Data in Brief 48 (2023) 109232

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Household electricity consumption in Greece: A dataset based on socio-economic features

Stavros Mischos[a,*], Nikolaos Virtsionis Gkalinikis[a], Aikaterini Manolopoulou[b], Eleana Dalagdi[a], Dimitrios Zaikis[a], Aristotelis Lazaridis[a], Danai Vlachava[c], Kostantinos Lagouvardos[d], Dimitrios Vrakas[a]

[a] *Aristotle University of Thessaloniki, School of Informatics, Greece*
[b] *Information Technologies Institute, Centre of Research & Technology-Hellas, Greece*
[c] *International Hellenic University, Greece*
[d] *Institute for Environmental Research and Sustainable Development, National Observatory of Athens, Greece*

ARTICLE INFO

Article history:
Received 10 February 2023
Revised 19 April 2023
Accepted 8 May 2023
Available online 12 May 2023

Keywords:
Energy consumption
Electricity
Occupants' behavior
Weather data
Dataset
Questionnaire
Greece
Household Electricity Consumption

ABSTRACT

The electricity consumption of a residence depends on many factors such as the habits and economical status of the occupants, the properties of the household and many more. To shed more light on the subject a data set for households was created. The data were collected in Greece through an anonymous survey that comprises 26 questions, resulting in 188 data points from 104 households from different time periods. Each data point contains attributes that are divided into four categories. In the first category, the information is about the household data such as the type and properties of the residence. Next, occupants' socio-economic features are gathered. In this category information for the number and type of the occupants, the employment status and the total income of the residents is included. The third category of attributes is about the energy-related occupants' behavior. Finally, the location of the household was provided from the users to estimate the weather conditions for the provided time. Data augmentation was performed to discover non-trivial relationships between the data points. Thus, a secondary set of features was computed based on the raw attributes and is also

* Corresponding author.
E-mail address: mischossg@csd.auth.gr (S. Mischos).
Social media: @stmischos (S. Mischos)

https://doi.org/10.1016/j.dib.2023.109232

included. The provided data set can be used to extract insights that could be valuable during the imminent energy crisis.

Specifications Table

Subject	Electricity Consumption of residential installations
Specific subject area	The electricity consumption in a household is affected by multiple factors such as the occupants' habits, their level of energy awareness, the weather conditions, etc. This information is usually known to the Distribution System Operators but is not publicly available. The collected data set contains several features regarding the characteristics of the household and its residents alongside with the weather conditions and the corresponding electricity consumption of a given period. Note that the data set contains the overall household electricity consumption that was obtained from the electricity bills. These data can be used to create user profiles with similar characteristics and consumptions to determine the most influential factors related to the energy overconsumption. Hence, unique insights and recommendations can be produced for each group of users.
Type of data	Table
How the data were acquired	The data were collected in Greece through an anonymous survey comprising 26 questions. The survey was created in Google Forms and was published online.
Data format	Raw, Analyzed
Description of data collection	The initial data was collected through an anonymous online survey that was distributed among colleagues at several universities and other public fora in Greece, while weather data provided by the National Observatory of Athens. The survey contained 26 questions.
Data source location	Country: Greece
Data accessibility	Repository name: Electricity Consumption Data Direct URL to data: or https://figshare.com/articles/dataset/Data_Sheet_Energy_xlsx/21867933 DOI: 10.6084/m9.figshare.21867933 The survey can be found in the following link: https://forms.gle/nx118et8hC8BtpUB8

Value of the Data

- These data constitute a unique, previously unseen source of information related to electricity consumption in Greece.
- This dataset can be used from machine learning researchers, electricity domain specialists, statistics organizations (regional, national, or international) and Industry companies.
- The collected data could provide some insights on energy awareness of the public based on various social and economic factors.
- The dataset can be used to extract information regarding the way that the electricity consumption of a residence is affected by its characteristics. Thus, targeted solutions could be proposed in order to avoid energy waste.

1. Objective

The generation of this dataset occurred because of the need to look for the factors that increase energy overconsumption of domestic consumers. Previous studies [4–6] have shown that energy consumption can be determined based on a number of factors including socio economic features of the residents. Furthermore, it is important to estimate the energy awareness level of people in Greece. Using these data, machine learning engineers can implement more targeted recommendations, or home energy management, systems, to tackle the environmental problem that now exists. Finally, these data can be used by energy companies for solar panel installation suggestions to customers in an optimal way, based on their socio-economic profile.

2. Data Description

This dataset was created based on the answers of 26 questions. The goal was to collect various information about households in Greece regarding their characteristics and the habits of their residents. All questions that were studied in this survey and their extracted features, have been classified into four categories:

(1) Household details
(2) Occupants' socio-economic characteristics
(3) Energy-related occupants' behaviors
(4) Electricity consumption

Moreover, some extra features about the weather conditions prevailing in each household's region have been extracted, based on provided information. For each category of questions, two tables are provided; one that summarizes the extracted variables and another one augmented with variables obtained after processing the data.

2.1. Household Details

First, the participants had to answer a couple of questions to provide some information about their household. This information relates mainly to the characteristics of the house, i.e., its type, its size, the number of bedrooms it has, its age, its location and the type of heating appliances it contains. It should be noted that the existence of electric heating affects greatly the electricity consumption of a household. These features are outlined in Table 1. To replace categorical variables with ordinal ones and facilitate processing, the Age, the Household size and the Dwelling type variables were grouped, and a sample of the dataset is presented in Table 2.

2.2. Occupants' Socio-Economic Characteristics

Another category of questions that were provided to the participants concerned the social and economic features of their residence. The questions were about the number of people who live in the house and their educational and financial status. The extracted variables are shown in Table 3. Another sample of the dataset is presented in Table 4, augmented by the following indices: A_{inc}, A_{dec}, A_{gauge}, Education Index (EI) which are designed to quantify the ages of the people living in a house.

(1) $A_{inc} = 0.5 * \text{Children} + 0.75 * \text{Teenagers} + 0.9 * \text{Adults} + 1.0 * \text{Elders}$, assuming people consume more energy as they grow older.
(2) $A_{dec} = 1.0 * \text{Children} + 0.9 * \text{Teenagers} + 0.75 * \text{Adults} + 0.5 * \text{Elders}$, assuming people consume less energy as they grow older.

Table 1
Household properties.

Variable	Type	Measurement Unit	Statistics	Grouping
Dwelling type	Categorical	1. Single family home 2. Semi-detached home 3. Townhome 4. Apartment	16 (9%) 34 (18%) 6 (3%) 132 (70%)	Single-family → 1 Semi-detached → 0.7 Townhome → 0.4 Apartment → 0
Household m^2	Numerical		Avg: 109, Std: 63.4	$0m^2$ - $60m^2$ → 0 $61m^2$ - $80m^2$ → 0.2 $81m^2$ - $100m^2$ → 0.4 $101m^2$ - $140m^2$ → 0.6 $141m^2$ - $200m^2$ → 0.8 >$200m^2$ → 1
Bedrooms	Numerical		Avg: 2.48, Std: 1	
Age (Grouped by Age grade)	Ordinal	1. 0 - 5 2. 6 - 15 3. 16 - 30 4. 30+-	30 (16%) 72 (38%) 50 (27%) 36 (19%)	0 - 5 → 0 6 - 15 → 0.33 16 - 30 → 0.66 30+ → 1
Electric Heating	Categorical	1. Yes 2. No	1. Yes: 55 (30%) 2. No: 133 (70%)	
Area code	Numerical		N/A	

Table 2
Augmented household data.

Dwelling	Dwelling Grade	Household m^2	Size Grade	Bedrooms	Age	Age Grade	Electric Heating	Area Code
Single family	1	117	0.6	3	16-30	0.33	No	57003
Semi-det.	0.7	100	0.4	2	16-30	0.33	Yes	54634
Townhome	0.4	25	0	1	30+	1	Yes	54640
Apartment	0	142	0.8	3	30+	1	No	54248
Apartment	0	75	0.2	2	16-30	0.66	Yes	54636
Apartment	0	35	0	2	0-5	0	Yes	54636

Table 3
Occupants' socio-economic features.

Variable	Type	Unit	Statistics	Details
Occupants	Numerical		Avg: 2.87, Std: 1.2	Total number of occupants
Children (0-10 years old)	Numerical		Avg: 0.25, Std: 0.59	Number of children
Teenagers (10-18 years old)	Numerical		Avg: 0.29, Std: 0.57	Number of teenagers
Adults (19-69 years old)	Numerical		Avg: 2.2, Std: 0.94	Number of adults
Elder (70+)	Numerical		Avg: 0.1, Std: 0.35	Number of elders
Full timers	Numerical		Avg: 1.33, Std: 0.8	Number of full timers
Part timers	Numerical		Avg: 0.22, Std: 0.46	Number of part timers
Grads	Numerical		Avg: 1.33, Std: 0.91	Number of people that graduated from university
Post grads	Numerical		Avg: 0.55, Std: 0.72	Number of people that received a post-graduate degree
Income	Ordinal	1. 0€ - 10.000€ 2. 10.001€ - 20.000€ 3. 20.001€ - 40.000€ 4. 40.000€ - 60.000€ 5. >60.000€	27 (14%) 63 (34%) 72 (38%) 17 (9%) 9 (5%)	Total annual income of the family

Table 4
Augmented Socio-economic data.

Occupants	Children	Teenagers	Adults	Elders	A_{inc}	A_{dec}	A_{gauge}	Full timers	Part timers	Grads	Post Grads	E.I.
3	1	0	2	0	2,3	2,5	2,5	2	0	2	1	0,64
2	0	0	2	0	1,8	1,5	2	2	0	2	2	1,00
4	2	0	2	0	2,8	3,5	3	2	0	2	2	1,00
4	0	2	2	0	3,3	3,3	3,5	2	0	0	0	0,83
5	3	2	2	0	4,8	6,3	5	1	1	0	0	0,83
4	2	0	2	0	2,8	3,5	3	1	0	2	2	1,00

(3) A_{gauge}= 0.5 * Children + 0.75 * Teenagers + 1.0 * Adults + 0.5 * Elders, assuming adults and teenagers consume more energy than children and elders.
(4) Education Index (EI): an index to measure the educational attainment of a group of people. The formula is based on two factors: Mean Years of Schooling (MYS) for adults and elders and Expected Years of Schooling (EYS) for children and teenagers. The EYS in Greece in 2019 was 17.91 [3] and this value was used in our calculations. The MYS was calculated under the assumptions that the average duration of university studies is 4 years for the first degree and 2 years for a master's degree or a PhD. The formula for calculating Educational Index was based on [2] and is the following:

$$\mathrm{EI} = \frac{\frac{\mathrm{EYS}}{18} + \frac{\mathrm{MYS}}{18}}{2}$$

2.3. Energy-Related Occupants' Behaviors

This category contained questions that were about participants' energy-related habits. The goal is to quantify how aware the occupants of each household were about their energy footprint. All variables of this category are ordinal, except from the water heater existence index which is a binary categorical one. Table 5 describes the variables of this category.

2.4. Weather Data

These data were exported from other sources after the participants defined their area code. The area code was used to find the closest weather station from the National Observatory of Athens stations network. Through this process, the 113 houses from the questionnaire were categorized into 24 clusters. Each cluster was directly linked into a specific weather station from which were obtained hourly recordings of temperature and humidity for a three-year period to cover the consumption periods of all the provided electricity bills. A sample from the central station's data, located in the city of Thessaloniki, is presented in Table 6.

Regarding the electricity consumption of the buildings various additional variables were calculated for each weather station and hourly record. The computed variables are summarized in Table 7. The first two additional variables are Heating Degree Hours (HDH) and Cooling Degree Hours (CDH) give a rough indication of Heating and Cooling loads of a residence [1]. HDH and CDH are calculated by subtracting the actual hourly temperature from base heating and cooling temperatures as shown in the following two definitions:

$$\mathrm{HDH} = \begin{cases} 0, & T > T_{\mathrm{HB}} \\ \mathrm{T_{HB}} - T, & T \leq \mathrm{T_{HB}} \end{cases}$$

$$\mathrm{CDH} = \begin{cases} 0, & T < T_{\mathrm{CB}} \\ T - \mathrm{T_{CB}}, & T \geq \mathrm{T_{CB}} \end{cases}$$

Table 5
Energy-related occupants' behavior.

Variable	Type	Measurement Unit	Statistics	Details
Recycling	Ordinal	1. Rarely 2. Sometimes 3. Often / Always	25 (13%) 45 (24%) 118 (63%)	Values that represent how often the occupants recycle their waste·
Energy class	Ordinal	1. Rarely 2. Sometimes 3. Often / Always	15 (8%) 73 (39%) 100 (53%)	Values that represent how often the occupants check the energy class of the products they buy.
Thermostat	Ordinal	1. Rarely 2. Sometimes 3. Often / Always	25 (13%) 15 (8%) 148 (79%)	Values that indicate how often the occupants reduce the desired temperature in the thermostat.
Smart plugs	Ordinal	1. Rarely 2. Sometimes 3. Often / Always	170 (90%) 12 (7%) 6 (3%)	Values that represent how often occupants use smart plugs to monitor their consumption.
Awareness	Ordinal	1. Rarely 2. Sometimes 3. Often / Always	102 (54%) 77 (41%) 9 (5%)	Values that represent how often the occupants compare their energy consumption with friends and family.
Water heater	Categorical	1. Yes 2. No	65 (35%) 123 (65%)	A binary variable that indicates if the house has a solar water heater or not.

Table 6
Initial weather data retrieved from the central station of Thessaloniki, Greece.

Date	Time	Temperature (°C)	Humidity
…			
11/05/2019	7:00	13.9	0.86
11/05/2019	8:00	14.9	0.84
11/05/2019	9:00	15.6	0.74
11/05/2019	10:00	19.4	0.62
11/05/2019	11:00	20.9	0.55
11/05/2019	12:00	21.9	0.45
11/05/2019	13:00	22.4	0.5

Table 7
Augmented weather data from the central station of Thessaloniki, Greece.

Date	Time	Temperature (°C)	Humidity	HDH	CDH
11/05/2019	7:00	13.9	0.86	0.07	0.00
11/05/2019	8:00	14.9	0.84	0.03	0.00
11/05/2019	9:00	15.6	0.74	0.00	0.00
11/05/2019	10:00	19.4	0.62	0.00	0.00
11/05/2019	11:00	20.9	0.55	0.00	0.00
11/05/2019	12:00	21.9	0.45	0.00	0.00
11/05/2019	13:00	22.4	0.5	0.00	0.02
11/05/2019	14:00	23.4	0.45	0.00	0.06

The base temperatures for heating, $T_{HB,}$ and cooling, T_{CB}, may vary depending on the location, type, year of construction and energy class of the building. For simplicity reasons, T_{HB} and T_{CB} were set to 15.5 and 22 °C respectively. As an example, Table 7 contains the weather data of Table 6 augmented with the aforementioned variables.

Table 8
Electricity consumption variables.

Variable	Type	Statistics	Details
Start Date	Date		The first day of the billing period
End Date	Date		The last day of the billing period
Days	Numerical	Avg: 106.4, Std: 32.7	The length of the billing period in days
Kwhs	Numerical	Avg: 1259.2, Std: 1030.8	Overall household consumption in the billing period

Table 9
Augmented electricity consumption data.

Start	End	Days	Kwhs	Kwhs/day	Kwhs/day/m^2	HDD	CDD
27/09/2021	26/01/2022	121	2267	18,73553719	0,073472695	570,387	3,408
30/07/2021	26/09/2021	58	887	15,29310345	0,059972955	1,3	285,55
26/08/2021	28/12/2021	124	651	5,25	0,027631579	557,829	84,787
25/04/2021	26/08/2021	123	583	4,739837398	0,024946513	64,7	439,212
21/10/2021	20/02/2022	122	1464	12	0,137931035	722,244	0
22/06/2021	20/10/2021	120	1584	13,2	0,151724138	18,633	547,275
30/09/2021	17/10/2021	17	52	3,058823529	0,047058824	15,696	0,3

2.5. Electricity Consumption

The participants of this survey were also asked to provide some details about their total electricity consumption in Kwhs from one or more past electricity bills along with the corresponding calendar period. Based on the total consumption and the period, additional information was extracted for each household. The features of this category are summarized in Table 8. After some processing, information about the consumption in Kwhs per day, the consumption in Kwhs per day per m^2 and also the Heating Degree Days (HDD) and the Cooling Degree Days (CDD). A sample of these data is presented in Table 9.

3. Experimental Design, Materials and Methods

To reach the data acquisition phase, the form of an online survey was decided to be used, due to its simplicity of implementation, straightforward completion by the participants, as well as the relatively easy collection of results. Even though the main attribute aimed to be obtained from the questionnaire was the electricity consumption, it was necessary to extract also related information that would provide more insights on the consumption profile of the participants. Therefore, the questionnaire was designed to extract various properties such as the geographical location, household attributes, income information, educational level and energy-related behaviors of the participants. Finally, some features were transformed into a different type of data than their initial one, to facilitate future usage in scientific implementations such as machine learning models.

4. Power Analysis

According to its definition, power analysis is the calculation to estimate the smallest sample size that is needed for an experiment. However, in our case no experiment was performed. This dataset was created based on answers from a questionnaire about participants' electricity consumption data and their personal habits that may affect this consumption. Therefore, a power analysis of the data is not useful to evaluate the quality of them.

Ethics Statements

There is no funding for the present effort. There is no conflict of interest. The data are available in public domain. No human/animal subject was involved in experiments. Data not retrieved from social media platforms.

For our survey, no ethical approval, from an appropriate committee, was necessary for us to conduct the study. Each participant provided information about the electricity consumption of their household, some details about the construction of their house, about the occupants of them and finally about their daily habits that might affect electricity consumption. Therefore, our dataset does not store identifiable data that require ethical review and it meets current Data Protection regulations.

Consent

The distributed questionnaire included questions about the electricity consumption of the participants and also about their daily habits that might affect this consumption. Each participant gave consent in the questionnaire's URL to begin the completion process of the form.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT Author Statement

Stavros Mischos: Data curation, Visualization, Investigation, Conceptualization, Validation, Writing – original draft; **Nikolaos Virtsionis Gkalinikis:** Data curation, Methodology, Conceptualization, Writing – original draft; **Aikaterini Manolopoulou:** Conceptualization, Methodology, Investigation; **Eleana Dalagdi:** Conceptualization, Methodology, Investigation; **Dimitrios Zaikis:** Visualization, Investigation, Writing – original draft; **Aristotelis Lazaridis:** Visualization, Investigation, Writing – original draft; **Danai Vlachava:** Conceptualization, Methodology, Investigation; **Kostantinos Lagouvardos:** Methodology, Resources, Validation, Writing – review & editing; **Dimitrios Vrakas:** Supervision, Project administration, Data curation, Resources, Writing – review & editing.

Acknowledgments

This research has been supported by the European Regional Development Fund of the European Union, and by Greek national funds through the operational program, Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH–CREATE–INNOVATE (project code: TDEK 00343 - Energy Controlling Voice Enabled Intelligent Smart Home Ecosystem).

References

[1] A.E. McGarity, A.I. Gorski, Passive generation and storage of winter ice for summer cooling applications, in: Energy Developments: New Forms, Renewables, Conservation, Elsevier, 1984, pp. 245–250. pages.

[2] A.A. Priemyshev, M.V. Arysheva, Z.I Lavrova, Analysis of the human development index calculation. comparison of the old and the new methods, НАУЧНО-ТЕХНИЧЕСКИЕ (1994) 85.

[3] UNDP (United Nations Development Programme), Human development report, UNDP (United Nations Development Programme), 2019.

[4] D. Kotsila, P. Polychronidou, Determinants of household electricity consumption in Greece: a statistical analysis, J. Innov. Entrep. 10 (2021) 19, doi:10.1186/s13731-021-00161-9.

[5] A. Karananos, A. Dimara, K. Arvanitis, C. Timplalexis, S. Krinidis, D. Tzovaras, (2019). Energy consumption patterns of residential users: a study in Greece. In: Tzovaras D., Giakoumis D., Vincze M., Argyros A. (eds) Computer Vision Systems. ICVS 2019. Lecture Notes in Computer Science, vol 11754. Springer, Cham. doi:10.1007/978-3-030-34995-0_58.

[6] I. Kostakis, Socio-demographic determinants of household electricity consumption: evidence from Greece using quantile regression analysis, Curr. Res. Environ. Sustain. 1 (2020) 23–30 VolumePagesISSN 2666-0490, doi:10.1016/j.crsust.2020.04.001.