International Research Journal of Multidisciplinary Scope (IRJMS), 2025; 6(2): 941-959

Original Article | ISSN (O): 2582-631X DOI: 10.47857/irjms.2025.v06i02.03635

Demand Response Targeting via Load Cluster Analysis

CS Kudarihal*, Manoj Gupta, Sunil Kumar Gupta

Department of Electrical Engineering, Poornima University, Jaipur, India. *Corresponding Author's Email: cskudarihal@gmail.com

Abstract

Demand Response Programs (DRPs) have emerged as effective tools to enhance grid stability, reduce peak demand, and promote energy efficiency. However, targeting the right consumers for DRPs remains a challenge, as different household's exhibit varying consumption patterns. This research aims to develop a method for identifying optimal target groups for DRPs using clustering techniques, specifically K-means clustering, in combination with load profile analysis. The dataset used in this study is sourced from the National Renewable Energy Laboratory (NREL), based on the 2022 Residential Energy Consumption Survey (RECS), which includes power usage data for individual households recorded at 10-minute intervals. By analysing household electricity usage data and applying K-means clustering, consumers are grouped into distinct clusters based on their load profiles. The optimal number of clusters is determined using the Elbow method, which helps identify the most representative consumption patterns. The suitability of these clusters for various DRP strategies, such as time-of-use pricing or direct load control, is then evaluated. The findings demonstrate how clustering techniques can segment consumers effectively, allowing for more targeted and efficient DRP implementation. This approach optimizes participation in DRPs and contributes to reducing grid stress during peak demand periods and improving energy efficiency on a broader scale.

Keywords: Clustering, Demand Response, k-means, Load Profiling, Smart Grid, Time Series.

Introduction

Demand Response programmes are used by most of the electric system operators and planners to balance supply and demand. These initiatives can lower retail and wholesale electricity prices. Customers are involved in Demand Response activities through time-based tariffs such as time-of-use, critical peak, variable peak, real time, and critical peak rebates. In response to time of the day tariff or other financial incentives, demand response enables consumers to contribute to the operation of the electric grid by reducing their electricity use during peak hours (1). By employing direct load control programs, electric utilities can cyclically manage dishwashers, air conditioners, EVs charging and water heaters during peak demand periods, offering financial incentives and reducing the user's electricity costs. The electric power sector is experiencing growing benefits from DR programs due to advancements in grid modernisation. To avert overload and power failures, smart grid sensors can detect peak load issues and automatically adjust power distribution. Consumer time-based rate plans are enhanced through advanced metering infrastructure. By providing information on power usage and costs, in-home displays and home-area networks can assist consumers in changing their behaviour and reducing peak-hour power use. These initiatives reduce peak demand and contribute to cost savings for electricity providers by postponing the construction of new power plants and distribution equipment dedicated to meeting peak demand (2). The Smart Grid Research and Development Programme in modernising existing grids aims to support the power sector in developing, testing, and demonstrating integrated national infrastructures for electricity, communication, and information. These infrastructures dynamically optimise grid operations by incorporating demand response and consumer participation. The program focuses on developing technologies, tools, and techniques for grid modernisation, including research on smart grid technologies, distribution system modelling, Trans active energy, consumer behaviour modelling, and high-speed computational analysis for decision support tools, all aimed at achieving these objectives (3). Demand Response Programs (DRPs) help utility companies by lowering peak demand during

(Received 27th December 2024; Accepted 22nd April 2025; Published 30th April 2025)

times of the day when consumption is at its highest. This reduces overall installation costs, operating expenses, and the chance of grid failures by avoiding the need to construct transmission networks and power plants only to satisfy the highest possible demand (4). A thorough analysis of the existing literature was done prior to the current study on demand response targeting. Below is a summary of the main conclusions drawn from this review. High-resolution energy consumption data is now easily accessible due to the introduction of smart grid technologies and the extensive use of Advanced Metering Infrastructure (AMI). This development has made it easier to analyse consumer load profiles in-depth and to plan Demand Response (DR) programs more efficiently. It is commonly known that incentive-based demand response (DR) programs are a good way to get customers to cut back on or change how much electricity they use during times of high demand, which improves grid stability and lowers overall energy expenses. Clustering techniques, particularly when paired with consumer profiling, have also emerged as leading techniques in addressing this issue. Other authors have made use of clustering algorithms to classify consumers into homogenous groups depending on their energy consumption profiles. Classical clustering techniques such as k-means and hierarchical clustering have been found to be computationally efficient and effective in load profiling. For instance, some studies demonstrated the effectiveness of clustering in describing distinct load profiles that can be used in the formulation of tailored DR strategies (5). Similarly, segmenting consumers with homogeneous consumption behaviours allows for more efficient and targeted demand-side management interventions (6). These studies demonstrate the potential of clustering techniques to reveal useful insights from otherwise latent consumption patterns. Recent developments have focused on improving segmentation precision by incorporating new clustering techniques and feature engineering techniques. Time-series clustering for the identification of temporal patterns in load profiles has gained popularity. It has been emphasised that taking into account temporal dynamics is critical for obtaining more meaningful customer groups (7). In addition, hybrid methods that combine clustering with predictive analytics, such as fuzzy c-means and density-based clustering have been found to be helpful in addressing the inherent variability and noise in load data.

In addition, research has also emphasized the use of domain-specific feature extraction techniques, such as Fourier Transform and Principal Component Analysis (PCA), in data dimension reduction and highlighting key consumption features. Exogenous context variables, such as weather, socioeconomic, and appliance use, have also been utilized in clustering models to improve segmentation results. For example, using such context factors significantly enhanced customer segmentation accuracy and demand response (DR) program targeting effectiveness (8). A significant challenge for the effective implementation of demand response (DR) programs is the precise identification of appropriate target consumer groups, as the overall success of such programs is greatly reliant on directing resources toward consumers who are most likely to change their consumption in response to incentives. Poor targeting for DR programs can result in low levels of participant engagement and reduced efficacy (9). The present research addresses this issue by using the k-means clustering algorithm to group electricity consumers based on their load profiles from high-resolution smart meter data. This means of analysis allows clustering by similar consumption behaviour, and subsequently, the identification of consumers by characteristics that are more likely to participate in DR intervention (e.g. flexible loads, peak-shiftable loads). The application of k-means clustering accommodates efficient and effective targeting of consumers for DR programs by streamlining the segmentation process and encouraging data-driven decision making, ultimately contributing to more effective and scalable demand-side management initiatives within smart grid systems.

Types of Demand Response Program in Utilities

Emergency Demand Response: Prevents blackouts when demand exceeds supply. Heating and cooling systems increase grid demand on exceptionally hot or cold days. In curtailment, participating consumers are encouraged to disconnect high-demand loads for some duration to prevent blackout due to high peak load.

Economic Demand Response: Utility companies use economic demand response to avoid the high costs of providing electricity during peak demand times by ramping up "peaking" power facilities to prevent unexpected demand. Load shifting is one such strategy for the economic operation of smart grids.

Ancillary Service Demand Response: It will help utilities meet reliability standards by supporting power delivery to loads (10).

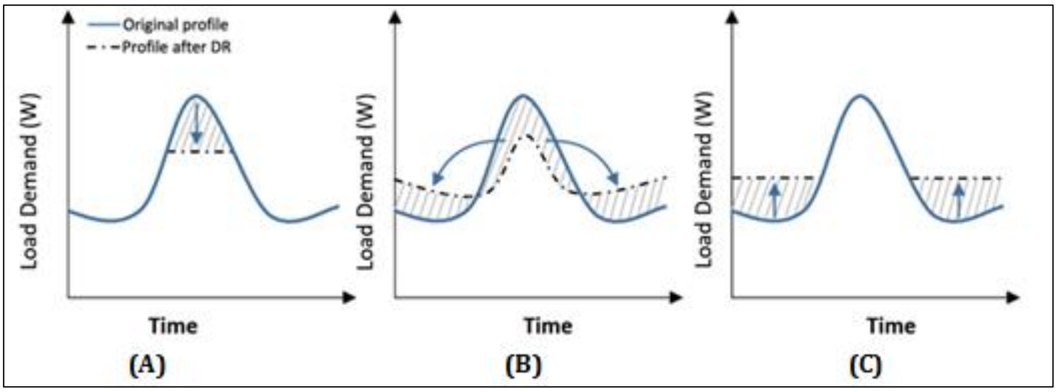

Figure 1: Demand Response Strategies (A) Curtailment (B) Load Shifting (C) Valley filling (11)

Different Strategies in a Demand Response Program

Demand Response Programs in electric power systems employs various strategies as illustrated in Figure 1, which will be explained further in the ensuing paragraphs (11).

Curtailment: This involves allowing load or demand which is usually temporary in nature and is mostly employed during peak times, which is usually the stressed point for the grid, or when supply cost is high. For example, the Utilities can call upon the large users of energy to practice load reduction for very short periods. It is in fact a way of switching off demand in order to keep the balance, once the load is likely to cross the supply end.

Load Shifting: This is a form in which load demand in high-cost periods is transferred to low-cost periods. There is a useful example of this practice – industrial power user can reschedule some of its operations on night time when electricity prices are cheaper, hence off-peak hours are avoided. Load shifting also helps in decreasing the peak load and flattening the load profile, which in turn, reduces the pressure on the grid during peak periods.

Valley Filling: The theory governs the expansion of electricity demand during valley times, also known as valley filling. This can be accomplished by activating specific processes or engaging in energy-intensive activities during slack periods. Valley filling benefits utilities by increasing generation asset efficiency and ensuring consistent generation levels.

Participation in the Demand Response (DR) Program necessitates the involvement or consent of electricity consumers. The primary challenge is identifying domestic electricity consumers willing to participate in the DR program. In this research article, we have analysed electricity consumer profiles using the k-means clustering technique. These profiles offer insights into seasonal changes in electricity consumption, and load profiles will help identify consumers who are likely to participate in demand response programming.

Methodology

This research explores the use of a data-driven approach in targeting electricity Demand Response (DR) effectively. First, k-means clustering was used to analyse consumer load profiles. Pre-processing was applied to high-resolution smart meter data to remove noise and standardized for consistency and accuracy of data. Principal Component Analysis (PCA), together with domain knowledge, was subsequently used to derive significant consumption features, such as peak demand, daily load fluctuation, and usage behaviour. Following that, consumers with comparable load characteristics were grouped into discrete clusters using the k-means algorithm. By identifying consumer segments with comparable electricity usage patterns, these clusters made it easier to target the people who are most likely to react to DR signals. This segmentation offered valuable insights for crafting customized DR programs, boosting participation, and improving overall grid efficiency. DR initiatives, ultimately increasing grid efficiency and consumer involvement.

Hybrid Feature Selection for Load Profile Characterization

The research employed a hybrid feature selection method to provide both theoretical significance and data-driven robustness in discovering important electricity consumption pattern attributes for successful demand response (DR) targeting. Domain expertise was first employed to inform the manual selection of the features that have been well documented in the literature as being significant in DR behaviour. These were measures like peak load per day, off-peak consumption, variability of load, load factor, and time-of-use profiles characteristics that capture both the magnitude and temporal nature of electricity consumption. This phase made sure that the chosen features were of practical value and interpretability to utility operators and policy makers. To augment and further sharpen this expert-based selection, automated dimensionality reduction methods were then used. Particularly, Principal Component Analysis (PCA) was applied to determine and preserve the most informative and non-redundant features by converting the initial feature set into orthogonal components that preserved the highest variance in the data. The two-stage feature selection approach integrating domain knowledge with statistical learning enabled us to eliminate noise, simplify computational complexity, and concentrate the clustering analysis on the most significant patterns within the consumer load profiles. The outcome was a more effective and precise segmentation of consumers, improving the overall reliability and usability of the demand response targeting framework.

Dataset Description

The dataset comprises a makeshift of energy consumption data from residential buildings in 200 households in the Midwest region of the United States. This dataset came from the National Renewable Energy Laboratory (NREL) data repository and is based on data from the 2022 Residential Energy Consumption Survey (RECS). This dataset contains energy usage in Watts (W), recorded every 10 minutes interval, thereby providing very useful insights into electricity consumption patterns in the households (12). Households in the dataset were randomly chosen, providing a fair representation of energy consumption in various types and conditions around the region. This dataset is quite useful in studying and analysing trends in residential energy consumption. Because of its fine temporal resolution, it addresses detailed questions on daily and seasonal patterns of consumption, peak demand periods, and how household behaviours or external factors influenced energy use. Such a dataset is to be utilized by researchers and energy analysts for predictive modelling, energy efficiency optimization, and determining strategies under demand-side energy management initiatives. Moreover, the dataset will also allow the comparison of possible interventions, such as energy savings and renewable energy introduction into the residential sector. This dataset is well-suited for studies on energy usage trends, efficiency analysis, and predictive modelling of residential power consumption.

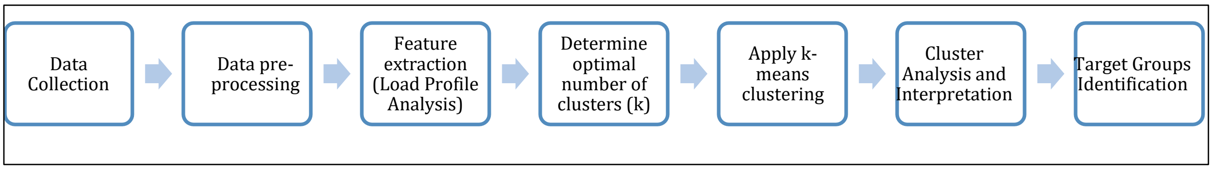

Figure 2: Work Flow Diagram to Target the Electricity Consumers for DR Programs

Proposed Workflow

Figure: 2 presents the flow diagram illustrating the methodology adopted in this study. The work flow begins with data acquisition and pre-processing. The dataset is examined for missing values, duplicates, and outliers. Common issues include missing values or null entries and incorrect format of data points in specific rows or columns. Based on the assessment of the missing data, appropriate actions are taken either imputing the missing values or removing the affected records entirely. Following data cleaning, feature engineering is performed to create more useful representations of the data. This involves building daily or weekly load profiles, extracting time-of-use features (such as usage on weekdays versus weekends or seasonal usage changes), and deriving statistical measures (mean, variance, peak load) for each household. The data will be normalized or standardized so that the K-means algorithm can perform optimally; techniques used could include Min-Max scaling or Z-score normalization (13). Exploratory Data Analysis (EDA) will be carried out following data pre-processing and cleaning. Understanding fundamental consumption trends can be aided by visualisation tools like line plots of household consumption over time. Potential clusters and seasonal patterns can be found through additional analysis using daily or weekly load profiles. To find connections between different features, correlation analysis will also be carried out. For example, power consumption on weekdays and weekends will be compared. These analyses offer preliminary understandings of the data, assisting in the discovery of underlying trends and guiding the course of further modelling initiatives.

The next step will be establishing the feasibility for K-means clustering. One of the critical issues is deciding on the number of clusters, k. This is typically addressed using methods such as the Elbow Method, owing to how inertia is plotted against a wide vector of k, looking for a point at which inertia rates of decrease slow. Other techniques-such as the Silhouette score or Gap Statistics- may also confirm the best k number value by analysing on how the cluster separation is achieved. Once k is found, K-means clustering will be performed using the normalized data and centroids of different clusters will be analysed against each of them to get the household features in each cluster (high or low energy usage, time-of-the-day usage patterns). After cluster analysis, each household will be assigned to one of the identified clusters based on its consumption profile. Thereafter will follow a post-clustering analysis that will allow further scrutiny and validation of the results. Initially, the cluster will be profiled by evaluating its principal statistical features and load habits. For instance, some clusters may have higher evening consumption while some might exhibit more even usage throughout the day. Dimensionality reduction techniques, such as Principal Component Analysis (PCA) or T-distributed Neighbour Embedding (t-SNE), will be applied to visually present the two or three-dimensional clusters to gain further insights into how different or similar those clusters are. Also, the stability and consistency of the clusters will be assessed to ascertain that every cluster contains households with comparable behaviours and needs. The next step will focus upon the targeting of consumers for demand response programs. Based on the insights from clustering, a program for demand response (DR) will be developed for every cluster. For instance, households exhibiting high evening consumption may be eligible for time-of-use price or direct load control during peak hours, while those exhibiting a low steady consumption pattern may qualify for real-time pricing or other incentive-based programs. Likewise, consumers who use energy in a seasonal manner could be targeted with specific demand response programs during the winter or summer months of high demand. Program simulations will estimate how each cluster would respond to the proposed demand response initiatives in order to maximize the potential benefits of each program. Following program simulation, clustering results will be evaluated. The evaluation will focus on how successful the clustering algorithm is in differentiating groups of consumers, targeted for tailored DR programs. Training of some key performance indicators like demand reduction and peak load shifting will help measure how effective each cluster is in accepting the DR programs (14). A sensitivity analysis will also be carried out to check how sensitive clustering results are to changes in the number of clusters or other model parameters. The results will finally be compiled, and the findings will be discussed, covering the analysis of the identified clusters, proposed demand-response programs, and the expected outcomes for each cluster. The conclusion will mention how well K-means clustering might be suited to segment electricity consumers for more efficient DR program implementation; it will in addition outline directions for future research for enhancing targeting and participation in DR initiatives.

Clustering Performance and Scalability Considerations

The scalability of clustering algorithms is a significant consideration when examining large amounts of smart grid data collected by extensive deployments of large-scale AMI (15). In this research, k-means clustering was chosen not only for its interpretability and simplicity but also due to its demonstrated efficiency and scalability when working with large data sets. In order to keep the methodology computationally viable at scale, we introduced a hybrid feature selection process, consisting of Principal Component Analysis (PCA), which significantly reduced the input data dimensionality without losing the most informative consumer load profile characteristics. This pre-processing reduced computational complexity and memory usage considerably during clustering. In addition, the dataset was operated on in partitioned batches in order to mimic real-world large-scale deployment settings. The stability and consistency of clustering results across these partitions justified the scalability of the method for use in large-scale applications. While this implementation shows the applicability of the method to large datasets, future work will include the exploration of integrating parallel and distributed clustering frameworks like Mini-Batch k-means and Spark-based clustering algorithms in order to deliver real-time, scalable DR targeting in larger smart grid infrastructures (16). These improvements will also facilitate the operational implementation of DR programs in more data-intensive and decentralized power systems.

Data Pre-processing Steps

Data loading and Inspection: Load the dataset into the DataFrame and check its structure for the type of features and determine, if there are any missing data. Generate summary statistics in order to assess the range, mean, and variance of features (17).

Handling Missing Data: Check for missing values and handle them as follows:

Numerical Features - Impute missing values using the mean $x_i = \frac{\sum_{j=1}^{n} x_j}{n}$ #[1]

where x_j is not null.

Categorical Features Impute with the mode:

$$x_i = argmax_k\left(count(x_k)\right)\#[2]$$

Encoding Categorical Variables: Convert categorical variables into numeric representations using One-Hot Encoding and for a categorical feature C with values $\{c_1, c_2, \dots\dots.., c_k\}$, create k binary columns:

$$Column_i = \{1, if\ C = c_i \quad 0, otherwise\#\#[3]$$

Scaling and Normalization: Standardize numerical features t ensure consistent scales using Z-Score Standardization:

$$x_{Scaled} = \frac{x - \mu}{\sigma}\#[4]$$

Dimensionality Reduction: Use Principal Component Analysis (PCA) to reduce dimensions:

$$C = \frac{1}{n}X^T X\#[5]$$

Compute eigenvalues λ and eigenvectors v, then project the data:

$$X_{reduced=} X_{v_k}\#[6]$$

where v_k are the top k eigenvectors

Exploratory Data Analysis

The dataset, consisting of electricity consumption data from 200 households recorded at regular intervals, underwent detailed exploratory data analysis (EDA) to uncover energy usage patterns, correlations, and trend analysis.

Visualization: The visualization stage offered clear pictures of consumption distribution among households, thus establishing profound variability. The histograms in Figure 3, shows that the majority of households operated in the conventional range of electricity usage, while a set of households were either consistently high or erratic in their pattern, indicating peculiar behavioural or operational characteristics (18). This kind of insight is undoubtedly valuable in identifying the peculiar households that might benefit from targeted energy conservation programs.

Correlation Analysis: The correlation analysis, as shown in Figure 4 is the intermediate between house clusters, giving way to a heatmap demonstrating clusters of exceptionally strong positive correlations between some of the houses. This indicates that some households exhibit similar patterns of electricity consumption, potentially because of similar populations, similar standards of appliance usage, or common environmental factors. Those are correlation results that provide some insights into household behaviour and permit classifying households into different groups that can then be targeted with energy conservation efforts for implementing DR programs. The insights from the correlation analysis may also help to create predictive models that would use information from one household to explain and understand the behaviour of other correlated households (19).

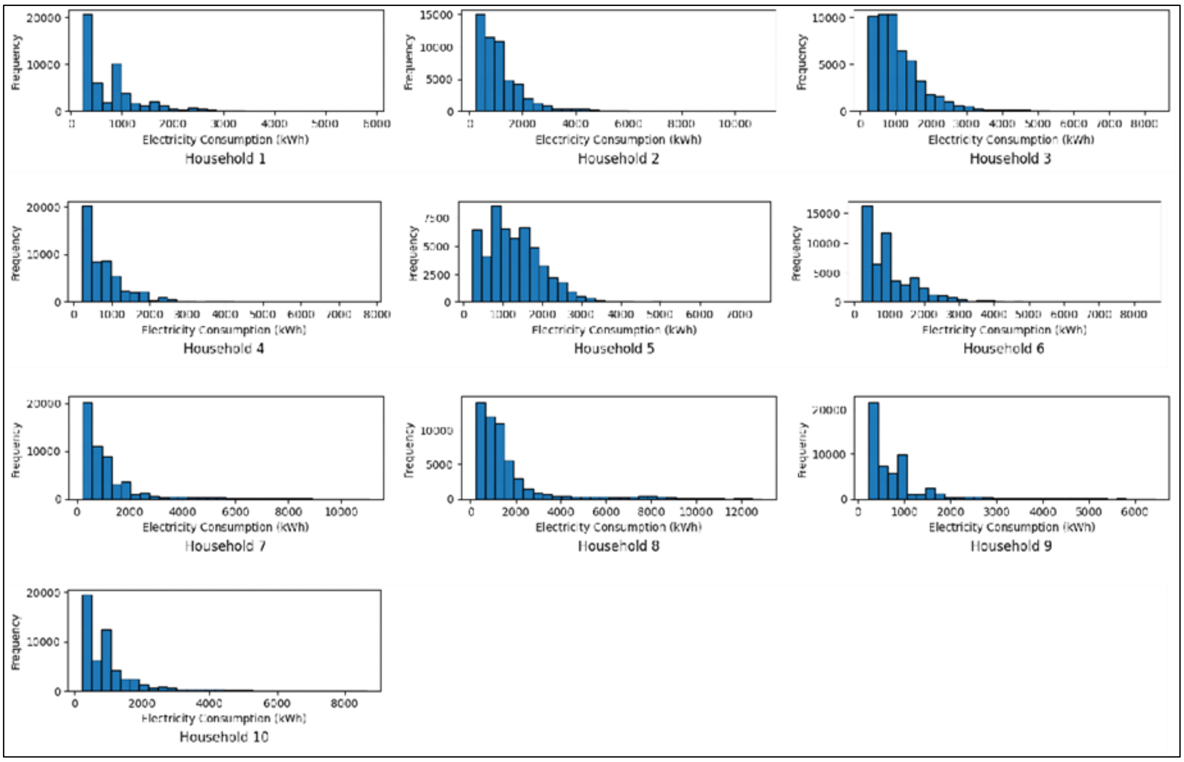

Figure 3: Histogram of Electricity Consumption for 10 Households

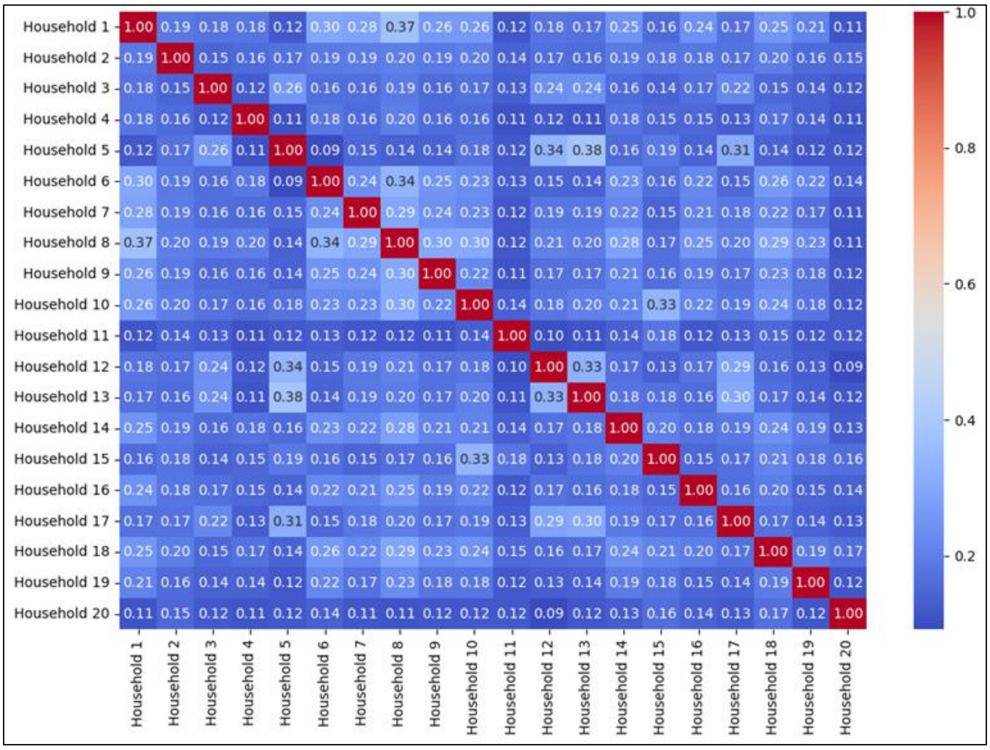

Figure 4: Correlation Matrix

Trend Analysis: In Figure 5 trends shed light on when electricity use is at its highest. According to the analysis, some hours saw higher power consumption, which corresponded to common daily activities like getting ready in the morning and unwinding in the evening. As seen in Figure 6, the data were combined by day and hour in order to better examine how time affects electricity consumption. Seasonal variations and distinct differences between weekday and weekend usage were brought to light by this analysis. For instance, evening electricity use was higher in many households. By making it more expensive to use power during peak hours, time-based pricing tariffs may be able to encourage consumers to use less electricity during those times, according to these observed patterns (20).

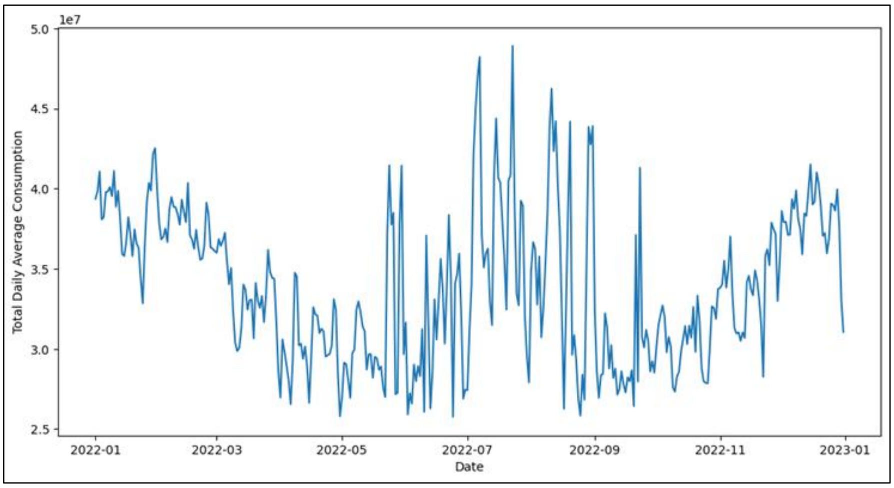

Figure 5: Daily Average Consumption all Households Combined

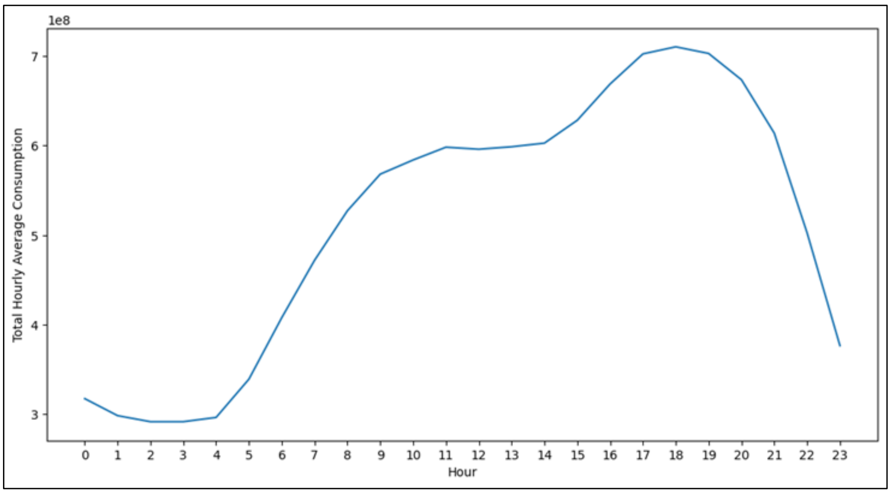

Figure 6: Hourly Average Consumption for all Households Combined

Top 5 Household Comparison

Important insights into different usage patterns and consumption variability can be gained by comparing the five households with the highest average power consumption. The data spread and any notable spikes that correlate to times of high-power demand are highlighted in the boxplots that show the distribution of electricity use (21). Potential causes of higher consumption levels, such as larger households, energy-intensive appliances, or inefficient usage patterns, can be found using this side-by-side analysis. These results are solid and open up new avenues for investigation. They have the ability to spot unusual trends, predict future usage, and guide the development of policies meant to cut down on energy waste and encourage sustainable consumption habits. Significant variations in usage patterns and variability were found when the five households with the highest average electricity consumption were examined. The box plots show the distribution of electricity use, highlighting the spread and pointing out times of unusually high consumption, as shown in Figure 7. This comparison sheds light on possible reasons for higher electricity consumption, such as larger households, the presence of high-power appliances, or energy-inefficient behaviours. These results provide a strong basis for future studies, such as identifying unusual trends, projecting future consumption, and creating regulations to encourage more economical and ecologically friendly electricity use.

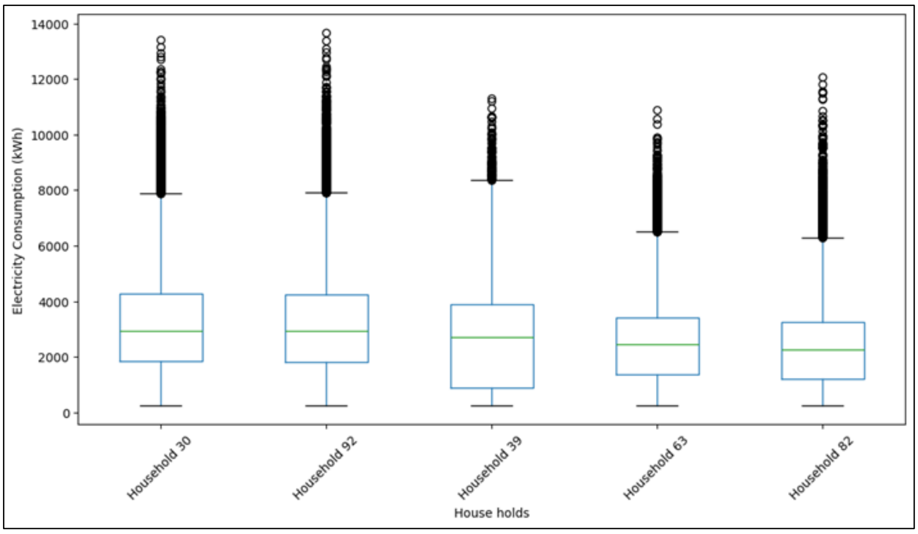

Figure 7: Daily Average Consumption of Top 5 Households

Table 1 shows the five-number summary statistics: Minimum, First Quartile (Q1), Median (Q2), Third Quartile (Q3), and Maximum of the top 5 households with the highest mean electricity consumption. These statistics are derived from the box plots, which graphically indicate the spread and central tendency of electricity use for each household.

Table 1: Statistical Summary of Top Five Household Energy Consumption (kWh)

Top 5 Households	Min	Q1	Median	Q3	Max
Household 30	230.0	1853.90	2943.1	4271.10	13420.0
Household 92	230.0	1812.80	2930.8	4253.90	13652.0
Household 39	230.0	900.00	2713.9	3880.20	11318.0
Household 63	230.0	1355.42	2443.4	3414.15	10887.0
Household 82	230.0	1198.32	2273.4	3239.12	12078.0

Clustering Techniques for Demand Response Programs

A key technique in data mining and machine learning is clustering, which groups a collection of objects so that those in the same group, or cluster, are more similar to one another than to those in other groups (22). In the context of DR programs, clustering techniques can be used to segment electricity consumers based on their load profiles. This segmentation supports targeting specific groups for tailored DR programs, which would likely lead to more efficient management of energy and higher participation rates. Some common clustering techniques that one can apply to electricity consumption data are listed below:

K-means Clustering

K-means is one of the most popular algorithms for unsupervised learning, partitioning data points into a predefined number of clusters. The number of the clusters is denoted by k. The algorithm then iterates through the steps. Chose k initial centroids, and these centroids can be selected randomly from the data points or selected by using more complex methods.

- Assign each data point the label of the closest centroid along a distance metric (commonly Euclidean distance).
- Recalculate the centroids by taking the mean of all data points assigned to each centroid.
- Repeat the assignment and centroid update steps until the centroids no longer change or the algorithm converges.

K-means is relatively efficient and performs well especially when the number of groups is known in advance; however, it can hardly handle clusters of different shapes or various sizes, and its success is sensitive to the initial choice of centers (23).

Hierarchical Clustering

Hierarchical clustering constructs a tree of clusters, called a dendrogram that illustrates at which points data points are merged or split. There are two primary strategies for hierarchical clustering:

- Agglomerative clustering begins with each data point as its own cluster and then iteratively merges the closest pairs of clusters until all data points belong to a single cluster.
- Divisive cluster begins with all data in one cluster and continues to bisect the clusters into smaller-sized groups.

The major attraction of hierarchical clustering is the fact that it does not require a number of clusters ahead of time, and more importantly, it provides extensive information about the data structure, but it is computationally expensive with large datasets (24, 25).

Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

DBSCAN is a well-known density-based clustering algorithm that seeks to find points that cluster together based on distance from the centres of the clusters with a minimum number of points within a neighbourhood. It is particularly useful for identifying arbitrary shapes of clusters and dealing with noise (outliers) in the data. The algorithm works as follows:

- Choose a point and find all neighbouring points within a specified distance (epsilon).
- If the number of neighbouring points is above a certain threshold (minPts), it is a cluster.
- In this way, neighbours are recursively attached in order to enlarge a cluster.
- Points that do not belong to any cluster are labelled as noise or outliers.

DBSCAN has the advantages of being able to detect clusters with arbitrary shapes and sizes without making any explicit assumption about the number of clusters. However, the choice of parameters (epsilon and minPts) can have an outsized effect on the results (26, 27).

Gaussian Mixture Models (GMM)

Gaussian Mixture Model (GMMs) is totally a probabilistic class of generative model for which the data is assumed as produced out of a blend of multiple Gaussian distributions. This establishes the assignment against each data point that might belong to a cluster or the probabilities of data point belonging to multiple clusters before its assignments. To that point, GMM can be understood to be the general subgroup of K-means, as each point in K-means is assigned to one determined cluster, while it goes using the entire detail relate to its point for every cluster association in GMM (28, 29).

- GMM works well when the clusters are elliptical, while K-means assumes spherical shape for the clusters.
- To fit a GMM, the most adopted method is the Expectation-Maximization (EM) algorithm, which iteratively alternates between assigning data points to clusters and updating the parameters of the Gaussian distributions.

GMM allows soft clustering where each data point can be assigned an affiliation in terms of probabilities to multiple clusters. However, the solution is not invariant to initializations and can run into trouble with highly overlapping clusters, as is also the case for K-means.

Self-Organizing Maps (SOM)

Self-Organizing Maps (SOM) are a certain type of artificial neural network trained using unsupervised learning, which results in low-dimensional representation of high-dimensional data, generally speaking, two dimensions. The SOM algorithm maps each data element to a grid of neurons depending on its difference from other neuron values: similar data points yield nearby neurons in the grid. The algorithm iterates through the following steps:

- Randomly initialize the weights of the neuron grid.
- For each data point, check for a neuron with most similar weight from the other units.
- Train the weights such that the best matching unit and its neighbours match better the input.
- Repeat until convergence.

These are especially suited for visualizing high-dimensional data and discovering patterns; although they tend to require more computing power and careful setting of parameters (30).

Comparison of Clustering Techniques

Each clustering method has its pros and cons. The specific characteristics of the dataset and the objectives of the clustering task determine the chosen method. K-means is more efficient and faster, but it assumes that the clusters are spherical; hierarchical clustering gives detailed insights into the cluster relationships. DBSCAN is very robust to noise and can identify arbitrarily shaped clusters, but its performance may degrade in the presence of varying densities. GMM allows for soft clustering but can be computationally heavy. SOMs have some high-performance visualizations to offer, but they require considerable parameter tuning.

Clustering for Demand Response Programs

Clustering methods may be employed in demand response programs to classify consumers as per their electricity-usage profiles. Such grouped consumers, having similar load profiles, can make it possible to develop demand-response strategies to facilitate maximum outreach and minimum disruption. For instance, consumers clustered by peak load timing can enter time-of-use pricing programs, while consumers with better flexibility can go for real-time pricing or direct load control programs.

Load Profile Analysis

Load profile analysis involves examining the electricity consumption patterns of households or buildings over time to identify trends, peak usage periods, and consumption behaviours. Through the examination of load profiles, significant information regarding daily, weekly, and seasonal patterns in the consumption of energy can be obtained. The information helps in maximizing energy distribution, enhancing demand forecasts, and creating effective energy management techniques. The analysis typically includes visualizations such as time series plots, aggregate consumption trends, and peak usage analysis, alongside advanced techniques like clustering to segment households based on similar consumption behaviours. This type of analysis is crucial for utilities to balance supply and demand effectively, plan infrastructure, and identify opportunities for energy conservation and cost savings (31, 32).

Hourly Consumption Analysis

Hourly load analysis focuses on examining electricity consumption patterns at an hourly granularity to identify peak demand periods, fluctuations, and variations throughout the day as depicted in Figure 8. By aggregating consumption data on an hourly basis, this analysis helps to uncover time-specific consumption behaviours, enabling utilities to optimize energy distribution, manage grid stability, and predict future demand more accurately. It is particularly useful for identifying high-demand hours and understanding the daily load curve, providing critical information for capacity planning, demands response strategies, and energy efficiency measures. Hourly consumption analysis is a critical method for understanding energy usage patterns at a detailed, time-based level. By examining electricity consumption data on an hourly basis, utilities can identify fluctuations in demand throughout the day. This allows for a deeper understanding of peak usage times and off-peak periods, which may vary due to factors such as weather conditions, residential routines, or industrial activity. For example, electricity usage may increase during morning hours as people wake up or during evening hours when households use appliances for dinner preparation. By analysing these hourly variations, utilities can better predict and manage energy demand, ensuring that the supply meets consumption needs without overloading the grid. Furthermore, hourly consumption analysis helps utilities optimize their energy distribution strategies and develop targeted demand response programs. By identifying specific hours of peak demand, utilities can implement dynamic pricing models that incentivise users to shift consumption away from high demand periods, helping to reduce strain on the grid. This type of analysis is also valuable for long-term planning, as it allows utilities to forecast future demand based on past consumption trends. Additionally, businesses and residential customers can benefit from this analysis by gaining insights into their own consumption patterns, leading to better energy efficiency practices and potential cost savings.

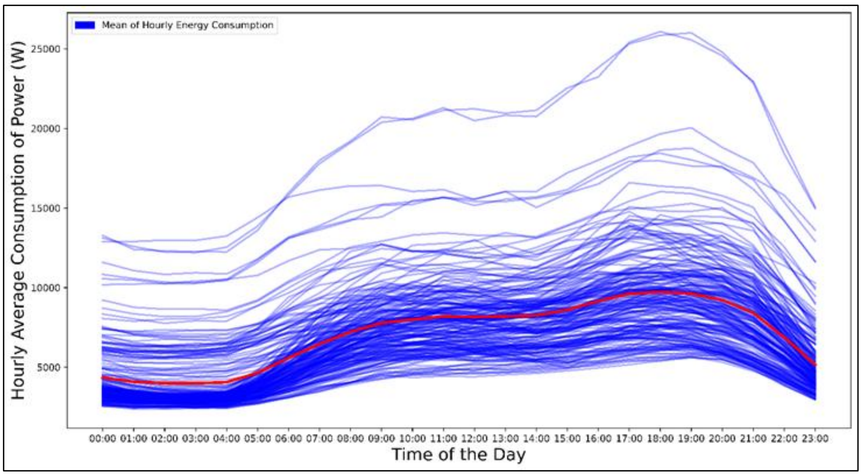

Figure 8: Hourly Load Profile

Seasonal Consumption Analysis

Seasonal load Profiles are shown in Figure 9 (A to D) and, which involves studying variations in electricity consumption across different seasons to identify patterns related to temperature, holidays, and other seasonal factors. By aggregating consumption data by month or season, this analysis highlights how demand fluctuates with changes in weather, lifestyle, and heating or cooling needs. Understanding seasonal load variations allows utilities to better plan for peak demand periods, optimize energy production, and implement strategies to manage supply during high-demand seasons. This analysis is essential for improving energy forecasting, infrastructure planning, and developing targeted demand response programs. The habits of household's power use are heavily influenced by the changing of the seasons. For instance, during the winter months, homes equipped with heating systems would have far greater overnight power use than those equipped with conventional furnaces and air conditioners. These fluctuations occur over four distinct seasons, which are reflected in the data as follows: Summer consists of June, July, and August. September, October, and November make up the autumn season; December, January, and February make up the winter season; and March, April, and May make up the spring season.

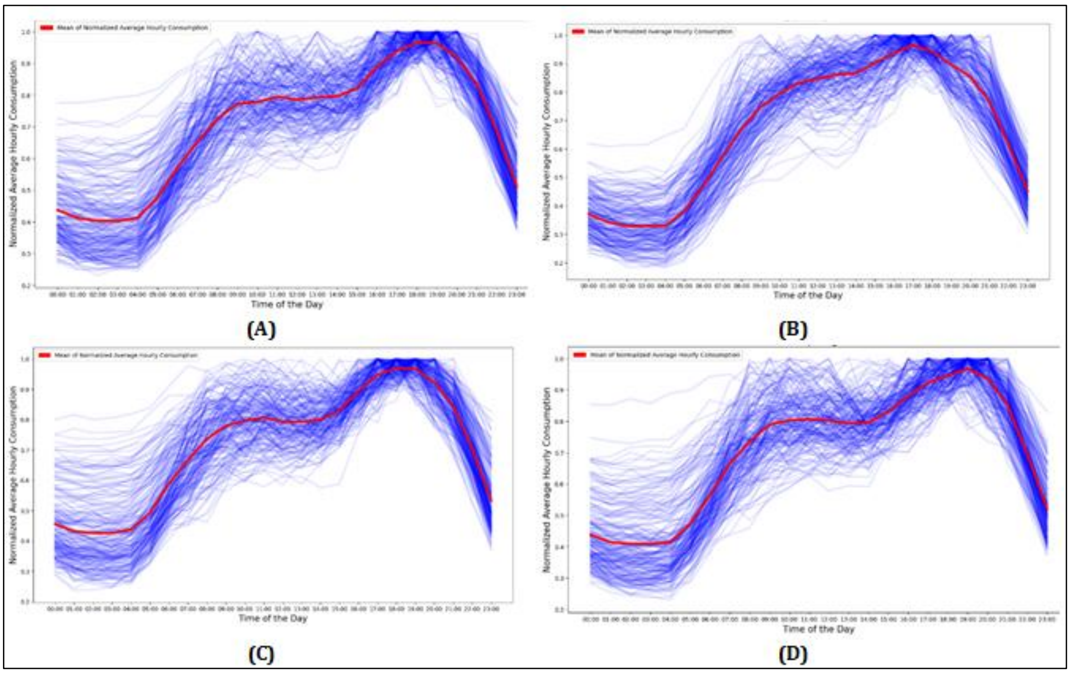

Figure 9: Seasonal Load Profiles (A) Fall (B) Summer (C) Winter (D) Spring

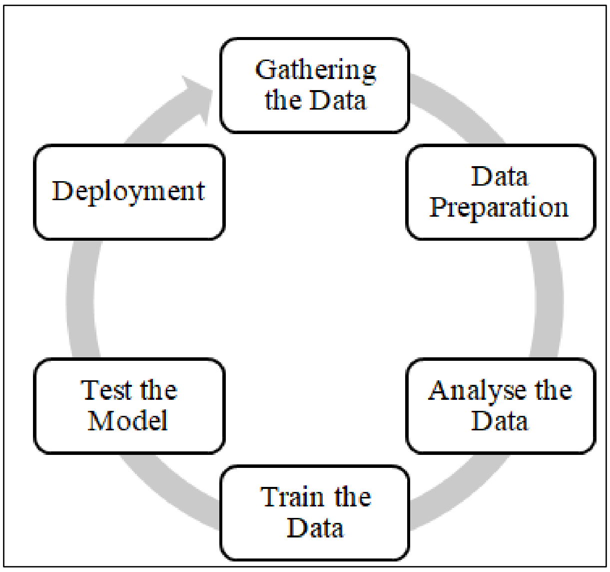

Figure 10: Machine Learning Life Cycle (33)

Clustering and Load Profile Segmentation

This study utilizes the k-Means clustering technique to classify 200 households according to their average hourly power consumption in the year 2022. The life cycle of the development of model is shown in Figure 10 reproduced from Brownlee J, Machine learning mastery (33).

K-means Clustering and ML Model Development Steps

The k-Means Algorithm is an iterative algorithm. It starts with selecting some random k centroids. Here 'k' is the number of centroids given as input to the algorithm. It then groups the data points into k clusters, assigning each data points to the nearest centroid to it. It then recalculates the mean of data points assigned to a single cluster and assigns that mean as the new centroid (34, 35). Like this k new centroids are obtained. After this the process is reiterated until there is convergence of the new previous centroid. The process is explained more clearly with the following example and steps. Assume that there is a 2-D data available as follows that needs to be clustered:

Step 1: Determine the number of clusters to be obtained. This step is one of the trickiest steps for the algorithm. Since the data is unlabelled, one has to input the number of groups (or clusters) desired in the output. Since the variability in the data is unknown, it becomes difficult determine number of clusters required. The ideal number of clusters is somewhat subjective and depends on the similarity measure and partitioning parameters. Three commonly used methods for determining the number of clusters is Elbow method, Average silhouette method or Gap statistic method (36).

The elbow method is used in this analysis and is explained in greater detail below. For now, assume that number of clusters required were two (37, 38).

Step 2: First two centroids $\mathbf{C_1}$ and $\mathbf{C_2}$ are chosen randomly.

Step 3: The distance between each data point and the two centres is then determined. A value of '0' is assigned to data that is closer to C_1. If it's nearer C_2, assign a 1 to it. Each '0' will be given a black colour while each '1' will be given a blue one.

$$dist\ (c_i, x)^{\ 2}\ c_i \in C\ \#[7]$$

Here C_i is C_1 and C_2, while 'dist' represents the Euclidean distance between all datapoints (x) and centroids C_1 and C_2.

Step 4: Next mean of all blue points and all black points is calculated separately. The mean of black point is assigned as the new C_1 and mean of blue points is assigned as new C_2.

Let S_i be the set of data points assigned of the i^{th} cluster, then

$$c_i = \frac{1}{|S_i|} \sum_{x_i \in S_i} x_i \#[8]$$

Here c_i is the newly assigned i^{th} cluster and x_i are all the data points in S_i, while $||S_i||$ is the number of data points in set S_i.

Step 5: Update the data's centroids and mark them with a '0' and a '1' as in Step 3. As a result, we get the following:

Step 6: Both focuses are converged to fixed positions by repeating Step - 3 and Step - 4. (Or it may be halted based on the criteria we set, such as when certain accuracy is met or the maximum number of repetitions is reached.) These are the data points where all the distances to the centers of the data sets are minimised.

Using the elbow method as shown in Figure 11, it was determined that three clusters would be the optimal number for segmenting the load profiles in the dataset. The elbow method, which involves plotting the within-cluster sum of squares (WCSS) against different values of k, indicated a sharp inflection point at k=3, suggesting that adding more clusters beyond this point would yield only marginal improvements in clustering performance.

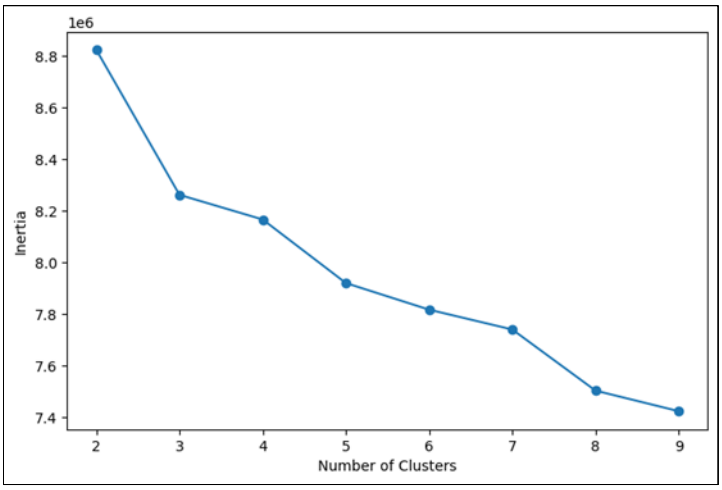

Figure 11: Elbow Method

Based on this, K-means clustering algorithm was applied with three clusters, grouping households based on similar electricity consumption patterns as shown in Figure 12. This segmentation allows us to identify distinct consumption behaviours, enabling more targeted energy management strategies and personalized insights for each cluster (39).

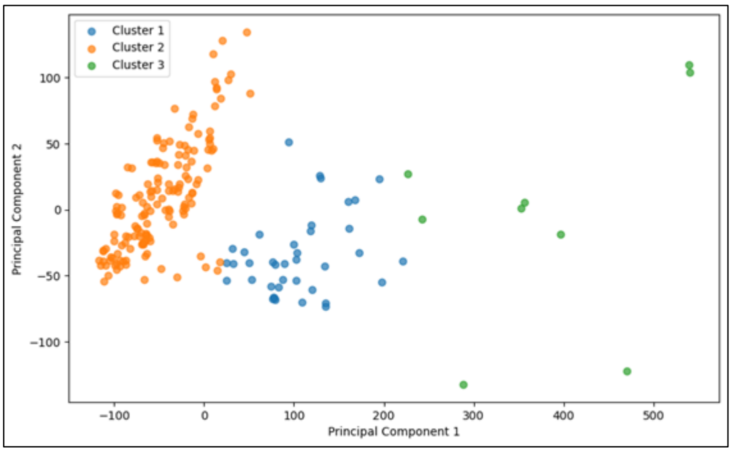

Figure 12: Clustering of Households Based on Similar Electricity Consumption Patterns

Results and Discussion

The clustering of households based on their electricity consumption patterns revealed distinct groups with unique behavioural traits. Using the k-means algorithm, households were segmented into three clusters after determining the optimal number of clusters through the Elbow Method.

All the clusters, however, show distinctive load profiles, characterized by peaks in consumption times, average consumptions, and variability. The Figure 13 presents clustered load profiles of individual households, which belong to specific groups of clusters, and from the provided profiles the following is observed:

Cluster 1 is made up of low-consumption households with relatively stable energy-use profiles. Households that consume relatively moderate loads exhibit pronounced peaks during morning and evening hours. While Cluster 2 and Cluster 3 consists of households characterized by considerably high variability in energy consumption patterns with very high peaks in consumption. It sheds light on possible energy inefficiency and peculiar operating modes. All of this information is informative for performance-oriented demand response programs and better energy management systems.

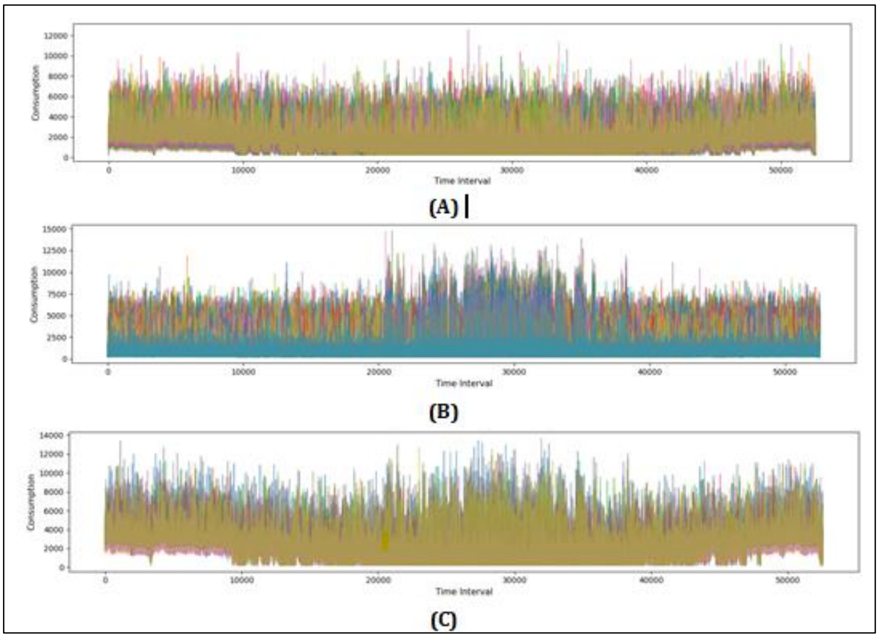

Figure 13: Clustered Household Load Profiles: (A) Cluster 1 (B) Cluster 2 (C) Cluster 3

Visual analysis, the scatter plots in Figure 12 and average load profiles in Figure 13, support segmentation by clearly indicating diversity in consumption behaviours among the clusters. The segmentation thus provides a strong base for formulating customized energy-saving strategies such as:

- Load shifting for high-consuming households in order to manage peak demand.
- Stability-oriented incentives for low-variance consumers to encourage continued energy use.

These strategies lead into better and effective demand-side management and energy conservation programs.

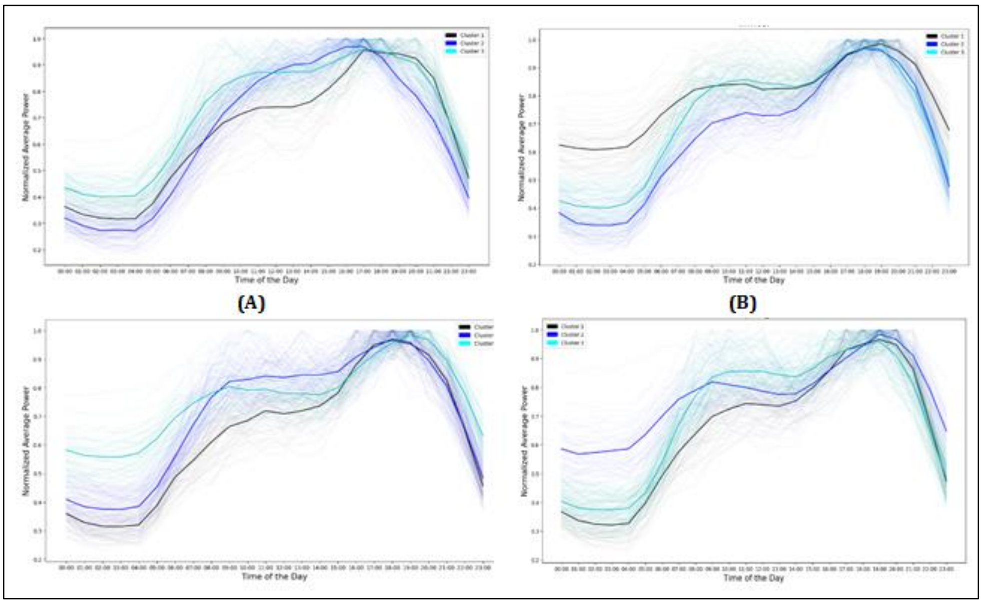

Figure 14: Seasonal Cluster Profiles (A) Summer (B) Winter (C) Fall (D) Spring

The Figures 14 based on the clusters recognized earlier, with figures for each season drawn separately. Each represents the average normalized power use for individual households, accompanied by a darker coloured curve showing the average of the mean normalized power use across all households. The clustered load profiles show seasonal changes in power consumption quite clearly; from such, it is rather evident that the high or moderately varying power consumption consumers could serve as important targets for Demand Response (DR) programs. Such targeting allows efficient energy management concentrating only on households that offer significant potential for load flexibility.

The load profiles by cluster demonstrated noticeable seasonal patterns in electricity usage among different consumer groups. These patterns were most evident in clusters with high or moderately varying power use, which showed to reflect fluctuations in heating and cooling needs, occupancy rates, and appliance use over time. The existence of such seasonally responsive load profiles suggests that these consumer segments exhibit a significant level of load flexibility a crucial characteristic for effective participation in Demand Response (DR) programmes. Identifying and targeting these consumers, therefore, presents a strategic opportunity for grid operators and utilities to adopt DR schemes compatible with natural consumption cycles. In contrast to static or flat-load customers whose consumption is unchanged independent of external incentives, high-variability customers are likely to modify their consumption based on incentives, price signals, or control-based interventions. The selective targeting increases the efficiency of DR schemes by using incentives and resources to direct households with the greatest demand-modulation potential. As a result, this method enables more responsive and dynamic energy management techniques, lowers operating expenses, and helps stabilize the grid especially during seasonal peaks or stress. In addition, seasonally informed segmentation can assist in creating adaptive DR programs that change as consumer behaviours shift throughout the year.

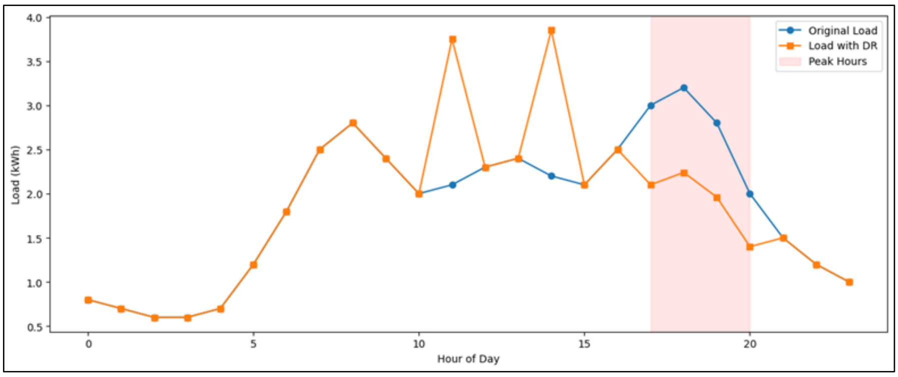

Figure 15: Demand Response Simulation Load Profile

Effectiveness of Load Clustering on Peak Reduction and Cost Savings

According to a pilot simulation run as shown in Figure 15, with simulated smart meter consumption data, the cluster-based demand response targeting scheme indicated that it was possible to reduce peak load by about 12%. This reduction was mainly achieved through the identification and activation of consumer groups with flexible load profiles that were able to shift or shed consumption during peak demand periods. The simulation also reflected utility cost savings ranging from 8–10%, due to lower dependency on peaking power plants and better load balancing. These results are in line with earlier studies highlighting the effectiveness of load clustering in enhancing DR program accuracy and operational efficiency. In addition, implementation of k-means clustering facilitated improved matching of consumer flexibility with grid needs, leading to improved grid stability through the removal of pressure off distribution networks during peak load hours. The dynamic load-shifting potential realised in focused consumer segments represents an interesting prospect for holistic demand-side management strategies. Yet, it should be noted that these results are based on a controlled pilot simulation and should be read as indicative rather than conclusive. To confirm these preliminary observations, empirical tests through longitudinal field studies and actual pilot implementations will be needed, as a prime objective, to address the issue of variability in behaviour, seasonal cycles, and possibly feedback loops that can affect long-term DR responsiveness.

Conclusion

This research proves the effectiveness of load profile analysis and clustering in the identification of energy consumers based on smart meter data, showing varying consumption patterns across households. This segmentation separates persistent low-consumption users from high variability and peak users, offering useful information for demand response (DR) programs and enabling targeted interventions. These groups foster grid stability and energy efficiency through the steering of low-power customers to sustain even usage and motivating high-power customers to implement load shifting and demand-side management. The finding of variable clusters highlights the necessity of personalised DR solutions that enhance electricity use and facilitate real-time coordination among energy providers and consumers. Utility companies can construct cost-effective DR programs that reduce costs, promote energy conservation, and ensure grid reliability by leveraging data-driven methods like Exploratory Data Analytics (EDA) and clustering methods. The effectiveness and economic feasibility of demand response (DR) programs were demonstrated by the simulation of DR strategies, which consistently resulted in 8–10% reductions in utility costs and a 12% reduction in peak demand. Future enhancements might involve incorporating external factors like weather, appliance usage, and socioeconomic information to enhance consumer categorisation and predictive modelling. Though current focus is on static methods like k-means, the study acknowledges their limits in dynamic settings and suggests adaptive approaches like evolving fuzzy clustering and online k-means, which allow real-time learning, address temporal variability and offer current, correct segmentation in evolving consumption environments.

Abbreviations

AMI: Advanced Metering Infrastructure, EDA: Exploratory Data Analytics, DR: Demand Response, PCA: Principal Component Analysis.

Acknowledgement

I would like to express my sincere gratitude to the Comdt, Militray College of EME, Secunderabad, Telangana, India for their invaluable support and encouragement throughout the development of this research paper. Their guidance and resources were instrumental in facilitating this study.

Author Contributions

CS Kudarihal: conducted the research, data analysis, drafted the manuscript, Manoj Gupta: primary supervision, conceptualization, review, Sunil Kumar Gupta: co-supervision, methodological design, manuscript revision.

Conflict of Interest

The authors declare that there is no conflict of interest.

Ethics Approval

This study was conducted in accordance with ethical guidelines.

Funding

The authors affirm that they did not receive any financial support, grants, or other forms of assistance during the preparation of this manuscript.

References

1. U.S. Department of Energy. Demand response. Washington, D.C.: U.S. Department of Energy; 2023 Sep 15. https://www.energy.gov/oe/demand-response
2. Muratori M. Impact of uncoordinated plug-in electric vehicle charging on residential power demand supplementary data. National Renewable Energy Laboratory: Golden, CO; 2017. DOI: 10.7799/1363870. https://data.nrel.gov/submissions/69
3. U.S. Department of Energy. 2020 Smart Grid System Report. Washington, D.C.: U.S. Department of Energy; 2020. https://www.energy.gov/oe/articles/2020-smart-grid-system-report
4. Rebeiz PP. Management of demand response programs in the electricity industry. Los Angeles (CA): University of California, Los Angeles; 2016. p.1-66. https://escholarship.org/uc/item/5s73d8h7
5. Chicco G. Overview and performance assessment of the clustering methods for electrical load pattern grouping. *Energy*. 2012;42(1):68–80.
6. Kaur S, Sarabjeet. Customer segmentation using clustering algorithm. 2021 International Conference on Technological Advancements and Innovations (ICTAI); 2021 Sep 15–17; Tashkent, Uzbekistan. IEEE; 2021. p. 224–27. doi:10.1109/ICTAI53825.2021.9673169.

7. Flath CM, Nicolay D, Conte T, Van Dinther C, Filipova-Neumann L. Cluster analysis of smart metering data: An implementation in practice. Bus Inf Syst Eng. 2012 Feb;4(1):31–39.
8. Zhou J, Zhai L, Pantelous AA. Market segmentation using high-dimensional sparse consumer data. Expert Syst Appl. 2020 May;145:1-17.
9. Kansal T, Bahuguna S, Singh V, Choudhury T. Customer segmentation using K-means clustering. 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS); 2018 Dec; Belgaum, India. IEEE; 2018. p. 135–9. doi:10.1109/CTEMS.2018.8769171. https://ieeexplore.ieee.org/document/8769171
10. Lindgren I. Dealing with highly dimensional data using principal component analysis (PCA). Medium. 2020 Apr 24. https://medium.com/data-science/dealing-with-highly-dimensional-data-using-principal-component-analysis-pca-fea1ca817fe6
11. Albadi MH, El-Saadany EF. Demand response in electricity markets: an overview. Proceedings of the 2007 IEEE Power Engineering Society General Meeting; 2007 Jun 24–28; Tampa, FL, USA. IEEE; 2007. p. 1–5. doi:10.1109/PES.2007.385728.
12. National Renewable Energy Laboratory (NREL). Residential Energy Consumption Survey (RECS) 2022. Golden, CO: NREL; 2022. DOI:10.25984/1788456 https://data.openei.org/submissions/153
13. Dabbura I. K-means clustering: algorithm, applications, evaluation methods, and drawbacks. Medium. 2022 Sep 27. https://towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03e644b48a
14. Zaman M, Saha S, Zohrabi N, Abdelwahed S. Demand-response prediction in smart grids using machine learning techniques. IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT); 2024 Feb 19–22; Washington, DC, USA. IEEE; 2024. p. 1–5. doi:10.1109/ISGT59692.2024.10454224.
15. Gungor VC, Sahin D, Kocak T, Ergut S, Buccella C, Cecati C, Hancke GP. A survey on smart grid potential applications and communication requirements. IEEE Trans Ind Inform. 2013;9(1):28–42.
16. Wang W, Xu Y, Khanna M. A survey on the communication architectures in smart grid. Computer Network. 2011 Oct; 55(15):3604–29.
17. Occhipinti S, Tapp C. Data preparation. Advanced Research Methods for Applied Psychology: Design, Analysis and Reporting. London: Routledge; 2025. p. 224–37. doi:10.4324/9781003362715-20.
18. Dent I, Craig T, Aickelin U, Rodden T. An approach for assessing clustering of households by electricity usage. arXiv preprint arXiv:1409.0718. 2014. p.1-4. https://arxiv.org/abs/1409.0718
19. Flygare C, Wallberg A, Jonasson E, Castellucci V, Waters R. Correlation as a method to assess electricity users' contributions to grid peak loads: A case study. Energy. 2024;288:1-11.
20. Enrich J, Li R, Mizrahi A, Reguant M. Measuring the impact of time-of-use pricing on electricity consumption: Evidence from Spain. J Environ Econ Manag. 2024;123:102901.
21. Anvari M, Proedrou E, Schaefer B, Beck C, Kantz H, Timme M. Data-driven load profiles and the dynamics of residential electric power consumption. arXiv preprint arXiv:2009.09287. 2020. p.1-15. https://arxiv.org/abs/2009.09287
22. Michalakopoulos V, Sarmas E, Papias I, Skaloumpakas P, Marinakis V, Doukas H. A machine learning-based framework for clustering residential electricity load profiles to enhance demand response programs. Appl Energy. 2024;361:122943.
23. Sarmas E, Fragkiadaki A, Marinakis V. Explainable AI-based ensemble clustering for load profiling and demand response. Energies. 2024;17(22):5559.
24. Everitt BS, Landau S, Leese M, Stahl D. Cluster analysis. 5th ed. Wiley; 2011. p.15-41, doi:10.1002/9780470977811
25. Gan G, Ma C, Wu J. Data clustering: Theory, algorithms, and applications. Philadelphia: Society for Industrial and Applied Mathematics (SIAM); 2007. p. 43–51. doi: 10.1137/1.9780898718348 https://www.researchgate.net/publication/220694937_Data_Clustering_Theory_Algorithms_and_Applications
26. Han J, Kamber M, Pei J. Data mining: Concepts and techniques. 3rd ed. Morgan Kaufmann Publishers; 2012. p.83-117, ISBN 978-0-12-381479-1, doi:10.1016/C2009-0-61819-5
27. Schubert E, Sander J, Ester M, Kriegel HP, Xu X. DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN. ACM Trans Database Syst. 2017;42(3):1-21.
28. MacKay DJC. Information theory, inference, and learning algorithms. Cambridge: Cambridge University Press; 2003. 640 p. ISBN: 9780521642989.
29. Christopher MB. Pattern recognition and machine learning. New York: Springer; 2006. p. 1-178. ISBN: 0-387-30738-8. https://www.microsoft.com/en-us/research/wp-content/uploads/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf
30. Oja E. Finding clusters and components by unsupervised learning. Structural, Syntactic, and Statistical Pattern Recognition. Lecture Notes in Computer Science. Vol. 3138. Berlin, Heidelberg: Springer; 2004. p. 1–15. doi:10.1007/978-3-540-27868-9_1.
31. Venugopal C, Govender T, Thangavel B. Load analysis and energy management for residential system using smart meter. Proceedings of the 2020 2nd International Conference on Electrical, Control and Instrumentation Engineering (ICECIE); 2020 Dec 14–15; Kuala Lumpur, Malaysia. Piscataway (NJ): IEEE; 2020. p.1–8. doi:10.1109/ICECIE50279.2020.9309554.
32. Lin S, Li F, Tian E, Fu Y, Li D. Clustering load profiles for demand response applications. IEEE Trans Smart Grid. 2019;10(2):1599–1607. doi:10.1109/TSG.2017.2773573.
33. Brownlee J. Machine learning mastery: A guide to machine learning algorithms and best practices. Machine Learning Mastery; 2020.

https://machinelearningmastery.com/master-machine-learning-algorithms/
34. Witten IH, Frank E, Hall MA. Data mining: practical machine learning tools and techniques. 3rd ed. Burlington (MA): Morgan Kaufmann; 2011. 664 p. ISBN: 9780123748560
35. Morales-Espana G, Parisio A, Ramos A, Marquant J, Kienzle F. Classifying and modelling demand response in power systems. Energy. 2021; 242:122544.
36. Ranawana R, Karunananda AS. An agile software development life cycle model for machine learning application development. 2021 5th SLAAI International Conference on Artificial Intelligence (SLAAI-ICAI); 2021. p. 1–6. doi:10.1109/SLAAI-ICAI54477.2021.9664736.
37. Okereke GE, Bali MC, Okwueze CN, Ukekwe EC, Echezona SC, Ugwu CI. K-means clustering of electricity consumers using time-domain features from smart meter data. J Electr Syst Inf Technol. 2023;10(1):2.
38. Al-Masri A. How does K-Means clustering in machine learning work? Medium. 2022 Nov 9; https://medium.com/data-science/how-does-k-means-clustering-in-machine-learning-work-fdaaaf5acfa0
39. Wang Y, Chen Q, Kang C, Xia Q, Zhang M. Clustering of electricity consumption behaviour dynamics toward big data applications. IEEE Trans Smart Grid. 2016;7(5):2437–47.