REPORT – ESC

IST 687: Introduction to Data Science

GROUP 2 | Prof. Riley

Optimizing Energy Consumption through Data-Driven Analytics for eSC

Abstract:

This report describes a joint venture between a team of data analysts and eSC, a forward-thinking energy company dedicated to sustainable energy delivery. The main objective was to create a service based on data analytics to optimize energy consumption, especially during high-demand periods in summer. Through sophisticated data analytics and predictive modeling, the project sought to improve eSC's operational effectiveness and commitment to environmental sustainability, without major new capital expenditures. The outcomes from the data analysis and models suggest considerable potential for energy conservation and enhanced management of energy demand, which supports wider environmental objectives.

Introduction:

The escalating global temperatures have underscored the importance of energy management for both sustainability and operational effectiveness. In response, eSC collaborated with data analysts to harness sophisticated data analytics techniques to refine energy consumption optimization. This endeavor is in sync with international efforts to tackle climate change, epitomizing a proactive strategy towards intelligent energy management. The primary aims of this project were to dissect energy usage trends, forecast peak demand periods, and devise strategies for energy consumption that are both effective and based on analytic insights. The collaboration also aimed to integrate renewable energy sources more effectively into the existing grid, thereby reducing reliance on non-renewable resources. Additionally, by understanding consumption patterns, eSC hopes to influence consumer behavior towards more energy-efficient practices, further contributing to environmental and economic sustainability.

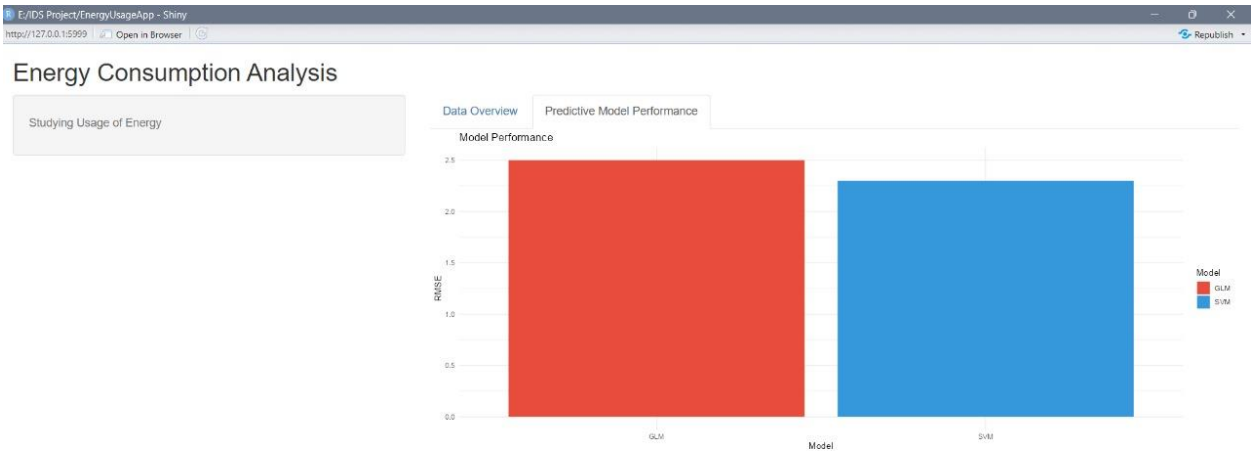
Methodology:

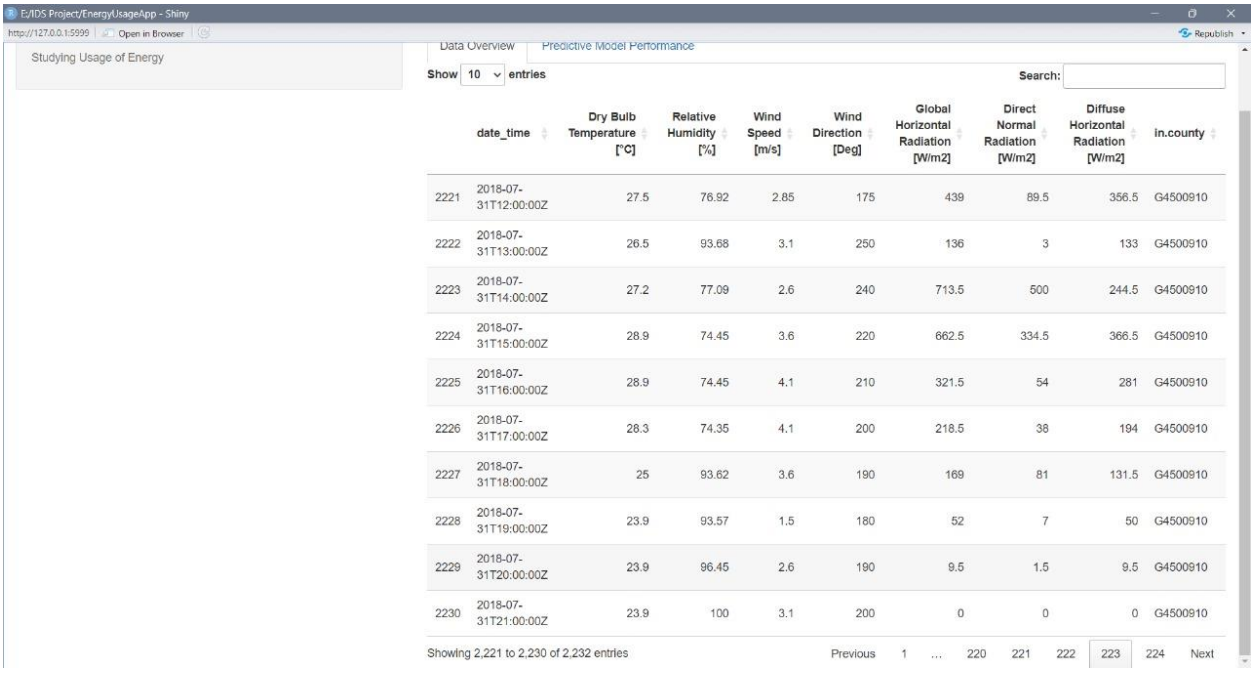
The methodology employed in this project included the collection and encryption of comprehensive datasets, which covered static housing data and hourly weather patterns from several counties across the Carolinas. This information detailed the characteristics of the buildings and the specific types of energy consumption, and it was supplemented by additional metadata related to weather and energy. A meticulous data cleansing process was carried out to refine these datasets for analytical purposes, concentrating on relevant building IDs and specific counties. The core of the data analysis involved the integration of two primary datasets: housing data and energy usage data, which were merged based on building ID and county parameters, enabling an in-depth examination of energy consumption trends. Additional steps included the integration of temporal data to examine energy usage fluctuations over time and the application of advanced filtering techniques to isolate significant patterns and anomalies. This comprehensive approach not only enhanced the accuracy of the analysis but also allowed for more targeted energy management recommendations.

Results:

In the results section, the merged datasets afforded a holistic perspective on energy consumption patterns across various building types and environmental conditions. The statistical analysis pointed out that electricity and natural gas were the chief sources of energy, with minimal contributions from other types. The use of visual analytics illuminated discrepancies in energy usage among different counties and building types, pinpointing areas where energy efficiency could be markedly improved. Additionally, the execution of R code facilitated not only the data cleaning and merging processes but also the application and demonstration of predictive modeling techniques, specifically Generalized Linear Models (GLM) and Support Vector Machines (SVM). The outputs from these models were crucial in refining our understanding and predictions of energy consumption trends. Furthermore, the results were dynamically displayed using Shiny applications, providing interactive visualizations that effectively communicated the findings and underscored the practical implications of the data-driven insights.

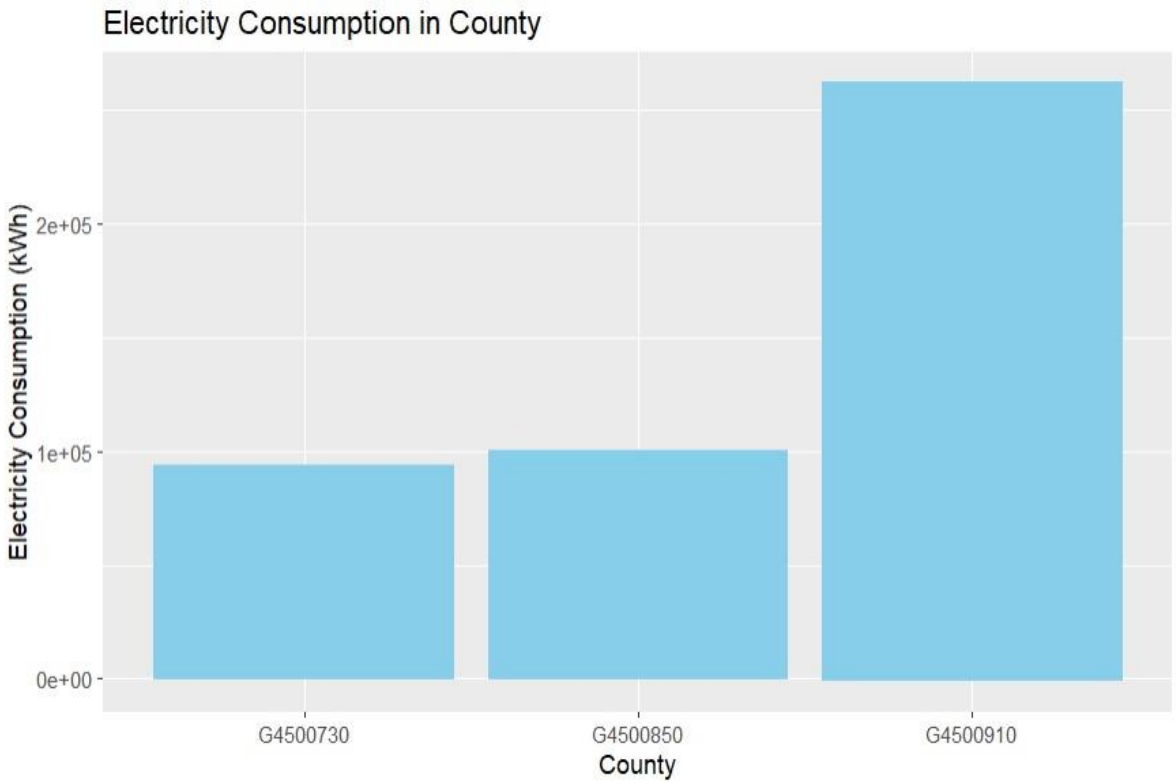
Shiny App Output

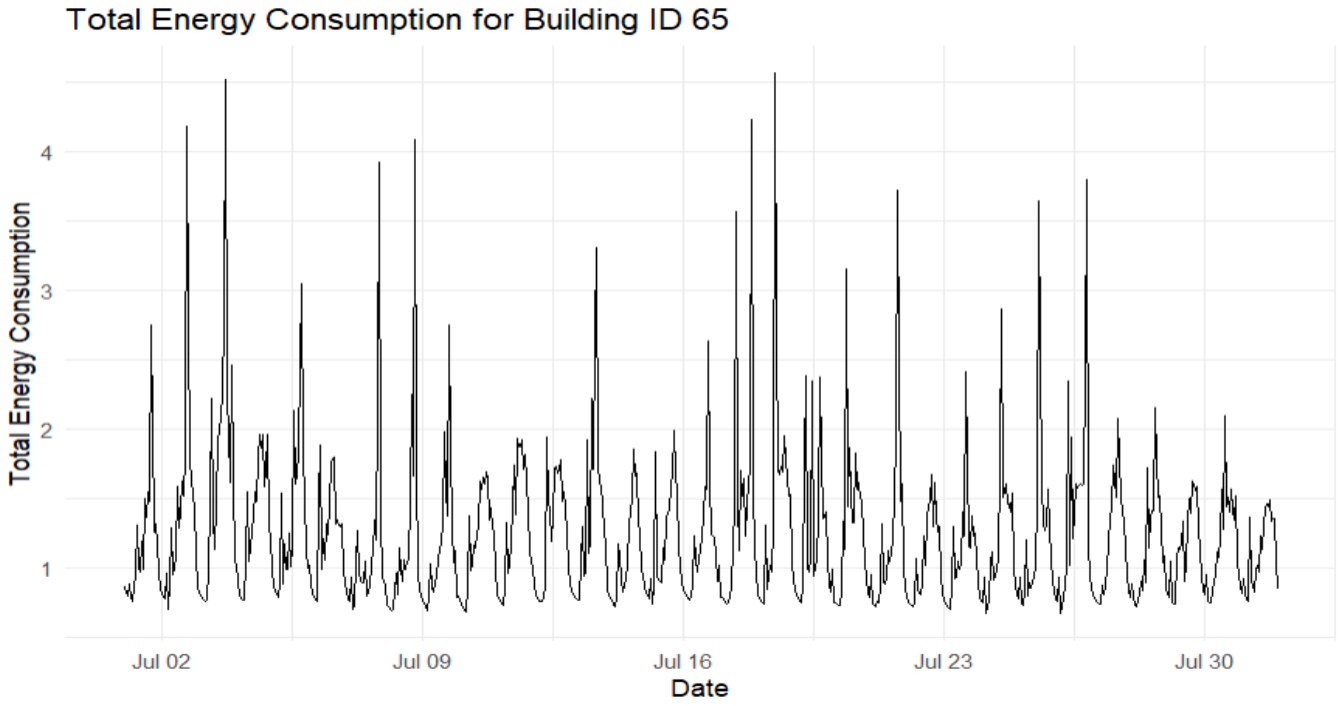


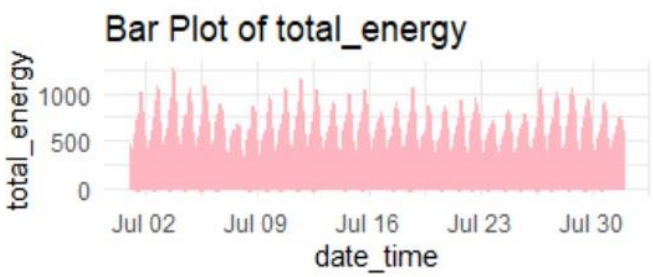


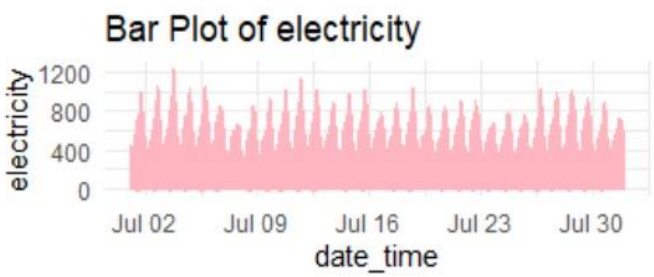
	date_time	Dry Bulb Temperature [°C]	Relative Humidity [%]	Wind Speed [m/s]	Wind Direction [Deg]	Global Horizontal Radiation [W/m2]	Direct Normal Radiation [W/m2]	Diffuse Horizontal Radiation [W/m2]	in.county
2221	2018-07-31T12:00:00Z	27.5	76.92	2.85	175	439	89.5	356.5	G4500910
2222	2018-07-31T13:00:00Z	26.5	93.68	3.1	250	136	3	133	G4500910
2223	2018-07-31T14:00:00Z	27.2	77.09	2.6	240	713.5	500	244.5	G4500910
2224	2018-07-31T15:00:00Z	28.9	74.45	3.6	220	662.5	334.5	366.5	G4500910
2225	2018-07-31T16:00:00Z	28.9	74.45	4.1	210	321.5	54	281	G4500910
2226	2018-07-31T17:00:00Z	28.3	74.35	4.1	200	218.5	38	194	G4500910
2227	2018-07-31T18:00:00Z	25	93.62	3.6	190	169	81	131.5	G4500910
2228	2018-07-31T19:00:00Z	23.9	93.57	1.5	180	52	7	50	G4500910
2229	2018-07-31T20:00:00Z	23.9	96.45	2.6	190	9.5	1.5	9.5	G4500910
2230	2018-07-31T21:00:00Z	23.9	100	3.1	200	0	0	0	G4500910

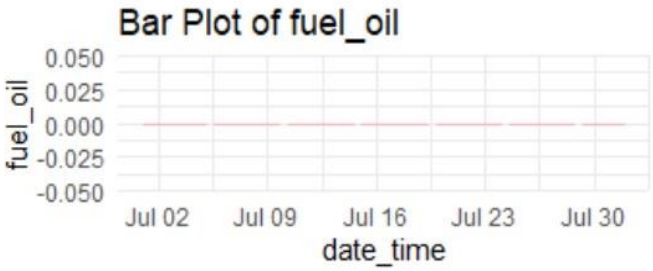
Merged Data Report

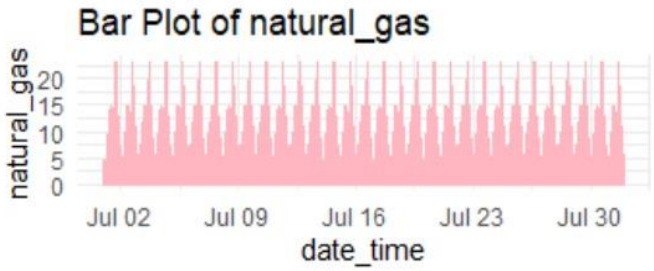


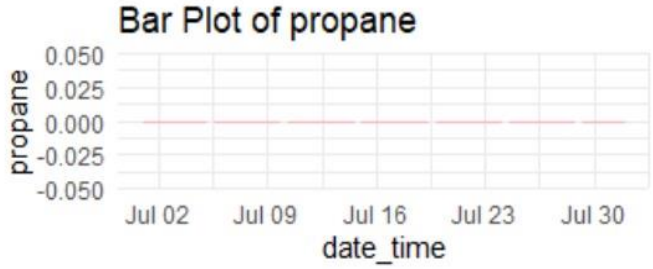


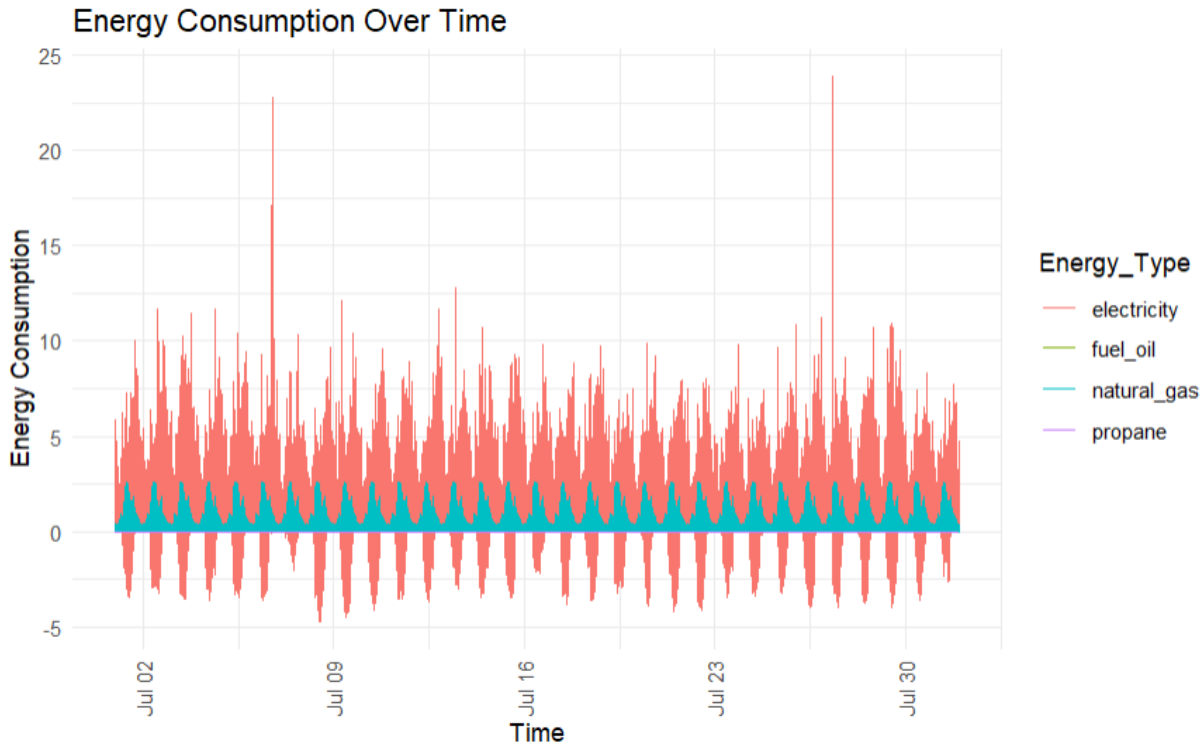

Bar Plot of total_energy
total_energy
1000
500
0
Jul 02
Jul 09
Jul 16
Jul 23
Jul 30
date_time


Bar Plot of electricity
electricity
1200
800
400
0
Jul 02
Jul 09
Jul 16
Jul 23
Jul 30
date_time

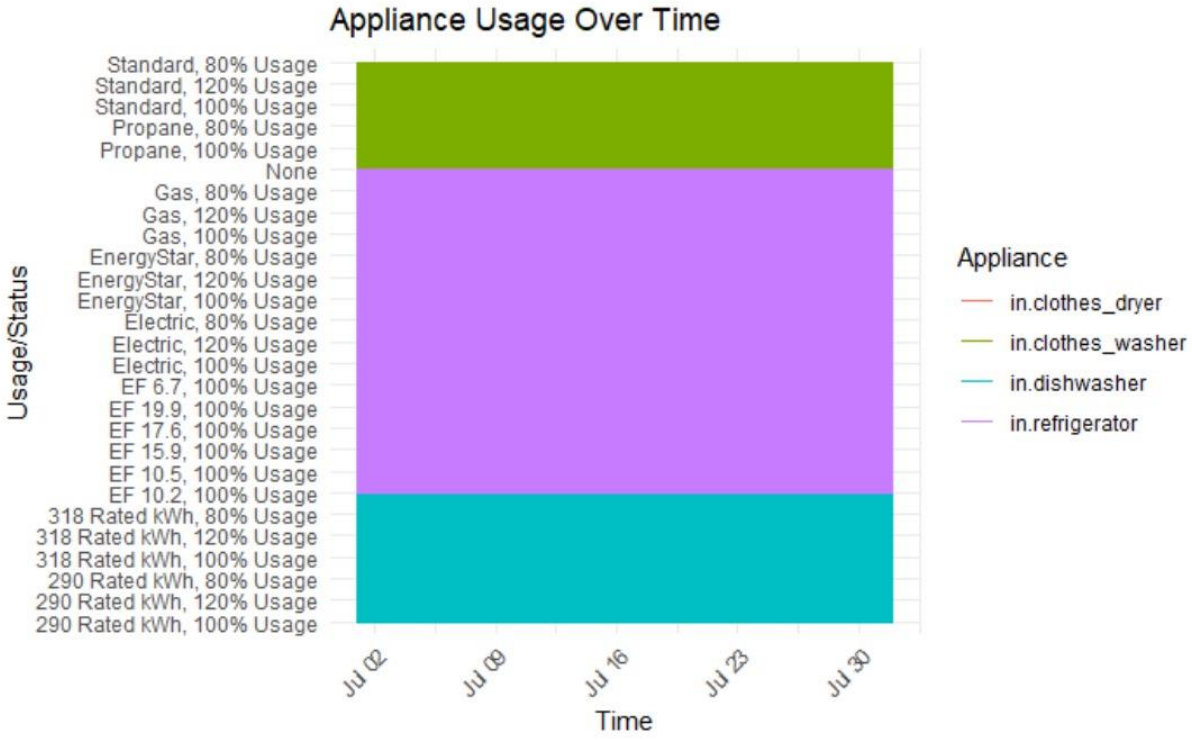

Bar Plot of fuel_oil
fuel_oil
0.050
0.025
0.000
-0.025
-0.050
Jul 02
Jul 09
Jul 16
Jul 23
Jul 30
date_time

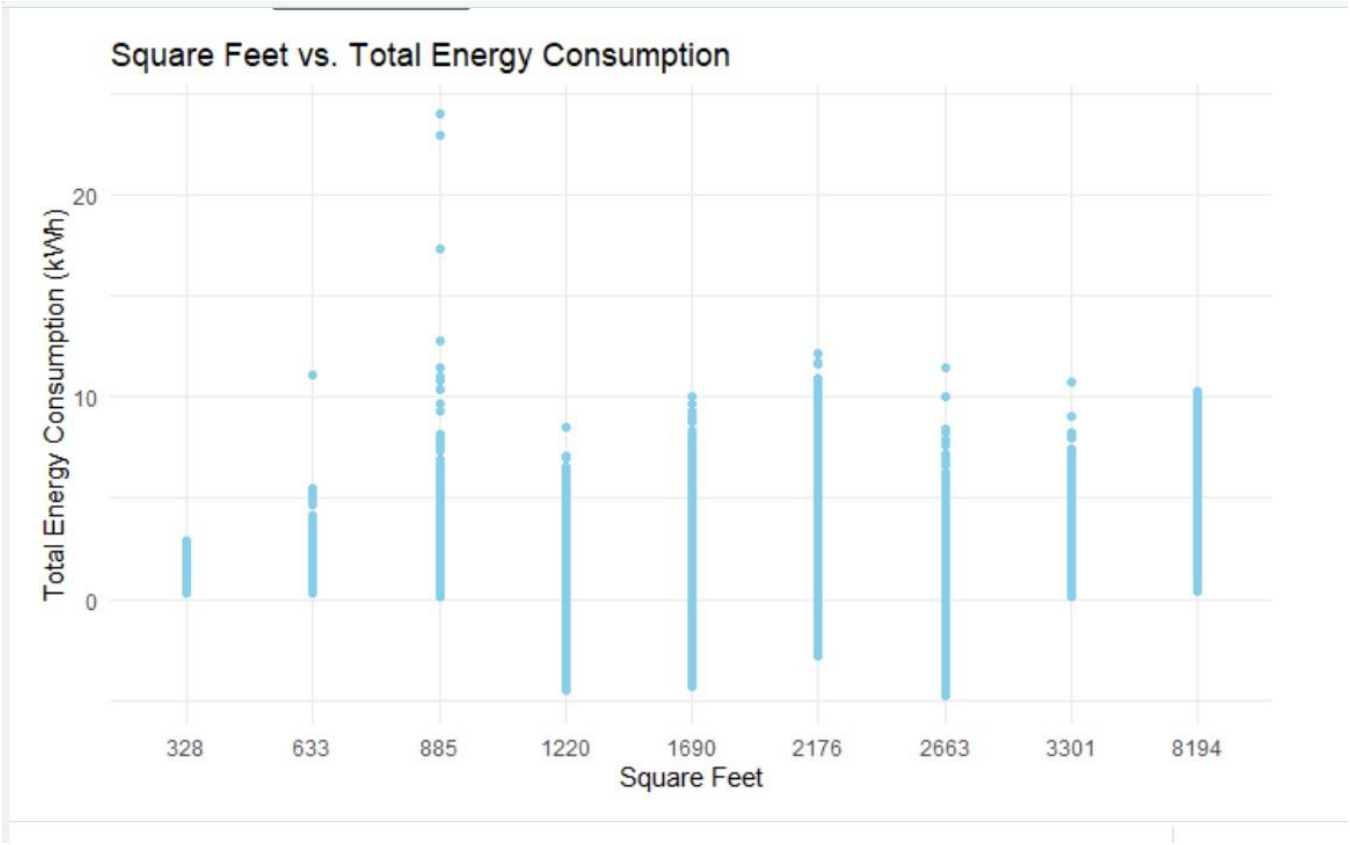

Bar Plot of natural_gas
natural_gas
20
15
10
5
0
Jul 02
Jul 09
Jul 16
Jul 23
Jul 30
date_time

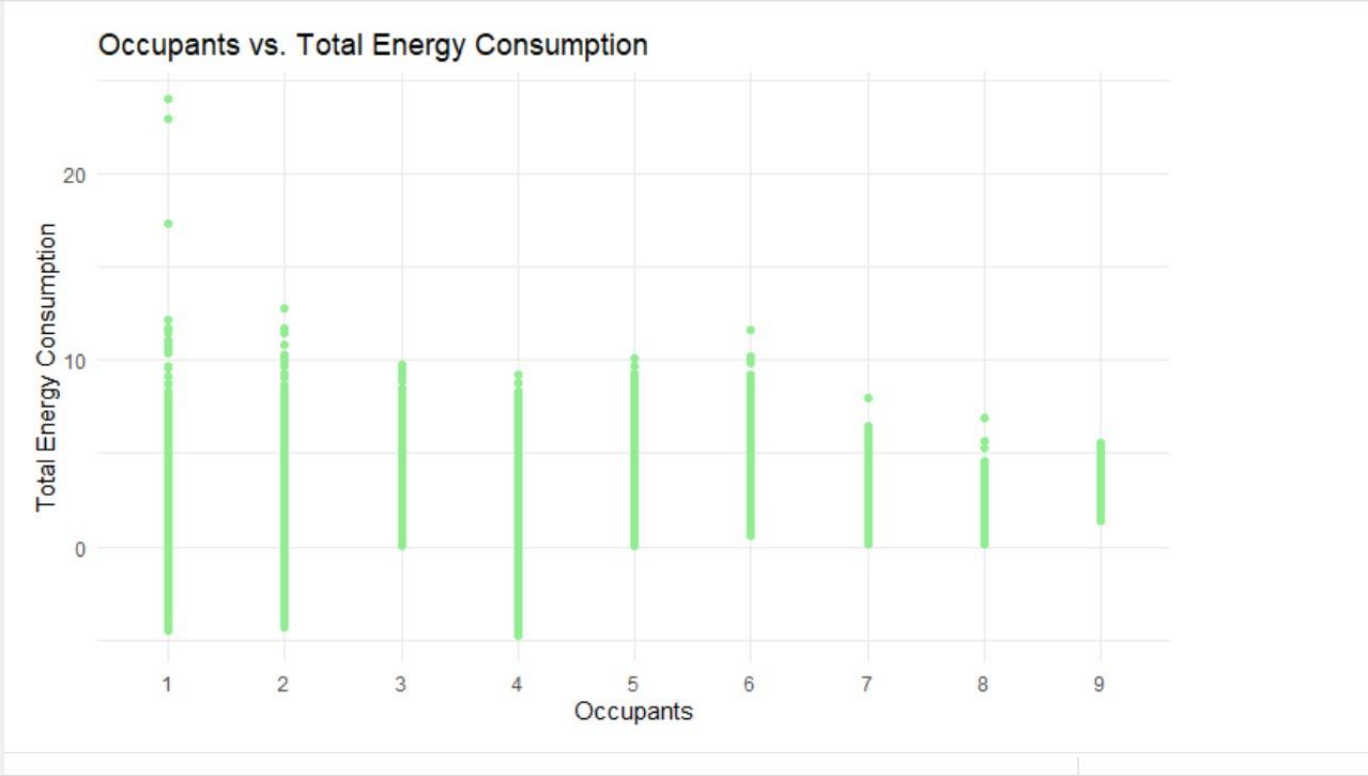

Bar Plot of propane
propane
0.050
0.025
0.000
-0.025
-0.050
Jul 02
Jul 09
Jul 16
Jul 23
Jul 30
date_time


Energy Consumption Over Time
Energy Consumption
25
20
15
10
5
0
-5
Jul 02
Jul 09
Jul 16
Jul 23
Jul 30
Time
Energy_Type
electricity
fuel_oil
natural_gas
propane

Predictive Models Output






Occupants vs. Total Energy Consumption
Total Energy Consumption
20
10
0
1
2
3
4
5
6
7
8
9
Occupants

Discussions:

The analysis facilitated the creation of predictive models employing Generalized Linear Models (GLM) and Support Vector Machines (SVM), selected due to their reliability and adaptability in managing varied datasets. These models demonstrated high predictive accuracy, as evidenced by low root mean square errors—0.5255 for GLM and 0.0784 for SVM. Such precision is crucial for forecasting peak energy demands and formulating dynamic energy management strategies. The accuracy of these models allows stakeholders to make informed decisions about energy distribution and conservation strategies in real-time. Additionally, the models' ability to handle complex and multifaceted datasets makes them invaluable tools for continuous improvement in energy efficiency initiatives. This predictive capability is especially important in anticipating and mitigating potential challenges in energy supply and demand, thereby enhancing sustainability and operational efficiency.

Conclusion:

The project effectively showcased the profound influence that data-driven analytics can have on optimizing energy management and enhancing efficiency. The deployment of predictive models equips eSC with robust tools to accurately predict and manage peak energy demands, thus circumventing the need for expensive new infrastructure developments and minimizing environmental repercussions. This partnership has established a model for how energy providers can utilize analytics to address modern challenges in energy consumption and sustainability. By integrating these analytical tools, eSC is now better positioned to adapt to fluctuating energy needs and implement proactive strategies that contribute to long-term sustainability goals. Moreover, the success of this initiative demonstrates the potential for broader adoption of similar strategies across the industry, potentially leading to widespread improvements in energy conservation and efficiency. This project not only advances eSC's operational goals but also contributes to a larger framework of sustainable practices in the energy sector.

References:

All data analysis and predictive modeling were carried out using specialized datasets and tools, detailed in the project documentation. Specific R packages and methodologies used are available on Blackboard Exams and Projects Bar. Additionally, the R code and Shiny application used in this project will be uploaded to the Blackboard platform along with this report, allowing for easy access and further exploration by researchers and students. This helps ensure transparency and fosters educational and practical applications of the developed methods.