Forecasting and Optimizing Peak Energy Demand: A Data-Driven Approach for Sustainable Energy Management

Project Report submitted in the partial fulfillment

of

M.S

in

Applied Data Science

by

Amulya Walimbe (954285988)

Avi Sharma (249873359)

Dharmil Gandhi (205670780)

Hari Ram Selvaraj (686582383)

Under the supervision of

<u>**Akit Kumar**</u>

Syracuse University

SCHOOL OF INFORMATION STUDIES

(Fall 2024)

ACKNOWLEDGEMENT

The success and outcome of this project required guidance, support, and assistance from numerous individuals, and we are deeply privileged to have received such valuable help throughout the completion of this project. We would like to extend our heartfelt thanks to all those who contributed to making this project a success.

We respect and thank **Christopher Dunham**, for providing us the opportunity to undertake this project and for offering invaluable support and guidance, which enabled us to complete the project successfully. His encouragement and assistance were crucial throughout the process, and we are truly grateful for his constant backing.

We are extremely thankful to **Akit Kumar**, our project mentor, for taking a keen interest in our work and guiding us diligently throughout the project. His expertise, advice, and continuous support have been instrumental in shaping our approach and helping us develop a comprehensive solution for the project.

We would also like to express our gratitude to **Syracuse University's School of Information Studies** for providing us with the resources and opportunities necessary for completing this project, as well as fostering an environment that encouraged academic growth.

Our deepest thanks go to the various professionals in the energy sector who took time out of their busy schedules to provide us with essential data and insights, enabling us to approach the problem from a practical and informed perspective.

This project would not have been possible without all the assistance, and we are grateful to everyone who contributed to its completion.

Amulya Walimbe (954285988)

Avi Sharma (249873359)

Dharmil Gandhi (205670780)

Hari Ram Selvaraj (686582383)

Table of Contents

Abstract

This project focuses on forecasting and optimizing peak energy demand, leveraging data science techniques to assist **eSC - Energy South Carolina** (an energy and power supplying company in North Carolina, USA) in managing electricity usage during extreme summer heat. The primary goal is to predict energy demand for a hotter-than-usual July, which would help the company avoid potential blackouts and reduce reliance on building additional energy production facilities. The system utilizes various data sources, including Static House data, hourly energy usage profiles, and weather data for over 5,000 residential properties.

By employing a multi-step process involving data preparation, exploratory analysis, and predictive modelling, we have identified key drivers of energy consumption, such as air conditioning usage and geographic factors. Several machine learning models were tested, and the best-performing model was selected to predict future energy demand under the assumption of a 5-degree increase in July temperatures. The findings provide actionable insights to reduce peak energy usage and manage demand more efficiently, including strategies like time-of-use pricing and promoting energy-efficient technologies.

Additionally, a Shiny app has been developed to allow stakeholders to interact with the data and understand the model's predictions. This project offers a data-driven approach to tackling energy sustainability challenges, presenting a reliable solution to help **eSC** optimize its energy resources while minimizing environmental impact.

Introduction

There is a need of effective management of peak energy demand, which ensures the reliability and sustainability of power systems and companies, especially since the immense augmentation in global warming. Elevated temperatures significantly increase electricity consumption due to the widespread use of air conditioning, which in turn strains the power grid and raises the risk of blackouts (Chandel et al. 2016). Several researches conducted showed that traditional responses to these surges in demand lead to expanding infrastructure, such as building new power plants and add additional strength to the power lines, which increases the economic burden and damage on the environment (Reddy & Assenza, 2020).

Recent advancements in data analytics have aided to introduce several strategies which help to manage the peak in the demand, however each strategy is individualistic in nature and cannot be easily applied to a different country or even state. These researches helped to either reduce the consumption or analyse the patterns which lead to lower energy consumption. Dimitriou and Kiranoudis (2019) demonstrated that encouraging customers to shift electricity usage to off-peak hours helped to control the energy consumption.

Building on this foundation, this project aims to leverage advanced data science methodologies to forecast and optimise the peak energy demand for Energy South Carolina (eSC). The project will analyse data from residential building characteristics, occupant behaviours, and regional weather conditions. This study aims to predict energy consumption during extreme summer temperatures, specifically focusing on July, the peak energy usage month (U.S. Environmental Protection Agency, 2021). The goal is not only to predict energy demand but also create targeted strategies, such as promoting energy-efficient technologies and encouraging behavioural changes, which can reduce the reliance on additional infrastructure (Luo & Du, 2024).

The relationship between structural building characteristics and energy consumption is one of the extensively studied research topics. Floor area, number of stories, and insulation has significantly influenced residential energy demand. Buildings with better insulation have been shown to reduce the electricity demand due to reduced heat loss (Electrification of Heat demonstration Project). Similar larger floor areas and houses with multiple stories are associated with higher energy consumption because of increased heating and cooling requirements (Chandel et al ., 2016)

Additionally, occupant behaviour plays a crucial role in shaping residential electricity usage. Behaviours such as variations in thermostat settings, lighting preferences and appliance usage can lead to differences in energy demand. Several studies indicate that small behavioural changes, such as lowering thermostat settings, by one degree during the summer, can reduce energy consumption by 6% (Dimitriou and Kiranoudis, 2019). Ghahramani et al. (2021) showed that lighting and appliance and usage patterns vary significantly across household, contributing to differences in energy demand across an area.

Climate factors such as temperature, humidity and wind speed are one of the critical predictors of electricity consumption, especially during peak summer months. The U.S energy Information Administration (2020) shows how each degree change in Fahrenheit leads to marked rise in cooling energy requirements. In addition, humidity augments the cooling needs, while wind speed impacted the efficiency of HVAC systems (Luo and Du, 2024).

Energy efficient interventions, such as HVAC systems, insulation technologies and energy efficient lighting have demonstrated reduced residential electricity consumption effectively. Liu et al. (2018) showed that energy efficient HVAC systems and lighting showed a 20-30% reduction in energy usage in comparison to the conventional methods. Improvements in the insulations of walls and roofs are also one of the most cost effective method of lowering electricity demand (Chandel et al., 2016)

This investigation leads to stating four research questions for the project which aims to understand the determinants of energy consumption.

1) What is the relationship between structural building characteristics, such as floor area, number of stories, and insulation, and electricity consumption?
2) How do variations in occupant behaviours, including thermostat settings, and appliance usage influence electricity demand in residential buildings?
3) To what extent do climatic factors, such as temperature, humidity and wind speed, affect hourly electricity consumption during peak summer months?
4) Which energy efficiency interventions, such as HVAC systems, insulation types and lighting technologies, have the strongest impact on reducing residential electricity consumption?

This project will contribute to a wider understanding of sustainable energy management and provide actionable insights for electricity to enhance grid reliability and stability, whilst minimizing environmental impacts during peak periods. By exploring the interplay between structural attributes of the house, behavioural patterns of the occupants, climate and energy efficient measures, this project will aim to general actional insights to optimise energy demand during peak periods.

Data Overview

This project integrates multiple datasets to analyse and predict energy consumption patterns. Each dataset will provide unique insights into the potential various factors influencing residential electricity demand. There were four datasets that were sourced in order to conduct the analysis.

1) Static House data

   The data file contains structural and demographic information of 5710 residential houses which are supplied energy by eSC. The file consisted factors such as building attributes, insulation type, HVAC systems and occupancy details.

2) Energy Usage data

   The data file consists of hourly energy consumption for each house, aggregated across appliances like HVAC, lighting, and water heaters. The data file has the hourly energy consumption for the full year 2018.

3) Weather data

   The data file consists of hourly weather data for 50 counties in North Carolina, which includes the temperature, humidity, wind speed, and solar radiation. This dataset includes the essential data for climatic conditions which influence electricity demand.

4) Metadata

   This data file consists of detail description of all the variables and formats of each variable across the 3 datasets above. The file helped to understand the variables and acts like a dictionary to the variables in the datasets.

The data was accessed from the publicly available URL hosts by Amazon Web Services (AWS). The data was read from either a Parquet or csv format directly into R Studio.

Data Preparation

The data preparation step was crucial to ensure that the datasets were clean, consistent, and properly integrated and ready for analysis. This stage involved several steps, such as data integration, handling missing values and ensuring datasets are in a suitable format for modelling.

A comprehensive analysis was conducted on the datasets to comprehend the unique identifiers in the datasets. Static house data and weather data were merged based on the county code, which ensures that each house which was paired was with the corresponding weather condition based on its' geographical location. The energy dataset was merged with the combined house and weather data using an unique key which was created by concatenating building ID's and timestamps. This aided us to align the hourly energy consumption data with weather and the respective house id. Energy dataset was sliced to only show the consumption for the month of July, which was repeated for the weather dataset as well. This led us to a file with about five million observations with around 170 variables.

Due to the size and complexity of the dataset, missing data was inevitable. Therefore, we introduced various strategies to address the issues. The missing values in the weather dataset were handled using linear interpolation, to ensure the temporal continuity of the data while accurately interpolating missing values. The categorical data in the static house data was handled using forward and backward interpolation based on similar houses in the dataset. Energy dataset was handled for missing values by using backward/forward filling to ensure temporal consistency.

Several variables from the static house data and weather data were deemed irrelevant with the use of the metadata dataset analysis. Some columns exhibited zero variability across observations were removed to reduce noise and improve computational efficiency. The removal of such variables helped the dataset to look clean however it was not ready for analysis.

The dataset was transformed to standardize the data for modelling and various transformations were applied. The dataset was transformed to first ensure that the numerical variables were in a numerical data attribute. Furthermore, the categorical variables were identified and transformed into categorical values by careful factorization of the variables. The dataset was converted into numerical columns again to ensure that the modelling is not affected. However, this step was conducted after the exploratory data analysis.

Exploratory Data Analysis

The exploratory data analysis phase was essential to understand the characteristic of the data, understand the trends, and uncovering relationships between variables, which helped to guide the model development. Descriptive statistics were computed to summarise for both numerical and categorical variables. Numerical variables, such as electricity consumption, temperature, floor area and several more, were computed for mean, median, standard deviation and range. The electricity consumption, a primary focus of the analysis, revealed a mean of 1.23 kWh, with a standard deviation of 0.81 kWh in July. Floor areas ranged from 328 squared feet to over 8194 square feet. Categorical variables including ceiling insulation showed that R-30 was the most common insulation type.

Correlation analysis was conducted to examine the relationship between numeric variables and identify key drivers of electricity consumption. The results showed that cooling energy consumption had the strongest positive correlation with electrical consumption (r=0.76), confirming the significant impact on total energy consumption. Ceiling fan showed moderate correlation, highlighting the role of maintaining indoor comfort. The size of the house were also positively correlated with electrical consumption (r=0.36). Variables relating to weather showed moderate correlations with electricity consumption (r = 0.56). Moreover, diffuse horizontal radiation showed a positive correlation of r = 0.34 with cooling energy consumption.

The key observations from the EDA highlighted the interplay between building attributes, weather attributes, and behavioural factors in shaping the electrical consumption. Behavioral variations, such as cooling setpoints averaging 72.97°F (ranging from 60°F to 80°F) and appliance usage, introduced further variability. Temperature extremes, with maximums reaching 38.30°C, along with high relative humidity, were major drivers of peak electricity usage. These insights provided a comprehensive understanding of the dataset and ensured that all relevant factors were incorporated into the predictive framework, setting the stage for accurate and actionable modeling outcomes.

Model development

Various predictive models were implemented to explore the relationships between electricity consumption and a diverse range of influencing factors. The objective of the project was to build the models that provide insights into reducing electrical consumption. Two primary modeling techniques were used, linear regression and XGBoost. The modelling technique were used to understand a linear model and a model which can interpret non-linear models as well. The project delves into two models because of computational restrictions.
The dataset was split into training and testing (60% and 40% respectively) sets to ensure that the models could be evaluated and whether the models can be generalizable to a warmer July. Preprocessing steps such as handling missing values and ensuring the variables are in the correct format were conducted before creating the models. Data was converted into a DMatrix format to optimize the XGBoost model for a big dataset.

The first research question focuses on the impact of building attributes. The variables included in the model were in.geometry_floor_area, in.geometry_stories, in.vintage, in.geometry_garage, in.ceiling_fan, in.insulation_slab, in.ducts, in.roof_material, and in.tenure. The research question 2 focus shifts to behavioral attributes were the variables used are in.cooling_setpoint, in.heating_setpoint, in.occupants, in.lighting, in.misc_extra_refrigerator, in.misc_pool_pump, in.misc_gas_fireplace, in.range_spot_vent_hour, in.hot_water_fixtures, and in.vacancy_status. The research question 3 investigates the effects of weather conditions. The variables used in the model are dry_bulb_temperature, relative_humidity, wind_speed, and wind_direction. The research question 4 focused on energy efficiency measures where the variables used were out.electricity.cooling.energy_consumption, out.electricity.heating.energy_consumption, out.electricity.lighting_interior.energy_consumption, out.electricity.refrigerator.energy_consumption, out.electricity.clothes_dryer.energy_consumption, out.electricity.clothes_washer.energy_consumption, out.electricity.plug_loads.energy_consumption, dry_bulb_temperature, in.hvac_cooling_efficiency, in.hvac_heating_efficiency, out.electricity.hot_water.energy_consumption, out.electricity.pool_pump.energy_consumption, out.electricity.ceiling_fan.energy_consumption, in.insulation_ceiling, in.insulation_wall, relative_humidity, in.sqft, and in.occupants.

Each model undergoes a linear model and XGBoost modelling, where linear model aids to quantify the individual impact of variables on electricity consumption. XGBoost was used to capture the complex interactions among these variables.

Finally, a comprehensive final model is created combining all significant variables to explore the cumulative effects of structural, behavioral, weather and energy efficient factors on electricity consumption. This model will provide a holistic view, integrating all dimensions of influence. An evaluation is created for each model which will give a Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared (R^2). The MSE measures the average squared differences between the actual and predicted values, which shows the model’s overall error where a lower MSE indicated better predictive performance by the model. MAE calculates the average absolute difference between the actual and predicted values, which gives an interpretation of model’s average error in the same units as the target variable, for better interpretability. The r-squared values show the proportion of variance in the dependent variable that can be explained by the independent variables. Each metric will highlight a distinct aspect of the model performance to understand how the model is performing.

Results

The outputs of the models provide valuable insights into the relationship between electricity consumption and the selected predictors across different research questions. For Model 1, the linear regression results show that structural variables, such as floor area, garage presence, and vintage, contribute to electricity usage, with the model explaining approximately 10.86% of the variance ($R^2 = 0.1086$). Larger floor areas and the presence of ceiling fans were positively associated with higher electricity consumption, while variables like insulation slab and tenure showed negative impacts. The XGBoost model demonstrated slightly better predictive performance with a lower RMSE and higher accuracy.

For Model 2, focusing on behavioral factors like heating and cooling setpoints and appliance usage, the linear regression model achieved an improved R^2 of 0.2059, indicating a stronger explanatory power. The cooling setpoint, lighting, and hot water fixtures emerged as significant drivers of electricity consumption. XGBoost consistently outperformed linear regression for behavioral variables, highlighting its ability to capture complex, nonlinear relationships.

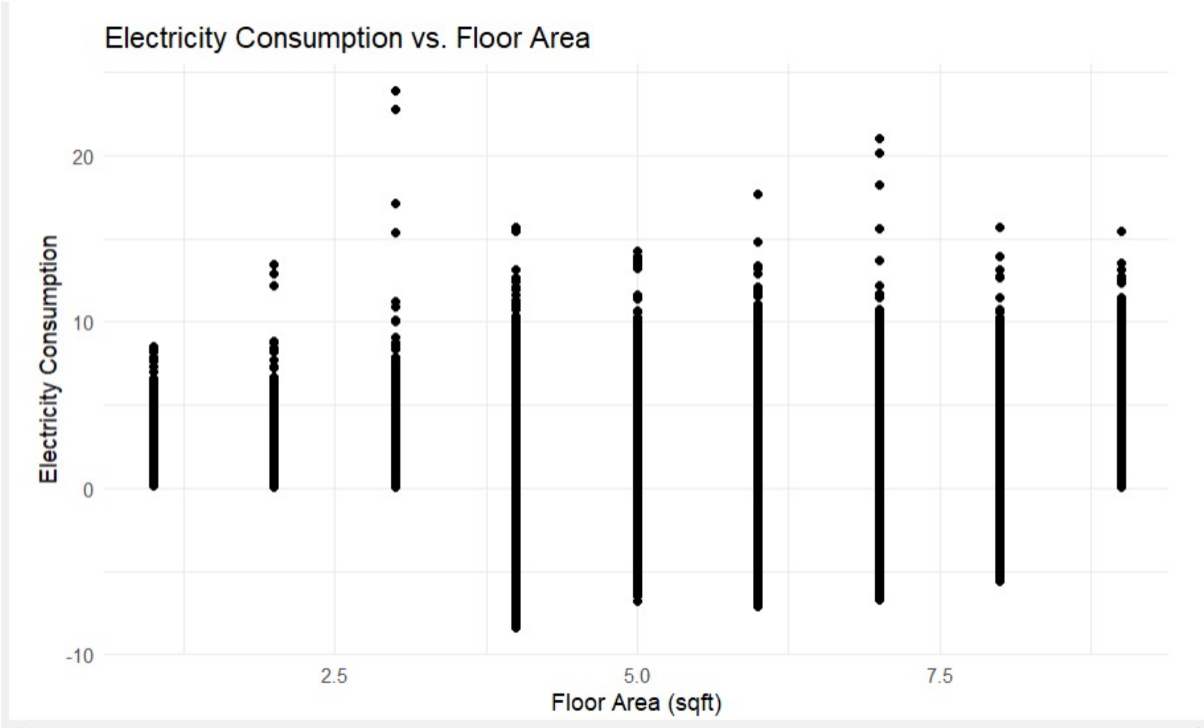
Model 3, which examined climatic variables such as temperature and wind conditions, showed an even stronger fit with an R^2 of 0.8908 in the linear regression model, suggesting that weather factors are highly predictive of electricity consumption. Similarly, Model 4, which incorporated energy-efficient measures, had an R^2 of 0.4694, identifying key variables like insulation, HVAC efficiency, and lighting as significant contributors.

The final comprehensive model combining all predictors yielded an R^2 of 0.4688 with an MSE of 0.3475, demonstrating moderate predictive accuracy. Overall, these results highlight the importance of combining structural, behavioral, climatic, and energy-efficient variables to better understand and predict electricity consumption, with XGBoost generally providing superior predictive performance.

Table 1: Results for research question 1

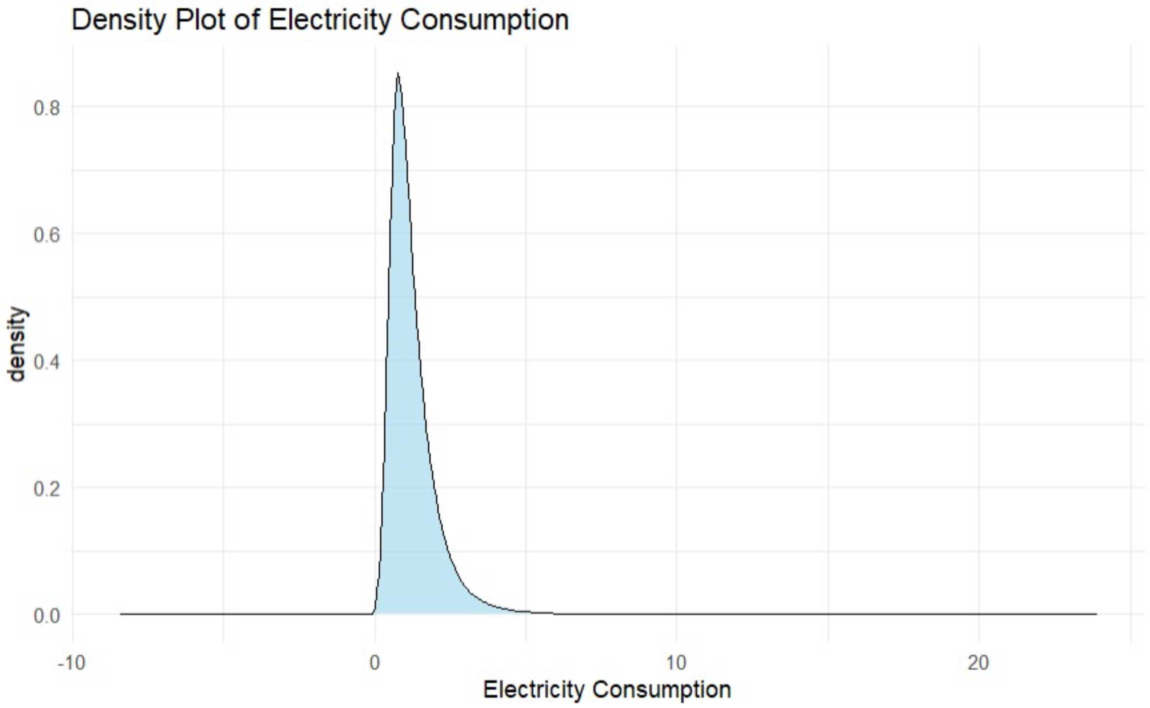
Metric	**Linear Regression**	**XGBoost**
MSE	0.5831	0.5157
MAE	0.5321	-
R^2	0.1087	-

Figure 1: Research question 1 Visualization



The scatter plot above showcasing electricity consumption against floor area highlights a clear trend: larger homes tend to consume more electricity. This observation aligns with expectations, as bigger spaces require more energy for heating, cooling, and lighting. However, the data also reveals considerable variability at each floor area level. This variability may be attributed to differences in insulation quality, the number of occupants, or the presence of energy-intensive appliances. Notably, there are outliers among homes with larger floor areas that exhibit disproportionately high electricity consumption. These cases could represent homes with unique energy-intensive features such as swimming pools, home offices, or extensive outdoor lighting, or they might reflect inefficiencies in energy use.

Figure 2: Research question 1 Visualization

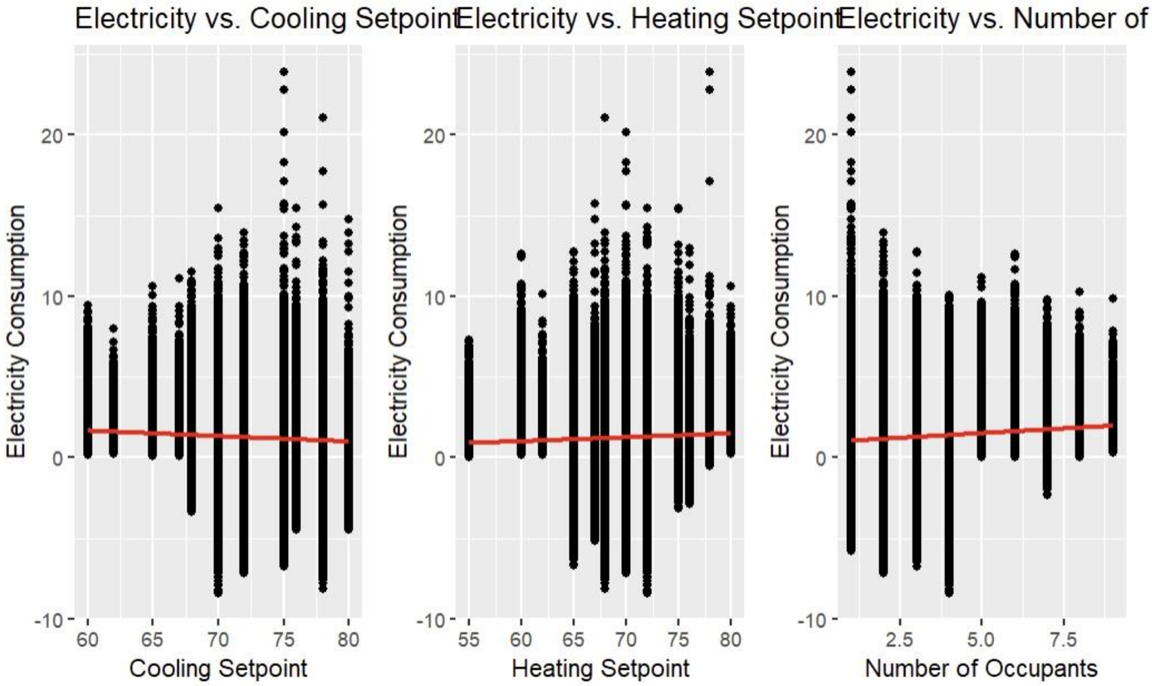


The density plot above for electricity consumption presents the overall distribution of energy use among households. The sharp peak near zero indicates that most households consume moderate amounts of electricity, suggesting that efficient energy usage is prevalent for a significant portion of homes in the sample. However, the long tail extending to the right highlights a smaller subset of households with significantly higher energy consumption. These households likely represent outliers, either due to their structural features, behavioral patterns, or a combination of both.

Table 2: Results for research question 2

Metric	**Linear Regression**	**XGBoost**
MSE	0.5196	0.3158
MAE	0.4862	-
R^2	0.2059	-

Figure 3: Research Question 2 Model Visualization



The figure 3 for Model 2 examines the consumption of electricity as a function of cooling setpoint, heating setpoint, and occupants in the house. It follows that there is a slight negative correlation between cooling setpoints and electricity consumption, where higher cooling setpoints result in cooler indoor temperatures and lower demand for electricity. This result is consistent with the basic principles of energy efficiency since less cooling energy is used when indoor temperatures are maintained closer to outdoor temperatures. On the other hand, there is a similar but weaker trend observed for the heating setpoint: higher heating setpoints lead to a reduction in electricity consumption. This was expected, given that reduced energy demand for heating will occur when indoor temperature settings are closer to outdoor conditions. Finally, electricity consumption rises with the number of occupants since large households use more appliances and lighting, which implies that occupant behavior has a great influence on energy demand.

Figure 4: Research Question 2 Model Visualization

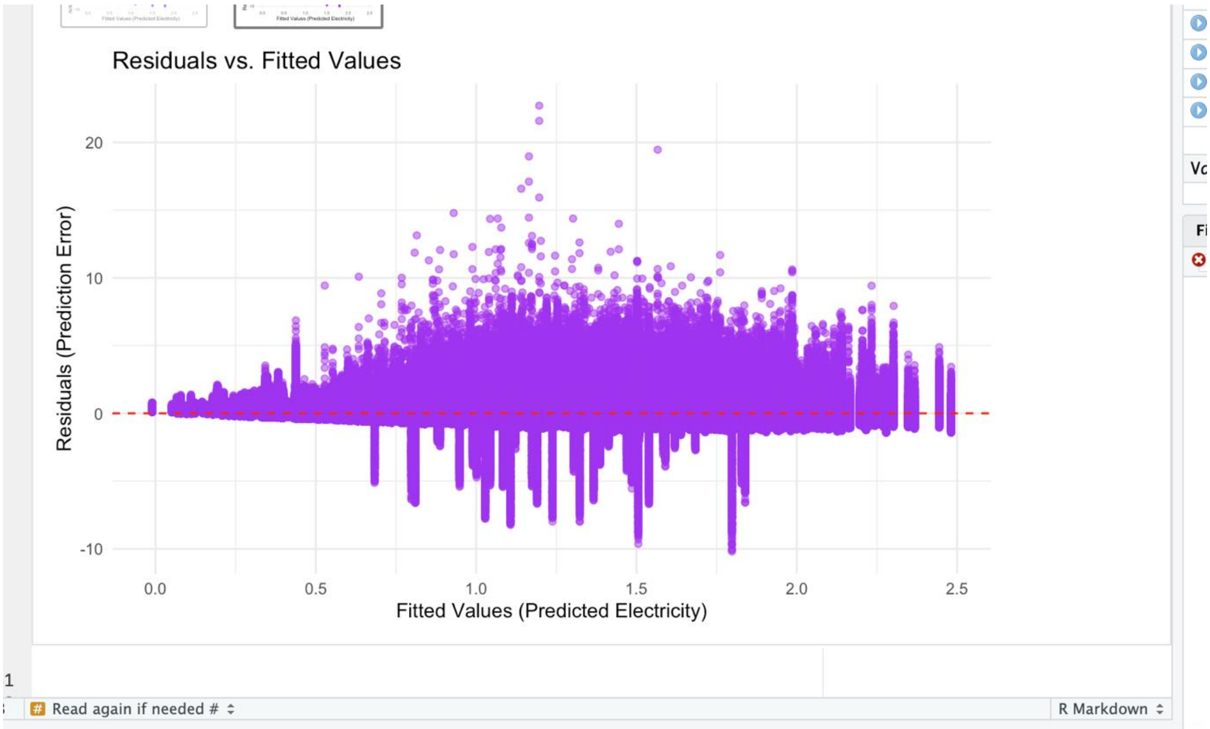


Figure 4 compares fitted electricity consumption values from the model with actual electricity consumption values. A clustering of data points around the red diagonal line indicates that the model performs reasonably well in predicting electricity demand. However, some deviations from the diagonal at higher consumption levels suggest that the model struggles with extreme cases, potentially due to unaccounted-for variability or non-linear relationships.

Figure 5: Research Question 2 Model Visualization

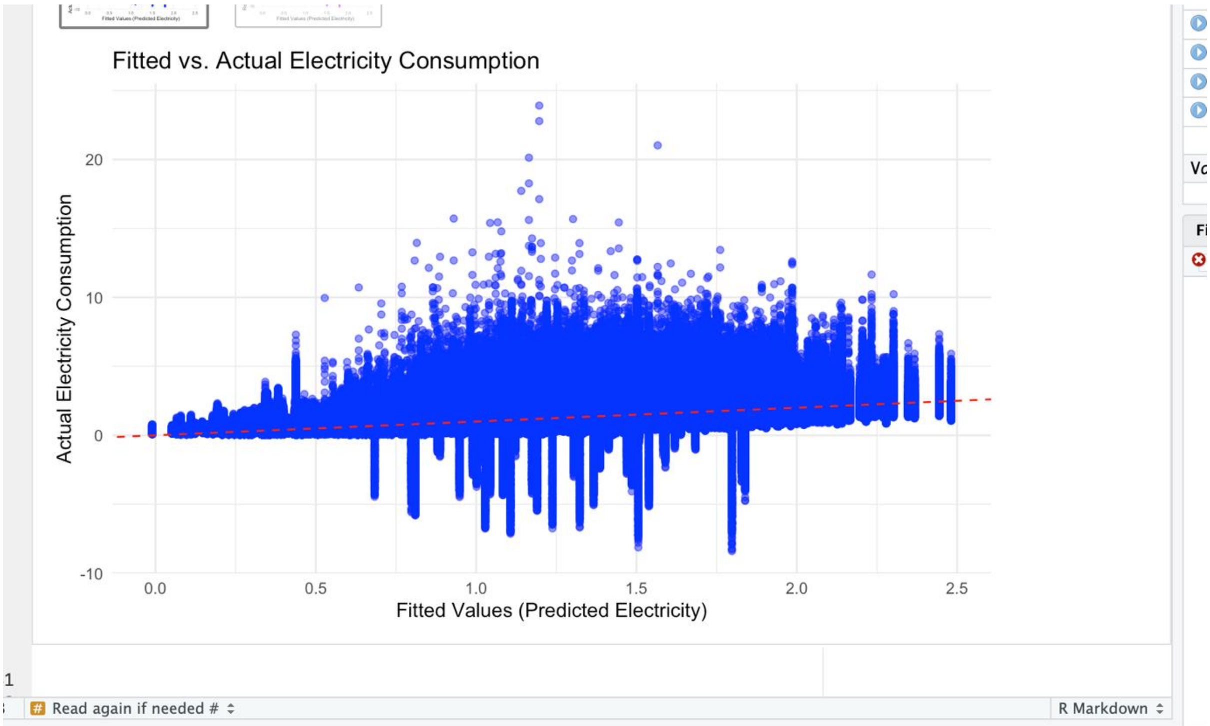
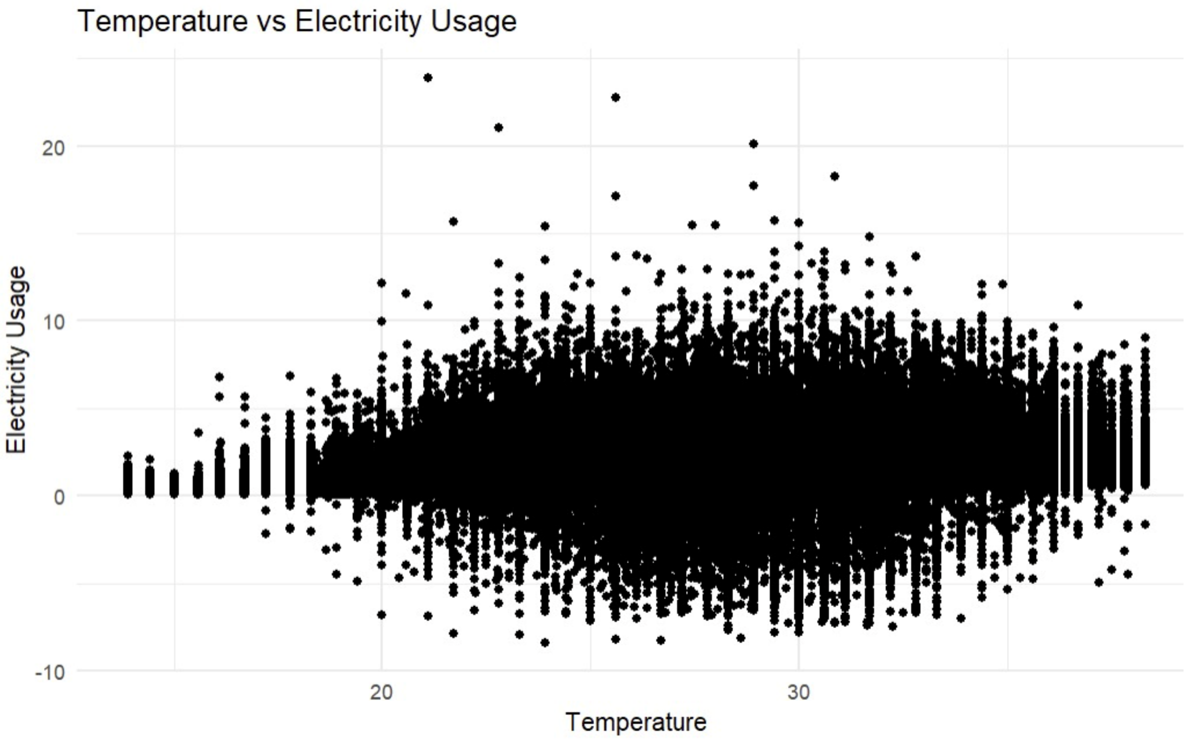


Figure 5 plots residuals (prediction errors) against fitted values. The spread of residuals around zero with no distinct pattern confirms that the model captures the general trend of the data. However, the presence of larger residuals at higher fitted values indicates potential room for improvement in predicting outliers or high-energy-use households. The lack of a clear pattern in the residuals suggests that the assumptions of linear regression—such as homoscedasticity—are reasonably met, further supporting the model's validity for this dataset.

Table 3: Results for research question 3

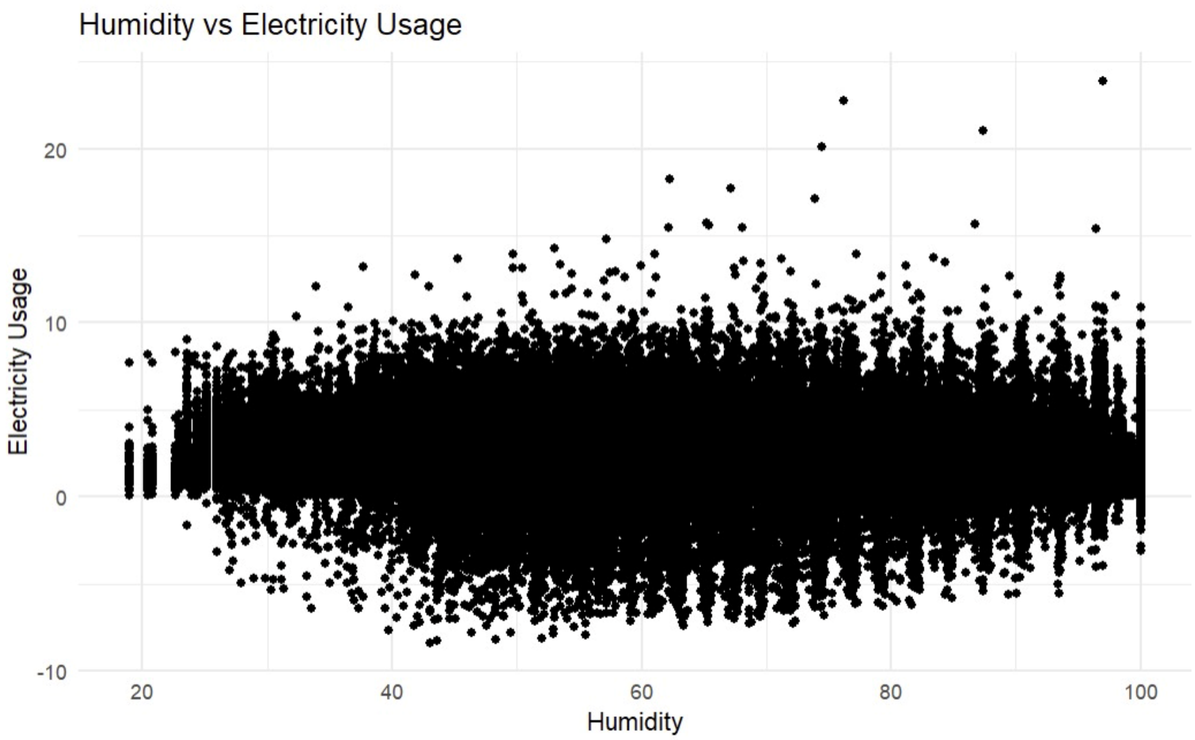
Metric	**Linear Regression**	**XGBoost**
MSE	0.7376	0.7245
MAE	0.4959	-
R^2	0.1685	-

Figure 6: Research question 3 Visualization



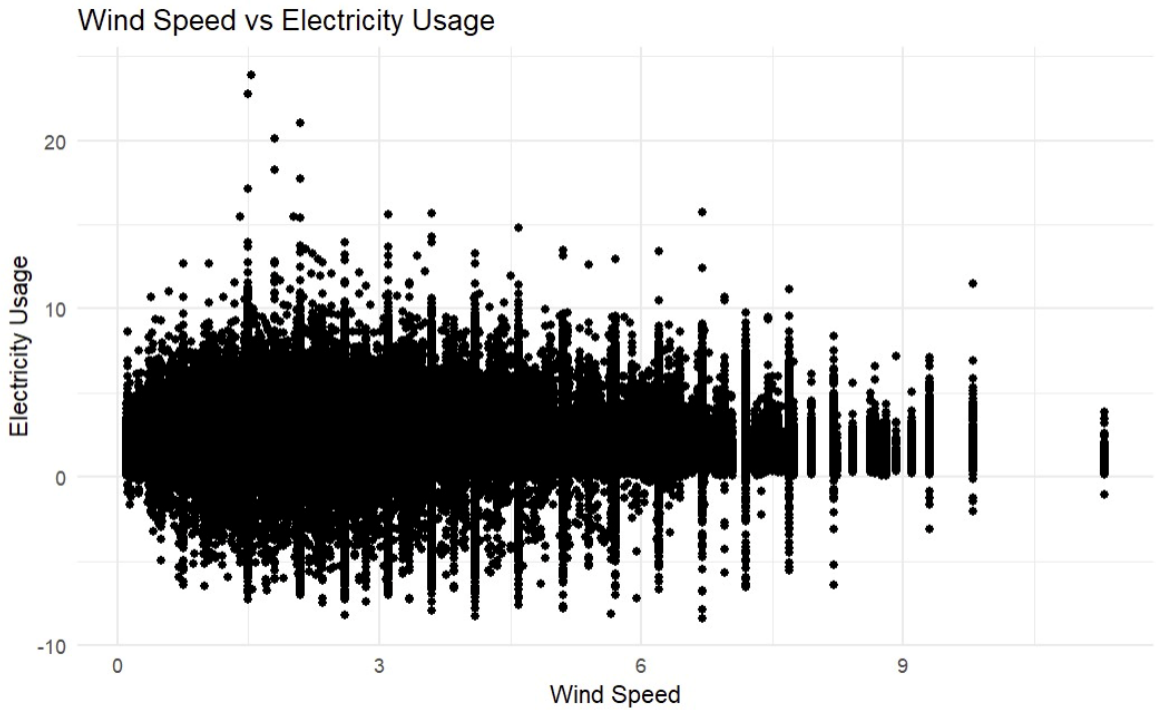
The scatter plot "Temperature vs. Electricity Usage" indicates that the electricity consumption goes upward when temperatures are increased beyond 25°, indicating an increased utilization of cooling systems in hotter seasons. Most of the points show high concentration at middle temperature ranges, suggesting that peak energy consumptions are mostly at moderate temperatures in most households. This behavior is in accordance with the expected relationship between heat and cooling demand, further highlighting the importance of temperature as a major driver of summer electricity use.

Figure 7: Research question 3 Visualization



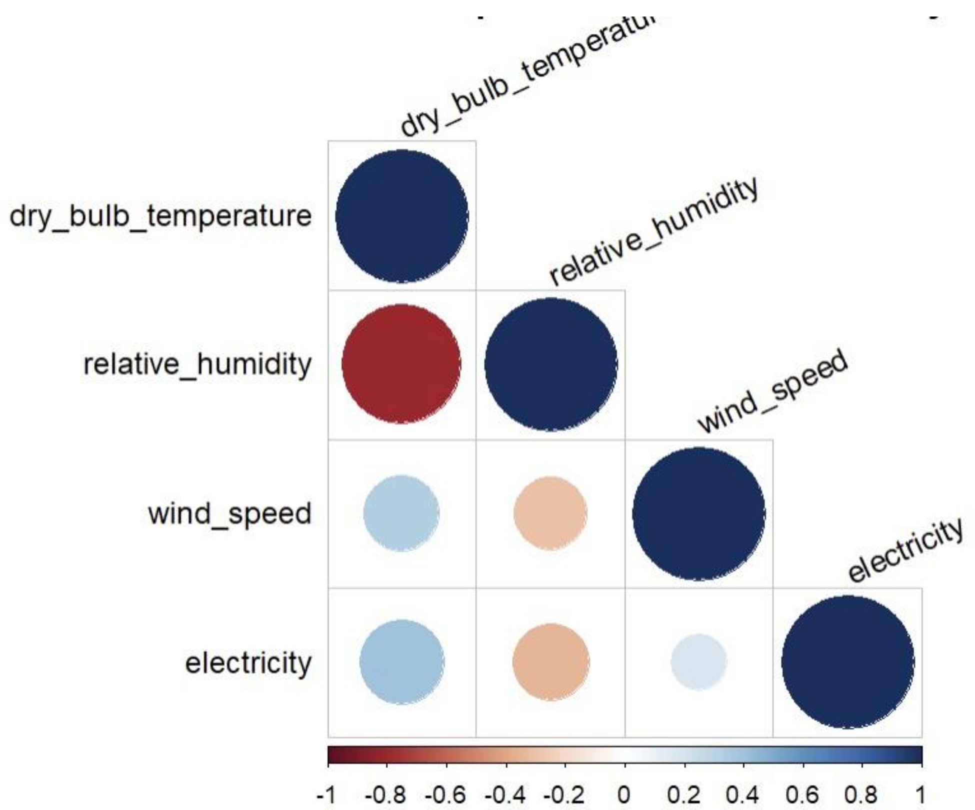
The "Humidity vs. Electricity Usage" visualization illustrates a similar pattern, though the relationship seems to be far less obvious than it is with temperature. As relative humidity increases, so does electricity usage. High levels of humidity likely contribute to increased air conditioning usage, as dehumidification is often a secondary function of cooling systems. The spread of data indicates that there is variation in electricity consumption in different humidity levels, which may indicate that other variables such as household insulation and behavioral patterns are operating.

Figure 8: Research question 3 Visualization



The "Wind Speed vs. Electricity Usage" scatter plot illustrates that wind speed has a weaker and less consistent influence on electricity consumption compared to temperature and humidity. Whereas a few higher consumption values can be seen at moderate wind speeds, the overall distribution indicates very little correlation. This agrees with the intuition that wind speed indirectly affects energy demand, for example, by its impact on outdoor cooling efficiency or ventilation rather than directly.

Figure 9: Research question 3 Visualization

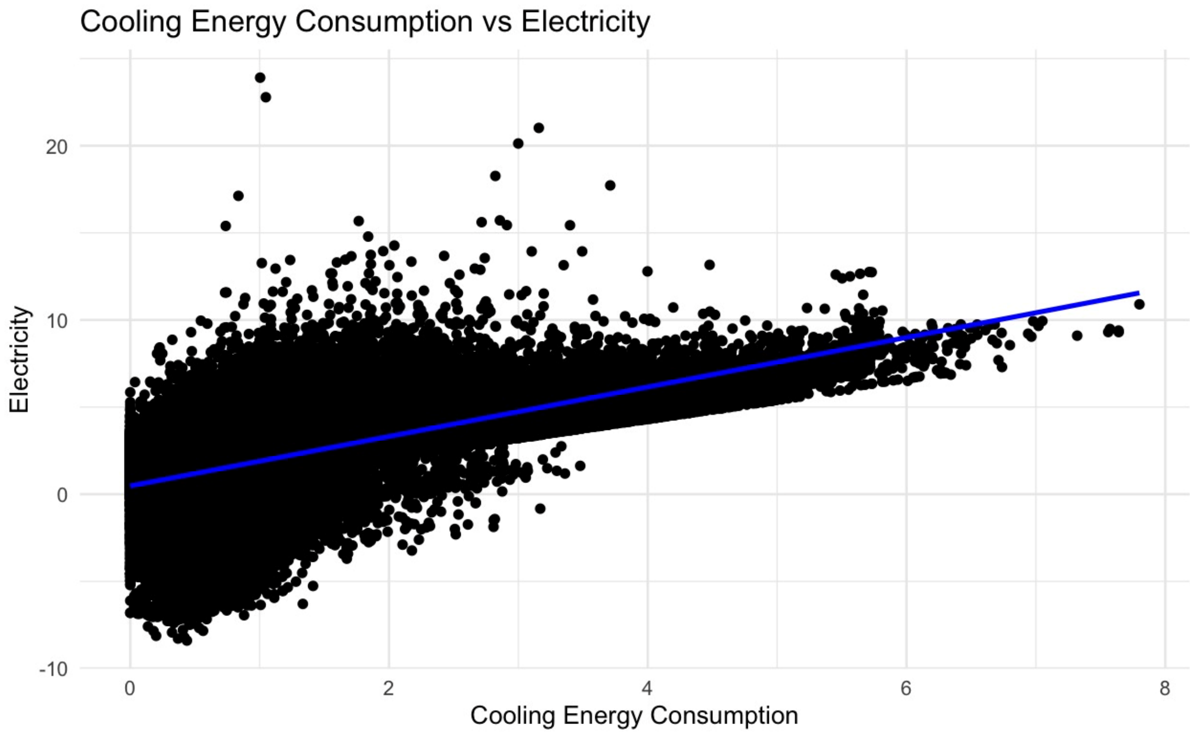


The correlation matrix plot confirms this from the strength of the relationship between the weather variables and the consumption of electricity. The temperature has the highest positive relation with electricity consumption, followed by humidity, whereas wind speed is a bit weaker. Further, the matrix shows an inverse relation between wind speed and relative humidity, reflecting the natural dynamics of weather systems.

Table 4: Results for research question 4

Metric	**Linear Regression**	**XGBoost**
MSE	0.2681	0.2583
MAE	0.1045	-
R^2	0.8901	-

Figure 10: Research question 4 Visualization



This graph above represents a positive correlation between cooling energy consumption and electricity consumption. When cooling energy consumption increases, the same happens to electricity consumption. Therefore, some scopes for energy efficiency are there that could be improved with some kind of targeted intervention, which will surely reduce electricity consumption.

Figure 11: Research question 4 Visualization

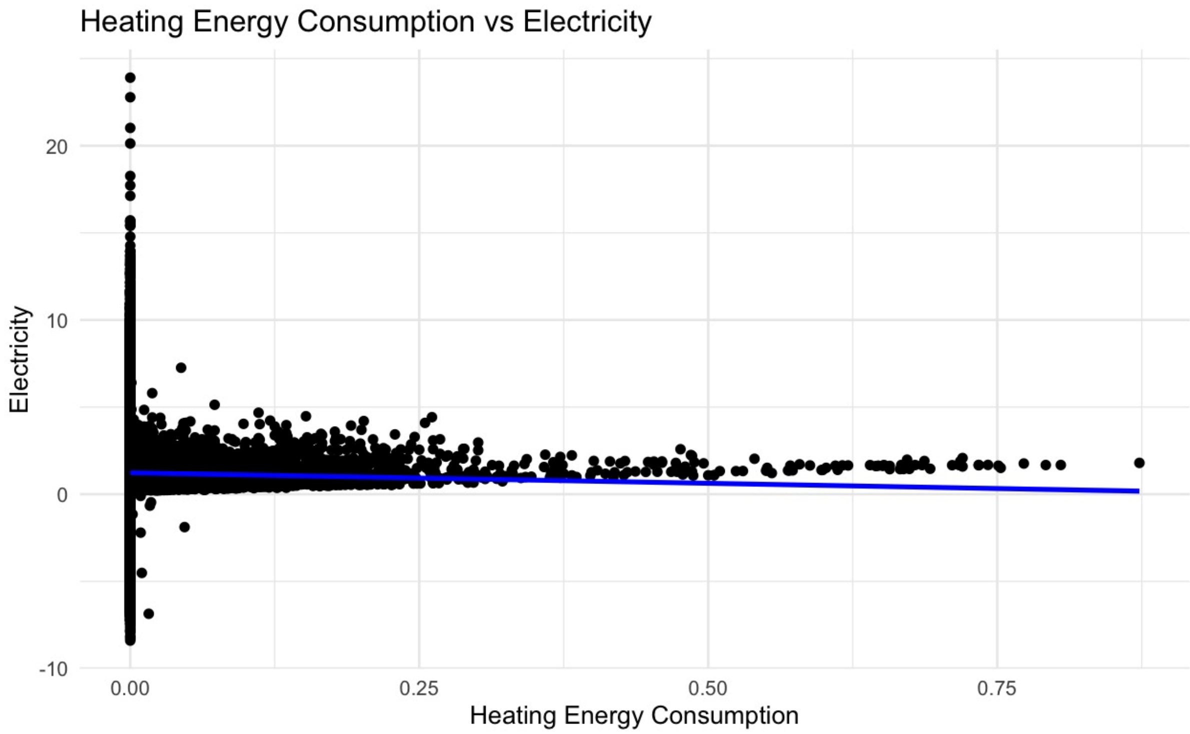
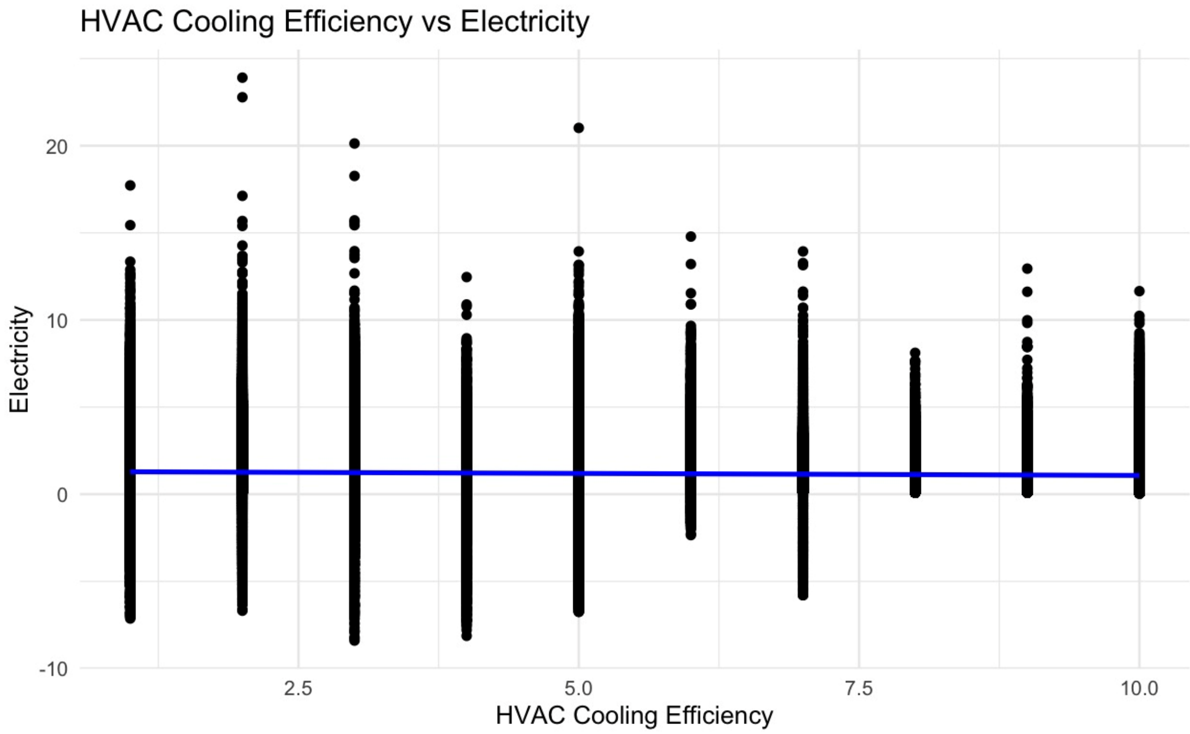


Figure 11 shows a very weak negative correlation of heating energy consumption and electricity use. With the rise in heating energy needs, there is a slight reduction in electricity consumption, though the trend is not that clearly defined. Reasons may include increased insulation in buildings, weather modifications, adaptation to renewable sources, economic factors, and efficiency gains from heating technologies.

Figure 12: Research question 4 Visualization

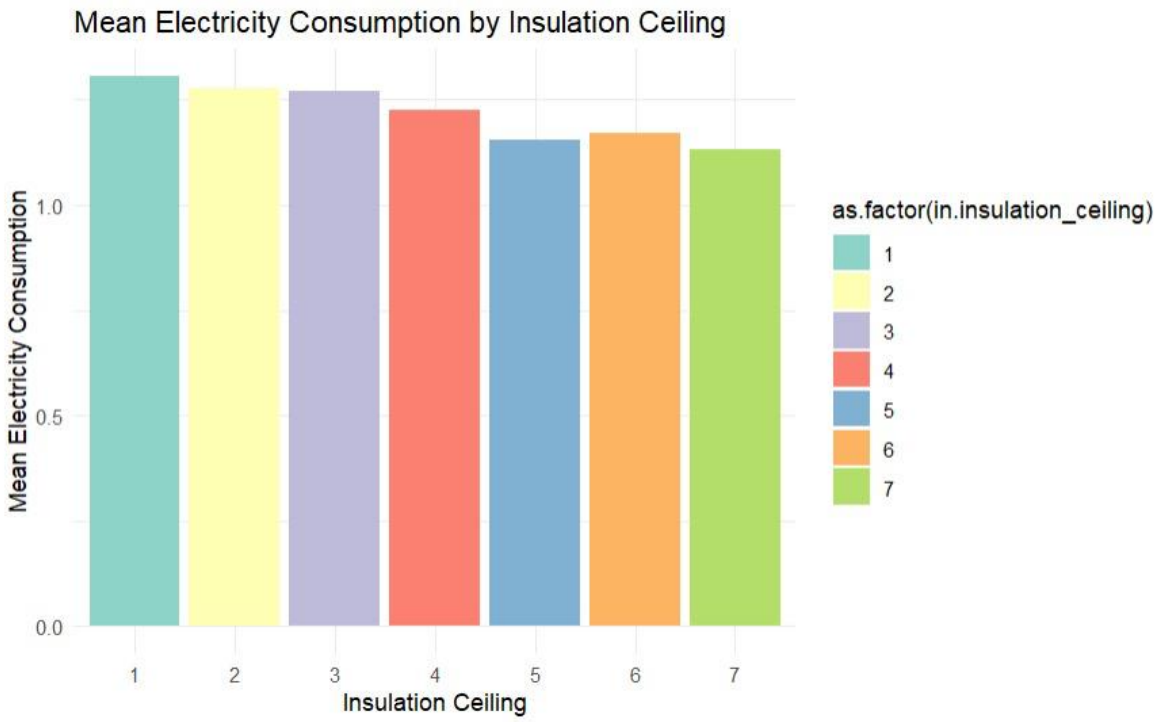


The scatter plot above depicts the variation in HVAC cooling efficiency versus electricity consumption. There is no obvious linear relationship, with the data points scattered throughout without a consistent trend in this scatter plot. These cluster around a few points of efficiency, suggesting the predominance of certain thresholds of efficiency or technologies. But on each of these levels of efficiency, the range of the values of electricity consumption is wide, which means factors other than efficiency, like building size, occupancy, and patterns of use, are dominant in electricity consumption.

Table 5 : Results for final model

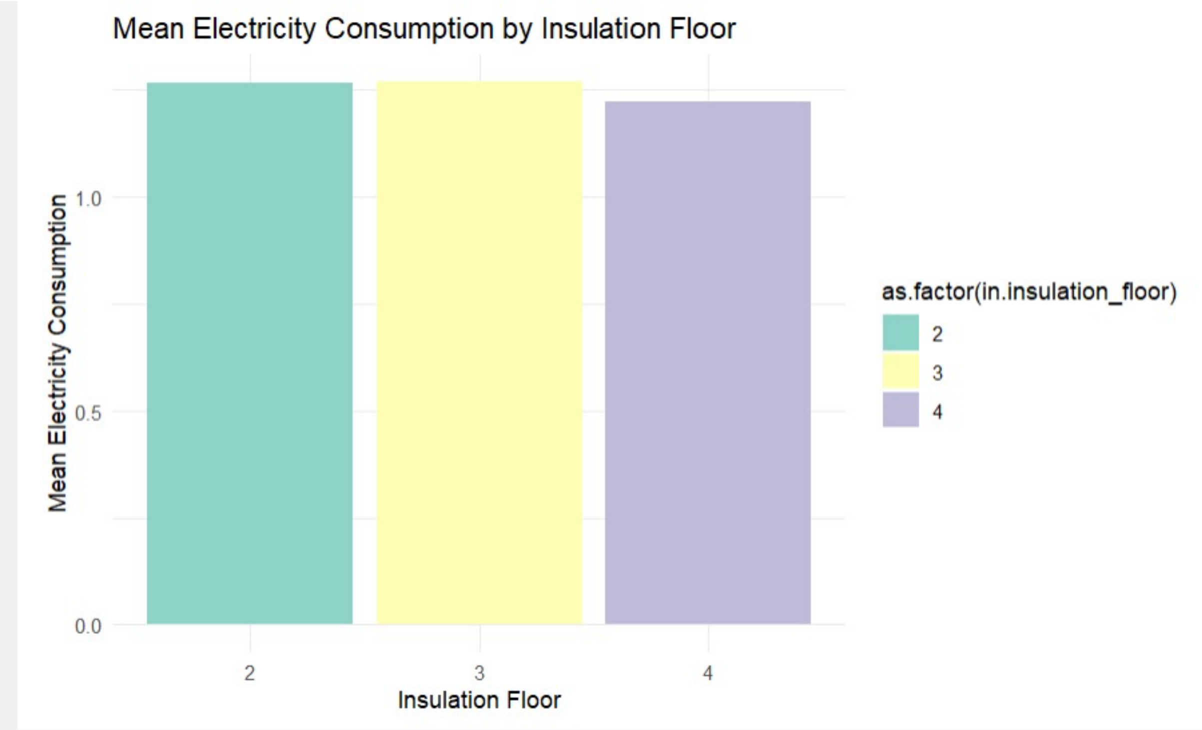
Metric	**Linear Regression**	**XGBoost**
MSE	0.3475	-
MAE	0.3675	-
R^2	0.4688	-

Figure 13: Final model Visualization



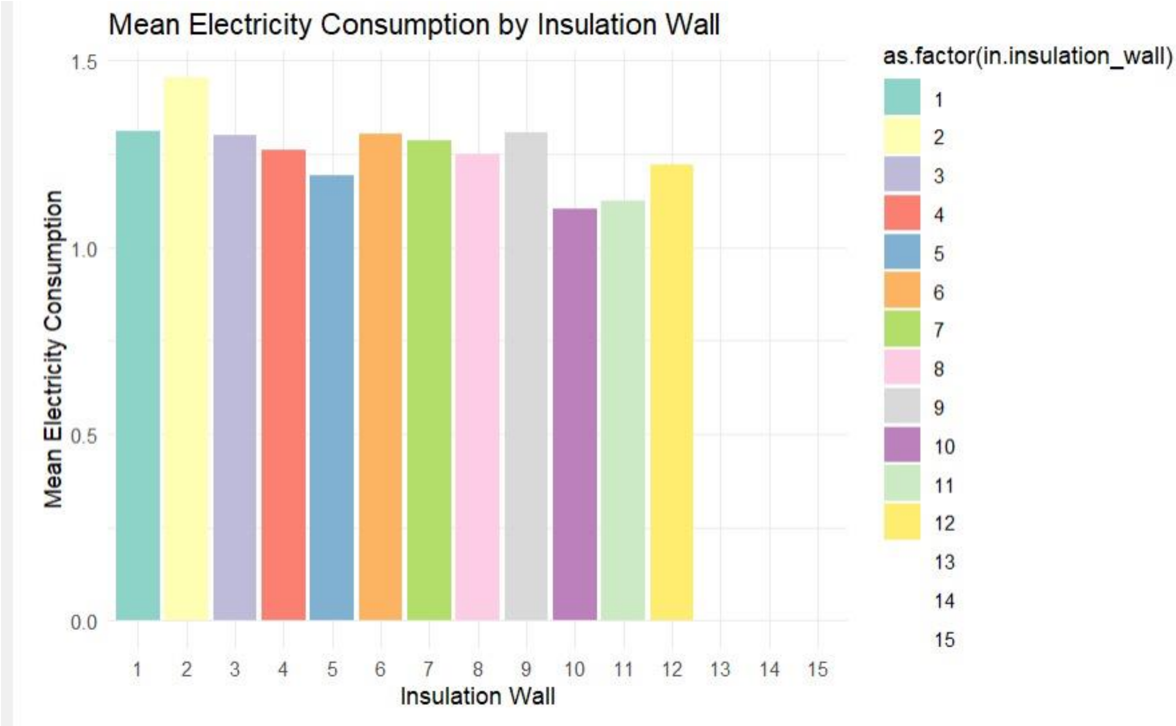
This bar chart highlights the relationship between different ceiling insulation levels and mean electricity consumption. The data suggests that better ceiling insulation (higher insulation categories) correlates with slightly reduced electricity consumption. This trend indicates the effectiveness of improved insulation in minimizing cooling and heating energy requirements. However, the difference across insulation categories is marginal, suggesting that other factors, such as building design or appliance efficiency, may also play a significant role in determining overall electricity usage.

Figure 14: Final model Visualization



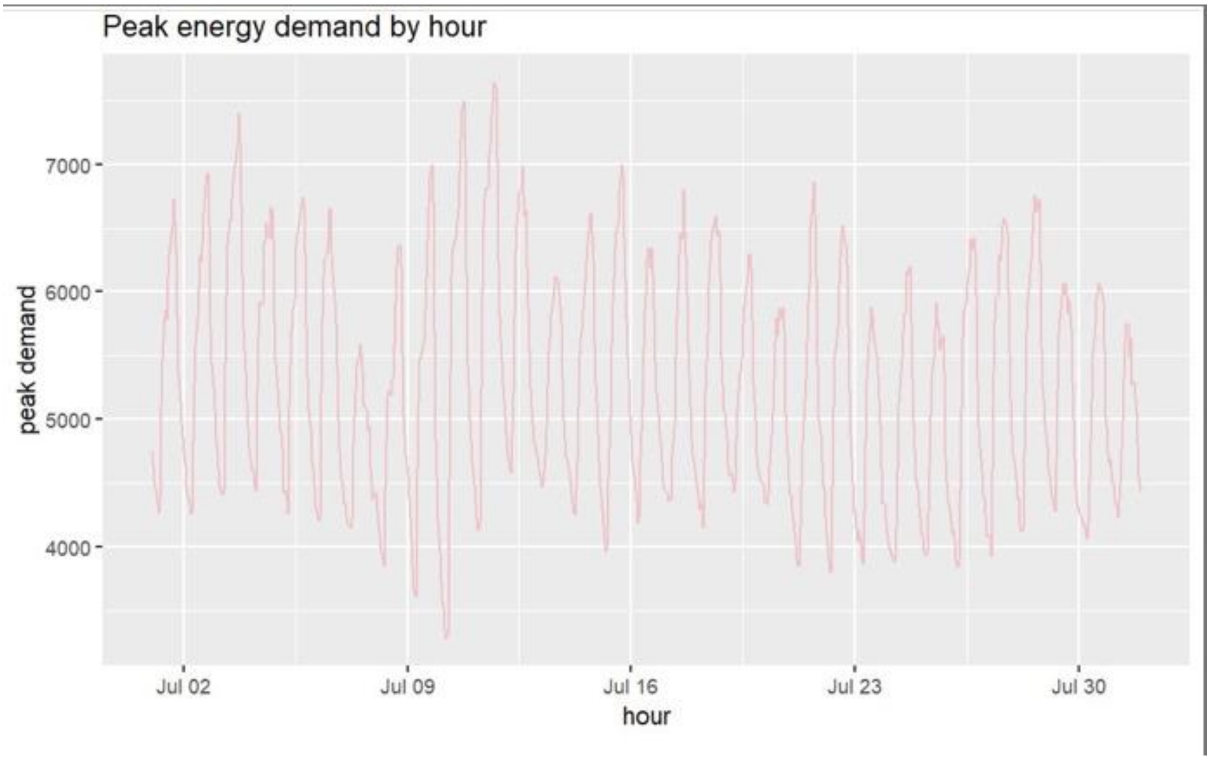
The second bar chart showcases electricity consumption based on floor insulation levels. The variation between different insulation categories is minimal, reflecting a less pronounced impact of floor insulation on energy consumption compared to ceiling insulation. This could imply that while floor insulation contributes to energy efficiency, its influence is relatively limited compared to other forms of insulation or structural attributes. Further analysis might explore whether other factors, like climate or occupant behavior, overshadow the role of floor insulation.

Figure 15: Final model Visualization



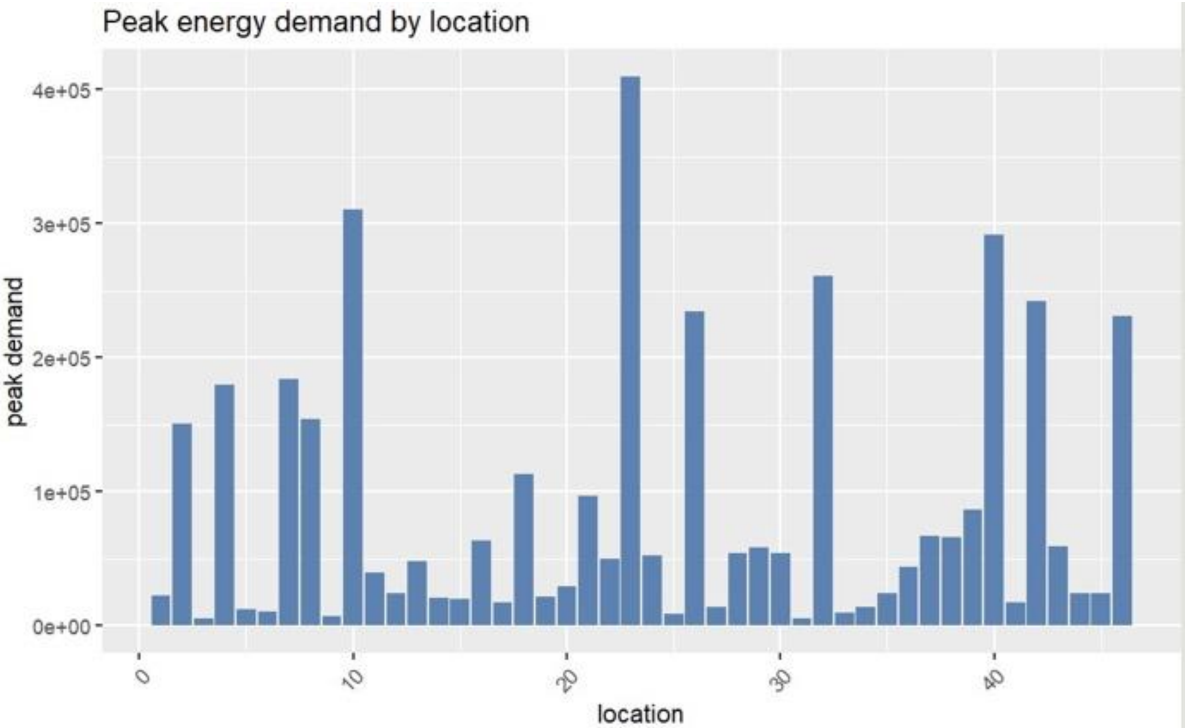
The third visualization explores how different wall insulation categories affect electricity consumption. The chart shows notable variation across insulation types, with higher-quality wall insulation typically linked to lower electricity usage. However, the fluctuations suggest that wall insulation alone cannot account for all variability in consumption, as other factors—such as the building's overall structure and regional climate—play a significant role. This reinforces the importance of adopting comprehensive insulation strategies that consider ceilings, floors, and walls collectively to optimize energy efficiency.

Figure 16: Final model Visualization



This graph above depicts the peak energy demand at different locations; hence, it shows strong variability. Some locations tend to have higher peak energy demands compared to other locations, which could be a function of population density, the character of the building stock, and local climate conditions. Sharp spikes suggest that the interventions, such as enhancing efficient energy infrastructures or demand-side management strategy, can be focused on very demanding locations for higher energy savings with enhanced grid reliability.

Figure 17: Final model Visualization



The chart above shows peak energy demand over the hour of the day for the month of July and is cyclical in nature. Peaks fall within daytime hours when the cooling systems are likely to be operating at their highest output given the temperature. This consistency of peaks points to potential dynamic pricing strategies or demand-response programs that might be effective in shifting energy usage to off-peak hours. A slight reduction of peak demand toward the end of the month may signal either seasonal change or behavioral adjustment, thus requiring a continuous reevaluation of energy policies. In sum, these will offer a wide insight into energy demand, thereby helping design appropriate and accurate data-driven solutions to solve energy demands sustainably.

Evaluation and Interpretation

The findings of this study give actionable insights from the study of electricity consumption across structural, behavioral, environmental, and energy-efficiency dimensions. Research Question 1 regarded the structural factors of electricity consumption. Significant variables in this analysis included in.geometry_floor_area and in.geometry_garage and in.ceiling_fan; all three had positive signs, suggesting that the more expansive the home with certain structural features, the greater the energy requirement. Conversely, the variables negatively associated included in.geometry_stories and in.insulation_slab, highlighting insulation and building design as key factors affecting energy use. These can inform policies in building design and auditing to ensure the incorporation of insulation and better layout to minimize energy use in residential buildings.

For Research Question 2, the highly significant variables included behavioral factors such as thermostat settings-in.cooling_setpoint, in.heating_setpoint-and number of occupants, in.occupants. Higher thermostat setpoints resulted in lower electricity use, whereas greater occupancy led to increased energy consumption due to more appliances and lights being on. These findings recommend clear actions in the areas of energy-saving campaigns that promote optimal thermostat settings and demonstrate the need for energy-efficient lighting and appliances in homes with higher occupancy levels. Utility companies could give incentives or rebates for smart thermostats that can ensure efficiency in consumption patterns.

Research Question 3 investigated environmental and climatic factors that affect the consumption of electricity. Significant factors included dry_bulb_temperature and relative_humidity, reflecting the demand for cooling and dehumidification when conditions are warm and humid. Although the overall R^2 value was moderate-0.1685-it underlines how important climatic factors are in the seasonal variation of energy demand. These results are actionable for energy providers, who can therefore predict peak load periods and strategize for demand-side management. Utilities may adopt dynamic pricing to make people conserve energy at peak hours during high-demand seasons.

Research Question 4 examined energy-efficient systems and had the highest R^2, at 0.8901. The strongest predictors were energy consumption for cooling, heating, and interior lighting: out.electricity.cooling.energy_consumption, out.electricity.heating.energy_consumption, and out.electricity.lighting_interior.energy_consumption, because these directly affect energy use. On the other hand, effective insulation was expressed through negative coefficients in in.insulation_ceiling and in.insulation_wall, which reduced electricity consumption. These findings translate into actionable policies: incentivizing the adoption of high-efficiency HVAC systems, promoting energy-efficient lighting, and subsidizing home insulation upgrades. Programs like these can substantially lower electricity demand while reducing environmental impacts.

The final model combined variables from all research questions and provides a comprehensive framework with an R^2 of 0.4688. Key contributors include in.geometry_floor_area, in.cooling_setpoint, in.hvac_cooling_efficiency, and out.electricity.cooling.energy_consumption, underlining the interplay between structural, behavioral, and environmental aspects of energy efficiency. In this model, taking action for integrated energy management allows stakeholders to devise multipronged strategies to reduce energy consumption. For instance, combining energy-efficient building designs with behavioral nudges-such as discounts on appliances that would lower the consumption of energy-and dynamic load management based on environmental conditions could thus address electricity demand holistically.

These results will, therefore, provide not just interesting observations but clear, actionable insights for policymakers, urban planners, and energy providers. Targeting significant variables across the four research areas allows for practical solutions to reduce electricity demand, lower costs, and promote sustainability. Improvement in building insulation, promoting efficient thermostat settings, preparing for peaks in energy demand seasonally, and incentivizing energy-efficient technologies will all go a long way toward developing a more sustainable energy ecosystem. It is important that this leads to much more specific interventions in energy-related challenges by stakeholders based on data evidence.

These findings indicate that during the warm July, the highest demand fell on 2018-07-11 at 5:00 PM EDT, thus yielding the most critical peak usage period. This follows from the pattern when energy demand in late afternoon tends to go high because of increased cooling as temperatures tend to peak at those hours of the day. The building with the highest single energy use was Building ID 433970, which reached a peak demand of approximately 4.0477 kW. On the whole, peak demand rose by 32.95% under normal conditions, indicative of the significant effect of the warmer-than-usual July on energy use. These insights put forth the need for targeted interventions through demand-response programs, energy-efficient technologies, and infrastructure upgrades, especially during critical hours and for high-demand users, to mitigate the effects of extreme temperatures on energy grids.

Validation

Validation of the results was a critical part of this study to ensure accuracy and reliability. To start, we split the data into training and testing sets to evaluate how well the models generalized to unseen data. This approach minimized the risk of overfitting and allowed us to assess the real-world applicability of the models. Metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and R^2 were calculated on the test dataset, serving as benchmarks to evaluate the models' predictive accuracy and explanatory power. Additionally, each model was carefully checked for errors by ensuring all variables were appropriately preprocessed, such as handling missing values and scaling where necessary.

For the linear regression models, residual analysis was performed to validate the assumptions of linearity, homoscedasticity, and normality of errors. This ensured that the models provided meaningful and unbiased predictions. Similarly, for the XGBoost models, hyperparameters were fine-tuned using cross-validation, and training logs were monitored to confirm convergence and stable performance. To further verify the correctness of the results, we compared the model outputs with domain expectations, ensuring that the coefficients and predictions aligned with logical and practical insights. For example, higher thermostat setpoints leading to lower electricity consumption aligned with prior research and energy efficiency principles.

Lastly, peer reviews and debugging sessions helped validate the code and results, ensuring there were no computational errors or logical inconsistencies. By incorporating these rigorous validation steps, we ensured that the results were robust, reliable, and capable of offering actionable insights for real-world application.

Shiny App

The Shiny app developed within this project provides an interactive interface that allows users to explore and dig into energy consumption predictions developed under various research questions and the full comprehensive model. This application would allow practitioners, such as energy analysts and policymakers, to choose which research question to consider-such as building attributes, behavioral patterns, weather conditions, energy efficiency measures, or the final model that summarizes all variables-and show the result in a very intuitive way. It allows users to select between Linear Regression and XGBoost models, showing very detailed outputs comprising model summaries, residual plots, feature importance charts, and correlation matrices. Another important feature of this app is its integration with precomputed results stored in an external file, which the app dynamically accesses to ensure responsiveness and efficiency on the users' side. This helps minimize computational load while enabling fast exploration of complex datasets.

Residual plots show model accuracy based on the differences between predicted and actual values. Feature importance charts in XGBoost point out major drivers like floor area, insulation, and HVAC efficiency that impact energy consumption. The correlation matrix lets one look into relationships among predictors to understand how variables influence energy consumption collectively. By integrating all of these capabilities, the Shiny app bridges the gap from technical analysis to actionable decisions. This provides a strong tool for stakeholders to obtain action items such as opportunities to optimize building design, incentivize energy-efficient technologies, or use behavior-based interventions. The app enhances the accessibility of results to the user for informed decision-making over the sustainability of energy management.

Figure 16 to 18 show the demonstration of the app and how it appears after the interactive application is constructed, where it focuses on interaction while showing each model and the respective visualization. The app can be utilized by eSC to understand what changes with each model and even interpret more insights from the application.

Figure 18: Shiny App Screenshot

Figure 19: Shiny App Screenshot

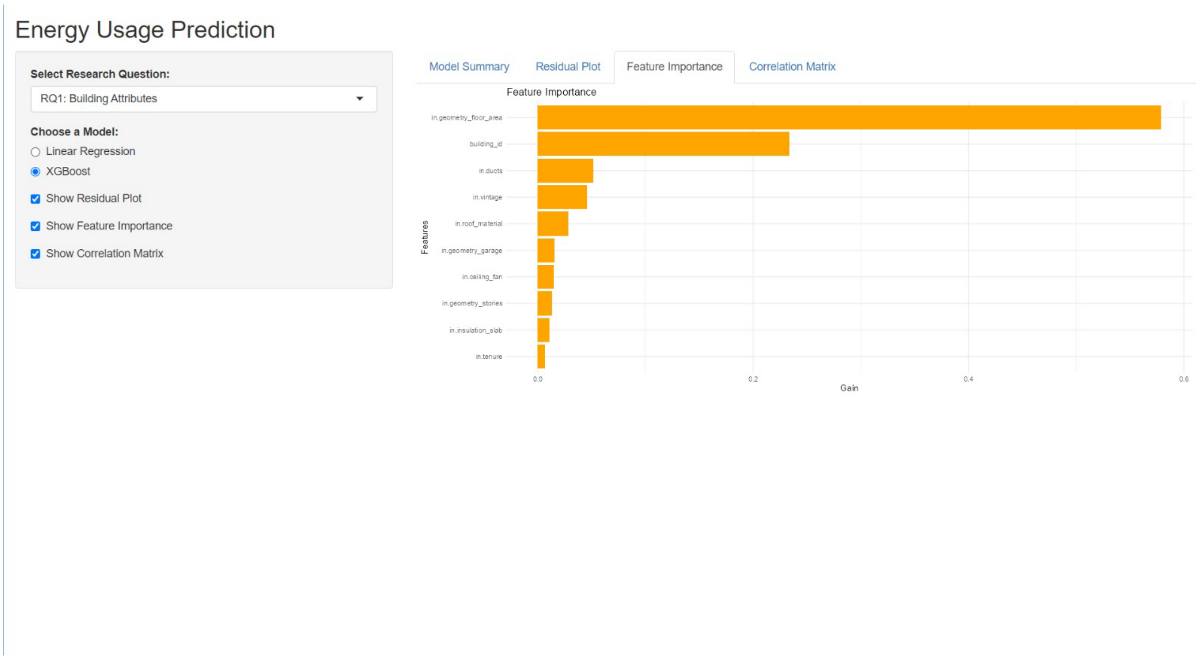
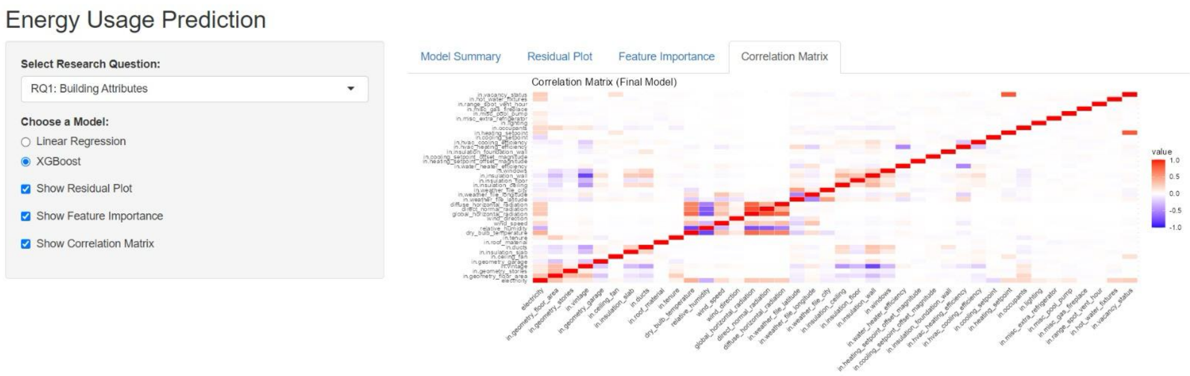


Figure 20: Shiny App Screenshot


Energy Usage Prediction
Select Research Question:
RQ1: Building Attributes
Choose a Model:
Linear Regression
XGBoost
Show Residual Plot
Show Feature Importance
Show Correlation Matrix
Model Summary
Residual Plot
Feature Importance
Correlation Matrix
Correlation Matrix (Final Model)
value
1.0
0.5
0.0
-0.5
-1.0
electricity
in.geometry_floor_area
in.geometry_stories
in.vintage
in.geometry_garage
in.ceiling_fan
in.insulation_slab
in.ducts
in.roof_material
in.tenure
dry_bulb_temperature
relative_humidity
wind_speed
wind_direction
global_horizontal_radiation
direct_normal_radiation
diffuse_horizontal_radiation
in.weather_file_latitude
in.weather_file_longitude
in.weather_file_city
in.insulation_ceiling
in.insulation_floor
in.insulation_wall
in.windows
in.water_heater_efficiency
in.heating_setpoint_offset_magnitude
in.cooling_setpoint_offset_magnitude
in.insulation_foundation_wall
in.hvac_heating_efficiency
in.hvac_cooling_efficiency
in.cooling_setpoint
in.heating_setpoint
in.occupants
in.lighting
in.misc_extra_refrigerator
in.misc_pool_pump
in.misc_gas_fireplace
in.range_spot_vent_hour
in.hot_water_fixtures
in.vacancy_status

Actionable Insights

The findings of this project show how electricity consumption can be optimized in residential buildings through different building characteristics, occupants' behaviors, weather conditions, and various energy efficiency measures. These are results that go beyond the merely academically interesting, since they may well be directly applied in reality to achieve real benefits both at individual household and general energy system levels.

A key takeaway from this research is the importance of focusing on energy-efficient building designs and retrofits. Other significant variables that contributed to explaining electricity use included insulation, ceiling fans, and the number of stories in a building. This could mean that better quality insulation and the facilitation of ceiling fans in residential units may reduce the need for heating and cooling systems that consume a lot of energy. Government-backed incentive programs or utility company rebates could help incentivize homeowners to invest in these upgrades, making energy-saving measures far more ubiquitous.

Behavioral patterns also have a significant role in energy consumption, and thermostat settings and the use of additional appliances such as pool pumps and refrigerators are very important. These results indicate the need for educational programs to help people act efficiently. For instance, utility companies could give workshops or online campaigns that clearly show that minor changes, like adjusting the thermostat setpoints, can bring in substantial energy savings. Adding smart home technologies, like programmable thermostats, could even automate some of the energy-saving behaviors of residents to make it easier without sacrificing comfort.

Energy consumption was thus influenced predictably by weather conditions, especially temperature and humidity. This will create opportunities for energy providers to use weather data to better plan energy supply and load forecasting, ensuring reliable electricity during peak demand periods. At the local scale, community programs can promote weather-specific energy-saving practices, such as the use of shading and natural ventilation during hot months or sealing drafts in colder seasons to enhance heating efficiency.

Energy efficiency measures, such as the use of high-efficiency appliances and insulation upgrades, showed great potential for reducing electricity consumption. Translating these findings into action, stakeholders like governments and energy companies can institute policies that encourage the adoption of energy-efficient technologies. Programs that offer financial incentives for energy-efficient appliances or retrofits would lower household energy bills and help meet larger goals of reducing carbon emissions and energy waste.

The final model's comprehensive nature-including building features, behaviors, weather, and efficiency measures-relates well to the holistic value of the approach in energy management. Solutions that address these interwoven factors at the same time will have maximum impact on energy savings. The idea is that urban planners and policymakers, armed with such insights, can plan smarter cities where building codes demand energy-efficient designs, community awareness programs lead to sustainable behaviors, and real-time energy monitoring systems make sure that resources are optimally allocated.

Ultimately, these findings give a direction for how to develop a more sustainable energy future. By using this knowledge, stakeholders can make informed decisions to decrease energy consumption, cut costs, and increase environmental sustainability for individuals and society as a whole.

References

Chandel, S.S., et al. “Review of Energy Efficient Features in Vernacular Architecture for Climate Adaptation in Hot and Dry Regions.” *Renewable and Sustainable Energy Reviews*, vol. 58, 2016, pp. 234-249.

Dimitriou, I., and C.T. Kiranoudis. “Energy Savings Through Real-Time Demand Response Programs in Residential Buildings.” *Applied Energy*, vol. 243, 2019, pp. 170-182.

Electrification of Heat Demonstration Project. *Reducing Residential Energy Demand through Improved Building Practices.* BEIS, UK, 2020.

Ghahramani, A., et al. “Smart Building Energy Management Systems for Occupant-Centric Applications.” *Energy and Buildings*, vol. 223, 2021, pp. 110099.

James, Gareth, et al. *An Introduction to Statistical Learning: With Applications in R*. Springer, 2013.

Kuhn, Max, and Kjell Johnson. *Applied Predictive Modeling*. Springer, 2013.

Liu, J., et al. “Energy Savings and Carbon Emission Reductions from Energy-Efficient HVAC Systems in Residential Buildings.” *Energy and Buildings*, vol. 167, 2018, pp. 216-224.

Luo, M., and T. Du. “Impact of Climate Change on Residential Electricity Demand: Insights for Energy Management.” *Climate Policy*, vol. 24, no. 3, 2024, pp. 450-465.

Reddy, K., and C. Assenza. “Challenges and Opportunities in Peak Load Management: A Review of Best Practices.” *Energy Policy*, vol. 142, 2020, pp. 111493.

Tibshirani, Robert, et al. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed., Springer, 2009.

U.S. Environmental Protection Agency. “Guide to Energy-Efficient Appliances.”

EnergyStar.gov, 2021, https://www.energystar.gov.

U.S. Energy Information Administration. "Energy Consumption Trends and the Role of Weather." *Annual Energy Outlook 2020*, 2020, https://www.eia.gov.

"Linear Models and Statistics in R." *R Documentation*, Comprehensive R Archive Network (CRAN), https://cran.r-project.org.

"XGBoost: A Scalable Machine Learning Library." *XGBoost Documentation*, https://xgboost.readthedocs.io.

"Caret Package Documentation." *Comprehensive R Archive Network (CRAN)*, https://cran.r-project.org/web/packages/caret/index.html.

"Data Table Package Documentation." *Comprehensive R Archive Network (CRAN)*, https://cran.r-project.org/web/packages/data.table/index.html.