IST 687 M009: Introduction to Data Science

Project on

"*Energy Consumption Forecast and Analysis*"

By Group Members:

Mansi Gopani

Nandita Ghildyal

Keerthi Aiyappan

Janella Bauske

Nagul Pandian Chinnasamy Ramkumar

Abstract

E.SC, the leading energy provider in South Carolina, is dedicated to delivering reliable and sustainable electricity to its customers. With the escalation of summer temperatures due to climate change, E.SC anticipates an increase in demand for electricity, particularly during peak heatwaves. This surge poses a threat to the power grid's stability, potentially resulting in disruptive blackouts.

To address this challenge, E.SC is actively initiating a comprehensive study to understand the primary determinants of residential electricity demand. The project seeks to uncover influential factors driving energy consumption, analyse its correlation with temperature changes in July, identify significant determinants, provide actionable recommendations for energy conservation, and demonstrate the efficacy of energy reduction strategies through predictive modelling.

1. Introduction

The project's primary objectives encompass uncovering the key factors influencing energy consumption, investigating its relationship with temperature variations in July, pinpointing significant determinants, offering practical recommendations for energy conservation, and demonstrating the effectiveness of energy reduction strategies using predictive modelling.

Our initial steps involve exploratory data analysis (EDA) and data cleaning, followed by the development of predictive models. Within the dataset under analysis lies a wealth of information about individual residences, encompassing unique building IDs and specific attributes associated with each property. Furthermore, the dataset provides intricate hour-by-hour details of energy consumption patterns in these residences.

Moreover, the dataset meticulously integrates hour-by-hour weather data, systematically organised according to geographic regions. This weather information serves as a critical external factor significantly shaping energy usage patterns across diverse regions.

The object of this project is to answer the following questions:

1. Identification of Factors Affecting Energy Consumption.
2. Analysis of Energy Consumption Variation with Temperature Increase in July.
3. Identification of Significant Factors Impacting Energy Consumption.
4. Recommendations for Decreasing Energy Consumption.
5. Demonstration of Energy Consumption Reduction via a Model.

2. Exploratory Data Analysis

To understand the most significant factors that are driving the consumption of energy we start cleaning the data. We began with the first dataset “static_housing” dataset by pulling it from a parquet file.

```r
library (arrow)
library(tidyverse)
library(writexl)
static_housing <- read_parquet("https://intro-datascience.s3.us-east-2.amazonaws.com/SC-data/static_house_info.parquet")
str(static_housing)
write_xlsx(static_housing, "static_housing.xlsx") #writng to excel for easier access (time consuming to pull repititively)
```

The static housing dataset, comprising 5710 rows and 171 columns,with around 8 numeric columns and rest 163-character types. Each row is unique based on its building id. Key variables such as building id, size of house labelled as in.sqft, and characteristics of the house infrastructure like wall to window ratio, cooling and heating system setting and a lot more that could contribute to energy consumption.

```
$ in.heating_fuel                                : chr  "Natural Gas" "Natural Gas" "Natural Gas" "Natural Gas" ...
$ in.heating_setpoint                            : chr  "70F" "65F" "70F" "68F" ...
$ in.heating_setpoint_has_offset                 : chr  "No" "Yes" "No" "Yes" ...
$ in.heating_setpoint_offset_magnitude           : chr  "0F" "3F" "0F" "3F" ...
$ in.heating_setpoint_offset_period              : chr  "None" "Night -4h" "None" "Night -3h" ...
$ in.holiday_lighting                            : chr  "No Exterior Use" "No Exterior Use" "No Exterior Use" "No Exterior Use" ...
$ in.hot_water_distribution                      : chr  "Uninsulated" "Uninsulated" "Uninsulated" "Uninsulated" ...
$ in.hot_water_fixtures                          : chr  "100% Usage" "100% Usage" "50% Usage" "50% Usage" ...
$ in.hvac_cooling_efficiency                     : chr  "AC, SEER 15" "AC, SEER 13" "AC, SEER 13" "None" ...
$ in.hvac_cooling_partial_space_conditioning     : chr  "100% Conditioned" "100% Conditioned" "100% Conditioned" "None" ...
$ in.hvac_cooling_type                           : chr  "Central AC" "Central AC" "Central AC" "None" ...
$ in.hvac_has_ducts                              : chr  "Yes" "Yes" "Yes" "No" ...
$ in.hvac_has_shared_system                      : chr  "None" "None" "None" "None" ...
$ in.hvac_has_zonal_electric_heating             : chr  "No" "No" "No" "No" ...
```

For example, it consists of a variable called “in.pv_system_size” which signifies how big the solar panels and other variables describing the insulation of the house.

```
in.pv_system_size
Length:5710
Class :character
Mode  :character
```

```
in.insulation_ceiling in.insulation_floor
Length:5710           Length:5710
Class :character      Class :character
Mode  :character      Mode  :character
```

On doing some descriptive analysis on the numerical columns we saw that the number of bedrooms and the size of house (in square feet) both had a roughly normal distribution.

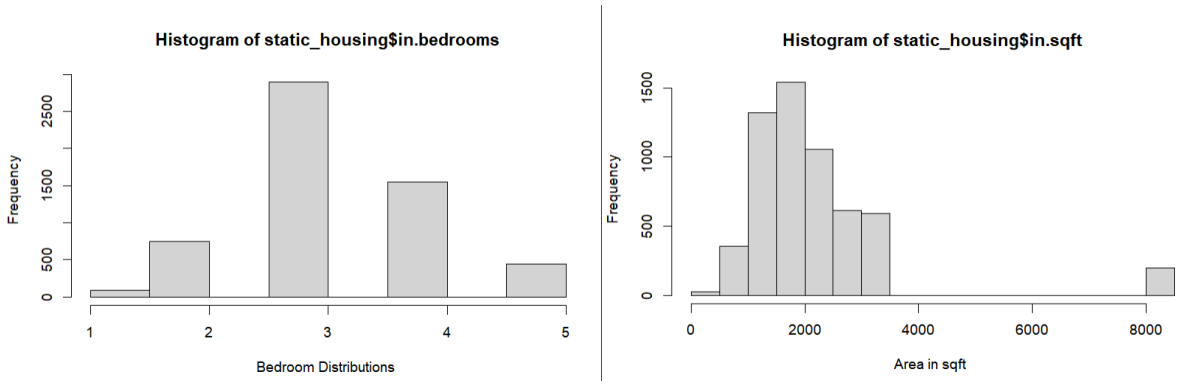


Fig. 2.1

Two other variables were discrete in nature. Regional Energy Deployment Type consisted of two categories: 95 and 96. The number of stories in a building had three possible values (1, 2, or 3) and displayed a right-skewed distribution (see Figure 2.3 in the Appendix).

To understand the correlation between numerical variables we created a correlation matrix. The analysis revealed a robust correlation between weather and reed, indicating a strong association between these variables. Consequently, we have opted to retain the variable at this stage of the analysis.

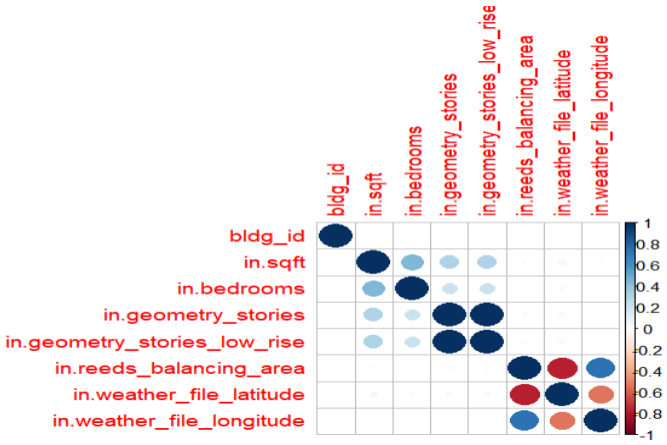


Fig. 2.4

We cleaned the data by checking for any NA columns and got rid of all the rows that consisted of just 1 value as that would imply no variability. This brought down the columns from 171 to 93 in number. Lastly, we checked for the percentage of blanks in columns. We saw lower than 5% of blanks and decided to leave it as it is.

We also plotted the density of buildings by counties and saw an uneven distribution, which led to the conclusion that we cannot aggregate any of these factors for energy consumption by county. We see that Greenville has the highest density, followed by Colleton, Georgetown, Horry and so on.

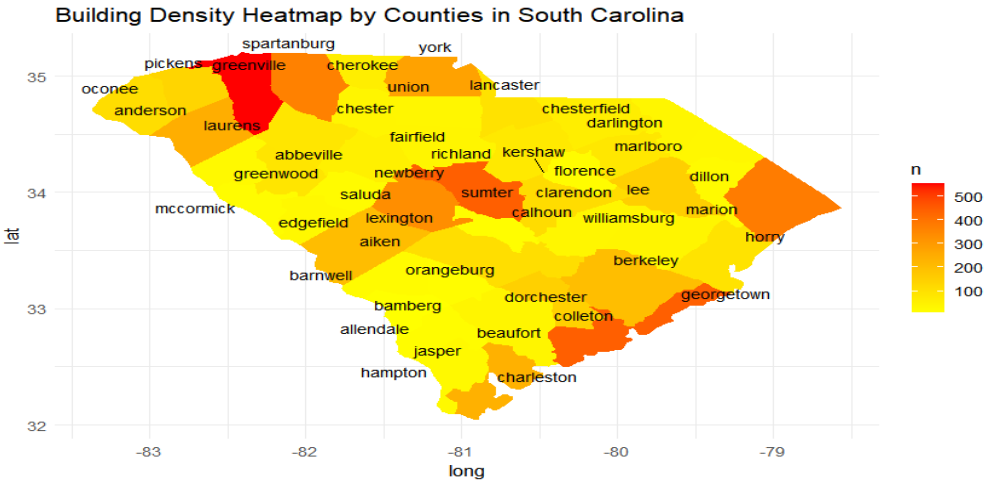


Fig. 2.5

Moving on to the energy consumption data. For each of the buildings there was a data record for each hour for each month for 2018. We decided to pick July of 2018 for our analysis. Our objective was to figure out an hourly increase in consumption of energy, so for each building house we aggregate the energy by hour for all of July. This left us with around 130k rows to work with across 137 variables.

```
Rows: 137,040
Columns: 136
$ bldg_id                          <dbl> 65, 65, 65, 65, 65, 65, 65, 65, 65, 65, 65, 65, 65, 65, 65, 65, 65, 65, 65, 65, 65, …
$ in.sqft                          <dbl> 885, 885, 885, 885, 885, 885, 885, 885, 885, 885, 885, 885, 885, 885, 885, 885, 8…
$ in.bathroom_spot_vent_hour       <chr> "Hour23", "Hour23", "Hour23", "Hour23", "Hour23", "Hour23", "Hour23", "Hour2…
$ in.bedrooms                      <dbl> 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 2, 2…
$ in.building_america_climate_zone <chr> "Mixed-Humid", "Mixed-Humid", "Mixed-Humid", "Mixed-Humid", "Mixed-Humid", "…
$ in.ceiling_fan                   <chr> "Standard Efficiency", "Standard Efficiency", "Standard Efficiency", "Standa…
$ in.city                          <chr> "SC, Rock Hill", "SC, Rock Hill", "SC, Rock Hill", "SC, Rock Hill", "SC, Roc…
$ in.clothes_dryer                 <chr> "Gas, 100% Usage", "Gas, 100% Usage", "Gas, 100% Usage", "Gas, 100% Usage", …
$ in.clothes_washer                <chr> "Standard, 100% Usage", "Standard, 100% Usage", "Standard, 100% Usage", "Sta…
$ in.clothes_washer_presence       <chr> "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes",…
$ in.cooking_range                 <chr> "Electric, 100% Usage", "Electric, 100% Usage", "Electric, 100% Usage", "Ele…
$ in.cooling_setpoint              <chr> "72F", "72F", "72F", "72F", "72F", "72F", "72F", "72F", "72F", "72F", "72F",…
```

We also noticed that output provided for a variety of utilities utilised in the month of July. Since we need the total energy consumed per hour per building we aggregated the energy columns.

Firstly, we investigated the total energy consumption by different variables. We saw that the energy consumption per hour in July was taking negative values for some homes. This was because of “in.pv” variables. They represented the solar power panels that generated a surplus of energy for various building ids.

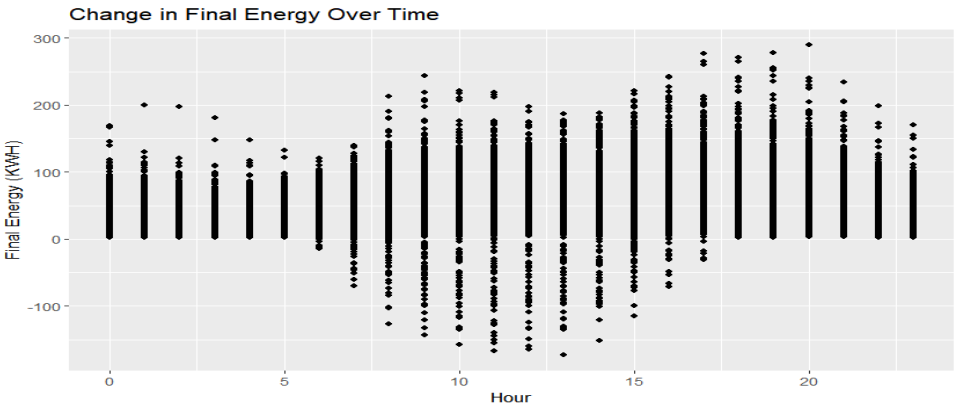


Fig. 2.6

Some of the other variables that had a strong positive correlation was the number of bedrooms and size of a building. Below are the graphs showing a fitted linear regression line to showcase the same.

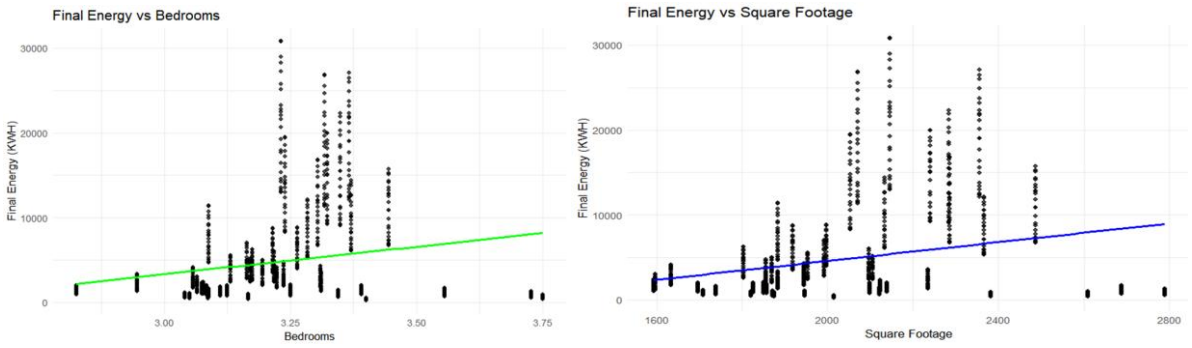


Fig. 2.7

There was also a similarity in pattern for energy consumption (by hour) for all the counties (in different magnitude).

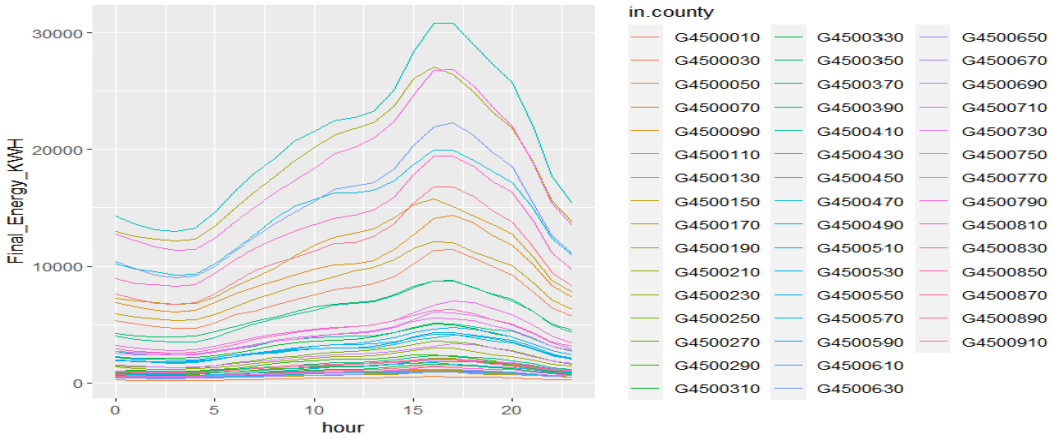


Fig. 2.8

We also explored the categorical variables to select features to put in the model. Like based of humid or dry weather how we saw the mean consumption change:

Category of Weather <fctr>	Frequency <int>	Mean_Value <dbl>
Hot-Humid	39336	41.10031
Mixed-Humid	97704	37.87466

We also tested for some of the variables that in theory would not contribute a lot to energy consumption. Like ceiling fans, but we found it to be contradictory.

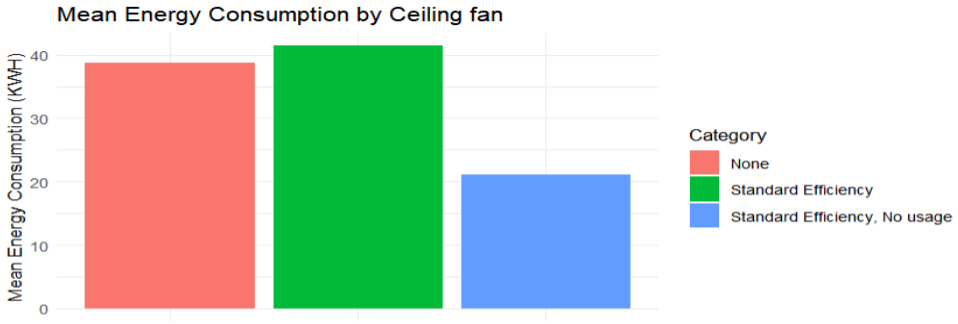


Fig. 2.9

Finally, weather data consisted of 1104 rows and 9 columns for every day of July, which were all numeric. For the sake of efficiency weather data was also averaged out for July to showcase daily averages of Temperature, Humidity, Wind and a variety of radiations.

```
tibble [1,104 × 9] (S3: tbl_df/tbl/data.frame)
 $ in.county                          : chr [1:1104] "G4500010" "G4500010" "G4500010" "G4500010" ...
 $ hour                               : num [1:1104] 0 1 2 3 4 5 6 7 8 9 ...
 $ Dry Bulb Temperature [°C]          : num [1:1104] 22.4 22.1 21.8 21.6 21.5 ...
 $ Relative Humidity [%]              : num [1:1104] 95.2 95.7 96.6 96.9 96.9 ...
 $ Wind Speed [m/s]                   : num [1:1104] 1.089 0.932 0.978 0.729 0.956 ...
 $ Wind Direction [Deg]               : num [1:1104] 125.6 104.2 127.4 86 83.5 ...
 $ Global Horizontal Radiation [W/m2] : num [1:1104] 0 0 0 0 0 ...
 $ Direct Normal Radiation [W/m2]     : num [1:1104] 0 0 0 0 0 ...
 $ Diffuse Horizontal Radiation [W/m2]: num [1:1104] 0 0 0 0 0 ...
```

This data was merged with the Energy data for each house to generate the final file utilised for modelling. The final dataset had 130k rows and over 102 variables.

```
Rows: 137,040
Columns: 102
$ in.county                                  <chr> "G4500010", "G4500010", "G4500010", "G4500010", "G4500010", "G4500010", "G4500010", "G4…
$ hour                                       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …
$ `Dry Bulb Temperature [°C]`                <dbl> 22.35581, 22.35581, 22.35581, 22.35581, 22.35581, 22.35581, 22.35581, 22.35581, 22.3558…
$ `Relative Humidity [%]`                    <dbl> 95.18613, 95.18613, 95.18613, 95.18613, 95.18613, 95.18613, 95.18613, 95.18613, 95.1861…
$ `Wind Speed [m/s]`                         <dbl> 1.0893548, 1.0893548, 1.0893548, 1.0893548, 1.0893548, 1.0893548, 1.0893548, 1.0893548,…
$ `Wind Direction [Deg]`                     <dbl> 125.5919, 125.5919, 125.5919, 125.5919, 125.5919, 125.5919, 125.5919, 125.5919, 125.591…
$ `Global Horizontal Radiation [W/m2]`       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …
$ `Direct Normal Radiation [W/m2]`           <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …
$ `Diffuse Horizontal Radiation [W/m2]`      <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …
$ bldg_id                                    <dbl> 410602, 465218, 473719, 29915, 102598, 402928, 450383, 350625, 107360, 376705, 32902, 2…
$ in.sqft                                    <dbl> 1220, 2176, 3301, 2663, 1690, 2176, 1220, 2176, 1690, 1220, 1690, 2176, 1220, 3301, 266…
$ in.bathroom_spot_vent_hour                 <chr> "Hour20", "Hour11", "Hour4", "Hour19", "Hour0", "Hour1", "Hour0", "Hour11", "Hour21", "…
$ in.bedrooms                                <dbl> 4, 4, 5, 3, 3, 4, 3, 4, 3, 2, 4, 3, 3, 5, 3, 3, 3, 3, 3, 3, 3, 4, 1, 5, 4, 3, 2, 4, …
$ in.building_america_climate_zone           <chr> "Mixed-Humid", "Mixed-Humid", "Mixed-Humid", "Mixed-Humid", "Mixed-Humid", "Mixed-Humid…
$ in.ceiling_fan                             <chr> "Standard Efficiency", "Standard Efficiency", "Standard Efficiency", "Standard Efficien…
$ in.city                                    <chr> "In another census Place", "Not in a census Place", "Not in a census Place", "Not in a …
```

Lastly, we referred to the meta data mainly while considering features for the model and being conscious of the units of variables while aggregating and interpreting the final merged data.

3. Feature Engineering

In analysing energy consumption patterns, feature engineering plays a pivotal role in preparing and enhancing the dataset for predictive modelling and analysis. It also allows relevant features from the dataset that might have a significant impact on energy consumption patterns.

Feature engineering also helps in selecting or creating the most relevant features, reducing dimensionality, and ensuring that the model does not over fit due to a feature that doesn't impact the model or incorrectly represents the energy consumption.

We looked at each of the 102 variables carefully and decided based on three major factors to decide if we wanted to keep that variable at all. First, we visualised the mean energy consumption using bar graphs for each categorical variable to discern potential variations. Subsequently, we delved into academic literature to investigate potential correlations between these variables and energy consumption. Finally, we subjected all the variables to a linear regression model to ascertain their individual statistical significance in relation to energy consumption.

For example, we saw a significant increase in average energy consumption for hot water fixtures in the house. This was unexpected as during the summer there isn't a need for the hot water fixtures to run at a 200% percent capacity let alone 100%.

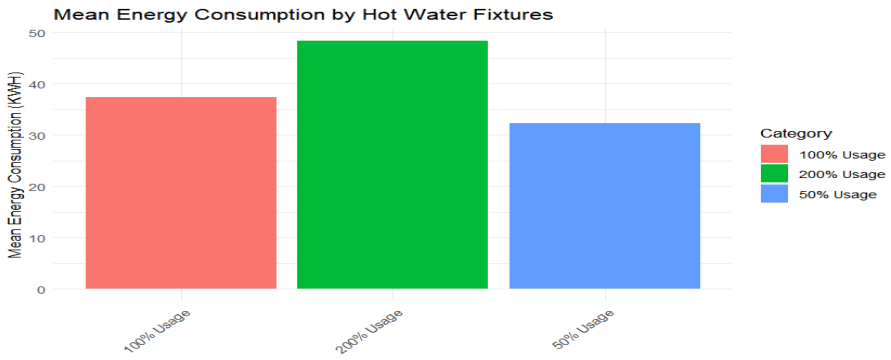


Fig. 3.0

We also saw a very low correlation for income (of about 0.04) and decided that it was not one of the major factors for energy consumption.

Now, one of the more interesting observations was for “in.infiltration” where we expected the energy consumption to increase with increase in infiltration (which is essentially a leakage in air insulation from outside to inside and vice versa).

However, our analysis revealed inverse relationships between the two variables. This finding prompted a deeper investigation, leading us to the conclusion that higher the “ACH” (air change per hour) lower the frequency at which the inside air is replaced with outside.

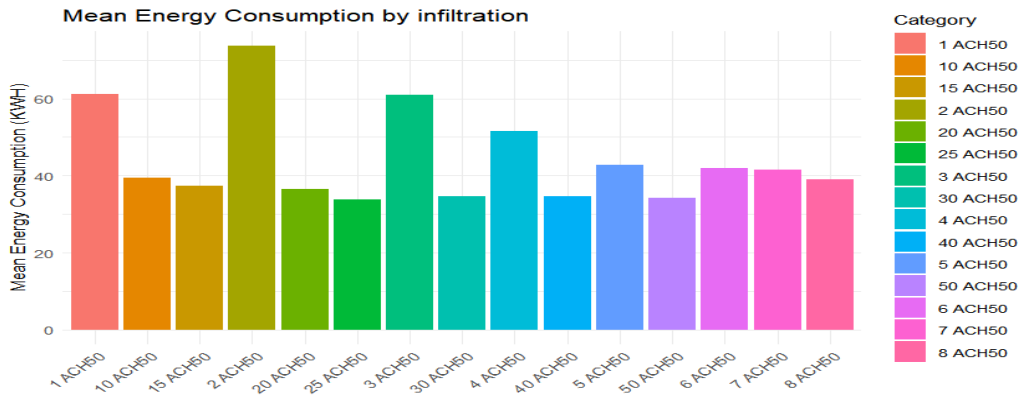


Fig. 3.1

With the same iterative process, we were able to choose around 40 variables. Finally, we explored the weather aspects of the dataset and found some strong linear relationships. (see Figure 3.2 in the Appendix)

Overall, all the weather-related variables had some sort of strong correlation with energy consumption. We also saw an inverse relation between energy and wind which was quite interesting. Upon constructing a correlation matrix, we observed a robust relationship among the variables. Consequently, we opted to retain all of them for our modelling phase.

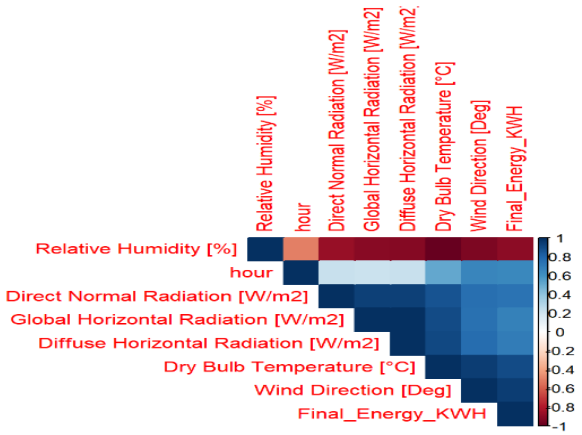


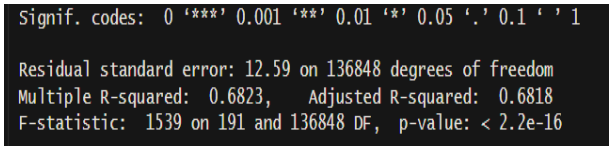
Fig. 3.3

4. Modelling

The modelling phase in our project involves testing out a variety of variables together, training and evaluation. This was an iterative process, and we mainly used linear regression, gradient boost and XGBoost to do so. Our findings of each of the models are listed below.

Linear Regression: establishes a linear relationship between the independent variables and the dependent variable. It aims to minimise the difference between the observed and predicted values by fitting a straight line to the data.

We began experimenting with a linear regression model to identify statistically significant variables influencing energy consumption. Over the course of testing up to 20 different models, we observed that adding more variables led to an increase in the multiple R-squared value but a decline in the adjusted R-squared value. This indicated that the model was being penalized for overfitting due to excessive variables. The adjusted R-squared value peaked at 69%, suggesting that the model's explanatory power plateaued. Based on this, we concluded that linear regression was not suitable for our dataset and transitioned to testing Gradient Boosting methods for improved performance.



Gradient Boosting: is an ensemble learning method that builds multiple decision trees sequentially. It minimises errors by iteratively training new models to correct the mistakes of previous models. The model produces a strong predictive model by combining multiple weak models (decision trees), focusing on the residuals (errors) of the previous models.

Gradient Boosting was an ideal choice for predicting Final Energy Consumption due to its ability to capture complex non-linear relationships in the data. It's an ensemble learning technique that combines the strengths of multiple models, making it robust against outliers and providing high predictive accuracy.

Additionally, Gradient Boosting handles both numerical and categorical variables effortlessly, reducing the need for extensive data preprocessing. Its built-in regularisation techniques prevent overfitting, and feature importance scores help identify key predictors of energy consumption. Overall, Gradient Boosting flexibility and predictive power make it a strong candidate for accurate energy consumption predictions.

```r
# Define the parameters for the Gradient Boosting model
gbm_params <- list(
  distribution = "gaussian",
  n.trees = nrounds,
  interaction.depth = 8,
  shrinkage = 0.1,
  bag.fraction = 0.5
)

# Train the Gradient Boosting model
gbm_model <- gbm(
  formula = Final_Energy_KWH ~ .,
  data = train_data,
  distribution = gbm_params$distribution,
  n.trees = gbm_params$n.trees,
  interaction.depth = gbm_params$interaction.depth,
  shrinkage = gbm_params$shrinkage,
  bag.fraction = gbm_params$bag.fraction,
  verbose = FALSE
)
```

We tested the goodness of our model by Root mean squared error and R-Squared going forward. RMSE is a measure of the average deviation of predicted values from the actual observed values. It represents the square root of the average of the squared differences between predicted and actual values. R-Squared talks about how much of the variability in the data the model accounts for.

In this case, the RMSE value of approximately 6.93 suggests that, on average, the predictions of the model are around 6.93 units away from the actual values. R-Squared had significantly improved for the model making it a better fit for the data than linear regression. This was a good result, but we wanted to see if RMSE could be lowered further so we tested out Extreme Gradient boost onto our dataset.

```r
> rmse_gbm <- sqrt(mean((predictions_gbm - test_data$Final_Energy_KWH)^2))
> print(paste("RMSE (GBM):", rmse_gbm))
[1] "RMSE (GBM): 6.93080183569941"
>
> # Compute R-squared
> SST_gbm <- sum((test_data$Final_Energy_KWH - mean(test_data$Final_Energy_KWH))^2)
> SSR_gbm <- sum((predictions_gbm - test_data$Final_Energy_KWH)^2)
> r_squared_gbm <- 1 - SSR_gbm/SST_gbm
> print(paste("R-squared (GBM):", r_squared_gbm))
[1] "R-squared (GBM): 0.902063067390577"
```

```r
> importance_summary <- summary(gbm_model)
> importance_summary
                                                                 var     rel.inf
in.sqft                                                       in.sqft 2.041027e+01
`Dry Bulb Temperature [°C]`                 `Dry Bulb Temperature [°C]` 1.213516e+01
in.vacancy_status                                     in.vacancy_status 9.165926e+00
hour                                                             hour 6.951033e+00
`Direct Normal Radiation [W/m2]`       `Direct Normal Radiation [W/m2]` 5.886754e+00
in.hot_water_fixtures                             in.hot_water_fixtures 5.767530e+00
in.county                                                     in.county 4.553274e+00
in.occupants                                               in.occupants 4.306346e+00
in.pv_system_size                                     in.pv_system_size 4.290164e+00
in.cooling_setpoint                                 in.cooling_setpoint 4.202246e+00
in.misc_pool_pump                                     in.misc_pool_pump 3.036000e+00
`Relative Humidity [%]`                     `Relative Humidity [%]` 2.747264e+00
in.ducts                                                       in.ducts 2.698718e+00
in.lighting                                                 in.lighting 1.895096e+00
in.infiltration                                         in.infiltration 1.876884e+00
in.cooling_setpoint_offset_magnitude in.cooling_setpoint_offset_magnitude 1.189164e+00
in.insulation_wall                                   in.insulation_wall 1.160132e+00
in.clothes_washer                                     in.clothes_washer 8.862936e-01
`Diffuse Horizontal Radiation [W/m2]`  `Diffuse Horizontal Radiation [W/m2]` 8.586879e-01
`Global Horizontal Radiation [W/m2]`    `Global Horizontal Radiation [W/m2]` 8.143707e-01
`Wind Speed [m/s]`                                 `Wind Speed [m/s]` 7.862158e-01
```

XGBoost (Extreme Gradient Boosting): is a powerful, scalable implementation of gradient boosting that improves accuracy and efficiency through techniques like regularization, parallel computing, and tree pruning. We selected XGBoost for its ability to address overfitting using built-in regularization, a critical feature given our model's complexity with numerous variables. It effectively handles non-linear relationships between predictors and targets, making it well-suited for our analysis.

Additionally, XGBoost is robust against outliers, such as high-energy consumption data points from air bases and airports. Its focus on prediction rather than inference aligns perfectly with the project's objectives. Moreover, it provides feature importance scores, enabling us to identify key drivers of energy consumption and optimize the model accordingly.

```
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 Radiation [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',
```

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

nrounds <- 3000  # Number of boosting
```

In this scenario, the RMSE value is approximately 6.31, the lowest among the two models evaluated. While the difference is not substantial, it could meaningfully improve the accuracy of energy consumption estimates in predictive analysis. The R-squared value, around 0.919 (91.9%), indicates that 91.9% of the variability in the dependent variable is explained by the independent variables, a slight improvement over the previous model.

This analysis highlights the key variables critical for predicting energy consumption. Factors such as building size, presence of ceiling fans, infiltration rates, solar panel size, and similar attributes emerge as significant contributors—controllable elements that notably impact consumption. Furthermore, most weather-related factors also stand out as primary drivers of energy usage, emphasizing their role in shaping predictive models.

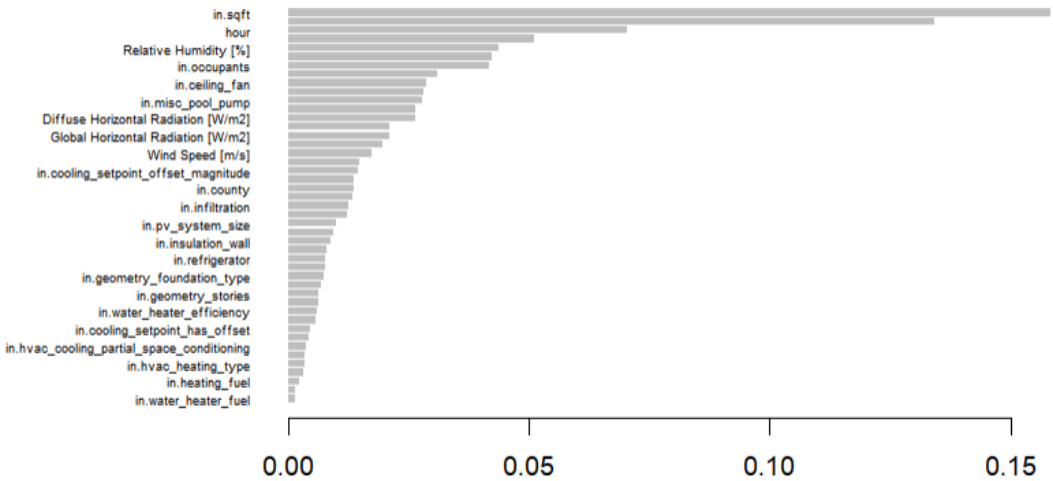


Fig. 4.1

With all the above in mind, we decided to choose XG Boost as our final model, since it worked the best with our dataset and had the lowest root mean squared error as well as the highest accuracy.

5. Energy Prediction at Warmer Temperatures

With the final model we predicted what the energy consumption will look like for all the buildings across the counties in South Carolina if there is a 5℃ increase in temperatures. Greenville has one of the highest energy consumptions followed by colleton and so on.

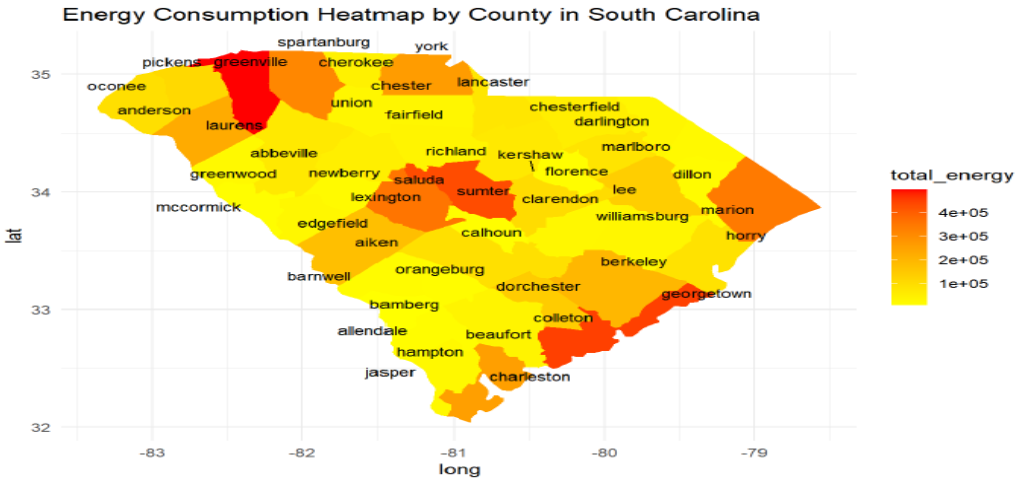


Fig. 4.2

This map is in line with the energy consumption we saw before the temperature increase which confirms no drastic change in energy consumption, just a general increase across the counties. We also looked at county wise percent increase in the energy usage. Again, Greenville has one of the largest increases in terms of magnitude. However, Horry has had the largest percentage increase of 33% across all the counties as shown below, making it the county with the largest impact with temperature change.

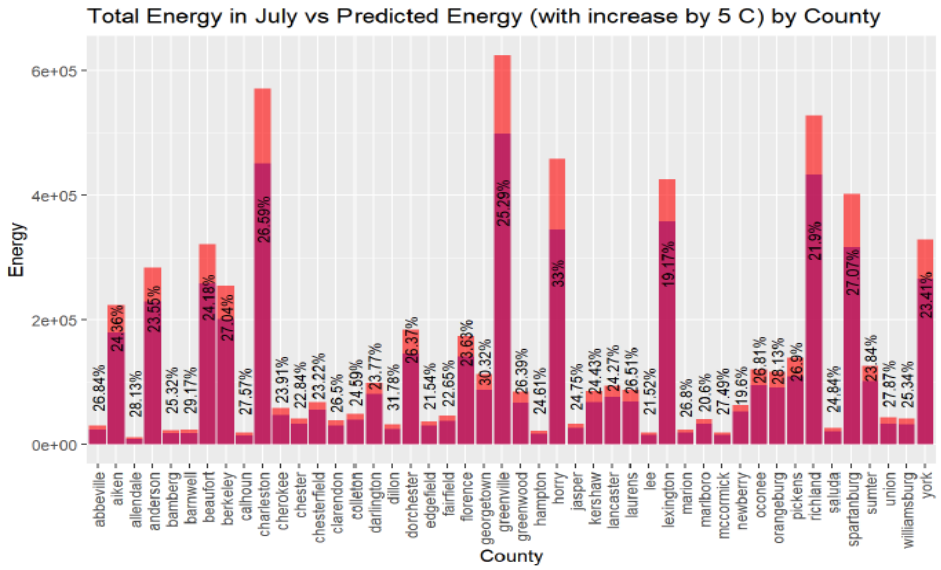


Fig. 4.3

This graph illustrates the hourly temperature increase throughout the month of July. We see the energy consumption peak at around 4pm and then it starts to come down from there.

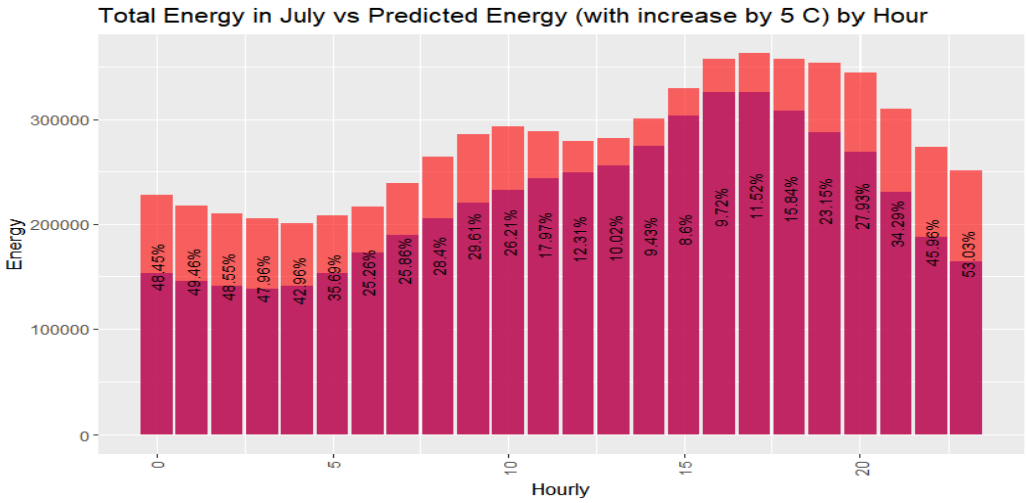


Fig. 4.4

Despite rising temperatures, the pattern of energy consumption remains consistent, with only the magnitude increasing. For instance, a 5°C increase led to a significant 30% surge in energy usage, underscoring the substantial impact of minor temperature variations on demand.

To manage this, we could implement load-shedding strategies by scheduling blackouts during peak demand periods to optimize energy distribution while minimizing disruptions. Communicating peak usage periods to consumers can encourage behavior changes that ease grid strain. Additionally, scheduling power grid maintenance during low-demand periods reduces inconvenience while ensuring infrastructure reliability.

These findings highlight the need for effective energy management strategies to address temperature-driven demand surges and support sustainable energy usage planning.

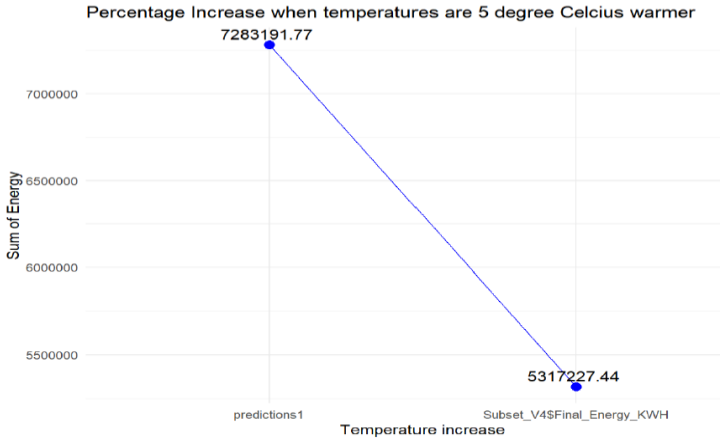


Fig. 4.5

6. Actionable Energy Efficiency Strategies

With the recent surge in energy consumption, particularly highlighted by a substantial 30% increase linked to just a 5-degree Celsius rise in temperature, the imperative for implementing actionable energy efficiency strategies becomes increasingly evident. The observed spike in energy demand not only underscores the sensitivity of energy consumption to minor temperature variations but also raises critical concerns about sustainability, resource management, and the ecological footprint of heightened energy usage.

Considering these developments, this section aims to explore and propose actionable energy efficiency strategies. These strategies, designed to curtail energy demand, mitigate environmental impact, and foster sustainability, are crucial for meeting the escalating energy needs while simultaneously striving for responsible resource utilisation.

We began by addressing some of the more apparent variables affecting energy consumption. Adjusting ceiling fan efficiency, optimizing hot water fixtures, and improving Air Changes per Hour (ACH) levels resulted in only a minimal 1% reduction in energy usage. Fig. 4.6 (left side) illustrates these changes, which, while cost-effective, present challenges in feasibility. State regulations require buildings to maintain a minimum ACH of 4, yet achieving this reduction would necessitate an unrealistically high level of 15 ACH—exceeding legal thresholds and rendering this approach impractical despite its cost benefits.

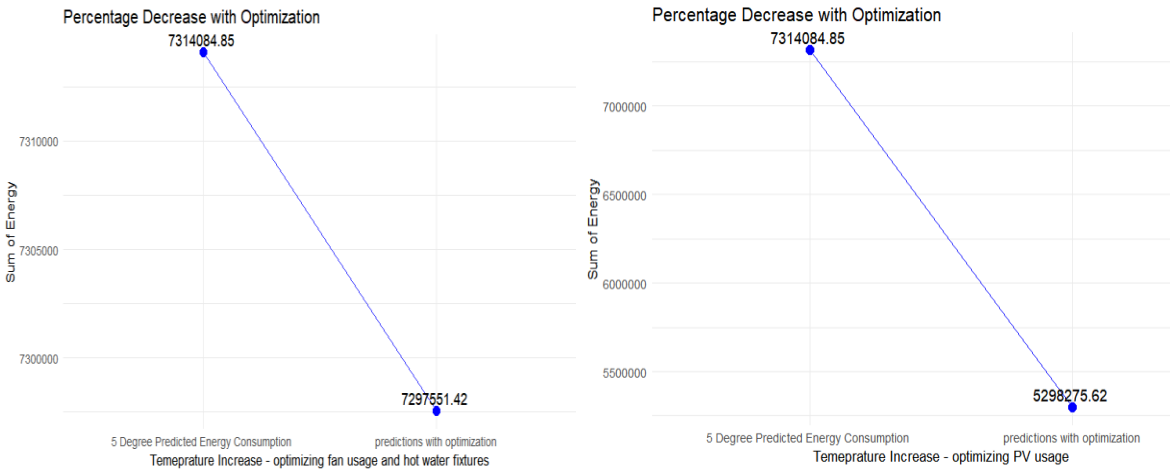


Fig. 4.6

To simplify energy reduction efforts, we shifted focus to the installation of solar panels. Rather than increasing panel sizes in buildings with existing solar systems, we recommend installing a "1KwDC" Solar Panel System in buildings currently without solar panels. This strategy balances efficiency and feasibility while achieving significant energy savings.

7. Conclusion

Based on our findings, installing '1KwDC' solar panels on buildings without existing solar panels proved to have the most significant impact, achieving a 28% reduction in energy consumption. This result highlights the cost-effectiveness of solar panel installation compared to alternatives such as upgrading house insulation or enhancing appliance efficiency.

Improving house insulation requires extensive structural modifications and material upgrades, often involving high initial investment costs. Similarly, enhancing appliance efficiency entails replacing existing devices with high-rated energy-efficient models, leading to substantial upfront expenses.

In contrast, installing '1KwDC' solar panels directly harnesses renewable solar energy, reducing reliance on traditional energy sources. This one-time investment provides long-term sustainable energy generation and minimizes dependency on grid-based electricity. Moreover, solar panel installations are scalable and modular, making them adaptable to diverse building sizes and energy requirements. While alternatives like insulation upgrades or appliance replacements remain viable, they are generally more intricate and expensive to implement.

Here’s a link to Shiny App: https://keerthikrishnaaiyappan.shinyapps.io/ShinyApp/

Appendix: Figures

Fig 2.3 Histogram of Building Stories Distribution

This histogram depicts the frequency of buildings categorized by the number of stories. The data shows a right-skewed distribution where 1-story buildings are most common (~3500), 2-story buildings are less frequent (~2000), and 3-story buildings are rare (close to 0)

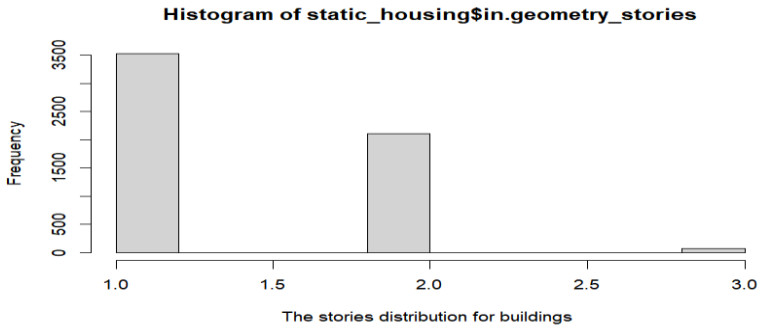


Fig. 3.2 F Scatter Plots of Weather Variables vs Energy Consumption

This figure presents four scatter plots highlighting strong linear relationships between weather variables and energy consumption in July:

- Dry Bulb Temperature [°C] vs Final Energy: Displays a positive linear correlation where higher dry bulb temperatures correspond to increased energy consumption.
- Relative Humidity [%] vs Total Energy: Shows a negative linear relationship, indicating that higher relative humidity leads to decreased energy consumption.
- Wind Direction vs Final Energy: Highlights a positive linear correlation with increased energy consumption as wind direction rises.
- Global Horizontal Radiation (Wh/m²) vs Final Energy: Demonstrates a positive linear relationship, with higher radiation levels linked to greater energy consumption.