12/10/2023

Energy Usage in the Carolinas

1. Abstract

This study investigates the impending challenge faced by eSC, which is an energy company operating in South Carolina and parts of North Carolina. The overall issue is focused on the demand for energy being used within the summer months, particularly in July. The primary concern is centered on the electrical grid due to cooling needs being higher from global warming being a stronger issue. The goal of eSC's research is to ascertain if energy use in suburbs leads to more energy waste or benefits. The overarching objective is for eSC to comprehend whether its customers are wasting more energy than they are using. This is to prepare eSC for the summer months with the energy being used.

This analysis focused the center of its study on the Insulation, AC, and Income of a house. These factors have a major importance when talking about energy usage in a home. Types of Insulation that were researched included roofing, foundation walls, ceiling, walls in general, floor, etc. When focusing on AC the variables that were looked at are the cooling setpoint of a house, whether the house had air ducts or not, the type of AC cooling, etc. There were three different Income variables, but this study focused on overall Income. Rather than just one year of income the variables were centered around 2015 and 2020. This would limit our analysis of the energy being used to just that one year. All these variables were looked into in detail. The methodology used is three separate linear models that show whether these variables are having a large effect on the cooling energy in July. Peak demand was focused on whether the peak demand changed based on the variable that was being looked at. As well as seeing if the peak demand of energy usage per household was too much or too little.

Contents

2. Introduction

Our team was tasked with understanding the data sets for an energy company, eSC, which provides electricity to residential properties in South Carolina and a small part of North Carolina for Summer 2024.Our final analysis aims to help eSC understand the key drivers of energy usage and how they could encourage their customers to save energy, ultimately reducing the demand and avoiding the need to build a new energy production facility.

2.1 Objective

The objective of this research is to predict the consumption of residential electricity in the state of North and South Carolina for Summer 2024.

2.2 Research Questions

a. Does type of insulation in houses have a major influence on the amount of energy being used?
b. How do ways of cooling in homes affect the amount of energy being used?
c. Does income have an effect on the energy being used?

2.3 Background of the Issue

The energy company eSC, responsible for supplying power to South Carolina and a part of North Carolina, is expressing concerns about potential power outages stemming from the escalating energy demand associated with the effects of global warming. Given that July is typically considered a high-demand period due to its classification as a peak summer month, there's a heightened urgency to thoroughly analyze and grasp the patterns and variations in energy consumption during this specific time frame. Understanding these consumption trends becomes pivotal in strategizing and preparing for potential strain on the power grid, thereby aiding in the prevention of power disruptions or outages.

2.3.1 Relevance of the research

The research is being conducted to help eSC understand the key drivers of energy usage and how they could encourage their customers to save energy. The goal is to reduce energy usage if next summer is ‘extra hot’ so that they can meet demand and avoid building a new energy production facility. This approach would also help the environment. The focus is on July energy usage as it is typically the highest energy usage month.

Additionally, there have not been many studies conducted in the states of the Carolinas to understand the energy usage patterns. As per Wilder and Willenborg (1975), the demand for electricity increases with an increase in the income in the state of South Carolina. While the results might be true in today’s scenario, one cannot be ignorant to accept that with the increase in demand, load management is becoming highly insignificant and alternative methods need to be adopted to cope with the increase in the demand (Corbett, Wardle & Chen, 2018).

3. Examining the Data Sets

The data which was provided consisted of four different types of data.

3.1 Static House Data

This dataset comprises fundamental details about a random selection of single-family homes serviced by eSC. It includes a comprehensive list of all houses in the dataset, with each house being described by various static attributes. These attributes range from the building ID (which is used to access the corresponding energy data) to other unchanging characteristics like the size of the house.

https://intro-datascience.s3.us-east-2.amazonaws.com/SC-data/static_house_info.parquet

3.2 Energy Usage Data

This dataset provides detailed energy consumption data for each house, collected on an hourly basis. Each house has a dedicated dataset file, containing calibrated and validated energy usage with 1-hour load profiles. This means that each file describes the energy consumption from various sources (e.g., air conditioning system, dryer) on an hourly basis for a specific house. Each file is named after the 'building ID', which serves as the unique identifier for the house. The files are stored in 'parquet' format (an optimized version of a CSV file for storage) and are all located in a single directory on Amazon AWS. For instance, the file for 'building_id' 102063 can be accessed via a specific URL. The directory contains data for approximately 5,000 houses, each represented by a unique building ID.

https://intro-datascience.s3.us-east-2.amazonaws.com/SC-data/2023-houseData/102063.parquet

3.3 Meta Data

The project includes a metadata file that serves as a guide to understanding the various fields present in the housing data files. This file is designed to be easily interpretable, providing clear explanations of the attributes found in both the static and energy usage datasets. The metadata file is essential for users to comprehend the data structure and content, ensuring accurate data analysis and interpretation. The file is accessible in CSV format and can be downloaded from the provided URL.

https://intro-datascience.s3.us-east-2.amazonaws.com/SC-data/data_dictionary.csv

3.4 Weather Data

The dataset includes hourly weather information for each county, organized by a specific county code. The corresponding county code for each house is listed in the 'in. county' column within the static house dataset. The weather data files are in a straightforward CSV format for ease of use. An example URL is provided to access the weather data for county 'G4500010'. The directory holds weather data for approximately 50 counties, facilitating comprehensive analysis when combined with house-specific energy usage data.

https://intro-datascience.s3.us-east-2.amazonaws.com/SC-data/weather/2023-weatherdata/G4500010.csv

4. Workflow

4.1 Data Cleaning

The code operates on various datasets, starting by extracting unique building IDs and filtering specific energy consumption variables for each building from fetched Parquet files. It also retrieves weather data for different counties and combines it into a single dataframe. The datasets are aligned by filtering for data from July and then merged based on a common 'time' column. The process involves renaming columns for better alignment, merging the energy and weather datasets, and subsequently cleaning the merged data by removing missing values, duplicates, and zero-value variables. Overall, this workflow consolidates diverse datasets, ensuring alignment and cleanliness, culminating in a unified dataset ready for detailed analysis of energy consumption patterns in relation to weather conditions for the selected buildings in July.

4.2 Exploratory Data Analysis

After cleaning the data, an exploratory analysis was conducted to understand the basic trends of the collected data. Exploratory analysis serves as a gateway to unravelling data's hidden tales, employing statistical and visual methods to unearth insights. Therefore, figures 1, 2, 3 and 4 show the summary of the cleaned data, where for July, the maximum temperature peaked was 38.30 degrees Celsius, while the lowest temperature was 13.89 degrees. Similarly, relative humidity reached 100%. As per the summary statistics, the highest energy consumption was on cooling of the house, which was 5.513 units.

```
     time                          Dry Bulb Temperature [°C]
Min.   :2018-07-01 00:00:00.00   Min.   :13.89
1st Qu.:2018-07-08 19:00:00.00   1st Qu.:23.42
Median :2018-07-16 15:00:00.00   Median :25.97
Mean   :2018-07-16 13:07:35.76   Mean   :26.42
3rd Qu.:2018-07-24 07:00:00.00   3rd Qu.:29.40
Max.   :2018-07-31 23:00:00.00   Max.   :38.30
Relative Humidity [%] Wind Speed [m/s] Wind Direction [Deg]
Min.   : 18.91        Min.   : 0.000   Min.   :  0.0
1st Qu.: 63.65        1st Qu.: 0.790   1st Qu.: 50.0
Median : 79.41        Median : 2.100   Median :120.0
Mean   : 76.68        Mean   : 2.164   Mean   :132.9
3rd Qu.: 90.99        3rd Qu.: 3.100   3rd Qu.:207.4
Max.   :100.00        Max.   :11.300   Max.   :360.0
Global Horizontal Radiation [W/m2] Direct Normal Radiation [W/m2]
Min.   :   0.0                     Min.   :  0.0
1st Qu.:   0.0                     1st Qu.:  0.0
Median :  78.5                     Median : 35.0
Mean   : 252.8                     Mean   :212.3
3rd Qu.: 499.5                     3rd Qu.:388.5
Max.   :1040.0                     Max.   :957.0
Diffuse Horizontal Radiation [W/m2]
Min.   :  0.0
1st Qu.:  0.0
Median : 53.0
Mean   :104.1
3rd Qu.:168.0
Max.   :488.5
out.electricity.ceiling_fan.energy_consumption
Min.   :0.008000
1st Qu.:0.008000
Median :0.008000
Mean   :0.009977
3rd Qu.:0.011000
Max.   :0.020000
out.electricity.clothes_dryer.energy_consumption
Min.   :0
1st Qu.:0
Median :0
Mean   :0
3rd Qu.:0
Max.   :0
out.electricity.clothes_washer.energy_consumption
Min.   :0
1st Qu.:0
Median :0
Mean   :0
3rd Qu.:0
```

Fig 1. A summary statistic of the cleaned variables used for the study

```
out.electricity.cooling_fans_pumps.energy_consumption
Min.   :0.006
1st Qu.:0.046
Median :0.101
Mean   :0.120
3rd Qu.:0.176
Max.   :0.400
out.electricity.cooling.energy_consumption
Min.   :0.151
1st Qu.:0.932
Median :1.791
Mean   :1.945
3rd Qu.:2.752
Max.   :5.513
out.electricity.dishwasher.energy_consumption
Min.   :0
1st Qu.:0
Median :0
Mean   :0
3rd Qu.:0
Max.   :0
out.electricity.freezer.energy_consumption
Min.   :0.03600
1st Qu.:0.03900
Median :0.04000
Mean   :0.04069
3rd Qu.:0.04400
Max.   :0.04800
out.electricity.heating_fans_pumps.energy_consumption
Min.   :0
1st Qu.:0
Median :0
Mean   :0
3rd Qu.:0
Max.   :0
out.electricity.heating_hp_bkup.energy_consumption
Min.   :0
1st Qu.:0
Median :0
Mean   :0
3rd Qu.:0
Max.   :0
out.electricity.heating.energy_consumption
Min.   :0
1st Qu.:0
Median :0
Mean   :0
```

Fig 2. A summary statistic of the cleaned variables used for the study (in continuation to Figure 1)

```
out.electricity.hot_tub_heater.energy_consumption
Min.   :0
1st Qu.:0
Median :0
Mean   :0
3rd Qu.:0
Max.   :0
out.electricity.hot_tub_pump.energy_consumption
Min.   :0
1st Qu.:0
Median :0
Mean   :0
3rd Qu.:0
Max.   :0
out.electricity.hot_water.energy_consumption
Min.   :0.00400
1st Qu.:0.00400
Median :0.00400
Mean   :0.04067
3rd Qu.:0.00400
Max.   :0.36200
out.electricity.lighting_exterior.energy_consumption
Min.   :0.00800
1st Qu.:0.00800
Median :0.01200
Mean   :0.01233
3rd Qu.:0.01600
Max.   :0.02000
out.electricity.lighting_garage.energy_consumption
Min.   :0.00000
1st Qu.:0.00000
Median :0.00400
Mean   :0.00249
3rd Qu.:0.00400
Max.   :0.00400
out.electricity.lighting_interior.energy_consumption
Min.   :0.0560
1st Qu.:0.0560
Median :0.2120
Mean   :0.2231
3rd Qu.:0.2930
Max.   :0.7790
out.electricity.mech_vent.energy_consumption
Min.   :0.000000
1st Qu.:0.000000
Median :0.000000
Mean   :0.002693
3rd Qu.:0.000000
```

Fig 3. A summary statistic of the cleaned variables used for the study (in continuation to Figure 1)

```
out.electricity.plug_loads.energy_consumption
Min.   :0.389
1st Qu.:0.424
Median :0.456
Mean   :0.461
3rd Qu.:0.482
Max.   :0.597
out.electricity.pool_heater.energy_consumption
Min.   :0
1st Qu.:0
Median :0
Mean   :0
3rd Qu.:0
Max.   :0
out.electricity.pool_pump.energy_consumption
Min.   :0.0280
1st Qu.:0.0340
Median :0.2220
Mean   :0.3976
3rd Qu.:0.7470
Max.   :1.1900
out.electricity.pv.energy_consumption
Min.   :0
1st Qu.:0
Median :0
Mean   :0
3rd Qu.:0
Max.   :0
out.electricity.range_oven.energy_consumption
Min.   :0.00000
1st Qu.:0.00000
Median :0.00000
Mean   :0.04926
3rd Qu.:0.00000
Max.   :0.86000
out.electricity.refrigerator.energy_consumption
Min.   :0.0480
1st Qu.:0.0520
Median :0.0530
Mean   :0.0542
3rd Qu.:0.0570
Max.   :0.0640
out.electricity.well_pump.energy_consumption
Min.   :0.02600
1st Qu.:0.05200
Median :0.05500
Mean   :0.06994
```

Fig 4. A summary statistic of the cleaned variables used for the study (in continuation to Figure 1)

Owing to the extensive scope of observations, the analysis concentrated solely on utilizing scatterplots and boxplots. Using both scatterplots and boxplots together would provide a more comprehensive understanding of the time series. Scatterplots help in understanding relationships and trends, while boxplots offer a concise summary and comparison of distributions and variability. Based on figure 5, it can be observed that there is a strong positive correlation between cooling energy consumption and temperature (0.64). This means that the higher the temperature the more usage of energy towards cooling the house. For the rest of the variables, there seems to be no linear relationship (figures 6,7,8).

```
Warning: the standard deviation is zero                                             Dry Bulb
Temperature [°C]
Dry Bulb Temperature [°C]                                            1.0000000
out.electricity.cooling.energy_consumption                           0.6444286
out.electricity.cooling_fans_pumps.energy_consumption                0.6182319
out.electricity.ceiling_fan.energy_consumption                       0.1713341
out.electricity.pv.energy_consumption                                       NA
                                                     out.electricity.cooling.energy_consumption
Dry Bulb Temperature [°C]                                                             0.6444286
out.electricity.cooling.energy_consumption                                            1.0000000
out.electricity.cooling_fans_pumps.energy_consumption                                 0.9969817
out.electricity.ceiling_fan.energy_consumption                                        0.1365108
out.electricity.pv.energy_consumption                                                        NA

out.electricity.cooling_fans_pumps.energy_consumption
Dry Bulb Temperature [°C]
0.6182319
out.electricity.cooling.energy_consumption
0.9969817
out.electricity.cooling_fans_pumps.energy_consumption
1.0000000
out.electricity.ceiling_fan.energy_consumption
0.1182625
out.electricity.pv.energy_consumption
NA
                                                     out.electricity.ceiling_fan.energy_consumption
Dry Bulb Temperature [°C]                                                                 0.1713341
out.electricity.cooling.energy_consumption                                                0.1365108
out.electricity.cooling_fans_pumps.energy_consumption                                     0.1182625
out.electricity.ceiling_fan.energy_consumption                                            1.0000000
out.electricity.pv.energy_consumption                                                            NA
                                                     out.electricity.pv.energy_consumption
Dry Bulb Temperature [°C]                                                               NA
out.electricity.cooling.energy_consumption                                              NA
out.electricity.cooling_fans_pumps.energy_consumption                                   NA
out.electricity.ceiling_fan.energy_consumption                                          NA
out.electricity.pv.energy_consumption                                                    1
Warning: the standard deviation is zero                                           Dry Bulb
Temperature [°C]
Dry Bulb Temperature [°C]                                         1
out.electricity.clothes_dryer.energy_consumption                 NA
out.electricity.clothes_washer.energy_consumption                NA
                                                  out.electricity.clothes_dryer.energy_consumption
Dry Bulb Temperature [°C]                                                                       NA
out.electricity.clothes_dryer.energy_consumption                                                 1
out.electricity.clothes_washer.energy_consumption                                               NA
                                                  out.electricity.clothes_washer.energy_consumption
```

Fig 5. Correlation matrix.

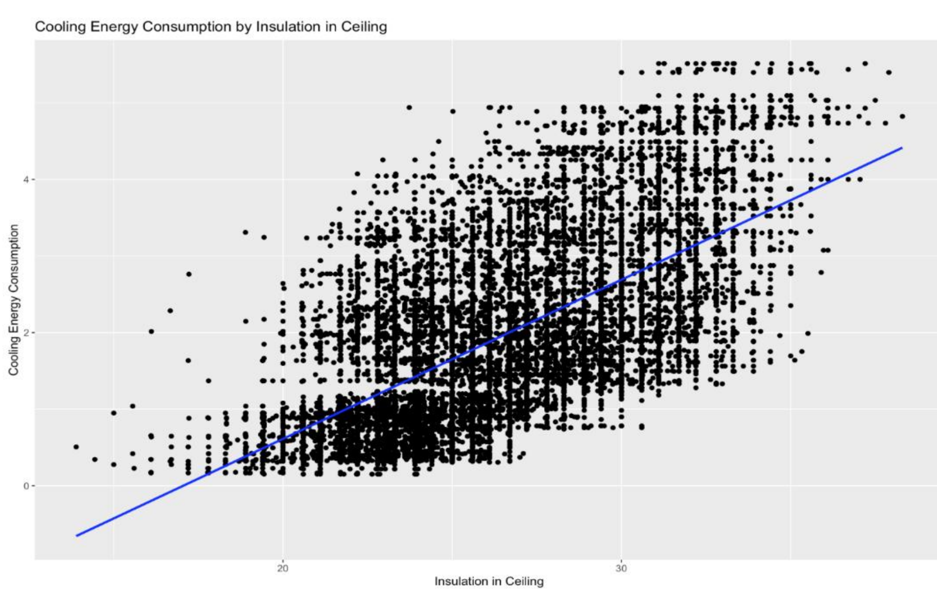

Fig 6. Scatterplot between cooling energy consumption and insulation in ceiling

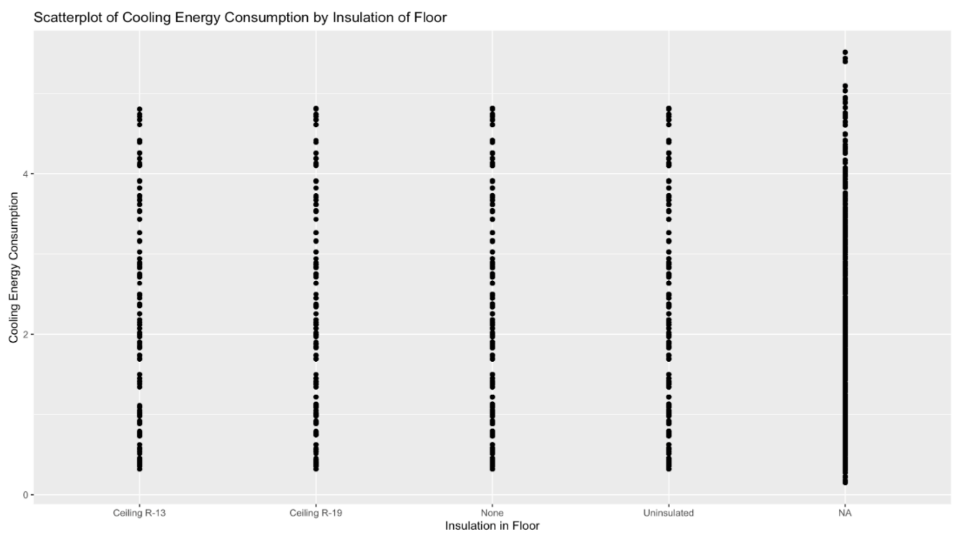

Fig 7. Scatterplot between cooling energy consumption and insulation of the floor

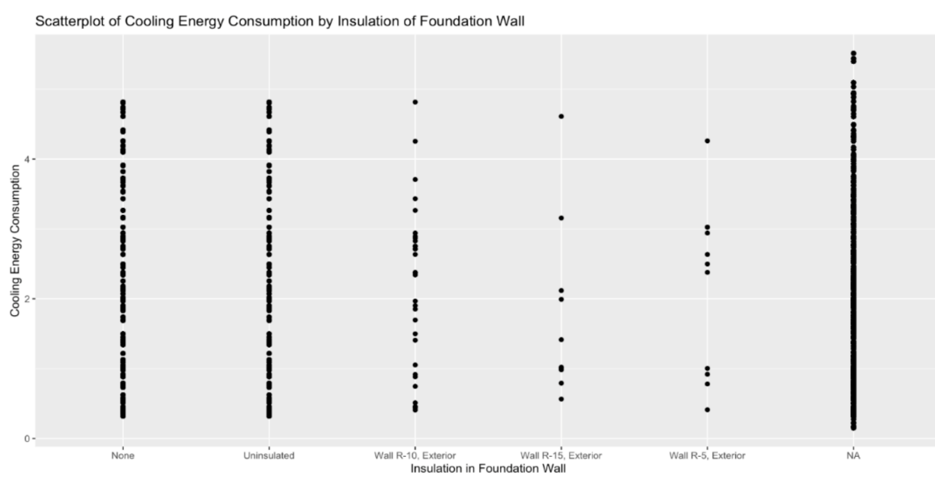

Fig 8. Scatterplot between cooling energy consumption and insulation in the foundation wall

On the other hand, to understand the outliers, boxplots were created, and it was found that only NA values had outliers (figures 9, 10, 11, 12), which were later removed before progressing with model analysis.

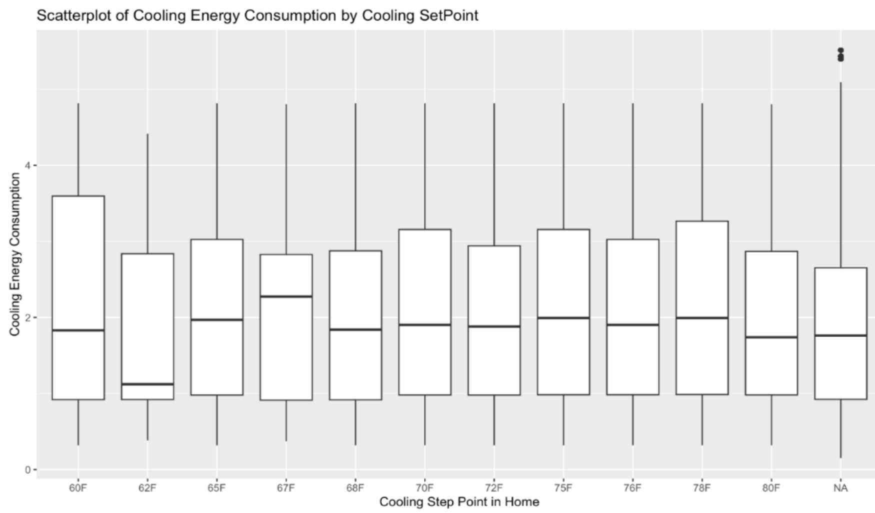

Fig 9. Boxplot between cooling energy consumption and cooling step point in the home.

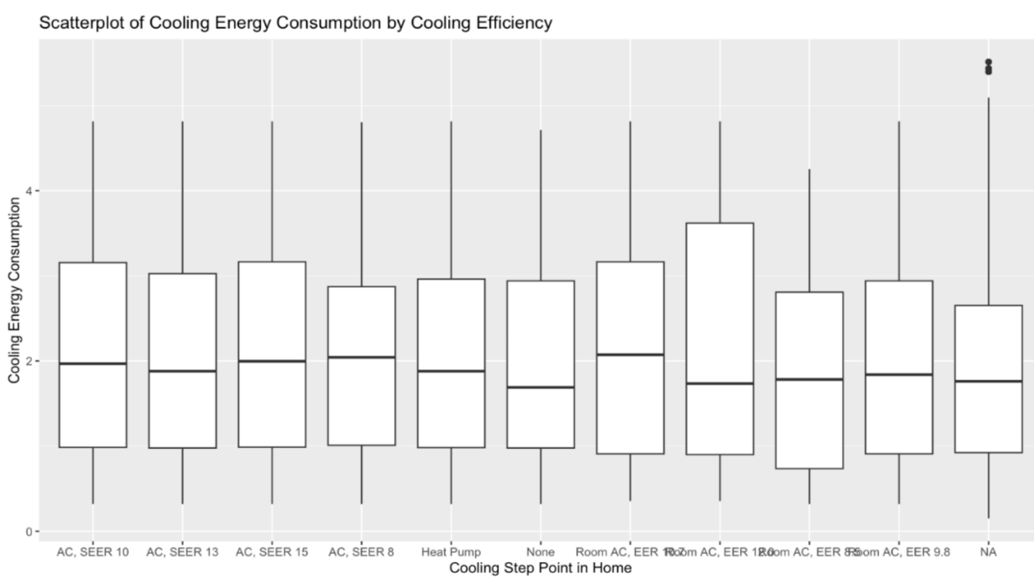

Fig 10. Boxplot between cooling energy consumption and cooling step point in home by cooling efficiency.

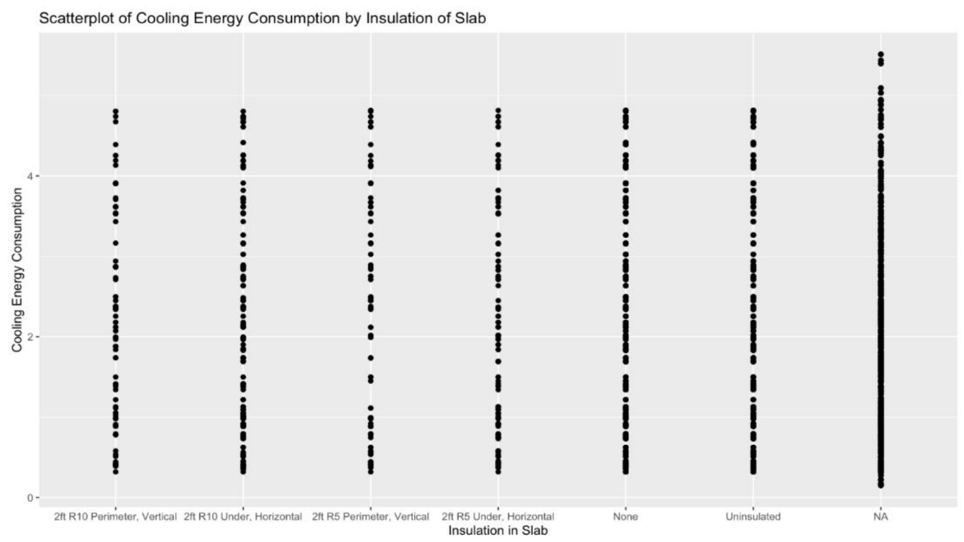

Fig 11. Boxplot between cooling energy consumption and insulation in slab

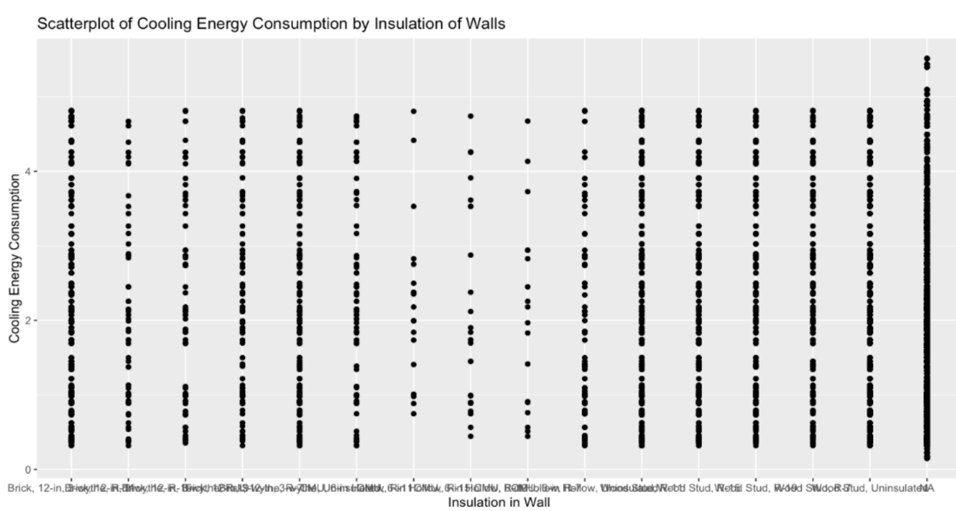

Fig 12. Boxplot between cooling energy consumption and insulation in wall

4.3 Assumptions

1. “Energy Consumption” and “electricity” are used synonymously for this project.

5. Method and Analysis

5.1 Poisson Distributions

Poisson Distributions, which is a statistical concept representing the probability of a given number of events occurring in a fixed interval, was used to begin with the analysis of the merged dataset. This involved the development of three distinct models, each tailored to address specific research questions. However, upon analysis, no discernible relationships were found among the variables within each model. Subsequently, a mixture of all three models was examined to see if there were any changes or relationships between the variables. Regrettably, this consolidated analysis yielded inconclusive findings, failing to reveal any significant relationships within the model.

```
Call:
glm(formula = out.electricity.cooling.energy_consumption ~ in.cooling_setpoint.x +
    in.hvac_has_ducts.x + in.hvac_cooling_type.x +
in.hvac_cooling_partial_space_conditioning.x +
    in.hvac_cooling_efficiency.x, family = poisson, data = merged_data)

Coefficients: (4 not defined because of singularities)
                                                        Estimate
(Intercept)                                             0.845414
in.cooling_setpoint.x62F                               -0.145796
in.cooling_setpoint.x65F                               -0.041416
in.cooling_setpoint.x67F                               -0.022859
in.cooling_setpoint.x68F                               -0.081838
in.cooling_setpoint.x70F                               -0.027500
in.cooling_setpoint.x72F                               -0.046488
in.cooling_setpoint.x75F                               -0.020361
in.cooling_setpoint.x76F                               -0.037494
in.cooling_setpoint.x78F                               -0.014730
in.cooling_setpoint.x80F                               -0.097349
in.hvac_has_ducts.xYes                                 -0.041107
in.hvac_cooling_type.xHeat Pump                        -0.037200
in.hvac_cooling_type.xNone                             -0.093987
in.hvac_cooling_type.xRoom AC                          -0.126168
in.hvac_cooling_partial_space_conditioning.x20% Conditioned  0.085388
in.hvac_cooling_partial_space_conditioning.x40% Conditioned  0.137634
in.hvac_cooling_partial_space_conditioning.x60% Conditioned -0.041327
in.hvac_cooling_partial_space_conditioning.x80% Conditioned  0.014143
in.hvac_cooling_partial_space_conditioning.xNone                   NA
in.hvac_cooling_efficiency.xAC, SEER 13                -0.030495
in.hvac_cooling_efficiency.xAC, SEER 15                 0.014793
in.hvac_cooling_efficiency.xAC, SEER 8                 -0.005739
in.hvac_cooling_efficiency.xHeat Pump                         NA
in.hvac_cooling_efficiency.xNone                              NA
in.hvac_cooling_efficiency.xRoom AC, EER 10.7           0.059806
in.hvac_cooling_efficiency.xRoom AC, EER 12.0           0.024277
in.hvac_cooling_efficiency.xRoom AC, EER 8.5           -0.142008
in.hvac_cooling_efficiency.xRoom AC, EER 9.8                  NA
                                                       Std. Error z value
(Intercept)                                              0.112363   7.524
in.cooling_setpoint.x62F                                 0.180409  -0.808
in.cooling_setpoint.x65F                                 0.107459  -0.385
in.cooling_setpoint.x67F                                 0.135006  -0.169
in.cooling_setpoint.x68F                                 0.097054  -0.843
in.cooling_setpoint.x70F                                 0.094263  -0.292
in.cooling_setpoint.x72F                                 0.094310  -0.493
in.cooling_setpoint.x75F                                 0.094042  -0.217
```

Fig 13. Results of Poisson Distribution

```
in.cooling_setpoint.x76F                                     0.097373  -0.385
in.cooling_setpoint.x78F                                     0.094723  -0.156
in.cooling_setpoint.x80F                                     0.106879  -0.911
in.hvac_has_ducts.xYes                                       0.061474  -0.669
in.hvac_cooling_type.xHeat Pump                              0.032489  -1.145
in.hvac_cooling_type.xNone                                   0.077056  -1.220
in.hvac_cooling_type.xRoom AC                                0.093582  -1.348
in.hvac_cooling_partial_space_conditioning.x20% Conditioned  0.080967   1.055
in.hvac_cooling_partial_space_conditioning.x40% Conditioned  0.087462   1.574
in.hvac_cooling_partial_space_conditioning.x60% Conditioned  0.063406  -0.652
in.hvac_cooling_partial_space_conditioning.x80% Conditioned  0.057609   0.246
in.hvac_cooling_partial_space_conditioning.xNone                   NA      NA
in.hvac_cooling_efficiency.xAC, SEER 13                      0.030078  -1.014
in.hvac_cooling_efficiency.xAC, SEER 15                      0.033947   0.436
in.hvac_cooling_efficiency.xAC, SEER 8                       0.076684  -0.075
in.hvac_cooling_efficiency.xHeat Pump                              NA      NA
in.hvac_cooling_efficiency.xNone                                   NA      NA
in.hvac_cooling_efficiency.xRoom AC, EER 10.7                0.063495   0.942
in.hvac_cooling_efficiency.xRoom AC, EER 12.0                0.088237   0.275
in.hvac_cooling_efficiency.xRoom AC, EER 8.5                 0.132260  -1.074
in.hvac_cooling_efficiency.xRoom AC, EER 9.8                       NA      NA
                                                            Pr(>|z|)
(Intercept)                                                 5.32e-14 ***
in.cooling_setpoint.x62F                                       0.419
in.cooling_setpoint.x65F                                       0.700
in.cooling_setpoint.x67F                                       0.866
in.cooling_setpoint.x68F                                       0.399
in.cooling_setpoint.x70F                                       0.770
in.cooling_setpoint.x72F                                       0.622
in.cooling_setpoint.x75F                                       0.829
in.cooling_setpoint.x76F                                       0.700
in.cooling_setpoint.x78F                                       0.876
in.cooling_setpoint.x80F                                       0.362
in.hvac_has_ducts.xYes                                         0.504
in.hvac_cooling_type.xHeat Pump                                0.252
in.hvac_cooling_type.xNone                                     0.223
in.hvac_cooling_type.xRoom AC                                  0.178
in.hvac_cooling_partial_space_conditioning.x20% Conditioned    0.292
in.hvac_cooling_partial_space_conditioning.x40% Conditioned    0.116
in.hvac_cooling_partial_space_conditioning.x60% Conditioned    0.515
in.hvac_cooling_partial_space_conditioning.x80% Conditioned    0.806
in.hvac_cooling_partial_space_conditioning.xNone                  NA
in.hvac_cooling_efficiency.xAC, SEER 13                        0.311
in.hvac_cooling_efficiency.xAC, SEER 15                        0.663
in.hvac_cooling_efficiency.xAC, SEER 8                         0.940
in.hvac_cooling_efficiency.xHeat Pump                             NA
in.hvac_cooling_efficiency.xNone                                  NA
```

Fig 14. Results of Poisson Distribution (continued figure 13)

5.2 Support Vector Machines

Furthermore, attempts were made to implement Support Vector Machine (SVM) models across three distinct datasets but encountered challenges stemming from the dataset's considerable size. The application of the SVM model was impeded as the training dataset failed to run due to the sheer volume of data, leading to limitations in effectively leveraging the model. Due to the vastness of the dataset, the model's functionality was constrained, utilizing only a fraction of the available data.

5.3 Linear Regression Model

A series of diverse linear models were executed as part of the analysis. The dependent variable across all models was the cooling energy consumption variable. Initially, the models were conducted utilizing various insulation variables related to floors, ceilings, foundation walls, and roofs. However, the outcomes revealed that solely the intercept variable exhibited statistical significance. Consequently, a mixture of all insulation variables was performed to create a comprehensive linear model, facilitating a more detailed investigation into the research question.

Similarly, an analogous approach was adopted concerning the AC model variables, encompassing factors such as the presence of air ducts, house temperature setpoints, and diverse types of AC cooling systems. Within the income model, emphasis was placed on the overall income variable, disregarding specific income variables pertaining to distinct periods, such as 2020 or 2015. Notably, the exploration revealed an absence of relationships among these disparate time-based income variables, attributed to their distinct temporal contexts. This consolidated model emerged as the cornerstone for drawing conclusions within our analysis.

```
Call:
lm(formula = out.electricity.cooling.energy_consumption ~ in.cooling_setpoint.x +
    in.hvac_has_ducts.x + in.hvac_cooling_type.x + in.hvac_cooling_partial_space_conditioning.x +
    in.hvac_cooling_efficiency.x, data = merged_data)

Residuals:
    Min      1Q  Median      3Q     Max
-2.0161 -1.1551 -0.1831  0.9659  2.8587

Coefficients: (4 not defined because of singularities)
                                                          Estimate
(Intercept)                                                2.31827
in.cooling_setpoint.x62F                                  -0.29669
in.cooling_setpoint.x65F                                  -0.08858
in.cooling_setpoint.x67F                                  -0.04924
in.cooling_setpoint.x68F                                  -0.17223
in.cooling_setpoint.x70F                                  -0.05897
in.cooling_setpoint.x72F                                  -0.09919
in.cooling_setpoint.x75F                                  -0.04405
in.cooling_setpoint.x76F                                  -0.08020
in.cooling_setpoint.x78F                                  -0.03174
in.cooling_setpoint.x80F                                  -0.20229
in.hvac_has_ducts.xYes                                    -0.08497
in.hvac_cooling_type.xHeat Pump                           -0.07878
in.hvac_cooling_type.xNone                                -0.19529
in.hvac_cooling_type.xRoom AC                             -0.25987
in.hvac_cooling_partial_space_conditioning.x20% Conditioned  0.17793
in.hvac_cooling_partial_space_conditioning.x40% Conditioned  0.29543
in.hvac_cooling_partial_space_conditioning.x60% Conditioned -0.08454
in.hvac_cooling_partial_space_conditioning.x80% Conditioned  0.02905
in.hvac_cooling_partial_space_conditioning.xNone                  NA
in.hvac_cooling_efficiency.xAC, SEER 13                   -0.06474
in.hvac_cooling_efficiency.xAC, SEER 15                    0.03208
in.hvac_cooling_efficiency.xAC, SEER 8                    -0.01242
in.hvac_cooling_efficiency.xHeat Pump                           NA
in.hvac_cooling_efficiency.xNone                                NA
in.hvac_cooling_efficiency.xRoom AC, EER 10.7              0.12634
in.hvac_cooling_efficiency.xRoom AC, EER 12.0              0.04980
in.hvac_cooling_efficiency.xRoom AC, EER 8.5              -0.27022
in.hvac_cooling_efficiency.xRoom AC, EER 9.8                    NA
                                                          Std. Error t value
(Intercept)                                                  0.21881  10.595
in.cooling_setpoint.x62F                                     0.33671  -0.881
in.cooling_setpoint.x65F                                     0.21032  -0.421
in.cooling_setpoint.x67F                                     0.26420  -0.186
in.cooling_setpoint.x68F                                     0.19035  -0.905
```

Fig 15. Results of Linear Model

```
.n.cooling_setpoint.x70F                                          0.18550  -0.318
.n.cooling_setpoint.x72F                                          0.18551  -0.535
.n.cooling_setpoint.x75F                                          0.18510  -0.238
.n.cooling_setpoint.x76F                                          0.19136  -0.419
.n.cooling_setpoint.x78F                                          0.18645  -0.170
.n.cooling_setpoint.x80F                                          0.20780  -0.973
.n.hvac_has_ducts.xYes                                            0.11648  -0.729
.n.hvac_cooling_type.xHeat Pump                                   0.06302  -1.250
.n.hvac_cooling_type.xNone                                        0.14567  -1.341
.n.hvac_cooling_type.xRoom AC                                     0.17733  -1.465
.n.hvac_cooling_partial_space_conditioning.x20% Conditioned       0.15587   1.142
.n.hvac_cooling_partial_space_conditioning.x40% Conditioned       0.17083   1.729
.n.hvac_cooling_partial_space_conditioning.x60% Conditioned       0.12045  -0.702
.n.hvac_cooling_partial_space_conditioning.x80% Conditioned       0.11152   0.261
.n.hvac_cooling_partial_space_conditioning.xNone                       NA      NA
.n.hvac_cooling_efficiency.xAC, SEER 13                           0.05848  -1.107
.n.hvac_cooling_efficiency.xAC, SEER 15                           0.06648   0.483
.n.hvac_cooling_efficiency.xAC, SEER 8                            0.14918  -0.083
.n.hvac_cooling_efficiency.xHeat Pump                                  NA      NA
.n.hvac_cooling_efficiency.xNone                                       NA      NA
.n.hvac_cooling_efficiency.xRoom AC, EER 10.7                     0.12243   1.032
.n.hvac_cooling_efficiency.xRoom AC, EER 12.0                     0.16951   0.294
.n.hvac_cooling_efficiency.xRoom AC, EER 8.5                      0.23669  -1.142
.n.hvac_cooling_efficiency.xRoom AC, EER 9.8                           NA      NA
                                                                Pr(>|t|)
(Intercept)                                                       <2e-16 ***
.n.cooling_setpoint.x62F                                          0.3783
.n.cooling_setpoint.x65F                                          0.6737
.n.cooling_setpoint.x67F                                          0.8522
.n.cooling_setpoint.x68F                                          0.3656
.n.cooling_setpoint.x70F                                          0.7506
.n.cooling_setpoint.x72F                                          0.5929
.n.cooling_setpoint.x75F                                          0.8119
.n.cooling_setpoint.x76F                                          0.6752
.n.cooling_setpoint.x78F                                          0.8648
.n.cooling_setpoint.x80F                                          0.3304
.n.hvac_has_ducts.xYes                                            0.4657
.n.hvac_cooling_type.xHeat Pump                                   0.2113
.n.hvac_cooling_type.xNone                                        0.1801
.n.hvac_cooling_type.xRoom AC                                     0.1429
.n.hvac_cooling_partial_space_conditioning.x20% Conditioned       0.2537
.n.hvac_cooling_partial_space_conditioning.x40% Conditioned       0.0838 .
.n.hvac_cooling_partial_space_conditioning.x60% Conditioned       0.4828
.n.hvac_cooling_partial_space_conditioning.x80% Conditioned       0.7945
.n.hvac_cooling_partial_space_conditioning.xNone                      NA
.n.hvac_cooling_efficiency.xAC, SEER 13                           0.2683
.n.hvac_cooling_efficiency.xAC, SEER 15                           0.6294
```

Fig 16. Results of Linear Model (Continued figure 15)

```
in.hvac_cooling_efficiency.xAC, SEER 15                    0.6294
in.hvac_cooling_efficiency.xAC, SEER 8                     0.9337
in.hvac_cooling_efficiency.xHeat Pump                          NA
in.hvac_cooling_efficiency.xNone                               NA
in.hvac_cooling_efficiency.xRoom AC, EER 10.7              0.3021
in.hvac_cooling_efficiency.xRoom AC, EER 12.0              0.7689
in.hvac_cooling_efficiency.xRoom AC, EER 8.5               0.2536
in.hvac_cooling_efficiency.xRoom AC, EER 9.8                   NA
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.33 on 5685 degrees of freedom
  (28514 observations deleted due to missingness)
Multiple R-squared:  0.004462,	Adjusted R-squared:  0.0002595 
F-statistic: 1.062 on 24 and 5685 DF,  p-value: 0.3804
```

Fig 17. Results of Linear Model (Continued figure 15)

6. Shiny Application

6.1 Website

URL - http://127.0.0.1:6872

6.2 Code

library(tidyverse)

library(arrow)

library(readr)

library(arrow)

library(dplyr)

library(shiny)

```
library(ggplot2)

library(rsconnect)


rsconnect::setAccountInfo(name='maryoelkers',

             token='6FA91355D32E9D5FBD75BFCB1FAEEE17',

             secret='kggXZbt16/Vry/Efn+le/JC8CC5rVcQMypJK9Kui')

# Define UI and Server for Shiny app

ui <- fluidPage(

 titlePanel("Energy Usage"),

 sidebarLayout(

  sidebarPanel(

   selectInput("model", "Select Model:",

         choices = c("Insulation", "Cooling", "Heating", "Income"),

         selected = "Insulation")

  ),

  mainPanel(

   tableOutput("merged_data"),

   verbatimTextOutput("model_summary")

  )

 )

)

# Define the server function

 server <- function(input, output) {

 # Read the static house information

 static_info <- arrow::read_parquet("https://intro-datascience.s3.us-east-2.amazonaws.com/SC-
data/static_house_info.parquet")

 # Extract unique building IDs from the static information

 building_ids <- unique(static_info$bldg_id)

 # Initialize an empty list to store filtered energy data for each building ID
```

```r
filtered_energy_data_list <- list()

# Filter specific datasets within energy_data (adjust conditions as needed)

selected_variables <- c("out.electricity.hot_water.energy_consumption",
"out.electricity.lighting_exterior.energy_consumption",
"out.electricity.plug_loads.energy_consumption",
"out.electricity.refrigerator.energy_consumption",
"out.fuel_pil.hot_water.energy_consumption",
"out.electricity.ceiling_fan.energy_consumption",
"out.electricity.clothes_dryer.energy_consumption",
"out.electricity.clothes_washer.energy_consumption",
"out.electricity.colling_fans_pumps.energy_consumption",
"out.electricity.freezer.energy_consumption",
"out.electricity.heating_fans_pumps.energy_consumption",
"out.electricity.heating.energy_coonsumption", "time")


# Fetch and filter specific energy data for each building ID

for (id in building_ids) {

  energy_url <- paste0("https://intro-datascience.s3.us-east-2.amazonaws.com/SC-data/2023-houseData/", id ,".parquet")

  # Read energy data for the current building ID

  energy_data <- arrow::read_parquet(energy_url)

  selected_energy_data <- energy_data %>%

    select(starts_with(selected_variables))

  # Store filtered data in the list

  filtered_energy_data_list[[id]] <- selected_energy_data

  # Process or store the filtered data as needed

  print(paste("Processed data for Building ID:", id))

  energy <- do.call(rbind, filtered_energy_data_list)

}

# This is bringing weather based on counties.

counties <- unique(static_info$in.county)

# Fetch weather data for each county

combined_weather_data <- data.frame()  # Initialize an empty dataframe
```

```
for (county in counties) {
  weather_url <- paste0("https://intro-datascience.s3.us-east-2.amazonaws.com/SC-data/weather/2023-weather-data/", county, ".csv")
  weather_data <- readr::read_csv(weather_url)
  weather_data$county <- county
  # Store weather data for each county
  combined_weather_data <- bind_rows(combined_weather_data, weather_data)
  # Process or store the weather data as needed
  print(paste("Processed weather data for", county))
}
# This is bringing weather based on counties.
counties <- unique(static_info$in.county)
for (county in counties) {
  weather_url <- paste0("https://intro-datascience.s3.us-east-2.amazonaws.com/SC-data/weather/2023-weather-data/", county, ".csv")
  weather_data <- readr::read_csv(weather_url)
  weather_data$county <- county
  # Store weather data for each county
  combined_weather_data <- bind_rows(combined_weather_data, weather_data)
  # Process or store the weather data as needed
  print(paste("Processed weather data for", county))
}
merged_data <- merge(combined_weather_data, energy, by = "time", all = TRUE)
merged_data <- merged_data %>%
  mutate(ID = row_number())
static_info <- static_info %>%
  mutate(ID = row_number())
merged_data <- merged_data %>%
  left_join(static_info %>%
```

```
        select(ID, in.insulation_ceiling),
      by = "ID")
merged_data <- merged_data %>%
  left_join(static_info %>%
        select(ID, in.insulation_floor),
      by = "ID")
merged_data <- merged_data %>%
  left_join(static_info %>%
        select(ID, in.insulation_rim_joist),
      by = "ID")
merged_data <- merged_data %>%
  left_join(static_info %>%
        select(ID, in.insulation_roof),
      by = "ID")
merged_data <- merged_data %>%
  left_join(static_info %>%
        select(ID, in.insulation_slab),
      by = "ID")
merged_data <- merged_data %>%
  left_join(static_info %>%
        select(ID, in.insulation_wall),
      by = "ID")
merged_data <- merged_data %>%
  left_join(static_info %>%
        select(ID, in.cooling_setpoint),
      by = "ID")
merged_data <- merged_data %>%
  left_join(static_info %>%
        select(ID, in.hvac_cooling_efficiency),
```

```
          by = "ID")
merged_data <- merged_data %>%
  left_join(static_info %>%
            select(ID, in.hvac_cooling_partial_space_conditioning),
          by = "ID")
merged_data <- merged_data %>%
  left_join(static_info %>%
            select(ID, in.hvac_cooling_type),
          by = "ID")
merged_data <- merged_data %>%
  left_join(static_info %>%
            select(ID, in.hvac_has_ducts),
          by = "ID")
merged_data <- merged_data %>%
  left_join(static_info %>%
            select(ID, in.income),
          by = "ID")
merged_data <- merged_data %>%
  left_join(static_info %>%
            select(ID, in.heating_fuel),
          by = "ID")
merged_data <- merged_data %>%
  left_join(static_info %>%
            select(ID, in.heating_setpoint),
          by = "ID")
merged_data <- merged_data %>%
  left_join(static_info %>%
            select(ID, in.heating_setpoint_has_offset),
          by = "ID")
```

```
merged_data <- merged_data %>%
  left_join(static_info %>%
            select(ID, in.heating_setpoint_offset_magnitude),
          by = "ID")
merged_data <- merged_data %>%
  left_join(static_info %>%
            select(ID, in.heating_setpoint_offset_period),
          by = "ID")
# Define your models as reactive expressions
model1 <- reactive({
  # Your code to create Model 1
  lm(out.electricity.cooling.energy_consumption ~ in.insulation_ceiling.x + in.insulation_floor.x + in.insulation_rim_joist.x + in.insulation_roof.x + in.insulation_slab.x + in.insulation_wall.x, data = merged_data)
})
model2 <- reactive({
  # Your code to create Model 2
  lm(out.electricity.cooling.energy_consumption ~ in.cooling_setpoint.x + in.hvac_has_ducts.x + in.hvac_cooling_type.x + in.hvac_cooling_partial_space_conditioning.x + in.hvac_cooling_efficiency.x + in.heating_setpoint_offset_period + in.heating_setpoint_offset_magnitude + in.heating_setpoint_has_offset + in.heating_setpoint + in.heating_fuel, data = merged_data)
})
model3 <- reactive({
  # Your code to create Model 3
  lm(out.electricity.heating.energy_consumption ~ in.cooling_setpoint.x + in.hvac_has_ducts.x + in.hvac_cooling_type.x + in.hvac_cooling_partial_space_conditioning.x + in.hvac_cooling_efficiency.x + in.heating_setpoint_offset_period + in.heating_setpoint_offset_magnitude + in.heating_setpoint_has_offset + in.heating_setpoint + in.heating_fuel, data = merged_data)
})
model4 <- reactive({
```

```
  # Your code to create Model 4
  lm(out.electricity.cooling.energy_consumption ~ in.income.x, data = merged_data)
 })
 output$model_summary <- renderPrint({
  selected_model <- switch(input$model,
                           "Model 1" = model1(),
                           "Model 2" = model2(),
                           "Model 3" = model3(),
                           "Model 4" = model4())
  output$model_plot <- renderPlot({
   selected_mod <- selected_model()
   plot(selected_mod) # Modify this based on what you want to visualize from the model
  })
 })
}

# Run the Shiny app
shinyApp(ui = ui, server = server)
```

7. Conclusion

Based on the exploratory data analysis and the models implemented, it can be concluded that for July, ceiling insulation has one of the major influences on the consumption of energy in the counties of South Carolina and a few counties of North Carolina. Additionally, cooling with the assistance of air conditioning has a major influence, with an accuracy of 2.1245. Whereas for the Insulation Model it was 1.892368, Income Model overall, it was 2.12864, while particular for 2015 it was 2.086296, and for 2020 it was 2.15492

Lastly, in order to predict the energy consumption for the next year, which is the summer 2024, a new variable was created with an assumption to increase the temperature by 5 degrees. Insulation caused an increase in energy and at its peak, in July, at 2.8581. AC on the other hand, caused an increase in energy and peaked, at 2.4484. Lastly, income caused an increase in energy at 2.2829 in July.

7.1 Potential Approach

One potential approach to address the challenges identified in this study is to implement targeted energy-saving initiatives and programs for residential customers. eSC can collaborate with homeowners to promote energy-efficient practices and technologies. By offering incentives and education on insulation improvement, homeowners can be encouraged to invest in better insulation materials and techniques. This approach can help reduce energy waste and enhance the overall energy efficiency of homes, particularly during peak demand periods in the summer.

Furthermore, eSC can focus on promoting and incentivizing the use of energy-efficient cooling systems. This can include providing information on energy-saving settings for air conditioning units, encouraging the use of programmable thermostats, and offering rebates or discounts on energy-efficient cooling equipment. By guiding homeowners towards more efficient cooling practices, eSC can help reduce energy consumption during high-demand months like July.

In addition to these measures, eSC can also explore the potential impact of income on energy usage. By conducting further research and analysis, the company can identify patterns and correlations between income levels and energy consumption. This information can inform targeted outreach programs, financial incentives, or energy assistance programs aimed at low-income households. By addressing the financial barriers that may contribute to excessive energy usage, eSC can promote energy conservation and reduce waste.

7.2 Impact Analysis

Implementing the potential approaches described above can have several positive impacts. Firstly, by promoting energy-efficient practices and technologies, eSC can reduce overall energy consumption and lower the strain on the electrical grid during peak demand periods. This can help avoid power disruptions or outages, ensuring a reliable energy supply for all customers.

Furthermore, encouraging energy-saving behaviors and investments can lead to cost savings for homeowners. By improving insulation and adopting energy-efficient cooling systems, households can reduce their energy bills and achieve long-term savings. This can contribute to increased customer satisfaction and loyalty to eSC, as homeowners recognize the financial benefits of energy conservation.

Additionally, addressing the impact of income on energy usage can have social and environmental benefits. By providing targeted support and assistance to low-income households, eSC can help alleviate energy burdens and improve energy affordability. This can enhance equity and inclusivity in energy access, ensuring that all customers can benefit from energy-saving initiatives.

Moreover, reducing energy waste and promoting energy efficiency aligns with broader environmental goals. By decreasing energy consumption, eSC can contribute to the reduction of greenhouse gas emissions and mitigate the effects of global warming. This supports sustainability efforts and helps create a cleaner and greener energy future for the region.

Overall, the potential approaches outlined above have the potential to positively impact energy usage, customer satisfaction, affordability, equity, and the environment. By implementing these strategies, eSC can proactively prepare for future challenges, promote responsible energy consumption, and contribute to the well-being of their customers and the communities they serve.

8. Contribution of the member

Mary Oelkers – Merging the Data, Exploratory Analysis, Analysis of Models, Results, Abstract Running of all the Code in RStudio, Shiny App, Presentation Design

Srishti RV Singh – Merging the data, various model selections, exploratory analysis interpretation and report compilation.

Adarsh Ashok Jha – Merging the data and report compilation.

Kartik Trisal – Result analysis and potential approach

9. References

Wilder, R. P., & Willenborg, J. F. (1975). Residential Demand for Electricity: A Consumer Panel Approach. Southern Economic Journal, 42(2), 212–217. https://doi.org/10.2307/1056769

Corbett, J., Wardle, K. & Chen, C. (2018). Toward a sustainable modern electricity grid: The effects of smart metering and program investments on demand-side management performance in the US electricity sector 2009-2012. IEEE Transactions on Engineering Management, 65(2), 252-263. doi: 10.1109/TEM.2017.2785315.