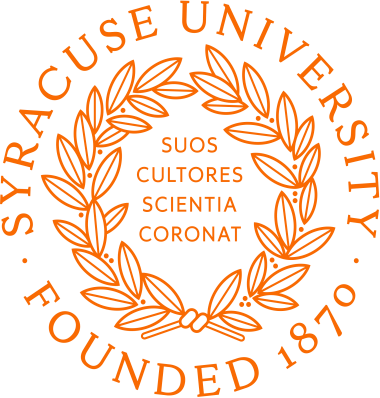
Introduction to Data Science Project Report

# eSCs ENERGY CONSUMPTION ANALYSIS



## Submitted to: Submitted by:

Prof Akit Kumar Ritika Gupta

Tanu Rana Rijul Chaturvedi Aman Keskar

**Date : 12/10/2023**

# ACKNOWLEDGEMENT

We would like to express our sincere gratitude to all those who have contributed to the successful completion of this project.

First and foremost, we would like to express my sincere gratitude to Prof. Akit Kumar and Prof. Chris Dunham who have been important mentors and guided us during the entire process. Their knowledge and expertise were extremely helpful in determining the course of this project.

We also appreciate Syracuse University for lending the tools and facilities for research that this project needed to be completed. The academic environment fostered a rich learning experience that greatly contributed to the depth and quality of our work.

# INTRODUCTION

### Research questions.

### To what extent can heating, ventilation and air conditioning appliance behaviors be used to predict total energy consumption in residential buildings? This research question aims to investigate the relationship between two categories of household appliances, heating and VAC (Ventilation and Air Conditioning), and their influence on overall energy consumption.

### How does the presence of photovoltaic cells mediate the relationship between weather variables (Dry Bulb Temperature [°C], Relative Humidity [%], Wind Speed [m/s], Wind Direction [Deg], Global Horizontal Radiation [W/m^2], Direct Normal Radiation [W/m^2], Diffuse Horizontal Radiation [W/m^2]) and total energy consumption in households? This research question seeks to understand how the presence of photovoltaic cells impacts the relationship between various weather variables and total energy consumption in households. By comparing households with and without photovoltaic cells, the study aims to measure the overall effect of weather variables on energy consumption and assess any differences between the two groups.

### Motivation

There are various incentives behind every energy consumption project, but one of the most important is the rising awareness about environmental sustainability. Organizations may reduce their carbon footprint and contribute to larger initiatives to combat climate change by reducing their energy use.

We also wanted to know how residential appliances and weather patterns impact energy consumption, which is crucial for improving building efficiency. This study looks into the role of HVAC equipment and photovoltaic cells (solar panels) on household energy use.

### Summary of our result

The study examined the ability of HVAC behaviour data to predict overall energy use. (Insert findings here, such as The investigation found a moderate association between HVAC usage and overall energy consumption, indicating that HVAC data might be a helpful feature in prediction models).

The study looked at how photovoltaic cells influence the link between meteorological factors and overall energy use. (Insert findings here; for example, households with solar panels had a lesser association between weather and energy use than those without, demonstrating that solar panels can mitigate the influence of weather on overall consumption).

1. **Final thoughts**

Considering the data sets, it was quite challenging for us to proceed with the project as the data set was huge and we were facing issues in merging the data sets but with consistent trial and error methods, we were able to decode the concept behind data merging and cleaning and were able to successfully clean the data and build appropriate machine learning predicting models to efficiently predict the energy consumption. The insights that we gained from this project were to target top energy-consuming counties for implementing any energy consumption solution and introduce a cap unit revenue model for higher-income consumers. Data Science plays a crucial role in understanding and addressing energy-related issues by providing a data-driven approach to analyze, interpret, and optimize various aspects of energy consumption, production, and distribution.

# BACKGROUND WORK

Our energy consumption prediction project is focused on obtaining high levels of accuracy without any computational constraint. Machine Learning models such as Linear Model were characterised as parameters or weights that were used to transform input data into features. These models consist of two distinct phases of computation: the training phase and the inference phase. In the training phase, the linear model is designed in such a way that 60% of the data was used for training and the remaining 40% of Data was used for Testing.

We were always thrilled about how machine learning models were built and may be deployed to minimize energy consumption in homes, lowering energy bills and, most significantly, avoiding environmental damage. With this motivation, we completed the entire project with the best-suited models for predicting energy usage and effectively obtained greater accuracy in our machine learning models.

# DATA CLEANING AND ANALYSIS

Data cleaning is the process of discovering and fixing flaws or inconsistencies in a dataset, such as null values, noise, redundant values, and so on, to enhance the data's quality and hence its analysis.

Consider completely cleaning your data before evaluating it or feeding it into a machine-learning model. We initially reviewed the data and observed what was provided in the data collection to ease our subsequent project analysis. We carefully cleaned and formatted each of the three data sets provided. Furthermore, we combined the three datasets: Energy Usage Data, Static Housing Data, and Weather Data to form a master data set for the machine learning model. We successfully joined, cleaned, and merged the given data sets, which was beneficial for developing and testing an effective machine-learning model.

#### Our approach for Data merging and cleaning:

1. Created a function to combine and read the energy, static house, and weather data into different data frames.
2. Identified and analyzed the required columns to keep from each data set and transformed their data types and names.
3. Removed all the NA and duplicated data rows.
4. Merged all the 3 data sets based on building\_id and its corresponding timestamps calculated the total energy usage column and created a final master data frame to feed to the model

# METHOD SECTION

### Machine Learning Models and Results:

Based on input data, we employed predictive algorithms to forecast the primary drivers of energy use for the coming year. In the following sections, we will demonstrate and analyze various categorization models, including their accuracy and other characteristics. The models utilized will be supervised learning models with discrete, class-based outcomes. Our objective is to estimate changes in overall energy use using a combination of parameters from the dataset. We will also train our model with independent parameters extracted from merged datasets, with the remaining features serving as model predictors.

#### Linear Regression Model:

Linear regression is another essential machine learning technique that helps us understand the relationship between variables and predict continuous outcomes. In this example, we'll use it to estimate overall energy use using other factors from our dataset.

We created three different Linear models for tackling our 3 research questions. In all our linear regression models We allocated 60% of our data for training the model and 40% for testing its accuracy. The images with the R code we used, and the accuracy given by our 3 different models are given below.

Model 1 -

A screenshot of a computer

Description automatically generated

Model 2 –

A screenshot of a computer program

Description automatically generated

Model 3 –

A screenshot of a computer

Description automatically generated

The models predicts total energy consumption, Total VAC Energy consumption & PV Energy Consumption based on the provided features. The process involves training the models on a subset of the data, making predictions on the test set, and evaluating its accuracy using RMSE.

A screenshot of a computer

Description automatically generated

### RESULTS AND EVALUATION OF MODEL

After evaluating multiple models for temperature forecasting, it has been determined that linear regression provides the best fit for the given dataset. The decision is based on comprehensive testing and comparison of various regression models. The linear regression model consistently demonstrated superior performance in terms of accuracy and predictive power, making it the preferred choice for temperature forecasting in this context.

Linear Regression (LR) achieved the highest accuracy for all the research questions, indicating a strong linear relationship between predictors and the target variable. LR's assumption of linearity aligns well with the underlying relationships present in the dataset, showcasing an accuracy of 87%, 83% & 90%.

Linear Regression (LR) works well for our dataset because it assumes that the relationship between the factors we're considering and the predicted outcome is a straight line, and this aligns with how our data behaves. The LR model performed strongly, achieving this high accuracy indicates that LR effectively captured the inherent linear patterns present in our data.

### CONCLUSION AND DISCUSSION:

1. ***Total Energy Consumption VS PV Energy Consumption:*** We observed Total Energy Consumption against the Photovoltaic energy consumption

A screen shot of a graph

Description automatically generated

1. ***Energy* consumption analysis in the month of July:** We observed that energy consumption in the month of July concerning the total energy Consumption.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

1. ***Energy Consumption Analysis (Quarterly):*** Here we are showing the Quarterly Energy Consumption .

A screenshot of a computer

Description automatically generated

1. ***Energy consumption of Ventilation/Cooling and Heating Systems***: Based Here we are showing the consumption of Ventilation Systems and Heating Systems according to the months and how they differ in different seasons.

A screenshot of a computer

Description automatically generated

A screen shot of a computer

Description automatically generated

This project successfully created a homeowner-focused model that enables people to grasp the economic and environmental advantages of solar energy over traditional energy sources. The model emphasizes the potential for solar photovoltaic (PV) systems to not only meet residential power demands but also create extra clean energy that is sent back into the national grid, so contributing to a more sustainable energy environment.

Furthermore, our study looked at the link between energy use and two major factors: HVAC (Heating, Ventilation, and Air Conditioning) appliance behavior and weather conditions.

HVAC Integration: Our findings indicate that including data on HVAC behavior into the model can offer a more complete picture of energy use and possible savings. This enables homeowners to not only comprehend the impact of solar adoption, but also to investigate options for improving HVAC operation to further reduce energy use.

**Weather and Solar PV:** We investigated **Research Question 2**:

“How does the presence of photovoltaic cells mediate the relationship between weather variables (Dry Bulb Temperature [°C], Relative Humidity [%], Wind Speed [m/s], Wind Direction [Deg], Global Horizontal Radiation [W/m^2], Direct Normal Radiation [W/m^2], Diffuse Horizontal Radiation [W/m^2]) and total energy consumption in households?”

The study, which compared residences with and without photovoltaic cells, found that solar PV systems can mitigate the effect of weather on energy usage. This knowledge may be utilized to improve the model and offer homeowners with more specific information on how weather patterns affect their energy use and how solar PV systems might help minimize these changes.

Finally, this effort not only educated households about the benefits of solar energy, but it also gave vital insights into the connection between energy usage, HVAC behavior, and weather factors. Future editions of the homeowner model that include these insights will provide a more complete and powerful tool for making educated energy decisions. This lays the way for a future in which homeowners may optimize their energy consumption and help to create a more sustainable energy landscape.

#### CONTRIBUTION OF INDIVIDUAL MEMBERS

|  |  |
| --- | --- |
| Ritika Gupta | Cleaning and merging of Energy and static dataset according to the building IDs, timestamps, and Square Feet along with the merging of Weather dataset and Shiny App Creation |
| Tanu Rana | Performed Analysis based on the research questions and helped Cleaning the data for Energy, Static and Weather dataset, and Shiny App Creation |
| Rijul Chaturvedi | Cleaning and merging of Energy and static dataset according to the building IDs, timestamps, and Square Feet along with the merging of Weather dataset and Shiny App Creation |
| Aman Keskar | Performed Analysis based on the research questions and helped Cleaning the data for Energy, Static and Weather dataset, and Shiny App Creation |

**Shiny App:**

The Shiny app in question is an interactive and visually structured platform designed for the analysis of weather data. Utilizing the R programming language and various libraries such as tidyverse, ggplot2, and shiny dashboard, the application offers a polished user interface. The dashboard prominently features a title and a sidebar incorporating a dropdown menu facilitating the selection of a July and the rest of the year. The main functionalities of the app include the dynamic presentation of DataTable showcasing weather information specific to July and the rest of the months, a visually informative plot illustrating the difference between the energy consumption of heating and VAC elements and Total energy produced, and a concise summary denoting the maximum total information, including the associated timestamp. In essence, this Shiny app provides a sophisticated yet accessible tool for users to delve into and interpret weather data, serving as a valuable resource for meteorological analysis.

We have deployed our app on shinyapps.io:- <https://iwbpol-tanu-rana.shinyapps.io/DSProj/>

### RECOMMENDATIONS:

1. **Energy Consumption Factors**: Our analysis reveals various factors influencing energy consumption.
2. **eSC's Concerns**: Addressing eSC's concerns about the impact of global warming on electricity demand, especially during hot summers, requires a strategic approach.
3. **Targeted Approach**: Focusing on the highest energy usage month, July, allows for a targeted strategy to reduce energy consumption and mitigate the risk of blackouts.
4. **Environmental Sustainability**: By proactively managing energy demand, eSC aligns with environmental sustainability goals, positioning itself as a responsible energy provider.
5. **Holistic Approach**: The proposed measures collectively form a holistic approach, addressing immediate concerns and contributing to a more reliable and eco-friendly future.