Final Project Report

Team Members: Abhinav Bhandari, Megh Kurlekar, Sumit Kharche, Yash Singh

Course: IST 687

Project Title: Predicting and Managing Energy Demand in Extreme Heat Conditions

Date: 2nd May 2025

Table of Contents

1. Introduction (Scope/Context/Background)

The southeastern United States, particularly South Carolina and neighboring regions, has witnessed a marked increase in summer temperatures over recent years. This climate trend poses a serious risk to the energy infrastructure, especially during July when residential air conditioning demand surges. The regional energy provider, eSC, faces the challenge of managing these increasing energy loads without constructing new power plants or significantly expanding the grid, which would be cost-prohibitive. In this project, our team aimed to support eSC by developing a data-driven framework to forecast energy usage during extreme summer heat, assess the impact of rising temperatures on peak electricity demand, and generate actionable recommendations to reduce the risk of blackouts. By integrating weather, energy consumption, and household characteristic data, we built predictive models and interactive tools to deliver meaningful insights to help optimize energy use during high-demand periods.

2. Business Questions Addressed

Our project focused on answering several critical business questions relevant to eSC's operational planning and consumer engagement strategies. We sought to understand how extreme summer heat influences household electricity consumption, particularly during peak times of day. We also aimed to identify which geographic regions and household characteristics contribute most to increased usage. Furthermore, we evaluated how advanced modeling techniques can be used to accurately forecast demand under future climate scenarios. Finally, we focused on determining what practical, cost-effective steps eSC could take to reduce energy consumption among its customers and mitigate the likelihood of service disruptions without requiring major capital investments.

3. Data Acquisition, Cleansing, Transformation, Munge

Sources and Structure:

- *Static House Data*: HVAC type, appliance types, square footage, number of bedrooms, etc.
- *Hourly Energy Data*: Consumption across 5,710 homes for 2018.

- *Weather Data*: Hourly county-level data including temperature and humidity.
- *Metadata*: Schema documentation.

Steps Taken:

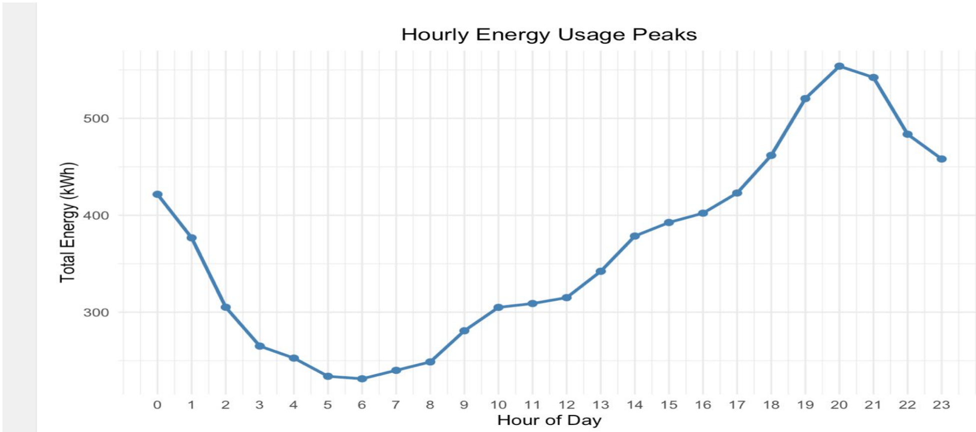
- Merged energy and weather data via timestamps and county identifiers.
- Linked static house data through unique building IDs.
- Focused on June to August to reduce data size and emphasize peak summer load.
- Handled missing values and inconsistencies.
- Standardized date-time formats filtered out incomplete records.

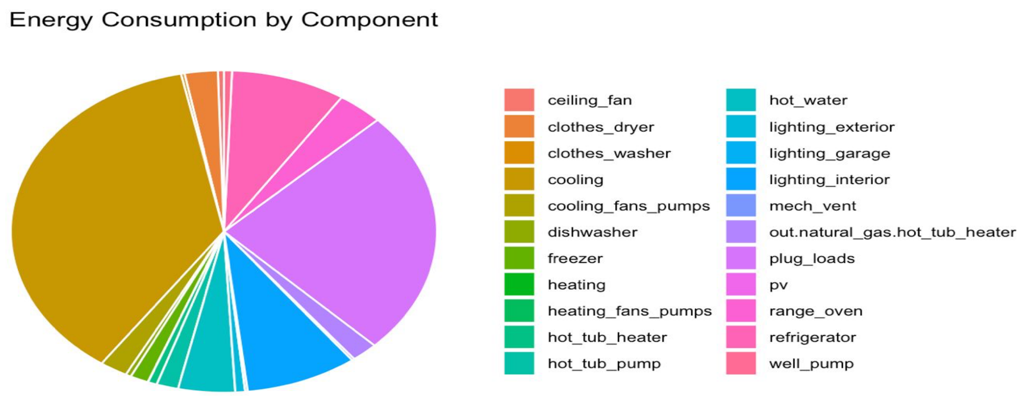
Final Dataset:

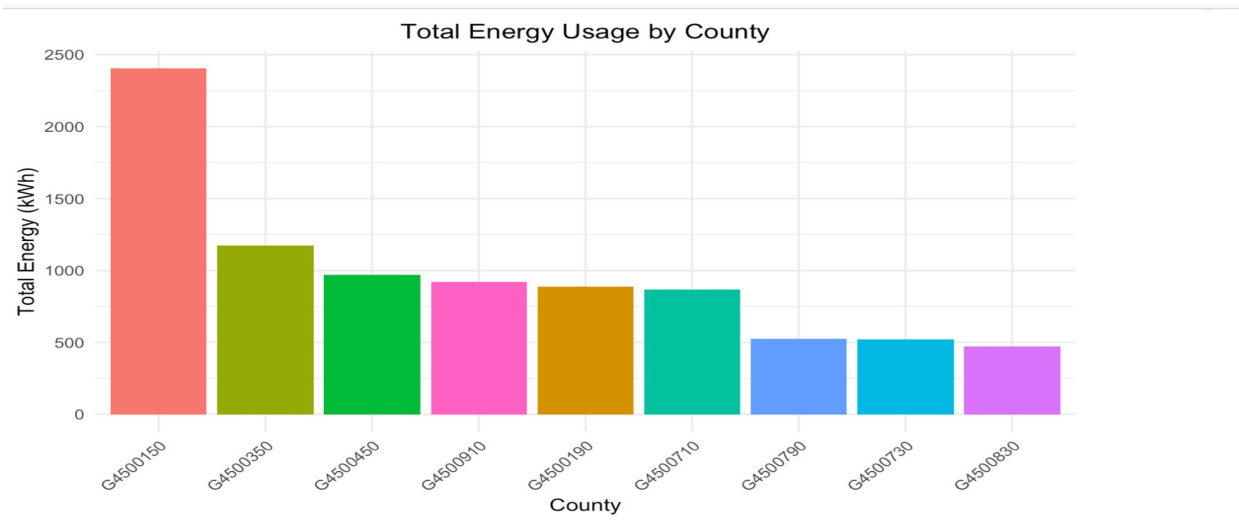
- 12,607,680 rows representing hourly energy consumption from June to August 2018.

4. Descriptive Statistics & Visualizations

Exploratory Data Analysis (EDA) provided key insights into patterns of energy usage across time, geography, and household characteristics. Hourly analysis revealed that the most intensive energy consumption occurred between 9:00 AM–10:00 AM and 7:00 PM–9:00 PM. This confirmed that peak usage aligns with common daily routines and high outside temperatures. Urban counties displayed higher average energy use compared to rural ones, reflecting population density and larger household infrastructure. Among all contributors to energy use, plug loads—including consumer electronics and appliances—were found to be the largest drivers, followed by cooling systems, lighting, and heating. Visualizations such as time-series line charts, county heatmaps, and bar plots helped highlight these differences and directed our feature engineering efforts. The descriptive analysis not only validated our initial assumptions but also helped frame the variables for deeper predictive modeling.


Hourly Energy Usage Peaks
Total Energy (kWh)
500
400
300
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Hour of Day


Energy Consumption by Component
ceiling_fan
clothes_dryer
clothes_washer
cooling
cooling_fans_pumps
dishwasher
freezer
heating
heating_fans_pumps
hot_tub_heater
hot_tub_pump
hot_water
lighting_exterior
lighting_garage
lighting_interior
mech_vent
out.natural_gas.hot_tub_heater
plug_loads
pv
range_oven
refrigerator
well_pump


Total Energy Usage by County
Total Energy (kWh)
2500
2000
1500
1000
500
0
G4500150
G4500350
G4500450
G4500910
G4500190
G4500710
G4500790
G4500730
G4500830
County

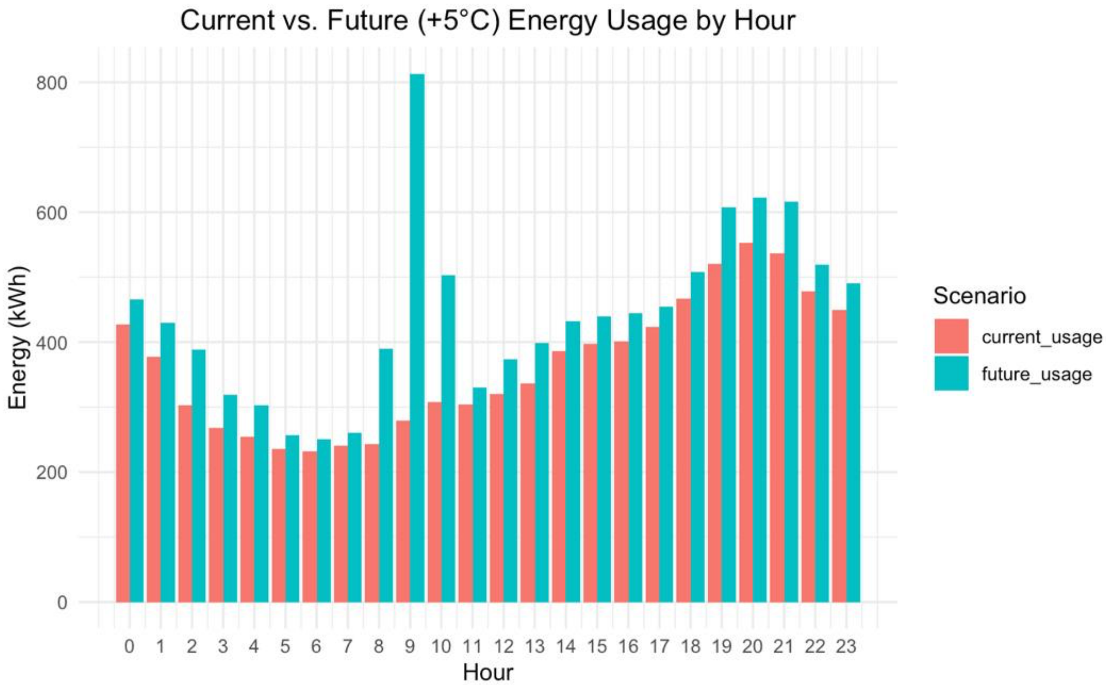
5. Use of Modeling Techniques & Visualizations

We explored three distinct modeling techniques to predict hourly energy consumption during the summer: Linear Regression, XGBoost, and Generalized Additive Models (GAM). The Linear Regression model served as a baseline and offered a straightforward interpretation of variables. It achieved an R^2 of 0.504 and a root mean square error (RMSE) of 0.591. However, it assumed linear relationships and lacked flexibility in handling nonlinear effects. XGBoost, a machine learning algorithm based on gradient-boosted decision trees, improved predictive performance (R^2 = 0.513, RMSE = 0.576) but sacrificed transparency, making it difficult to derive actionable insights. The most successful approach was the Generalized Additive Model (GAM), which combined interpretability with improved performance. The GAM explained 51.2% of the variance with an RMSE of 0.462 and allowed us to visualize the impact of variables such as temperature, square footage, and cooling setpoints using smoothed functions. Visualization of partial dependence plots confirmed the nonlinear effect of rising temperatures on energy consumption, validating the choice of GAM as the best model for both prediction and strategic planning.

Model	RMSE	R^2
Linear Regression	0.589	0.205
GAM (Best Model)	0.462	0.512
XGBoost	0.547	0.313

6. Actionable Insights / Overall Interpretation of Results

The predictive modeling, particularly using the GAM framework, yielded several critical findings. Firstly, temperature increases above 27–28°C led to disproportionately higher energy consumption, suggesting that upcoming hotter summers could overwhelm the energy grid. Square footage of homes was another strong predictor of energy demand, with usage increasing non-linearly in larger homes, especially those above 8,000 sq. ft. Time-of-day coefficients showed significant spikes in consumption during evenings, pointing to the need for consumer engagement around peak load reduction. From an appliance perspective, switching from incandescent to LED lighting could save up to 0.17 kWh per hour per household, while households using room AC units consumed less energy than those using central systems. Scenario analysis simulating a 5°C increase in temperature projected a 26.48% rise in average energy usage, with peak hour demand increasing by as much as 34.6%. These findings led to a series of practical recommendations for eSC.



First, implementing time-of-use pricing would shift nonessential appliance usage to off-peak hours. Second, offering rebates for smart thermostats could help households better manage their cooling needs. Third, public education campaigns promoting energy-efficient lighting and unplugging idle devices could lead to immediate reductions in load. Lastly, rebate programs for efficient HVAC and kitchen appliances would support long-term demand reduction. Together, these strategies provide a roadmap for eSC to reduce peak demand and prevent summer blackouts without costly infrastructure upgrades.

Interactive Shiny Application

As part of our project's evolution, we initially developed a basic Shiny application that accessed a pre-loaded dataset to demonstrate the capabilities of our energy demand prediction model. While this version allowed us to generate predictions and view some basic outputs, it lacked flexibility—users could not upload their own data, filter by specific conditions, or customize their interaction with the model.

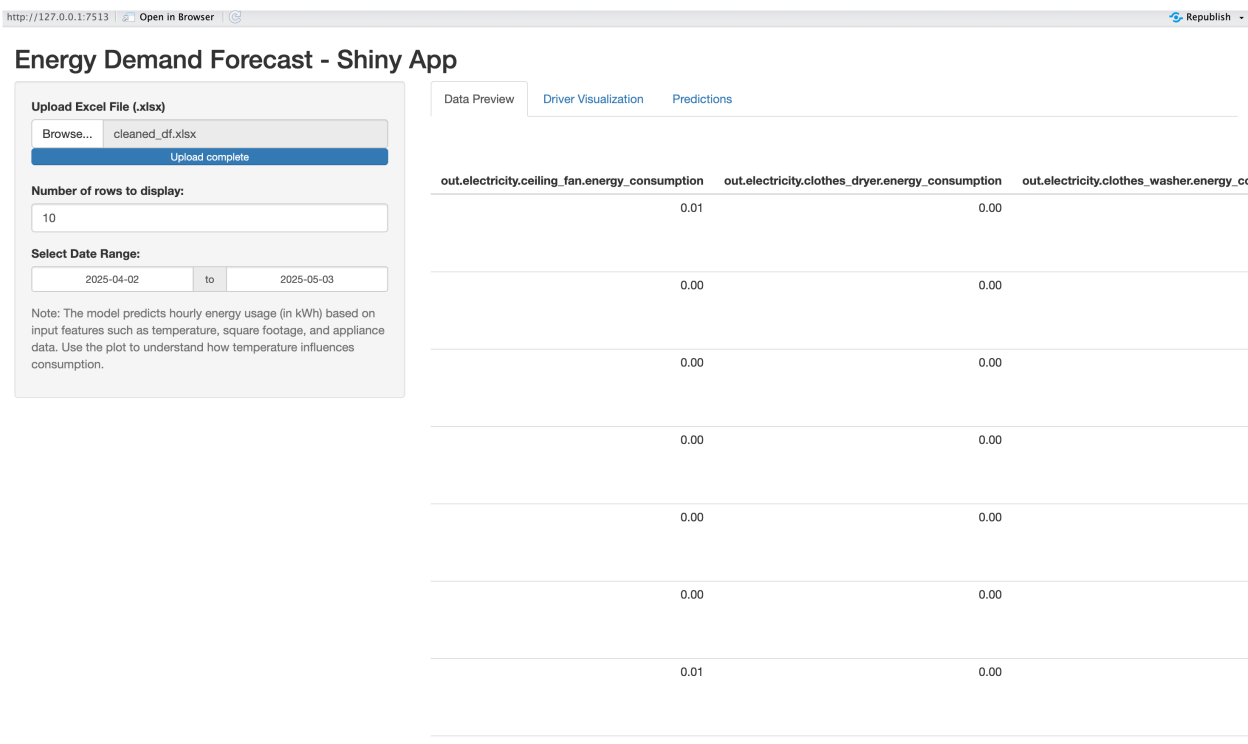
To address these limitations, we redesigned and significantly enhanced the application to support real-world use cases. In the upgraded version, users can upload their own hourly household energy usage data in .xlsx format and view a customizable number of rows.

Building on this foundation, we significantly enhanced the application to include advanced features that support greater interactivity and usability. In the current version, users can filter data by a selectable date range, visualize how temperature correlates with actual energy usage, and analyze how energy demand varies by hour. In addition to dynamic data preview and model prediction, the app now offers a downloadable results feature, a real-time energy usage prediction chart by hour, and a cleaned, column-specific output table tailored for decision-making. These upgrades were informed by our desire to bridge the gap between analytical rigor and user-centered design, making the tool practical for use by energy analysts, policy makers, and grid operators.

Through this iterative development process, our Shiny app evolved from a simple technical demonstrator into a more comprehensive and intuitive forecasting tool that highlights the insights derived from our energy demand model.

Link for shiny app:

https://sumit-kharche.shinyapps.io/IDSFINALAPP/



http://127.0.0.1:7513 | Open in Browser | Republish

Energy Demand Forecast - Shiny App

Upload Excel File (.xlsx)

Browse... cleaned_df.xlsx

Upload complete

Number of rows to display:

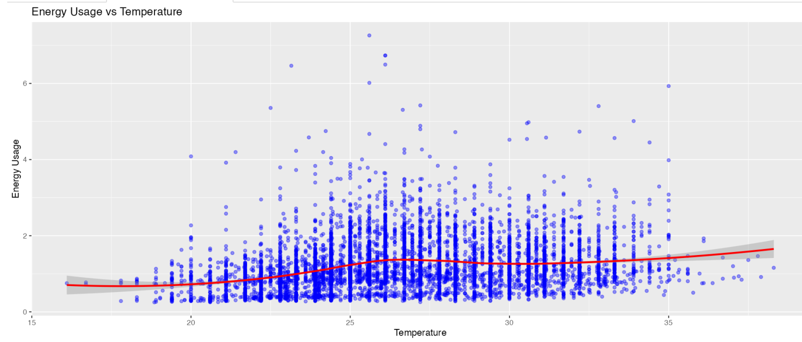
10

Select Date Range:

2018-07-01 to 2018-07-31

Note: The model predicts hourly energy usage (in kWh) based on input features such as temperature, square footage, and appliance data. Use the plot to understand how temperature influences consumption.

Data Preview | Driver Visualization | Predictions



http://127.0.0.1:7513 | Open in Browser

Energy Demand Forecast - Shiny App

Upload Excel File (.xlsx)

Browse... cleaned_df.xlsx

Upload complete

Number of rows to display:

10

Select Date Range:

2018-07-01 to 2018-07-31

Note: The model predicts hourly energy usage (in kWh) based on input features such as temperature, square footage, and appliance data. Use the plot to understand how temperature influences consumption.

Data Preview | Driver Visualization | Predictions

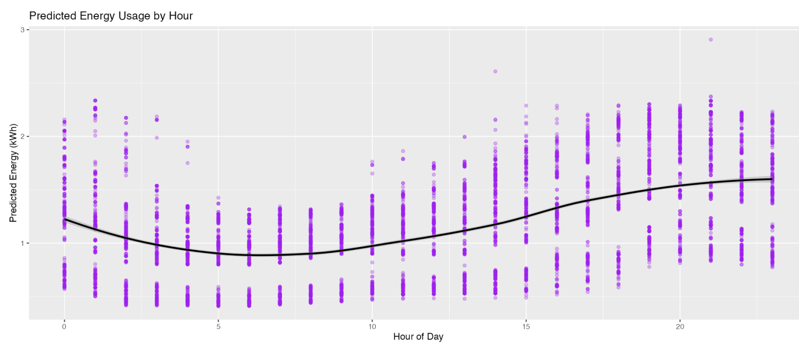
Download Predictions

temp	sqft	occupants	stories	hour	county	Predicted_Energy
22.20	885.00	3	1.00	4	G4500910	0.90
22.20	885.00	3	1.00	5	G4500910	0.89
22.80	885.00	3	1.00	6	G4500910	0.89
24.15	885.00	3	1.00	7	G4500910	0.91
25.35	885.00	3	1.00	8	G4500910	0.96
27.25	885.00	3	1.00	9	G4500910	1.03
29.07	885.00	3	1.00	10	G4500910	1.11
30.00	885.00	3	1.00	11	G4500910	1.13
30.60	885.00	3	1.00	12	G4500910	1.18
30.60	885.00	3	1.00	13	G4500910	1.19
32.20	885.00	3	1.00	14	G4500910	1.38
32.20	885.00	3	1.00	15	G4500910	1.38
31.10	885.00	3	1.00	16	G4500910	1.36
32.20	885.00	3	1.00	17	G4500910	1.56
26.67	885.00	3	1.00	18	G4500910	1.63
25.93	885.00	3	1.00	19	G4500910	1.75
25.00	885.00	3	1.00	20	G4500910	1.76
25.60	885.00	3	1.00	21	G4500910	1.80
24.40	885.00	3	1.00	22	G4500910	1.68
24.03	885.00	3	1.00	23	G4500910	1.61

http://127.0.0.1:7513 Open in Browser Republish

Note: The model predicts hourly energy usage (in kWh) based on input features such as temperature, square footage, and appliance data. Use the plot to understand how temperature influences consumption.

27.25	885.00	3	1.00	9	G4500910	1.03
29.07	885.00	3	1.00	10	G4500910	1.11
30.00	885.00	3	1.00	11	G4500910	1.13
30.60	885.00	3	1.00	12	G4500910	1.18
30.60	885.00	3	1.00	13	G4500910	1.19
32.20	885.00	3	1.00	14	G4500910	1.38
32.20	885.00	3	1.00	15	G4500910	1.38
31.10	885.00	3	1.00	16	G4500910	1.36
32.20	885.00	3	1.00	17	G4500910	1.56
26.67	885.00	3	1.00	18	G4500910	1.63
25.93	885.00	3	1.00	19	G4500910	1.75
25.00	885.00	3	1.00	20	G4500910	1.76
25.60	885.00	3	1.00	21	G4500910	1.80
24.40	885.00	3	1.00	22	G4500910	1.68
24.03	885.00	3	1.00	23	G4500910	1.61



Conclusion

Our project successfully demonstrated how data science can be applied to address a critical real-world challenge: managing residential energy demand during extreme summer heat. By integrating multiple datasets—ranging from hourly weather and energy usage to detailed household characteristics—we built a comprehensive analytical framework to explore the drivers of peak electricity consumption. Among the modeling techniques explored, the Generalized Additive Model (GAM) proved to be the most effective in balancing predictive accuracy with interpretability, making it the ideal tool for both analysis and policy recommendation.

The insights derived from our model validated the concerns of the energy provider, eSC, particularly regarding the projected impact of rising temperatures on grid stability. With an expected 26–34% increase in peak demand under a 5°C temperature rise, few proactive measures are needed to avoid service disruptions. Our findings provide a data-backed foundation for targeted interventions, including smart thermostat programs, dynamic pricing, public awareness campaigns, and appliance efficiency incentives.

Ultimately, our work underscores the value of data-driven decision-making in energy policy and infrastructure planning. With continued model refinement, broader data inclusion, and stakeholder engagement through tools like our Shiny app, eSC can lead the way in building a more resilient and efficient energy ecosystem—one that adapts to climate change without sacrificing affordability or reliability.