INTRODUCTION TO DATA SCIENCE

Team Members: Sakshi Singh & Shivani Pandeti

1. INTRODUCTION

1.1 Project Overview

Electricity is supplied to residential properties throughout South Carolina and portions of North Carolina by the South Carolina-based energy business eSC. Since July usually sees the biggest demand for cooling homes, eSC is concerned about the possible implications of global warming and is working to understand how rising temperatures could affect power usage throughout the summer months.

Making sure their electricity grid can manage peak energy demand during an especially hot summer without turning to expensive solutions like constructing new power plants is a crucial concern for the corporation. To lower demand during these peak periods, they instead seek to promote energy-saving practices among consumers. In addition to preventing blackouts, this strategy helps the environment by lowering the total amount of energy that must be produced.

1.2 Scope of the Project

To identify trends in power consumption, this project will analyze a number of datasets, such as weather data, hourly energy usage data, and static house data. The research aims to anticipate future energy demand under various situations, including temperatures that are higher than typical, by creating predictive models. The information gathered will assist eSC in pinpointing the main causes of energy consumption and investigating workable plans to successfully lower demand.

1.3 Problem Statement

eSC provides its consumers with dependable electricity while taking environmental issues seriously. However, their capacity to satisfy demand during the hottest summer months is under jeopardy due to rising global temperatures. Thousands of inhabitants' lives could be disrupted by blackouts if energy consumption surpasses the grid's capacity. eSC wishes to

avoid the costly and time-consuming solution of building a new power plant. Rather, they require a data-driven strategy to comprehend and control energy use during periods of intense heat.

1.4 Goals and Objectives

This project's main objective is to assist eSC in getting ready for possible high-demand situations by identifying important variables affecting energy consumption and suggesting workable remedies. The goals specifically include:

- Creating forecasting models to estimate July's hourly energy use.
- Examining the potential effects of higher temperatures (5 degrees warmer in July) on energy use.
- Developing a visually appealing interactive application to help eSC promote energy-saving practices among its clients by offering actionable findings.

In the end, this initiative seeks to provide eSC with the information and resources necessary to make wise choices, guaranteeing grid dependability and advancing sustainability.

2. BUSINESS QUESTIONS

Key Questions to Address

1. **What are the primary factors driving residential energy usage in July?**
   With the help of correlational matrix we identified the primary factors driving residential energy usage
2. **How can peak energy demand be accurately predicted during hotter-than-average conditions?**
   Using the prediction model (Linear Regression Model in our case).
3. **What strategies can reduce peak energy demand without compromising customer comfort?**
   Encouraging Smart Home Technology and Increasing the use of Renewable Energy

4. **How does energy demand vary across different geographic regions and household attributes?**
   Understanding these variations will help prioritize areas and customer groups for targeted interventions.

3. DATA PREPARATION

3.1 Data Sources Overview

The following are the datasets provided by eSC:

- **Static House Data**: This file holds permanent house characteristics like house size and geographic coordinates for approximately 5000 houses.
- **Energy Usage Data**: Each household has a separate file which contains hourly consumption data, and building id contained in each file.
- **Weather Data**: Weather information is available for every hour for each county and is stored in separate files as per county codes.
- **Meta Data**: A data dictionary outlining definitions of data fields across the different datasets to facilitate correct understanding of the data.

3.2 Data Merging Process

In this section, we describe the process of merging the datasets for energy, static house information, and weather data. Below are the steps taken:

1. **Loading and Filtering Static House Data:** The static house data was loaded from a Parquet file using the read_parquet() function. The dataset was filtered to include relevant columns such as building ID (bldg_id), square footage (in.sqft), and county (in.county).
2. **Reading and Filtering Energy Data:** Energy data for each building was retrieved dynamically using the building IDs. The energy data was filtered for the month of July, focusing on electricity usage (out.electricity.* columns). The dataset was also enhanced by adding the building ID (bldg_id) to each record.
3. **Merging Energy Data:** After filtering and cleaning, the energy data for all buildings was combined using the rbindlist() function, which merges the datasets row-wise.

4. **Weather Data Processing:** Weather data for various counties was obtained from multiple files. These files were read and cleaned by renaming columns for consistency, filtering for the month of July, and adding the county code to identify the corresponding location. Duplicate weather records were removed to ensure unique entries for each timestamp and county.

5. **Final Merging:**
    - The static house data was merged with the energy data using the bldg_id.
    - The merged energy dataset was then combined with the weather data based on the matching date, time_of_day, and county_code (which was renamed to in.county in the static data).
6. **Calculating Total Energy Consumption:** After the merge, a new column for total energy consumption was computed by summing up all electricity consumption columns for each record.

The final dataset, july_data, contains the necessary variables for further analysis, including static house details, energy consumption data, and weather data for the month of July.

3.3 Data Cleaning and Handling Missing Values

The data cleaning method consisted of various phases to assure the dataset's quality and completeness. Here is a summary of the steps taken:

1. **Removing Negative Total Energy Consumption:** Initially, the dataset was filtered to eliminate rows with negative total energy usage. This was accomplished by requiring only records with a non-negative value for the total_energy_consumption column. After cleaning, the number of rows was checked again.
2. **Identifying Missing Values:** To identify missing values in the dataset, we used the sapply() method to count the number of NA values in each column. This allowed us to discover columns with missing data that needed to be corrected.
3. **Interpolation of Missing Weather Data:** For weather variables (Dry, Bulb, Temperature, Relative).Humidity and wind speed—missing values were handled using interpolation. The na.approx() method from the zoo package was used for linear interpolation. This method fills in missing data by guessing values based on adjacent data points, ensuring that the weather data is consistent.

4. **Forward and Backward Filling:** In addition to interpolation, we used forward filling (na.locf()) for missing values at the beginning of the dataset. This approach uses the latest available value to fill gaps. Backward filling was used to fill gaps in the dataset caused by missing values at the end.
5. **Rechecking Missing Values:** To make sure all gaps were filled, we rechecked the dataset for any remaining missing values after using interpolation and filling procedures.
6. **Summary of Cleaned Data:** The cleaned dataset was summarized using the summary() function, which also provided basic statistics for each variable, such as the mean, maximum, and minimum values.

```
   bldg_id       out.electricity.ceiling_fan.energy_consumption out.electricity.clothes_dryer.energy_consumption
Min.   :    65   Min.   :0.000000                               Min.   : 0.00000
1st Qu.:137915   1st Qu.:0.000000                               1st Qu.: 0.00000
Median :277716   Median :0.008000                               Median : 0.00000
Mean   :276573   Mean   :0.005726                               Mean   : 0.02879
3rd Qu.:411406   3rd Qu.:0.008000                               3rd Qu.: 0.00000
Max.   :549916   Max.   :0.024000                               Max.   :19.97200
out.electricity.clothes_washer.energy_consumption out.electricity.cooling_fans_pumps.energy_consumption
Min.   :0.000000                                  Min.   :0.00000
1st Qu.:0.000000                                  1st Qu.:0.01200
Median :0.000000                                  Median :0.02100
Mean   :0.003034                                  Mean   :0.02949
3rd Qu.:0.000000                                  3rd Qu.:0.03600
Max.   :1.691000                                  Max.   :0.63600
out.electricity.cooling.energy_consumption out.electricity.dishwasher.energy_consumption out.electricity.freezer.energy_consumption
Min.   :0.0000                             Min.   :0.000000                               Min.   :0.00000
1st Qu.:0.2690                             1st Qu.:0.000000                               1st Qu.:0.00000
Median :0.4210                             Median :0.000000                               Median :0.00000
Mean   :0.5301                             Mean   :0.005128                               Mean   :0.01564
3rd Qu.:0.6600                             3rd Qu.:0.000000                               3rd Qu.:0.04000
Max.   :8.3000                             Max.   :4.871000                               Max.   :0.04800
out.electricity.heating_fans_pumps.energy_consumption out.electricity.heating_hp_bkup.energy_consumption
Min.   :0.00e+00                                      Min.   :0
1st Qu.:0.00e+00                                      1st Qu.:0
Median :0.00e+00                                      Median :0
Mean   :9.67e-06                                      Mean   :0
3rd Qu.:0.00e+00                                      3rd Qu.:0
Max.   :5.30e-02                                      Max.   :0
out.electricity.heating.energy_consumption out.electricity.hot_tub_heater.energy_consumption
Min.   :0.0000000                          Min.   :0.000000
1st Qu.:0.0000000                          1st Qu.:0.000000
Median :0.0000000                          Median :0.000000
Mean   :0.0001541                          Mean   :0.006127
3rd Qu.:0.0000000                          3rd Qu.:0.000000
Max.   :0.8730000                          Max.   :0.807000
out.electricity.hot_tub_pump.energy_consumption out.electricity.hot_water.energy_consumption
Min.   :0.000000                                Min.   :0.00000
1st Qu.:0.000000                                1st Qu.:0.00400
Median :0.000000                                Median :0.00400
Mean   :0.009109                                Mean   :0.05174
3rd Qu.:0.000000                                3rd Qu.:0.00400
Max.   :0.899000                                Max.   :3.54800
```

```
out.electricity.lighting_exterior.energy_consumption out.electricity.lighting_garage.energy_consumption
Min.   :0.000000                                     Min.   :0.000000
1st Qu.:0.004000                                     1st Qu.:0.000000
Median :0.004000                                     Median :0.000000
Mean   :0.008943                                     Mean   :0.002263
3rd Qu.:0.013000                                     3rd Qu.:0.004000
Max.   :0.080000                                     Max.   :0.016000
out.electricity.lighting_interior.energy_consumption out.electricity.mech_vent.energy_consumption
Min.   :0.0000                                       Min.   :0.000000
1st Qu.:0.0160                                       1st Qu.:0.000000
Median :0.0560                                       Median :0.000000
Mean   :0.1079                                       Mean   :0.002603
3rd Qu.:0.1280                                       3rd Qu.:0.000000
Max.   :5.0020                                       Max.   :0.057000
out.electricity.plug_loads.energy_consumption out.electricity.pool_heater.energy_consumption
Min.   :0.0000                                Min.   :0.000000
1st Qu.:0.1400                                1st Qu.:0.000000
Median :0.2480                                Median :0.000000
Mean   :0.2792                                Mean   :0.002344
3rd Qu.:0.3820                                3rd Qu.:0.000000
Max.   :1.5770                                Max.   :1.178000
out.electricity.pool_pump.energy_consumption out.electricity.pv.energy_consumption out.electricity.range_oven.energy_consumption
Min.   :0.00000                              Min.   :-7.839000                    Min.   :0.00000
1st Qu.:0.00000                              1st Qu.: 0.000000                    1st Qu.:0.00000
Median :0.00000                              Median : 0.000000                    Median :0.00000
Mean   :0.02592                              Mean   :-0.001553                    Mean   :0.03766
3rd Qu.:0.00000                              3rd Qu.: 0.000000                    3rd Qu.:0.00000
Max.   :1.57700                              Max.   : 0.000000                    Max.   :8.88000
out.electricity.refrigerator.energy_consumption out.electricity.well_pump.energy_consumption      date            time_of_day
Min.   :0.00000                                 Min.   :0.000000                                 Min.   :2018-07-01   Length:4237697
1st Qu.:0.05200                                 1st Qu.:0.000000                                 1st Qu.:2018-07-08   Class :character
Median :0.06000                                 Median :0.000000                                 Median :2018-07-16   Mode  :character
Mean   :0.08225                                 Mean   :0.004659                                 Mean   :2018-07-16
3rd Qu.:0.10700                                 3rd Qu.:0.000000                                 3rd Qu.:2018-07-24
Max.   :0.40400                                 Max.   :0.197000                                 Max.   :2018-08-01
   in.sqft      in.county         Dry.Bulb.Temperature Relative.Humidity   Wind.Speed     total_energy_consumption
Min.   : 328   Length:4237697     Min.   :13.89        Min.   : 18.91    Min.   : 0.000   Min.   : 0.000
1st Qu.:1220   Class :character   1st Qu.:23.90        1st Qu.: 65.02    1st Qu.: 1.050   1st Qu.: 0.720
Median :1690   Mode  :character   Median :26.10        Median : 80.50    Median : 2.100   Median : 1.046
Mean   :2114                      Mean   :26.47        Mean   : 77.17    Mean   : 2.284   Mean   : 1.237
3rd Qu.:2176                      3rd Qu.:29.15        3rd Qu.: 90.99    3rd Qu.: 3.350   3rd Qu.: 1.536
Max.   :8194                      Max.   :38.30        Max.   :100.00    Max.   :11.300   Max.   :23.906
```

4. EXPLORATORY DATA ANALYSIS (EDA)

4.1 Overview of Data Characteristics

The dataset includes various essential factors relating to energy consumption and meteorological conditions. Temperature, humidity, wind speed, and energy use metrics are included in the data, which is categorized by time of day. The dataset also provides detailed breakdowns of energy usage by heating, cooling, and other electricity-related activities.

Key variables:

- **Energy Usage**: Overall energy usage (kWh) and category-specific consumption, such as cooling and heating.

- **Weather Data**: Includes factors like Dry.Bulb.Temperature is relative.Wind and humidity.Speed provides insight into the environmental factors that influence energy consumption.

4.2 Visualizations

- **Total Energy Consumption Over Time**: A line plot was created to visualize the total energy consumption over time.

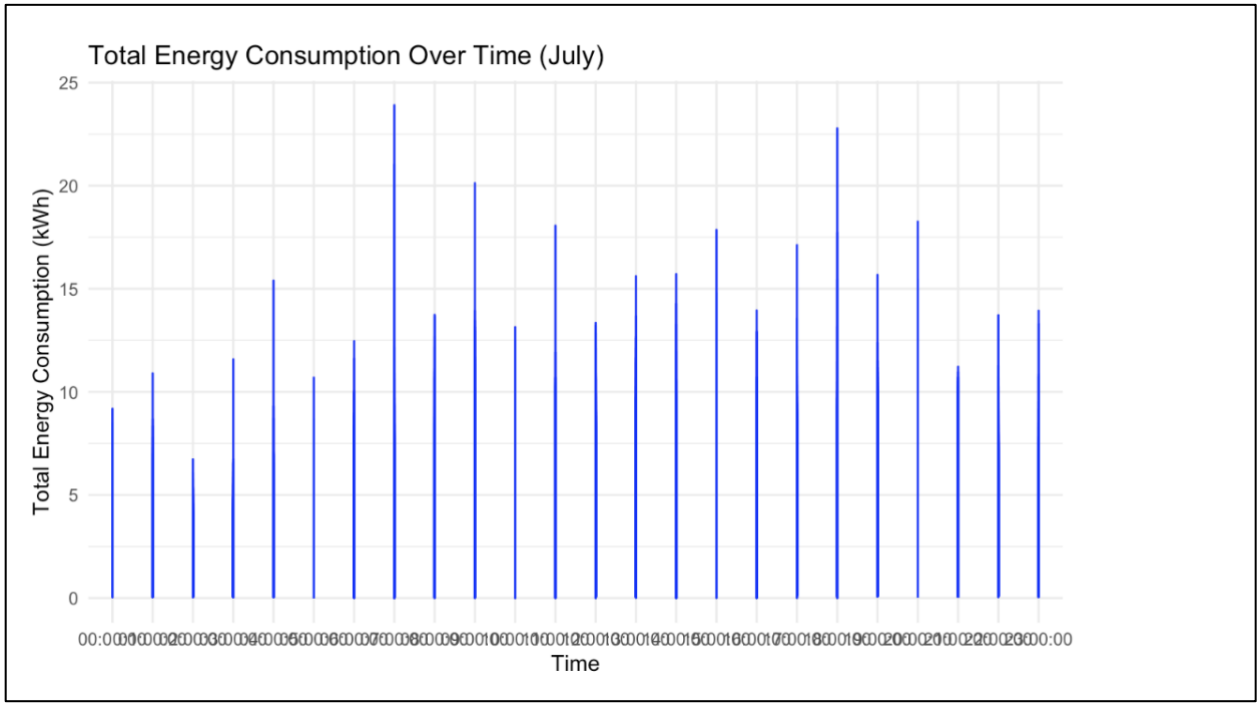


Fig 4.2.1: Total Energy Consumption Over Time (July)

- **Distribution of Total Energy Consumption:** A histogram was used to examine the distribution of total energy consumption. This allows us to understand how energy consumption is spread across different values.

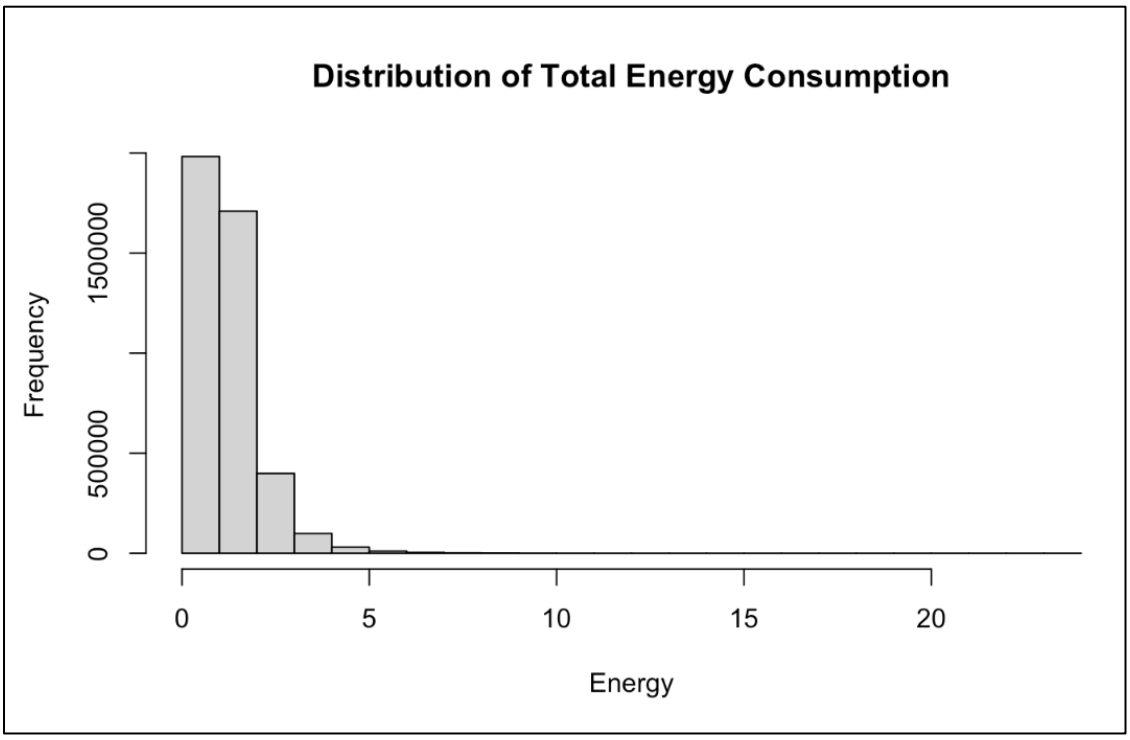


Fig 4.2.2: Distribution of Total Energy Consumption

- **Temperature Distribution**: A histogram was also plotted for the Dry.Bulb.Temperature to show the distribution of temperatures.

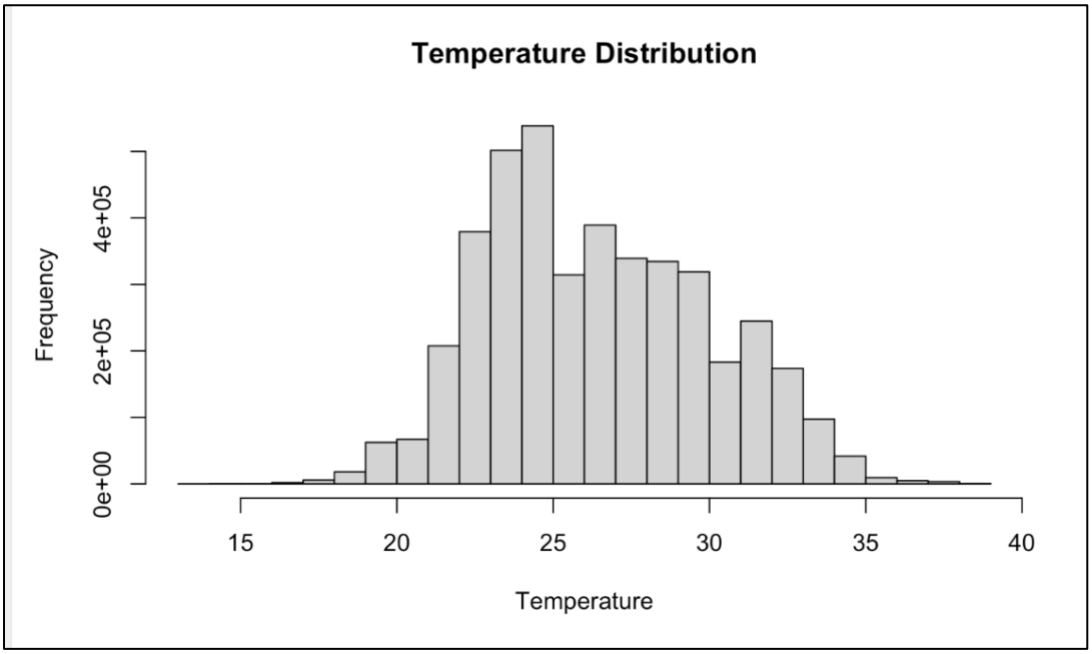


Fig 4.2.3: Temperature Distribution

- **Energy vs Temperature**: A scatter plot was created to explore the relationship between temperature and total energy consumption. This visualization helps us understand how temperature fluctuations affect energy usage.

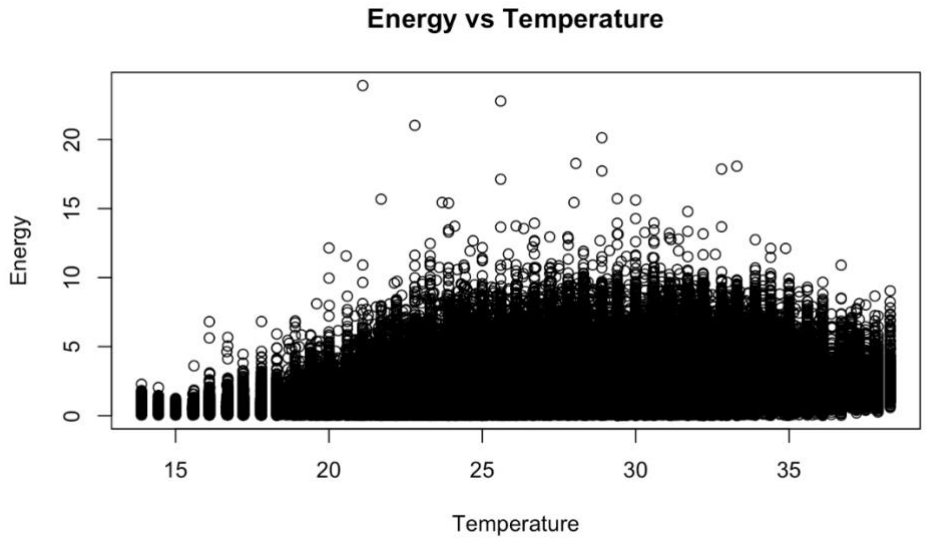


Fig 4.2.4: Energy vs Temperature

- **Energy vs Humidity**: Another scatter plot was created to visualize the relationship between humidity and energy consumption. This plot helps in understanding how humidity impacts energy usage.

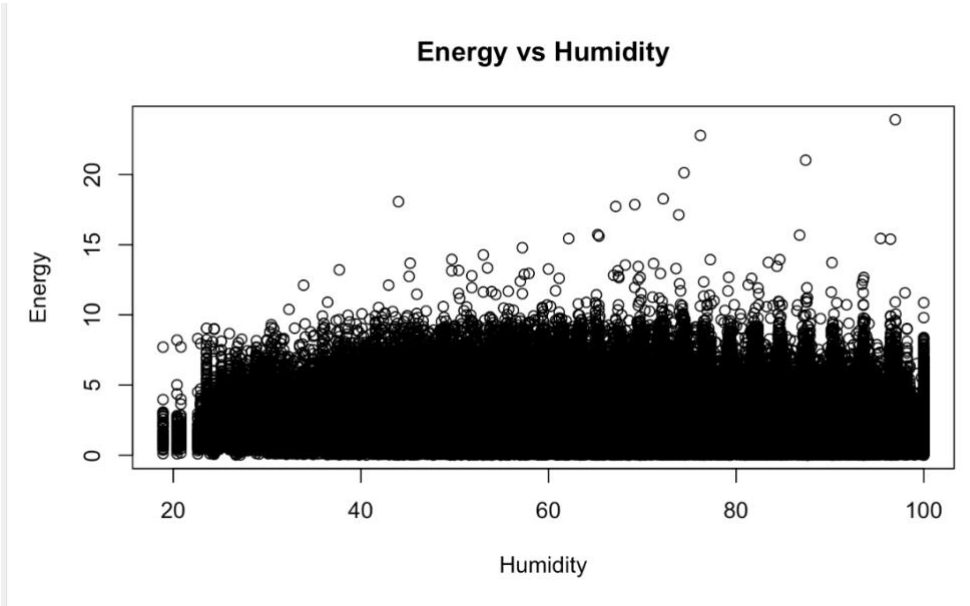


Fig 4.2.5: Energy vs Humidity

- **Average Energy Consumption by Category**: The average energy consumption was summarized by category, and a bar plot was generated to visualize the energy usage by category.

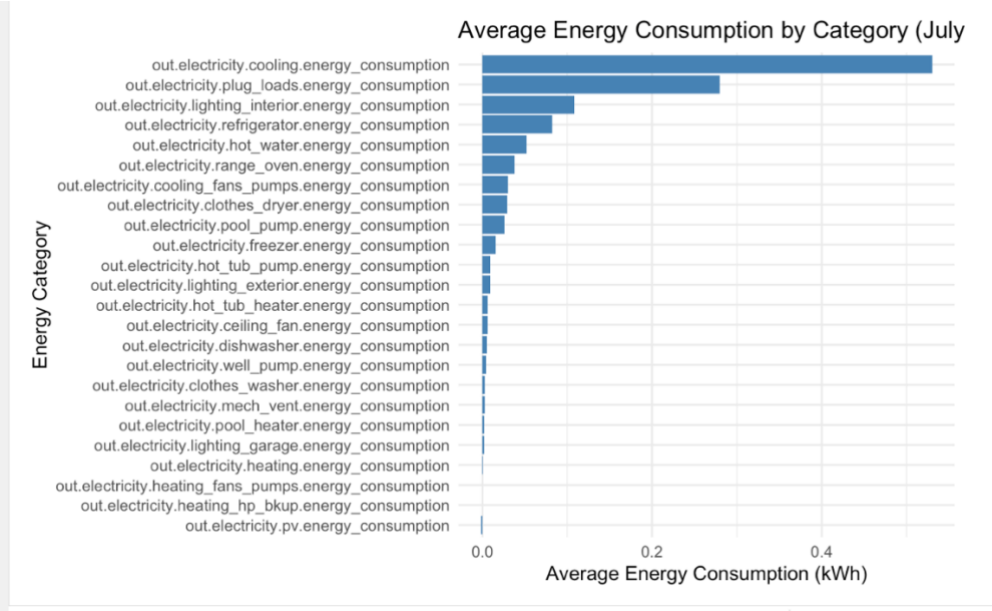


Fig 4.2.5: Average Energy Consumption

- **Cooling vs Heating Energy Usage**: A line plot comparing cooling and heating energy consumption over time was also created. This helps to observe any patterns or trends in the consumption of energy for cooling versus heating needs.

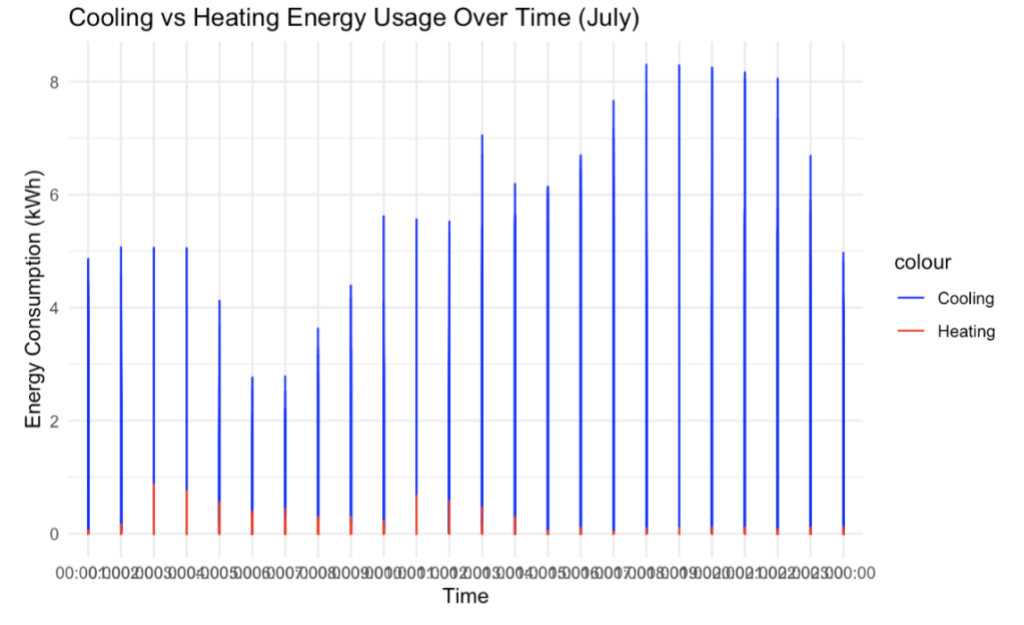


Fig 4.2.6: Cooling vs Heating Energy Usage

5. MODELING TECHNIQUES AND EVALUATION

5.1 Data Preprocessing

In the preprocessing stage, we identified and removed columns with zero variance. These columns were deemed irrelevant as they did not contribute any variability to the dataset, which could potentially skew the model's performance. The remaining numeric columns were selected for further analysis.

Next, we first performed data sampling. From the preprocessed dataset, 1000 rows were chosen at random as a sample. This allowed the study to concentrate on a reasonable subset of the data that was still representative of the overall dataset, thereby reducing the complexity of the data without sacrificing its representativeness.

After sampling, we created a new feature by adding 5 degrees to the existing temperature column. This raised temperature feature was used to simulate the effects of a hotter future climate, which is critical for evaluating energy demand under changing conditions.

5.2 Feature Selection

We explored the correlation between features and the target variable, total_energy_consumption, using a correlation matrix. This analysis helped identify variables that exhibited moderate correlations (ranging between 0.3 and 0.8) with the target variable. The selected features included various categories of energy consumption, temperature-related variables, and time-of-day features. These were chosen because they were expected to have meaningful relationships with energy consumption, which are key to improving the model's predictive accuracy.

5.3 Evaluation Metrics

Predictive model performance is evaluated using evaluation measures. They offer quantifiable indicators of a model's predictive accuracy or fit to the data. In this investigation, the following metrics were employed:

- Root Mean Squared Error (RMSE): By computing the square root of the difference between expected and observed values, the RSME determines the average size of errors. Better model performance is shown by a lower RMSE.

- R-squared (R^2): The proportion of the dependent variable's variance that can be accounted for by the independent variables. It has a range of 0 to 1, where higher performance is indicated by closer values.

5.4 Model Exploration

Two machine learning models were used to predict energy consumption during July, with a focus on the effects of temperature increase:

1. **Linear Regression:**
   A simple and interpretable model that assumes a linear relationship between the dependent variable (energy consumption) and independent variables (appliance usage, temperature, etc.). Linear regression was used as a baseline model to assess the initial performance and understand basic relationships in the data.

2. **Random Forest:**
   An ensemble model that uses multiple decision trees to make predictions. It is well-suited for datasets with complex interactions and non-linear relationships. Random Forest was expected to handle the complexity of the data better than Linear Regression, especially when accounting for interactions between variables like temperature, appliance usage, and time of day. Its ability to model non-linear relationships makes it more robust for this analysis.

Both models were trained using the same training data, which was preprocessed and sampled.

5.5 Model Comparisons and Selection

The performance of both models was evaluated and compared based on key metrics—Root Mean Squared Error (RMSE) and R-squared—using a random sample of 1000 rows from the preprocessed dataset.

Linear Regression:

- RMSE: 0.211
- R-squared: 0.910
- Evaluation: With a high R-squared value and good performance, linear regression was able to account for 91% of the variation in energy use. While it assumes linear relationships between variables, the model proved effective for this analysis and offered a straightforward, interpretable approach.

Random Forest:

- RMSE: 0.297
- R-squared: 0.823
- Evaluation: The Random Forest model explained about 82% of the variance in energy usage, which was less than the R-squared value of Linear Regression. Its RMSE was higher, indicating less precise predictions.

Model Selection:
These comparisons led to the selection of **Linear Regression** as the project's final model. It was the best fit for this analysis due to its lower RMSE and better R-squared value, as well as its ease of use and interpretability. It is crucial to remember that the random sampling procedure may cause the model's output to differ somewhat with each run.

5.6 Predictions Using the Best Model

The Linear Regression model was used to predict energy consumption during July, with a focus on the effects of a 5-degree increase in temperature. The predictions made by the model were evaluated under this future temperature scenario to assess potential energy demand increases.

6. Scenario Analysis with Adjusted Weather Dataset

6.1 Methodology to Create Warmer Weather Dataset

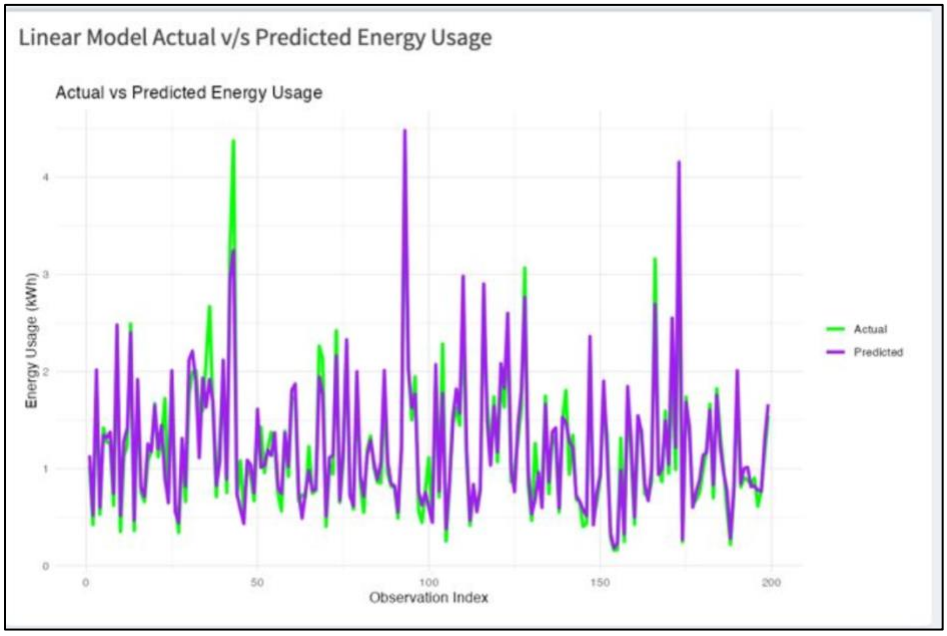
To simulate the potential impact of rising temperatures on energy consumption, we created a warmer weather dataset by adjusting the existing weather data. Specifically, we added 5 degrees to the temperature values for the month of July, reflecting a scenario where future temperatures could be higher due to global warming. This new dataset was designed to help predict how a temperature increase would affect energy demand, particularly for cooling systems such as air conditioners and fans.

- **Step-by-step Process**:
    - The original weather data, including temperature values for each day in July, was retrieved.
    - A new column was added to the dataset, which represents the adjusted temperatures by adding 5 degrees to the existing values.
    - The adjusted temperature values were integrated with the energy usage data, enabling us to analyze the impact of higher temperatures on energy consumption patterns.

This adjusted weather dataset allowed us to model future energy consumption under warmer conditions, helping to evaluate how changes in weather would influence energy demand across various regions.

Using the **Linear Regression** model, we predicted energy consumption under the adjusted weather scenario (with +5 degrees added to the temperature). The results were analyzed to understand the predicted impact on energy usage, particularly during peak hours.

- **Increased Energy Demand**:
  The predictions indicated a significant rise in energy consumption, especially in the afternoon and evening when cooling systems are typically used the most. The model showed that as temperatures increased by 5 degrees, the demand for energy, spiked, leading to higher overall energy consumption at certain times.



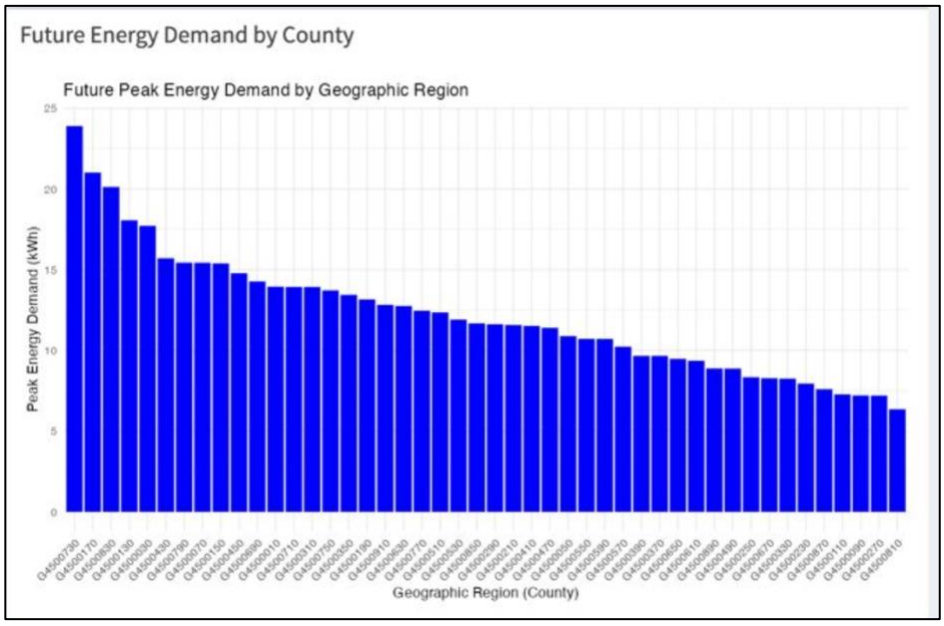
7. Peak Energy Demand Prediction

7.1 Total Demand Predictions for Different Dimensions

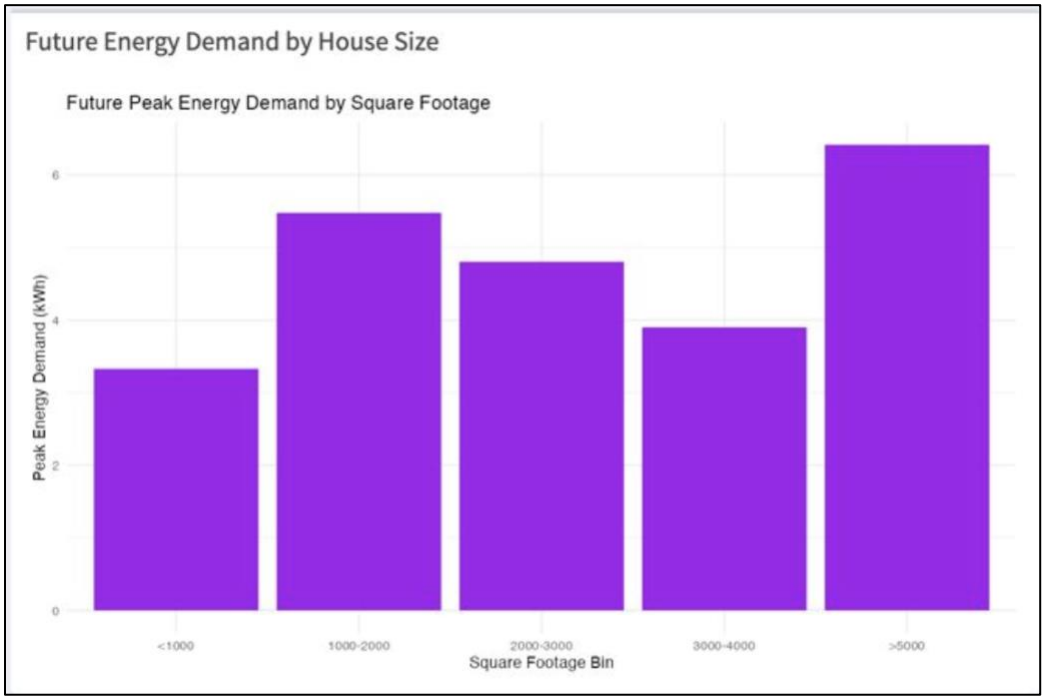
To predict the future peak energy demand, we focused on two key dimensions: geographic regions (counties) and house attributes (square footage). Both dimensions were analyzed to understand how energy consumption is distributed and how different factors contribute to peak demand. These predictions are based on the original dataset and model outputs, not incorporating the raised temperature scenario.

1. **Geographic Regions (County-wise Energy Demand)**:
   The **geographic region** was analyzed by aggregating energy demand data based on the in.county attribute. For each region, we calculated the maximum energy demand, representing the predicted peak demand for that county.

   The bar chart shows that certain counties exhibit higher peak energy demand than others. These regions, often characterized by higher population density or more frequent use of cooling systems, are expected to face greater challenges in managing peak energy consumption. Identifying these high-demand regions can help target energy-saving initiatives or infrastructure improvements.



1. **House Attributes (Energy Demand by Square Footage)**: We also examined how **square footage** (as a house attribute) influences energy demand. Larger homes typically require more energy to cool or heat, so we grouped homes by their square footage into bins and analyzed their peak energy demand.
   The bar chart shows a clear trend: larger homes tend to have higher peak energy demands. Homes with square footage greater than 4000 sq. ft. exhibited the highest peak demand, while smaller homes (less than 1000 sq. ft.) had lower energy consumption peaks. This highlights that energy efficiency measures in larger homes could play a significant role in reducing peak demand.



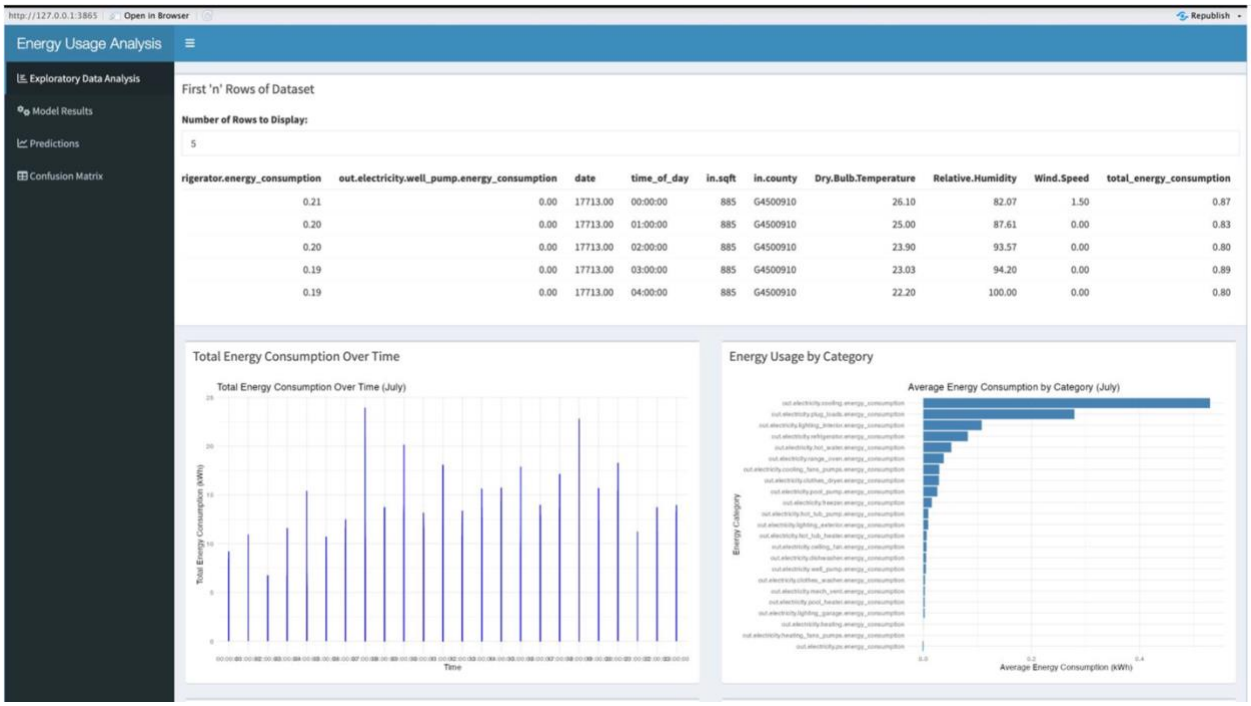
8. Interactive Application (Shiny App)

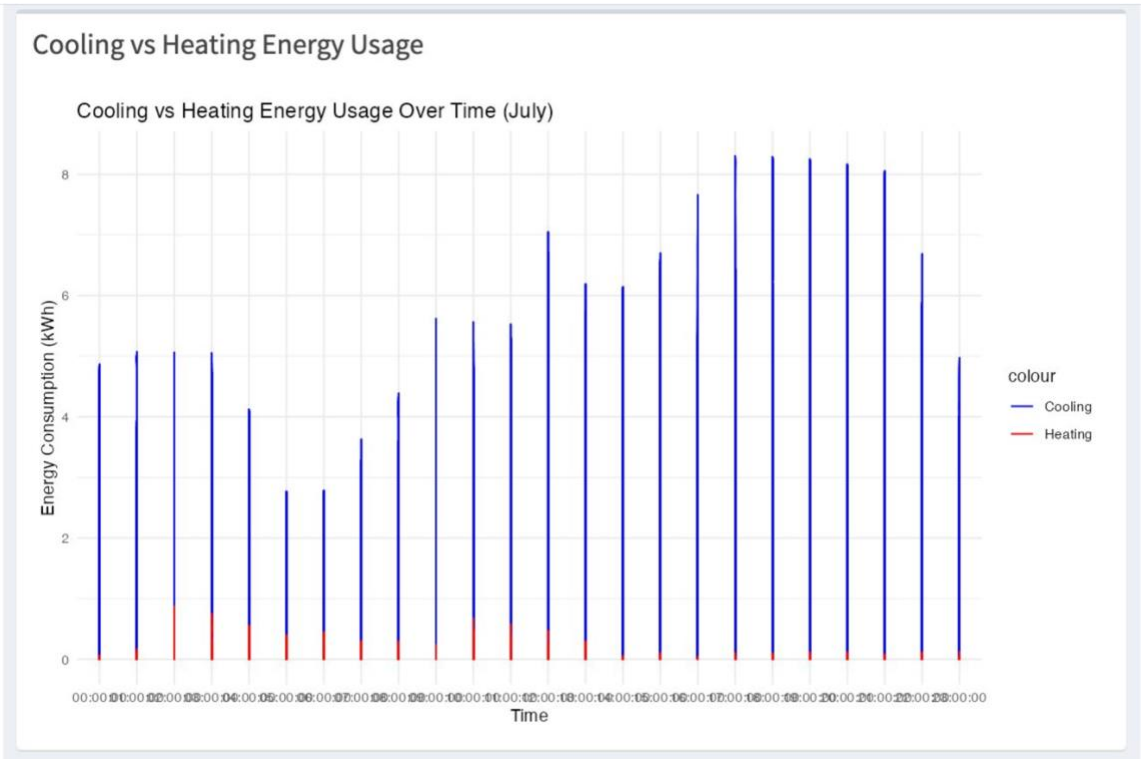
The Shiny App created for this project functions as a dynamic interface for comparing model results, showing forecasts, and analyzing data on energy consumption. The application offers a user-friendly platform for exploring important insights from the information by integrating interactive visualizations and summaries.

8.1 App Overview and Features

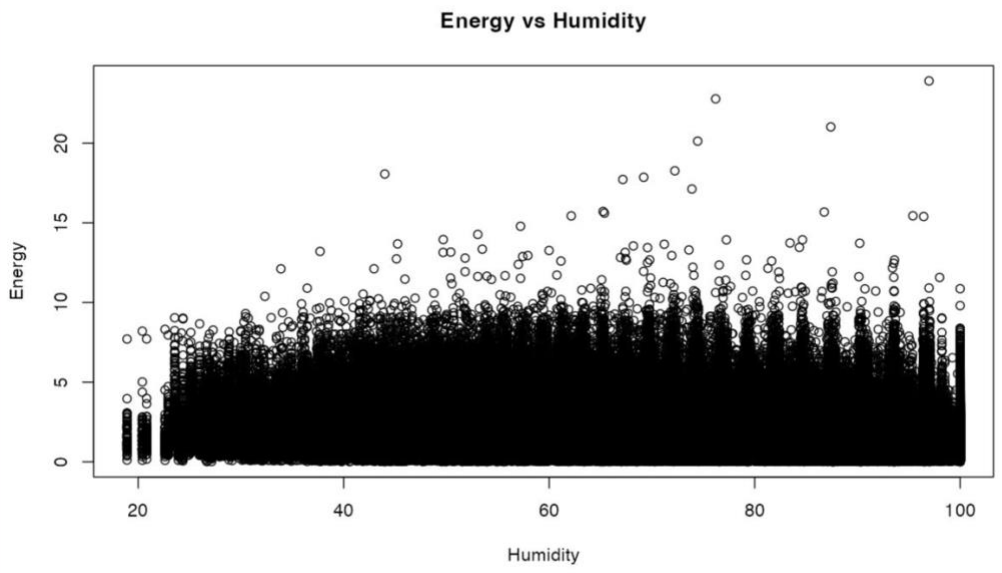
Each of the four primary tabs that make up the Shiny App focuses on a distinct facet of the analysis:

1. **Exploratory Data Analysis (EDA)**:
    - A summary of the dataset is given, along with visualizations showing trends in energy consumption and the connections between important variables like temperature and energy use.
    - By choosing how many rows to show, users can see the dataset's first few rows.
    - Interactive visualizations include:
        - Total energy consumption over time (line plot).
        - Energy usage by category (bar chart).
        - Cooling vs. heating energy usage trends.
        - Relationships between temperature/humidity and energy consumption.

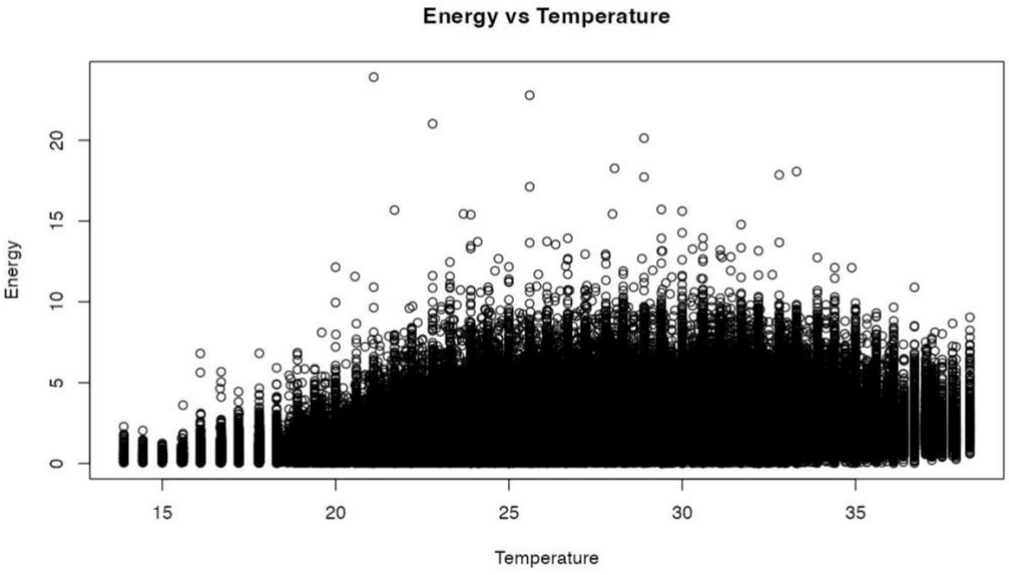

http://127.0.0.1:3865
Open in Browser
Republish
Energy Usage Analysis
Exploratory Data Analysis
Model Results
Predictions
Confusion Matrix
First 'n' Rows of Dataset
Number of Rows to Display:
5
rigerator.energy_consumption
out.electricity.well_pump.energy_consumption
date
time_of_day
in.sqft
in.county
Dry.Bulb.Temperature
Relative.Humidity
Wind.Speed
total_energy_consumption
0.21 0.00 17713.00 00:00:00 885 G4500910 26.10 82.07 1.50 0.87
0.20 0.00 17713.00 01:00:00 885 G4500910 25.00 87.61 0.00 0.83
0.20 0.00 17713.00 02:00:00 885 G4500910 23.90 93.57 0.00 0.80
0.19 0.00 17713.00 03:00:00 885 G4500910 23.03 94.20 0.00 0.89
0.19 0.00 17713.00 04:00:00 885 G4500910 22.20 100.00 0.00 0.80
Total Energy Consumption Over Time
Total Energy Consumption Over Time (July)
Total Energy Consumption (kWh)
Time
Energy Usage by Category
Average Energy Consumption by Category (July)
Energy Category
Average Energy Consumption (kWh)


Cooling vs Heating Energy Usage
Cooling vs Heating Energy Usage Over Time (July)
Energy Consumption (kWh)
0
2
4
6
8
Time
colour
Cooling
Heating

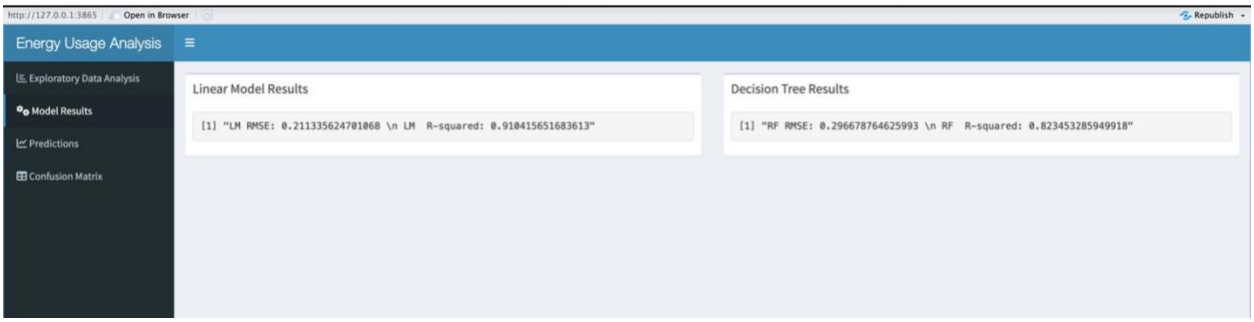
Humidity vs Energy Usage



Temperature vs Energy Usage



2. **Model Results**:
    - Displays RMSE and R-squared values for both the Linear Regression and Random Forest models.
    - Provides a summary of model performance, showing Linear Regression as the preferred model due to its better performance metrics.



3. **Predictions**:

- **Actual vs. Predicted Energy Usage (Line Graph)**:

A line graph compares the **actual energy consumption** values from the test set with the **predicted energy usage** values generated by the **Linear Regression model**. This visualization shows how well the model captures the patterns in the energy consumption data, highlighting its accuracy and deviations. The close alignment of the two lines suggests the model's effectiveness in predicting energy demand. **Dataset**: This graph is based on the test set data (20% of the sampled 1000 rows) and predictions generated using the trained Linear Regression model.

- **Geographic Regions (County-wise Energy Demand)**:

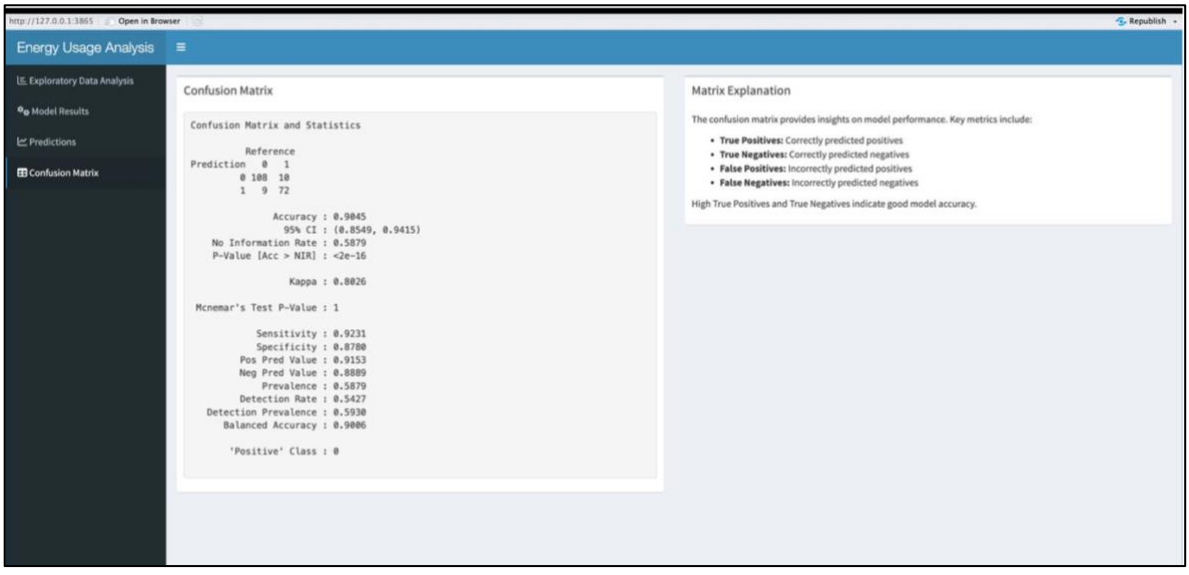
A bar chart displays the **maximum energy demand** for each county, aggregated using the in.county attribute from the **original dataset**. This visualization highlights counties with the highest predicted energy usage, enabling identification of regions that may face challenges in managing peak demand.

- **House Attributes (Energy Demand by Square Footage)**:

A bar chart groups homes into size categories (e.g., "<1000 sq. ft.", "1000-2000 sq. ft.") and displays the **maximum energy demand** for each category. Larger homes, which generally require more energy for cooling and heating, are shown to have higher peak demands.

This visualization for geographic regions and house attributes uses the original dataset, focusing on existing energy consumption patterns, and does not incorporate the raised temperature scenario.

4. **Confusion Matrix**:
    - Summarizes the classification performance of the Linear Regression model.
    - Includes a detailed confusion matrix and explanations for metrics such as accuracy, sensitivity, and specificity.



9. Deployment of the Shiny Application

The Shiny application developed for this project was deployed to enable access via a web browser, making it available for stakeholders to explore and analyze energy consumption trends, model results, and predictions interactively. The deployment process and its current state are outlined below:

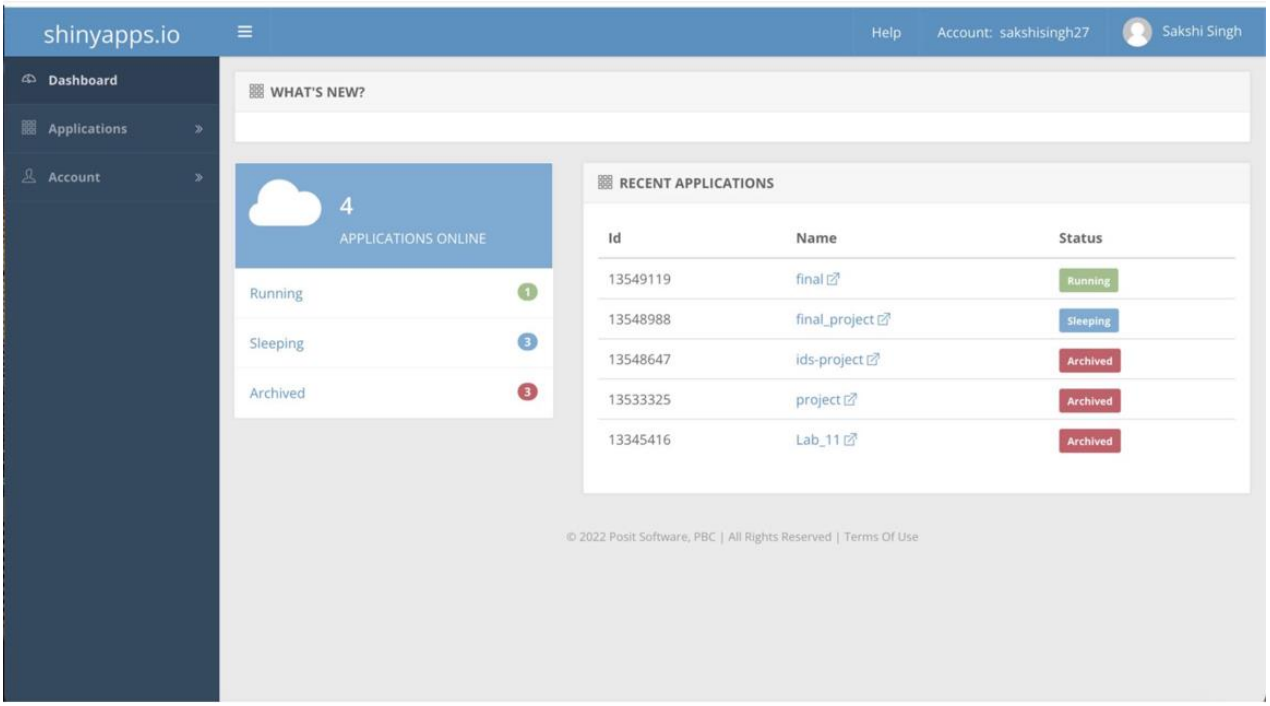
9.1 Deployment Process

The Shiny application was deployed using **RStudio's Shiny Server**, following these steps:

1. **Preparation for Deployment**:
    - All necessary libraries, datasets, and scripts were finalized and tested locally to ensure the app's functionality.
    - The application was bundled into a deployable format, including all dependencies.
2. **Deployment to Shiny Server**:
    - The application was published to ShinyApps.io via the **RStudio Console**, which confirmed the successful deployment with a status message.

```
— Preparing for deployment ————————————————————————————————————
✔ Re-deploying "final" using "server: shinyapps.io / username: sakshisingh27"
ℹ Looking up application with id "13549119"...
✔ Found application <https://sakshisingh27.shinyapps.io/final/>
ℹ Bundling 2 files: app.R and july_data.rds
ℹ Capturing R dependencies
✔ Found 113 dependencies
✔ Created 50,834,800b bundle
ℹ Uploading bundle...
✔ Uploaded bundle with id 9481520
— Deploying to server ——————————————————————————————————————————
Waiting for task: 1486728923
  building: Processing bundle: 9481520
  building: Building image: 11642438
  building: Installing system dependencies
  building: Fetching packages
  building: Installing packages
  building: Installing files
  building: Pushing image: 11642438
  deploying: Starting instances
  unstaging: Stopping old instances
— Deployment complete ————————————————————————————————————————
✔ Successfully deployed to <https://sakshisingh27.shinyapps.io/final/>
Deployment completed: https://sakshisingh27.shinyapps.io/final/
```

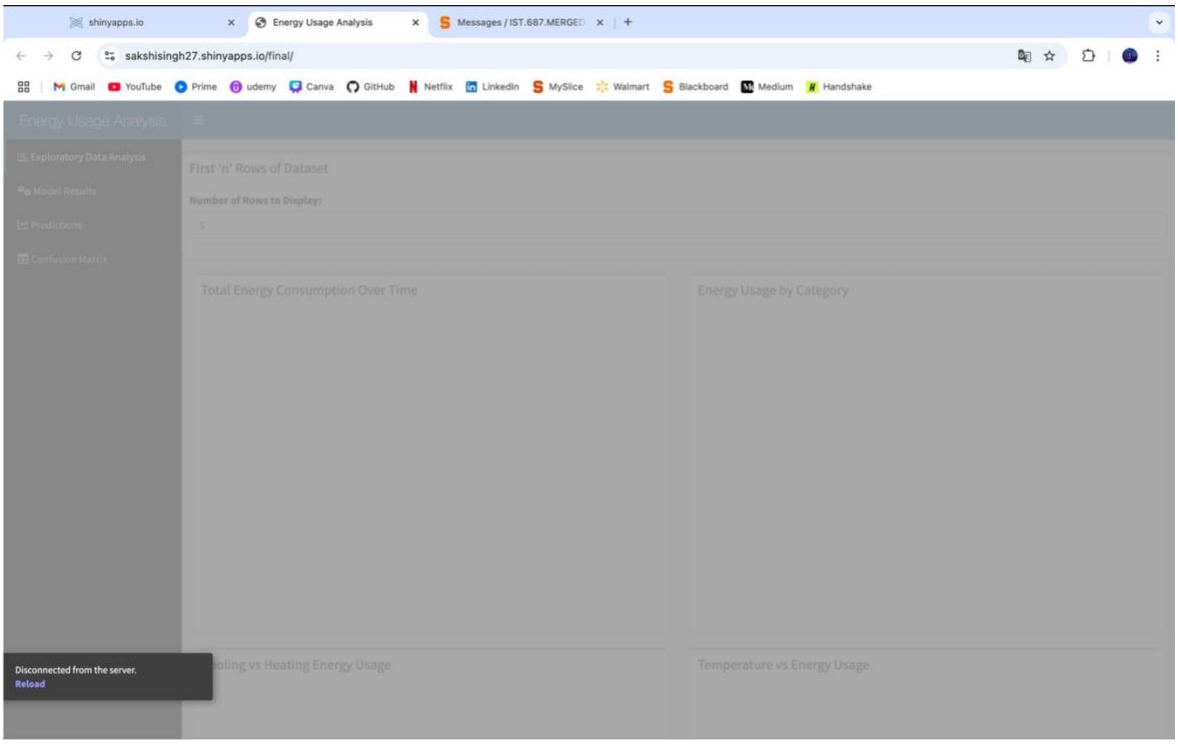
3. **Accessible URL**:
    - Upon successful deployment, the application was assigned a web URL where it could be accessed.
    - The website displayed the app's title ("final") as running on the ShinyApps.io server.



9.2 Current Status and Issues

Despite successful deployment and the application showing as "running," the website currently encounters a **"Disconnected from the server"** error every time it is accessed. This error prevents the app from displaying any content or functioning as intended.

- **Error Message**:
  The site repeatedly shows the error message, "Disconnected from the server. Please reload." Reloading the page does not resolve the issue.


shinyapps.io
Energy Usage Analysis
Messages / IST.687.MERGED
sakshisingh27.shinyapps.io/final/
Gmail
YouTube
Prime
udemy
Canva
GitHub
Netflix
LinkedIn
MySlice
Walmart
Blackboard
Medium
Handshake
Energy Usage Analysis
Exploratory Data Analysis
Model Results
Predictions
Confusion Matrix
First 'n' Rows of Dataset
Number of Rows to Display:
5
Total Energy Consumption Over Time
Energy Usage by Category
oling vs Heating Energy Usage
Temperature vs Energy Usage
Disconnected from the server.
Reload

CONCLUSION:

In conclusion, this project analyzed energy consumption patterns, evaluated predictive models, and developed an interactive Shiny application to explore energy demand. Through rigorous data preprocessing and modeling, Linear Regression emerged as the best-performing model, demonstrating strong predictive accuracy with interpretable results. Predictions revealed that regions with higher population densities and larger homes face greater peak energy demands, emphasizing the need for targeted energy efficiency measures.

An interactive Shiny application was created to visualize the analysis and predictions dynamically. However, deployment challenges, including server disconnection errors, limit its current functionality. Addressing these issues will enhance the app's usability and accessibility. Overall, this project provides valuable insights into future energy demands and offers a foundation for improving energy management and planning.