Technical Explanation Document

IST 687 M009 Group 1

Description

Our work on this project was focused on finding the energy usage of households throughout the different counties that eSC provides power to. We aimed to find out what contributed most to power consumption in households and also to design a model that could accurately predict energy consumption in case of a hotter summer. The model will help eSC decide if the current infrastructure can hold up to a potential temperature increase. We will also provide potential ideals to help people reduce power consumption in their households.

Objectives

1. **Data Preprocessing & Merging** - Done by Nawazish Shaik

   a. **Data Integration and Refinement**

      i. Data Consolidation:

      - Static housing information was retrieved from AWS, resulting in a consolidated data frame termed "static_house."
      - All character columns were converted into factors for later analytical use.
      - The dataset was refined by excluding columns that exhibited minimal variance, identified by having only one unique value.
      - Data quality was further enhanced by removing columns with high intercorrelation, specifically those with correlation coefficients above 0.8, to mitigate multicollinearity issues in future modeling.

   b. **Data Assembly and Synchronization:**

      - A comprehensive data collection process was initiated, targeting energy usage details for households during July.
      - Weather data was synchronized with energy usage statistics, ensuring both datasets were aligned to Eastern Time standards.
      - A final, extensive dataset was created by merging housing, energy, and weather data on shared identifiers - building ID, county, and time - yielding over 4 million records across more than 100 variables.

   c. **Data Streamlining:**

- The large dataset was analyzed to evaluate energy consumption across counties.
- A statistical approach was used to assess the distribution of buildings based on their electrical appliance usage.
- A mean-based sampling method was implemented to effectively reduce the dataset size while retaining representative samples from each county.
- An iterative process was employed to apply this sampling method across counties, significantly reducing the dataset.

2. Descriptive Analysis & Visualization - Done by Pranav Mekal and Bryce Anthony

Data Preparation and Prediction:

The initial part of the process involves preparing future data for analysis and utilizing an XGBoost model for predictions. This step is crucial in understanding and forecasting energy demand. The data is typically transformed into a format suitable for the model's requirements, ensuring optimal performance during the prediction phase.

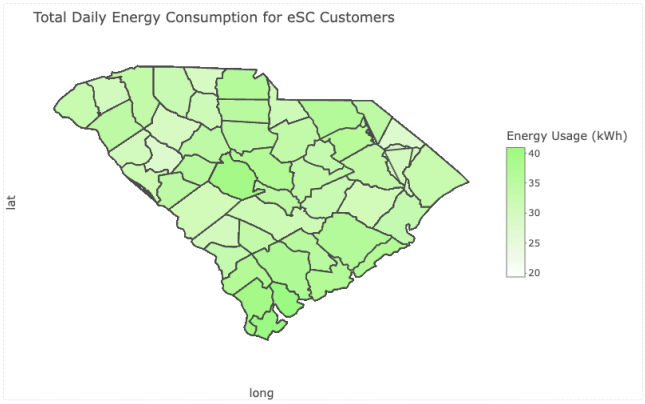
Data Processing and Output:

Following the prediction phase, the results are processed to make them more meaningful and actionable. Predicted demand values are incorporated back into the original dataset, providing a comprehensive view of the expected demand patterns. This augmented dataset is often saved in a structured format, such as a CSV file, for further analysis and dissemination.

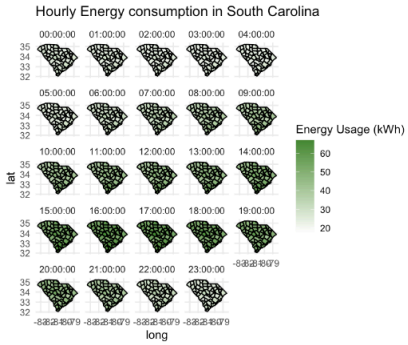
Exploratory Data Analysis (EDA) and Visualization:

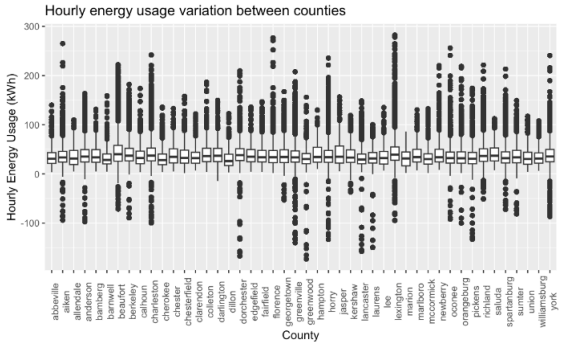
The subsequent steps involve exploratory data analysis (EDA) and visualization to gain insights from the predicted data. Graphical representations, such as line plots and heatmaps, are often used to illustrate trends and patterns. These visualizations enhance the interpretability of the data and make it easier for stakeholders to comprehend complex information.

Exploratory Data Analysis Visualizations:
These were made to explore the data and look for regional trends and differences since we knew we would have access to weather data. Although we didn't find any regional trends, We did learn about negative energy production via solar panels and about the somewhat cyclical trend of hourly energy usage that seemed to fit most of the data. These visualizations also gave us insights into the relative importance of various features in the dataset.

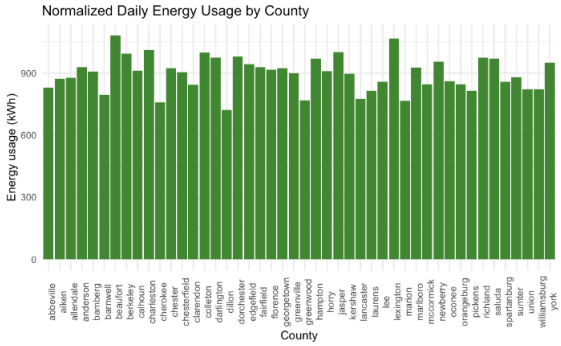

Total Daily Energy Consumption for eSC Customers
Energy Usage (kWh)
40
35
30
25
20
lat
long

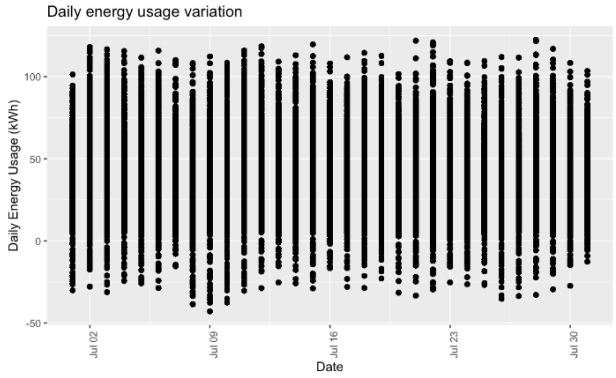
```{r}
```


Hourly Energy consumption in South Carolina
00:00:00
01:00:00
02:00:00
03:00:00
04:00:00
05:00:00
06:00:00
07:00:00
08:00:00
09:00:00
10:00:00
11:00:00
12:00:00
13:00:00
14:00:00
15:00:00
16:00:00
17:00:00
18:00:00
19:00:00
20:00:00
21:00:00
22:00:00
23:00:00
Energy Usage (kWh)
60
50
40
30
20
lat
long

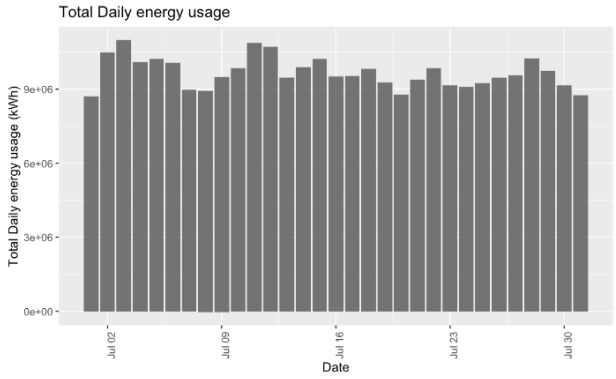

Hourly energy usage variation between counties
Hourly Energy Usage (kWh)
County

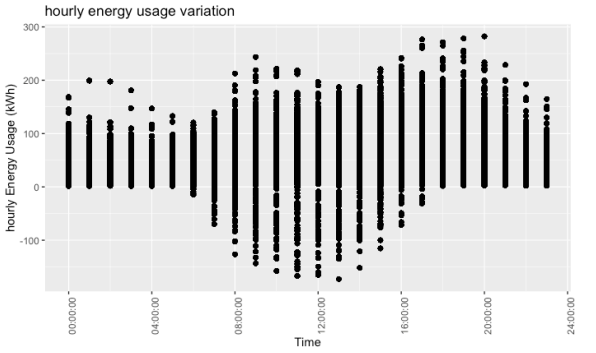
```{r}
```


Normalized Daily Energy Usage by County
Energy usage (kWh)
County

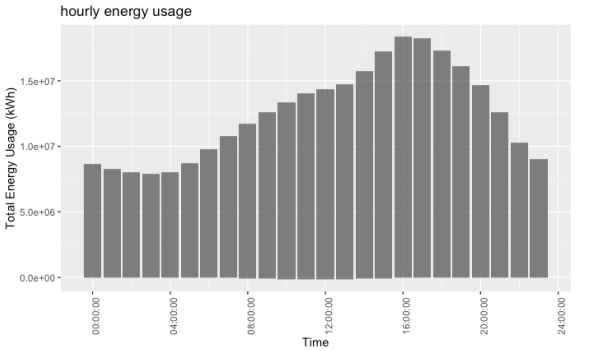

Daily energy usage variation
Daily Energy Usage (kWh)
100
50
0
-50
Jul 02
Jul 09
Jul 16
Jul 23
Jul 30
Date

```{r}
```


Total Daily energy usage
Total Daily energy usage (kWh)
9e+06
6e+06
3e+06
0e+00
Jul 02
Jul 09
Jul 16
Jul 23
Jul 30
Date

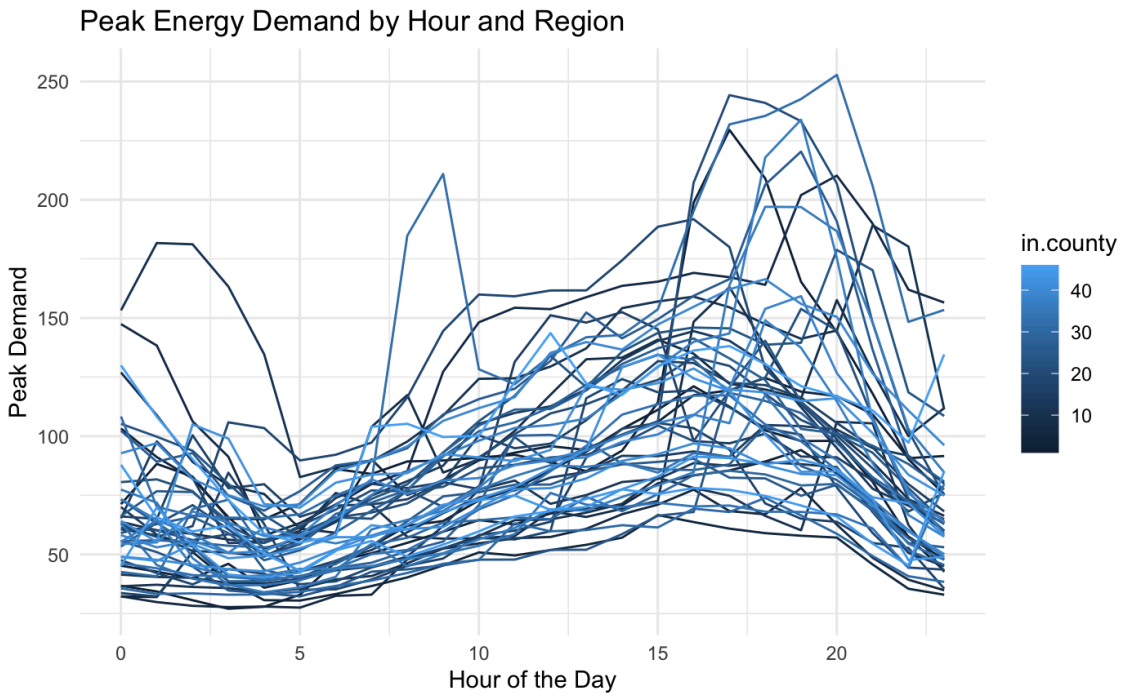



```{r}
```



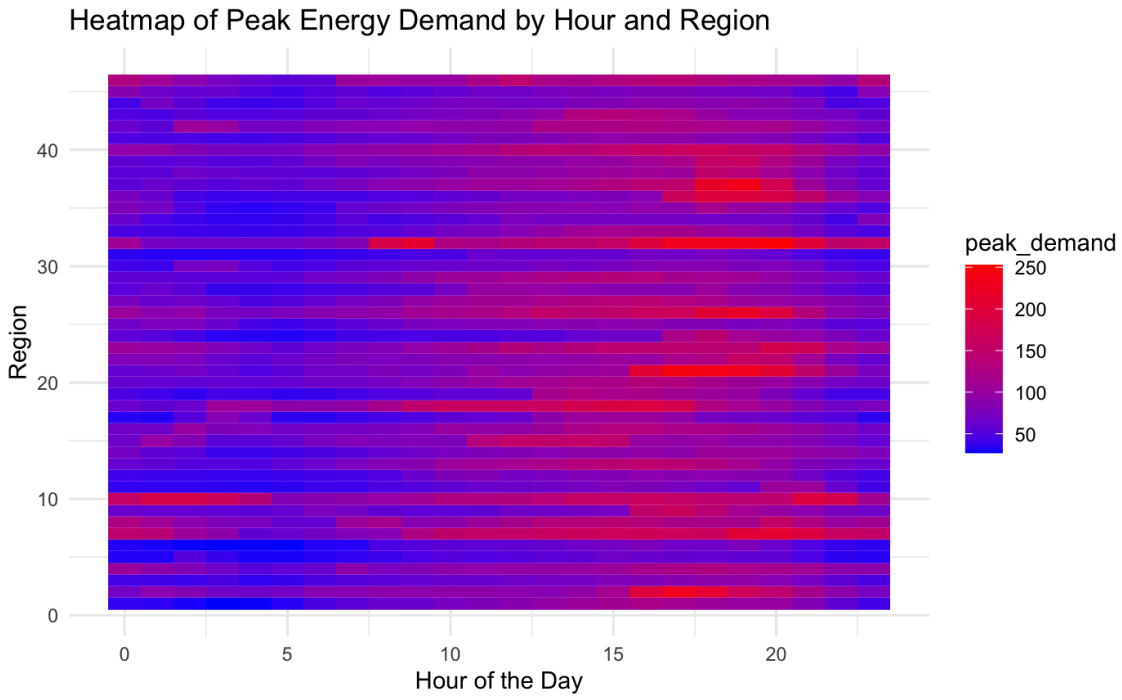
Line Plot of Peak Demand by Hour and Region:

One of the visualizations typically includes a line plot showcasing the variation in peak energy demand across different hours of the day and regions. Different colors may be used to represent distinct regions, facilitating a clear understanding of the trends.



Heatmap of Peak Demand by Hour and Region:

Another common visualization is a heatmap, which provides a comprehensive view of peak energy demand. This grid-based representation uses color intensity to convey the level of demand, with a gradient from low (blue) to high (red). This format is effective in highlighting patterns and variations across both temporal and regional dimensions.



Conclusion:

In conclusion, the overall process encompasses data preparation, model prediction, processing of results, and visualization to gain insights into predicted energy demand. The combination of data processing and visualization aids in making informed decisions and understanding the intricate dynamics of energy consumption patterns across different regions and times.

3. <u>Modeling Preparation & Attribute Selection</u> - Done by Pranav Mekal, Nawazish Shaik, and Bryce Anthony

Introduction:

This analysis revolves around leveraging XGBoost, a powerful machine learning algorithm, to predict total energy consumption. The process encompasses several key stages, including data preprocessing, model training, evaluation, and the application of the trained model to make predictions on future data.

1. Data Preparation:

The initial phase focuses on preparing the dataset for optimal model performance. This involves converting categorical and textual data into a numeric format, ensuring compatibility with the

XGBoost algorithm. Irrelevant columns, such as identifiers and certain energy consumption features, are strategically removed to streamline the dataset for modeling.

2. Model Training and Evaluation:

2.1 Dataset Splitting:

To assess the model's generalization capability, the dataset is divided into training and testing sets. This partitioning allows the model to be trained on a subset of the data and evaluated on unseen data, providing valuable insights into its predictive accuracy.

2.2 XGBoost Model Construction:

The XGBoost model is constructed with careful consideration of various parameters, including the learning rate, maximum tree depth, and the number of boosting rounds. The implementation includes a mechanism for early stopping to prevent overfitting and enhance model robustness.

2.3 Model Evaluation:

The trained XGBoost model undergoes evaluation using standard regression metrics. Metrics such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared gauge the model's performance, revealing how well it predicts total energy consumption on the test set.

3. Feature Importance Analysis:

An important facet of the analysis involves determining the significance of each feature in the predictive process. Feature importance is assessed using the XGBoost algorithm, providing a clear understanding of which variables contribute most to the model's predictions.

4. Future Data Prediction:

The script demonstrates the practical application of the trained model by making predictions on future data. A new dataset is prepared, aligned with the model's requirements, and predictions are generated. This predictive capability is crucial for anticipating energy consumption trends in real-world scenarios.

Conclusion:

In conclusion, the outlined workflow underscores the effectiveness of XGBoost in predictive modeling for total energy consumption. From data preprocessing and model training to evaluation and future predictions, the methodology is designed for robust and insightful outcomes. Feature importance analysis enhances interpretability, while the model's application to new data exemplifies its utility in practical scenarios. This report encapsulates a holistic

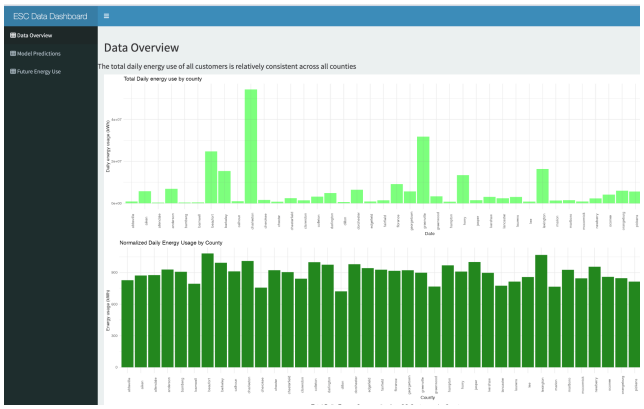
approach to predictive modeling with XGBoost, providing a foundation for informed decision-making in the realm of energy consumption forecasting.

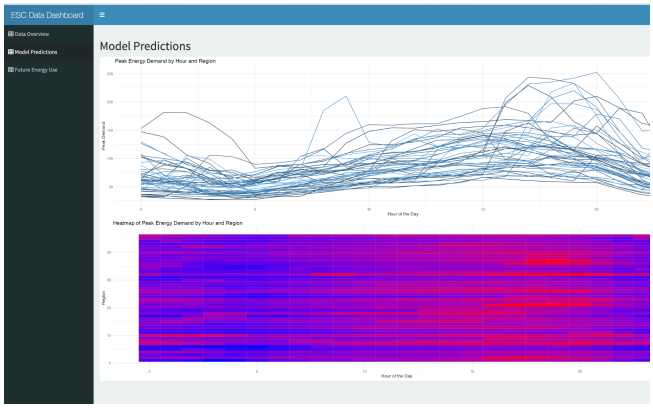
4. Shiny App:

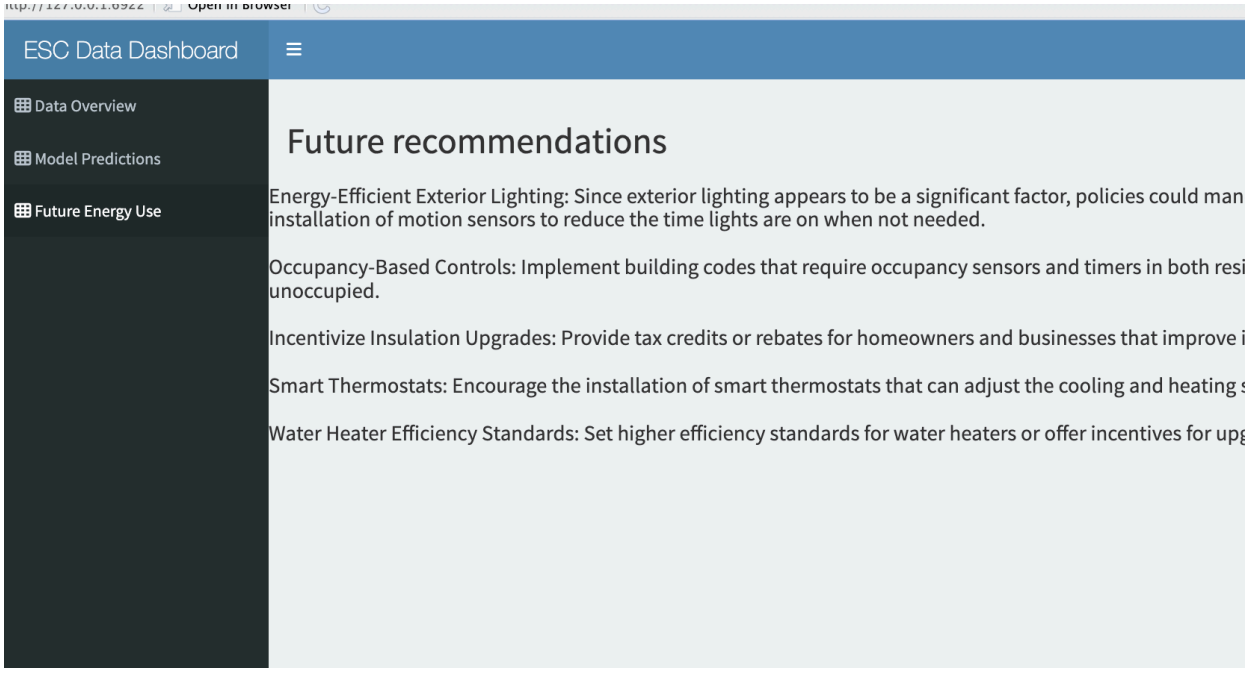
Built by Bryce and John: https://bryce-ka.shinyapps.io/ids_final_g1/

Developed a Shiny Dashboard to showcase the data we used to inform our analyses and to highlight predicted trends in future energy use. The application has a sidebar with 3 tabs. The first tab, Data overview, has a series of plots guiding the user through our exploratory data analysis of daily and hourly energy use across county's and individual building units. The second tab, Model predictions highlights the future peak energy demands as predicted by the model and the third tab, Future Energy Use, goes over our recommendations on reducing peak energy needs in the coming years.

Screenshots:







5. Conclusion:

1. A temperature increase of 5 degrees would increase power consumption. Reasons for this could be increased use of utilities like air conditioning and appliances like fans.
2. Homeowners that have viable houses and the funds to afford it should invest in solar panels for their houses to reduce their energy consumption from the established infrastructure and to also significantly reduce the cost of their power bill.

6. Suggestions:

1. Energy efficient exterior lighting
2. Motion sensors for lights
3. Incentivised insulation upgrades
4. Smart thermostats
5. Energy efficient water heaters
6. Install energy efficient windows