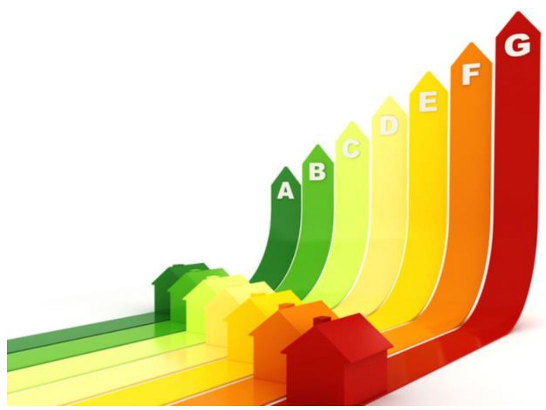
IST 687-Introduction to Data Science

Group –1

SC Energy Analysis



Submitted by:

Sai Subrahmanya Akhil Badampudi
Jash Nimesh Dharia Jaliwala
Masaki Hosohata
Sampada Regmi
Neha Sharma

TABLE OF CONTENTS

PROJECT DESCRIPTION

The project's objective is to address the challenges faced by an energy company (eSC) operating in South Carolina and a portion of North Carolina during the upcoming summer, particularly in the month of July. The company is anticipating an increase in demand for electricity due to rising temperatures associated with global warming. The concern is that this increased demand could potentially strain the electrical grid, leading to the risk of blackouts during exceptionally hot periods.

Rather than opting for a traditional approach of expanding their energy production infrastructure, which can be costly and may have environmental implications, the energy company is taking a proactive stance. The primary focus is on understanding the key factors driving energy consumption and developing strategies to encourage residential customers to adopt energy-saving practices.

By gaining insights into the drivers of energy consumption, the company aims to implement targeted measures that will help reduce overall demand on the electrical grid. This proactive approach not only ensures a reliable electricity supply during peak periods but also aligns with sustainability goals, reducing the environmental impact associated with increased energy production.

Strategies to encourage energy-saving practices among residential customers may include educational campaigns, incentives for adopting energy-efficient technologies, and promoting responsible energy consumption habits. By engaging with the community and raising awareness about the importance of energy conservation, the company aims to create a culture of responsible energy use.

Ultimately, the goal of the project is to mitigate the risk of blackouts, improve the reliability of the electricity supply, and contribute to environmental sustainability. By avoiding the need for additional power plants and focusing on demand-side management, the energy company aims to meet the growing energy needs of the community in a more sustainable and efficient manner.

PROJECT SCOPE

Objective

The primary objective of this project is to address the anticipated increase in electricity demand during the summer months, particularly in July, for an energy company (eSC) operating in South Carolina and parts of North Carolina. The project aims to mitigate the risk of blackouts caused by the strain on the electrical grid due to rising temperatures and increased energy consumption.

Scope Components

Data Analysis and Insight Generation:

- Analyzing historical energy consumption data to identify key drivers of increased usage.
- Understanding patterns and trends related to energy consumption, especially during peak demand periods.
- Assessing the impact of temperature rise on energy demand.

Demand-Side Management Strategies:

- Developing and implementing strategies to encourage energy-saving practices among residential customers.
- Identifying and promoting the adoption of energy-efficient technologies.
- Initiating educational campaigns to raise awareness about energy conservation.

Community Engagement and Behavioral Change:

- Engaging with local communities to foster a culture of responsible energy use.
- Implementing incentive programs to encourage energy-efficient behaviors.
- Monitoring the effectiveness of community engagement initiatives in reducing energy consumption.

Monitoring and Evaluation:

- Establishing metrics to measure the success of energy-saving strategies.
- Conducting regular assessments to ensure project goals are being met.

PROJECT DELIVERABLES

To effectively address the anticipated increase in electricity demand during the summer months for the energy company (eSC) in South Carolina and parts of North Carolina, our project will adopt a comprehensive and multi-faceted approach. Here's an outline of our planned activities:

Data Collection and Cleaning:

- Gather historical data on electricity consumption, weather patterns, and demographic variables.
- Conduct rigorous data cleaning to ensure the dataset is free from invalid or missing fields, ensuring reliability in the subsequent analysis.

Energy Consumption Analysis:

- Utilize linear regression models to identify key factors that significantly impact energy consumption, particularly during peak summer months.
- Analyze these factors in-depth to understand their contribution to the overall energy demand.

Predictive Modeling for Demand Forecasting:

- Implement advanced predictive models, such as Support Vector Machines (SVM), to forecast electricity demand for the upcoming summer.
- Generate actionable insights that can inform the energy company's planning and response strategies.

Shiny apps:

Our analysis setup is designed to give detailed information about electricity use in South Carolina, with a special feature that lets us focus on individual cities. Here's how it works in simpler terms:

-Choose Cities to Look At: We can pick specific cities within South Carolina to see how much electricity they're using. This is helpful if you want to understand what's going on in one particular place.

-See the Big Picture: Not only can we look at specific cities, but we can also see how much electricity the whole state of South Carolina is using.

-Individual City Details: Apart from the state-wide view, we can dive into details for each city. This means we can figure out the unique energy needs and usage patterns of each city, which can be quite different from one place to another.

DESCRIPTION OF DATA

2A. Static House Data:

Description:
The Static House Data consists of information regarding a random sample of single-family houses served by eSC (Energy Service Company). This dataset provides a comprehensive list of all houses included in the study. Each house is uniquely identified by a 'building id,' which serves as a key to access the corresponding energy usage data. The information includes various house attributes that remain constant over time, such as the size of the house.

Format:
The data is stored in 'parquet' format, an optimized storage format for tabular data.

Size:
The dataset comprises around 5,000 houses, each represented by a unique entry in the file.

2B. Energy Usage Data:

Description:
The Energy Usage Data is a collection of datasets, each corresponding to a specific house in the Static House Data. The dataset captures hourly energy usage for each house, providing calibrated and validated energy consumption data. The data encompasses 1-hour load profiles, detailing the usage of energy from various sources within the house (e.g., air conditioning system, dryer). The 'building ID' serves as the filename, uniquely identifying each house.

Format:
Similar to the Static House Data, each file in the Energy Usage Data is in 'parquet' format, optimizing storage efficiency.

Size:
There are approximately 5,000 individual datasets, each corresponding to a different house ('building ID')

2C. Meta Data:

Description:
The Meta Data file serves as a data description document, offering a human-readable guide to the fields used across the Static House Data and Energy Usage Data. It provides clarity on the attributes present in both datasets, aiding in the interpretation and understanding of the data.

Format:

This file is likely to be in a simple, human-readable format, such as plain text or a CSV File.

2D. Weather Data:

Description:
The Weather Data comprises hourly weather information, with one file dedicated to each geographic area or county. The data is timestamped and aligned with the hourly format of the other datasets. The 'in.county' column in the Static House Data links each house to its corresponding county code.

Format:

The Weather Data is stored in a simple CSV format, facilitating ease of access and interpretation.

Size:

There are approximately 50 weather files, each associated with a unique county code.

DATA PREPARATION

1. **City Filtering:**
Objective: Segmenting cities into "not in census place" and "in census place" categories.

Explanation: Here we have filtered cities based on whether they were classified as "not in census place" or "in census place." This categorization provided differences in energy consumption patterns between these two groups.

2. **Date Conversion:**
Objective: Converting the date column into a datetime format.

Explanation: The date information in the dataset was converted into a datetime column. This conversion ensures uniformity in handling date-related operations.

3. **Data Merging:**
Objective: Merging the Sample House Data (SHD) and Energy Data (ED) using the building data as a key.
Explanation: The Sample House Data and Energy Data were integrated by merging them through the building data. This enables a unified dataset that combines static house information with hourly energy consumption profiles.

4. **County-Level Aggregation:**
Objective: Aggregateing total humidity and temperature data at the county level.
Explanation: County-level data aggregation was performed on total humidity and temperature information. This aggregated data was then merged with the Energy and Housing datasets. This approach allows for an analysis of energy consumption patterns in relation to climatic conditions on a broader geographic scale.

5. **Data Integration:**
Objective: Integrating county-level data with individual house-level data.
Explanation: The aggregated county-level data was merged with the individual house-level data from the Sample House Data and Energy Data. This integration provides a comprehensive dataset that combines both micro and macro environmental factors, offering a view of the relationship between weather conditions and energy consumption.

6. **Grouping Entire Counties:**
Objective: Taking entire county data and group total humidity and temperature.

Explanation: The approach involves grouping entire counties' weather data and summarizing total humidity and temperature. This aggregated weather information is then linked back to the Energy and Housing datasets, creating a more manageable and informative dataset for analysis.

EXPLANATION OF DATA PREPARATION

- **Data Loading**:
  - Static Housing Data (**shd**), Building Electricity Data (**ed**), and Weather Data (**wd**) are loaded using the **arrow** and **tidyverse** libraries.
  - These datasets are initially explored using the **summary()** function and checked for missing values.

- **Data Cleaning and Pruning**:
  - Irrelevant columns in the Static Housing Data are removed to streamline the dataset.
  - The housing data is filtered to exclude non-relevant city entries, such as 'Not in a census Place' and 'In another census Place'.

- **Creation of Unique Lists**:
  - Lists of unique building IDs and counties are generated from the Static Housing Data for further processing.

- **Custom Functions for Data Filtering**:
  - Functions **julyfilter** and **julyfilter_wd** are created to filter energy and weather data specifically for the month of July.

- **Dynamic Data Loading in Loops**:
  - Loops are used to dynamically load and process Building Electricity Data and Weather Data for each building ID and county, respectively.

- **Data frame Construction for Energy and Weather Data**:
  - Separate dataframes for energy (**fullEDf**) and weather data (**fullwd_df** and **fullwd_inc_df**) are constructed by combining data from all relevant sources.
  - **fullwd_inc_df** is created by adding 5 degrees to the weather data to simulate increased temperatures.

- **Removal of Zero Consumption Columns**:

    - Columns in the energy data with zero total consumption are identified and removed to ensure data relevancy.

- **Merging Datasets**:
    - The Static Housing Data and Building Electricity Data are merged based on building IDs.
    - A new **date** column is created in the merged datasets for easier analysis.

- **Aggregation and Summarization**:
    - Total energy consumption is calculated for each row (representing per building, per hour) in the merged dataset.
    - A separate dataframe (**geo_merged_df**) is created to summarize energy consumption and group it by state, city, county, latitude, longitude, and date.

Weather Data Integration:

- Average temperature and humidity data are calculated and merged with the energy consumption data to create **merged_geo_avgwd_df** and **merged_geo_avgwd_inc_df**.
- These datasets provide a comprehensive view of energy consumption patterns along with corresponding weather conditions

In conclusion, this process effectively cleans, merges, and structures a complex set of housing, energy, and weather data, providing a robust foundation for subsequent analysis and modeling, particularly focusing on energy consumption patterns in South Carolina.

Exploratory Data Analysis

We made several visualization plots:

1) Daily Energy Consumption Trend: Using the ggplot2 library in R, we visualized the daily energy consumption trend. This line chart helps in understanding the fluctuations in energy usage over time, offering insights into daily patterns or anomalies.

Here is the detail:

X-Axis (Date):

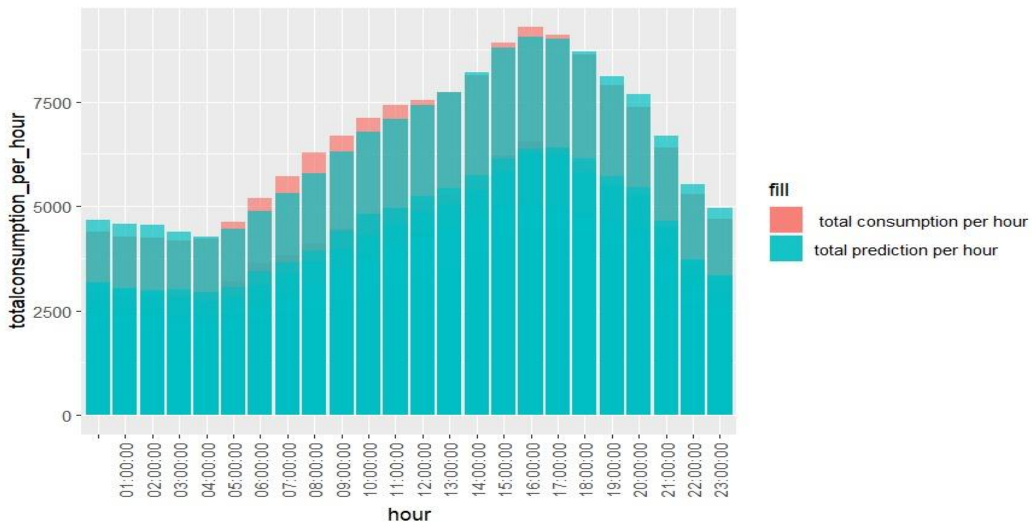
The x-axis of the chart represents time, specifically dates. Each point on the x-axis corresponds to a unique day in the dataset. This chronological arrangement allows us to follow the energy consumption trend over time, making it easy to spot any temporal patterns or anomalies.

Y-Axis (Daily Energy Consumption):

The y-axis indicates the total energy consumption for each day. It quantifies the amount of energy used and is likely measured in a standard unit of energy (like kilowatt-hours).The scale of the y-axis is determined based on the range of energy consumption values in the dataset, allowing for a detailed view of fluctuations in energy use.

Line Graph Representation: In this visualization, a line connects data points from consecutive days, illustrating how energy consumption rises or falls from one day to the next.

By observing the line's movement, one can identify patterns such as cyclical trends (like higher usage on specific days of the week), spikes (days with unusually high energy consumption), or troughs (days with particularly low usage).



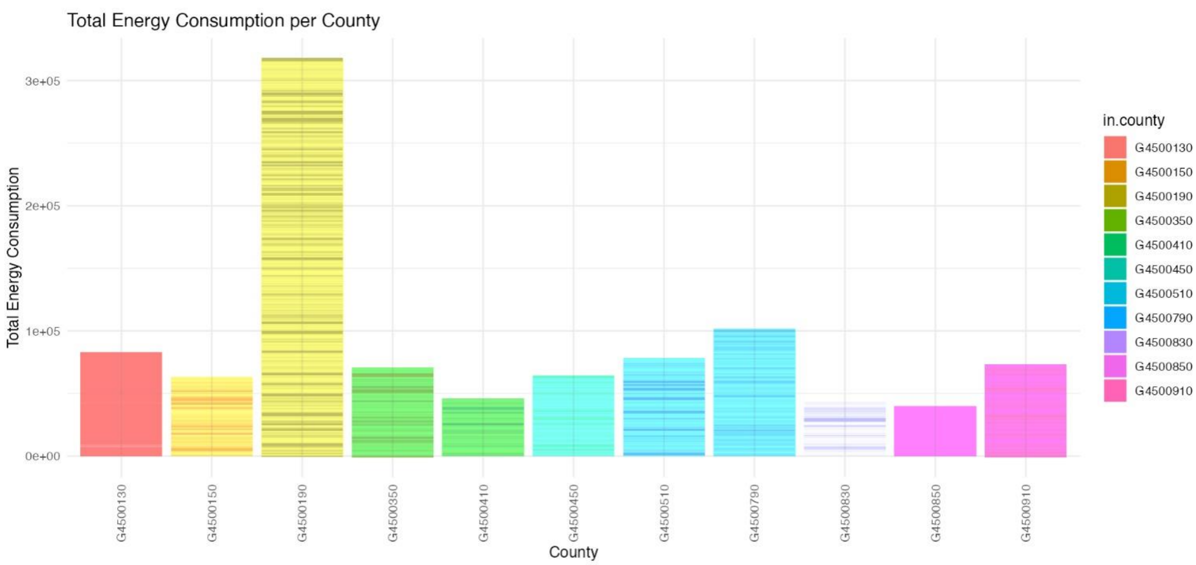
2)**County-wise Total Energy Consumption**: A bar chart was created to compare the total energy consumption across different counties. This visualization helps in identifying which counties have higher energy demands.

Here is the detail:

x = in.county: The x-axis represents different counties. Each bar in the chart will correspond to a specific county.

y = total consumption: The y-axis shows the total energy consumption for each county

Each bar's height corresponds to the total energy consumption of the respective county.



This visualization provides a clear comparison of total energy consumption across different counties. By representing each county as a separate bar, the chart allows for easy comparison of energy usage between counties, helping to identify which ones have higher or lower consumption levels.

Result:

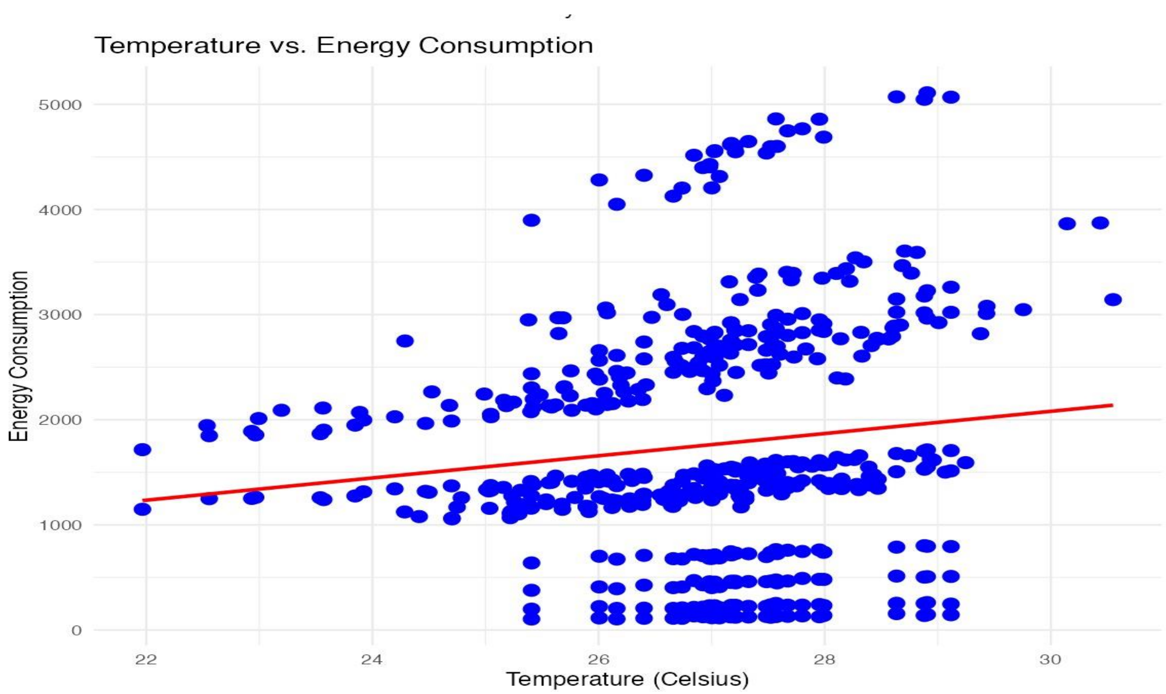
From the visualization bar graph, we can see that county G4500190 has the highest energy consumption and county G4500850 has the lowest energy consumption.

3) **Temperature vs. Energy Consumption**: A scatter plot with a regression line was used to explore the relationship between temperature and energy consumption. This analysis helps in understanding if higher temperatures relate with increased energy usage.

- Data Filtering: Here we have filtered the data2 dataset to include only the data from South Carolina (SC) and North Carolina (NC).
- Plotting: The ggplot function from the ggplot2 package is used to create a scatter plot. The aes function specifies the aesthetics of the plot, mapping average temperature (avg_temp) to the x-axis and total energy consumption to the y-axis.

Scatter Plot Points: This adds blue points to the scatter plot, each representing an observation in the filtered data set.

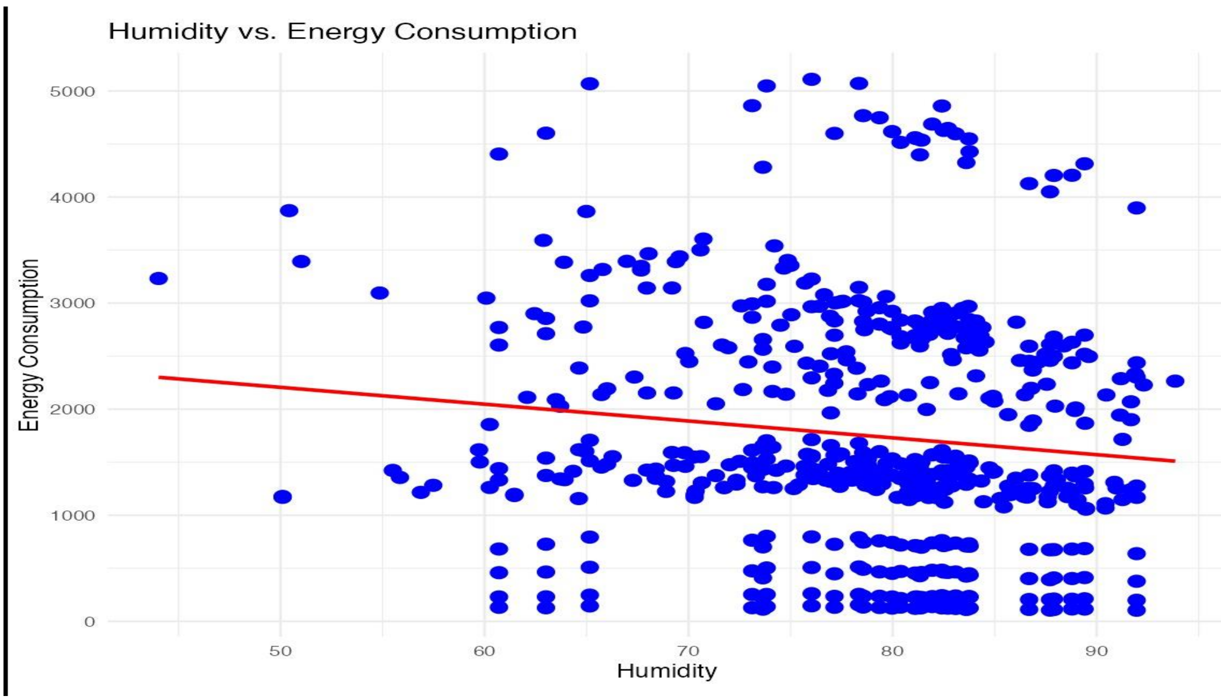
Linear Model: We have added a linear regression line to the plot, indicating the trend between temperature and energy consumption. The method = "lm" specifies that a linear model should be used, and se = FALSE means that the standard error bands around the line are not shown.



Result: The analysis gives us a positive correlation between temperature and energy consumption within the specified data set, focusing on South Carolina and North Carolina. This positive trend indicates that as temperatures rise, there is a corresponding increase in energy consumption in these regions.

4) **Humidity vs. Energy Consumption:** Like the temperature analysis, this scatter plot examines the relationship between humidity levels and energy consumption, identifying any correlation.

- Data Filtering: We filtered the dataset to include only data from SC and NC. to refine the dataset based on the specified states.
- Plotting Setup: The aes function defines the plot's aesthetics, assigning average humidity (avgHumidity) to the x-axis and total energy consumption (total consumption) to the y-axis.
- Scatter Plot –The scatter plot visually represents the relationship between humidity and energy consumption.



Result: The analysis conducted on the dataset reveals a negative correlation between humidity and energy consumption in South Carolina (SC) and North Carolina (NC). This negative correlation implies that as humidity levels increase, there is a tendency for energy consumption to decrease in these specific regions. It suggests that humidity plays a significant role in influencing energy consumption patterns in SC and NC

5) Percentage Distribution of Usage Levels by Climate Zone and City: This visualization used a bar chart with a facet wrap to show how energy usage levels vary by city and climate zone. It provides a detailed breakdown of energy consumption patterns across different geographical and climatic conditions.

-Data Plotting- The aesthetic function **aes** sets the x-axis to represent different cities which represent different categories or levels of usage.

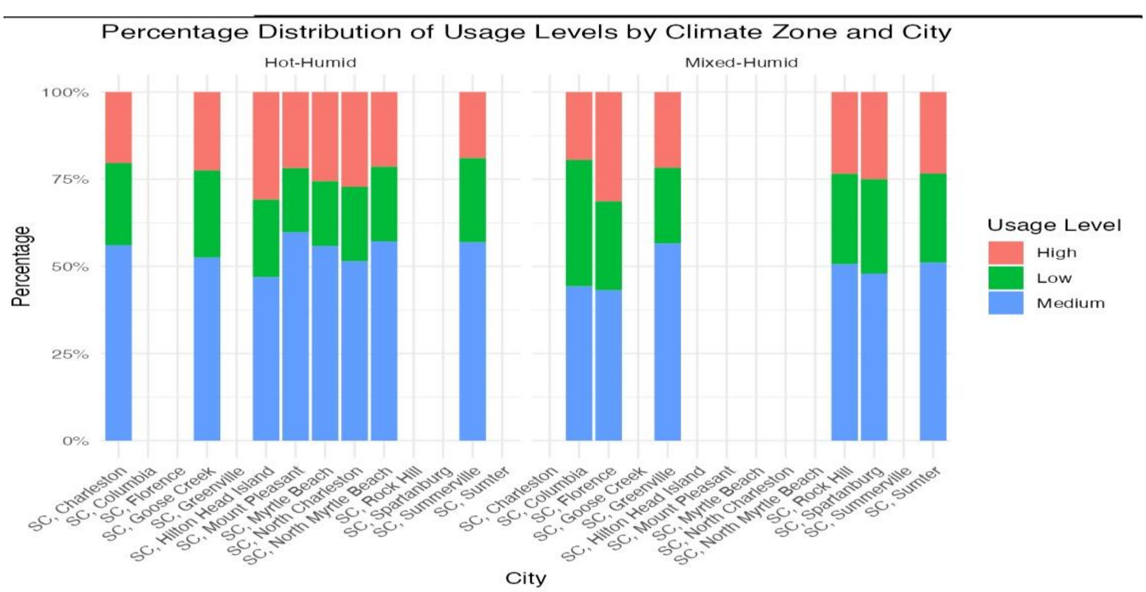
-Bar Chart –We have created a stacked bar chart. Each bar will represent the percentage distribution of usage levels within each city.

-Faceting - Each panel represents a different climate zone, allowing for a comparison across these zones.

-Labels and Titles (labs): The **labs** function adds a title to the chart and labels for the x-axis, y-axis, and the legend (fill). The title indicates that the chart displays the percentage distribution of usage levels by climate zone and city.

- **Y-Axis Scale: This** adjusts the y-axis to display values as percentages, making it easier to interpret the proportional data.

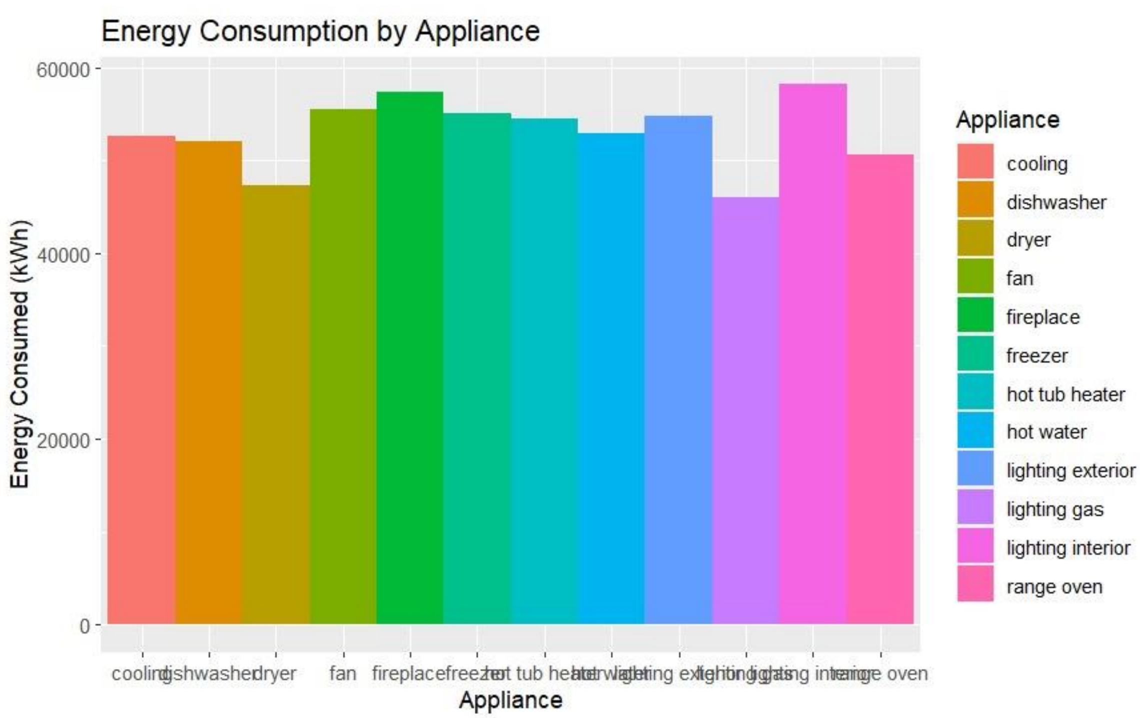
This visualization is particularly useful for understanding how usage levels vary across different cities within various climate zones, presented in a percentage format for straightforward comparison. The facetted design allows for an easy comparison across different climate zones, highlighting regional differences or similarities in usage patterns.



5)Energy Consumption by Appliances

The graph shows the energy consumption of different appliances by type, in kilowatt-hours (kWh). The appliances are arranged in a rainbow color scheme, which means that they are all different types of appliances. The most common type of appliance is a cooling appliance, which uses the most energy to cool food and air. The second most common type of appliance is a dishwasher, which uses the most energy to wash dishes. The third most common type of appliance is a dryer, which uses the most energy to dry clothes.

This visualization is particularly useful for understanding how much these appliances are consuming energy



MODELING

The analysis presented here involves a series of linear regression models created to understand the factors influencing total energy consumption in buildings. The final model (model5) was chosen after a careful examination of various factors and their significance in predicting energy consumption. Here's a brief explanation of why model5 was selected:

- **Iterative Refinement**: Initially, a comprehensive model including a wide range of variables was created. This model considered various factors such as square footage, number of bedrooms, number of stories, number of occupants, and energy consumption by different appliances and systems.

- **Variable Significance**: We figured out a key step in refining the model involved examining the significance of each variable. Variables that didn’t have any energy consumption in the first to third quartile were not used and taken out of the model(model3). These variables didn’t affect the model because most of the dataset didn’t have any energy consumption for those variables.

- **Simplification and Focus**: The final model (model5) was developed by simplifying the previous models and focusing on the most significant variables. It includes average temperature and sums of energy consumption by various appliances and systems. This model strikes a balance between complexity and explanatory power, making it more robust and reliable for prediction.

- **Predictive Power and Generalizability**: Model5 was chosen for its improved predictive power and generalizability. By focusing on key variables that affect energy consumption, the model can more accurately predict total energy consumption across different buildings and conditions.

- **Practical Relevance**: The selected variables in model5 (like average temperature, energy consumption by ceiling fans, cooling systems, freezers, etc.) are practically relevant. They represent common factors that are likely to vary across different buildings and directly influence energy consumption

- **Visualization and Application**: The model's utility is further demonstrated through visualizations in Parts F and G, where predictions of daily energy consumption and city-by-city energy consumption are presented. These visualizations help in understanding the practical implications of the model.

In summary, model5 was chosen for its balance of simplicity, accuracy, and relevance, making it a robust tool for understanding and predicting energy consumption in buildings

CODE EXPLANATION

This code is an R script designed for a comprehensive analysis of energy consumption data. It involves data processing, statistical modeling, and data visualization. The script is divided into several parts, each serving a specific purpose in the analysis:

Data Preparation:

- The tidyverse and car libraries are loaded for data manipulation and advanced regression modeling, respectively.
- The data is read into the R environment using read.csv.
- str(data) is used to display the structure of the dataset.

Initial Model Building (Part C):

Variables are extracted from the dataset to be used in linear regression models. The first model uses all variables that factor into energy consumption. We run a summary of the predictor variables and decide it isn't practical for predicting energy consumption.

We look at a summary of all the variables and examine that some variables have a value of 0 energy consumption in the first to third quartiles. We can take these variables out of our original model because they don't affect energy consumption.

We are left with model3 with the factors that don't affect energy consumption taken out. These include factors like square footage, number of bedrooms, stories, occupants, and various categories of energy consumption (e.g., ceiling fan, cooling, freezer).

An extensive linear regression model (model) is created with total consumption as the dependent variable and a wide range of independent variables.

After examining the summary of data, less significant variables (those with little variation in energy consumption) are removed, leading to a more refined model (model3).

Further Model Refinement:

model4 is created using a subset of variables from model3, focusing on those factors deemed most relevant to energy consumption.

Incorporating Additional Data:

New datasets (data2inc and data2) are read into R, likely including temperature data and other relevant information.

A final model (model5) is built, incorporating average temperature and sums of various types of energy consumption.

Prediction and Visualization (Part F and G):

Predictions are made using model5 and the new dataset (data2inc). These predictions are added to data2inc as a new column.

Visualization 1: A bar chart is created using ggplot2 to show predicted daily energy consumption over time. We used this to find which days had the highest peak usage in energy consumption.

Visualization 2: Another bar chart displays the predicted energy consumption by city. This barchart takes a look at the different energy consumption by region. By splitting the columns up by city, we can see which cities use a higher consumption vs a lower consumption.

Visualization 3: A scatter plot with a linear regression line is used to explore the relationship between average temperature and total energy consumption. We added a line of best fit to show the general relationship between the two variables is positive.

Summary: This describes how the analysis started with a comprehensive model considering various potential predictors of energy consumption. It should then explain how the model was refined by removing less significant variables, resulting in a more focused model. The integration of additional data, particularly temperature, should be highlighted as a key step in enhancing the model's predictive power. Finally, the report should discuss how visualizations were used to present the predictions and explore the relationships between key variables and energy consumption.

```
> summary(model4)

Call:
lm(formula = totalconsumption ~ sqft + bedrooms + stories + occupants +
    ceiling_fan + cooling + freezer + hot_water + lighting_exterior +
    lighting_interior + plug + refrigerator, data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-9.5291 -0.1394 -0.0843 -0.0133 13.4741

Coefficients:
                    Estimate Std. Error  t value Pr(>|t|)
(Intercept)        3.268e-02  2.446e-03   13.363   <2e-16 ***
sqft              -2.149e-05  3.900e-07  -55.092   <2e-16 ***
bedrooms           7.139e-03  6.469e-04   11.035   <2e-16 ***
stories           -8.343e-03  9.238e-04   -9.032   <2e-16 ***
occupants          8.294e-03  3.382e-04   24.525   <2e-16 ***
ceiling_fan        1.799e+00  9.545e-02   18.846   <2e-16 ***
cooling            1.188e+00  1.090e-03 1089.508   <2e-16 ***
freezer            1.361e+00  2.259e-02   60.256   <2e-16 ***
hot_water          1.184e+00  4.512e-03  262.322   <2e-16 ***
lighting_exterior -2.341e+00  7.237e-02  -32.351   <2e-16 ***
lighting_interior  1.353e+00  4.433e-03  305.193   <2e-16 ***
plug               1.073e+00  2.613e-03  410.725   <2e-16 ***
refrigerator       9.871e-01  9.892e-03   99.792   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3925 on 758867 degrees of freedom
Multiple R-squared:  0.778,     Adjusted R-squared:  0.778
F-statistic: 2.217e+05 on 12 and 758867 DF,  p-value: < 2.2e-16
```

Result for 1st model:

Statistical Significance:

All the predictor variables included in the model are statistically significant, as indicated by their p-values being less than the conventional alpha level of 0.05 (in fact, they are all far less than 0.01).

This suggests that variables like square footage, number of bedrooms, stories, occupants, ceiling fan, cooling, freezer, hot water, lighting exterior, lighting interior, plug, and refrigerator have a statistically significant relationship with total energy consumption.

Accuracy of the Model:

The Multiple R-squared value is 0.778, indicating approximately 77.8% of the variability in the dependent variable (total consumption).

The Adjusted R-squared value is also 0.778, which adjusts the R-squared value for the number of predictors in the model relative to the number of observations. Since it is virtually the same as the Multiple R-squared, it indicates that the number of predictors is appropriate for the number of observations and that the model isn't overly complex.

The F-statistic is about 2.217e+05, , with a p-value of less than 2.2e-16, indicating that the model as a whole has a statistically significant joint effect on the predicted variable.

```
> summary(model5)

Call:
lm(formula = totalconsumption ~ avg_temp + sum_ceiling_fan + 
    sum_cooling + sum_freezer + sum_hot_water + sum_lighting_exterior + 
    sum_lighting_interior + sum_plug_loads + sum_refrigerator, 
    data = data2)

Residuals:
   Min     1Q Median     3Q    Max 
-95.58 -25.48   3.43  20.53 102.34 

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)    
(Intercept)          -39.92566   61.59678  -0.648 0.517142    
avg_temp               1.43572    2.26057   0.635 0.525619    
sum_ceiling_fan       14.33815    1.56182   9.180  < 2e-16 ***
sum_cooling            1.01590    0.01662  61.122  < 2e-16 ***
sum_freezer            6.70968    0.86012   7.801 3.13e-14 ***
sum_hot_water          0.74427    0.35427   2.101 0.036112 *  
sum_lighting_exterior  9.62700    2.72539   3.532 0.000447 ***
sum_lighting_interior  1.00754    0.16291   6.185 1.22e-09 ***
sum_plug_loads         0.80591    0.07276  11.076  < 2e-16 ***
sum_refrigerator       0.79514    0.20521   3.875 0.000120 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 35.15 on 548 degrees of freedom
Multiple R-squared:  0.999,	Adjusted R-squared:  0.999 
F-statistic: 6.365e+04 on 9 and 548 DF,  p-value: < 2.2e-16
```

Result for 2nd model:

Based on the p-values, Heres the detail explanation of the result:

Statistical Significance: Several predictors show a statistically significant relationship with the dependent variable (total consumption). A p-value less than 0.05 is commonly significant.

- sum_ceiling_fan,sum_cooling,sum_freezer,sum_lighting_exterior, sum_lighting_interior, and sum_plug_loads have p-values significantly less than 0.05, indicating strong evidence against the null hypothesis and suggesting that these factors have a statistically significant impact on the total consumption.
- sum_hot_water and sum_refrigerator have p-values slightly above 0.01 but well below 0.05, which also indicates statistical significance.
- avg_temp has a p-value of 0.525619, which is not statistically significant, suggesting that average temperature is not a significant predictor of total consumption within the scope of this model.

Accuracy of the Model:

The Multiple R-squared value is 0.99, which suggests that 99 % of the variability in the dependent variable can be explained by the model.

The Adjusted R-squared is also 0.999, which takes into account the number of predictors in the model relative to the number of observations,

The F-statistic is a measure of the total variation explained by the model relative to the variation unexplained. Here, the F-statistic is approximately 63650 with a p-value of less than 2.2e-16, which means the model as a whole is statistically significant.

SHINY APPS

We used 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.

User Interface (UI):

1. Dashboard page includes a header with the title "SC Energy Analysis" and a sidebar with an introductory note on the app's primary objective.
2. The sidebar also includes a dropdown menu (selectInput) allowing us to filter the displayed data by city.
3. The dashboard body contains four box elements, each destined to display different outputs:
4. Two boxes are dedicated to textual outputs: one for "Total Energy Consumption" and another for "Total Future Energy Consumption", each showing the respective values for the selected city, or all cities combined.
5. The remaining four boxes are placeholders for plots (plotOutput), which will visualize various aspects of the data through interactive charts.

Output for total current and future consumption is calculated and rendered as text in the designated boxes.

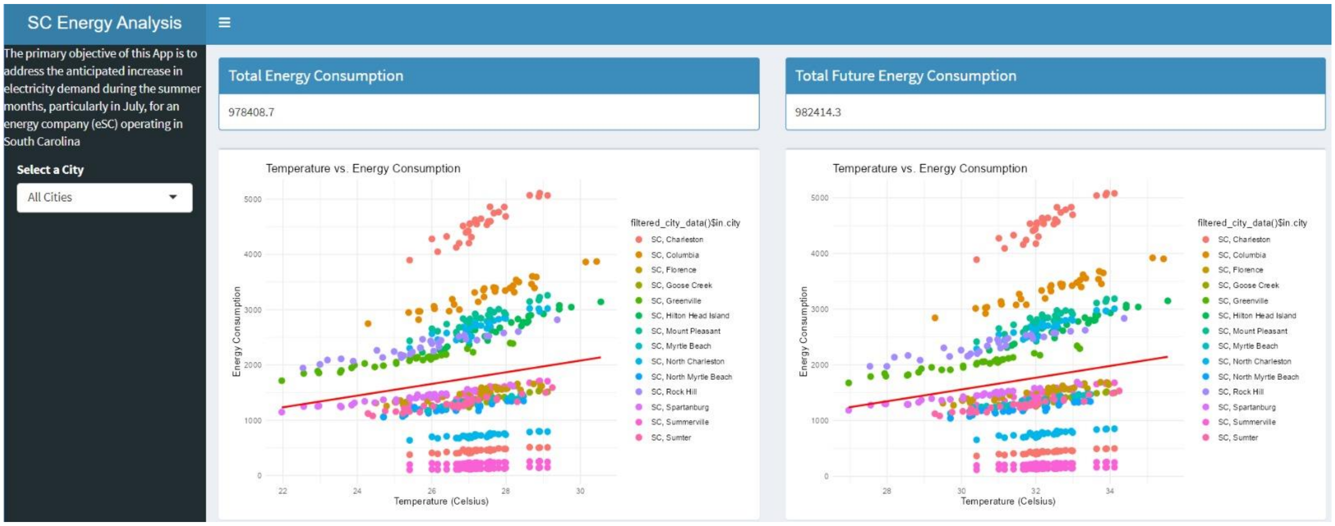
Four plots are generated to visualize the data:

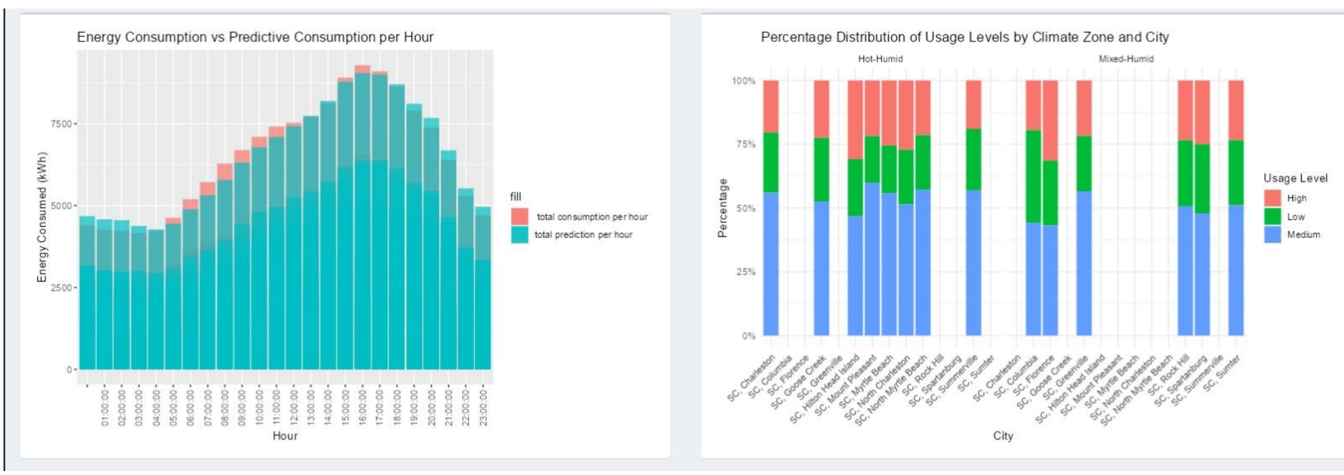
plot1 - Shows the percentage distribution of usage levels by climate zone and city.

plot2 - Displays a scatter plot of temperature versus energy consumption for the selected city or cities.

plot3 - Like plot2, but likely to display predictions of future energy consumption against temperature.

plot4 - Compares actual energy consumption per hour with predictive consumption, providing insights into hourly demand patterns.





https://sbadampuakhil.shinyapps.io/IDSShiney1/

\

Strategies for Peak Energy Demand Management: Optimizing Ceiling Fans, Exterior Lighting, and Freezers through Demand Response Programs

Reducing peak energy demand is a critical goal for energy providers and consumers alike. Peak demand often leads to higher energy costs, increased strain on the grid, and a greater likelihood of blackouts or brownouts. To address this issue, one potential approach is to implement demand response programs.

Understanding Peak Energy Demand

Peak energy demand refers to the periods when electricity consumption is at its highest during the day or year. These peaks are often predictable, occurring during specific times such as hot summer afternoons or cold winter evenings when people use energy-intensive appliances and systems to maintain comfort or carry out daily activities.

The Role of Ceiling Fans, Exterior Lights, and Freezers

Ceiling Fans: Ceiling fans are a common household appliance used for cooling purposes. While they consume less electricity compared to air conditioners, they still contribute to peak energy demand, especially in regions with hot and humid climates. During the summer, many people rely on ceiling fans to maintain comfort, resulting in an increased load on the grid during peak hours.

Exterior Lights: Exterior lighting serves multiple purposes, including safety and security. However, many outdoor lighting systems are not equipped with energy-efficient technologies, leading to unnecessary energy consumption. Exterior lights left on overnight or during daylight hours can significantly contribute to peak demand.

Freezers: Freezers are essential appliances in households and commercial establishments for preserving food. These appliances run continuously and consume a considerable amount of energy, making them a significant contributor to peak demand, particularly during the summer when cooling requirements are higher.

Reducing Peak Energy Demand through Demand Response

Demand response programs are designed to address peak energy demand by encouraging consumers to reduce their electricity consumption during peak periods. Here's how these programs work and how they can be applied to the top three variables contributing to energy consumption:

1. *Ceiling Fans:*

- Smart Ceiling Fan Control: Installing smart ceiling fan controls can enable users to schedule fan operation based on their preferences. These devices can also be integrated into demand response programs, allowing fans to automatically adjust their speed or turn off during peak hours.

- Incentives and Rewards: Energy providers can offer incentives, such as bill credits or discounts, to customers who voluntarily reduce their ceiling fan usage during peak demand periods.

2. *Exterior Lights:*

- Motion Sensors and Timers: Installing motion sensors and timers for outdoor lighting can ensure that lights are only active when needed. This reduces unnecessary energy consumption and aligns with demand response efforts.

- Rebate Programs: Energy companies can offer rebates or discounts to customers who upgrade their exterior lighting to energy-efficient LED fixtures, which consume less energy and have longer lifespans.

3. *Freezers:*

- Load Shifting: Load shifting involves adjusting the operation of appliances like freezers to off-peak hours when electricity demand is lower. Smart freezer controls can enable this feature, allowing freezers to consume less energy during peak periods.

- Peak Time Pricing: Energy providers can implement peak time pricing, where electricity costs more during peak hours. This incentivizes consumers to use their freezers during off-peak times, reducing overall peak demand.

4. *Fire place*

Installing programmable thermostats to better control the heating in home. This way, we can set the temperature to be lower when we are not home or during the night, reducing the need to use the fireplace

Optimizing Energy Efficiency: A Comprehensive Guide to Modeling Demand Response Impact on Peak Energy Usage

Modeling the impact of implementing demand response programs to reduce peak energy demand, particularly with a focus on ceiling fans, exterior lights, and freezers, requires a structured approach that considers various factors and quantifies the potential benefits. Here's a suggested method for modeling and explaining the impact:

1. Data Collection and Baseline Analysis:

- Start by collecting historical data on energy consumption for the target appliances (ceiling fans, exterior lights, freezers) in a specific region or area.

- Identify the peak demand periods based on this historical data.

2. Demand Response Program Implementation:

- Introduce demand response programs that encourage consumers to reduce usage during peak hours for the selected appliances.

- Implement technologies like smart controls, motion sensors, timers, and peak time pricing as mentioned in the previous article.

3. Monitoring and Measurement:

- Continuously monitor energy consumption patterns for the appliances before and after the implementation of demand response measures.

- Collect data on consumer participation rates in the demand response programs.

4. Quantifying Impact:

- Calculate the reduction in energy consumption during peak periods for each appliance category. This can be expressed in kilowatt-hours (kWh) or as a percentage decrease compared to the baseline consumption.

- Evaluate the demand reduction's impact on the electrical grid by assessing the reduction in peak load, measured in kilowatts (kW).

- Determine the cost savings for consumers who participated in the demand response programs.

5. Economic Analysis:

- Estimate the economic benefits of demand response, considering factors like reduced energy costs for consumers, potential grid infrastructure savings, and the cost of implementing the demand response programs.

- Calculate the return on investment (ROI) for energy providers and consumers participating in the programs.

Explaining the Impact:

When explaining the impact of implementing demand response programs, it's crucial to communicate the key findings and benefits clearly to various stakeholders:

1. Consumers: Emphasize the cost savings and convenience of participating in demand response programs. Highlight how their actions contribute to grid reliability and reduce their own energy bills.

2. Energy Providers: Showcase the reduced strain on the grid during peak hours, which can lead to lower infrastructure costs and improved overall grid performance. Explain how demand response programs can enhance customer satisfaction and loyalty.

3. Environmental Benefits: Describe the reduction in greenhouse gas emissions resulting from decreased energy consumption during peak periods, emphasizing the positive environmental impact.

4.Economic Benefits: Highlight the potential for economic growth, job creation, and increased energy efficiency resulting from demand response initiatives.

In summary, modeling and explaining the impact of demand response programs involves a comprehensive approach that considers data analysis, economic factors, public engagement, and policy recommendations. Effectively communicating the benefits to stakeholders is essential for driving adoption and ensuring the long-term success of these initiatives.

CONCLUSION

Reducing peak energy demand is essential for ensuring a reliable and sustainable energy supply. By targeting the top three variables contributing to energy consumption—ceiling fans, exterior lights, and freezers—through demand response programs and energy-efficient technologies, we can make significant progress in curbing peak demand. These initiatives not only benefit the environment but also help consumers save on energy bills and enhance the overall stability of the electrical grid.

Group Members

Name	Task
Sai Subrahmanya Akhil Badampudi	-Data Cleaning -Merging -Machine Learning predictions
Jash Nimesh Dharia Jaliwala	-Data cleaning -Exploratory data Analysis
Neha Chandrakant Sharma	-Exploratory Data Analysis -PPT
Masaki	-Machine Learning Predictions -PPT
Sampada Regmi	-Shiny App Development -Report