IST 687 - Introduction to Data Science

Group 3 - Energy Analysis & Prediction

Submitted By:-

Khushwant

Prashant

Shubham

Yaswanth

Table of Contents

Project Overview

- The energy company serving South Carolina and parts of North Carolina is gearing up for a challenging summer. With July temperatures expected to soar due to global warming, there's a likely surge in electricity demand, which could stress the electrical grid and possibly lead to blackouts during peak heat waves.
- Instead of the costly and environmentally taxing option of expanding their infrastructure, the company is choosing to tackle the issue proactively. They're diving deep into understanding what drives energy consumption and are crafting strategies to encourage households to embrace energy-saving habits.
- The company's efforts are centered on dissecting the factors that influence how much energy people use. With this knowledge, they plan to roll out targeted initiatives aimed at lowering the strain on the grid. This approach isn't just about keeping the lights on during the hottest days; it's also about adhering to broader sustainability objectives, thereby lessening the environmental toll of higher energy production.
- To promote energy conservation, the company is considering various strategies. These include educational campaigns that explain the benefits of reduced energy use, incentives for installing energy-efficient appliances, and encouraging habits that contribute to energy savings. The idea is to foster a community-wide ethos of mindful energy consumption.
- The overarching aim here is to diminish the risk of blackouts, enhance the reliability of the electricity supply, and support environmental stewardship. By sidestepping the need for new power plants and focusing instead on managing demand, the company hopes to address the increasing energy needs of its communities in a sustainable, effective way.

Project Objectives and Scope

Objective

The purpose of this project is to preemptively manage and relieve the strain on the electrical grid during these high-demand periods without having to build new energy generating facilities. Instead, the emphasis is on identifying the fundamental determinants that influence energy use, developing predictive models to forecast peak energy demand, and devising strategies to encourage energy saving among residential consumers.

Scope Components

Data Analysis and Insight Generation:

- Analyze past energy usage data to uncover the primary factors contributing to increased demand.

- Examine trends and patterns in energy consumption, with a particular focus on peak demand times.

- Evaluate the effects of temperature increases on energy consumption, enhancing our predictive capabilities.

Demand-Side Management Strategies:

- Craft and execute strategies aimed at encouraging energy-saving habits among homeowners.

- Promote the use of energy-efficient technologies through targeted campaigns and incentives.

- Launch educational initiatives to heighten community awareness about the benefits of energy conservation.

Community Engagement and Behavioral Change:

- Work closely with local communities to build a strong culture of responsible energy use.

- Roll out incentive programs designed to foster energy-efficient behaviors among consumers.

- Continuously monitor the impact of these community efforts on reducing overall energy consumption.

Monitoring and Evaluation:

- Develop specific metrics to gauge the effectiveness of implemented energy-saving strategies.

- Regularly perform evaluations to ensure that the strategies are achieving the desired outcomes and adjust as needed to meet the project's objectives.

Project Outcomes

To manage the expected increase in electricity demand during the summer months for the energy company (eSC) operating in South Carolina and parts of North Carolina, our project team is preparing a detailed and layered strategy. Here's a breakdown of our approach:

Data Collection and Preparation:

- We will compile historical electricity usage, weather data, and demographic information to form a comprehensive dataset.

- Our team will undertake thorough data cleaning to eliminate any inaccuracies or gaps, ensuring the data's integrity for reliable analysis.

Analysis of Energy Consumption:

- We plan to use linear regression models to pinpoint critical factors influencing energy consumption, especially during the peak of summer.

- A deeper analysis of these factors will help us understand their roles in driving overall demand.

Predictive Modeling for Demand Forecasting:

- Advanced predictive techniques, such as XGboost, will be employed to project future electricity demands for the coming summer.

- These forecasts will provide essential insights to help the energy company optimize its operational and strategic responses.

Interactive Data Visualization with Shiny Apps:

- Our analysis tool is tailored to provide a detailed view of electricity consumption across South Carolina, with enhanced functionality for various parameters..

-X variable Selection: Users can select x variable to examine local electricity usage, offering valuable insights into localized demand dynamics.

-Y variable Selection: Users can select y variable to examine various parameters effect on electricity usage, offering valuable insights into localized demand dynamics.

Detailed City Insights: Beyond statewide data, the tool drills down into each selected city’s unique energy consumption patterns and needs, highlighting how these vary across different locations.

Data Overview

Datasets Used:

A. Static House Data:

Description:

The Static House Data encompasses a random selection of single-family homes serviced by eSC (Energy Service Company). This comprehensive dataset lists all houses involved in the study, each identified by a unique 'building id' that also links to their respective energy usage records. The data includes various static attributes of each house, such as its square footage.

Format:

The information is stored in a 'parquet' format, which is ideal for managing large tabular data because of its efficiency and performance.

Size:

The dataset contains information on approximately 5,000 houses, each documented as a separate entry.

B. Energy Usage Data:

Description:

This dataset is a compilation of energy consumption records for the houses listed in the Static House Data. It details hourly energy usage, providing validated measurements of how much energy is used in different areas of the house, like air conditioning units and dryers. Each dataset is named after the 'building ID' of the house it represents.

Format:

The Energy Usage Data is stored in the 'parquet' format to ensure efficient storage and access.

Size:

The collection includes around 5,000 individual datasets, one for each house in the Static House Data.

C. Meta Data:

Description:

The Meta Data file acts as an explanatory guide to the attributes and fields included in the Static House Data and Energy Usage Data. It helps users understand and analyze the information contained in these datasets.

Format:

This document is typically presented in a straightforward, easy-to-read format such as plain text or CSV.

Size:

The Meta Data is concise, designed to be clear and informative without overwhelming the user.

D. Weather Data:

Description:

The Weather Data includes hourly weather records for each geographic area or county corresponding to the houses in the Static House Data. Each data point is timestamped to match the hourly energy usage data, and houses are linked to specific weather records via the 'in.county' column.

Format:

This dataset is maintained in a CSV format, which makes it easily accessible and simple to work with.

Data Processing Steps

1. House Classification:

Objective: To classify houses based on their built area “<900” and climate zone= “Hot-Humid”

Explanation: Cities were categorized according to whether they fell within designated census places. This distinction helps to identify variations in energy consumption patterns between these two types of urban classifications.

2. Date Formatting:

Objective: To standardize the format of the date column into a datetime format.

Explanation: The dates in the dataset were transformed into a uniform datetime format, facilitating more consistent and accurate date-time operations across the dataset.

3. Data Consolidation:

Objective: To merge the Sample House Data (SHD) with the Energy Data (ED) using the building data as a linkage key.

Explanation: The Sample House Data, which contains static details of houses, was combined with the dynamic Energy Data, which records hourly energy use, by linking them through a common building identifier. This merger creates a cohesive dataset that encompasses both structural and usage information.

4. County-Wide Data Aggregation:

Objective: To compile and aggregate total humidity and temperature data at the county level.

Explanation: Aggregation of humidity and temperature data was carried out at the county level. This consolidated climatic data was then integrated with the energy and housing datasets to facilitate analyses that explore how weather conditions impact energy consumption across larger geographic areas.

5. Comprehensive Data Integration:

Objective: To merge county-level environmental data with detailed house-level data.

Explanation: The aggregated environmental data from various counties was combined with detailed energy and housing data from individual homes. This comprehensive integration melds macro-level climatic influences with micro-level energy usage data, providing a nuanced perspective on how weather conditions affect household energy consumption.

6. County Grouping and Data Summarization:

Objective: To organize and summarize weather data across entire counties, focusing on total humidity and temperature.

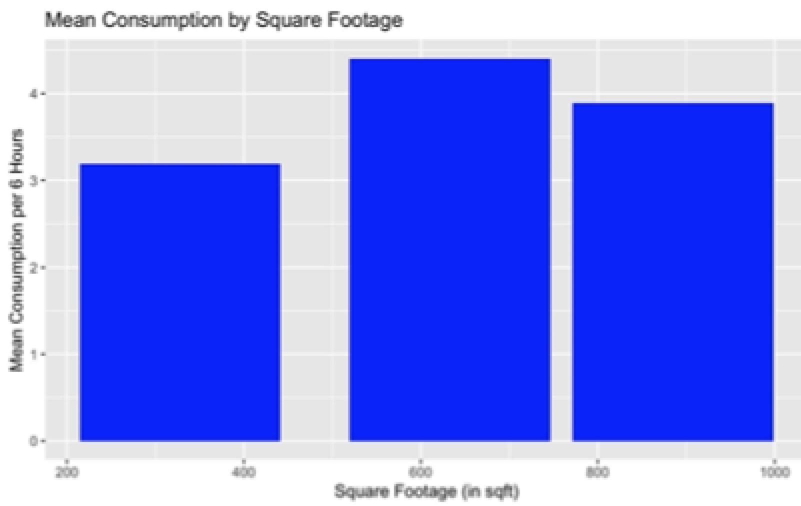
Explanation: This process involved collecting and summarizing weather data for entire counties, specifically looking at overall humidity and temperature levels. This summarized data was then linked back to the broader Energy and Housing datasets, enhancing the dataset's utility for broader analytical purposes by simplifying the data structure.

Analytical Insights and Models

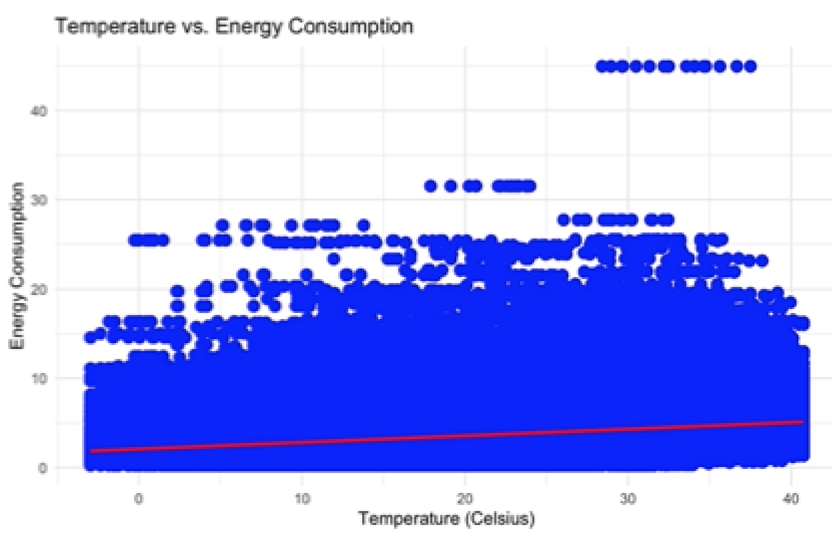
Analytical Insights:

From the project presentation and the provided data analysis, several key insights into residential energy consumption were derived:

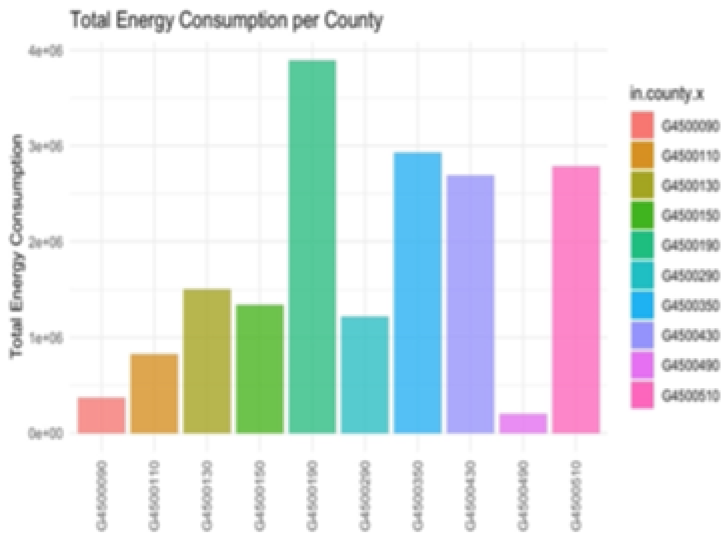
1. Impact of Building Characteristics: The analysis identified a positive correlation between building size (square footage) and energy consumption. Larger buildings inherently require more energy for heating and cooling, impacting overall energy usage patterns.



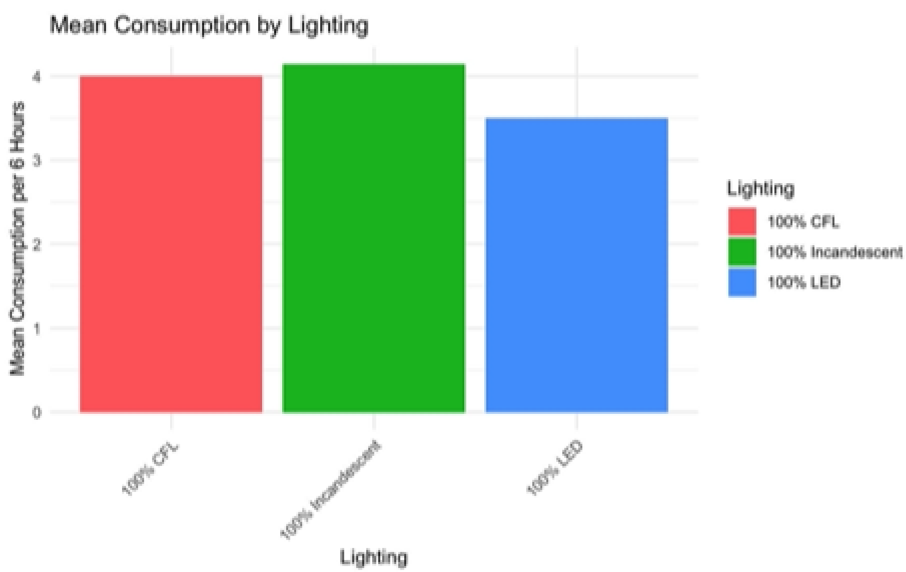
2. Influence of Weather Conditions: The data showed a direct relationship between temperature and energy consumption, with higher temperatures leading to increased use of cooling systems. This relationship was highlighted by the positive slope observed in the graph correlating temperature with energy usage, reflecting higher energy demands during hotter periods.



3. Variations by County and Insulation: Energy usage varied significantly across different counties, likely influenced by population density, local climate, and industrial activities. Additionally, different types of insulation and window setups affected energy efficiency, with Low-E glass and better-insulated walls showing lower energy consumption.



4. Lighting Efficiency: The comparison of lighting types revealed that LED lights are the most energy-efficient, followed by CFLs and incandescent bulbs. This highlights the potential energy savings from adopting more efficient lighting solutions in residential settings.



Predictive Models:

Two main predictive models were developed to forecast energy consumption and assist in managing peak energy demands:

1. Linear Regression Model: This model was used to establish a statistical relationship between energy consumption (dependent variable) and several independent variables such as square footage, weather conditions, and household characteristics. The model aimed to predict energy usage based on these factors, providing a foundational tool for understanding and managing energy demands.

````r
# 1.LR MODEL

```{r}
library(caret)
library(dplyr)

# Assuming you have a data frame named grouped_data
# Replace this with the actual code or data that defines grouped_data
grouped_data <- read.csv("final_all_merged.csv")  # Replace with the actual file or data loading code

set.seed(123)

# Check the number of rows in grouped_data
print(nrow(grouped_data))

# Splitting the data into a train-test data using a 70-30 split
trainList <- createDataPartition(y = final_merged_dataset$consumption_per_6hrs, p = 0.70, list = FALSE)
trainSet <- final_merged_dataset[trainList,]
testSet <- final_merged_dataset[-trainList,]

# Multiple Linear model
linear_model <- lm(consumption_per_6hrs ~ . , data = trainSet)
summary(linear_model)

# Predictions using the trained linear regression model
predictions <- predict(linear_model, newdata = testSet)

# These are the actual total energy values of test data
actual_values <- testSet$consumption_per_6hrs
# Obtain the R-Squared error.
R_squared <- 1 - (sum((actual_values - predictions)^2) / sum((actual_values - mean(actual_values))^2))
cat ("R-squared (R):", R_squared, "\n")
cat ("Accuracy of Linear regression model is : ", R_squared * 100, "\n")
```
````

2. XGBoost Model: An advanced machine learning model, XGBoost was utilized to handle non-linear relationships and interactions between variables more effectively than linear regression. This model was tuned with hyperparameters to optimize its performance, focusing on accurately predicting energy consumption during peak periods. The XGBoost model's accuracy was demonstrated through its R-squared value, indicating a strong ability to predict based on the inputs.

````r
# 3. XG BOOST - Using Hyperparameters
```{r}

# Convert logical columns to numeric for both trainSet and testSet
trainSet[] <- lapply(trainSet, function(x) if(is.logical(x)) as.numeric(x) else x)
testSet[] <- lapply(testSet, function(x) if(is.logical(x)) as.numeric(x) else x)



# Ensure all columns except the target are numeric for the model input
train_data_matrix <- as.matrix(trainSet[, -which(names(trainSet) == "consumption_per_6hrs")] %>% select_if(function(col) is.numeric(col)))
test_data_matrix <- as.matrix(testSet[, -which(names(testSet) == "consumption_per_6hrs")] %>% select_if(function(col) is.numeric(col)))

library(xgboost)
library(dplyr)
# Train the XGBoost model
model_xgb <- xgboost(
  data = train_data_matrix,
  label = trainSet$consumption_per_6hrs,
  objective = "reg:squarederror",
  nrounds = 100  # Number of boosting rounds; consider tuning this and other hyperparameters
)

# Predict on the test set
predictions <- predict(model_xgb, test_data_matrix)

# Calculate R-squared for the model's predictions
rsquared <- 1 - (sum((predictions - testSet$consumption_per_6hrs)^2) / sum((mean(testSet$consumption_per_6hrs) -
testSet$consumption_per_6hrs)^2))
cat("R-squared:", rsquared, "\n")
cat("Accuracy of XG BOOST with hyperparameters model is : ", rsquared * 100, "\n")
```
````

Interactive Visualization Applications

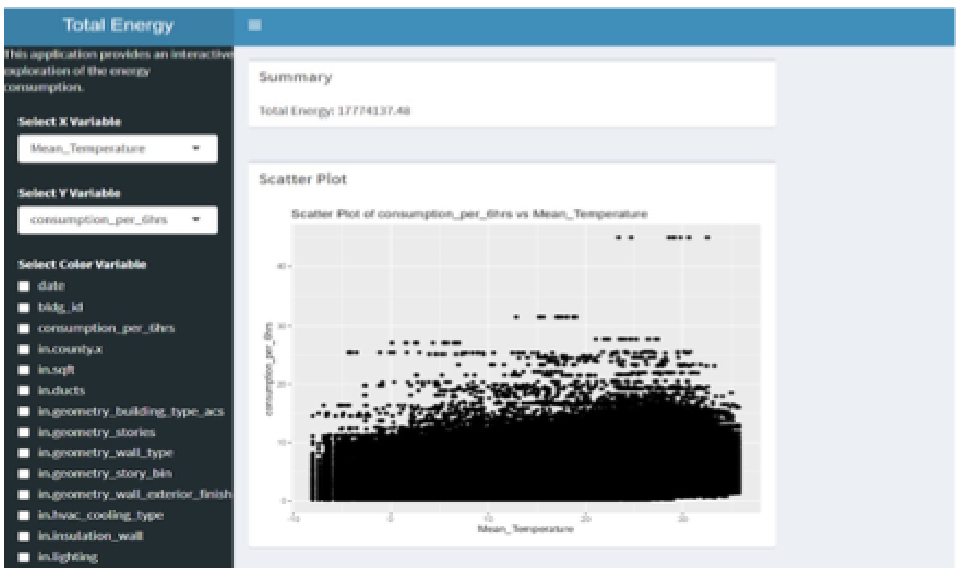
(Shiny App)

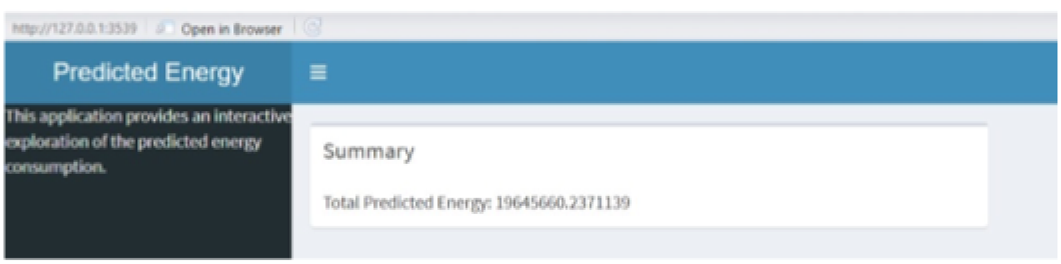
We used a Shiny web application to analyze and visualize energy consumption data in South Carolina (SC), with a particular focus on anticipating electricity demand during the peak month of July.

1. Scatter Plot Interface:
   - Main Visualization: The scatter plot demonstrates how energy consumption varies with changes in temperature. This visualization helps in understanding the relationship between these two variables visually.
   - Control Panel: On the left side, there's a panel where users can select different variables for the x-axis, y-axis, and the color of the data points in the scatter plot. This allows for dynamic exploration of the data depending on the user's interests or specific analysis needs. For example, they might choose to plot energy use against time of day or different weather conditions.
   - Summary Box: Above the scatter plot, there is a summary box displaying the total energy used, providing a quick reference that aggregates the data points visually represented in the scatter plot.
2. Predicted Energy Interface:
   - Summary Information: This simpler interface provides a summary of predicted energy consumption. It features a headline with "Total Predicted Energy" followed by a numeric value. This straightforward presentation focuses on delivering a key piece of information without additional context or visualization, which is useful for users looking for quick insights or conclusions derived from predictive models.
3. Graphical Navigation and Usability:
   - Interactive Elements: The interface includes interactive elements such as dropdown menus for selecting variables, which enhances user engagement by allowing personalized data exploration.

- Visual Design: The design uses a clean and minimalistic approach, with clear labeling and ample white space to prevent visual clutter, thus making the data and controls easy to read and use.

4. Functional Descriptions:
    - The accompanying text provides context for the user, explaining that the scatter plot depicts variance in energy use as temperature increases, which is an essential feature for educational or operational insights into how weather conditions affect energy demand.





URL: https://idsteam.shinyapps.io/Team3Project/

Strategies for Peak Energy Management and Energy Efficiency

Demand Response Initiatives:
Strategies for Managing Peak Energy Demand: Focusing on Ceiling Fans, Outdoor Lighting, and Freezers with Demand Response Programs

Managing peak energy demand is crucial for both energy providers and consumers. High peak demand can lead to increased energy costs, additional stress on the electrical grid, and a higher risk of blackouts or brownouts. Implementing demand response programs is one effective strategy to mitigate these issues.

Understanding Peak Energy Demand :

Peak energy demand occurs when electricity usage reaches its highest point, typically during specific periods such as hot summer afternoons or chilly winter evenings. During these times, households and businesses often use energy-intensive appliances and systems to stay comfortable or to perform daily tasks.

The Impact of LED Lighting, Insulation, and Weather-Responsive Energy Management

LED Lighting: LED (Light Emitting Diode) lighting is one of the most energy-efficient lighting technologies available today. LEDs consume significantly less electricity compared to traditional incandescent and CFL (Compact Fluorescent Lamp) bulbs, translating into lower energy bills and reduced demand on the power grid. The efficiency of LED lighting is primarily due to its ability to convert electricity into light with minimal heat loss, making it highly effective for both indoor and outdoor use. Moreover, LEDs have a longer lifespan, which decreases the frequency of replacement and the associated environmental impact of manufacturing, transporting, and disposing of light fixtures.

Insulation: Proper insulation in buildings plays a critical role in enhancing energy efficiency by maintaining a stable indoor temperature, thereby reducing the demand for heating in winter and cooling in summer. Insulation materials like fiberglass, foam, and cellulose help prevent heat transfer through walls, roofs, and floors. Higher R-values, indicating greater insulative properties, can significantly diminish the energy required to heat or cool a space. Effective insulation not only reduces energy consumption but also lowers greenhouse gas emissions and enhances occupant comfort. Upgrading insulation is considered a cost-effective measure that contributes to energy savings over the long term.

Weather-Responsive Energy Management: Weather-responsive energy management involves the use of technologies and strategies that adjust energy consumption based on current weather conditions. This approach typically utilizes smart thermostats and automated building management systems that can react in real-time to changes in temperature, humidity, and sunlight. For instance, smart thermostats can reduce air conditioning usage automatically during cooler days or adjust settings based on the humidity levels to optimize comfort and energy use. Such systems are integral to reducing energy consumption during peak demand periods, particularly in regions experiencing wide temperature fluctuations. Additionally, weather-responsive systems can enhance the overall efficiency of renewable energy sources like solar panels by adjusting energy use based on available solar power, further supporting sustainable energy practices.

Mitigating Peak Energy Demand Through Demand Response Programs

Demand response programs are crafted to mitigate peak energy demand by encouraging consumers to minimize their electricity usage during high-demand periods. Here's a closer look at how these programs can be effectively applied to the primary contributors to energy consumption:

- Adopting LED Lighting: Switching to LED lighting across households to reduce energy consumption significantly, as demonstrated by the lower energy usage of LEDs compared to other lighting types.
- Improving Insulation: Upgrading insulation materials in buildings to higher R-values to enhance thermal efficiency and reduce the need for heating and cooling, thereby lowering energy consumption.
- Weather-Responsive Energy Management: Implementing smart thermostats and energy management systems that adjust energy usage based on real-time weather data to optimize consumption during peak demand periods.

Conclusion

The project successfully highlighted the critical factors affecting residential energy consumption and developed predictive models to forecast energy usage. The findings underscore the importance of targeted interventions such as upgrading insulation and lighting to reduce energy consumption. These insights and recommendations can help utility providers and policymakers implement more effective energy management strategies, promoting sustainability and reducing the strain on energy infrastructures during peak periods.