Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Towards a data-driven profiles of household energy consumption in collective housing: Machine learning clustering with insights from Bab Ezzouar (Algiers)

Kahina Ouazar

kouazar@usthb.dz

Department of Geography and Territorial Planning, LREAU Laboratory, University of Sciences and Technology Houari Boumediene (USTHB), Algeria

Guglielmina Mutani

Department of Energy, Politecnico Di Torino, Italy

Leila Abbas Mansour

Department of Geography and Territorial Planning, LREAU Laboratory, University of Sciences and Technology Houari Boumediene (USTHB), Algeria

---

Research Article

Keywords: Data-driven, Electricity and Gas consumption, Residential buildings, Machine Learning, Multi-energy clustering, Households profiling, Algeria

Posted Date: April 7th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9289979/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations:
The authors declare no competing interests.

Ethics statement: The household survey conducted for this study relied on the voluntary participation of respondents. Informed consent was obtained from all participants prior to data collection. The survey responses were collected anonymously and used exclusively for academic research purposes. This research was carried out as part of a national doctoral thesis within an academic framework under the supervision of the Directorate-General for Scientific Research and Technological Development and

Houari Boumediene University of Science and Technology, which approved and oversaw the overall research process.

Towards a data-driven profiles of household energy consumption in collective housing: Machine learning clustering with insights from Bab Ezzouar (Algiers).

Kahina Ouazar [1]*, Guglielmina Mutani[2], Leila Abbas Mansour [3]

[1, 3] Department of Geography and Territorial Planning, University of Sciences and Technology Houari Boumediene (USTHB), Algiers 16000, Algeria

[2] Department of Energy, Politecnico Di Torino, 10129 Torino, Italy

Correspondent author: Kahina Ouazar(kouazar@usthb.dz)

Abstract

This study develops a multi-criterion, multi-energy clustering framework for residential energy profiling, integrating five years of electricity and natural gas consumption data with household survey data collected from occupants of collective residential buildings (valid sample: n = 380). The proposed approach explicitly incorporates socio-demographic characteristics, occupant perceptions and behaviours, dwelling geometry, and envelope thermal properties, enabling a comprehensive representation of residential energy systems. The methodology combines univariate statistical analysis and principal component analysis (PCA) for dimensionality reduction, followed by a comparative evaluation of multiple Machine Learning (ML) clustering algorithms, including both hierarchical and non-hierarchical methods, to assess the robustness and sensitivity of the results. Among the tested models, K-means demonstrates the best performance (average Silhouette score = 0.54), consistently identifying an optimal structure of three household clusters with satisfactory internal cohesion. The findings highlight the critical role of integrating behavioural and physical dimensions in capturing the complexity of residential energy consumption patterns. Applied to the municipality of Bab Ezzouar in Algiers, Algeria, the study provides novel empirical insights from an underrepresented North African context. It contributes to the literature by advancing multi-dimensional clustering approaches and supports the development of targeted and context-sensitive energy efficiency strategies in collective housing.

Keywords: Data-driven; Electricity and Gas consumption; Residential buildings; Machine Learning; Multi-energy clustering; Households profiling; Algeria

1. Introduction

The building sector is central to global energy transition, accounting for over one-third of final energy consumption worldwide [1]. In residential contexts, improving efficiency remains challenging due to the combined influence of building attributes, energy systems, and occupant-related factors [2]. Capturing this socio-technical complexity is essential for reliable energy assessment, policy design, and targeted demand-side interventions [3,4]. Residential energy use emerges from the interplay of physical, technical, and behavioural determinants [5] . Physically, envelope characteristics: thermal transmittance, glazing, window-to-wall ratio, and shading; largely govern heat gains and losses, while geometry, orientation, and urban form modulate solar exposure and energy demand [6–9] . Technical systems, including HVAC, lighting, and on-site renewables, further shape performance [10–12] alongside climatic variability [13].

Occupant behaviour, comprising conscious and unconscious interactions with buildings and energy systems [14] ; remains a major source of variability and uncertainty [15]. Daily presence, thermostat settings, window operation, and appliance use generate substantial gaps between predicted and observed consumption [16–18]. This behavioural heterogeneity is context-dependent and dynamic, challenging conventional energy modelling [19], and motivating analytical approaches capable of structuring occupant diversity [18].Clustering techniques have therefore emerged as a key tool for profiling residential energy use, reducing complexity while retaining meaningful heterogeneity[20].

Related work

Unsupervised machine learning has been widely employed for residential energy profiling, including K-means [21,22], K-medoids [23], hierarchical clustering [24,25] , Gaussian mixture models [26,27], DBSCAN [21], and self-organising maps [28]. Hierarchical clustering is valued for interpretability, whereas K-means is preferred for efficiency and scalability [29]. Comparative frameworks are common, evaluating multiple algorithms to select the most suitable approach. Recent studies increasingly adopt hybrid methods, combining clustering with statistical learning, simulation, or deep

learning [30–33] integrating diverse data sources such as high-resolution smart meters and survey-based socio-demographics. While the former provides temporal granularity, the latter offer broader contextual insight.

Feature engineering critically affects clustering outcomes. Beyond time-series indicators (peak demand, load variability) [34], contextual variables: building attributes, energy systems, and socio-demographics; are increasingly incorporated, shifting analyses from signal-centric to socio-technical frameworks. These approaches better capture occupant, building interactions but increase complexity. Clustering has revealed behavioural patterns linked to routines and socio-demographics, with appliance use and air-conditioning behaviour strongly influencing energy demand [22,35,36]. Yet context-specific behavioural models remain scarce [37].

Current limitations include reliance on single energy carriers, short-term datasets, partial multi-criteria integration, and geographic focus on developed regions [38–40].

To address these gaps, this study proposes a multi-criteria, multi-energy clustering framework combining long-term electricity and natural gas consumption with household survey data. Inputs explicitly integrate socio-demographic characteristics, occupant behaviours, dwelling geometry, and envelope thermal properties. Applied to a residential case study in Algiers, Algeria, the methodology evaluates hierarchical and non-hierarchical algorithms to assess sensitivity to methodological choices. This work advances residential energy profiling by: (i) integrating socio-technical and physical variables, (ii) extending clustering to multi-energy, long-term datasets, and (iii) providing methodological guidance for underrepresented contexts. It delivers novel empirical insights from North Africa and methodological benchmarks for multi-dimensional clustering in data-scarce regions.

2. Material and Methods

2.1 Case study

The study focuses on Bab Ezzouar, in eastern Algiers (Fig. 1). It is a municipality originally developed as a “cité-dortoir”, primarily residential; and now characterised by dense collective housing estates.

It represents one of the most illustrative examples of large-scale social housing implemented under national programmes since the 1970s, with early developments (1970s-1980s) coexisting alongside more recent projects (2000s-2015), reflecting sustained urban expansion. Today, the municipality comprises eight major urban units and hosts over 100,000 inhabitants across more than 17,000 collective housing units [41]. The study area lies within climatic zone A, as defined by DTR 2.3/4 ([42], corresponding to Algeria's coastal regions. It is characterised by a Mediterranean climate, with hot, dry summers and mild, relatively wet winters, which significantly influence residential energy demand[43].

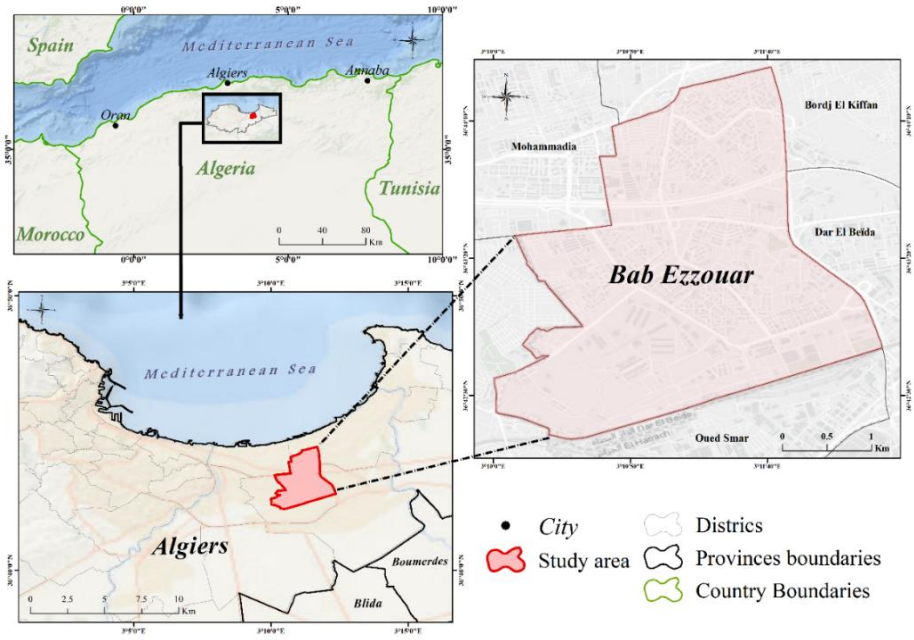


Fig. 1: Case study location.

At the national level, the residential sector represents a major share of final energy consumption[44], primarily supplied by electricity and natural gas [45]. Natural gas is primarily utilised for space heating, domestic hot water production, and cooking, whereas electricity is predominantly used for air conditioning, lighting, and household appliances. In public collective housing, particularly in older developments; the absence of centralised or hybrid energy systems means that energy provision relies primarily on individual appliances at the dwelling level, resulting in fragmented and often suboptimal patterns of energy use.

From a morphological perspective, the housing stock is highly standardised, comprising primarily F3 (two-bedroom) and F4 (three-bedroom) apartments within mid-rise (G+4) and high-rise (G+9 to G+14) buildings, with Table 1 summarising the key characteristics of these typical dwelling types. These buildings were constructed using similar materials and construction techniques, typically featuring reinforced concrete structural frames with double-leaf external walls made of hollow bricks and floor slabs composed of solid or hollow-core concrete elements. This architectural homogeneity is accompanied by relatively uniform socio-economic conditions, characterised mainly by middle-income households and comparatively large family sizes. Nevertheless, variations in occupants' daily practices and behavioural patterns introduce additional diversity in residential energy consumption.

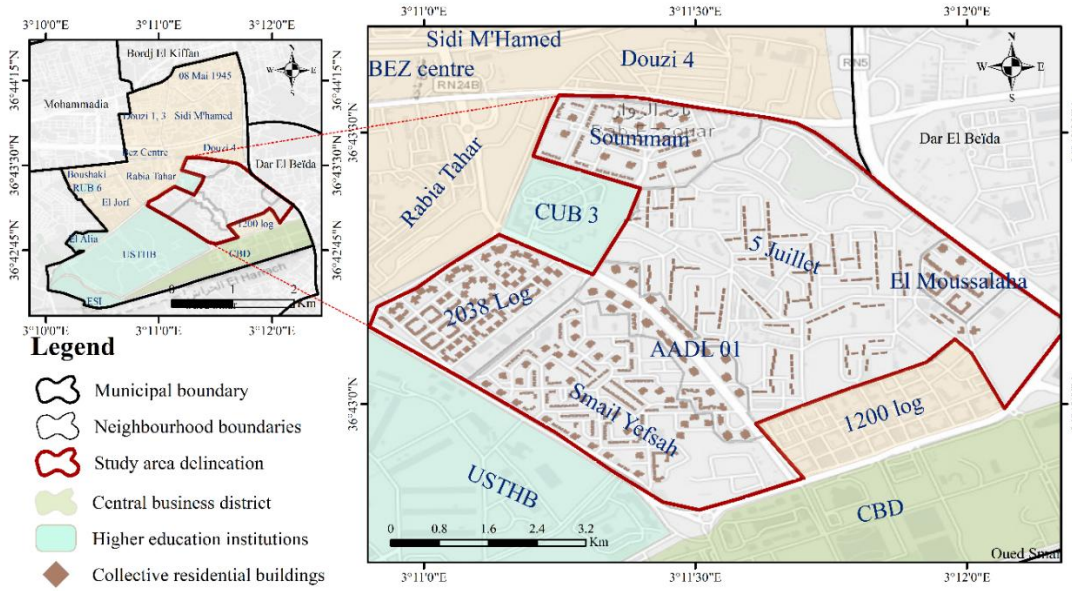
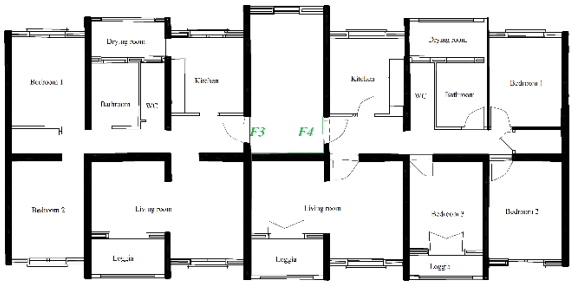


Fig. 2: Study area delineation with 3d building typologies and typical apartment layouts.

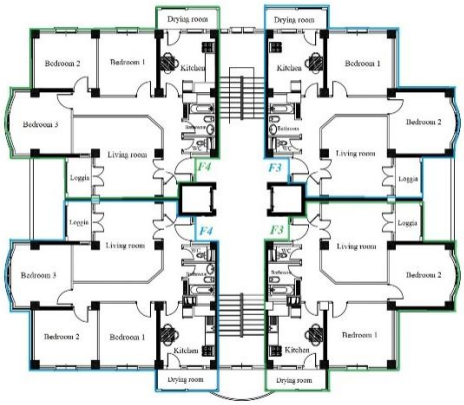
Given these characteristics, Bab Ezzouar provides a representative context for analysing residential energy use in Algerian collective housing. The analysis focuses on five eastern districts selected for their high density and availability of architectural data, reflecting the dominant housing typologies within the municipality (Fig. 2).

Table 1: Characteristics of typical dwellings in the study area.

Floor plan	Dimensions		
	Type 01: 5-story multi-family residential buildings with 2 dwelling units per floor, built using brick masonry construction (1970-1995)		
		F3	**F4**
	Total useful area	69	79

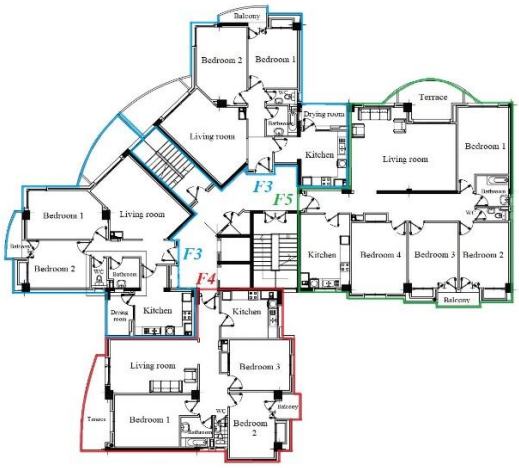


Wall area	46	55
Ceiling height	2.6	2.6
Total windows area	13	16
S/V	0.4	0.45



Type 02: A 14-storey multi-family residential building with four dwelling units per floor, built using brick masonry construction (2001-2003)

	F3	**F4**
Total useful area	72	87
Wall area	50	70
Ceiling height	2.7	2.7
Total windows area	8	10
S/V	0.4	0.5



Type 03: A 9-story multi-family residential building with four dwelling units per floor, built using brick masonry construction (2015)

	F3	**F4**	**F5**
Total useful area	75	88	93
Wall area	65	78	84
Ceiling height	3	3	3
Total windows area	13.2	12.8	19
S/V	0.4	0.5	0.44

2.2 Methodology

To identify household energy consumption profiles in collective housing, a structured workflow was followed: data collection, preprocessing, PCA-based dimensionality reduction, clustering with multiple algorithms, internal validation via indices, and cluster interpretation (Fig. 3)

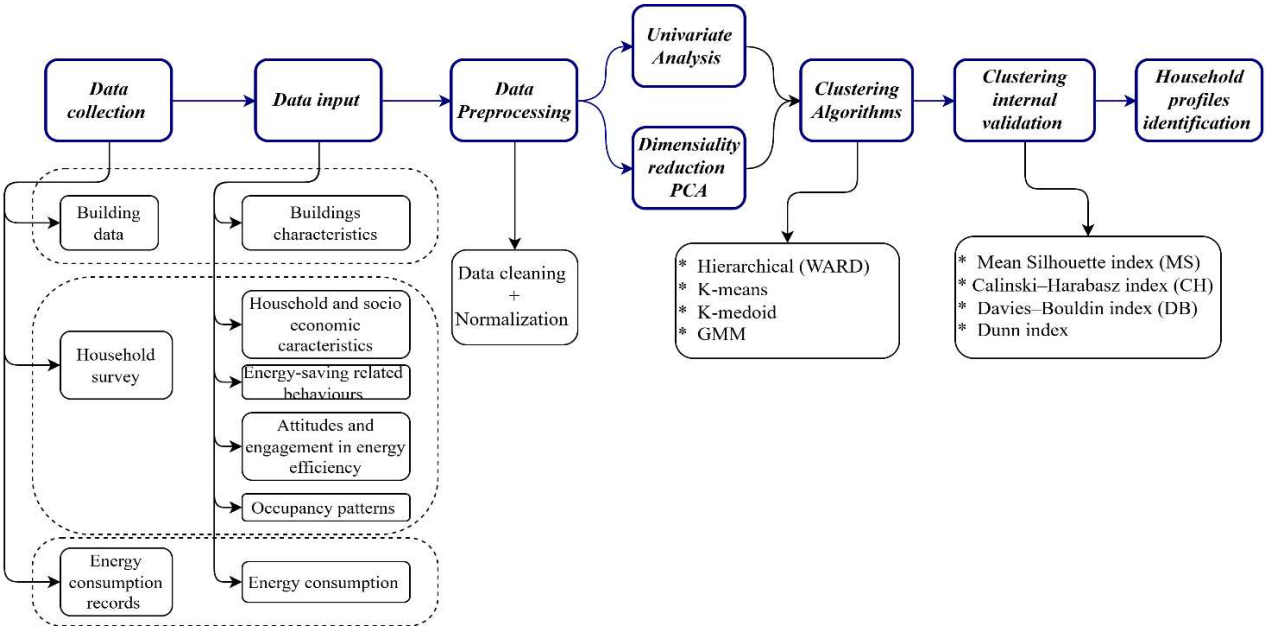


Fig. 3*:* Methodological framework adopted in the study

2.2.1 Data collection

- **Questionnaire survey**

A survey was conducted between November 2024 and June 2025 across the study area, focusing on public collective housing. A spatially targeted sampling approach was adopted, concentrating on districts with a high density of collective residential buildings and available building documentation. Household responses were collected through door-to-door surveys based on voluntary participation, potentially introducing self-selection bias. The survey was administered orally through face-to-face interviews using a structured questionnaire comprising six thematic sections (illustrated in Fig. 4 and detailed in Appendix A.), delivered by trained personnel. It covered basic household characteristics, energy systems and appliances in use, adoption of energy-saving measures, attitudes and perceptions regarding energy efficiency, as well as occupancy patterns and housing (inspired from [4,20,46,47] and adapted to the local context. The survey yielded 570 valid responses. All procedures were carried out by trained personnel and adhered to established ethical research standards, ensuring informed consent, confidentiality, and participant anonymity. The study complied with the Declaration of Helinski [48] and did not require formal ethics approval under Algerian regulations governing anonymous, non-interventional research.

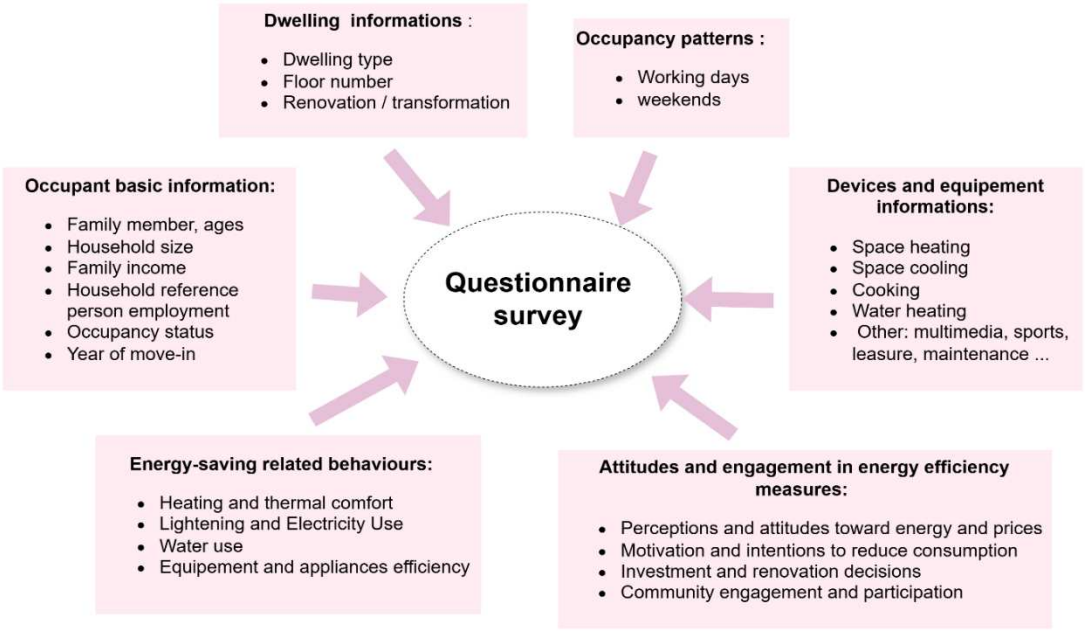


Fig. 4: Questionnaire design used in the study

- **Dwelling information**: includes the geometry and the characteristics of the construction materials, obtained from the institutions responsible for the construction and management of the residential stock (public institutions and design offices). Information was extracted from the architectural plans (AutoCAD files) and civil engineering documents relating to the analysed buildings. The thermal properties of the building envelopes were determined in accordance with Algerian building thermal regulations (DTR C3.2/4)[42], enabling the calculation of thermal coefficients for two distinct periods: summer and winter.

- **Energy consumption records:** Alongside the survey, we collected longitudinal data on electricity and gas usage from SONELGAZ, the state-owned company responsible for energy supply and management in Algeria [49]. The dataset spans a five-year period (2020-2024) providing consumption values. For each household, the consumption is recorded in kilowatt-hours (kWh) for electricity and Therms (Th) for gas, with a total of 16 quarterly values per energy type (four per year). Gas consumption was then converted to kWh using a specific conversion factor. Subsequently, annual primary energy consumption values were computed by summing electricity and converted gas consumption for each year, resulting in five annual energy consumption figures per household.

2.2.2 Preprocessing data

Data were cleaned to address missing values and inconsistencies, resulting in a validated sample of 380 households from the initial 570, ensuring reliable information on energy consumption and household characteristics. New indicators were created to capture behavioural, perceptual, and physical aspects, and survey variables were recoded numerically. Both electricity and natural gas consumptions were converted into primary energy (kWh) over the period 2020-2024 and subsequently averaged across the five years in order to obtain a robust and representative estimate of long-term household energy use. This methodological choice mitigates the influence of short-term variability arising from climatic fluctuations, behavioural changes, or transient occupancy patterns. It is therefore consistent with the study's objective of identifying stable socio-technical consumption profiles, rather than capturing short-term or episodic variations.

Energy consumption values were log-transformed (base 10) to stabilise variance, while other X variables were standardised using the z-score, using average μ and standard deviation σ (Eq.1), to ensure comparability and statistical consistency for subsequent analyses.)

$$Z = (X - \mu)/\sigma \quad (1)$$

2.2.3 Filtering, verification and dimensionality reduction of variables

The initial dataset consisted of a range of explanatory variables (see Table 1.A in Appendix), which were processed, coded, and structured into seven distinct categories. To mitigate redundancy caused by multicollinearity, a Pearson correlation analysis [50] was first carried out, adopting a 0.95 threshold [51] Subsequently, a Principal Component Analysis (PCA) was employed to further reduce the dimensionality of the data[52].

Prior to running the PCA, two standard diagnostic tests were performed to verify the adequacy of the dataset: The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy[53], calculated using Eq. 2, and Bartlett's test of sphericity [54]**.** Both confirmed the relevance of inter-variable correlations and

supported the use of PCA. Variables presenting an individual KMO value below 0.7 were excluded to ensure the robustness of the factor structure and improve the quality of the retained components [53].

$$KMO = \frac{\sum_{i \neq j} r_{ij}^2}{\sum_{i \neq j} r_{ij}^2 + \sum_{i \neq j} p_{ij}^2} \quad (2)$$

where r_{ij} denotes the correlation coefficient between variables i and j, p_{ij} represents the corresponding partial correlation coefficient, and $\sum_{i \neq j}$ denotes the summation over all variable pairs.

2.2.4 Profiles clustering

Variables derived from PCA formed the basis for clustering, with multiple unsupervised machine learning algorithms: hierarchical and non-hierarchical; applied to determine the optimal number of groups.

- **Hierarchical clustering** (Ward's method, Ward.D2 in *R*) was first used to explore the intrinsic data structure without predefining the number of clusters. By iteratively merging the most similar groups based on squared Euclidean distances [55], it minimized within-cluster variance and produced a dendrogram that visually guided the selection of the optimal cluster configuration [56].
- **K-means clustering** was then applied for efficient partitioning and as a benchmark method. It minimizes within-cluster variance to produce compact, well-separated groups suitable for large datasets[57]. The optimal number of clusters (k) was determined using two complementary criteria: the Elbow, silhouette and average silhouette methods; ensuring robustness through multiple calibration perspectives.
- **K-medoids clustering** (Partitioning Around Medoids; PAM algorithm)[58] offered greater resistance to outliers by representing each cluster with a medoid; an actual data point, rather than a centroid[59]. The best k was selected using the average silhouette method to ensure balanced cohesion and separation[60].

- Finally, **Gaussian Mixture Models (GMM)** provided a probabilistic approach, assigning each observation a likelihood of belonging to different clusters[61] Each cluster was defined by its own mean and covariance matrix, with model selection based on the Bayesian Information Criterion (BIC) Additional refinements, outlier removal, covariance testing, and variable selection; further improved stability and accuracy.

Together, these methods offered complementary insights: hierarchical clustering revealed structure, K-means and K-medoids ensured robust partitioning, and GMM captured uncertainty in cluster membership, reinforcing the overall reliability of the analysis.

2.2.5 Clustering comparison and validation

The clustering algorithms were evaluated using four complementary internal validation indices (Table 2), capturing cohesion, separation, and compactness. Combined, these metrics provide an objective and consistent framework to assess algorithm performance and identify the most robust cluster structure.

Table 2: Internal validation indices used to evaluate clustering performance

Index	Function	Formula	Reference
Average silhouette coefficient	Assesses intra-cluster cohesion and inter-cluster separation.	$S_i = \frac{(b_i - a_i)}{\max(a_i, b_i)}$ (3) a_i is the average intra-cluster distance for observation i, and b_i is its minimum average distance to the nearest cluster.	[60]
Calinski-Harabasz index (CH)	Evaluates the ratio of between-cluster to within-cluster dispersion.	$CH(K) = \frac{B(K)/(K-1)}{W(k)/(n-K)}$ (4) $B(K)$ represents the between-cluster dispersion, $W(K)$ the within-cluster dispersion, K the number of clusters, and n the total number of observations.	[62]
The Davies–Bouldin index (DB)	Measures the average similarity between clusters based on compactness and separation.	$DB = \frac{1}{k}\sum_{i=1}^{k} max_{(j \neq i)} \left(\frac{S_i + S_j}{d_{ij}}\right)$ (5) k is the number of clusters, S_i the average intra-cluster distance for cluster i, and d_{ij} the inter-centroid distance between clusters i and j	[63]
Dunn index	Favors cluster configurations with high inter-cluster separation and low intra-cluster dispersion	$DI = \frac{\min_{i \neq j} \mathrm{d}(c_i, c_j)}{\max_k \Delta_k}$ (6) $d(c_i, c_j)$ is the distance between the centroids of clusters i and j, and Δ_k denotes the maximum distance between observations within cluster k	[64]

2.2.6 Profiles identification

After identifying the optimal clusters, household profiles were defined via socio-demographics, energy use, and behaviours, with cluster means forming an operational typology. Analyses were performed in RStudio using *psych, FactoMineR, factoextra, mclust, clusterCrit, fpc,* and *tidyverse.*

3. Results

The initial dataset, combining survey responses, energy consumption, and dwelling characteristics, comprised 70 variables grouped into seven categories: Socio-Economic Characteristics, Building Characteristics, Energy Consumption, Occupancy Patterns, Systems and Equipment Ownership, Behavioural Actions, and Perceptions and Engagement in Energy Efficiency (see Table B1, Appendix B).

Feature engineering added several indicators, including the Energy Sobriety Index (ESI)**,** Perceptions and Willingness for Energy Efficiency (WEE)**,** Household Equipment Index (HEI), and occupancy ratios for weekdays and weekends. U-values for windows, walls, doors, ceilings, and floors were also calculated. The energy consumption variable was defined as the average of the annual electricity and natural gas consumption over the five-year period (2020-2024), providing a stable long-term representation of household demand. After these transformations, a refined dataset of 47 coded and distinct variables was obtained for the subsequent analyses (univariate and PCA).

3.1 Univariate analysis and PCA

A $\log_{10}$ transformation was applied to the energy consumption variable (Fig. 5), and the dataset was standardised accordingly. Subsequently, variables exhibiting high collinearity ($p>0.05$) were removed to reduce redundancy.

The correlation matrix (Fig. 6) was used to assess data adequacy through the Kaiser-Meye-Olkin (KMO) test. The initial KMO value of 0.65 indicated moderate suitability. To improve reliability, variables with individual KMO values below 0.70 were removed, eliminating 20 variables. The remaining 26 variables yielded a higher global KMO of 0.79 (Table 3).

The Bartlett test ($\chi^2 = 4630.214$, df = 325, $p < 0.001$) confirmed significant inter-variable correlations, validating the dataset for PCA. The analysis retained ten components explaining 74.5 % of total variance, effectively capturing the latent structure and reducing dimensionality for subsequent clustering and modelling (Fig. 7).

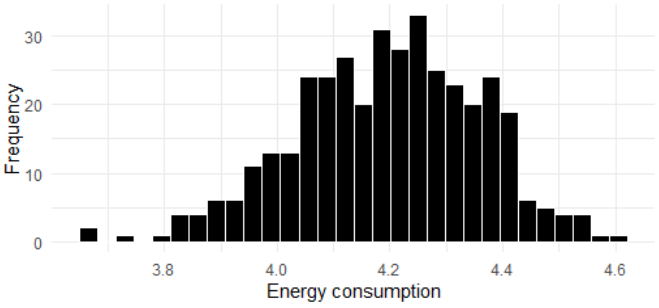


Fig. 5: Log Five-year (2020-2024) average annual energy consumption (Gas & Electricity) (kWh)

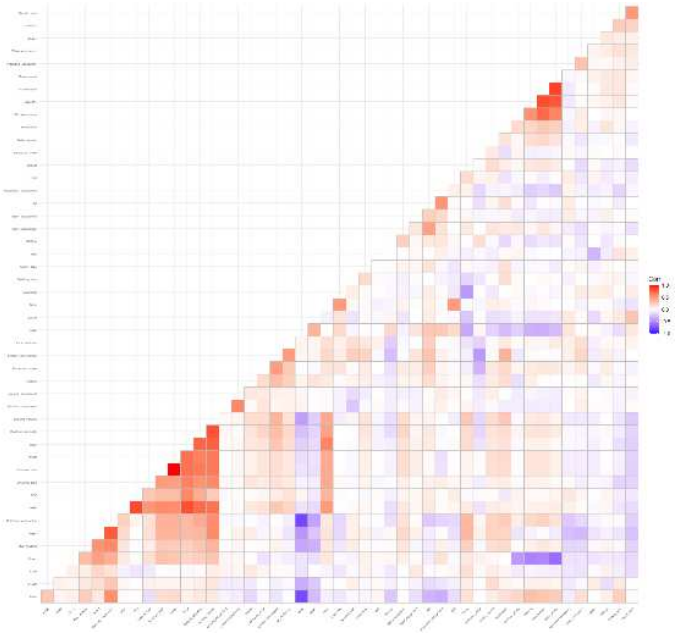


Fig. 6: Correlation matrix

Table 3: Individual KMO scores of the dataset's variables

Overall MSA = 0.79					
Energy consumption	Willingness and Engagement score	Energy sobriety index	Year_ housing	Building construction period	Dwelling type
0.71	0.79	0.7	0.83	0.77	0.82
Number floor	Dwelling perimeter	THLP	THLA	WWR	U_{value}windows
0.73	0.89	0.78	0.83	0.84	0.72
U_{value}door	Age 18-29	Age 30-49	Income level	HRPjob	HEI
0.75	0.74	0.86	0.74	0.73	0.77
Space Cooling (air-conditioning)	Retrofit interventions	Household size	Thermal insulation measures	Windows substitution	Door replacement
0.72	0.74	0.7	0.74	0.77	0.8
Balcony extension	Water heating system				
0.83	0.73				

MSA: mean silhouette average

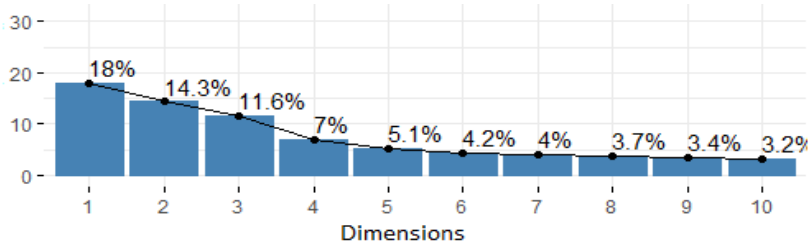


Fig. 7: Distribution of Explained Variance across the First Ten Principal Components

This initial stage, comprising variable transformation, univariate statistical analysis, and principal component analysis (PCA); is illustrated in Fig. 8, which provides a schematic overview of the evolution of variable categories throughout the study, from raw input data to pre-processed variables and, ultimately, to the final set of features retained after dimensionality reduction.

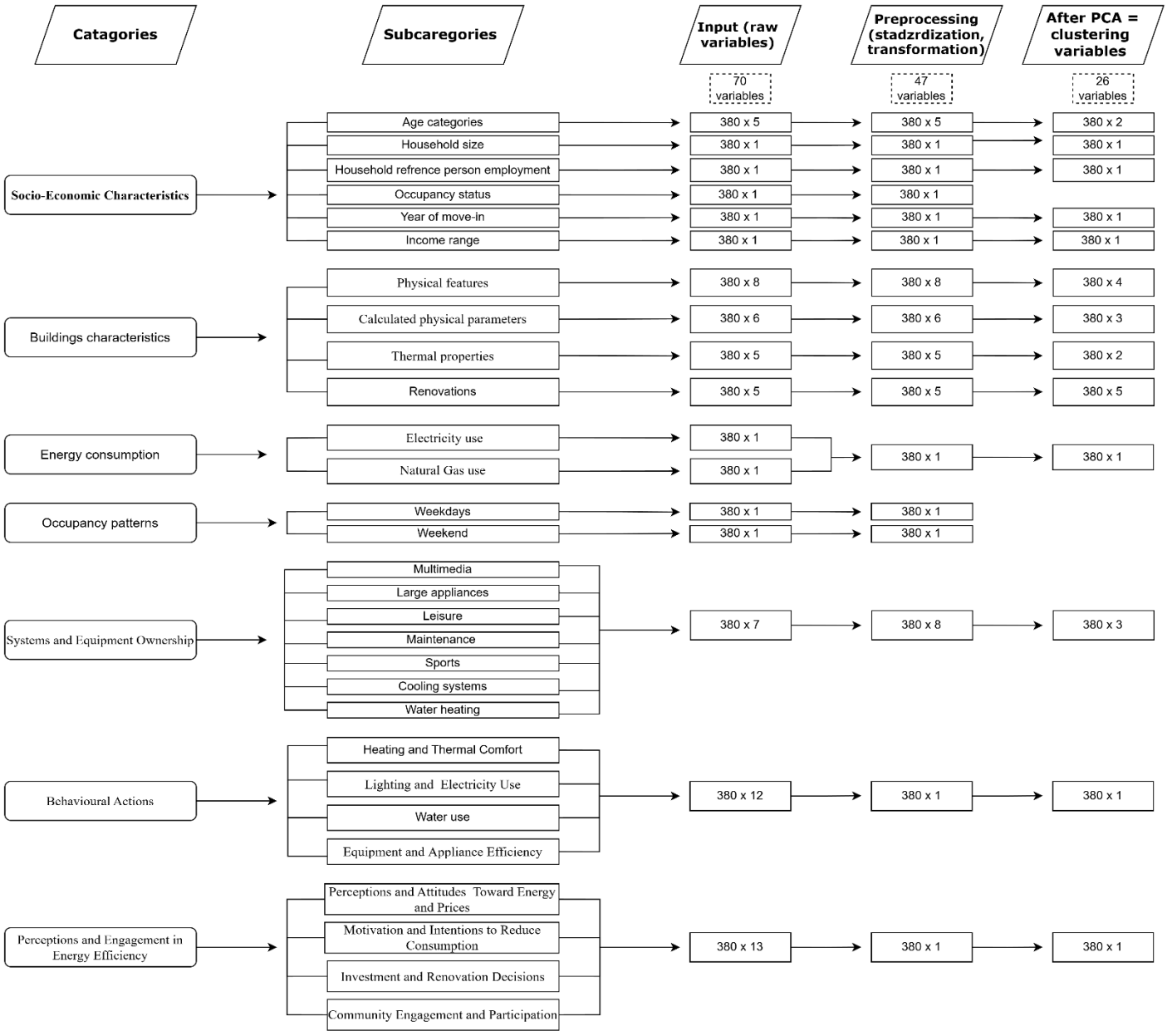


Fig. 8: Workflow of Variable Transformation and Dimensionality Reduction Across Data Processing Stages

1.1 Clustering

The dataset obtained at this initial stage subsequently forms the basis for the clustering analyses

1.1.1 Hierarchical clustering

Ward's method was applied to explore the dataset's internal structure. The average silhouette analysis (Fig. 9) showed maxima at k = 3 and k = 4, with a slightly higher score for k = 3. Therefore, a three-cluster solution was retained, providing the best balance between within-cluster cohesion and between-cluster separation for a robust and interpretable partitioning.

Hierarchical clustering with k = 3 produced the visualization shown in Fig. 10(a). As a hard clustering method, each observation is assigned to a single cluster, with no overlap or outliers. Fig. 10(b) displays the distribution of observations via Ward's method, illustrating cluster cohesion and the relative similarity among observations, facilitating a clear and interpretable cluster structure.

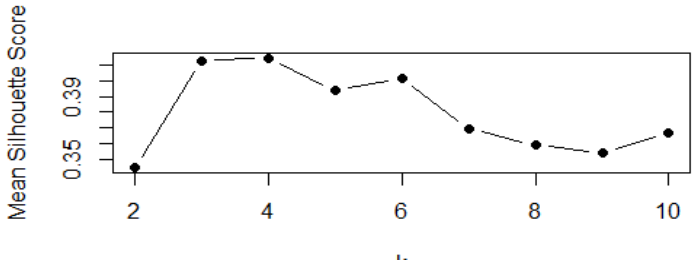


Fig. 9: Value of the optimal number of clusters for hierarchical clustering

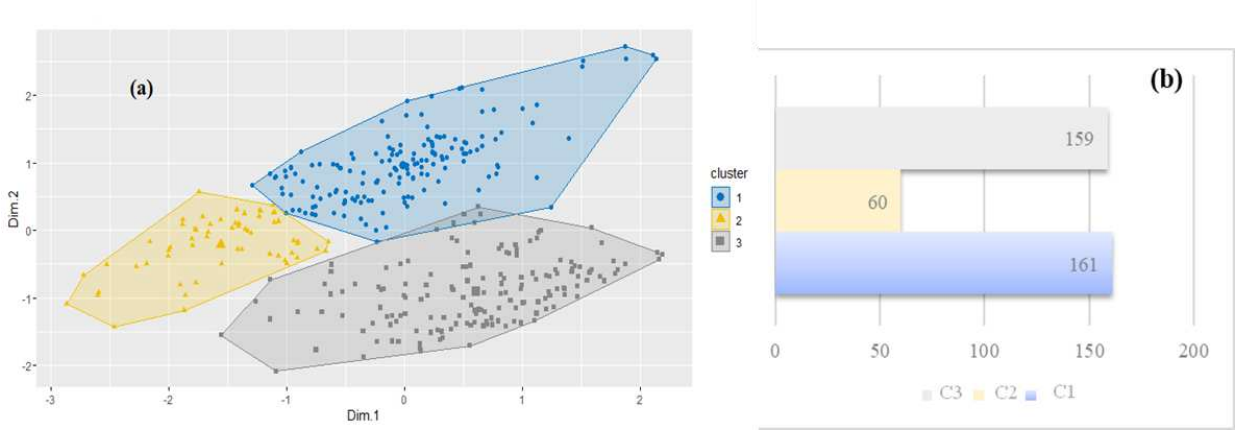


Fig. 10: Hierarchical clustering plot results; **(a)** clusters visualisation, **(b)** observations distribution in the clusters

1.1.2 K-means

The optimal number of clusters for K-means was determined using the Elbow (Fig. 11.a) and average silhouette methods Fig. 11.b), both indicating k = 3 with a mean silhouette near 0.5, supporting a clear and stable partition. Final clustering strictly assigned each observation to one cluster, with Fig. 12 illustrating both the overall distribution and the cluster-wise observation spread, confirming interpretability and robustness.

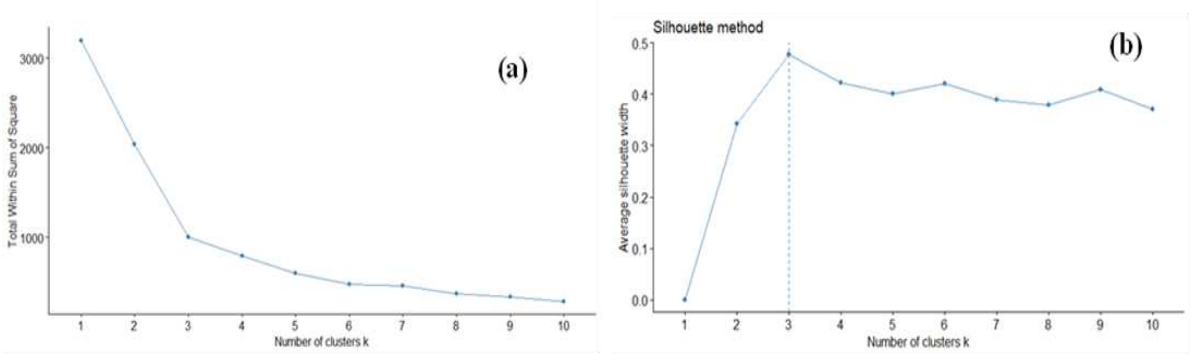


Fig. 11: Optimal number of clusters with **(a)** the Elbow method; **b)** average silhouette method

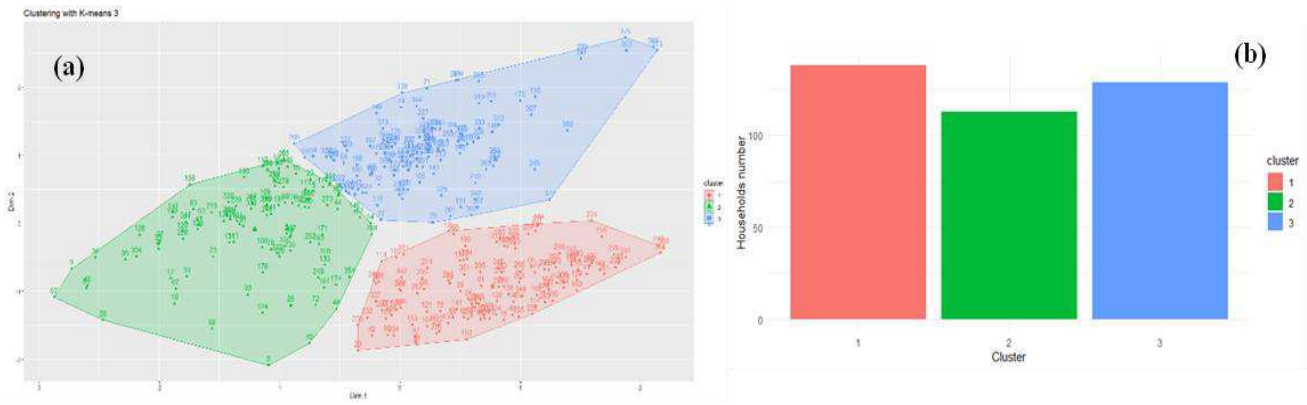


Fig. 12: K-means clustering results **(a)** clusters partition, **(b)** observations distribution in the clusters.

1.1.3 K-medoid

A **(**PAM algorithm) was applied as a robust alternative to K-means, defining clusters around actual observations (medoids) rather than centroids, which reduces sensitivity to outliers. Using the average silhouette criterion, the silhouette curve Fig. 13(a) clearly peaks at $k = 3$, indicating a compact and well-defined partition. The resulting three stable clusters, shown in Fig. 13(b), confirm the method's effectiveness in capturing the dataset's underlying structure without outliers.

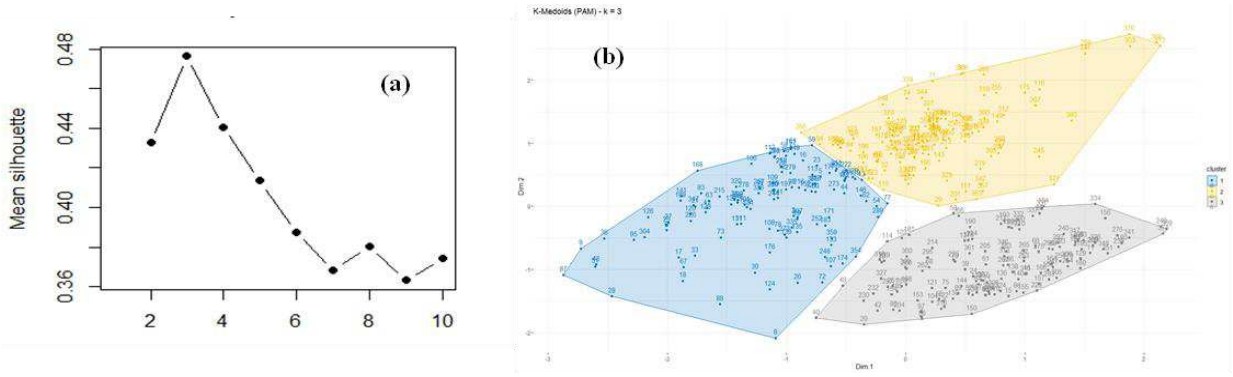


Fig. 13: (a) search for optimal k number with K-medoid clustering, **(b)** K-medoid clustering results.

1.1.4 The Gaussian Mixture Model (GMM)

The GMM was fitted using the Expectation-Maximization algorithm, with the Bayesian Information Criterion (BIC) selecting a VVE covariance structure (ellipsoidal clusters with equal orientation) and three components as optimal. The model yielded a log-likelihood of –904.57, BIC = –1898.25, and ICL = –1942.89, confirming partition stability. The 380 observations were assigned to three groups (n = 75, 220, and 85), consistent with previous clustering results, demonstrating the robustness of the three-cluster solution and GMM's ability to capture clusters of varying shapes and volumes. Fig. 14 illustrates the classification**.**

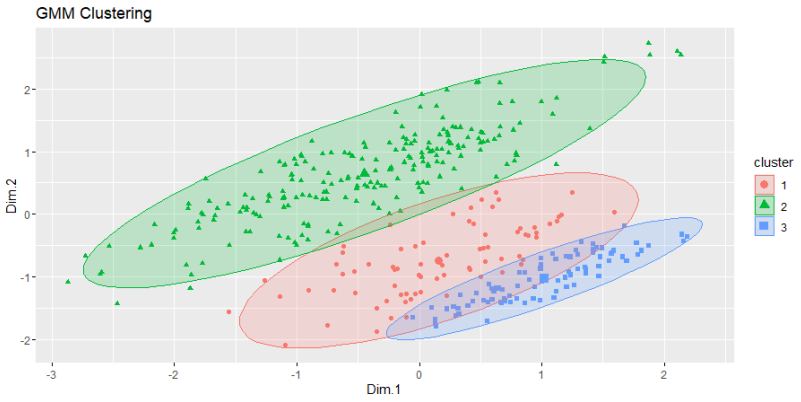


Fig. 14: GMM clustering results

1.2 Evaluation and Comparison of Clustering Methods

All algorithms identified k = 3 clusters, but partitions varied, prompting evaluation via multiple validation indices (Tables 4). K-means achieved the highest mean silhouette (0.51), reflecting

superior intra-cluster cohesion, while k-medoids excelled in Calinski-Harabasz (416) and balanced inter- versus intra-cluster dispersion. Davies-Bouldin values were lowest for K-means and k-medoids (0.71), indicating well-separated clusters, whereas GMM showed poor performance (>1.3). Ward's method yielded the highest Dunn index (0.25), suggesting slightly clearer boundaries but weaker cohesion. Overall, K-means provided the most robust and interpretable clustering, combining cohesion and separation effectively. (Fig. 15; Table 5).

Table 4: clustering indexes values

Clustering Algorithm	Mean silhouette (MS)	Calinski-Harabasz (CH)	Davies-Bouldin (DB)	Dunn Index
Hierarchical by Ward	**0.488**	**342.117**	0.728	**0.025**
K-means	**0.513**	411.447	**0.710**	0.019
K-medoid	0.477	**416.666**	0.715	0.018
GMM	0.281	168.310	1.354	0.008

Table 5: Silhouette Means and Cluster Sizes (K-Means)

	Cluster 01	Cluster 02	Cluster 03
Silhouette width	0.56	0.44	0.52
Observations	138	113	129

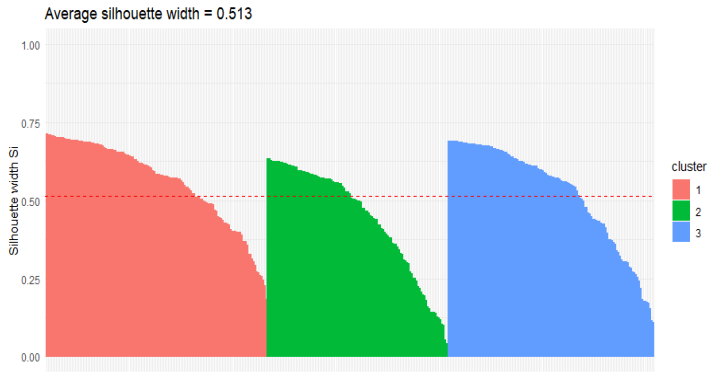


Fig. 15: average silhouette width of K-means clustering

1.3 Profiles identification

The clustering analysis identified three household profiles characterised by distinct combinations of energy consumption, dwelling characteristics, socio-economic attributes, and behavioural indicators. Comparison of cluster centroids (Table 6) reveals a clear gradient in energy demand, ranging from

high-consumption households (Cluster 03, n = 129) to low-consumption households (Cluster 02, n = 113), with Cluster 01 (n = 138) occupying an intermediate position.

Table 6: The mean values of analysed variables in each cluster

	Cluster 01	**Cluster 02**	**Cluster 03**
Energy consumption (kWh)	16870.5	14244.8	19055.7
Willingness and engagement to energy efficiency index	6.52	5.45	6.52
Energy sobriety index (ESI)	0.23	0.54	0.20
Building's period of construction	1999	1990	1999
Dwelling type	3.48	3.10	3.68
Dwelling area (m^2)	79.97	69.25	84.26
Dwelling perimeter (m)	42.27	36.13	44.12
Windows area (m^2)	12.26	9.02	13.04
Volume gross (m^3)	237.92	198.59	253.96
THLP (m)	34.40	27.38	38.05
THLA (m^2)	115.59	93.48	131.08
U windows ($W/m^2/°C$)	2.98	2.30	2.60
U door ($W/m^2/°C$)	2.01	2.68	4.09
Household income range	2.52	2.27	2.58
HRP job	0.38	0.65	0.47
HEI (household equipment index)	9.53	9.76	9.91
Space cooling (Air conditioner)	0.88	0.77	0.87
Retrofit interventions	0.88	0.49	0.84
Family size	4.67	4.45	5.22
Windows substitution	0.75	0.24	0.09
Weekend occupancy	0.82	0.91	0.82
Week Occupancy	0.80	0.81	0.77

Cluster 03: Energy-intensive households, less sustainable in behaviour but with a stronger willingness to engage in efficiency measures.

Represents the most energy-intensive households, with an average annual consumption of approximately 19,000 kWh (Fig. 16). These dwellings are the largest in the sample (≈84 m^2) and display the highest building volumes and envelope dimensions. Larger family sizes (5.22 persons), high equipment ownership (HEI = 9.91), and widespread air-conditioning use (0.87) are also observed. Although households exhibit the lowest energy sobriety index (0.20), their willingness to engage in energy efficiency measures is relatively high (6.52). Retrofit interventions are relatively

frequent (0.84), though window replacement remains limited (0.09), suggesting that improvements may focus more on equipment upgrades than on envelope retrofits.

Cluster 02: Energy-sober households, low consumption, and limited willingness to pay for efficiency improvements.

Corresponds to the lowest energy-consuming households, with an average annual consumption of about 14,245 kWh. These dwellings are smaller (≈69 m²) and exhibit lower building volumes and envelope dimensions. Households also show slightly smaller family sizes (4.45 persons) and lower equipment ownership. Despite recording the highest Energy Sobriety Index (0.54), engagement in energy efficiency measures is comparatively lower (5.45), and retrofit activities are less frequent. Combined with lower income levels, these characteristics suggest that reduced energy consumption may partly reflect structural and economic constraints rather than proactive efficiency behaviour.

Cluster 01: Intermediate households, with moderate consumption and a strong readiness to invest in energy efficiency.

Represents households with intermediate energy consumption (≈16,870 kWh annually). Dwellings are of medium size (≈80 m²), with construction periods concentrated around the late 1990s. Households exhibit moderate energy sobriety (0.23) but relatively strong willingness to engage in energy efficiency (6.52). Retrofit activity (0.88) and window replacement (0.75) are comparatively common, indicating a degree of proactive improvement of building components. Equipment ownership and air-conditioning use remain moderate but contribute to overall energy demand.

**A comparison of the three clusters reveals a socio-technical gradient in residential energy demand (Fig. 16) Cluster 03 exhibits higher consumption due to larger dwellings, bigger households, and higher incomes, while Cluster 02 comprises lower-income households in smaller homes with fewer appliances. Occupancy patterns differ, with Cluster 02 more present at weekends and Cluster 03 slightly higher on weekdays. These findings indicate that structural and socio-economic factors, dwelling size, household composition, and equipment ownership; largely drive energy use, as similar

levels of willingness to adopt efficiency measures do not translate into equivalent consumption.

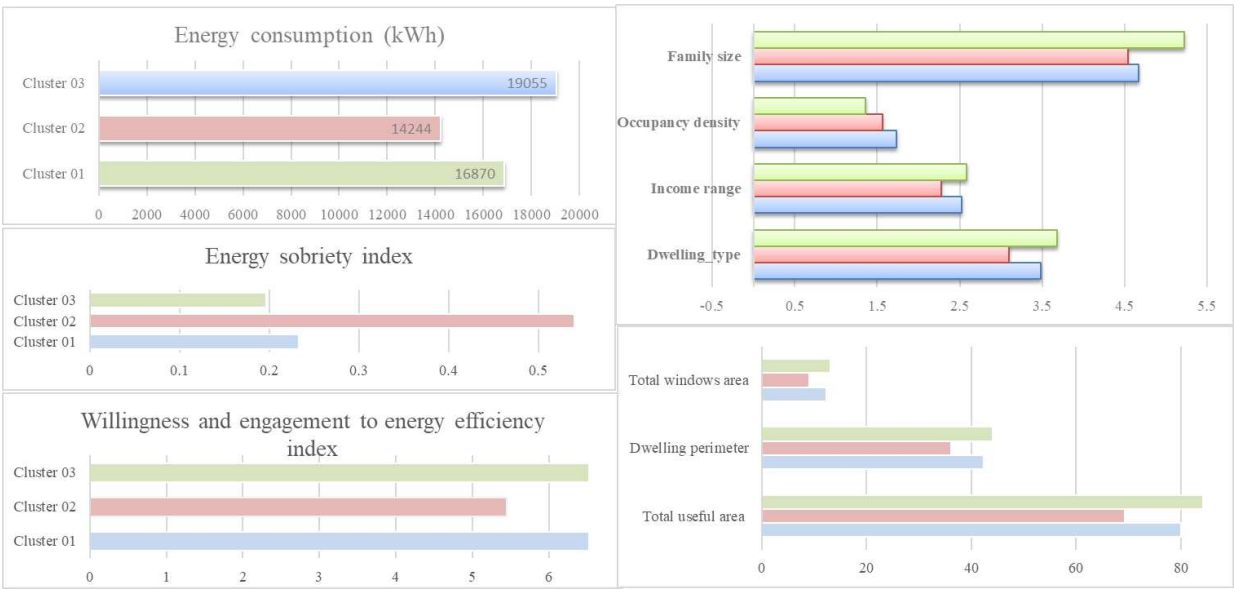


Fig. 16: Multi-dimensional characterisation of household clusters

2. Discussion

2.1 Optimal approach for complex, multi-dimensional residential energy data

The dataset combines behavioural, perceptual, and socio-technical variables with physical and thermal attributes, reflecting high dimensionality and heterogeneity. All clustering methods consistently converge towards a three-cluster solution (k=3), indicating a stable underlying structure. K-means outperforms alternative approaches, achieving higher silhouette scores and strong intra-cluster cohesion, resulting in well-separated and interpretable profiles. This confirms the effectiveness of partition-based methods for complex, mixed datasets following appropriate preprocessing.

These findings are consistent with prior studies demonstrating the robustness of K-means across high-resolution, high-dimensional, and socio-technical datasets [22,24,25,65] , supporting its suitability for residential energy profiling.

4.2 Socio-demographics and energy consumption

Socio-demographic characteristics structure residential energy use through household size, income, equipment ownership, and occupancy patterns. High-consumption households combine greater

affluence, larger size, and higher appliance density, whereas low-consumption groups are associated with smaller households and reduced equipment intensity. A transitional segment with moderate consumption and strong readiness for efficiency measures emerges as a key target for policy interventions. Overall, energy use is driven by household resources and structural living conditions, supporting differentiated, household-centred strategies [66]

4.3 Human factors driving residential energy intensity: behavioural and perceptual influences

Beyond structural factors, energy consumption is primarily driven by behavioural practices and usage intensity. High demand reflects lifestyle patterns rather than limited awareness, confirming the role of daily routines in shaping energy use [67,68]. Conversely, low consumption often indicates "involuntary efficiency" linked to economic constraints, consistent with evidence on income-driven consumption and behavioural inertia under rising costs [69,70]. Households with moderate consumption and high behavioural readiness constitute a strategic leverage point for demand-side interventions and energy transition policies [71,72]

4.4 Building attributes driving residential energy consumption pattern

Residential energy consumption is strongly influenced by building geometry, with larger surface areas, volumes, and window perimeters associated with higher demand, reflecting increased heating, hot water, and ventilation needs [73,74] . However, thermal indicators such as U-values provide only partial insight [47]. High consumption may persist despite moderate window performance when combined with poor door insulation and intensive occupancy, highlighting the interaction between envelope weaknesses and usage patterns.

These results emphasise the role of partial renovations and localised inefficiencies, showing that energy performance cannot be inferred from isolated thermal metrics. Overall, energy demand emerges from the combined effects of geometry, envelope performance, and occupant behaviour [14].

4.5 Leveraging household profiles to guide energy efficiency strategies in collective housing

Household typologies provide a foundation for designing targeted interventions in collective housing. Low energy consumption may reflect financial constraints rather than proactive environmental behaviour, highlighting the need for equitable strategies that ensure both adequate and affordable energy access. Effective approaches should combine targeted retrofits, behavioural measures, and opportunities for renewable energy adoption, recognising that some households aim to reduce consumption while others seek to produce energy through distributed renewables. Distinguishing voluntary from involuntary efficiency, and integrating comfort, affordability, and renewable energy engagement into evaluation frameworks, is essential to avoid penalising vulnerable households and to capture the full spectrum of household energy practices

3. Conclusion

This study presents a comprehensive multi-criteria framework for profiling residential energy consumption in collective housing, integrating socio-demographic, behavioural, perceptual, geometric, and thermal building characteristics. By applying and comparing four clustering algorithms: K-means, Ward hierarchical, Gaussian Mixture Models, and K-medoids; across a dataset of Algerian households, three consistent clusters were identified, with K-means demonstrating superior performance across multiple evaluation indices.

Energy consumption is not solely determined by building attributes but strongly driven by occupant behaviour, practices, and perceptions. High-consumption households exhibit less sustainable usage patterns, highlighting the need to combine behavioural interventions with physical upgrades. At the same time, dwelling geometry and envelope performance remain key structural drivers of demand.

These findings support the design of targeted efficiency strategies in Algerian collective housing by linking household typologies with socio-technical and building characteristics. The study provides a robust basis for policy development and context-specific interventions in underrepresented settings.

Overall, this work advances residential energy profiling by integrating multi-domain data and comparative clustering, demonstrating the value of combining behavioural and physical dimensions for more effective, context-adapted energy strategies.

Limitations of the study

The study is constrained by a modest sample, limited to collective housing and voluntary participation, which may introduce response bias. Building performance assessment relied solely on U-values, omitting factors such as thermal inertia, air infiltration, and solar gains. While the exploratory comparative clustering approach identifies household profiles, it does not distinguish between natural gas consumption (for space heating and cooking) and electricity use (for cooling and appliances), nor does it incorporate predictive modelling, building energy simulations, or the potential integration of renewable energy systems to strengthen the robustness and applicability of the findings. Integration of qualitative and quantitative variables relied on conventional methods; future work could employ advanced techniques, including deep learning, to capture complex non-linear relationships across multi-domain datasets.

Declaration of AI-assisted technologies in Scientific writing

During the preparation of this manuscript, the authors used Grammarly to improve language accuracy and readability. All content was carefully reviewed and edited by the authors. The analyses, interpretations, and conclusions are entirely the authors' own work. AI tools were used solely to enhance presentation, under full human oversight, and the authors take full responsibility for the content of the published article.

Declaration of competing interest

The authors declare that they have no financial or personal conflicts of interest that could inappropriately influence or bias their work.

Data availability

The datasets used and/or analysed during the current study are available from the corresponding author upon reasonable request.

References

[1] IEA, IEA – International Energy Agency - IEA, (2023). https://www.iea.org/data-and-statistics (accessed January 2, 2026).

[2] X. Xu, H. Yu, Q. Sun, V.W.Y. Tam, A critical review of occupant energy consumption behavior in buildings: How we got here, where we are, and where we are headed, Renewable and Sustainable Energy Reviews 182 (2023) 113396. https://doi.org/10.1016/J.RSER.2023.113396.

[3] S. Hu, X. Zhou, D. Yan, F. Guo, T. Hong, Y. Jiang, A systematic review of building energy sufficiency towards energy and climate targets, Renewable and Sustainable Energy Reviews 181 (2023) 113316. https://doi.org/10.1016/J.RSER.2023.113316.

[4] K. Ouazar, L.A. Mansour, E. Eveno, K. Rahmani, A. Zemmit, A.A. Arafat, B.A. Jemberie, M. Alsharef, A. Althobaiti, A GIS and statistical based approach to assessing citizen awareness of the energy transition and climate change across algerian climatic zones, Scientific Reports 2026 (2026). https://doi.org/10.1038/s41598-026-43948-0.

[5] H. Yoshino, T. Hong, N. Nord, IEA EBC annex 53: Total energy use in buildings—Analysis and evaluation methods, Energy Build. 152 (2017) 124–136. https://doi.org/10.1016/J.ENBUILD.2017.07.038.

[6] H. Fang, W.K. Li, P. Dai, Parametric analysis of passive ultra-low energy building envelope performance in existing residential buildings, Scientific Reports 2025 15:1 15 (2025) 23407-. https://doi.org/10.1038/s41598-025-07421-8.

[7] X. Chen, Z. Xu, Y. Qiu, H. Hu, X. Wang, China's newest design of apartment buildings with modernized façade: A comparative evaluation of its energy performance in five major climate zones, Sustain. Cities Soc. 117 (2024) 105954. https://doi.org/10.1016/J.SCS.2024.105954.

[8] H. Samuelson, S. Claussnitzer, A. Goyal, Y. Chen, A. Romo-Castillo, Parametric energy simulation in early design: High-rise residential buildings in urban contexts, Build. Environ. 101 (2016) 19–31. https://doi.org/10.1016/J.BUILDENV.2016.02.018.

[9] H. Li, J. Zheng, Y. Zhang, R. Zhang, Urban morphology and residential building net-zero energy performance: A global meta-analysis, Energy Build. 346 (2025) 116142. https://doi.org/10.1016/J.ENBUILD.2025.116142.

[10] R. Alimohamadi, M.H. Jahangir, Multi-Objective optimization of energy consumption pattern in order to provide thermal comfort and reduce costs in a residential building, Energy Convers. Manag. 305 (2024) 118214. https://doi.org/10.1016/J.ENCONMAN.2024.118214.

[11] R. Lee, M. Choi, J. Yoon, D. Kim, Impacts of lighting and plug load variations on residential building energy consumption targeting zero energy building goals, Journal of Building Engineering 75 (2023) 106962. https://doi.org/10.1016/J.JOBE.2023.106962.

[12] S. Habibi, Energy-efficient building design optimization and renewable energy integration for zero-energy residential buildings using simulations, Energy & Environment (2025). https://doi.org/10.1177/0958305X251343057.

[13] A. Remizov, S.A. Memon, J.R. Kim, Novel building energy performance-based climate zoning enhanced with spatial constraint, Appl. Energy 355 (2024) 122238. https://doi.org/10.1016/J.APENERGY.2023.122238.

[14] B.F. Balvedi, E. Ghisi, R. Lamberts, A review of occupant behaviour in residential buildings, Energy Build. 174 (2018) 495–505. https://doi.org/10.1016/J.ENBUILD.2018.06.049.

[15] B. Yan, W. Yang, F. He, W. Zeng, Occupant behavior impact in buildings and the artificial intelligence-based techniques and data-driven approach solutions, Renewable and Sustainable Energy Reviews 184 (2023) 113372. https://doi.org/10.1016/J.RSER.2023.113372.

[16] M.N. Uddin, H.-H. Wei, H.-L. Chi, M. Ni, N. Tamanna, Building Layout Influence on Occupant's Energy Consumption Behaviour: An Agent-Based Modeling Approach, in: The 9th International Conference on Sustainable Development, MDPI, Basel Switzerland, 2022: p. 22. https://doi.org/10.3390/environsciproc2022015022.

[17] Y. Yuan, L. Gao, K. Zeng, Y. Chen, Space-Level air conditioner electricity consumption and occupant behavior analysis on a university campus, Energy Build. 300 (2023) 113646. https://doi.org/10.1016/J.ENBUILD.2023.113646.

[18] A. Mylonas, A. Tsangrassoulis, J. Pascual, Modelling occupant behaviour in residential buildings: A systematic literature review, Build. Environ. 265 (2024) 111959. https://doi.org/10.1016/J.BUILDENV.2024.111959.

[19] J. Rouleau, L. Gosselin, P. Blanchet, Robustness of energy consumption and comfort in high-performance residential building with respect to occupant behavior, Energy 188 (2019) 115978. https://doi.org/10.1016/J.ENERGY.2019.115978.

[20] J. Duan, N. Li, J. Peng, Q. Liu, T. Peng, S. Wang, Clustering and prediction of space cooling and heating energy consumption in high-rise residential buildings with the influence of occupant behaviour: Evidence from a survey in Changsha, China, Journal of Building Engineering 76 (2023) 107418. https://doi.org/10.1016/J.JOBE.2023.107418.

[21] V. Michalakopoulos, E. Sarmas, I. Papias, P. Skaloumpakas, V. Marinakis, H. Doukas, A machine learning-based framework for clustering residential electricity load profiles to enhance demand response programs, Appl. Energy 361 (2024) 122943. https://doi.org/10.1016/j.apenergy.2024.122943.

[22] X. Cui, M. Lee, M.N. Uddin, X. Zhang, V.G. Zakka, Analyzing different household energy use patterns using clustering and machine learning, Renewable and Sustainable Energy Reviews 212 (2025) 115335. https://doi.org/10.1016/j.rser.2025.115335.

[23] W. Tang, H. Wang, X.L. Lee, H.T. Yang, Machine learning approach to uncovering residential energy consumption patterns based on socioeconomic and smart meter data, Energy 240 (2022) 122500. https://doi.org/10.1016/j.energy.2021.122500.

[24] L. Czétány, V. Vámos, M. Horváth, Z. Szalay, A. Mota-Babiloni, Z. Deme-Bélafi, T. Csoknyai, Development of electricity consumption profiles of residential buildings based on smart meter data clustering, Energy Build. 252 (2021) 111376. https://doi.org/10.1016/j.enbuild.2021.111376.

[25] E. Zaidan, A. Abulibdeh, A. Alban, R. Jabbar, Motivation, preference, socioeconomic, and building features: New paradigm of analyzing electricity consumption in residential buildings, Build. Environ. 219 (2022) 109177. https://doi.org/10.1016/j.buildenv.2022.109177.

[26] K. Li, Z. Ma, D. Robinson, J. Ma, Identification of typical building daily electricity usage profiles using Gaussian mixture model-based clustering and hierarchical clustering, Appl. Energy 231 (2018) 331–342. https://doi.org/10.1016/J.APENERGY.2018.09.050.

[27] S. Kallel, M. Amayri, N. Bouguila, Clustering and Interpretability of Residential Electricity Demand Profiles, Sensors 2025, Vol. 25, 25 (2025). https://doi.org/10.3390/s25072026.

[28] G. Petricli, T. Inkaya, G. Gokay Emel, Identifying green citizen typologies by mining household-level survey data, Renewable and Sustainable Energy Reviews 189 (2024) 113957. https://doi.org/10.1016/J.RSER.2023.113957.

[29] Y. Wei, X. Zhang, Y. Shi, L. Xia, S. Pan, J. Wu, M. Han, X. Zhao, A review of data-driven approaches for prediction and classification of building energy consumption, Renewable and Sustainable Energy Reviews 82 (2018) 1027–1047. https://doi.org/10.1016/j.rser.2017.09.108.

[30] Y. Wang, T. Cao, H. Wang, Q. Fu, J. Chen, Y. Lu, An interactive clustering method via metric learning for analyzing energy consumption patterns, Journal of Building Engineering 98 (2024) 111143. https://doi.org/10.1016/j.jobe.2024.111143.

[31] V.R. Chifu, T. Cioara, I. Danci, C.B. Pop, I. Anghel, Whale optimization-based clustering for buildings' energy consumption profiles, Energy Reports 14 (2025) 1207–1221. https://doi.org/10.1016/j.egyr.2025.07.034.

[32] B. Kim, Y. Yamaguchi, Y. Shimoda, Physics-based modeling of electricity load profile of commercial building stock considering building system composition and occupancy profile, Energy Build. 304 (2024) 113813. https://doi.org/10.1016/J.ENBUILD.2023.113813.

[33] M.H. Khajedehi, J. Vivian, A. Zarrella, Identification of Italian user archetypes for building energy simulations based on a national survey, Energy Build. 344 (2025) 115924. https://doi.org/10.1016/j.enbuild.2025.115924.

[34] R. Kaur, D. Gabrijelčič, Behavior segmentation of electricity consumption patterns: A cluster analytical approach, Knowl. Based. Syst. 251 (2022) 109236. https://doi.org/10.1016/j.knosys.2022.109236.

[35] N. Carbonare, T. Pflug, A. Wagner, Clustering the occupant behavior in residential buildings: a method comparison, Bauphysik 40 (2018) 427–433. https://doi.org/10.1002/bapi.201800023.

[36] R. Tolas, R. Portase, R. Potolea, From Individual Device Usage to Household Energy Consumption Profiling, Electronics 2024, Vol. 13, 13 (2024). https://doi.org/10.3390/electronics13122325.

[37] D. Krishnan, S. Kelly, Y. Kim, A Meta-Analysis Review of Occupant Behaviour Models for Assessing Demand-Side Energy Consumption, Energies 2022, Vol. 15, 15 (2022). https://doi.org/10.3390/en15031219.

[38] G. Ruellan, S. Attia, G. Haesbroeck, Clustering of archetypal building-inhabitant pairs to improve energy efficiency: The case of the Walloon region in Belgium, Energy Build. 335 (2025) 115549. https://doi.org/10.1016/j.enbuild.2025.115549.

[39] M. De Simone, L. Callea, G. Fajilla, Surveys and inferential statistics to analyze contextual and personal factors influencing domestic hot water systems and usage profiles in residential buildings of Southern Italy, Energy Build. 255 (2022) 111660. https://doi.org/10.1016/j.enbuild.2021.111660.

[40] T. Malatesta, J.K. Breadsell, Identifying Home System of Practices for Energy Use with K-Means Clustering Techniques, Sustainability 2022, Vol. 14, 14 (2022). https://doi.org/10.3390/su14159017.

[41] ONS (2008), National Statistics Office, (n.d.). https://www.ons.dz/ (accessed March 26, 2026).

[42] CNERIB (2016) DTR C3.2/4, Building Thermal Regulation, (n.d.). https://www.cnerib.edu.dz/ (accessed December 1, 2025).

[43] M. Traich, A. Rahmane, A Decomposition Analysis of Algeria's Residential Energy Consumption Change: (2000–2020), Strategic Planning for Energy and the Environment 43 (2024) 151–176. https://doi.org/10.13052/spee1048-5236.4317.

[44] APRUE, National Agency for the Promotion and Rationalisation of Energy Use, (2023). https://aprue.org.dz/fr/ (accessed March 28, 2026).

[45] M. Afaifia, K.A. Djiar, N. Bich-Ngoc, J. Teller, An energy consumption model for the Algerian residential building's stock, based on a triangular approach: Geographic Information System (GIS), regression analysis and hierarchical cluster analysis, Sustain. Cities Soc. 74 (2021) 103191. https://doi.org/10.1016/j.scs.2021.103191.

[46] S. Hu, D. Yan, S. Guo, Y. Cui, B. Dong, A survey on energy consumption and energy usage behavior of households and residential building in urban China, Energy Build. 148 (2017) 366–378. https://doi.org/10.1016/J.ENBUILD.2017.03.064.

[47] L. Li, Y. Wang, M. Wang, W. Hu, Y. Sun, Impacts of multiple factors on energy consumption of aging residential buildings based on a system dynamics model--Taking Northwest China as an example, Journal of Building Engineering 44 (2021) 102595. https://doi.org/10.1016/j.jobe.2021.102595.

[48] Declaration of Helsinki, WMA Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Participants – WMA – The World Medical Association, (1964). https://www.wma.net/policies-post/wma-declaration-of-helsinki/ (accessed March 26, 2026).

[49] SONELGAZ, Sonelgaz, (2024). https://www.sonelgaz.dz/fr (accessed December 1, 2025).

[50] K. Pearson, VII. Note on regression and inheritance in the case of two parents, Proceedings of the Royal Society of London 58 (1895) 240–242. https://doi.org/10.1098/RSPL.1895.0041.

[51] G. Mutani, S. Santantonio, A. Tartaglia, Statistical Data Analysis for Energy Communities, TECNICA ITALIANA-Italian Journal of Engineering Science 64 (2020) 385–397. https://doi.org/10.18280/ti-ijes.642-438.

[52] I.T. Jolliffe, J. Cadima, Principal component analysis: a review and recent developments, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 374 (2016) 20150202. https://doi.org/10.1098/rsta.2015.0202.

[53] H.F. Kaiser, An index of factorial simplicity, Psychometrika 39 (1974) 31–36. https://doi.org/10.1007/BF02291575/METRICS.

[54] M.S. Bartlett, The Effect of Standardization on a χ 2 Approximation in Factor Analysis, Biometrika 38 (1951) 337. https://doi.org/10.2307/2332580.

[55] J.H. Ward, Hierarchical Grouping to Optimize an Objective Function, J. Am. Stat. Assoc. 58 (1963) 236. https://doi.org/10.2307/2282967.

[56] F. Murtagh, P. Legendre, Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion?, J. Classif. 31 (2014) 274–295. https://doi.org/10.1007/s00357-014-9161-z.

[57] A. Likas, N. Vlassis, J. J. Verbeek, The global k-means clustering algorithm, Pattern Recognit. 36 (2003) 451–461. https://doi.org/10.1016/S0031-3203(02)00060-2.

[58] Leonard. Kaufman, P.J.. Rousseeuw, Finding Groups in Data : an Introduction to Cluster Analysis, (1990) 368. https://doi.org/10.1002/9780470316801.

[59] F. Bin, A. Abid, A Novel Approach for PAM Clustering Method, Int. J. Comput. Appl. 86 (2014) 1–5. https://doi.org/10.5120/15074-3039.

[60] P.J. Rousseeuw, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, J. Comput. Appl. Math. 20 (1987) 53–65. https://doi.org/10.1016/0377-0427(87)90125-7.

[61] G. McLachlan, D. Peel, Finite Mixture Models, Wiley, 2000. https://doi.org/10.1002/0471721182.

[62] T. Caliński, J. Harabasz, A dendrite method for cluster analysis, Communications in Statistics 3 (1974) 1–27. https://doi.org/10.1080/03610927408827101.

[63] D.L. Davies, D.W. Bouldin, A Cluster Separation Measure, IEEE Trans. Pattern Anal. Mach. Intell. PAMI-1 (1979) 224–227. https://doi.org/10.1109/TPAMI.1979.4766909.

[64] J.C. Dunn†, Well-Separated Clusters and Optimal Fuzzy Partitions, Journal of Cybernetics 4 (1974) 95–104. https://doi.org/10.1080/01969727408546059.

[65] M. Pullinger, E. Zapata-Webborn, J. Kilgour, S. Elam, J. Few, N. Goddard, C. Hanmer, E. McKenna, T. Oreszczyn, L. Webb, Capturing variation in daily energy demand profiles over time with cluster analysis in British homes (September 2019 – August 2022), Appl. Energy 360 (2024) 122683. https://doi.org/10.1016/j.apenergy.2024.122683.

[66] M. Shamsaiee, U. Eicker, Socio-Demographic insights on urban building energy consumption, Energy Build. 352 (2026) 116793. https://doi.org/10.1016/j.enbuild.2025.116793.

[67] Y. Zhang, X. Bai, F.P. Mills, Characterizing energy-related occupant behavior in residential buildings: Evidence from a survey in Beijing, China, Energy Build. 214 (2020) 109823. https://doi.org/10.1016/j.enbuild.2020.109823.

[68] S. Chen, G. Zhang, X. Xia, Y. Chen, S. Setunge, L. Shi, The impacts of occupant behavior on building energy consumption: A review, Sustainable Energy Technologies and Assessments 45 (2021) 101212. https://doi.org/10.1016/j.seta.2021.101212.

[69] J. Shim, D. Song, Unveiling energy inefficiencies: A study on building energy consumption in single-person households, Renewable and Sustainable Energy Reviews 214 (2025) 115546. https://doi.org/10.1016/j.rser.2025.115546.

[70] R.K. Moussa, D.K. Kanga, Elicitation of electricity consumption habit formation among new subscribers, Heliyon 10 (2024). https://doi.org/10.1016/j.heliyon.2024.e28360.

[71] M.M. Jaber, D. Qasim, A. Ghaith, What Drives Sustainable Energy Use? A Case Study from Jordanian Households, Environments 2025, Vol. 12, 12 (2025). https://doi.org/10.3390/environments12120452.

[72] K. Chadwick, R. Russell-Bennett, N. Biddle, The role of human influences on adoption and rejection of energy technology: A systematised critical review of the literature on household energy transitions, Energy Res. Soc. Sci. 89 (2022). https://doi.org/10.1016/j.erss.2022.102528.

[73] S.J. Lee, S.Y. Song, Determinants of residential end-use energy: Effects of buildings, sociodemographics, and household appliances, Energy Build. 257 (2022). https://doi.org/10.1016/j.enbuild.2021.111782.

[74] E.L. Ofetotse, E.A. Essah, R. Yao, Evaluating the determinants of household electricity consumption using cluster analysis, Journal of Building Engineering 43 (2021) 102487. https://doi.org/10.1016/j.jobe.2021.102487.