Household-Level Day-Ahead Peak Electricity Demand Forecast

An extended research project report submitted to the University of Manchester for the degree of MSc Data Science in the Faculty of Humanities

2024

Student ID: [redacted]

School of Social Sciences

Table of Contents

Word Count: 7442

List of Figures

List of Tables

Abstract

This study develops and evaluates machine learning models for forecasting day-ahead peak electricity demand at the household level, addressing a gap in the literature by integrating diverse household-level data sources. Using the Low Carbon London project dataset and weather data, the research applies XGBoost models with various feature sets. The optimal model achieves a test WAPE of 29.63% with 15 features. Recent consumption patterns emerge as the most significant predictors, while household characteristics play a less prominent role than hypothesized. The study reveals that carefully selected subsets of features can achieve comparable performance to larger feature sets, suggesting potential for efficient, streamlined forecasting models. Limitations include underestimation of high consumption events and potential bias from opt-in recruitment. The findings have implications for data collection strategies and model development in the energy sector, emphasizing the importance of high-quality recent consumption data for short-term forecasting. Future research directions are proposed to address limitations and further refine forecasting techniques.

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Dr. Fanlin Meng, for his invaluable guidance and constant encouragement throughout the research and writing process. I am also grateful for the knowledge gained from his Python module, which has significantly contributed to my technical skills.

I would also like to thank my friends, both in the UK and abroad, who have been a constant source of support and motivation.

A special note of appreciation goes to my family and girlfriend. Their unwavering support and confidence in me were instrumental in the completion of this project.

Declaration

No portion of the work referred to in this extended research project report has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Intellectual Property Statement

1 INTRODUCTION

The global energy landscape is undergoing a drastic transformation as the world aims to achieve net-zero emissions by 2050 (United Nations, 2023). With fossil fuels comprising 84% of the global energy mix in 2020, the challenge to reduce this to less than 20% by 2050 to reach the net-zero emissions goal is monumental. This transition is further complicated by projections indicating a substantial increase in electricity consumption—approximately 50% by 2036 and more than doubling by 2050 in countries like the UK and US (National Grid, 2022).

Within this context, accurately forecasting peak electricity demand emerges as a crucial task for power system operators and policymakers. This forecasting capability has a direct impact on grid stability, economic efficiency, and environmental sustainability. Exceeding the power supply capacity during peak hours can lead to system breakdowns, which forces power companies to procure additional power at premium rates, ultimately resulting in higher tariffs for consumers (Mishra and Palanisamy, 2018). Moreover, during these peak times, peaking plants are often operated to meet demand. These peaking units, however, often run for only a short period of time, and have the highest variable operating costs (U.S. Energy Information Administration, 2012). In addition, they tend to utilize diesel and natural gas to operate, resulting in high carbon emissions and environmental degradation (Mishra and Palanisamy, 2018).

The transition to renewable energy sources adds another layer of complexity. While sources like wind, solar, and hydroelectric power have low operating costs, their output is variable and dependent on weather conditions, making accurate electricity demand forecasting even more critical (U.S. Energy Information Administration, 2012).

Given these challenges, this research aims to develop and evaluate machine learning models for forecasting day-ahead peak electricity demand at the household level, seeking to provide a granular understanding of peak demand drivers for more effective power system management and policymaking. The study leverages diverse data sources, including the Low Carbon London project dataset (Strbac et al., 2024) for historical electricity consumption, dynamic tariff information, and household survey data, as well as detailed weather data for London from Visual Crossing (2024).

This research addresses a gap in existing literature by integrating diverse household-level data sources. By incorporating household-specific information, including attitudes and behaviors of residents, pricing data, historical consumption, and weather data, this study aims to enhance the accuracy and interpretability of peak demand forecasts at the household level.

The structure of this paper is as follows: Section 2 reviews the literature on peak electricity demand forecasting. Section 3 describes the methodology used in this study, including data preprocessing, feature engineering, and model development. Section 4 presents and analyzes the results, comparing different models and discussing the implications of the findings. Finally, Section 5 concludes with a summary and suggestions for future research directions.

2 RELATED WORK

This review focuses primarily on short-term forecasting of peak electricity demand, while also covering medium-term and long-term forecasting. The emphasis on short-term forecasting reflects the nature and emphasis of this research project. The review covers definitions, time horizons, input variables, forecasting models, and evaluating metrics.

2.1 Definition of Peak Electricity Demand Forecasting

Peak electricity demand forecasting is a specialized type of load forecasting that focuses on predicting the maximum electricity consumption within a specified time frame. Unlike standard load forecasting, which aims to predict the entire load profile over a given period, peak demand forecasting targets the magnitude and timing of maximum electricity usage (McSharry et al., 2005). This distinction is crucial as peak demand often presents different challenges to power systems, requiring specialized forecasting approaches due to their volatile nature and susceptibility to extreme events or conditions (Dai et al., 2021).

2.2 Time Horizons in Peak Electricity Demand Forecasting

Peak electricity demand forecasting is typically categorized into three main time horizons:

- Short-term peak demand forecasting (STPDF): This generally covers a period from several hours up to one week ahead (Fan and Hyndman, 2011; McSharry et al., 2005). STPDF is crucial for day-to-day power system operations, helping utilities avoid blackouts and provide continuous power supply (Mughees et al., 2021).

- Medium-term peak demand forecasting (MTPDF): This typically spans from one week to one year (Wu et al., 2018; Yalcinoz and Eminoglu, 2005). MTPDF is particularly valuable for planning of electric utility assets such as power plants and generators, as well as optimizing resource allocation over extended periods (Lee and Hong, 2015).

- Long-term peak demand forecasting (LTPDF): This generally covers a period from one year to several years or even decades (Hyndman and Fan, 2010; Wu et al., 2018). LTPDF is integral to scheduling the construction of new generation facilities, developing transmission and distribution systems, and upgrading existing infrastructure (Abderrezak Laouafi et al., 2016).

It is worth noting that there is some variation in how different studies define these periods. For instance, McSharry et al. (2005) suggest that short-term forecasts cover five minutes to one week ahead, while Yalcinoz and Eminoglu (2005) define it as 30 minutes to one week ahead. These varying definitions reflect the diverse perspectives within the field of peak electricity demand forecasting.

2.3 Input Variables

The choice of input variables for peak demand forecasting models often depends on the forecast horizon. For example, season weather patterns and economic indicators play a significant role in MTPDF (Lee and Hong, 2015), while sociodemographic and macroeconomic variables like population, Consumer Price Index (CPI), and Gross Domestic Product (GDP) are crucial for LTPDF. As for STPDF, the focus of this paper, key input variables include:

- Weather variables: Temperature is crucial, as evident in numerous studies, including Fan and Hyndman (2011), Barochiner et al. (2022), and Yalcinoz and Eminoglu (2005), which incorporate temperature data in their models. Other weather features, such as windspeed, humidity, and precipitation, are also commonly used as predictors, as shown in studies by Fu et al. (2022), Alduailij et al. (2020), and Berrisch et al. (2023).
- Calendar variables: Variables capturing day of the week and holidays play a significant role (McSharry et al., 2005; Morales-Mareco et al., 2023; Abderrezak Laouafi et al., 2016; Amara-Ouali et al., 2023).
- Recent historical consumption data: This is utilized across all time horizons but particularly important for STPDF (Mughees et al., 2021; Fu et al., 2022; Barochiner et al., 2022).
- Household characteristics: Several studies have highlighted the effects of household characteristics and sociodemographic factors on electricity consumption, which are particularly relevant for household-level peak demand forecasting:
    - Jones et al. (2015) identified 62 factors potentially affecting domestic electricity use, categorized into socio-economic, dwelling, and appliance-related factors. They found positive correlations between electricity consumption and factors such as household income, number of occupants, and ownership of high-consumption devices.

- Frederiks et al. (2015) focused on individual-level predictors, including sociodemographic and psychological factors. They noted that while factors such as household size and income generally show positive relationships with energy consumption, psychological factors like knowledge, attitudes, and personal norms often have weak or inconsistent relationships with energy use.
- Fan et al. (2017) specifically addressed peak demand, finding that factors like air-conditioning ownership, number of occupants, swimming pool ownership, and clothes dryer usage were significant contributors to peak demand at the household level.

These studies strongly justify the use of household-level survey data in studying electricity consumption and peak demand. As Fan et al. (2017, p. 205) note, "A better understanding of the key drivers of residential peak electricity demand could assist in better managing peak demand growth through options including demand-side participation and energy efficiency programs."

Despite the availability of comprehensive datasets such as the Irish Smart Metering Electricity Customer Behaviour Trials (Commission for Energy Regulation, 2012) and the Low Carbon London project dataset (Strbac et al., 2024), the integration of household-level survey data with high-resolution consumption data for peak electricity demand forecasting remains underexplored. While recent studies have made significant contributions to the field of electricity demand forecasting, they do not focus on peak electricity demand forecasting. For instance, Mondal and Das (2023) focused on short-term demand forecasting using aggregated smart meter data, while Rausser et al. (2018) investigated the impact of smart meter installations on overall household electricity consumption in Ireland.

Notably, Rausser et al. (2018) utilized household-level data and survey responses. However, their study focused on general electricity consumption patterns and attitudes towards smart meters, rather than specifically addressing peak demand forecasting. Additionally, the feature set they employed, while valuable, was not as comprehensive as the one used in this study.

This gap presents an opportunity for developing more nuanced and accurate forecasting models that can better capture the complex interactions between various factors influencing peak electricity demand. A comprehensive approach that integrates household-level survey data, historical consumption data, weather data, and dynamic pricing data remains novel in the field of peak demand forecasting and could potentially fill this gap.

2.4 Evaluation Metrics

The performance of peak demand forecasting models is typically assessed using the following commonly used metrics:

- Mean Absolute Percentage Error (MAPE):

$$\text{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{1}$$

  where y_i is the actual value and $\hat{y}_i$ is the predicted value.

- Root Mean Square Error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{2}$$

- Mean Absolute Error (MAE):

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{3}$$

Some studies also focus on both peak value and peak value timing prediction. Fu et al. (2022) developed an approach to predict the probability of the next day being the monthly peak day and the probability of an hour being the peak hour of the day. Similarly, McSharry et al. (2005) developed a model for generating probabilistic forecasts of both the timing and magnitude of peak demand for lead times of one year.

In general, the error tends to be significantly higher for individual forecasts at the household level compared to aggregated forecasts. For example, Xia et al. (2024) report that load forecasting accuracy can decrease by a factor of 28 when moving from 1000 smart meters to a single smart meter. Goncalves Da Silva et al. (2014) observed in their studies that for a group of 50 consumers, the average individual MAPE was 44.62%, while the group's MAPE was just 10.09%. In a larger study with 183 participants, they found that individual MAPEs averaged 48.53%, whereas the group's MAPE averaged only 10.59%. This discrepancy highlights the challenges in forecasting at the individual level, where electricity demand across households can vary greatly.

2.5 Forecasting Models

Peak demand forecasting models can be broadly categorized into statistical methods, machine learning methods, and hybrid approaches.

Statistical methods, popular among earlier studies, include regression models, time series models such as Autoregressive Integrated Moving Average (ARIMA) and Seasonal Autoregressive Integrated Moving Average (SARIMA), exponential smoothing methods, and semi-parametric models. Fan and Hyndman (2011) proposed a semi-parametric additive model for day-ahead peak demand forecasting on a regional scale within the Australian National Electricity Market (NEM), achieving a MAPE of 1.88%. Similarly, Sigauke and Chikobvu (2011) developed a Regression-Seasonal AutoRegressive Integrated Moving Average-Generalized Autoregressive Conditional Heteroskedasticity (Regression-SARIMA-GARCH) model for predicting daily peak electricity demand forecasting in South Africa, achieving a Mean Absolute Percentage Error (MAPE) of 1.42%. In a different approach, McSharry et al. (2005) utilized a regression model with weather simulation for probabilistic forecasts, achieving a MAPE of 2.52%. These methods often perform well for aggregated forecasts and short-term forecasts, especially when relationships between variables are relatively linear.

Machine learning methods have gained popularity in recent years due to their ability to capture complex, non-linear relationships. Artificial Neural Networks (ANNs) have shown promise, with Khan et al. (2011) achieving MAPE values as low as 1.38% in autumn seasons for short-term daily peak load forecasting in the UK. Additionally, Gradient Boosting methods, particularly eXtreme Gradient Boosting (XGBoost), have proven effective, with Barochiner et al. (2022) outperforming staff predictions for 1-week-ahead forecasting (RMSE of 289 compared to 429 for human forecasters using a traditional similar day approach).

Deep Learning techniques have also made significant contributions. Long Short-Term Memory (LSTM) networks (Ai et al., 2019) and Bidirectional LSTM (Mughees et al., 2021) have shown great promise in capturing complex temporal patterns. Morales-Mareco et al. (2023) found that a Bidirectional Gated Recurrent Unit (BiGRU) model achieved the lowest RMSE (231.57 MW) and MAPE (8.1%) among compared methods, including statistical, machine learning, and other deep learning methods, for short-term forecasting in Paraguay.

Hybrid approaches, which combine multiple techniques to leverage their strengths, have also performed well. Fu et al. (2022) developed an ensemble model using Random Forest, Gradient Boosting Machine, and Logistic Regression, successfully capturing 69 out of 72 peak days in their test period. Amara-Ouali et al. (2023) proposed a multi-resolution approach combining Generalized Additive Models (GAMs) and Neural Networks (NNs), achieving a MAPE of 1.41% for peak demand forecasting.

Direct comparisons of results from different studies are challenging due to variations in data granularity, time horizons, and other dataset characteristics. Nevertheless, literature suggests that hybrid approaches and machine learning methods, particularly ensemble and deep learning approaches, have shown superior performance in peak demand forecasting across various contexts.

2.6 Addressing the Gap in Literature

This review of peak electricity demand forecasting literature highlights the field's evolution from traditional statistical methods to more sophisticated machine learning and hybrid approaches. These advancements have significantly improved the ability to predict peak electricity demand.

However, a significant gap remains in the comprehensive use of household-level survey data, particularly when combined with other data sources such as dynamic pricing and high-resolution weather data. While aggregated data is useful for capturing the impact of temperature and temporal factors, it does not provide reliable information about other potential driving factors such as household characteristics.

This gap is particularly important as the electricity sector continues to evolve with widespread adoption of renewable energy and the rollout of smart grid technologies. More accurate and granular peak demand forecasting could ensure grid stability, enable more efficient power system operations and better-targeted demand management strategies, ultimately leading to more sustainable and cost-effective electrical systems.

This study aims to address this gap by developing a comprehensive model that incorporates household-level survey data, historical electricity consumption data, weather information, and dynamic pricing data to forecast day-ahead peak electricity demand. By integrating these diverse data sources and applying advanced machine learning techniques (specifically XGBoost), this research seeks to provide new insights into the drivers of short-term peak electricity demand at the household level. This approach has the potential to not only improve forecast accuracy but also inform more effective demand management strategies.

The following methodology section will discuss key research questions and detail the approach to data preprocessing, feature engineering, and model development and evaluation.

3 METHODOLOGY

This study aims to develop and evaluate machine learning models for forecasting day-ahead peak electricity demand at the household level, focusing on the key drivers of peak demand.

It addresses the following research questions:

- How accurately can machine learning models predict day-ahead peak electricity demand at the household level?
- What are the most significant factors influencing this demand?
- How does the inclusion of household-specific survey data and dynamic pricing information impact forecast accuracy?

To investigate these questions, this study employs a comprehensive methodology involving data preprocessing, feature engineering, and advanced machine learning techniques. The following section describes the data sources, preprocessing steps, feature engineering approaches, and modeling techniques used in the study. All code needed to reproduce this study is available in the GitHub repository shared as part of the additional materials submission.[1]

3.1 Data Sources and Description

This study utilizes two primary data sources: the Low Carbon London (LCL) project dataset (Strbac et al., 2024) and London weather data for 2013 (Visual Crossing, 2024).

3.1.1 Low Carbon London Project Dataset

The Low Carbon London project was a £28m research program led by UK Power Networks that investigated the impact of low carbon technologies on London's electricity distribution network. As part of this project, the UK's first residential dynamic time-of-use (dToU) electricity pricing trial was conducted in 2013.

This trial involved 5,567 households in the London area, divided into two groups. Group D consisted of 1,122 households who opted in to receive the experimental dToU tariff and completed two surveys: an appliance survey focused on household characteristics and appliance ownership, and an attitude survey assessing energy consumption behaviors and attitudes towards dToU pricing. The remaining 4,445 households remained on their static tariffs and only completed the appliance survey.

[1] https://github.com/sthant3/electrical-peak-demand-forecast

The dataset, curated by researchers from Imperial College London, consists of half-hourly smart meter readings for both groups, dToU tariff details for Group D, and household survey data. The data is available through the UK Data Service (Study Number 7857), released under their Safeguarded Access protocol.[2]

For this research, only Group D's data was used. The choice was made to focus on households that were exposed to dynamic pricing and completed both surveys, enabling a more thorough analysis of the factors influencing peak electricity demand.

3.1.2 London Weather Data

The weather data for this study was obtained through Visual Crossing, a platform that aggregates and processes weather information from multiple authoritative sources, such as the Integrated Surface Database (ISD), National Oceanic and Atmospheric Administration (NOAA), and the UK Met Office (Visual Crossing, 2024).

The weather dataset includes an hourly record for London throughout 2013, aligning with the period of the LCL project. It includes variables such as temperature, humidity, wind speed, and precipitation.

3.2 Data Preprocessing

3.2.1 Consumption, Tariff, and Weather Data Processing

The consumption and tariff data, initially recorded at 30-minute intervals, was aggregated to hourly granularity to align with the weather data. Changing the time granularity from 30 minutes to 1 hour for consumption required summing up the half-hourly values, necessitating imputation before merging. The distribution of missing data in these half-hourly values revealed a clear divide between households with minimal missing data (less than 5%) and those with substantial missing data (over 20%). To balance data quality with sample size, a threshold of 5% was chosen. Consequently, households with more than 5% of consumption values missing were removed from the dataset. Remaining missing values (0.05% of the dataset) were imputed using interpolation.

Weather data for London throughout 2013 was consolidated from multiple time periods, retaining key variables like solar radiation, wind speed, temperature, precipitation, and humidity. No missing values were present in the dataset, and an anomaly related to daylight savings time was addressed through interpolation.

[2] Dataset is available at: https://beta.ukdataservice.ac.uk/datacatalogue/doi/?id=7857

3.2.2 Initial Survey Data Cleaning and Filtering

The survey data was cleaned and filtered, retaining 166 of the 344 initial questions most relevant to household characteristics, energy-related behaviors, and attitudes. Individual comments and less pertinent questions were removed.

Misalignments between question numbers in the answer and question sheets were corrected to ensure accurate data mapping. A notable issue arose with key appliance questions, where the option "Our household DOES NOT HAVE this appliance" was present in the question but never recorded in the answer sheet. This ambiguity made it difficult to distinguish between non-responses and non-ownership, leading to the exclusion of appliance flexibility-related features. In cases where inconsistencies could not be resolved due to missing columns for certain answers, the affected data was removed to ensure the integrity of the dataset.

Lastly, households with no responses to either the attitudes or appliances survey, those with over 20% non-response rate, and those with completely missing or inconsistent demographic information were removed. This process narrowed the list of households suitable for analysis to 608.

3.2.3 Feature Extraction and Addressing Missing Values

Due to the complex nature of the survey responses, it was necessary to extract features from the responses before merging the survey data with other datasets. The survey questions often required aggregation and transformation to create meaningful features. For example, questions like “Household Member 1 Gender” and “Household Member 2 Gender” were aggregated into “male_occupants” and “female_occupants.” Similarly, questions such as “TV Size 1,” “TV Size 2,” “TV Type 1,” and “TV Type 2” were aggregated into more informative features such as “count_tv” and “tv_energy_score.”[3]

Missing values were addressed next. Insulation-related features, with at least 25% of the values missing, were removed. Experimentation revealed that these features also did not improve model performance. The remaining features had their missing values imputed using the Iterative Imputer method, chosen for its ability to handle multiple variable types and complex relationships in the data. To validate the imputation process, histograms, box plots, and summary statistics were generated for each imputed variable, comparing distributions before and after imputation.

[3] Additional information on these features is reported in Table 2 of the Appendix

Fifteen households with extreme values in household characteristics were identified and corrected. These included implausible numbers of rooms, bedrooms, high-efficiency bulbs, and fridges/freezers. Adjustments were made assuming data entry errors to maintain data integrity.

3.2.4 Data Integration

Following feature extraction, the datasets were merged to create a comprehensive dataset that combined hourly electricity consumption data, dynamic tariff data, extracted features from the household surveys, and hourly weather data. Each consumption data point was associated with corresponding tariff and weather conditions, while household characteristics remained constant across timestamps.

After completing the data preprocessing steps, an exploratory data analysis was conducted to gain insights into the data patterns and relationships.[4]

3.3 Feature Engineering

Additional feature engineering was necessary to create a richer set of predictors for the modeling phase. This process involved creating new features at the hourly level, aggregating to daily granularity, and finally dropping the less relevant hourly features.

3.3.1 Temporal Features

Several temporal features were created to capture time-based patterns. Day of the week and week of the month were cyclically encoded using since and cosine transformations, preserving their circular nature. Additionally, binary features "is_weekend" and "is_holiday" were created to account for potential shifts in consumption patterns during those periods. Finally, one-hot encoded "season" variable was created.

3.3.2 Weather Features

New weather-related features were introduced to capture both the overall weather conditions of each day and the specific conditions during peak electricity usage. For temperature, humidity, and wind speed, the daily minimum, maximum, and mean values were calculated, along with the value at the time of peak consumption. Precipitation and solar radiation were represented by their daily sum and the value at peak consumption time.

[4] For key visualizations from the exploratory data analysis, see Figures 8-13 in the Appendix

3.3.3 Consumption-related Features

Features engineered from historical consumption patterns included previous day's and previous week's same-day peak consumption. Rolling averages and maximum values of peak consumption over 3-day and 7-day windows were calculated to capture short-term trends. Furthermore, the current day's total consumption was included, alongside its minimum and peak consumption, to provide a full picture of each household's electricity usage patterns leading up to the prediction day.

3.3.4 Price-related Features

Features related to dynamic pricing included the proportion of hours with low, normal, and high tariff rates for each day, as well as the tariff rate at the time of peak consumption. Other features such as “had_high_rate” and “avg_high_rate_duration” were initially created but were later excluded due to their negative impact on model performance.

3.3.5 Aggregation to Daily Granularity

After creating these features at the hourly level, the data was aggregated to daily granularity. The daily peak electricity consumption value for each day was extracted to serve as the target variable for the day-ahead prediction. The average (mean) of these daily peak electricity consumption across all households over time is shown in Figure 1. Variables were categorized as numerical, categorical, binary, or count, and their data types were optimized accordingly to improve memory usage and computational efficiency. This process significantly reduced dataset size while retaining key information for daily peak demand forecasting.

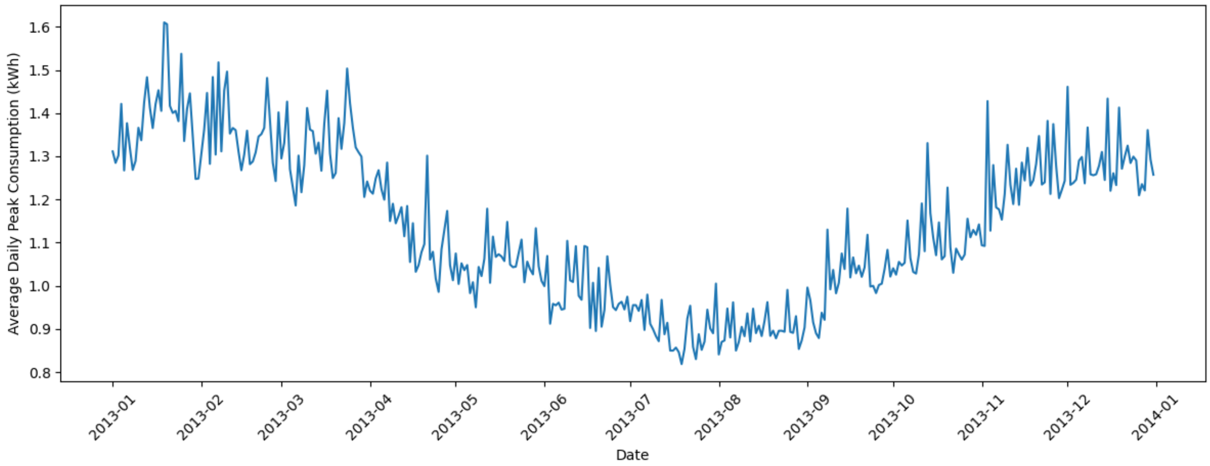

Figure 1: Mean daily peak electricity consumption across all households over time

3.4 Feature Selection

For feature selection, a two-step approach was employed:

- Correlation analysis: Highly correlated numerical features (correlation coefficient > 0.8) were identified. Some were removed while others were retained based on their perceived importance to the prediction task based on domain knowledge. Experimentation showed that this process improved model performance.
- Mutual information analysis: Remaining features were ranked based on their mutual information with the target variable. However, no features were dropped to retain potentially valuable information, allowing the subsequent modeling process to determine feature importance instead.

While Principal Component Analysis (PCA) was considered for further dimensionality reduction, it was not implemented to maintain the interpretability of individual features in the final model.

3.5 Model Development

3.5.1 Model Selection

XGBoost was chosen for its ability to handle non-linear relationships, complex feature interactions, and computational efficiency, while providing interpretable results through feature importance rankings.

While other methods were considered, they were not selected due to various reasons. Machine learning models such as RandomForest and LSTM networks, while robust, require much higher computational power, which this study lacked. As for traditional time series models, they were not chosen as they are primarily designed for univariate time series forecasting tasks and may struggle to handle the wide range of features included in this study. Additionally, they often assume linear relationships among variables, which is not the case for electricity demand. Therefore, XGBoost emerged as the most appropriate choice, balancing performance, interpretability, and computational feasibility.

A series of models were developed to explore different feature sets:

- Model 1: Temporal and consumption features
- Model 2: Model 1 features + weather features

- Model 3: Model 2 features + household, pricing, attitudinal, behavioral, and appliance features
- Model 4: Top 15 features from Model 3
- Model 5: Model 3 features + newly aggregated household features + household features used for aggregation
  These new features included a heating control score, total appliance count, and schedule rigidity score, among others, derived from existing household characteristics.[5]
- Model 6: Model 3 features + only newly aggregated household features
- Model 7: Top 15 features from Model 6

Model 5-7 were developed based on insights from previous models' results, and the best model was selected for further analysis. The details will be discussed in the Results section.

3.5.2 Model Training and Validation

The dataset was split using a time series approach, with three progressive folds for cross-validation spanning January to November 2013:

- Fold 1: Train (2013-01-08 to 2013-08-31), Validate (2013-09-01 to 2013-09-30)
- Fold 2: Train (2013-01-08 to 2013-09-30), Validate (2013-10-01 to 2013-10-31)
- Fold 3: Train (2013-01-08 to 2013-10-31), Validate (2013-11-01 to 2013-11-30)

The final test set was reserved for December 2013 (2013-12-01 to 2013-12-30). Time series cross-validation was chosen to maintain the temporal order of the data. Ideally, the test period would cover different seasons throughout the year to assess the model's performance across different time periods. However, due to the limitation of only having 2013 data, the folds were structured to maximize training data while still allowing for multiple validation periods.

Hyperparameter optimization was conducted using Random Search with time series cross-validation. For Model 5-7, the search space was refined based on the performance of earlier models, and the number of iterations was increased from 100 to 200 to explore the hyperparameter space more thoroughly. Key hyperparameters tuned included the number of estimators, maximum tree depth, learning rate, subsample ratio, and regularization terms.[6]

Models were trained to predict day-ahead peak consumption at the individual household level. As the target variable had a right-skewed distribution, shown in Figure 2 along with the

[5] View Table 3 of the Appendix for more information about the aggregated household features
[6] View Table 4 of the Appendix for detail on the hyperparameters

transformed distributions (log, square root, and cube root), the process was repeated using different target variable transformations. Each transformation was tested with and without a StandardScaler, and cube root transformation without a StandardScaler emerged as the best approach.

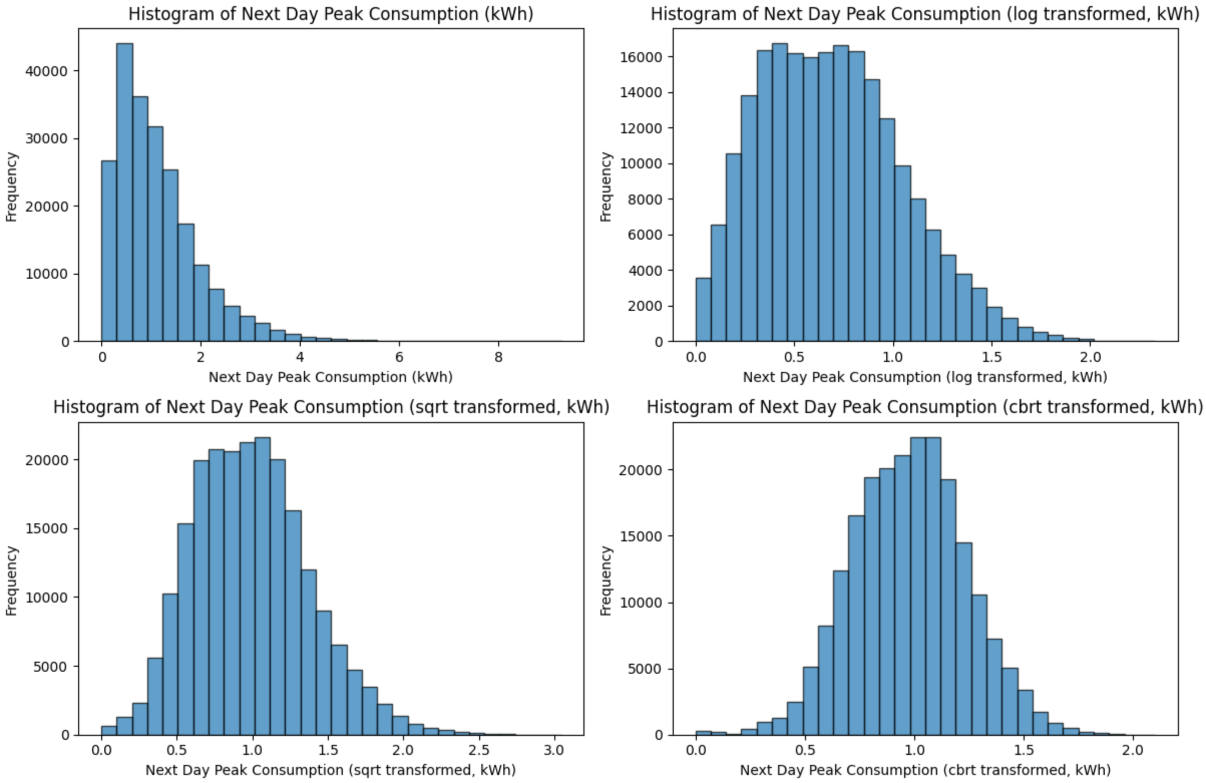

Figure 2: Distribution of next day's (day-ahead) peak consumption (different transformations applied)

3.5.3 Model Evaluation Metrics

The primary evaluation metric chosen for this study was the Weighted Absolute Percentage Error (WAPE). WAPE is defined as:

$$\mathrm{WAPE} = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{\sum_{i=1}^{n} |y_i|} \times 100\% \qquad (4)$$

where y_i is the actual value and $\hat{y}_i$ is the predicted value.

WAPE was selected over MAPE due to its robustness in handling values close to zero. Unlike MAPE, which can become inflated with many values close to zero, WAPE provides a more stable metric by weighting the errors by the magnitude of the actual values, which is particularly suitable for models predicting electricity consumption, where values can approach zero during low-usage periods.

While WAPE served as the primary metric, additional metrics such as MAE, MAPE, and R^2 were also reported to provide a comprehensive evaluation of each model's performance.[7]

[7] View Table 5 in the Appendix for a comprehensive model evaluation

4 RESULTS AND ANALYSIS

This section presents the results and analysis of the day-ahead peak electricity demand forecasting models developed in this study. While all models will be briefly discussed for an initial comparison, the focus will be on Model 7, selected as the optimal model based on the balance of performance, interpretability, and efficiency.

4.1 Model Performance Overview

Initial analysis using Models 1-4 revealed that Model 4, using the cube root transformation of the target variable without StandardScaler, offered the best balance of performance and complexity. Feature importance analysis of these models showed that individual household characteristics had relatively low importance, leading to the creation of additional aggregated household features such as “electric_heating_score”. This approach aimed to explore whether further aggregation of household features could capture broader patterns in household characteristics and enhance model performance.

Based on this finding, Models 5-7 were trained using the cube root transformation without scaling and incorporating aggregated household features. Table 1 presents the performance comparison of all models using this approach.

Table 1: Model performance comparison (cube root transformation without scaling)

Model	Number of Features	Average Training WAPE	Average Validation WAPE	Test WAPE
Model 1 (Temporal and consumption features)	18	29.43%	31.33%	29.94%
Model 2 (Model 1 Features + Weather Features)	30	29.38%	31.38%	30.17%
Model 3 (Model 2 Features + Survey Features)	88	26.66%	30.22%	29.27%
Model 4 (Model 3's Top 15 Features)	15	28.16%	31.01%	29.65%
Model 5 (Model 3 + Aggregated Household Features + Features Used for Aggregation)	94	26.59%	30.18%	29.23%
Model 6 (Model 5 Without Features Used for Aggregation)	57	26.99%	30.44%	29.44%
Model 7 (Model 6's Top 15 Features)	15	27.82%	30.92%	29.63%

Going from Model 2 to Model 3 led to a notable improvement in performance, with the test WAPE decreasing from 31.07% to 29.27% This indicates the addition of these features likely captures important household-specific patterns and behaviors that influence peak electricity demand, providing the model with valuable information not offered by temporal and weather data alone.

Model 5 emerged as the best-performing model, slightly outperforming the second-best model, Model 3, by 0.04 percentage points (29.23% vs 29.27%) in terms of test WAPE. Model 5, which incorporates aggregated household features in addition to the features used in Model 3, demonstrates that these aggregated features provide a marginal improvement in predictive accuracy.

However, after considering both the performance metrics and model complexity, Model 7 was selected as the optimal choice. Despite using only 15 features compared to Model 5’s 94, Model 7 achieved a competitive test WAPE of 29.63%. This demonstrates that a carefully selected subset of features can achieve comparable performance to that of a much larger feature set, adhering to the principle of Occam's Razor. The limited feature set also enhances model efficiency and interpretability, which is crucial for practical application and stakeholder communication. While Model 7 shows some variability between its validation WAPE (30.92%) and test WAPE (29.63%), this difference is not substantial enough to negate its other advantages.

Given this decision, the subsequent sections will provide a comprehensive analysis of Model 7, focusing on model performance, feature importance, limitations and potential improvements, and implications for peak electricity demand forecasting and energy system management.

4.2 Analysis of Best Model

4.2.1 Model Performance

Model 7 achieved a test WAPE of 29.63%, a reasonable performance given the challenges of individual household-level peak demand forecasting, which Xia et al. (2024) note is challenging due to the randomness of consumer behavior. In this context, the model’s performance is noteworthy, especially considering it had only 2013 data, which limits the model’s exposure to seasonal variations and long-term trends.

Figure 3 provides the distribution of prediction errors, which offers additional insights into the model's performance. The errors follow a roughly normal distribution with a slight right skew, indicating the model's tendency to underestimate peak demand in some cases. This skew suggests that while the model performs well on average, it may struggle with certain extreme cases. This tendency to underestimate the peak demand, especially during high-consumption periods, could be a concern for grid management, where underestimation is usually more problematic than overestimation.

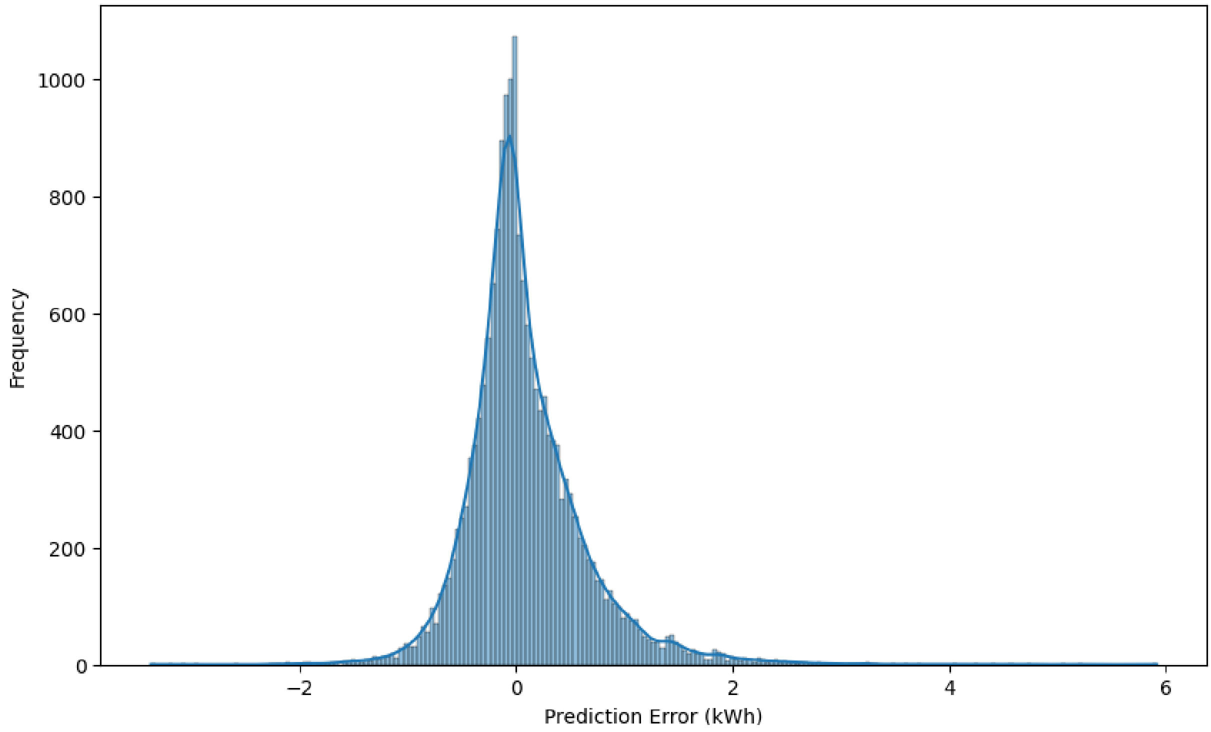

Figure 3: Distribution of prediction errors of Model 7

To further illustrate the model's performance, Figure 4 presents a scatter plot of actual versus predicted peak consumption values. This visualization reveals a clear positive correlation between actual and predicted values, indicating that the model captures the general trend effectively. However, it also confirms the model's tendency to underestimate high consumption values, as evidenced by the majority of points below the diagonal line for higher actual consumption, starting from around 2 kWh.

Moreover, the model's predictions seem to be mostly confined to a range of about 0-5 kWh despite actual values extending beyond 8 kWh, further illustrating its limitation in capturing high peak consumption events. This could indicate the model is heavily reliant on average consumption patterns and fails to account for factors driving unusually high peak demand.

Furthermore, the increasing scatter at higher consumption levels implies growing uncertainty in predictions for high values. This could be due to inherent variability in high-consumption events or the absence of relevant predictors for distinguishing between moderate and extreme peak demand drivers. In either case, this limitation is worrying in the context of grid stability and capacity planning, where accurate prediction of peak demand is crucial for ensuring a stable power supply.

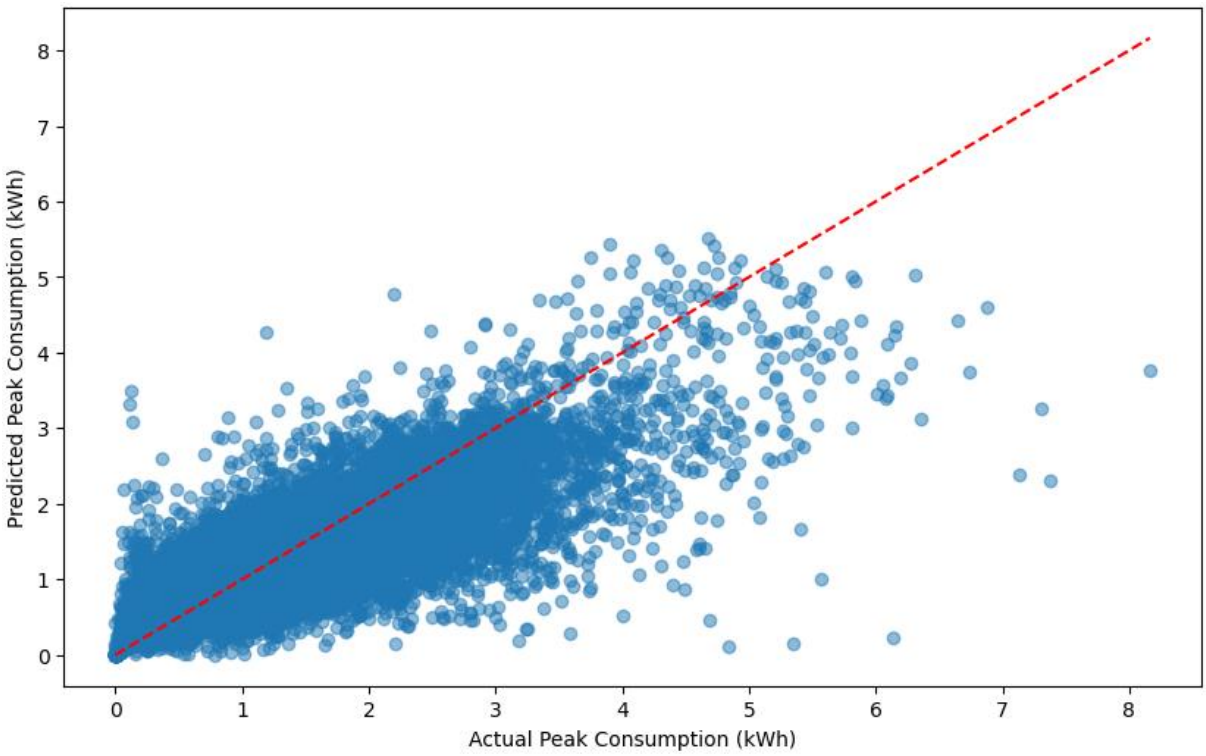

Figure 4: Scatter plot of actual versus predicted peak electricity consumption values of Model 7

Test set predictions for three randomly chosen households, shown in Figure 5, show Model 7's performance across different consumption levels. In all three cases, the model manages to capture the general trend, especially for households D0115 and D0393. However, the model also often fails to capture abrupt peaks and valleys. Notably, the model tends to underestimate high values, aligning with previous observations. This tendency is especially noticeable in predictions for D0257, which had a lot of high values, including sudden peaks. This further indicates the model's limitation in predicting high consumption events.

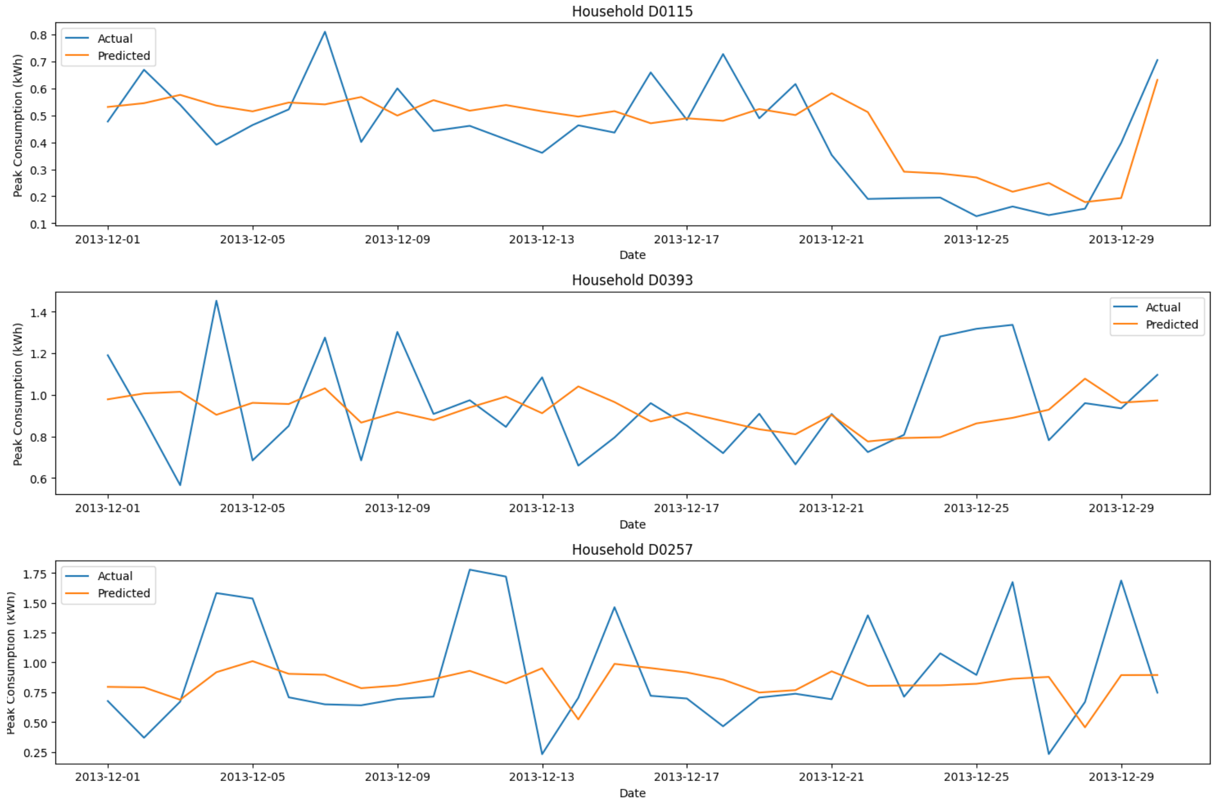

Figure 5: Actual vs predicted daily peak electricity consumption values of Model 7 for three randomly selected households

Despite these consistent tendencies in the model's performance, it is important to note the households still show widely varied consumption patterns. This observation highlights the complexity of forecasting peak demand at the household level. It suggests that improvements in forecasting accuracy might be achieved through strategies such as household segmentation/clustering or the development of separate models for different consumption levels. These approaches could potentially address the model's current limitations in capturing extreme events and account for varying consumption patterns across households.

The learning curve for Model 7, shown in Figure 6, reveals an interesting dynamic in model performance as the training set size increases. The validation WAPE shows significant improvement throughout the process, while Training WAPE stays relatively stable. This stability in training WAPE suggests that the model quickly reaches its capacity to fit the training data, indicating that the chosen model complexity is appropriate for this task.

The persistent gap between the training and validation WAPE indicates some degree of overfitting, an occurrence observed in all models explored. However, the gradual convergence of these curves suggests that additional data might bridge the gap further, improving generalization. This observation highlights a key limitation of this study - only using one year of data. A longer time series could potentially lead to better model performance and generalization. However, due to limited data availability, this possibility could not be tested.

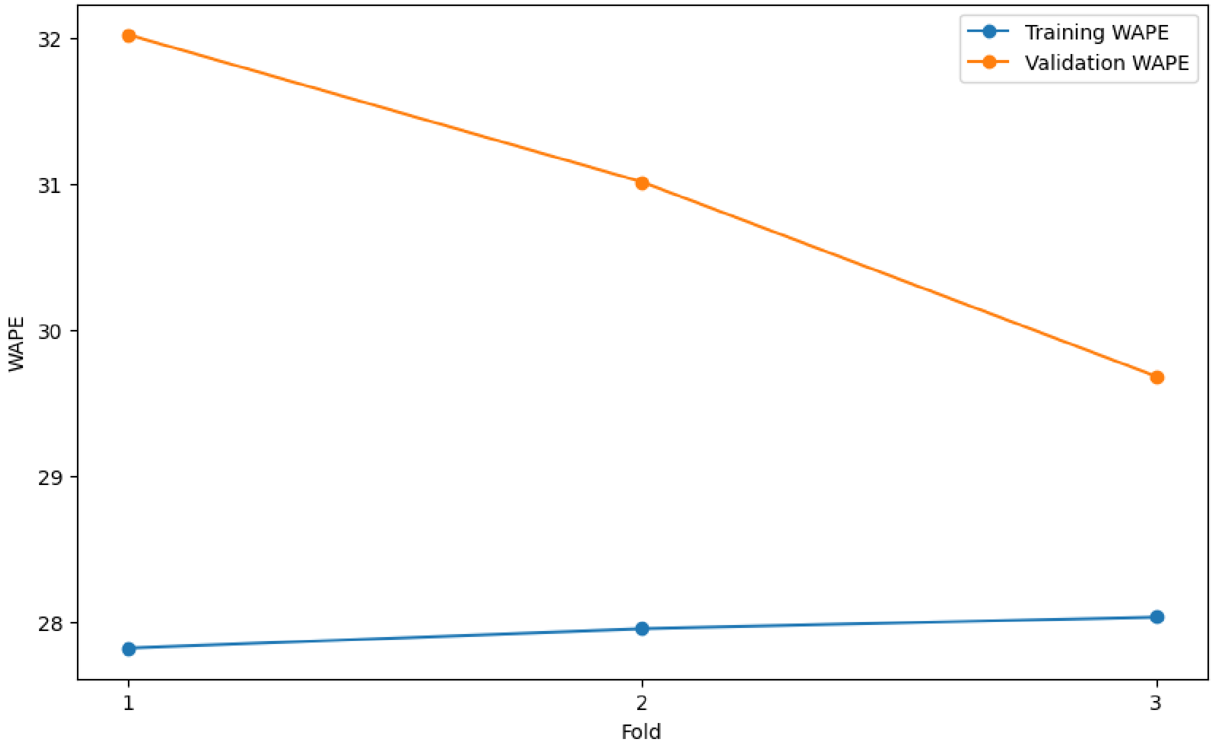

Figure 6: Learning curve of Model 7

Having examined the model's performance, the next step involves analyzing the features that contribute the most to these predictions. The following section on feature importance provides insights into the key drivers of day-ahead peak electricity demand at the household level.

4.2.2 Feature Importance

Feature importances for Model 7, illustrated in Figure 7, highlight the dominance of recent consumption patterns in predicting day-ahead peak demand, a pattern consistent across all models. The top three features – "avg_peak_7d", "avg_peak_3d", and "consumption_peak_today" - account for nearly 80% of the total feature importance, indicating that short-term historical consumption data is by far the most crucial predictor.

Surprisingly, household characteristics like "electric_heating_score," "total_appliance_count," and demographic factors such as "male_occupants" have low importance scores, all below 0.02. This indicates that household and demographic features may not be as predictive as anticipated.

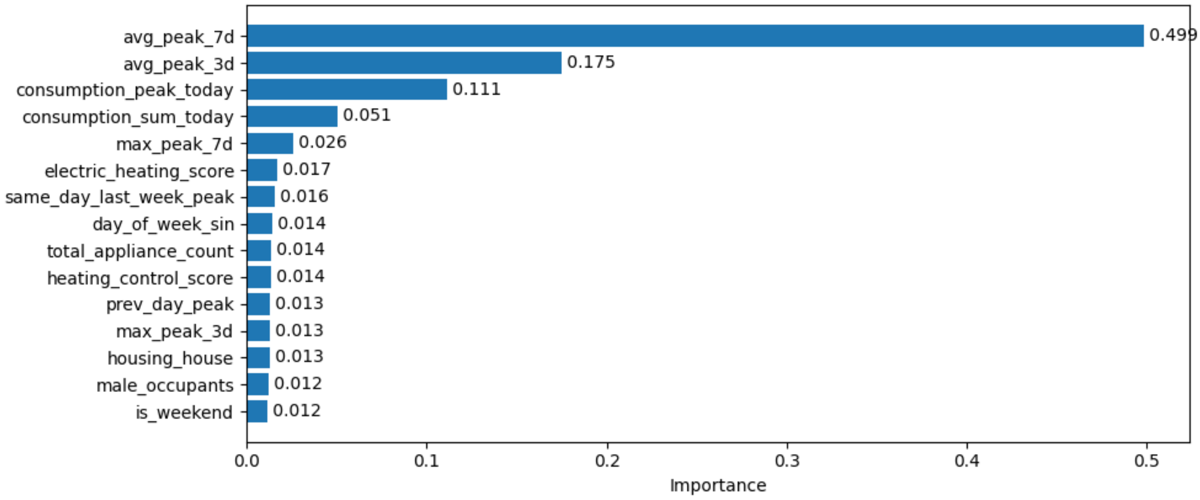

Figure 7: Feature importances of Model 7

While these results might seem disappointing in terms of the limited impact of household characteristics, they provide valuable insights. The strong predictive power of recent consumption features suggests that households tend to have consistent short-term energy use behaviors. Moreover, the model's heavy reliance on recent consumption patterns might make it more adaptable to changing household behaviors over time, as it does not heavily depend on household characteristics.

This adaptability is particularly valuable in the context of this study, where 79.2% of households in the sample are rented. In environments with a high proportion of renters, frequent changes in tenancy can lead to significant variations in electricity consumption patterns. The model's focus on recent consumption data allows it to adjust quickly to these changes, crucial for maintaining forecast accuracy in a dynamic environment with a high proportion of renters and high tenant turnover rates.

These findings have significant implications for peak demand forecasting strategies. They suggest that the collection of detailed household data may not be essential for short-term peak demand prediction, potentially reducing data collection costs and privacy concerns. Instead, the emphasis on recent consumption data appears to offer a more effective approach. For example, forecasting systems could prioritize the collection and processing of high-quality consumption data over extensive household surveys. This approach aligns with the dynamic nature of household energy consumption and offers practical advantages in data collection efficiency and model adaptability.

4.2.3 Model Limitations and Potential Improvements

While Model 7 demonstrates promising performance in forecasting day-ahead peak electricity demand at the household level, it has limitations and room for improvement. A significant constraint is the limited seasonal coverage. As the model was trained and tested on data from only 2013, it may struggle to generalize across different years and capture long-term trends in electricity consumption patterns. Additionally, while the model was adept at capturing the general trend in its predictions, the model tends to underestimate peak demand during high consumption events, as observed in the scatter plot and individual household predictions.

Another limitation is the potential bias introduced by the opt-in recruitment of Group D households. Participants who voluntarily chose to be part of the dToU program may not be representative of the general population. These households might be more energy-conscious, more willing to adjust their consumption patterns, or more interested in potential cost savings than the average household. This self-selection bias could affect the generalizability of the findings, as the model's performance might be influenced by the characteristics of this potentially more engaged and flexible group of consumers.

To improve the model's performance and generalizability, several approaches could be considered in future research. Considering the diverse consumption patterns observed across households, the use of clustering methods and development of separate models for different household types may potentially address some of these limitations. Moreover, although XGBoost performed well in this study, exploring other advanced machine learning techniques could enhance model performance. Deep learning models, such as LSTM networks, might better capture long-term dependencies in the time series data. Hybrid approaches that leverage the strengths of different models (e.g., SARIMA and ANN) could also be considered. Furthermore, implementing ensemble methods that combine predictions from multiple models optimized for different electricity consumption ranges could potentially improve model accuracy, especially for high consumption events.

Due to computational constraints, testing other feature engineering methods and alternative models was not feasible in this study. These limitations, along with the need to validate findings with a more diverse, randomly selected group of households, present opportunities for future research with more resources. Such research could help validate the broader applicability of these findings and lead to better, more generalized models for peak electricity demand forecasting at the household level.

4.2.4 Implications of the Research

Despite the limitations, this study's findings have key implications for peak electricity demand forecasting and energy system management. The strong predictive power of recent consumption patterns suggests that high-quality, recent historical consumption data should be prioritized in data collection efforts for short-term forecasting. The model's strong reliance on recent consumption data makes it potentially more adaptable to changing household behaviors, which is particularly valuable in areas with high tenant turnover rates.

While household characteristics were not strong predictors of peak demand in this study, their importance for medium-term and long-term remains unclear given this study's focus on short-term forecasting. Further research would be needed to determine their relevance in longer-term prediction models.

Nonetheless, this study offers insights into effective feature selection for peak electricity demand forecasting. It demonstrates that a carefully selected subset of features can achieve comparable performance to larger feature sets, highlighting the potential for efficient, streamlined models without significant accuracy loss. These implications are especially relevant in resource-constrained contexts and provide valuable insights for energy providers, policymakers, and researchers working towards more efficient and sustainable energy systems.

These findings highlight the potential of data-driven, household-level approaches in managing peak electricity demand and optimizing grid operations. However, the potential bias towards more engaged customers in this study's sample necessitates caution when interpreting these results and generalizing these strategies. Future research should address this limitation, along with the other limitations identified, to further refine these forecasting approaches, potentially leading to more robust, generalizable findings.

5 CONCLUSIONS AND FUTURE DIRECTIONS

This study has developed and evaluated machine learning models for forecasting day-ahead peak electricity demand at the household level, focusing on the key drivers of peak demand. The research demonstrated that machine learning models, particularly XGBoost, can predict household-level peak demand with reasonable accuracy, with the optimal model achieving a test WAPE of 29.63% with only 15 features. Notably, this model outperformed the model that only used temporal, consumption, and weather features (Model 2), achieving a test WAPE of 29.63% compared to 30.17%, despite using a smaller feature set (15 features versus 30). These findings indicate that including household characteristics results in a marginal but meaningful increase in model performance.

The study addressed a significant gap in the literature by integrating diverse data sources at the household level, including survey data, dynamic pricing information, and high-resolution weather data. The research revealed that recent consumption patterns are the most significant factors influencing day-ahead peak demand, while household characteristics and survey data play a less prominent role than initially hypothesized. This finding has important implications for data collection strategies and model development in the energy sector, suggesting that high-quality, recent consumption data should be prioritized over extensive household surveys for short-term forecasting.

While the model shows promise, the study's limitations, including the underestimation of high consumption events, limited seasonal coverage, and opt-in nature of the participant group, highlight areas for future research. Future studies could address these limitations by exploring advanced feature engineering techniques, implementing household segmentation, and investigating hybrid modeling approaches. Moreover, future research should strive to include randomly selected households and reduce the potential bias in the sample. This approach would provide a more representative view of household electricity consumption levels and potentially lead to more generalizable models and insights.

By addressing these limitations and building on the insights gained from this study, future research can further refine short-term peak electricity forecasting techniques. This advancement could ultimately contribute to the development of more efficient, sustainable, and widely applicable energy systems, benefiting energy providers, policymakers, and consumers alike.

6 REFERENCES

Abderrezak Laouafi, Mourad Mordjaoui, Farida Laouafi and Taqiy Eddine Boukelia (2016). Daily peak electricity demand forecasting based on an adaptive hybrid two-stage methodology. *International Journal of Electrical Power & Energy Systems*, [online] 77, pp.136–144. doi:https://doi.org/10.1016/j.ijepes.2015.11.046.

Alduailij, M.A., Petri, I., Rana, O., Alduailij, M.A. and Aldawood, A.S. (2020). Forecasting Peak Energy Demand for Smart Buildings. *The Journal of Supercomputing*, [online] 77(6), pp.6356–6380. doi:https://doi.org/10.1007/s11227-020-03540-3.

Barochiner, D., Lado, R., Carletti, L. and Pintar, F. (2022). A Machine Learning Approach to Address 1-week-ahead Peak Demand Forecasting Using the XGBoost Algorithm. *2022 IEEE Biennial Congress of Argentina (ARGENCON)*. [online] doi:https://doi.org/10.1109/argencon55245.2022.9939986.

Berrisch, J., Narajewski, M. and Ziel, F. (2023). High-resolution Peak Demand Estimation Using Generalized Additive Models and Deep Neural Networks. *Energy and AI*, [online] 13, p.100236. doi:https://doi.org/10.1016/j.egyai.2023.100236.

Commission for Energy Regulation (CER). (2012). *CER Smart Metering Project - Electricity Customer Behaviour Trial, 2009-2010 [dataset]. 1st Edition.* Irish Social Science Data Archive. SN: 0012-00. Available at: https://www.ucd.ie/issda/data/commissionforenergyregulationcer/ [Accessed 2 Jul. 2024].

Dai, S., Meng, F., Dai, H., Wang, Q. and Chen, X. (2021). Electrical Peak Demand forecasting- A Review. *arXiv (Cornell University)*. [online] doi:https://doi.org/10.48550/arxiv.2108.01393.

Fan, H., MacGill, I.F. and Sproul, A.B. (2017). Statistical Analysis of Drivers of Residential Peak Electricity Demand. *Energy and Buildings*, [online] 141, pp.205–217. doi:https://doi.org/10.1016/j.enbuild.2017.02.030.

Fan, S. and Hyndman, R.J. (2011). Short-Term Load Forecasting Based on a Semi-Parametric Additive Model. *IEEE Transactions on Power Systems*, [online] 27(1), pp.134–141. doi:https://doi.org/10.1109/tpwrs.2011.2162082.

Fu, T., Zhou, H., Ma, X., Hou, Z.J. and Wu, D. (2022). Predicting Peak Day and Peak Hour of Electricity Demand with Ensemble Machine Learning. *Frontiers in Energy Research*, [online] 10. doi:https://doi.org/10.3389/fenrg.2022.944804.

Goncalves Da Silva, P., Ilic, D. and Karnouskos, S. (2014). The Impact of Smart Grid Prosumer Grouping on Forecasting Accuracy and Its Benefits for Local Electricity Market Trading. *IEEE Transactions on Smart Grid*, 5(1), pp.402–410. doi:https://doi.org/10.1109/tsg.2013.2278868.

Hyndman, R.J. and Fan, S. (2010). Density Forecasting for Long-Term Peak Electricity Demand. *IEEE Transactions on Power Systems*, [online] 25(2), pp.1142–1153. doi:https://doi.org/10.1109/tpwrs.2009.2036017.

Jones, R.V., Fuertes, A. and Lomas, K.J. (2015). The Socio-economic, Dwelling and Appliance Related Factors Affecting Electricity Consumption in Domestic Buildings. *Renewable and Sustainable Energy Reviews*, [online] 43, pp.901–917. doi:https://doi.org/10.1016/j.rser.2014.11.084.

Lee, W.-J. and Hong, J. (2015). A Hybrid Dynamic and Fuzzy Time Series Model for mid-term Power Load Forecasting. *International Journal of Electrical Power & Energy Systems*, [online] 64, pp.1057–1062. doi:https://doi.org/10.1016/j.ijepes.2014.08.006.

McSharry, P.E., Bouwman, S. and Bloemhof, G. (2005). Probabilistic Forecasts of the Magnitude and Timing of Peak Electricity Demand. *IEEE Transactions on Power Systems*, [online] 20(2), pp.1166–1172. doi:https://doi.org/10.1109/tpwrs.2005.846071.

Mishra, S. and Palanisamy, P. (2018). Efficient Power Flow Management and Peak Shaving in a Microgrid-PV System. *IEEE Energy Conversion Congress and Exposition (ECCE)*, [online] pp.3792–3798. doi:https://doi.org/10.1109/ecce.2018.8558312.

Mondal, A. and Das, S. (2023). *Parametric Time-series Modelling of London Smart Meter Data for Short-term Demand Forecasting*. IEEE. Available at: http://hdl.handle.net/10871/134043 [Accessed 31 Aug. 2024].

Morales-Mareco, F., Carlos Sauer Ayala, Stalder, D.H., Salgueiro, L. and Grillo, S.A. (2023). Short-Term Electricity Demand Forecasting: Evaluating the Effectiveness of Statistical, Machine Learning, and Deep Learning Models. *2023 XLIX Latin American Computer Conference (CLEI)*. [online] doi:https://doi.org/10.1109/clei60451.2023.10346118.

Mughees, N., Mohsin, S.A., Mughees, A. and Mughees, A. (2021). Deep Sequence to Sequence Bi-LSTM Neural Networks for day-ahead Peak Load Forecasting. *Expert Systems with Applications*, [online] 175, p.114844. doi:https://doi.org/10.1016/j.eswa.2021.114844.

National Grid (2022). *How Will Our Electricity Supply Change in the future? | National Grid Group*. [online] National Grid. Available at: https://www.nationalgrid.com/stories/energy-explained/how-will-our-electricity-supply-change-future [Accessed 23 Aug. 2024].

Rausser, G., Strielkowski, W. and Streimikiene, D. (2017). Smart meters and household electricity consumption: A case study in Ireland. *Energy & Environment*, 29, 0958305X17741385. doi: 10.1177/0958305X17741385.

Sigauke, C. and Chikobvu, D. (2011). Prediction of daily peak electricity demand in South Africa using volatility forecasting models. *Energy Economics*, [online] 33(5), pp.882–888. doi:https://doi.org/10.1016/j.eneco.2011.02.013.

Strbac, G., Tindemans, S., Woolf, M., Bilton, M., Carmichael, R., Schofield, J. R, 2024, *Low Carbon London Project: Data from the Dynamic Time-of-Use Electricity Pricing Trial, 2013*, [data collection], UK Data Service, Accessed 2 June 2024. SN: 7857, doi: http://doi.org/10.5255/UKDA-SN-7857-2

U.S. Energy Information Administration (2012). *Electric Generator Dispatch Depends on System Demand and the Relative Cost of Operation - Today in Energy - U.S. Energy Information Administration (EIA)*. [online] U.S. Energy Information Administration. Available at: https://www.eia.gov/todayinenergy/detail.php?id=7590 [Accessed 23 Aug. 2024].

United Nations (2023). *For a Livable climate: Net-zero Commitments Must Be Backed by Credible Action*. [online] United Nations. Available at: https://www.un.org/en/climatechange/net-zero-coalition [Accessed 23 Aug. 2024].

Visual Crossing (2024) *Weather data services*. Available at: https://www.visualcrossing.com/weather/weather-data-services (Accessed: 5 July 2024). Data retrieved via the querying tool on the website.

Wu, D.-C., Amini, A., Razban, A. and Chen, J. (2018). ARC algorithm: A Novel Approach to Forecast and Manage Daily Electrical Maximum Demand. *Energy*, [online] 154, pp.383–389. doi:https://doi.org/10.1016/j.energy.2018.04.117.

Yalcinoz, T. and Eminoglu, U. (2005). Short Term and Medium Term Power Distribution Load Forecasting by Neural Networks. *Energy Conversion and Management*, [online] 46(9-10), pp.1393–1405. doi:https://doi.org/10.1016/j.enconman.2004.07.005.

7 APPENDIX

This section contains supplementary figures and tables that offer more information about the study. While not essential for understanding the main report, these materials provide deeper insights into the study for readers interested in exploring the work further.

For those seeking even more comprehensive information, the complete codebase in the form of well-annotated Jupyter notebooks is available in the GitHub repository, which has been shared as part of the additional materials submission.

Section 7.1 is on the modeling process while 7.2 focuses on the exploratory data analysis portion of the study.

7.1 Supplementary Information on Model Development and Evaluation

Table 2 on the next page contains additional information on the features used as input variables for the modeling phase (Model 1-4). Table 3 presents the newly aggregated features introduced in Model 5.

Table 2: Features used as input variables

Feature Category	Type	Number of Features	Example
Fixed Schedule	Binary	11	washing_machine_fixed_schedule
Timer Use	Categorical	5	tumble_dryer_timer_use
Appliance Ownership	Categorical	5	washer-dryer_combined_ownership
Demographics	Count	5	count_children
Home ownership	Categorical (one-hot encoded)	3	ownership_owned
Work from home	Categorical	1	work_from_home
Housing Type	Categorical (one-hot encoded)	2	housing_house
Room count	Count	2	count_rooms
Electric Central Heating	Binary	1	electric_central_heating
Heating control method	Binary	5	heating_manual_boiler
Electric heater use	Binary	1	uses_electric_heater
Light Bulb Count	Count	3	count_low_efficiency_bulbs
Other Appliance Count	Count	9	count_tv
Appliance Energy Score	Numerical	2	tv_energy_score
Behavioral and Attitudinal Characteristics	Numerical and Categorical	7	climate_change_concern
Temporal	Numerical and Binary	9	is_holiday
Weather	Numerical	17	humidity_max_today
Consumption	Numerical	10	consumption_peak_today
Pricing	Numerical and Categorical	4	prop_low_price

Table 3: Newly aggregated features introduced in Model 5

Newly Aggregated Feature	**Features Used for Aggregation**
heating_control_score	heating_manual_boiler, heating_thermostatic_valves, heating_auto_set_times, heating_auto_temp_control, heating_not_sure
total_appliance_count	count_cooking_appliances, count_laundry_appliances, count_kitchen_appliances, count_heating_water_appliances, count_entertainment_devices, count_computing_devices, count_tv, washer-dryer_combined_ownership, washing_machine_ownership, tumble_dryer_ownership, dishwasher_ownership, electric_space_heating_ownership
schedule_rigidity	washing_machine_fixed_schedule, tumble_dryer_fixed_schedule, dishwasher_fixed_schedule, immersion_water_heater_fixed_schedule, electric_oven_fixed_schedule, electric_hob_fixed_schedule, ironing_fixed_schedule, electric_shower_fixed_schedule, kettle_fixed_schedule, lighting_fixed_schedule, electric_heater_fixed_schedule
timer_use_score	washer-dryer_combined_timer_use, washing_machine_timer_use, tumble_dryer_timer_use, dishwasher_timer_use, electric_space_heating_timer_use
people_per_room	household_size, count_rooms
electric_heating_score	electric_central_heating, uses_electric_heater

Table 4 outlines the hyperparameters tuned for Model 7, including a brief description of each hyperparameter, the range of values explored during the optimization process, and the optimal value selected for the final model.

Table 4: Hyperparameter search range and optimal values for model 7

Hyperparameter	Description	Search Range	Best Hyperparameter
n_estimators	Number of gradient boosted trees	randint(800, 2000)	1449
max_depth	Maximum tree depth for base learners	randint(4, 12)	11
learning_rate	Boosting learning rate	loguniform(0.01, 0.29)	0.016540578683031364
subsample	Subsample ratio of the training instances	uniform(0.5, 0.5)	0.669615333889757
colsample_bytree	Subsample ratio of columns when constructing each tree	uniform(0.5, 0.5)	0.8241563780815623
min_child_weight	Minimum sum of instance weight needed in a child	randint(1, 10)	5
gamma	Minimum loss reduction required to make a further partition on a leaf node	uniform(0, 0.3)	0.05098221426517733
reg_alpha	L1 regularization term on weights	loguniform(1e-3, 1)	0.8845998752143264
reg_lambda	L2 regularization term on weights	loguniform(1e-3, 1)	0.6724110524452301

Table 5 on the following page presents a detailed comparison of performance metrics for all models. "Tr" represents Training data, "Val" represents Validation data, and "Test" represents Test data. All models show signs of overfitting despite the rigorous hyperparameter tuning process, as evidenced by the discrepancy between training (Tr) and validation (Val) or test performance metrics. This suggests that insufficient training data may be the reason for this phenomenon.

Table 5: Comprehensive performance metrics across all models

Model Name	Tr WAPE	Val WAPE	Test WAPE	Tr MAPE	Val MAPE	Test MAPE	Tr MAE	Val MAE	Test MAE	Tr R^2	Val R^2	Test R^2
Model 1 (Temporal + Consumption Features)	29.43%	31.33%	29.94%	41.34%	45.43%	48.61%	0.328	0.351	0.384	0.68	0.59	0.64
Model 2 (Model 1 Features + Weather Features)	29.38%	31.38%	30.17%	41.60%	45.76%	49.70%	0.327	0.352	0.387	0.68	0.59	0.63
Model 3 (Model 2 Features + Survey Features)	26.66%	30.22%	29.27%	36.94%	44.07%	48.53%	0.297	0.339	0.375	0.74	0.62	0.65
Model 4 (Model 3's Top 15 Features)	28.16%	31.01%	29.65%	39.16%	44.31%	47.03%	0.313	0.348	0.380	0.71	0.59	0.64
Model 5 (Model 3 + Aggregated Household Features + Features Used for Aggregation)	26.59%	30.18%	29.23%	35.24%	43.74%	48.25%	0.286	0.338	0.374	0.76	0.62	0.66
Model 6 (Model 5 Without Features Used for Aggregation)	26.99%	30.44%	29.44%	37.37%	44.30%	48.77%	0.300	0.341	0.378	0.73	0.61	0.65
Model 7 (Model 6's Top 15 Features)	27.82%	30.92%	29.63%	38.46%	44.07%	46.32%	0.310	0.347	0.380	0.72	0.60	0.64

7.2 Additional Insights from Exploratory Data Analysis

Figure 8 and Figure 9 show a scatter plot of maximum temperature for current day vs peak consumption for the same day as well as peak consumption for the next day respectively. Interestingly, both plots only show a weak negative trend. In addition to having a non-linear relationship, colder temperatures tend to have higher peaks. This makes sense for the UK, where electricity consumption tends to be higher during colder weather (for heating).

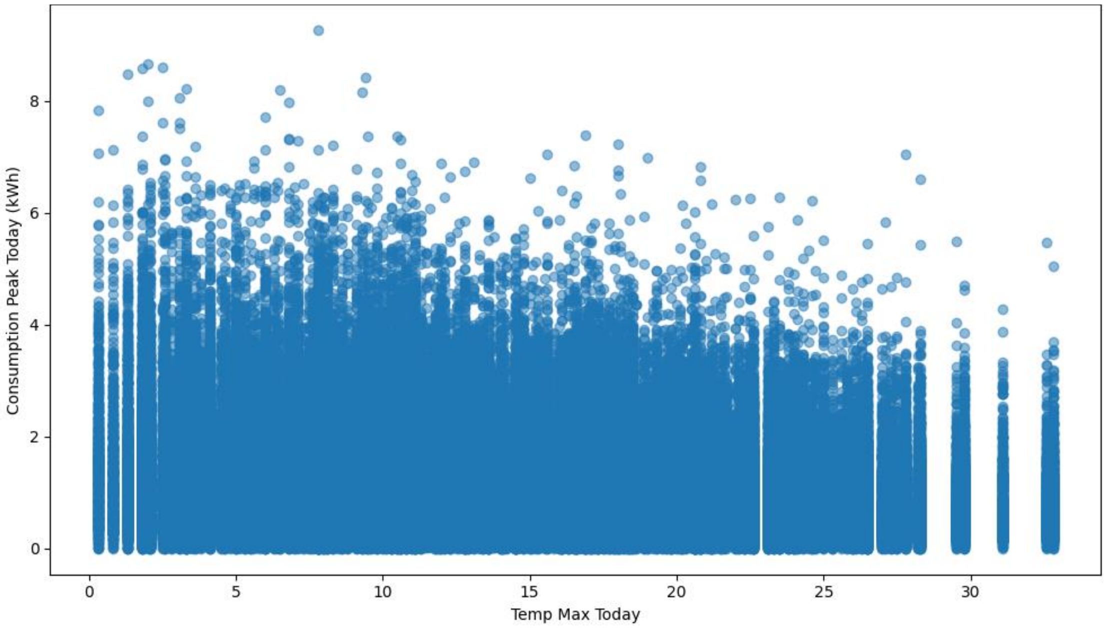

Figure 8: Scatter plot of current day's maximum temperature vs current day's peak electricity consumption

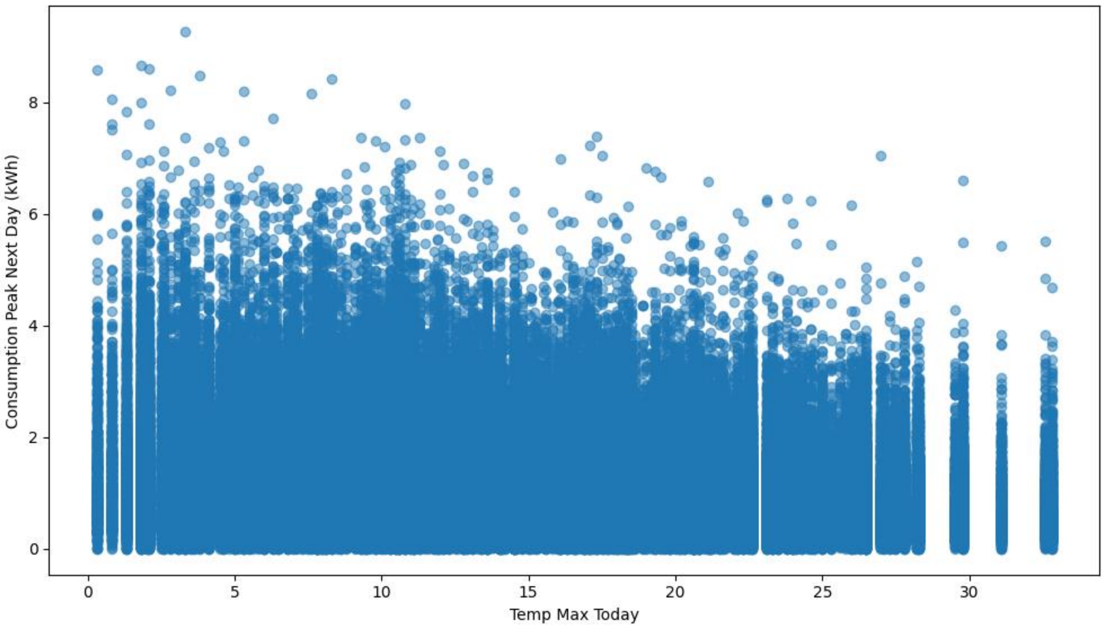

Figure 9: Scatter plot of current day's maximum temperature vs next day's peak electricity consumption

Figure 10 reveals that the daily peak consumption tends to be slightly higher on Saturdays, with the other days having similar consumption levels.

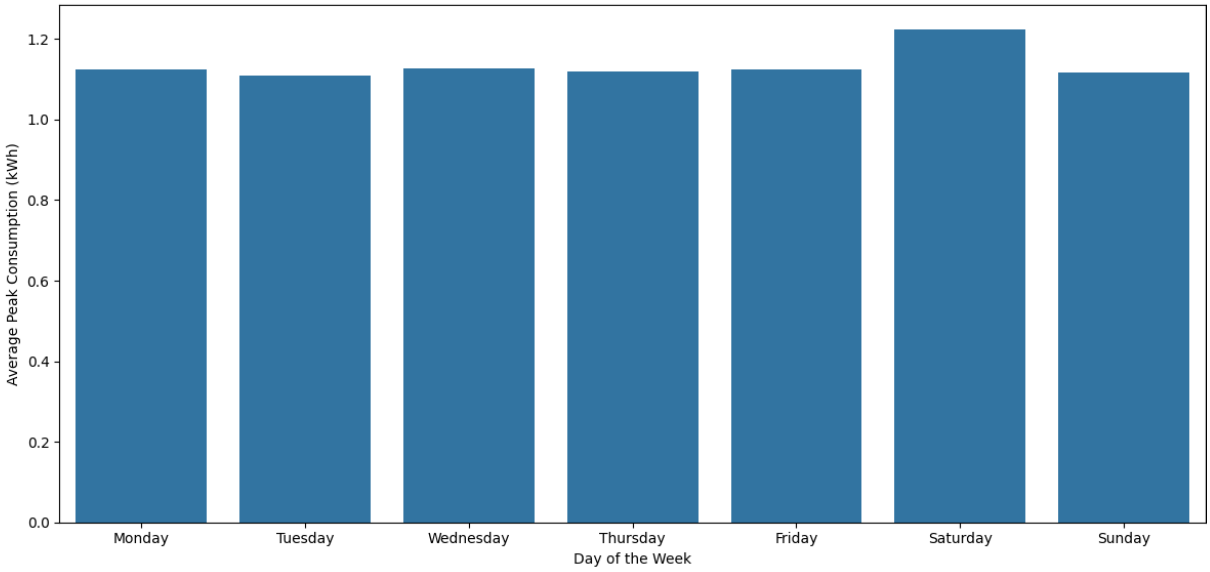

Figure 10: Average (mean) daily peak consumption by day of the week

In Figure 11, the variation in daily peak consumption throughout the year can be observed, with winter months having the highest consumption and summer months having the lowest.

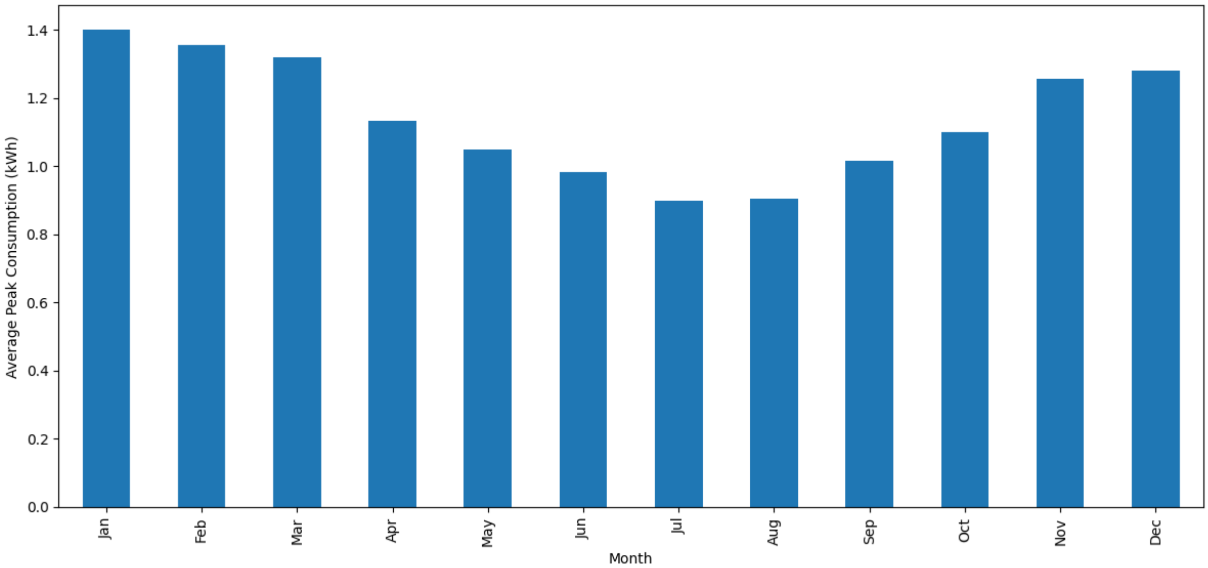

Figure 11: Average (mean) daily peak consumption by month

The following boxplot (Figure 12) shows the relationship between the target variable and interest in renewable energy. Surprisingly, those who are not at all interested in renewable energy usually had a lower peak demand than the other groups.

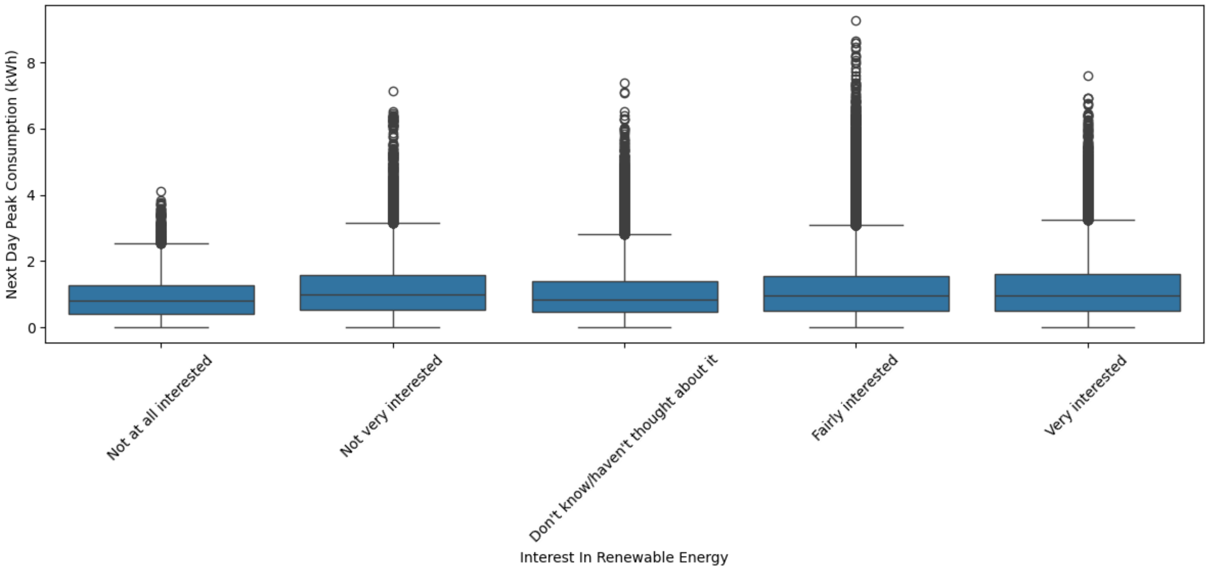

Figure 12: Peak consumption next day of households grouped by interest in renewable energy

Figure 13 shows a correlation matrix of a few selected variables highlighting a strong positive correlation between recent consumption patterns and peak consumption of the following day. Household characteristics, in contrast, do not have the same magnitude of correlation. The full correlation matrix, along with the Spearman correlation matrix for categorical variables and the mutual information scores between categorical variables and the next day's peak consumption, is provided in the GitHub repository shared as part of the additional materials submission.

Figure 13: Correlation matrix of a few selected variables