Consulting Recommendations

Prepared for

Created by

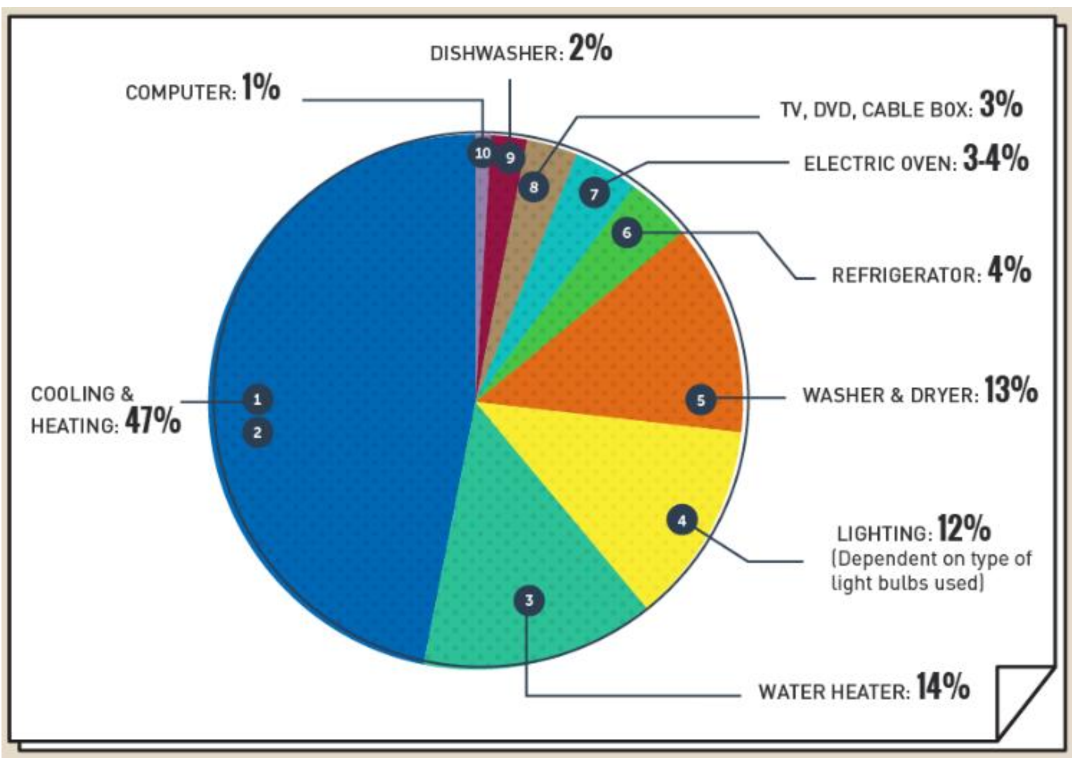
1. Introduction

As the premier and first green energy company to serve Texas, Green Mountain Energy understands accurately predicting residential energy consumption is crucial for optimizing resource allocation and promoting the adoption of sustainable solutions. This recommendation document outlines a solution that utilizes both linear regression and deep neural network models to predict residential energy consumption. By leveraging these advanced analytical techniques, Green Mountain can make informed decisions, improve operational efficiency, and support a greener future.

2. Data Collection

A comprehensive data collection process is crucial to understanding residential energy consumption. One of the main factors that influences energy usage is the weather. When it's hot, people tend to turn on their air conditioners, and when it's cold, they use heating systems. In fact, it is estimated that about 50% of residential energy consumption is due to heating and cooling, as I'm sure you are aware (see Figure 1 below). Therefore, our team decided to focus on weather conditions for our study.

Figure 1. General proportion of energy consumption by appliance in the home

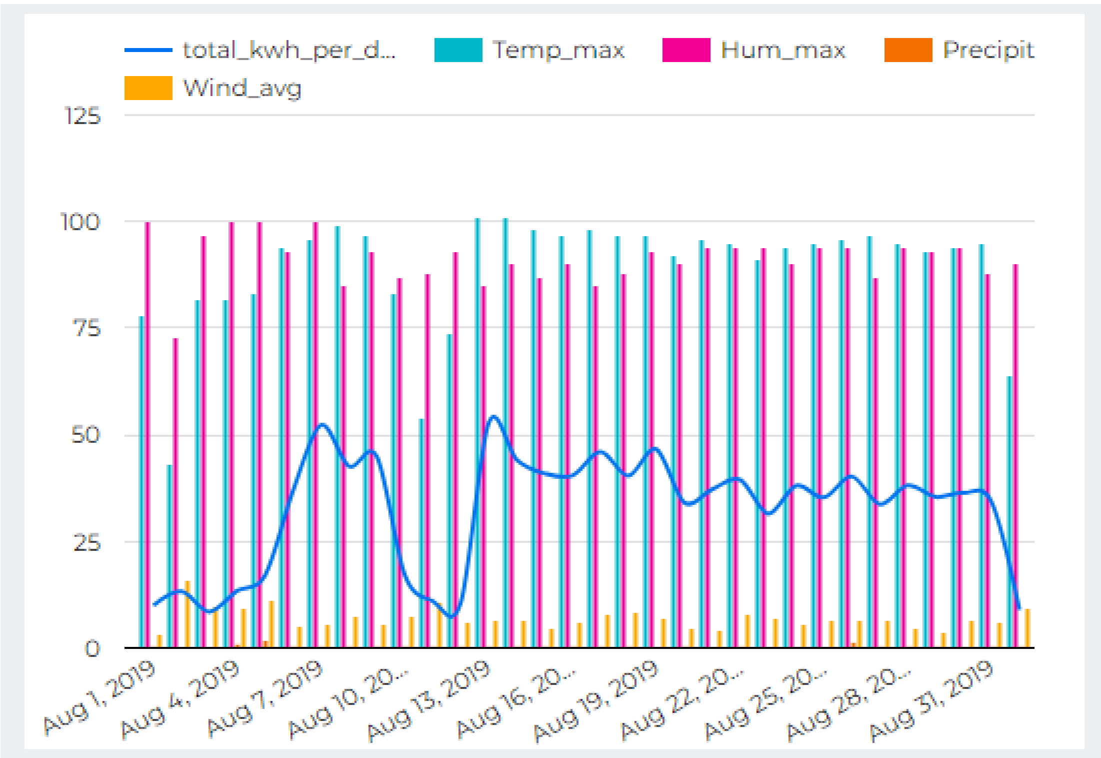


(*source: "What Uses the Most Energy in Your Home?" by Jeff Desjardins of the Visual Capitalist*)

To gather the necessary data, we collected historical weather readings including temperature, wind, and precipitation. We also obtained real-time data and forecasts to match with our collected sample of residential meter readings. By comparing weather patterns with energy usage, we aimed to identify any correlations or patterns between the two.

This approach allowed us to analyze the impact of weather on residential energy consumption more accurately. By considering weather factors alongside other variables, we can gain a better understanding of how different weather conditions affect energy usage in homes. In Figure 2 below we see the primary dimensions we used in developing our model, overlaid by the actual energy consumption we gathered over the course of a month.

Figure 2. Weather parameters vs. daily average kwh consumed, August 2019



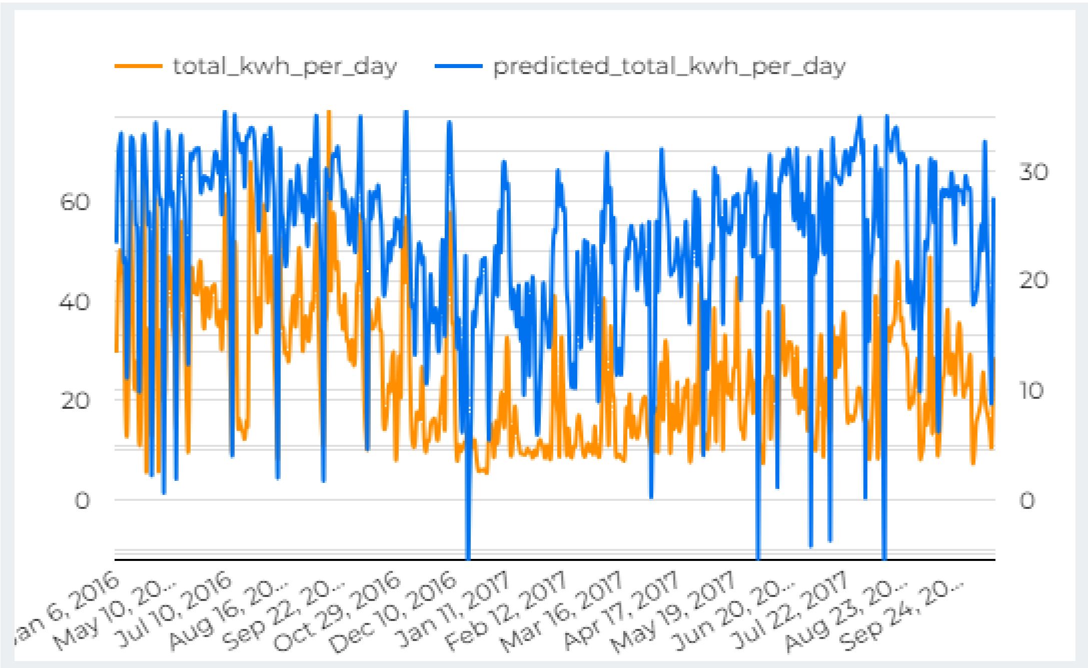
3. Data Preprocessing

Once the data was collected, thorough preprocessing was required to ensure data quality and compatibility across different sources. This involved cleaning the data, to include identifying and addressing outliers and anomalies; handling missing values; and data relational joining for a comprehensive dataset.

4. Linear Regression Model

Linear regression can serve as an initial predictive model for residential energy consumption due to its simplicity and interpretability. By our gathered data on weather conditions, we developed a linear regression model which could estimate energy consumption accurately. Below is an example of how the model predicted consumption versus our historical actuals (Figure 3).

Figure 3. Linear regression predicted consumption vs. actual January 2016 - September 2017

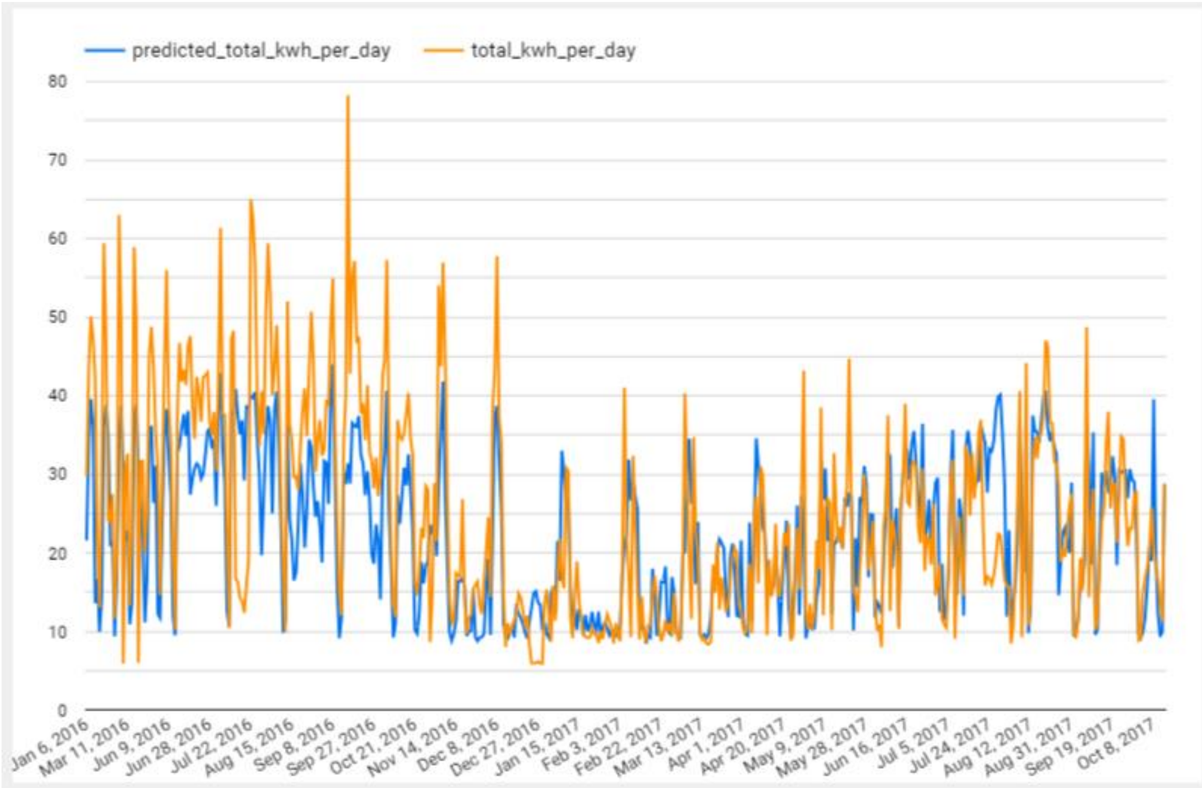


Overall, the model was a moderately good fit, although it was a big conservative when it came to the highs of energy consumption, a scalar adding to the constant for the model, may be beneficial for some greater cushion in action.

5. Deep Neural Network Model

To capture more complex relationships and nonlinear patterns, a deep neural network (DNN) model was developed. DNNs excel at handling large amounts of data and can automatically learn intricate features from the input variables. By utilizing multiple hidden layers, activation functions, and appropriate optimization algorithms, DNNs can effectively predict residential energy consumption with significant accuracy. Below (Figure 4) was the result for our DNN model, over the same time as our linear regression sample.

Figure 4. DNN predicted consumption vs. actual January 2016 - September 2017



6. Model Training and Evaluation

The dataset was randomly split into training and testing sets to evaluate the performance of both the linear regression and DNN models. The models were trained using various techniques such as cross-validation, regularization, and hyperparameter tuning to optimize their performance. Below Evaluation metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared can be used to compare and assess the models' predictive capabilities.

Figure 5. Evaluation metrics

Metric	Linear Regression	DNN
MAE (0 is best)	5.57	5.32
RMSE (0 is best)	8.11	7.77
R-squared (1 is best)	0.63	0.66

It should be noted that visually (in Figures 3 and 4) and by our evaluation metrics, the DNN model performs.

7. Insights and Recommendations:

The predictions generated by the linear regression and DNN models can provide valuable insights into residential energy consumption patterns. By analyzing the model outputs, Green Mountain can identify peak energy demand periods, target energy-saving initiatives, optimize resource allocation, and develop tailored energy efficiency programs for Houston, Texas, as well as the other regions you are involved in to include New York and Pennsylvania.

8. Ways Forward

Gaining access to Green Mountain's personal data and metrics, smart meters, and other sources can provide a more diverse and extensive dataset for training and testing the models.