude : chr [1:137040] "0F" "0F" "0F" "0F" ...
## $ in.clothes_dryer : chr [1:137040] "Electric, 120% Usage" "Gas,
100% Usage" "Electric, 80% Usage" "Propane, 100% Usage" ...
## $ in.clothes_washer : chr [1:137040] "EnergyStar, 120% Usage" "En
ergyStar, 100% Usage" "Standard, 80% Usage" "EnergyStar, 100% Usage" ...
## $ in.ducts : chr [1:137040] "20% Leakage, R-4" "20% Leak
age, R-8" "20% Leakage, R-4" "20% Leakage, R-4" ...
## $ in.geometry_foundation_type : chr [1:137040] "Slab" "Slab" "Slab" "Slab"
...
## $ in.geometry_wall_type : chr [1:137040] "Wood Frame" "Wood Frame" "W
ood Frame" "Steel Frame" ...
## $ in.has_pv : chr [1:137040] "No" "No" "No" "No" ...
## $ in.heating_fuel : chr [1:137040] "Electricity" "Electricity"
"Propane" "Electricity" ...
## $ in.hot_water_fixtures : chr [1:137040] "200% Usage" "100% Usage" "5
0% Usage" "100% Usage" ...
## $ in.hvac_cooling_partial_space_conditioning : chr [1:137040] "100% Conditioned" "100% Con
ditioned" "100% Conditioned" "100% Conditioned" ...
## $ in.hvac_cooling_type : chr [1:137040] "Central AC" "Heat Pump" "Ce
ntral AC" "Heat Pump" ...
## $ in.hvac_heating_type : chr [1:137040] "Ducted Heating" "Ducted Hea
t Pump" "Ducted Heating" "Ducted Heat Pump" ...
## $ in.insulation_ceiling : chr [1:137040] "R-30" "R-30" "R-7" "R-30"
...
## $ in.insulation_wall : chr [1:137040] "Wood Stud, Uninsulated" "Wo
od Stud, R-15" "Wood Stud, Uninsulated" "Wood Stud, R-11" ...
## $ in.lighting : chr [1:137040] "100% Incandescent" "100% In
candescent" "100% LED" "100% CFL" ...
## $ in.misc_extra_refrigerator : chr [1:137040] "EF 15.9" "None" "None" "Non
e" ...
## $ in.misc_freezer : chr [1:137040] "None" "EF 12, National Avera
ge" "None" "EF 12, National Average" ...
## $ in.misc_pool_pump : chr [1:137040] "None" "None" "None" "None"
...
## $ in.occupants : chr [1:137040] "1" "5" "4" "2" ...
## $ in.pv_system_size : chr [1:137040] "None" "None" "None" "None"
...
## $ in.refrigerator : chr [1:137040] "EF 17.6, 100% Usage" "EF 1
7.6, 100% Usage" "EF 17.6, 100% Usage" "EF 17.6, 100% Usage" ...
## $ in.roof_material : chr [1:137040] "Composition Shingles" "Wood
Shingles" "Composition Shingles" "Composition Shingles" ...
## $ in.usage_level : chr [1:137040] "High" "Medium" "Low" "Mediu
m" ...
```

```
## $ in.vacancy_status : chr [1:137040] "Occupied" "Occupied" "Occupied" "Vacant" ...
## $ in.water_heater_efficiency : chr [1:137040] "Electric Standard" "Electric Standard" "Electric Standard" ...
## $ in.water_heater_fuel : chr [1:137040] "Electricity" "Electricity" "Electricity" ...
## $ Final_Energy_KWH : num [1:137040] 24.9 36 19 17 28.1 ...
```

```
non_numeric_cols <- sapply(Subset_V2, function(x) !is.numeric(x))
Subset_V2[non_numeric_cols] <- lapply(Subset_V2[non_numeric_cols], as.factor)
str(Subset_V2)
```



```
## tibble [137,040 × 36] (S3: tbl_df/tbl/data.frame)
## $ Dry Bulb Temperature [°C] : num [1:137040] 22.4 22.4 22.4 22.4 22.4 ...
## $ Relative Humidity [%] : num [1:137040] 95.2 95.2 95.2 95.2 95.2 ...
## $ Global Horizontal Radiation [W/m2] : num [1:137040] 0 0 0 0 0 0 0 0 0 ...
## $ in.sqft : num [1:137040] 1220 2176 3301 2663 1690 ...
## $ in.bedrooms : num [1:137040] 4 4 5 3 3 4 3 4 3 2 ...
## $ in.building_america_climate_zone : Factor w/ 2 levels "Hot-Humid","Mixed-Humi
d": 2 2 2 2 2 2 2 2 2 ...
## $ in.ceiling_fan : Factor w/ 3 levels "None","Standard Efficie
ncy",...: 2 2 2 3 2 2 2 3 2 2 ...
## $ in.cooling_setpoint : Factor w/ 11 levels "60F","62F","65F",...: 8
6 8 8 10 10 7 6 8 7 ...
## $ in.cooling_setpoint_has_offset : Factor w/ 2 levels "No","Yes": 1 1 1 1 2 1
2 2 1 1 ...
## $ in.cooling_setpoint_offset_magnitude : Factor w/ 4 levels "0F","2F","5F",...: 1 1 1
1 4 1 4 4 1 1 ...
## $ in.clothes_dryer : Factor w/ 10 levels "Electric, 100% Usag
e",...: 2 4 3 8 3 2 1 1 1 1 ...
## $ in.clothes_washer : Factor w/ 7 levels "EnergyStar, 100% Usag
e",...: 2 1 7 1 7 6 5 5 1 5 ...
## $ in.ducts : Factor w/ 14 levels "0% Leakage, Uninsulate
d",...: 6 8 6 6 13 10 2 13 9 2 ...
## $ in.geometry_foundation_type : Factor w/ 6 levels "Ambient","Heated Baseme
nt",...: 3 3 3 3 5 3 3 6 6 1 ...
## $ in.geometry_wall_type : Factor w/ 4 levels "Brick","Concrete",...: 4
4 4 3 4 1 4 1 4 4 ...
## $ in.has_pv : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1
1 1 1 ...
## $ in.heating_fuel : Factor w/ 6 levels "Electricity",...: 1 1 6
1 3 1 1 1 3 3 ...
## $ in.hot_water_fixtures : Factor w/ 3 levels "100% Usage","200% Usag
e",...: 2 1 3 1 3 2 1 1 1 1 ...
## $ in.hvac_cooling_partial_space_conditioning: Factor w/ 6 levels "100% Conditioned",...: 1
1 1 1 1 1 1 1 1 1 ...
## $ in.hvac_cooling_type : Factor w/ 4 levels "Central AC","Heat Pum
p",...: 1 2 1 2 1 2 1 2 1 1 ...
## $ in.hvac_heating_type : Factor w/ 4 levels "Ducted Heat Pump",...: 2
1 2 1 2 1 2 1 2 2 ...
## $ in.insulation_ceiling : Factor w/ 8 levels "None","R-13",...: 4 4 7
4 4 7 2 2 4 5 ...
## $ in.insulation_wall : Factor w/ 15 levels "Brick, 12-in, 3-wythe,
R-11",...: 15 12 15 11 13 5 14 5 15 15 ...
## $ in.lighting : Factor w/ 3 levels "100% CFL","100% Incande
scent",...: 2 2 3 1 1 2 3 1 3 3 ...
## $ in.misc_extra_refrigerator : Factor w/ 7 levels "EF 10.2","EF 10.5",...:
3 7 7 7 4 7 7 4 7 7 ...
## $ in.misc_freezer : Factor w/ 2 levels "EF 12, National Averag
e",...: 2 1 2 1 2 2 2 2 2 2 ...
## $ in.misc_pool_pump : Factor w/ 2 levels "1.0 HP Pump",...: 2 2 2
2 2 2 2 2 ...
## $ in.occupants : Factor w/ 10 levels "1","10+","2",...: 1 6 5
3 3 3 3 8 3 3 ...
## $ in.pv_system_size : Factor w/ 8 levels "1.0 kWDC","11.0 kWDC",...:
8 8 8 8 8 8 8 8 8 ...
## $ in.refrigerator : Factor w/ 7 levels "EF 10.2, 100% Usag
```

```

e",...: 4 4 4 4 4 4 4 5 4 ...
## $ in.roof_material          : Factor w/ 7 levels "Asphalt Shingles, Mediu
m",...: 2 7 2 2 1 2 2 5 2 2 ...
## $ in.usage_level            : Factor w/ 3 levels "High","Low","Medium": 1
3 2 3 2 1 3 3 3 3 ...
## $ in.vacancy_status         : Factor w/ 2 levels "Occupied","Vacant": 1 1
1 2 1 1 1 2 1 1 ...
## $ in.water_heater_efficiency : Factor w/ 12 levels "Electric Heat Pump, 80
gal",...: 3 3 3 3 8 3 12 3 7 7 ...
## $ in.water_heater_fuel       : Factor w/ 5 levels "Electricity",...: 1 1 1
1 3 1 5 1 3 3 ...
## $ Final_Energy_KWH          : num [1,127040] 24 0 26 10 17 28 1

```

```

# Example assuming 'energy_consumption' is the target variable
model_lm_2 <- lm( Final_Energy_KWH~ ., data = Subset_V2)
summary(model_lm_2)

```

```
##
## Call:
## lm(formula = Final_Energy_KWH ~ ., data = Subset_V2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -158.492   -6.910   -1.128    5.297   189.168
##
## Coefficients: (13 not defined because of singularities)
##
##              Estimate Std. Error
## (Intercept)      8.139e+00  3.085e+00
## `Dry Bulb Temperature [°C]`      2.641e+00  5.138e-02
## `Relative Humidity [%]`      -4.472e-01  1.282e-02
## `Global Horizontal Radiation [W/m2]`      -1.484e-02  2.498e-04
## in.sqft      6.347e-03  3.037e-05
## in.bedrooms      4.810e-01  4.908e-02
## in.building_america_climate_zoneMixed-Humid      -8.259e-01  1.286e-01
## in.ceiling_fanStandard Efficiency      4.347e-01  8.466e-02
## in.ceiling_fanStandard Efficiency, No usage      1.347e+00  2.286e-01
## in.cooling_setpoint62F      -3.483e+00  6.837e-01
## in.cooling_setpoint65F      -5.291e+00  4.246e-01
## in.cooling_setpoint67F      -9.243e+00  5.329e-01
## in.cooling_setpoint68F      -8.625e+00  3.841e-01
## in.cooling_setpoint70F      -1.049e+01  3.746e-01
## in.cooling_setpoint72F      -1.227e+01  3.744e-01
## in.cooling_setpoint75F      -1.530e+01  3.736e-01
## in.cooling_setpoint76F      -1.647e+01  3.864e-01
## in.cooling_setpoint78F      -1.852e+01  3.766e-01
## in.cooling_setpoint80F      -2.122e+01  4.198e-01
## in.cooling_setpoint_has_offsetYes      2.241e+00  1.706e-01
## in.cooling_setpoint_offset_magnitude2F      -2.472e+00  1.821e-01
## in.cooling_setpoint_offset_magnitude5F      -9.242e-01  1.977e-01
## in.cooling_setpoint_offset_magnitude9F      NA      NA
## in.clothes_dryerElectric, 120% Usage      1.262e+00  6.721e-01
## in.clothes_dryerElectric, 80% Usage      1.576e+00  7.132e-01
## in.clothes_dryerGas, 100% Usage      -3.675e-01  2.326e-01
## in.clothes_dryerGas, 120% Usage      1.091e+00  7.360e-01
## in.clothes_dryerGas, 80% Usage      2.019e+00  7.708e-01
## in.clothes_dryerNone      3.540e-01  3.985e-01
## in.clothes_dryerPropane, 100% Usage      -1.546e+00  4.870e-01
## in.clothes_dryerPropane, 120% Usage      2.023e+00  1.053e+00
## in.clothes_dryerPropane, 80% Usage      2.033e+00  9.498e-01
## in.clothes_washerEnergyStar, 120% Usage      -3.833e-01  9.241e-01
## in.clothes_washerEnergyStar, 80% Usage      -2.738e+00  9.163e-01
## in.clothes_washerNone      -1.161e+00  5.328e-01
## in.clothes_washerStandard, 100% Usage      8.311e-01  1.051e-01
## in.clothes_washerStandard, 120% Usage      1.067e+00  9.306e-01
## in.clothes_washerStandard, 80% Usage      -2.243e+00  9.238e-01
## in.ducts10% Leakage, R-4      3.940e+00  1.596e+00
## in.ducts10% Leakage, R-6      6.251e-01  1.612e+00
## in.ducts10% Leakage, R-8      2.560e+00  1.598e+00
## in.ducts10% Leakage, Uninsulated      1.890e+00  1.597e+00
## in.ducts20% Leakage, R-4      3.858e+00  1.594e+00
## in.ducts20% Leakage, R-6      7.509e-01  1.602e+00
## in.ducts20% Leakage, R-8      2.228e+00  1.595e+00
```

## in.ducts20% Leakage, Uninsulated	2.298e+00	1.594e+00
## in.ducts30% Leakage, R-4	3.508e+00	1.595e+00
## in.ducts30% Leakage, R-6	1.070e+00	1.609e+00
## in.ducts30% Leakage, R-8	2.033e+00	1.598e+00
## in.ducts30% Leakage, Uninsulated	4.351e+00	1.596e+00
## in.ductsNone	-1.273e+00	1.611e+00
## in.geometry_foundation_typeHeated Basement	4.090e+00	1.572e+00
## in.geometry_foundation_typeSlab	1.924e+00	1.551e-01
## in.geometry_foundation_typeUnheated Basement	4.754e+00	3.184e-01
## in.geometry_foundation_typeUnvented Crawlspace	1.538e+00	3.998e-01
## in.geometry_foundation_typeVented Crawlspace	1.353e+00	2.651e-01
## in.geometry_wall_typeConcrete	3.006e+00	3.417e-01
## in.geometry_wall_typeSteel Frame	7.142e-01	3.857e-01
## in.geometry_wall_typeWood Frame	6.992e-01	1.715e-01
## in.has_pvYes	-9.295e-01	1.922e+00
## in.heating_fuelFuel Oil	-1.262e+00	4.556e-01
## in.heating_fuelNatural Gas	3.116e-01	1.206e-01
## in.heating_fuelNone	-2.904e+00	1.572e+00
## in.heating_fuelOther Fuel	1.166e-01	3.724e-01
## in.heating_fuelPropane	5.802e-02	1.917e-01
## in.hot_water_fixtures200% Usage	8.991e+00	6.392e-01
## in.hot_water_fixtures50% Usage	-3.772e+00	5.819e-01
## in.hvac_cooling_partial_space_conditioning20% Conditioned	-1.101e-02	3.168e-01
## in.hvac_cooling_partial_space_conditioning40% Conditioned	-2.424e-02	3.444e-01
## in.hvac_cooling_partial_space_conditioning60% Conditioned	1.008e+00	2.434e-01
## in.hvac_cooling_partial_space_conditioning80% Conditioned	-8.947e-01	2.260e-01
## in.hvac_cooling_partial_space_conditioningNone	-7.382e-01	2.799e-01
## in.hvac_cooling_typeHeat Pump	-9.610e-02	2.200e-01
## in.hvac_cooling_typeNone	NA	NA
## in.hvac_cooling_typeRoom AC	-9.842e-01	3.079e-01
## in.hvac_heating_typeDucted Heating	-1.258e-01	2.088e-01
## in.hvac_heating_typeNon-Ducted Heating	NA	NA
## in.hvac_heating_typeNone	NA	NA
## in.insulation_ceilingR-13	2.408e+00	3.162e-01
## in.insulation_ceilingR-19	2.034e+00	3.123e-01
## in.insulation_ceilingR-30	2.158e+00	3.057e-01
## in.insulation_ceilingR-38	2.292e+00	3.233e-01
## in.insulation_ceilingR-49	-2.194e-01	5.991e-01
## in.insulation_ceilingR-7	2.417e+00	3.304e-01
## in.insulation_ceilingUninsulated	2.540e+00	3.534e-01
## in.insulation_wallBrick, 12-in, 3-wythe, R-15	-1.640e+00	3.550e-01
## in.insulation_wallBrick, 12-in, 3-wythe, R-19	3.800e-01	3.304e-01
## in.insulation_wallBrick, 12-in, 3-wythe, R-7	6.713e-01	2.849e-01
## in.insulation_wallBrick, 12-in, 3-wythe, Uninsulated	1.643e+00	2.070e-01
## in.insulation_wallCMU, 6-in Hollow, R-11	-2.910e+00	4.113e-01
## in.insulation_wallCMU, 6-in Hollow, R-15	-4.983e+00	6.913e-01
## in.insulation_wallCMU, 6-in Hollow, R-19	-2.086e+00	6.196e-01
## in.insulation_wallCMU, 6-in Hollow, R-7	-3.560e+00	7.011e-01
## in.insulation_wallCMU, 6-in Hollow, Uninsulated	NA	NA
## in.insulation_wallWood Stud, R-11	-6.276e-01	1.161e-01
## in.insulation_wallWood Stud, R-15	-2.625e-01	1.657e-01
## in.insulation_wallWood Stud, R-19	-9.650e-01	1.721e-01
## in.insulation_wallWood Stud, R-7	3.907e-01	1.732e-01
## in.insulation_wallWood Stud, Uninsulated	NA	NA
## in.lighting100% Incandescent	5.125e+00	8.285e-02
## in.lighting100% LED	-4.523e-01	9.224e-02

## in.misc_extra_refrigeratorEF 10.5	-2.721e-01	5.009e-01
## in.misc_extra_refrigeratorEF 15.9	-8.621e-01	4.713e-01
## in.misc_extra_refrigeratorEF 17.6	-1.059e+00	4.629e-01
## in.misc_extra_refrigeratorEF 19.9	-1.041e+00	5.318e-01
## in.misc_extra_refrigeratorEF 6.7	2.638e+00	4.960e-01
## in.misc_extra_refrigeratorNone	-2.746e+00	4.564e-01
## in.misc_freezerNone	-1.588e+00	7.327e-02
## in.misc_pool_pumpNone	-9.647e+00	1.123e-01
## in.occupants10+	3.208e+01	9.657e-01
## in.occupants2	2.274e+00	9.580e-02
## in.occupants3	5.168e+00	1.177e-01
## in.occupants4	7.584e+00	1.263e-01
## in.occupants5	1.027e+01	1.685e-01
## in.occupants6	1.215e+01	2.653e-01
## in.occupants7	1.613e+01	3.643e-01
## in.occupants8	1.693e+01	5.668e-01
## in.occupants9	2.144e+01	1.019e+00
## in.pv_system_size11.0 kWDC	-5.713e+01	2.273e+00
## in.pv_system_size13.0 kWDC	-6.406e+01	2.706e+00
## in.pv_system_size3.0 kWDC	-1.626e+01	2.173e+00
## in.pv_system_size5.0 kWDC	-2.644e+01	2.071e+00
## in.pv_system_size7.0 kWDC	-3.582e+01	2.119e+00
## in.pv_system_size9.0 kWDC	-4.650e+01	2.142e+00
## in.pv_system_sizeNone	NA	NA
## in.refrigeratorEF 10.5, 100% Usage	-6.523e-01	5.481e-01
## in.refrigeratorEF 15.9, 100% Usage	-2.410e+00	5.238e-01
## in.refrigeratorEF 17.6, 100% Usage	-2.456e+00	5.187e-01
## in.refrigeratorEF 19.9, 100% Usage	-3.180e+00	5.254e-01
## in.refrigeratorEF 6.7, 100% Usage	1.738e+00	5.609e-01
## in.refrigeratorNone	-4.025e+00	9.318e-01
## in.roof_materialComposition Shingles	2.196e-01	1.089e-01
## in.roof_materialMetal, Dark	5.262e-01	1.620e-01
## in.roof_materialSlate	-1.416e-02	3.790e-01
## in.roof_materialTile, Clay or Ceramic	-5.712e-02	2.533e-01
## in.roof_materialTile, Concrete	1.408e-01	4.126e-01
## in.roof_materialWood Shingles	4.157e-01	1.590e-01
## in.usage_levelLow	NA	NA
## in.usage_levelMedium	NA	NA
## in.vacancy_statusVacant	-2.095e+01	2.088e-01
## in.water_heater_efficiencyElectric Premium	5.334e-01	6.773e-01
## in.water_heater_efficiencyElectric Standard	-1.298e-01	6.633e-01
## in.water_heater_efficiencyElectric Tankless	5.592e+00	7.599e-01
## in.water_heater_efficiencyFuel Oil Standard	-3.255e+00	1.712e+00
## in.water_heater_efficiencyNatural Gas Premium	3.863e-02	7.001e-01
## in.water_heater_efficiencyNatural Gas Standard	-7.699e-02	6.705e-01
## in.water_heater_efficiencyNatural Gas Tankless	3.230e+00	8.627e-01
## in.water_heater_efficiencyOther Fuel	6.457e-01	8.706e-01
## in.water_heater_efficiencyPropane Premium	-1.217e+00	9.995e-01
## in.water_heater_efficiencyPropane Standard	-2.201e-03	7.041e-01
## in.water_heater_efficiencyPropane Tankless	-5.023e-02	9.456e-01
## in.water_heater_fuelFuel Oil	NA	NA
## in.water_heater_fuelNatural Gas	NA	NA
## in.water_heater_fuelOther Fuel	NA	NA
## in.water_heater_fuelPropane	NA	NA
##	t value	Pr(> t)
## (Intercept)	2.638	0.008338 **

## `Dry Bulb Temperature [°C]`	51.402	< 2e-16	***
## `Relative Humidity [%]`	-34.889	< 2e-16	***
## `Global Horizontal Radiation [W/m2]`	-59.420	< 2e-16	***
## in.sqft	208.998	< 2e-16	***
## in.bedrooms	9.800	< 2e-16	***
## in.building_america_climate_zoneMixed-Humid	-6.420	1.37e-10	***
## in.ceiling_fanStandard Efficiency	5.135	2.82e-07	***
## in.ceiling_fanStandard Efficiency, No usage	5.892	3.82e-09	***
## in.cooling_setpoint62F	-5.094	3.51e-07	***
## in.cooling_setpoint65F	-12.461	< 2e-16	***
## in.cooling_setpoint67F	-17.346	< 2e-16	***
## in.cooling_setpoint68F	-22.458	< 2e-16	***
## in.cooling_setpoint70F	-27.988	< 2e-16	***
## in.cooling_setpoint72F	-32.767	< 2e-16	***
## in.cooling_setpoint75F	-40.964	< 2e-16	***
## in.cooling_setpoint76F	-42.625	< 2e-16	***
## in.cooling_setpoint78F	-49.178	< 2e-16	***
## in.cooling_setpoint80F	-50.550	< 2e-16	***
## in.cooling_setpoint_has_offsetYes	13.139	< 2e-16	***
## in.cooling_setpoint_offset_magnitude2F	-13.578	< 2e-16	***
## in.cooling_setpoint_offset_magnitude5F	-4.674	2.95e-06	***
## in.cooling_setpoint_offset_magnitude9F	NA	NA	
## in.clothes_dryerElectric, 120% Usage	1.878	0.060380	.
## in.clothes_dryerElectric, 80% Usage	2.209	0.027151	*
## in.clothes_dryerGas, 100% Usage	-1.580	0.114024	
## in.clothes_dryerGas, 120% Usage	1.483	0.138142	
## in.clothes_dryerGas, 80% Usage	2.619	0.008826	**
## in.clothes_dryerNone	0.888	0.374468	
## in.clothes_dryerPropane, 100% Usage	-3.175	0.001499	**
## in.clothes_dryerPropane, 120% Usage	1.921	0.054717	.
## in.clothes_dryerPropane, 80% Usage	2.140	0.032338	*
## in.clothes_washerEnergyStar, 120% Usage	-0.415	0.678275	
## in.clothes_washerEnergyStar, 80% Usage	-2.989	0.002803	**
## in.clothes_washerNone	-2.179	0.029360	*
## in.clothes_washerStandard, 100% Usage	7.909	2.61e-15	***
## in.clothes_washerStandard, 120% Usage	1.147	0.251401	
## in.clothes_washerStandard, 80% Usage	-2.428	0.015179	*
## in.ducts10% Leakage, R-4	2.468	0.013576	*
## in.ducts10% Leakage, R-6	0.388	0.698162	
## in.ducts10% Leakage, R-8	1.602	0.109156	
## in.ducts10% Leakage, Uninsulated	1.184	0.236434	
## in.ducts20% Leakage, R-4	2.421	0.015471	*
## in.ducts20% Leakage, R-6	0.469	0.639198	
## in.ducts20% Leakage, R-8	1.397	0.162501	
## in.ducts20% Leakage, Uninsulated	1.441	0.149561	
## in.ducts30% Leakage, R-4	2.199	0.027886	*
## in.ducts30% Leakage, R-6	0.665	0.506168	
## in.ducts30% Leakage, R-8	1.272	0.203210	
## in.ducts30% Leakage, Uninsulated	2.726	0.006415	**
## in.ductsNone	-0.790	0.429481	
## in.geometry_foundation_typeHeated Basement	2.602	0.009270	**
## in.geometry_foundation_typeSlab	12.405	< 2e-16	***
## in.geometry_foundation_typeUnheated Basement	14.928	< 2e-16	***
## in.geometry_foundation_typeUnvented Crawlspace	3.847	0.000120	***
## in.geometry_foundation_typeVented Crawlspace	5.103	3.35e-07	***
## in.geometry_wall_typeConcrete	8.798	< 2e-16	***

## in.geometry_wall_typeSteel Frame	1.852	0.064048	.
## in.geometry_wall_typeWood Frame	4.077	4.57e-05	***
## in.has_pvYes	-0.484	0.628633	
## in.heating_fuelFuel Oil	-2.771	0.005588	**
## in.heating_fuelNatural Gas	2.584	0.009775	**
## in.heating_fuelNone	-1.847	0.064806	.
## in.heating_fuelOther Fuel	0.313	0.754225	
## in.heating_fuelPropane	0.303	0.762187	
## in.hot_water_fixtures200% Usage	14.066	< 2e-16	***
## in.hot_water_fixtures50% Usage	-6.482	9.08e-11	***
## in.hvac_cooling_partial_space_conditioning20% Conditioned	-0.035	0.972267	
## in.hvac_cooling_partial_space_conditioning40% Conditioned	-0.070	0.943897	
## in.hvac_cooling_partial_space_conditioning60% Conditioned	4.141	3.46e-05	***
## in.hvac_cooling_partial_space_conditioning80% Conditioned	-3.958	7.55e-05	***
## in.hvac_cooling_partial_space_conditioningNone	-2.638	0.008346	**
## in.hvac_cooling_typeHeat Pump	-0.437	0.662196	
## in.hvac_cooling_typeNone	NA	NA	
## in.hvac_cooling_typeRoom AC	-3.197	0.001391	**
## in.hvac_heating_typeDucted Heating	-0.602	0.546889	
## in.hvac_heating_typeNon-Ducted Heating	NA	NA	
## in.hvac_heating_typeNone	NA	NA	
## in.insulation_ceilingR-13	7.614	2.68e-14	***
## in.insulation_ceilingR-19	6.511	7.49e-11	***
## in.insulation_ceilingR-30	7.059	1.68e-12	***
## in.insulation_ceilingR-38	7.091	1.34e-12	***
## in.insulation_ceilingR-49	-0.366	0.714190	
## in.insulation_ceilingR-7	7.316	2.56e-13	***
## in.insulation_ceilingUninsulated	7.188	6.62e-13	***
## in.insulation_wallBrick, 12-in, 3-wythe, R-15	-4.618	3.87e-06	***
## in.insulation_wallBrick, 12-in, 3-wythe, R-19	1.150	0.250007	
## in.insulation_wallBrick, 12-in, 3-wythe, R-7	2.357	0.018432	*
## in.insulation_wallBrick, 12-in, 3-wythe, Uninsulated	7.935	2.11e-15	***
## in.insulation_wallCMU, 6-in Hollow, R-11	-7.075	1.50e-12	***
## in.insulation_wallCMU, 6-in Hollow, R-15	-7.209	5.65e-13	***
## in.insulation_wallCMU, 6-in Hollow, R-19	-3.366	0.000762	***
## in.insulation_wallCMU, 6-in Hollow, R-7	-5.077	3.84e-07	***
## in.insulation_wallCMU, 6-in Hollow, Uninsulated	NA	NA	
## in.insulation_wallWood Stud, R-11	-5.408	6.39e-08	***
## in.insulation_wallWood Stud, R-15	-1.584	0.113305	
## in.insulation_wallWood Stud, R-19	-5.609	2.04e-08	***
## in.insulation_wallWood Stud, R-7	2.256	0.024087	*
## in.insulation_wallWood Stud, Uninsulated	NA	NA	
## in.lighting100% Incandescent	61.863	< 2e-16	***
## in.lighting100% LED	-4.903	9.43e-07	***
## in.misc_extra_refrigeratorEF 10.5	-0.543	0.586938	
## in.misc_extra_refrigeratorEF 15.9	-1.829	0.067364	.
## in.misc_extra_refrigeratorEF 17.6	-2.287	0.022181	*
## in.misc_extra_refrigeratorEF 19.9	-1.958	0.050202	.
## in.misc_extra_refrigeratorEF 6.7	5.319	1.04e-07	***
## in.misc_extra_refrigeratorNone	-6.016	1.79e-09	***
## in.misc_freezerNone	-21.677	< 2e-16	***
## in.misc_pool_pumpNone	-85.936	< 2e-16	***
## in.occupants10+	33.216	< 2e-16	***
## in.occupants2	23.740	< 2e-16	***
## in.occupants3	43.913	< 2e-16	***
## in.occupants4	60.067	< 2e-16	***

```

## in.occupants5                60.943 < 2e-16 ***
## in.occupants6                45.777 < 2e-16 ***
## in.occupants7                44.281 < 2e-16 ***
## in.occupants8                29.868 < 2e-16 ***
## in.occupants9                21.036 < 2e-16 ***
## in.pv_system_size11.0 kWDC   -25.131 < 2e-16 ***
## in.pv_system_size13.0 kWDC   -23.669 < 2e-16 ***
## in.pv_system_size3.0 kWDC    -7.484 7.27e-14 ***
## in.pv_system_size5.0 kWDC    -12.768 < 2e-16 ***
## in.pv_system_size7.0 kWDC    -16.904 < 2e-16 ***
## in.pv_system_size9.0 kWDC    -21.716 < 2e-16 ***
## in.pv_system_sizeNone        NA      NA
## in.refrigeratorEF 10.5, 100% Usage -1.190 0.234050
## in.refrigeratorEF 15.9, 100% Usage -4.600 4.22e-06 ***
## in.refrigeratorEF 17.6, 100% Usage -4.735 2.20e-06 ***
## in.refrigeratorEF 19.9, 100% Usage -6.053 1.43e-09 ***
## in.refrigeratorEF 6.7, 100% Usage  3.099 0.001943 **
## in.refrigeratorNone          -4.320 1.56e-05 ***
## in.roof_materialComposition Shingles 2.016 0.043828 *
## in.roof_materialMetal, Dark  3.248 0.001163 **
## in.roof_materialSlate        -0.037 0.970203
## in.roof_materialTile, Clay or Ceramic -0.226 0.821557
## in.roof_materialTile, Concrete  0.341 0.732866
## in.roof_materialWood Shingles  2.614 0.008948 **
## in.usage_levelLow            NA      NA
## in.usage_levelMedium         NA      NA
## in.vacancy_statusVacant      -100.338 < 2e-16 ***
## in.water_heater_efficiencyElectric Premium 0.788 0.430966
## in.water_heater_efficiencyElectric Standard -0.196 0.844893
## in.water_heater_efficiencyElectric Tankless 7.360 1.85e-13 ***
## in.water_heater_efficiencyFuel Oil Standard -1.901 0.057244 .
## in.water_heater_efficiencyNatural Gas Premium 0.055 0.955997
## in.water_heater_efficiencyNatural Gas Standard -0.115 0.908582
## in.water_heater_efficiencyNatural Gas Tankless 3.745 0.000181 ***
## in.water_heater_efficiencyOther Fuel 0.742 0.458248
## in.water_heater_efficiencyPropane Premium -1.217 0.223569
## in.water_heater_efficiencyPropane Standard -0.003 0.997506
## in.water_heater_efficiencyPropane Tankless -0.053 0.957641
## in.water_heater_fuelFuel Oil NA      NA
## in.water_heater_fuelNatural Gas NA      NA
## in.water_heater_fuelOther Fuel NA      NA
## in.water_heater_fuelPropane NA      NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.02 on 136898 degrees of freedom
## Multiple R-squared:  0.6603, Adjusted R-squared:  0.6599
## F-statistic: 1887 on 141 and 136898 DF, p-value: < 2.2e-16

```

#Model 3


```
cols_3<-c(
  'Dry Bulb Temperature [°C]',
  'Relative Humidity [%]',
  'in.county',
  'hour',
  'Global Horizontal Radiation [W/m2]',
  'in.sqft',
  'in.bedrooms',
  'in.building_america_climate_zone',
  'in.ceiling_fan',
  'in.cooling_setpoint',
  'in.cooling_setpoint_has_offset',
  'in.cooling_setpoint_offset_magnitude',
  #-----
  'in.clothes_dryer',
  'in.clothes_washer',
  'in.insulation_slab',
  'Wind Speed [m/s]',
  #-----
  'in.ducts',
  'in.geometry_foundation_type',
  'in.geometry_wall_type',
  'in.has_pv',
  'in.heating_fuel',
  'in.hot_water_fixtures',
  'in.hvac_cooling_partial_space_conditioning',
  'in.hvac_cooling_type',
  'in.hvac_heating_type',
  #'in.hvac_heating_type_and_fuel',
  'in.insulation_ceiling',
  'in.insulation_wall',
  'in.lighting',
  'in.misc_extra_refrigerator',
  'in.misc_freezer',
  'in.misc_pool_pump',
  'in.occupants',
  'in.pv_system_size',
  'in.refrigerator',
  'in.roof_material',
  'in.usage_level',
  'in.vacancy_status',
  'in.water_heater_efficiency',
  'in.water_heater_fuel',
  'Final_Energy_KWH'
)

Subset_V3<-Merged_Final[,cols_3]
```

```
str(Substet_V3)
```

```
## tibble [137,040 × 40] (S3: tbl_df/tbl/data.frame)
## $ Dry Bulb Temperature [°C] : num [1:137040] 22.4 22.4 22.4 22.4 22.4 ...
## $ Relative Humidity [%] : num [1:137040] 95.2 95.2 95.2 95.2 95.2 ...
## $ in.county : chr [1:137040] "G4500010" "G4500010" "G4500
010" "G4500010" ...
## $ hour : num [1:137040] 0 0 0 0 0 0 0 0 0 ...
## $ Global Horizontal Radiation [W/m2] : num [1:137040] 0 0 0 0 0 0 0 0 0 ...
## $ in.sqft : num [1:137040] 1220 2176 3301 2663 1690 ...
## $ in.bedrooms : num [1:137040] 4 4 5 3 3 4 3 4 3 2 ...
## $ in.building_america_climate_zone : chr [1:137040] "Mixed-Humid" "Mixed-Humid"
"Mixed-Humid" "Mixed-Humid" ...
## $ in.ceiling_fan : chr [1:137040] "Standard Efficiency" "Stand
ard Efficiency" "Standard Efficiency" "Standard Efficiency, No usage" ...
## $ in.cooling_setpoint : chr [1:137040] "75F" "70F" "75F" "75F" ...
## $ in.cooling_setpoint_has_offset : chr [1:137040] "No" "No" "No" "No" ...
## $ in.cooling_setpoint_offset_magnitude : chr [1:137040] "0F" "0F" "0F" "0F" ...
## $ in.clothes_dryer : chr [1:137040] "Electric, 120% Usage" "Gas,
100% Usage" "Electric, 80% Usage" "Propane, 100% Usage" ...
## $ in.clothes_washer : chr [1:137040] "EnergyStar, 120% Usage" "En
ergyStar, 100% Usage" "Standard, 80% Usage" "EnergyStar, 100% Usage" ...
## $ in.insulation_slab : chr [1:137040] "Uninsulated" "2ft R10 Unde
r, Horizontal" "Uninsulated" "Uninsulated" ...
## $ Wind Speed [m/s] : num [1:137040] 1.09 1.09 1.09 1.09 1.09 ...
## $ in.ducts : chr [1:137040] "20% Leakage, R-4" "20% Leak
age, R-8" "20% Leakage, R-4" "20% Leakage, R-4" ...
## $ in.geometry_foundation_type : chr [1:137040] "Slab" "Slab" "Slab" "Slab"
...
## $ in.geometry_wall_type : chr [1:137040] "Wood Frame" "Wood Frame" "W
ood Frame" "Steel Frame" ...
## $ in.has_pv : chr [1:137040] "No" "No" "No" "No" ...
## $ in.heating_fuel : chr [1:137040] "Electricity" "Electricity"
"Propane" "Electricity" ...
## $ in.hot_water_fixtures : chr [1:137040] "200% Usage" "100% Usage" "5
0% Usage" "100% Usage" ...
## $ in.hvac_cooling_partial_space_conditioning : chr [1:137040] "100% Conditioned" "100% Con
ditioned" "100% Conditioned" "100% Conditioned" ...
## $ in.hvac_cooling_type : chr [1:137040] "Central AC" "Heat Pump" "Ce
ntral AC" "Heat Pump" ...
## $ in.hvac_heating_type : chr [1:137040] "Ducted Heating" "Ducted Hea
t Pump" "Ducted Heating" "Ducted Heat Pump" ...
## $ in.insulation_ceiling : chr [1:137040] "R-30" "R-30" "R-7" "R-30"
...
## $ in.insulation_wall : chr [1:137040] "Wood Stud, Uninsulated" "Wo
od Stud, R-15" "Wood Stud, Uninsulated" "Wood Stud, R-11" ...
## $ in.lighting : chr [1:137040] "100% Incandescent" "100% In
candescent" "100% LED" "100% CFL" ...
## $ in.misc_extra_refrigerator : chr [1:137040] "EF 15.9" "None" "None" "Non
e" ...
## $ in.misc_freezer : chr [1:137040] "None" "EF 12, National Aver
age" "None" "EF 12, National Average" ...
## $ in.misc_pool_pump : chr [1:137040] "None" "None" "None" "None"
...
## $ in.occupants : chr [1:137040] "1" "5" "4" "2" ...
## $ in.pv_system_size : chr [1:137040] "None" "None" "None" "None"
...
```

```
## $ in.refrigerator : chr [1:137040] "EF 17.6, 100% Usage" "EF 17.6, 100% Usage" "EF 17.6, 100% Usage" ...
## $ in.roof_material : chr [1:137040] "Composition Shingles" "Wood Shingles" "Composition Shingles" "Composition Shingles" ...
## $ in.usage_level : chr [1:137040] "High" "Medium" "Low" "Medium" ...
## $ in.vacancy_status : chr [1:137040] "Occupied" "Occupied" "Occupied" "Vacant" ...
## $ in.water_heater_efficiency : chr [1:137040] "Electric Standard" "Electric Standard" "Electric Standard" ...
## $ in.water_heater_fuel : chr [1:137040] "Electricity" "Electricity" "Electricity" "Electricity" ...
## $ Final_Energy_KWH : num [1:137040] 24.9 36 19 17 28.1 ...
```

```
non_numeric_cols <- sapply(Subset_V3, function(x) !is.numeric(x))
Subset_V3[non_numeric_cols] <- lapply(Subset_V3[non_numeric_cols], as.factor)
str(Subset_V3)
```

```

## tibble [137,040 × 40] (S3: tbl_df/tbl/data.frame)
## $ Dry Bulb Temperature [°C] : num [1:137040] 22.4 22.4 22.4 22.4 22.4 ...
## $ Relative Humidity [%] : num [1:137040] 95.2 95.2 95.2 95.2 95.2 ...
## $ in.county : Factor w/ 46 levels "G4500010","G450003
0",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ hour : num [1:137040] 0 0 0 0 0 0 0 0 0 0 ...
## $ Global Horizontal Radiation [W/m2] : num [1:137040] 0 0 0 0 0 0 0 0 0 0 ...
## $ in.sqft : num [1:137040] 1220 2176 3301 2663 1690 ...
## $ in.bedrooms : num [1:137040] 4 4 5 3 3 4 3 4 3 2 ...
## $ in.building_america_climate_zone : Factor w/ 2 levels "Hot-Humid","Mixed-Humi
d": 2 2 2 2 2 2 2 2 2 2 ...
## $ in.ceiling_fan : Factor w/ 3 levels "None","Standard Efficie
ncy",...: 2 2 2 3 2 2 2 3 2 2 ...
## $ in.cooling_setpoint : Factor w/ 11 levels "60F","62F","65F",...: 8
6 8 8 10 10 7 6 8 7 ...
## $ in.cooling_setpoint_has_offset : Factor w/ 2 levels "No","Yes": 1 1 1 1 2 1
2 2 1 1 ...
## $ in.cooling_setpoint_offset_magnitude : Factor w/ 4 levels "0F","2F","5F",...: 1 1 1
1 4 1 4 4 1 1 ...
## $ in.clothes_dryer : Factor w/ 10 levels "Electric, 100% Usag
e",...: 2 4 3 8 3 2 1 1 1 1 ...
## $ in.clothes_washer : Factor w/ 7 levels "EnergyStar, 100% Usag
e",...: 2 1 7 1 7 6 5 5 1 5 ...
## $ in.insulation_slab : Factor w/ 6 levels "2ft R10 Perimeter, Vert
ical",...: 6 2 6 6 5 6 6 5 5 5 ...
## $ Wind Speed [m/s] : num [1:137040] 1.09 1.09 1.09 1.09 1.09 ...
## $ in.ducts : Factor w/ 14 levels "0% Leakage, Uninsulate
d",...: 6 8 6 6 13 10 2 13 9 2 ...
## $ in.geometry_foundation_type : Factor w/ 6 levels "Ambient","Heated Baseme
nt",...: 3 3 3 3 5 3 3 6 6 1 ...
## $ in.geometry_wall_type : Factor w/ 4 levels "Brick","Concrete",...: 4
4 4 3 4 1 4 1 4 4 ...
## $ in.has_pv : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1
1 1 1 1 ...
## $ in.heating_fuel : Factor w/ 6 levels "Electricity",...: 1 1 6
1 3 1 1 1 3 3 ...
## $ in.hot_water_fixtures : Factor w/ 3 levels "100% Usage","200% Usag
e",...: 2 1 3 1 3 2 1 1 1 1 ...
## $ in.hvac_cooling_partial_space_conditioning: Factor w/ 6 levels "100% Conditioned",...: 1
1 1 1 1 1 1 1 1 1 1 ...
## $ in.hvac_cooling_type : Factor w/ 4 levels "Central AC","Heat Pum
p",...: 1 2 1 2 1 2 1 2 1 1 ...
## $ in.hvac_heating_type : Factor w/ 4 levels "Ducted Heat Pump",...: 2
1 2 1 2 1 2 1 2 2 ...
## $ in.insulation_ceiling : Factor w/ 8 levels "None","R-13",...: 4 4 7
4 4 7 2 2 4 5 ...
## $ in.insulation_wall : Factor w/ 15 levels "Brick, 12-in, 3-wythe,
R-11",...: 15 12 15 11 13 5 14 5 15 15 ...
## $ in.lighting : Factor w/ 3 levels "100% CFL","100% Incande
scent",...: 2 2 3 1 1 2 3 1 3 3 ...
## $ in.misc_extra_refrigerator : Factor w/ 7 levels "EF 10.2","EF 10.5",...:
3 7 7 7 4 7 7 4 7 7 ...
## $ in.misc_freezer : Factor w/ 2 levels "EF 12, National Averag
e",...: 2 1 2 1 2 2 2 2 2 2 ...
## $ in.misc_pool_pump : Factor w/ 2 levels "1.0 HP Pump",...: 2 2 2

```

```

2 2 2 2 2 2 2 ...
## $ in.occupants           : Factor w/ 10 levels "1","10+","2",...: 1 6 5
3 3 3 3 8 3 3 ...
## $ in.pv_system_size      : Factor w/ 8 levels "1.0 kWDC","11.0 kWDC",...: 8 8 8 8 8 8 8 8 ...
## $ in.refrigerator        : Factor w/ 7 levels "EF 10.2, 100% Usag
e",...: 4 4 4 4 4 4 4 5 4 ...
## $ in.roof_material       : Factor w/ 7 levels "Asphalt Shingles, Mediu
m",...: 2 7 2 2 1 2 2 5 2 2 ...
## $ in.usage_level         : Factor w/ 3 levels "High","Low","Medium": 1
3 2 3 2 1 3 3 3 3 ...
## $ in.vacancy_status      : Factor w/ 2 levels "Occupied","Vacant": 1 1
1 2 1 1 1 2 1 1 ...
## $ in.water_heater_efficiency : Factor w/ 12 levels "Electric Heat Pump, 80
gal",...: 3 3 3 3 8 3 12 3 7 7 ...
## $ in.water_heater_fuel    : Factor w/ 5 levels "Electricity",...: 1 1 1
1 3 1 5 1 3 3 ...
## $ Final_Energy_KWH       : num [1,127040] 24 0 26 10 17 28 1

```

```

# Example assuming 'energy_consumption' is the target variable
model_lm_3 <- lm( Final_Energy_KWH~ ., data = Subset_V3)
summary(model_lm_3)

```

```
##
## Call:
## lm(formula = Final_Energy_KWH ~ ., data = Subset_V3)
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
	-157.908	-6.464	-1.105	4.906	187.829

```
##
## Coefficients: (15 not defined because of singularities)
##
## (Intercept)
## `Dry Bulb Temperature [°C]`
## `Relative Humidity [%]`
## in.countyG450030
## in.countyG450050
## in.countyG450070
## in.countyG450090
## in.countyG450110
## in.countyG450130
## in.countyG450150
## in.countyG450170
## in.countyG450190
## in.countyG450210
## in.countyG450230
## in.countyG450250
## in.countyG450270
## in.countyG450290
## in.countyG450310
## in.countyG450330
## in.countyG450350
## in.countyG450370
## in.countyG450390
## in.countyG450410
## in.countyG450430
## in.countyG450450
## in.countyG450470
## in.countyG450490
## in.countyG450510
## in.countyG450530
## in.countyG450550
## in.countyG450570
## in.countyG450590
## in.countyG450610
## in.countyG450630
## in.countyG450650
## in.countyG450670
## in.countyG450690
## in.countyG450710
## in.countyG450730
## in.countyG450750
## in.countyG450770
## in.countyG450790
## in.countyG450810
## in.countyG450830
## in.countyG450850
```

	Estimate	Std. Error
(Intercept)	8.221e+01	7.272e+00
`Dry Bulb Temperature [°C]`	1.538e-01	1.600e-01
`Relative Humidity [%]`	-7.067e-01	3.725e-02
in.countyG450030	1.344e-01	5.452e-01
in.countyG450050	-1.649e+00	9.585e-01
in.countyG450070	-5.319e+00	5.307e-01
in.countyG450090	-1.480e+00	7.590e-01
in.countyG450110	-3.638e+00	7.299e-01
in.countyG450130	2.257e+00	6.676e-01
in.countyG450150	-5.014e+00	6.006e-01
in.countyG450170	-2.824e+00	8.129e-01
in.countyG450190	-4.544e+00	5.850e-01
in.countyG450210	4.701e+00	5.986e-01
in.countyG450230	9.592e-01	6.526e-01
in.countyG450250	-8.846e+00	6.560e-01
in.countyG450270	-6.914e+00	6.558e-01
in.countyG450290	4.600e+00	7.685e-01
in.countyG450310	-6.400e+00	5.721e-01
in.countyG450330	-9.242e+00	6.666e-01
in.countyG450350	-5.263e+00	6.116e-01
in.countyG450370	7.911e-01	6.877e-01
in.countyG450390	-1.598e+00	6.516e-01
in.countyG450410	-6.789e+00	5.378e-01
in.countyG450430	-1.079e+01	6.009e-01
in.countyG450450	-3.867e+00	5.092e-01
in.countyG450470	4.067e-01	5.585e-01
in.countyG450490	2.654e+00	8.910e-01
in.countyG450510	-1.177e+01	5.551e-01
in.countyG450530	7.181e-01	8.190e-01
in.countyG450550	-6.822e+00	5.757e-01
in.countyG450570	-9.116e+00	6.258e-01
in.countyG450590	1.286e+00	5.618e-01
in.countyG450610	-5.967e+00	7.864e-01
in.countyG450630	-4.695e+00	5.387e-01
in.countyG450650	1.110e+00	7.811e-01
in.countyG450670	-7.766e+00	7.168e-01
in.countyG450690	-8.200e+00	6.580e-01
in.countyG450710	-5.575e+00	6.253e-01
in.countyG450730	-8.132e-01	5.616e-01
in.countyG450750	-3.000e+00	5.486e-01
in.countyG450770	-2.313e+00	5.529e-01
in.countyG450790	-2.349e+00	5.214e-01
in.countyG450810	-2.680e-01	7.377e-01
in.countyG450830	-4.479e+00	5.144e-01
in.countyG450850	-5.997e+00	5.481e-01

```

## in.pv_system_size5.0 kWDC -12.192 < 2e-16 ***
## in.pv_system_size7.0 kWDC -16.889 < 2e-16 ***
## in.pv_system_size9.0 kWDC -21.640 < 2e-16 ***
## in.pv_system_sizeNone NA NA
## in.refrigeratorEF 10.5, 100% Usage -0.958 0.338152
## in.refrigeratorEF 15.9, 100% Usage -4.479 7.52e-06 ***
## in.refrigeratorEF 17.6, 100% Usage -4.281 1.86e-05 ***
## in.refrigeratorEF 19.9, 100% Usage -5.691 1.26e-08 ***
## in.refrigeratorEF 6.7, 100% Usage 3.642 0.000271 ***
## in.refrigeratorNone -4.530 5.92e-06 ***
## in.roof_materialComposition Shingles 2.070 0.038456 *
## in.roof_materialMetal, Dark 3.727 0.000194 ***
## in.roof_materialSlate 0.959 0.337496
## in.roof_materialTile, Clay or Ceramic -0.989 0.322724
## in.roof_materialTile, Concrete 0.555 0.578902
## in.roof_materialWood Shingles 2.829 0.004675 **
## in.usage_levelLow NA NA
## in.usage_levelMedium NA NA
## in.vacancy_statusVacant -102.611 < 2e-16 ***
## in.water_heater_efficiencyElectric Premium 1.276 0.202133
## in.water_heater_efficiencyElectric Standard 0.170 0.865336
## in.water_heater_efficiencyElectric Tankless 7.973 1.56e-15 ***
## in.water_heater_efficiencyFuel Oil Standard -1.703 0.088597 .
## in.water_heater_efficiencyNatural Gas Premium -0.074 0.940918
## in.water_heater_efficiencyNatural Gas Standard 0.059 0.953329
## in.water_heater_efficiencyNatural Gas Tankless 3.272 0.001067 **
## in.water_heater_efficiencyOther Fuel 1.150 0.250173
## in.water_heater_efficiencyPropane Premium -1.463 0.143392
## in.water_heater_efficiencyPropane Standard 0.494 0.621319
## in.water_heater_efficiencyPropane Tankless -0.723 0.469791
## in.water_heater_fuelFuel Oil NA NA
## in.water_heater_fuelNatural Gas NA NA
## in.water_heater_fuelOther Fuel NA NA
## in.water_heater_fuelPropane NA NA
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.59 on 136848 degrees of freedom
## Multiple R-squared: 0.6823, Adjusted R-squared: 0.6818
## F-statistic: 1539 on 191 and 136848 DF, p-value: < 2.2e-16

```

Final Model

```

library(arrow)
library(tidyverse)

cols_4<-c('hour',
          'in.county',
          'Dry Bulb Temperature [°C]', 'Relative Humidity [%]', 'Wind Speed [m/s]',
          'Wind Direction [Deg]', 'Direct Normal Radiation [W/m2]', 'Diffuse Horizontal Radiati
on [W/m2]',
          'Global Horizontal Radiation [W/m2]', 'in.sqft',
          'in.bedrooms',
          'in.building_america_climate_zone',
          'in.ceiling_fan',
          'in.clothes_dryer',
          'in.clothes_washer',
          'in.cooling_setpoint',
          'in.cooling_setpoint_has_offset',
          'in.cooling_setpoint_offset_magnitude',
          'in.dishwasher',
          'in.ducts',
          'in.geometry_foundation_type',
          'in.geometry_wall_type',
          'in.geometry_stories',
          'in.has_pv',
          'in.heating_fuel',
          'in.hot_water_fixtures',
          'in.hvac_cooling_partial_space_conditioning',
          'in.hvac_cooling_type',
          'in.hvac_heating_type',
          'in.hvac_heating_type_and_fuel',
          'in.infiltration',
          'in.insulation_ceiling',
          'in.insulation_wall',
          'in.lighting',
          'in.misc_extra_refrigerator',
          'in.misc_freezer',
          'in.misc_pool_pump',
          'in.occupants',
          'in.pv_system_size',
          'in.refrigerator',
          'in.roof_material',
          'in.usage_level',
          'in.vacancy_status',
          'in.water_heater_efficiency',
          'in.water_heater_fuel',
          'Final_Energy_KWH'

)

Subset_V4<-Merged_Final[,cols_4]

non_numeric_cols <- sapply(Substet_V4, function(x) !is.numeric(x))
Subset_V4[non_numeric_cols] <- lapply(Substet_V4[non_numeric_cols], as.factor)

```



```
#xGBoost Model
set.seed(123)

# Split data into training and test sets (e.g., 80% training, 20% test)
train_indices <- sample(1:nrow(Subset_V4), size = 0.7 * nrow(Subset_V4))
train_data <- Subset_V4[train_indices, ]
test_data <- Subset_V4[-train_indices, ]

library (xgboost)
```

```
## Warning: package 'xgboost' was built under R version 4.3.2
```

```
##
## Attaching package: 'xgboost'
```

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

```
# Convert training data to DMatrix format
dtrain <- xgb.DMatrix(data = data.matrix(train_data[, -which(names(train_data) == "Final_Energy_KWH")]),
                      label = train_data$Final_Energy_KWH)

params <- list(
  objective = "reg:squarederror",
  eta = 0.1,
  max_depth = 8,
  subsample = 0.5,
  colsample_bytree = 0.5
)

nrounds <- 3000 # Number of boosting rounds. Adjust based on your dataset and needs

xgb_model <- xgboost(params = params, data = dtrain, nrounds = nrounds)
```

```
## [2968] train-rmse:1.006519
## [2969] train-rmse:1.006065
## [2970] train-rmse:1.005599
## [2971] train-rmse:1.005191
## [2972] train-rmse:1.004762
## [2973] train-rmse:1.004264
## [2974] train-rmse:1.003765
## [2975] train-rmse:1.003328
## [2976] train-rmse:1.003082
## [2977] train-rmse:1.002697
## [2978] train-rmse:1.002335
## [2979] train-rmse:1.001887
## [2980] train-rmse:1.001607
## [2981] train-rmse:1.001198
## [2982] train-rmse:1.000703
## [2983] train-rmse:1.000167
## [2984] train-rmse:0.999723
## [2985] train-rmse:0.999183
## [2986] train-rmse:0.998777
## [2987] train-rmse:0.998336
## [2988] train-rmse:0.997923
## [2989] train-rmse:0.997582
## [2990] train-rmse:0.997263
## [2991] train-rmse:0.996861
## [2992] train-rmse:0.996386
## [2993] train-rmse:0.995986
## [2994] train-rmse:0.995568
## [2995] train-rmse:0.995237
## [2996] train-rmse:0.994799
## [2997] train-rmse:0.994577
## [2998] train-rmse:0.994053
## [2999] train-rmse:0.993743
## [3000] train-rmse:0.993360
```

```
summary(xgb_model)
```

##	Length	Class	Mode
## handle	1	xgb.Booster.handle	externalptr
## raw	47439846	-none-	raw
## niter	1	-none-	numeric
## evaluation_log	2	data.table	list
## call	13	-none-	call
## params	6	-none-	list
## callbacks	2	-none-	list
## feature_names	45	-none-	character
## nfeatures	1	-none-	numeric

```
# Assuming you have a trained XGBoost model 'xgb_model' and a test set 'test_data'

# Predict on the test set
dtest <- xgb.DMatrix(data = data.matrix(test_data[, -which(names(test_data) == "Final_Energy_
KWH")]))
predictions1 <- predict(xgb_model, dtest)

# Compute RMSE
rmse <- sqrt(mean((predictions1 - test_data$Final_Energy_KWH)^2))
#print(paste("RMSE:", rmse))

# Compute R-squared
SST <- sum((test_data$Final_Energy_KWH - mean(test_data$Final_Energy_KWH))^2)
SSR <- sum((predictions1 - test_data$Final_Energy_KWH)^2)
r_squared <- 1 - SSR/SST
print(paste("R-squared:", r_squared))
```

```
## [1] "R-squared: 0.918774005776453"
```

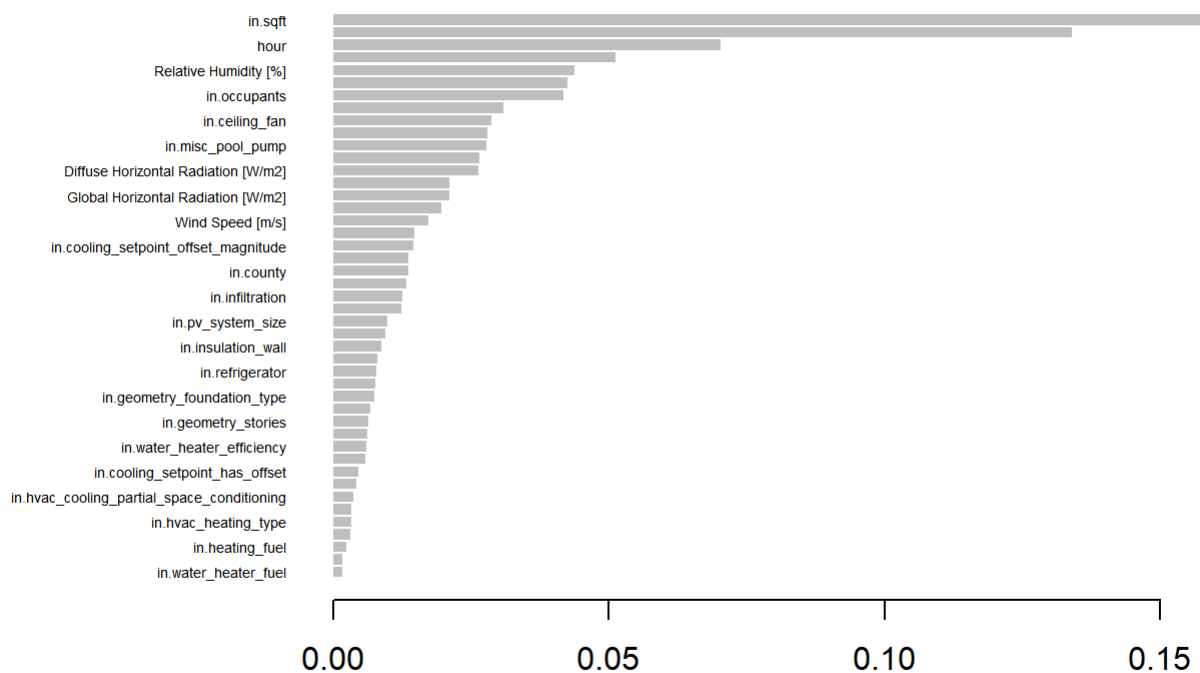
```
#range(predictions1-test_data$Final_Energy_KWH)
#summary(predictions1-test_data$Final_Energy_KWH)
# Visualize feature importance
```

```
importance_matrix <- xgb.importance(model = xgb_model)
print(importance_matrix)
```

##	Feature	Gain	Cover
## 1:	in.sqft	0.158177489	0.033965064
## 2:	Dry Bulb Temperature [°C]	0.134084002	0.069716993
## 3:	hour	0.070327063	0.034350854
## 4:	in.vacancy_status	0.051136147	0.004614631
## 5:	Relative Humidity [%]	0.043782236	0.066849953
## 6:	in.cooling_setpoint	0.042394795	0.030129696
## 7:	in.occupants	0.041805860	0.027784610
## 8:	in.usage_level	0.030905286	0.007420890
## 9:	in.ceiling_fan	0.028719415	0.009261957
## 10:	in.hot_water_fixtures	0.028027720	0.007087027
## 11:	in.misc_pool_pump	0.027747947	0.008502793
## 12:	in.has_pv	0.026476178	0.002185749
## 13:	Diffuse Horizontal Radiation [W/m2]	0.026324658	0.037677688
## 14:	in.lighting	0.021123791	0.013279487
## 15:	Global Horizontal Radiation [W/m2]	0.021013058	0.034260479
## 16:	in.bedrooms	0.019538042	0.016220458
## 17:	Wind Speed [m/s]	0.017243818	0.073707550
## 18:	Direct Normal Radiation [W/m2]	0.014680574	0.047107451
## 19:	in.cooling_setpoint_offset_magnitude	0.014570460	0.028221685
## 20:	in.clothes_dryer	0.013617540	0.012943325
## 21:	in.county	0.013545079	0.044087776
## 22:	in.ducts	0.013262754	0.034186903
## 23:	in.infiltration	0.012573600	0.036375677
## 24:	Wind Direction [Deg]	0.012256148	0.067447838
## 25:	in.pv_system_size	0.009841358	0.004205402
## 26:	in.clothes_washer	0.009336975	0.020165367
## 27:	in.insulation_wall	0.008691952	0.026202826
## 28:	in.insulation_ceiling	0.007876470	0.021353204
## 29:	in.refrigerator	0.007766143	0.017650896
## 30:	in.misc_extra_refrigerator	0.007600425	0.015955038
## 31:	in.geometry_foundation_type	0.007442659	0.015981068
## 32:	in.dishwasher	0.006719948	0.018027180
## 33:	in.geometry_stories	0.006332543	0.008442002
## 34:	in.roof_material	0.006110545	0.015426848
## 35:	in.water_heater_efficiency	0.005955791	0.016615164
## 36:	in.hvac_heating_type_and_fuel	0.005696044	0.014423808
## 37:	in.cooling_setpoint_has_offset	0.004577795	0.010023987
## 38:	in.hvac_cooling_type	0.004103965	0.006512882
## 39:	in.hvac_cooling_partial_space_conditioning	0.003673111	0.007434931
## 40:	in.misc_freezer	0.003274584	0.006725249
## 41:	in.hvac_heating_type	0.003263433	0.007528100
## 42:	in.geometry_wall_type	0.003083014	0.006954731
## 43:	in.heating_fuel	0.002250814	0.005583242
## 44:	in.building_america_climate_zone	0.001539205	0.003679797
## 45:	in.water_heater_fuel	0.001529566	0.003721744
##	Feature	Gain	Cover
##	Frequency		
## 1:	0.037852894		
## 2:	0.059497254		
## 3:	0.042519669		
## 4:	0.003675853		
## 5:	0.053061512		
## 6:	0.035223430		
## 7:	0.028204529		

```
## 8: 0.008260176
## 9: 0.014616464
## 10: 0.008645451
## 11: 0.008014320
## 12: 0.001238277
## 13: 0.029687163
## 14: 0.016442398
## 15: 0.024659099
## 16: 0.023501775
## 17: 0.052799165
## 18: 0.035152971
## 19: 0.022974083
## 20: 0.020424072
## 21: 0.042084922
## 22: 0.039597875
## 23: 0.040128565
## 24: 0.049109820
## 25: 0.001383692
## 26: 0.026628949
## 27: 0.028186539
## 28: 0.026143232
## 29: 0.019782447
## 30: 0.017565242
## 31: 0.018145403
## 32: 0.023800101
## 33: 0.011036555
## 34: 0.019449641
## 35: 0.017051042
## 36: 0.015827757
## 37: 0.012625627
## 38: 0.009873234
## 39: 0.009867238
## 40: 0.009252596
## 41: 0.009356036
## 42: 0.009628877
## 43: 0.009528435
## 44: 0.002636960
## 45: 0.004858663
##      Frequency
```

```
# Visualize feature importance
xgb.plot.importance(importance_matrix)
```



```

Test_Optimied_Variables <- Subset_V4
#Test_Optimied_Variables$in.insulation_wall<-"Brick, 12-in, 3-wythe, R-7"
#Test_Optimied_Variables$in.hvac_cooling_partial_space_conditioning<-"40% Conditioned"
#Test_Optimied_Variables$in.usage_Level<-"Low"
Test_Optimied_Variables$`Dry Bulb Temperature [°C]`<-Test_Optimied_Variables$`Dry Bulb Temperature [°C]`+5

dtest2 <- xgb.DMatrix(data = data.matrix(Test_Optimied_Variables[, -which(names(test_data) == "Final_Energy_KWH")]))

predictions1 <- predict(xgb_model, dtest2)
#actual vs predicted reduced due to upgrades
df_new = data.frame(predictions1,Subset_V4$Final_Energy_KWH)
#df_new
#sum(predictions1)
#sum(Subset_V4$Final_Energy_KWH)

data <- data.frame(
  Category = rep(c("Predicted Energy", "Current Energy"),each=nrow(predictions1)),
  Energy_Value = c(predictions1, test_data$Final_Energy_KWH)
)

```

```

## Warning in rep(c("Predicted Energy", "Current Energy"), each =
## nrow(predictions1)): first element used of 'each' argument

```

```
Test_Optimied_Variables$predictions1<-predictions1
```

```
#glimpse(Test_Optimied_Variables)
```

```
# Load necessary libraries
```

```
library(ggplot2)
```

```
library(dplyr)
```

```
library(maps)
```

```
##
```

```
## Attaching package: 'maps'
```

```
## The following object is masked from 'package:purrr':
```

```
##
```

```
##      map
```

```
#install.packages("mapdata")
```

```
library(mapdata)
```

```
## Warning: package 'mapdata' was built under R version 4.3.2
```

```

ICPSRNAME = c("ABBEVILLE", "AIKEN", "ALLENDALE", "ANDERSON", "BAMBERG", "BARNWELL", "BEAUFORT", "BERKELEY", "CALHOUN", "CHARLESTON",
               "CHEROKEE", "CHESTER", "CHESTERFIELD", "CLARENDON", "COLLETON", "DARLINGTON",
               "DILLON", "DORCHESTER", "EDGEFIELD",
               "FAIRFIELD", "FLORENCE", "GEORGETOWN", "GREENVILLE", "GREENWOOD", "HAMPTON",
               "HORRY", "JASPER", "KERSHAW", "LANCASTER",
               "LAURENS", "LEE", "LEXINGTON", "MARION", "MARLBORO", "MCCORMICK", "NEWBERRY",
               "OCONEE", "ORANGEBURG", "PICKENS",
               "RICHLAND", "SALUDA", "SPARTANBURG", "SUMTER", "UNION", "WILLIAMSBURG", "YORK")

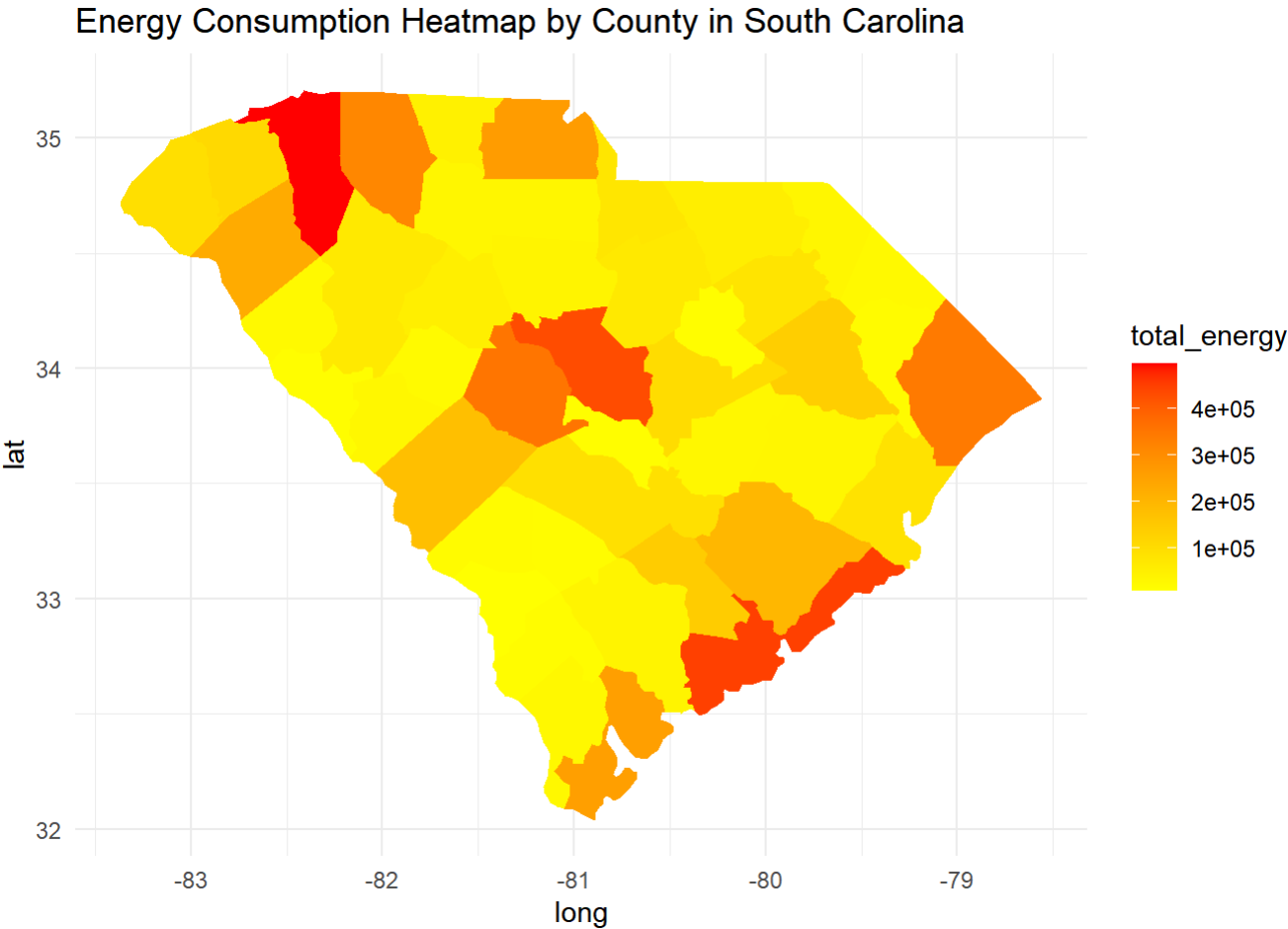
GISJOIN = c("G4500010", "G4500030", "G4500050", "G4500070", "G4500090", "G4500110", "G4500130", "G4500150", "G4500170", "G4500190",
            "G4500210", "G4500230", "G4500250", "G4500270", "G4500290", "G4500310", "G4500330", "G4500350", "G4500370", "G4500390",
            "G4500410", "G4500430", "G4500450", "G4500470", "G4500490", "G4500510", "G4500530", "G4500550", "G4500570", "G4500590",
            "G4500610", "G4500630", "G4500670", "G4500690", "G4500650", "G4500710", "G4500730", "G4500750", "G4500770", "G4500790",
            "G4500810", "G4500830", "G4500850", "G4500870", "G4500890", "G4500910")

# Calculate total energy by county
List_Name<-data.frame(tolower(ICPSRNAME),(GISJOIN))
#List_Name
energy_data <- Subset_V4 %>%
  group_by(in.county) %>%
  summarize(total_energy = sum(Final_Energy_KWH, na.rm = TRUE))
energy_data$County_name<-List_Name$tolower.ICPSRNAME.[match(energy_data$in.county,List_Name$X.GISJOIN.)]

county_map <- map_data("county", region = "south carolina")
county_map$subregion<-tolower(county_map$subregion)
energy_data$in.county<-tolower(energy_data$in.county)

# Merge energy data with the county map
merged_data <- merge(county_map, energy_data, by.x = "subregion", by.y = "County_name", all.x = TRUE)
#merged_data
# Create the heatmap
ggplot(merged_data, aes(x = long, y = lat, group = group, fill = total_energy)) +
  geom_polygon() +
  scale_fill_gradient(low = "yellow", high = "red") +
  labs(title = "Energy Consumption Heatmap by County in South Carolina") +
  theme_minimal()

```

```

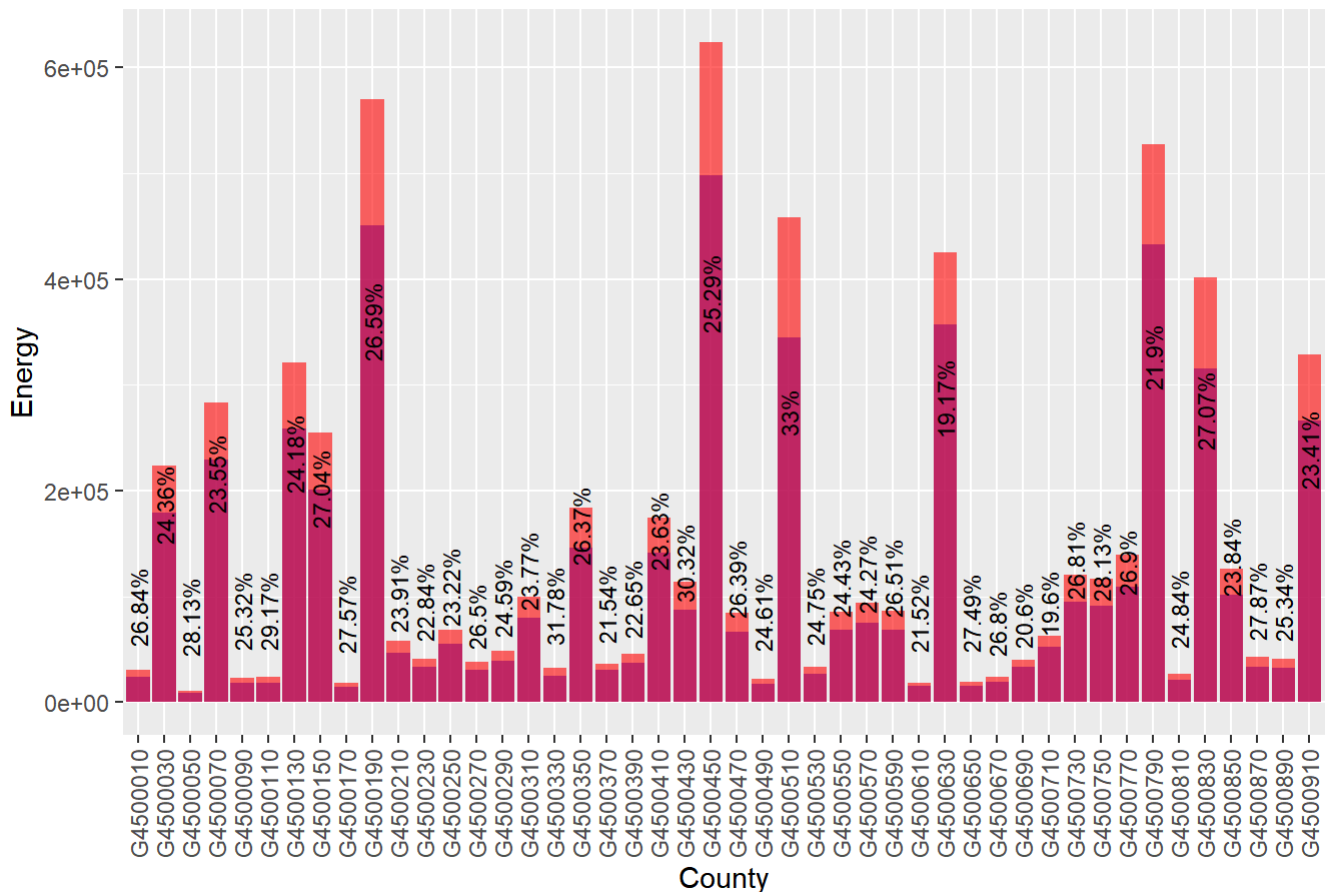
Summarize_Predictions<-Test_Optimied_Variables %>%group_by(in.county) %>%
  summarize(total_energy = sum(Final_Energy_KWH, na.rm = TRUE),predicted_energy=sum(predictio
ns1,na.rm=TRUE))
#str(Summarize_Predictions)
#
#
# library(ggplot2)
#
# # Create a bar plot
# ggplot(data = Summarize_Predictions, aes(x = in.county)) +
#   geom_bar(aes(y = total_energy), stat = "identity", fill = "blue", alpha = 0.6) +
#   geom_bar(aes(y = predicted_energy), stat = "identity", fill = "red", alpha = 0.6) +
#   labs(title = "Total Energy in July vs Predicted Energy (with increase by 5 C) by County",
#         x = "County", y = "Energy") +
#   theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))
#

# Calculate percentage difference
Summarize_Predictions$percentage_diff <- ( (Summarize_Predictions$predicted_energy -Summarize
_Predictions$total_energy)/Summarize_Predictions$total_energy) * 100

# Create a bar plot with percentage difference labels
ggplot(data = Summarize_Predictions, aes(x = in.county)) +
  geom_bar(aes(y = total_energy), stat = "identity", fill = "blue", alpha = 0.6) +
  geom_bar(aes(y = predicted_energy), stat = "identity", fill = "red", alpha = 0.6) +
  geom_text(aes(y = pmax(predicted_energy, total_energy),
                    label = paste0(round(percentage_diff, 2), "%")),
            position = position_stack(vjust = 0.5),
            size = 3,
            color = "black",
            angle = 90,
            hjust = -0.5) +
  labs(title = "Total Energy in July vs Predicted Energy (with increase by 5 C) by County",
        x = "County", y = "Energy") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))

```

Total Energy in July vs Predicted Energy (with increase by 5 C) by County



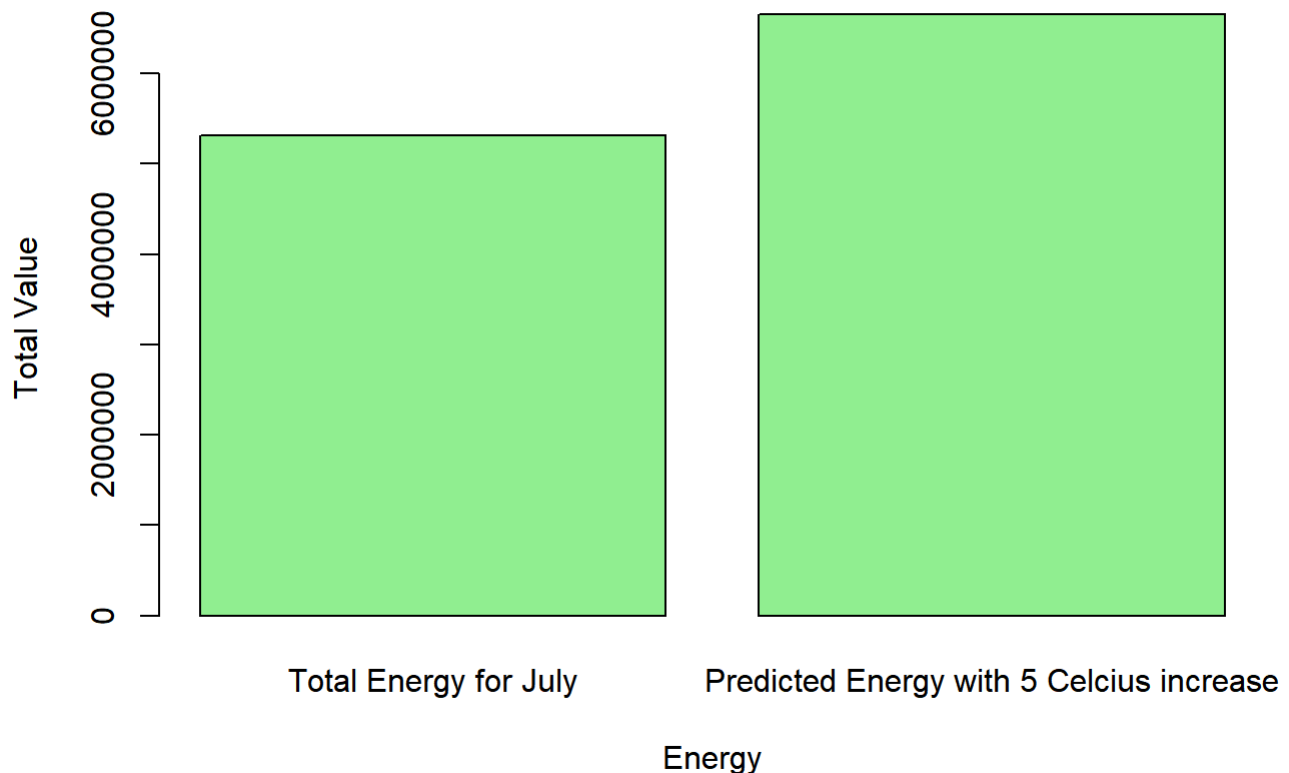
```

sum_energy<-sum(Summarize_Predictions$total_energy)
Predicted<-sum(Summarize_Predictions$predicted_energy)
values <- c(sum_energy, Predicted)

options(scipen = 999)
# Creating a bar plot
barplot(values, names.arg = c("Total Energy for July", "Predicted Energy with 5 Celcius incre
ase"), col = "lightgreen",
        xlab = "Energy", ylab = "Total Value", main = "Comparison of Total and Predicted Ener
gy")

```

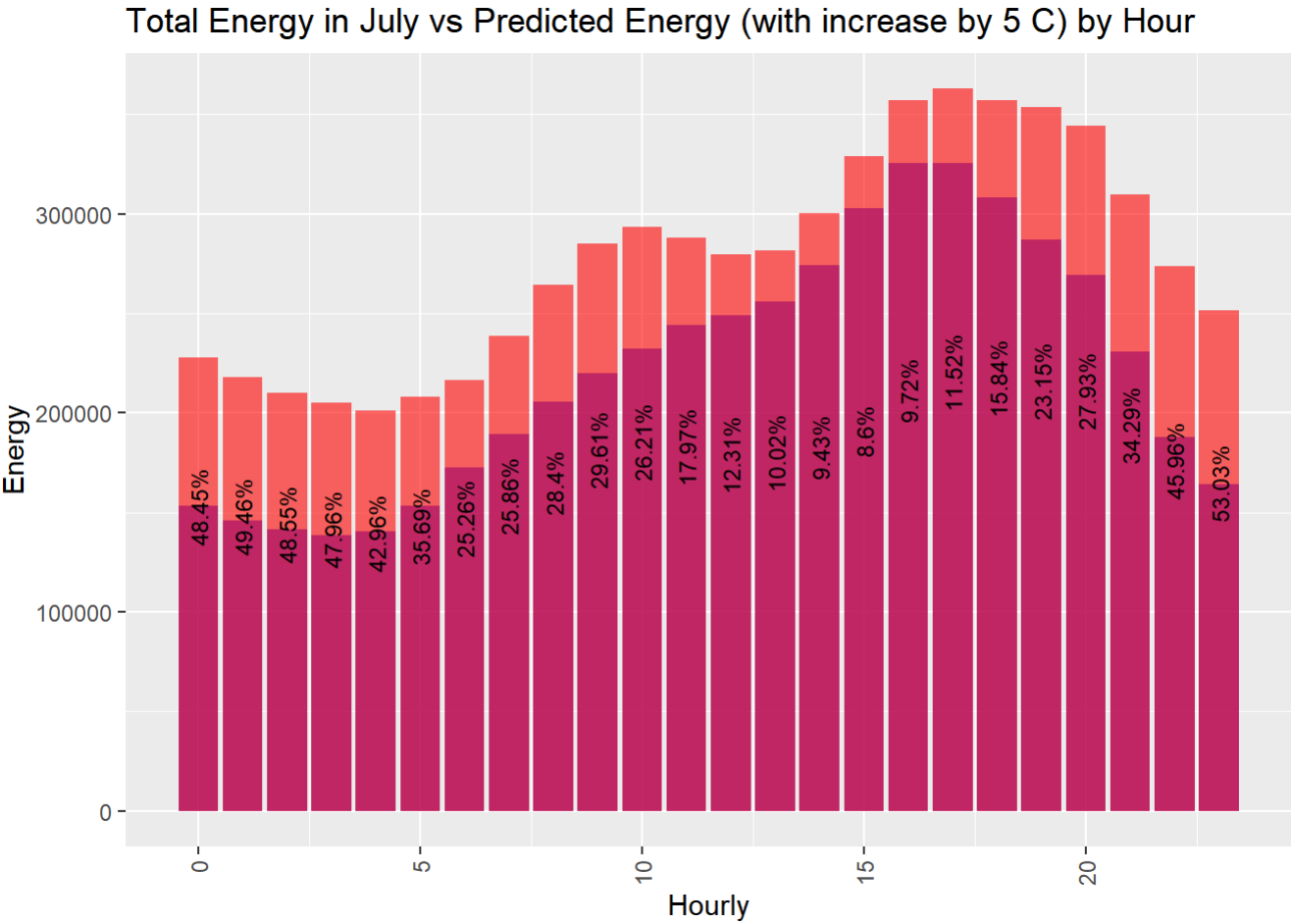
Comparison of Total and Predicted Energy



```
Predictions_hour<-Test_Optimied_Variables %>%group_by(hour) %>%
  summarize(total_energy = sum(Final_Energy_KWH, na.rm = TRUE),predicted_energy=sum(predictio
ns1,na.rm=TRUE))

# Calculate percentage difference
Predictions_hour$percentage_diff <- ((Predictions_hour$predicted_energy - Predictions_hour$to
tal_energy) / Predictions_hour$total_energy) * 100

#since temp increse people keep appliances on often
ggplot(data = Predictions_hour, aes(x = hour)) +
  geom_bar(aes(y = total_energy), stat = "identity", fill = "blue", alpha = 0.6) +
  geom_bar(aes(y = predicted_energy), stat = "identity", fill = "red", alpha = 0.6) +
  geom_text(aes(y = pmax(predicted_energy, total_energy),
    label = paste0(round(percentage_diff, 2), "%")),
    position = position_stack(vjust = 0.5),
    size = 3,
    color = "black",
    angle = 90,
    hjust = -0.5) +
  labs(title = "Total Energy in July vs Predicted Energy (with increase by 5 C) by Hour",
    x = "Hourly", y = "Energy") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))
```



```

Test_Optimied_Variables_reduce <-Subset_V4
Test_Optimied_Variables_reduce$`Dry Bulb Temperature [°C]`<-Test_Optimied_Variables$`Dry Bulb Temperature [°C]`+5
#Test_Optimied_Variables_reduce$in.ceiling_fan<-"Standard Efficiency, No usage"
#Test_Optimied_Variables$in.insulation_wall<-"Brick, 12-in, 3-wythe, R-7"
#Test_Optimied_Variables$in.hvac_cooling_partial_space_conditioning<-"40% Conditioned"
# Test_Optimied_Variables$in.usage_level<-"Low"
#Test_Optimied_Variables_reduce$in.pv_system_size<-"1.0 kWDC"

# Assuming 'Test_Optimized_Variables_reduce' is your dataset

# Replace "none" with "1kw" in the 'in.pv_system_size' column
#Test_Optimied_Variables_reduce$in.pv_system_size <- ifelse(Test_Optimied_Variables_reduce$in.pv_system_size == "None" , "1.0 kWDC",Test_Optimied_Variables_reduce$in.pv_system_size)

#unique(Test_Optimied_Variables_reduce$in.hvac_cooling_type)
#Test_Optimied_Variables_reduce$in.hvac_cooling_type<-"Central AC"
#Test_Optimied_Variables_reduce$in.hvac_cooling_partial_space_conditioning<-"40% Conditioned"
#Test_Optimied_Variables_reduce$in.ducts<-"None"
#Test_Optimied_Variables_reduce$in.hot_water_fixtures<-"50% Usage"
dtest2 <- xgb.DMatrix(data = data.matrix(Test_Optimied_Variables_reduce[, -which(names(test_data) == "Final_Energy_KWH")]))

predictions1 <- predict(xgb_model, dtest2)
#actual vs predicted reduced due to upgrades
df_new = data.frame(predictions1,Subset_V4$Final_Energy_KWH)
#df_new
sum(predictions1)

```

```
## [1] 7283192
```

```
sum(Subset_V4$Final_Energy_KWH)
```

```
## [1] 5317227
```

```

Test_Optimied_Variables$predictions1<-predictions1
# Calculate the sum of predictions1 and Subset_V4$Final_Energy_KWH
sum_predictions <- sum(predictions1)
sum_final_energy <- sum(Subset_V4$Final_Energy_KWH)

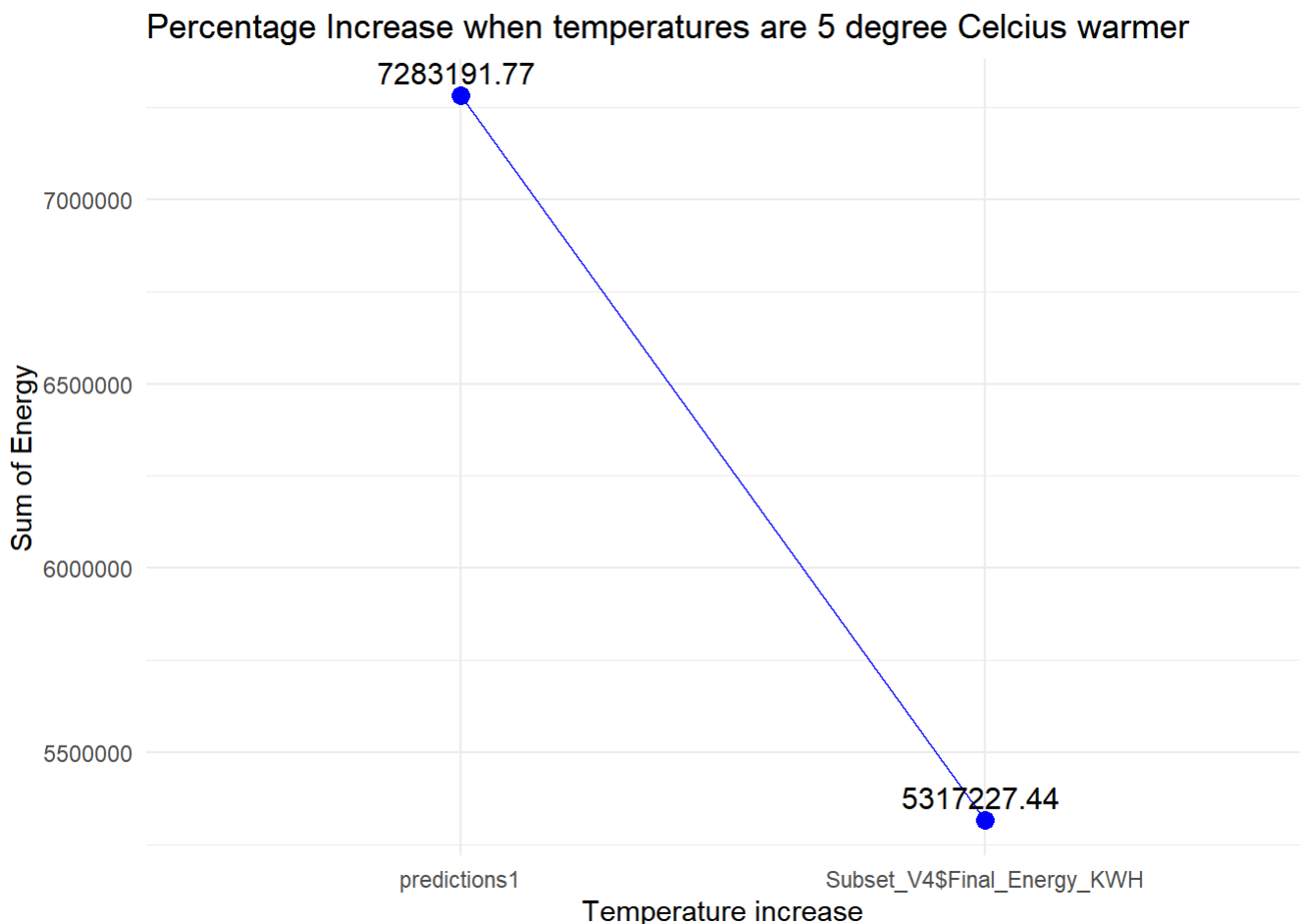
# Calculate the percentage increase
percent_increase <- ((sum_final_energy - sum_predictions) / sum_predictions) * 100

# Create a data frame for plotting
data <- data.frame(
  Variable = c("predictions1", "Subset_V4$Final_Energy_KWH"),
  Sum = c(sum_predictions, sum_final_energy)
)

# Load necessary libraries
library(ggplot2)

# Create a line plot
ggplot(data, aes(x = Variable, y = Sum, group = 1)) +
  geom_line(color = "blue") +
  geom_point(color = "blue", size = 3) +
  geom_text(aes(label = paste(round(Sum, 2), "")), vjust = -0.5, size = 4) +
  labs(title = "Percentage Increase when temperatures are 5 degree Celcius warmer",
       x = "Temperature increase",
       y = "Sum of Energy") +
  theme_minimal()

```



```

Test_Optimied_Variables_reduce <-Subset_V4
Test_Optimied_Variables_reduce$`Dry Bulb Temperature [°C]`<-Test_Optimied_Variables$`Dry Bulb Temperature [°C]`+5
Test_Optimied_Variables_reduce$in.ceiling_fan<-"Standard Efficiency, No usage"
#Test_Optimied_Variables$in.insulation_wall<-"Brick, 12-in, 3-wythe, R-7"
#Test_Optimied_Variables$in.hvac_cooling_partial_space_conditioning<-"40% Conditioned"
# Test_Optimied_Variables$in.usage_level<-"Low"
#Test_Optimied_Variables_reduce$in.cooling_setpoint<-"80F"
#Test_Optimied_Variables_reduce$in.pv_system_size<-"1.0 kWDC"

# Assuming 'Test_Optimized_Variables_reduce' is your dataset

# Replace "none" with "1kw" in the 'in.pv_system_size' column
#Test_Optimied_Variables_reduce$in.pv_system_size <- ifelse(Test_Optimied_Variables_reduce$in.pv_system_size == "None" , "1.0 kWDC",Test_Optimied_Variables_reduce$in.pv_system_size)

#unique(Test_Optimied_Variables_reduce$in.hvac_cooling_type)
#Test_Optimied_Variables_reduce$in.hvac_cooling_type<-"Central AC"
#Test_Optimied_Variables_reduce$in.hvac_cooling_partial_space_conditioning<-"40% Conditioned"
#Test_Optimied_Variables_reduce$in.ducts<-"None"
Test_Optimied_Variables_reduce$in.hot_water_fixtures<-"50% Usage"
dtest2 <- xgb.DMatrix(data = data.matrix(Test_Optimied_Variables_reduce[, -which(names(test_data) == "Final_Energy_KWH")]))

predictions1 <- predict(xgb_model, dtest2)
#actual vs predicted reduced due to upgrades
df_new = data.frame(predictions1,Subset_V4$Final_Energy_KWH)
#df_new
#sum(predictions1)
#sum(Subset_V4$Final_Energy_KWH)

Test_Optimied_Variables$predictions1<-predictions1
# Calculate the sum of predictions1 and Subset_V4$Final_Energy_KWH
sum_predictions <- sum(predictions1)
Final_temp_increase <-sum_predictions
sum_final_energy <- sum(Subset_V4$Final_Energy_KWH)

# Calculate the percentage increase
percent_increase <- ((sum_final_energy - sum_predictions) / sum_predictions) * 100

# Create a data frame for plotting
data <- data.frame(
  Variable = c("predictions1", "Subset_V4$Final_Energy_KWH"),
  Sum = c(sum_predictions, sum_final_energy)
)

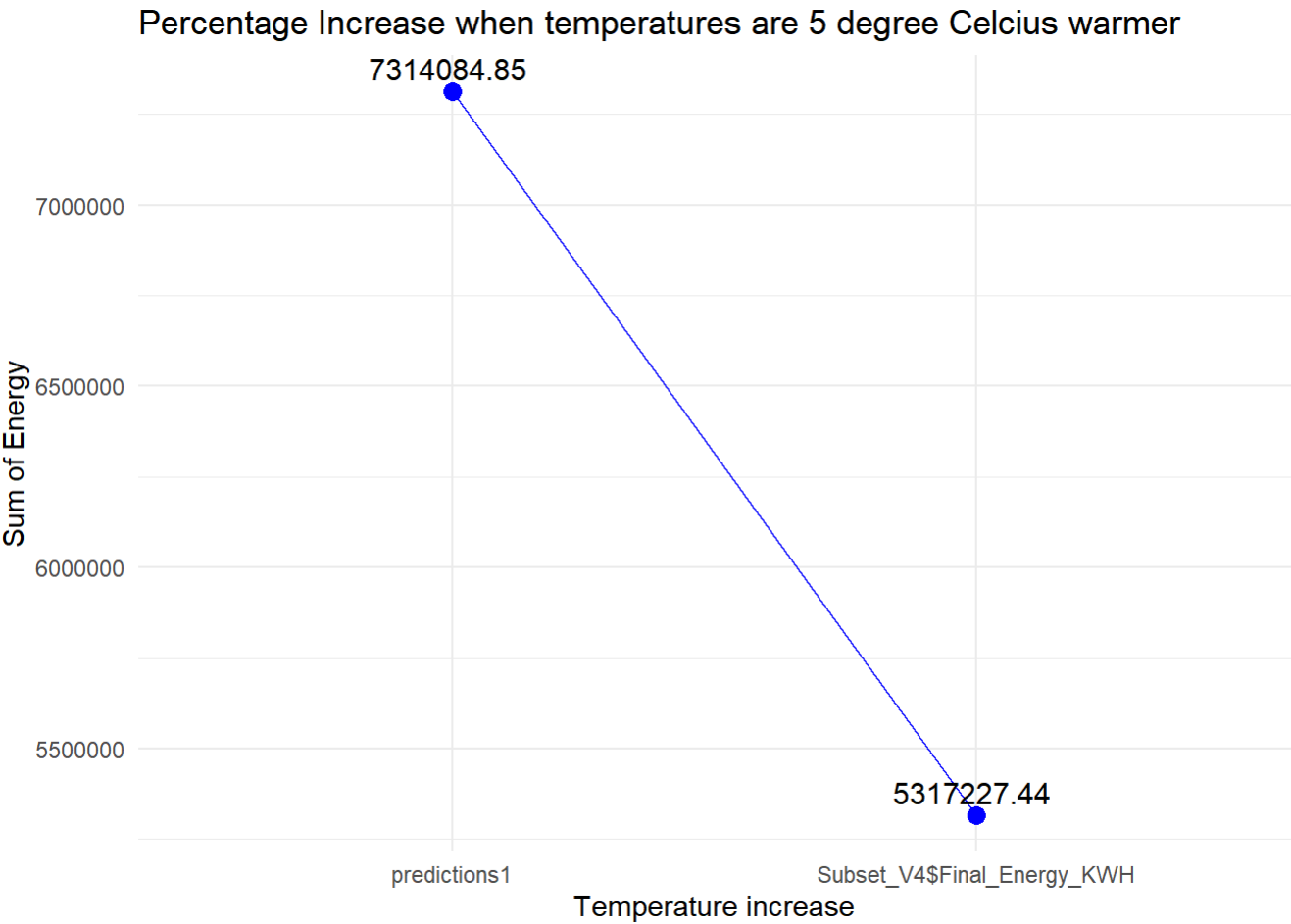
# Load necessary libraries
library(ggplot2)

# Create a line plot
ggplot(data, aes(x = Variable, y = Sum, group = 1)) +
  geom_line(color = "blue") +
  geom_point(color = "blue", size = 3) +
  geom_text(aes(label = paste(round(Sum, 2), "")), vjust = -0.5, size = 4) +
  labs(title = "Percentage Increase when temperatures are 5 degree Celcius warmer",

```



```
x = "Temperature increase",
y = "Sum of Energy") +
theme_minimal()
```



```

Test_Optimied_Variables_reduce <-Subset_V4
Test_Optimied_Variables_reduce$`Dry Bulb Temperature [°C]`<-Test_Optimied_Variables$`Dry Bulb Temperature [°C]`+5
Test_Optimied_Variables_reduce$in.ceiling_fan<-"Standard Efficiency, No usage"
#Test_Optimied_Variables$in.insulation_wall<-"Brick, 12-in, 3-wythe, R-7"
#Test_Optimied_Variables$in.hvac_cooling_partial_space_conditioning<-"40% Conditioned"
# Test_Optimied_Variables$in.usage_level<-"Low"
#Test_Optimied_Variables_reduce$in.cooling_setpoint<-"80F"
#Test_Optimied_Variables_reduce$in.pv_system_size<-"1.0 kWDC"

# Assuming 'Test_Optimized_Variables_reduce' is your dataset

# Replace "none" with "1kw" in the 'in.pv_system_size' column
#Test_Optimied_Variables_reduce$in.pv_system_size <- ifelse(Test_Optimied_Variables_reduce$in.pv_system_size == "None" , "1.0 kWDC",Test_Optimied_Variables_reduce$in.pv_system_size)

#unique(Test_Optimied_Variables_reduce$in.hvac_cooling_type)
#Test_Optimied_Variables_reduce$in.hvac_cooling_type<-"Central AC"
#Test_Optimied_Variables_reduce$in.hvac_cooling_partial_space_conditioning<-"40% Conditioned"
#Test_Optimied_Variables_reduce$in.ducts<-"None"
Test_Optimied_Variables_reduce$in.infiltration<-"ACH50 15"

Test_Optimied_Variables_reduce$in.hot_water_fixtures<-"50% Usage"
dtest2 <- xgb.DMatrix(data = data.matrix(Test_Optimied_Variables_reduce[, -which(names(test_data) == "Final_Energy_KWH")]))

predictions1 <- predict(xgb_model, dtest2)
#actual vs predicted reduced due to upgrades
df_new = data.frame(predictions1,Final_temp_increase )
#df_new
#sum(predictions1)
#sum(Subset_V4$Final_Energy_KWH)

Test_Optimied_Variables$predictions1<-predictions1
# Calculate the sum of predictions1 and Subset_V4$Final_Energy_KWH
sum_predictions <- sum(predictions1)
sum_final_energy <- Final_temp_increase

# Calculate the percentage increase
percent_increase <- ((sum_final_energy - sum_predictions) / sum_predictions) * 100

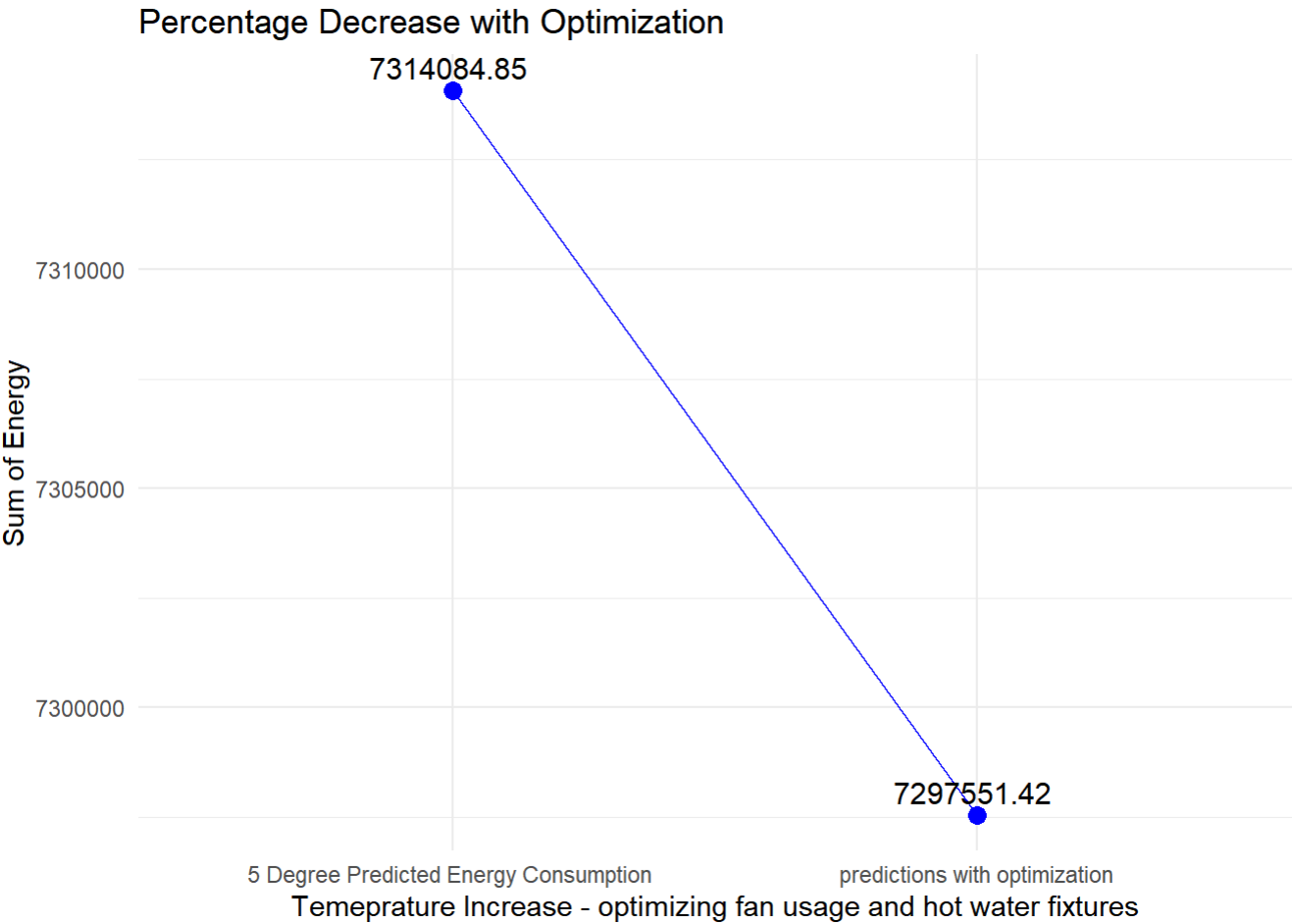
# Create a data frame for plotting
data <- data.frame(
  Variable = c("predictions with optimization", "5 Degree Predicted Energy Consumption "),
  Sum = c(sum_predictions, sum_final_energy)
)

# Load necessary libraries
library(ggplot2)

# Create a Line plot
ggplot(data, aes(x = Variable, y = Sum, group = 1)) +
  geom_line(color = "blue") +
  geom_point(color = "blue", size = 3) +
  geom_text(aes(label = paste(round(Sum, 2), "")), vjust = -0.5, size = 4) +

```

```
labs(title = "Percentage Decrease with Optimization",
      x = "Tempeprature Increase - optimizing fan usage and hot water fixtures",
      y = "Sum of Energy") +
theme_minimal()
```



```

Test_Optimied_Variables_reduce <-Subset_V4
Test_Optimied_Variables_reduce$`Dry Bulb Temperature [°C]`<-Test_Optimied_Variables$`Dry Bulb Temperature [°C]`+5
#Test_Optimied_Variables_reduce$in.ceiling_fan<-"Standard Efficiency, No usage"
#Test_Optimied_Variables$in.insulation_wall<-"Brick, 12-in, 3-wythe, R-7"
#Test_Optimied_Variables$in.hvac_cooling_partial_space_conditioning<-"40% Conditioned"
# Test_Optimied_Variables$in.usage_level<-"Low"
#Test_Optimied_Variables_reduce$in.cooling_setpoint<-"80F"
#Test_Optimied_Variables_reduce$in.pv_system_size<-"1.0 kWDC"

# Assuming 'Test_Optimized_Variables_reduce' is your dataset

# Replace "none" with "1kw" in the 'in.pv_system_size' column
Test_Optimied_Variables_reduce$in.pv_system_size <- ifelse(Test_Optimied_Variables_reduce$in.pv_system_size == "None" , "1.0 kWDC",Test_Optimied_Variables_reduce$in.pv_system_size)

#unique(Test_Optimied_Variables_reduce$in.hvac_cooling_type)
#Test_Optimied_Variables_reduce$in.hvac_cooling_type<-"Central AC"
#Test_Optimied_Variables_reduce$in.hvac_cooling_partial_space_conditioning<-"40% Conditioned"
#Test_Optimied_Variables_reduce$in.ducts<-"None"
#Test_Optimied_Variables_reduce$in.infiltration<-"ACH50 15"

#Test_Optimied_Variables_reduce$in.hot_water_fixtures<-"50% Usage"
dtest2 <- xgb.DMatrix(data = data.matrix(Test_Optimied_Variables_reduce[, -which(names(test_data) == "Final_Energy_KWH"))))

predictions1 <- predict(xgb_model, dtest2)
#actual vs predicted reduced due to upgrades
df_new = data.frame(predictions1,Final_temp_increase )
#df_new
#sum(predictions1)
#sum(Subset_V4$Final_Energy_KWH)

Test_Optimied_Variables$predictions1<-predictions1
# Calculate the sum of predictions1 and Subset_V4$Final_Energy_KWH
sum_predictions <- sum(predictions1)
sum_final_energy <- Final_temp_increase

# Calculate the percentage increase
percent_increase <- ((sum_final_energy - sum_predictions) / sum_predictions) * 100

# Create a data frame for plotting
data <- data.frame(
  Variable = c("predictions with optimization", "5 Degree Predicted Energy Consumption "),
  Sum = c(sum_predictions, sum_final_energy)
)

# Load necessary libraries
library(ggplot2)

# Create a Line plot
ggplot(data, aes(x = Variable, y = Sum, group = 1)) +
  geom_line(color = "blue") +
  geom_point(color = "blue", size = 3) +
  geom_text(aes(label = paste(round(Sum, 2), "")), vjust = -0.5, size = 4) +

```

```
labs(title = "Percentage Decrease with Optimization",
      x = "Tempeprature Increase - optimizing PV usage",
      y = "Sum of Energy") +
theme_minimal()
```

