Ref. Ares(2024)5507522 - 30/07/2024

ENCLUDE
Energy Citizens for Inclusive Decarbonization

D4.2 – Report on the identification of citizens' clusters for decarbonization

WP4 – Identification of citizens' clusters for decarbonization

Sobhan Naderian, Gioia Falcone, Anastasia Ioannou

University of Glasgow

Version: 1.0

www.encludeproject.eu

Sobhan Naderian	University of Glasgow
sobhan.naderian@glasgow.ac.uk	https://www.gla.ac.uk/schools/engineering/staff/sobhannaderian/

Gioia Falcone	University of Glasgow
Gioia.Falcone@glasgow.ac.uk	https://www.gla.ac.uk/schools/engineering/staff/gioiafalcone/

Anastasia Ioannou	University of Glasgow
Anastasia.Ioannou@glasgow.ac.uk	(now at the Technical University of Denmark)

Disclaimer

Copyright Message

The ENCLUDE project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 101022791

ENCLUDE PROJECT & DELIVERABLE PROFILE

Project Acronym and Full Name:	ENCLUDE - Energy Citizens for Inclusive Decarbonization
Grant Agreement No.:	101022791
Programme:	H2020-EU.3.3.6. - Robust Decision Making and Public Engagement
Topic:	LC-SC3-CC-1-2018-2019-2020 – Social Sciences and Humanities (SSH) aspects of the Clean-Energy Transition
Funding Type:	RIA - Research and Innovation Action
Deliverable:	D4.2 – Report on the identification of citizens' clusters for decarbonization
Work Package:	WP4 – Identification of citizens' clusters for decarbonization
Deliverable Due Date:	Project month M38 (31/07/2024)
Actual Date of Submission:	31/07/2024
Dissemination Level:	Confidential
Lead Beneficiary:	UOG
Responsible Author:	Sobhan Naderian (UOG)
Contributor(s):	Sobhan Naderian (UOG), Gioia Falcone (UOG), Anastasia Ioannou (UOG, now at the Technical University of Denmark)
Internal Reviewers:	Vassilis Stavrakas (UPRC), Andreas Tuerk (JR)

Preface

The overall vision of ENCLUDE is to help the EU to fulfil its promise of a just and inclusive decarbonization pathway through sharing and co-creating new knowledge and practices that maximize the number and diversity of citizens who are willing and able to contribute to the energy transition. Motivated by achieving an equitable and sustainable future and the fulfilment of individual potential, ENCLUDE will contribute to the upcoming transformation of energy use by: (1) Assembling, aligning, and adapting disparate energy citizenship concepts for diverse communities of citizens and for different scales of policy making, lowering the barrier for action. (2) Operationalizing the energy citizenship concept at all scales of policy making for decarbonization. (3) Catalyzing a chain reaction of decarbonization actions across the EU.

1. Changes with respect to the DoA

This deliverable was originally due at M32. However, following the no-cost extension to the project, the new deadline was set to M38.

2. Dissemination and uptake

This deliverable constitutes the basis for further research within the project, enabling the statistical approaches and modelling instruments of WP4 and WP5. It will be disseminated via the project's webpage and added to the Enlighten repository of the University of Glasgow. In addition, its release will be advertised through the media channels of the consortium and of the University of Glasgow. It is expected that this deliverable will contribute to further research into individual and collective citizens' actions in the context of mitigation of and also adaptation to climate change.

3. Short Summary of results (<250 words)

This deliverable presents key results on clustering citizens based on their energy consumption and carbon footprint. In the first WP4 deliverable (D4.1), earlier studies were explored based on the patterns of sociodemographic variables, psychological characteristics, and energy consumption data, and how these could be used to cluster groups of citizens. In this report, three countries (UK, Greece, and The Netherlands) have been selected from the ECHOES ("Energy CHOices supporting the Energy Union and the SET-Plan") project to implement proposed clustering method.

A machine learning model is applied to the ECHOES data to identify distinct clusters. The model workflow starts with an estimation of people's energy consumption by applying a Life Cycle Assessment (LCA) method for the housing and mobility sectors, and the corresponding CO_2 footprint is then calculated for those same sectors. Finally, the k-means clustering method is applied to the energy consumption and CO_2 footprint data on both housing and mobility sectors to identify distinct citizens' groups in the UK, Greece, and the Netherlands.

Clustering results shows that the majority of citizens have low to medium energy consumption and CO_2 footprint on both housing and mobility for the selected countries. However, there are differences in the identified clusters across the three countries on medium to high energy consumption and CO_2 footprint. The findings could be very informative for policymakers, suggesting, for example, that they might consider prioritising the decarbonisation of the mobility sector in the UK and Greece, and that of the housing sector in the Netherlands.

4. Evidence of accomplishment

This report serves as evidence of accomplishment.

LIST OF PARTICIPANTS

	Participant Name	Short Name	Country	Logo
1	TECHNISCHE UNIVERSITEIT DELFT	TUD	Netherlands	TU Delft
2	UNIVERSITY OF PIRAEUS RESEARCH CENTER	UPRC	Greece	TEESlab
3	EIDGENOESSISCHE TECHNISCHE HOCHSCHULE ZUERICH	ETHZ	Switzerland	ETH zürich
4	UNIVERSITY COLLEGE CORK - NATIONAL UNIVERSITY OF IRELAND, CORK.	UCC	Ireland	UCC
5	UNIVERSITY OF GLASGOW	UOG	United Kingdom	University of Glasgow
6	JOANNEUM RESEARCH FORSCHUNGSGESELLSCHAFT MBH	JR	Austria	JOANNEUM RESEARCH LIFE
7	TH!NK E	THNK	Belgium	Th!nk E
8	UNIVERSITEIT UTRECHT	UU	Netherlands	Utrecht University
9	GREEN PARTNERS SRL	GP	Romania	Green Partners
10	ZDRUZENIE CENTAR ZA ISTRAZUVANJE I INFORMIRANJE ZA ZIVOTNA SREDINA EKO-SVEST	Eko-svest	North Macedonia	Еко-свест
11	MISSIONS PUBLIQUES	MP	France	MISSIONS PUBLIQUES
12	HOLISTIC IKE	HOLISTIC	Greece	HOLISTIC
13	UNIVERSITY OF VICTORIA	Uvic	Canada	University of Victoria

Executive Summary

This deliverable presents key results on clustering citizens based on their energy consumption and carbon footprint. In the first Work Package 4 deliverable ("*Deliverable 4.1: Report on qualified clustering input attributes*"), earlier studies were explored based on the patterns of socio-demographic variables, psychological characteristics, and energy consumption data. Outputs of D4.1 were presented at two levels, according to what level the data clusters refer to, namely individual and collective.

- At an individual level, major variables for clustering energy behaviours were categorized as socio-economic and demographic, psychological, energy consumption/ environmental patterns across different areas of life (housing, transport, etc.), and other contextual variables.
- At a collective level, major variables were categorized as socio-economic and demographic, energy infrastructure variables, energy consumption profiles, environmental performance, and other contextual factors.

Based on above patterns, attributes for clustering are proposed in this deliverable for three countries (UK, Greece, and The Netherlands), using data from the ECHOES ("Energy CHOices supporting the Energy Union and the SET-Plan") project.

A machine learning model is applied to the ECHOES data to identify distinct clusters. The model workflow starts with an estimation of people's energy consumption by applying a Life Cycle Assessment (LCA) method for housing and mobility, and the corresponding CO_2 footprint is then calculated for those same sectors. Finally, the k-means clustering method is applied to the energy consumption and associated CO_2 footprint for both housing and mobility to identify distinct citizens' groups in the UK, Greece, and the Netherlands (as representative European countries for which relevant data already exist).

Clustering results show that most of the citizens have low to medium energy consumption and CO_2 footprint on both housing and mobility sectors for the selected countries. Although, the majority of the citizens consume less energy and consequently produce less CO_2 footprint, there is group of citizens in the selected countries, with medium to high energy consumption and CO_2 footprint on both sectors, which can be insightful for policy makers. For instance, Dutch citizens have high carbon footprint on housing, while British citizens have high carbon footprint on mobility.

Contents

Figures

Tables

1 Introduction

1.1 Background

In the transformation of the energy system, the role of its citizens is becoming increasingly important through their engagement, involvement, and shaping of the future energy landscape. The importance of the influence of the public on energy consumption is reflected in the term "energy citizen" (Lennon & Dunphy, 2024). Citizens consume significant amounts of energy each year and their total energy consumption is increasing because of population growth, and this is in turn associated with increased emissions. This places "citizens at the heart of energy transitions" and it is crucial to create a framework to measure their carbon footprint and develop a pathway to decrease their influence on emissions (Blasch et al., 2022; Naderian et al., 2022).

Considering energy consumption and usage profiles to cluster citizens into focus groups has advantages and disadvantages. Energy consumption data can be potentially accessible, providing a simple and straightforward information source for decision makers. Assessing different temporal resolutions of energy usage could help improve the understanding of citizens' habits and routines. In this regard, using data with better temporal resolution than daily and monthly data could help policy makers to develop roadmaps and pathways for reducing citizens' CO_2 footprint in the long-term and to plan for permanent behavioural change. Specifically, using hourly electricity usage data could help decision makers to identify citizens' daily habits and to define required action plans for decreasing their CO_2 impact.

Most of the studies reviewed in *Deliverable 4.1* ("*D4.1: Report on qualified clustering input attributes*") focused on a specified perspective for clustering citizens (i.e., sociodemographic, psychological, energy consumption), but few studies considered a combination of perspectives. Also, the CO_2 footprint of citizens is not taken into account, generally (Naderian et al., 2022). This study aims to create more comprehensive insights into energy consumption behaviour and CO_2 footprint of a nation's citizens. The main contributions of this study can be summarized as follows:

- k-means clustering of citizens based on their energy consumption on mobility and housing in the UK, Greece, and the Netherlands (as representative European countries for which relevant data already exist).
- k-means clustering of citizens based on their CO_2 footprint on housing and mobility in the UK, Greece, and the Netherlands.
- Cross-analysis of the clustering results to shed light on the relationship between energy consumption and CO_2 footprint on housing and mobility for the UK, Greece, and the Netherlands.

1.2 Context

Within ENCLUDE, the aim of Work Package (WP) 4 is to identify groupings of citizens (clusters for decarbonization) at different scales of analysis, so that they can be more effectively mobilized by policies to accelerate energy system decarbonization.

Individual clusters for decarbonization are not necessarily groups of citizens with common demographic characteristics; rather, they may involve demographically diverse groups sharing common characteristics of energy behaviour, including readiness to embrace energy citizenship actions, which in turn result in different amounts of carbon footprint. Such groupings can inform policy makers of the citizens' profiles that are expected to be more responsive to energy citizenship policy initiatives, as these should be targeted due to their decarbonization potential.

The first task of WP4 was to identify key variables relevant to the emergence and fostering of energy citizenship that will accordingly enable the strategic clustering of citizens for decarbonization.

To this end, more than 180 journal and conference proceedings articles were reviewed in D4.1Click or tap here to enter text. to create a database of qualitative and quantitative attributes for the clustering process. The second task of WP4 (and focus of this deliverable) is the development of a machine learning model for citizens clustering to extract meaningful outcomes for policy makers.

1.3 Structure of the report

This report summarizes the findings from the clustering results, based on energy consumption data from the H2020 project ECHOES (2019), and subsequently, by calculating and analysing the associated CO_2 footprint for citizens of the UK, Greece, and the Netherlands.

The report is structured as follows: firstly, the method of estimating energy consumption and CO_2 footprint is presented in Sections 2.1 and 2.2, respectively; then the results of applying k-means clustering to energy consumption and CO_2 footprint in the UK, Greece, and the Netherlands are shown and discussed in Section 3; and finally, summary and conclusions are presented in Section 4.

2 Energy Consumption and CO_2 Footprint

In the ECHOES project, a Europe-wide fully anonymous representative survey covering 18,000 citizens provided rich information on energy attitudes and behaviour of citizens (Reichl et al., 2019). In this work, each country's subset of the ECHOES dataset has been selected to categorize citizens based on their energy consumption and the associated CO_2 footprint. Energy consumption (in megajoule) and CO_2 footprint calculations (in $KgCO_2$) have been reported per year per citizen. Due to the continuous nature of the measured values, k-means clustering has been used to group citizens.

2.1 Energy Consumption

The energy consumption of citizens entails different aspects of their lives, such as energy consumption from housing, mobility, food, goods, and leisure (Bird et al., 2020). Most CO_2 emissions are produced from the energy consumed in the housing and mobility sectors; based on a recently published report by the International Energy Agency (IEA), buildings and transport are responsible for 16.6% and 21.8% of global CO_2 emissions, respectively (IEA, 2022). There are different methods for estimating energy consumption in these two sectors, but Life Cycle Assessment (LCA) is regarded as one of the most efficient techniques as it considers individuals' attributes as a whole, including their behaviour and the equipment used in their energy consumption (Notter et al., 2013). To this end, different effective parameters such as fuel type (fossil fuels, renewable sources) and consumption patterns are accounted for. Also, instead of focusing on direct energy consumption, the LCA method considers the life cycle of energy consumption, including any behaviours, products or services that use energy directly or indirectly.

In the following subsections, the LCA methodology applied for estimating energy demand for the housing and mobility sectors is presented, as adopted from Schwarzinger et al. (2019a) and Bird et al. (2019).

2.1.1 Energy consumption in the housing sector

Most energy in houses is consumed by heating, cooling, hot water, cooking and appliances (Schwarzinger et al., 2019b).

2.1.1.1 Heating

The Tabula Web Tool (Loga et al., 2016) is used to estimate heat energy demand per square meter for different types of housing and the year of construction. The estimate is based on effective parameters such as size of the house, heating technology, number of residents, and preferred temperature. The ECHOES data for the UK, Greece, and the Netherlands, are assumed to be meaningful values and are substituted in Equation 2, as follows (Schwarzinger et al., 2019b):

$$Energy_{Heating} = f(House\ size\ , Fuel\ type, Number\ of\ occupants, Preferred\ temperature) \quad (1)$$

$$Energy_{Heating}(MJ) = \frac{House\ size\ (m^2)}{Number\ of\ Occupants}\ C_{en}\left(\frac{MJ}{m^2}\right)\ (1 + \frac{\Delta T}{Te_{avg} - Ti_{avg}}) \quad (2)$$

where the value of C_{en} is based on Tabula Web Tool data; ΔT is the room temperature preference of the respondent; Te_{avg} is the average outdoor temperature during the warm season [°K]; and Ti_{avg} is the average indoor temperature (assumed to be 21°C, or 293°K).

2.1.1.2 Cooling

Due to climate change, summers are becoming hotter compared to past decades and the demand for cooling systems and air conditioners (AC) is increasing. To estimate the energy demand for cooling in a specific house, different parameters could be considered as follows:

$$Energy_{Cooling} = f(House\ size\ , Number\ of\ occupants, Preferred\ temperature) \tag{3}$$

Cooling energy estimation is calculated as follows, based on (Stanciu et al., 2016):

$$Energy_{Cooling}(MJ) = \frac{House\ size\ (m^2)}{Number\ of\ Occupants} C_{en} \left(\frac{MJ}{m^2}\right) \frac{CDD}{HDD} \left(1 - \frac{T_{set}}{Te_{avg}(>24\ ℃) - Ti_{avg}}\right) \tag{4}$$

where HDD is the number of heating degree days, which equates to 2,654; CDD is the cooling coefficient, which is calculated using $22\ ℃$ as a base temperature and considered to be 5.4; T_{set} is the preferred indoor temperature setting; and the rate of AC usage chosen by the respondent in the ECHOES questionnaire, Te_{avg} and Ti_{avg}, are the same values considered for heating. Heating and cooling degree days are calculated based on National Oceanic and Atmospheric Administration (NOAA) city specific temperature data for the period 1990 – 2017 (NOAA, 2018).

2.1.1.3 Hot water

It is considered that a bath needs 175 litres, and a shower requires 55 litres of hot water and e the corresponding energy demand can be estimated with the following expression (Eurostat, 2018):

$$Energy_{Hot\ water}(MJ) = \frac{Bath\ or\ Shower\ (litres)}{Bath\ or\ Shower_{avg}(litres)} Energy_{\ Hot\ water}\ (MJ)\ C_{en}\left(\frac{MJ}{MJ}\right) \tag{5}$$

where $Bath\ or\ Shower$ is the amount of required water based on hot water usage type answered by the respondent; $Bath\ or\ Shower_{avg}$ is the average hot water used in the UK, Greece, and the Netherlands; $Energy_{\ Hot\ water}$ is the average energy consumption for hot water in the UK, Greece, and the Netherlands; and C_{en} is the energy coefficient, based on the fuel type and the technology of the household water heating appliances (Eurostat, 2018).

2.1.1.4 Cooking

The amount of energy consumed in the cooking process can be estimated using a procedure similar to that for hot water usage, as shown below (Bird et al., 2019):

$$Energy_{Cooking}(MJ) = \frac{No.\ of\ hot\ meals}{No.\ of\ hot\ meals_{avg}} Energy_{\ Cooking}\ (MJ)\ C_{en}\left(\frac{MJ}{MJ}\right) \tag{6}$$

where $No. of\ hot\ meals$ is the total number of hot meals each respondent has weekly; $No. of\ hot\ meals_{avg}$ is the average number of hot meal in the UK, Greece, and the Netherlands; $Energy_{\ Cooking}$ is the average energy consumption for cooking in the selected countries; and C_{en} is the energy coefficient, based on the fuel type and the technology used for cooking (Eurostat, 2018) .

2.1.1.5 Lighting and appliances

The household energy consumed by lighting and electrical appliances can be estimated from the following expression (Bird et al., 2019):

$$Energy_{Lighting\ \&\ Appliances}(MJ) = \ Energy\ _{LA_{avg}}\ (MJ)\ (0.65\ +\ 0.15C_l + 0.05C_s) \tag{7}$$

where $Energy\ _{LA_{avg}}$ is the average electricity for lighting and appliances in the UK, Greece, and the Netherlands; C_l and C_s represent each respondent's usage for lighting and standby appliances, respectively.

2.1.2 Energy consumption in the mobility sector

In this section, the energy demand estimation for mobility sector is presented, which includes the energy consumed on private transportation, public transportation, and aviation.

2.1.2.1 Private transportation

The energy demand for private transportation is estimated from the following expression (Bird et al., 2019):

$$Energy_{Private\ transportion}(MJ) = \frac{\frac{L(km)}{100}\ Eff_{car}\left(\frac{liter}{100km}\right) C_{fuel}(\frac{MJ}{liter})}{passengers_{avg}} \tag{8}$$

where L is the estimated weekly commuting distance of respondents; Eff_{car} is the car fuel consumption per 100 km of respondents; C_{fuel} is the fuel coefficient which represents the amount of energy per liter for a specified fuel of respondents; and the average number of occupants in each vehicle, $passengers_{avg}$, is calculated as follows:

$$passengers_{avg} = Usage_{vehicle}(\%) + \left(1 -\ Usage_{vehicle}(\%)\right)Number\_of\_passengers \tag{9}$$

$Usage_{vehicle}$ is how often the respondent drives alone and $Number\_of\_passengers$ is the number of passengers in a vehicle when the respondent is not alone.

2.1.2.2 Motorcycle

The energy demand for motorcycle usage is estimated from the following expression:

$$Energy_{Motorcycle}(MJ) = \frac{L(km)\ Eff_{Motorcycle}\ (liter/km)}{Motorcycle\_efficiency\ _{avg}(liter/km)}\ 1.626(\frac{MJ}{km}) \tag{10}$$

Similar to private vehicles, L represents the weekly commuting distance of respondents; $Eff_{Motorcycle}$ is the efficiency of motorcycle fuel consumption; and $Motorcycle\_efficiency\ _{avg}$ is the average efficiency of motorcycles in the selected countries.

2.1.2.3 Public transport

The energy demand for private transportation is estimated from the following expression (Bird et al., 2019):

$$Energy_{Public\ transportion}(MJ) = \sum_{Transport\ mode} C_{en}\left(\frac{MJ}{km}\right)\frac{time_{mode}(min)}{60}\ V_{mode}\left(\frac{km}{hr}\right) \tag{11}$$

Each respondent was asked about the type of public transportation they used and the duration of their usage. In equation (11), C_{en} is the energy coefficient based on the type of transportation; $time_{mode}$ is the approximate time of using public transport each week; and V_{mode} is the average speed of different types of public transportation, where 17.3, 15.0 and 32.5 km/hr are considered for bus, train, and underground, respectively.

2.1.2.4 Aviation

In this section, the energy demand estimation is presented for non-business and business flight travel which is based on each respondent's answer from the ECHOES data. For non-business flights, the following expressions are used to estimate the energy consumed:

For non-business flights of less than 10 hours:

$$Energy_{short\ flight}(MJ) = T(hr)\,V_{short}\left(\frac{km}{hr}\right)C_{en\_short}\left(\frac{MJ}{km}\right) \tag{12}$$

For non-business flights longer than 10 hours:

$$Energy_{long\ flight}(MJ) = T_{short}(hr)\left(\frac{km}{hr}\right)C_{en_{short}}\left(\frac{MJ}{km}\right) + \ (T_{long}\ - T_{short})\,(hr)\,V_{long}C_{en_{long}}\left(\frac{MJ}{km}\right) \tag{13}$$

where T_{short} and T_{long} are the durations of the short and long flights for the respondent, respectively; V_{short} and V_{long} are the speeds of the airplane for short and long flights and are considered to be $523\frac{km}{hr}$ and $750\frac{km}{hr}$, respectively; $C_{en_{short}}$ and $C_{en_{long}}$ are the energy coefficients defined for the short and long flights in the UK and they are assumed to be $0.7\frac{MJ}{km}$ and $0.56\ \frac{MJ}{km}$, respectively (Bird et al., 2020).

Business flight energy demand estimation has been implemented as follows:

$$Energy_{business\ flight}(MJ) = 2 * Energy_{long\ flight}(MJ) \tag{14}$$

2.2 CO_2 footprint

In 2021, the UK residential sector emitted 68.1 $MtCO_2$, i.e. 19.9% of the total UK CO_2 emissions (Department for Business, 2021). In order to gain a better understanding of the amount of CO_2 emissions attributable to UK citizens' energy consumption habits, the responses collected through the ECHOES survey were initially translated into corresponding energy demand (see section 2.1). The energy demand values were subsequently translated into amounts of CO_2 emissions based on conversion factors for greenhouse gas (GHG) emissions recently published by the UK government (Department for Environment, 2024) . The CO_2 emission factors for the mobility and housing sectors are provided in **Tables 1 and 2**, respectively, along with the measurement units and a breakdown of the fuel types considered. In these tables, Scope 1 emissions are considered for passenger vehicles and fuels, and Scope 3 emissions for flights. Full electrification of the rail system is assumed here, even though only ~40% of the rail network is currently electrified in the UK (and ~60% in Europe); this is because it was not possible to allocate specific rail lines use to respondents. Similar coefficients for Greece and the Netherlands are presented in **Tables 3-6** (European Environment Agency, 2022). Different sub-sectors in housing and mobility are considered to allow a more detailed breakdown of the CO_2 footprint of each respondent (only electric cooling and cooking are considered) (European Environment Agency, 2022).

Table 1- Emission factors for different subcategories of housing based on fuel type and energy consumption for the UK (Department for Environment, 2024).

Housing subcategory	Measuring Parameter	Fuel type				
		Gas	Oil	Wood	Heat pump	Electricity
Heating	$kgCO_2/kwh$	0.2	0.29	0.34	0.19338	0.19338
Cooling	$kgCO_2/kwh$	-	-	-	-	0.19338
Hot water	$kgCO_2/kwh$	0.2	0.29	0.34	-	0.19338
Cooking & appliances	$kgCO_2/kwh$	-	-	-	-	0.19338

Table 2 - Emission factors for different subcategories of mobility based on fuel type and energy consumption for the UK (Department for Environment, 2024).

Mobility subcategory	Measuring Parameter	Fuel type				
		Petrol	Diesel	Hybrid	Electricity	Gas
Car (Medium)	$kgCO_2$/passenger/km	0.1781	0.1671	0.1090	0	0.1566
Motorcycle (Average)	$kgCO_2$/passenger/km	0.11367	-	-	0	-
Train	$kgCO_2$/passenger/km	-	-	-	0	-
Underground	$kgCO_2$/passenger/km	-	-	-	0	-
Bus (Average)	$kgCO_2$/passenger/km	0.089 or 0 if bus uses electricity				
Non business flights	$kgCO_2$/passenger/km	0.18592 for short flights and 0.26128 for long flights				
Business flight	$kgCO_2$/passenger/km	0.27430 for short flights and 0.58029 for long flights				

Table 3 - Emission factors for different subcategories of housing based on fuel type and energy consumption for Greece (European Environment Agency, 2022).

Housing subcategory	Measuring Parameter	Fuel type				
		Gas	Oil	Wood	Heat pump	Electric-ity
Heating	$kgCO_2/kwh$	0.2008	0.2665	0	0.344	0.344
Cooling	$kgCO_2/kwh$	-	-	-	-	0.344
Hot water	$kgCO_2/kwh$	0.2008	0.2665	0	-	0.344
Cooking & appliances	$kgCO_2/kwh$	-	-	-	-	0.344

Table 4 - Emission factors for different subcategories of mobility based on fuel type and energy consumption for Greece (European Environment Agency, 2022).

Mobility subcategory	Measuring Parameter	Fuel type				
		Petrol	Diesel	Hybrid	Electricity	Gas
Car (Average)	$kgCO_2$/passenger/km	0.1265	0.1259	0.03985	0	0.1019
Motorcycle (Average)	$kgCO_2$/passenger/km	0.103	-	-	0	-
Train	$kgCO_2$/passenger/km	-	-	-	0	-
Underground	$kgCO_2$/passenger/km	-	-	-	0	-
Bus (Average)	$kgCO_2$/passenger/km	0.105 or 0 if bus uses electricity				
Non business flights	$kgCO_2$/passenger/km	0.077033 for short flights and 0.1062 for long flights				
Business flight	$kgCO_2$/passenger/km	0.2311 for short flights and 0.3186 for long flights				

Table 5 - Emission factors for different subcategories of housing based on fuel type and energy consumption for the Netherlands (European Environment Agency, 2022).

Housing subcategory	Measuring Parameter	Fuel type				
		Gas	Oil	Wood	Heat pump	Electric-ity
Heating	$kgCO_2/kwh$	0.2008	0.2665	0	0.354	0.354
Cooling	$kgCO_2/kwh$	-	-	-	-	0.354
Hot water	$kgCO_2/kwh$	0.2008	0.2665	0	-	0.354
Cooking & appliances	$kgCO_2/kwh$	-	-	-	-	0.354

Table 6– Emission factors for different subcategories of mobility based on fuel type and energy consumption for the Netherlands (European Environment Agency, 2022).

Mobility subcategory	Measuring Parameter	Fuel type				
		Petrol	Diesel	Hybrid	Electric-ity	Gas
Car (Average)	$kgCO_2$/passenger/km	0.1440	0.1268	0.0361	0	0.1060
Motorcycle (Average)	$kgCO_2$/passenger/km	0.103	-	-	0	-
Train	$kgCO_2$/passenger/km	-	-	-	0	-
Underground	$kgCO_2$/passenger/km	-	-	-	0	-
Bus (Average)	$kgCO_2$/passenger/km	0.105 or 0 if bus uses electricity				
Non business flights	$kgCO_2$/passenger/km	0.077033 for short flights and 0.1062 for long flights				
Business flight	$kgCO_2$/passenger/km	0.2311 for short flights and 0.3186 for long flights				

2.3 K-means Clustering Method

Clustering is a technique that groups similar data to uncover valuable insights and aid in predictive modelling. In this study, k-means clustering is used to identify similarities in the energy consumption data and CO_2 emission footprint of citizens from the selected countries in the ECHOES data set.

The k-means algorithm is a machine learning-based method that divides data into predefined groups or clusters based on their similarity (J MacQueen, 1967). To achieve this, the algorithm calculates the Euclidean distance between data points, allowing them to be grouped together into clusters with similar characteristics. **Figure 1** shows the k-means method for clustering unlabeled data.

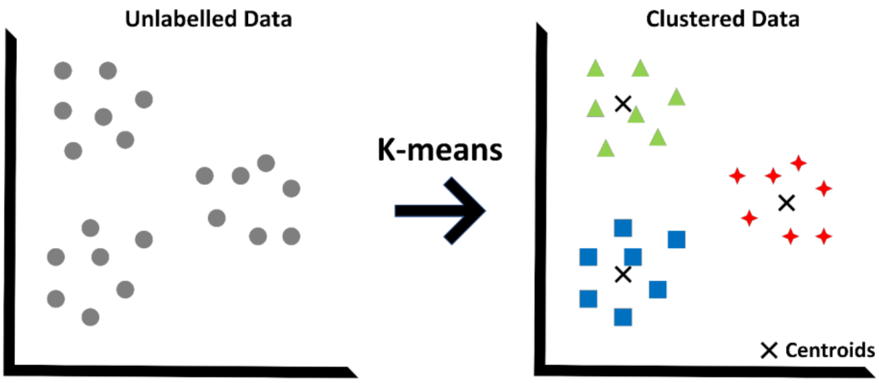

Figure 1 – K-means clustering method.

The steps of the k-means clustering method include the following:

1. Data points $(x_j, j = 1,2, \ldots n)$ are randomly assigned as clusters' centroids.
2. The measure criterion (i.e., the Euclidean distance) is calculated between each data point and each cluster's centroid $(\mu_k, k = 1,2, \ldots m)$. x_j is considered to be a member of cluster C_i if the measured distance between it and cluster's centroid μ_k is less than for the other centroids.
3. The mentioned procedure is repeated to create m clusters; each iteration will see centroids being re-assigned based on the average of new data points located in that cluster.
4. The whole process of the k-means clustering algorithm (distance measurement, dividing data points into groups, and centroid calculation) is repeated until there is no change in the centroid's position (Frédéric Magoules & Hai-Xiang Zhao, 2016).

To measure the convergence of the k-means clustering method, the following expression was used:

$$F = \sum_{k=1}^{m} \sum_{x_j^{(i)} \in C_i} (x_j^{(i)} - \mu_k)^2 \tag{15}$$

where $x_j^{(i)}$ is a data point and is a member of cluster k. The flowchart of the algorithm is presented in **Figure 2**. The optimal number of clusters is important and should be figured out before clustering. This study uses the elbow method to optimise the number of clusters which is determined from calculating the Within Cluster Sum of Squares (WCSS), which is a measure of the sum of the squared distance between data points in a specified cluster and the cluster centroid (Thorndike, 1953). The clustering algorithm is implemented in Python using the scikit-learn library (Buitinck et al., 2013).

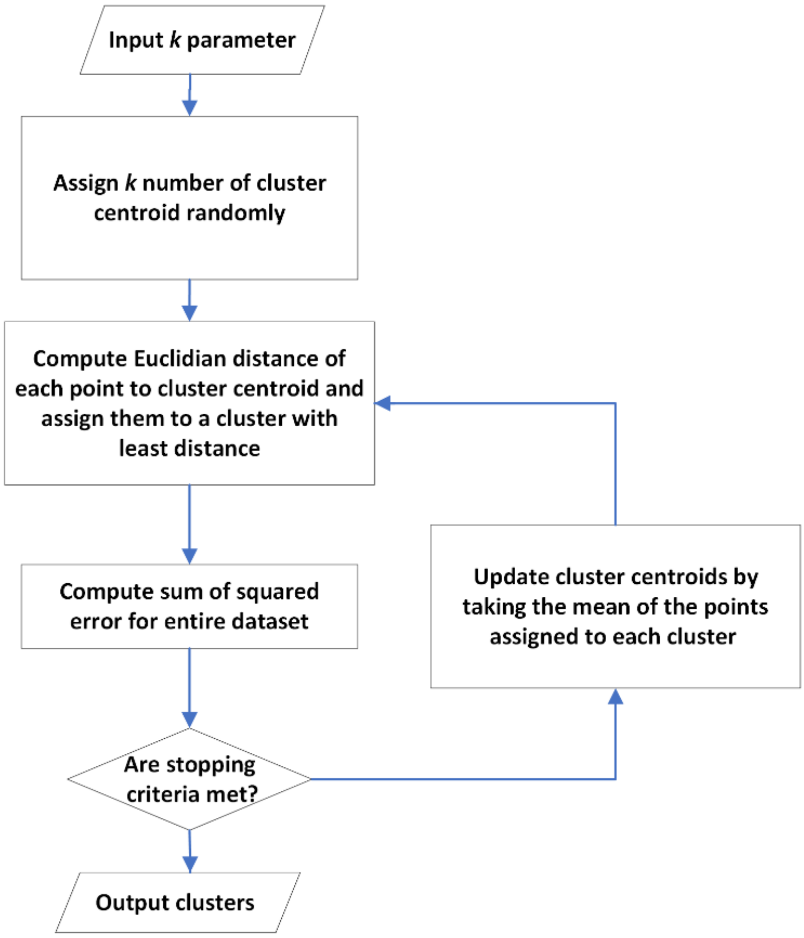

Figure 2 - Flowchart of the k-means clustering algorithm.

3 Clustering results and discussion

This section provides the results and discussion of the machine learning model developed for clustering the energy consumption and CO_2 footprint of citizens in the UK, Greece, and the Netherlands.

3.1 Clustering based on energy consumption for citizens in the UK

A data set of energy consumption for each citizen was created based on the energy demand calculation method presented in Section 2.1. The k-means clustering method is then applied to categorise the data into distinct groups. To create more stable centroids for the k-means algorithm, an optimal number of groups (k) is important and, to this end, the performance of the k-means technique is measured for different numbers of clusters. The elbow method has been applied for different values of k to find the optimum number of clusters (Buitinck et al., 2013). The elbow method performs well for small values of k and it measures the squared difference for different k values. As the k value increases, the number of data points in each cluster will decrease and the centroids will be closer to each other. As a decision criterion, the WCSS provides a measure of variance within each cluster such that a lower WCSS could create more efficient clustering results. **Figure 3** presents the elbow method results, which suggests that five clusters is considered optimal, as above that number the WCSS value does not change significantly.

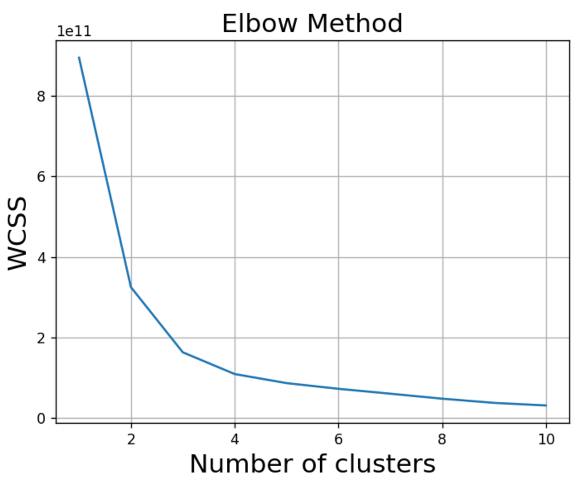

Figure 3 – Elbow method results for different number of clusters for energy consumption clustering in the UK

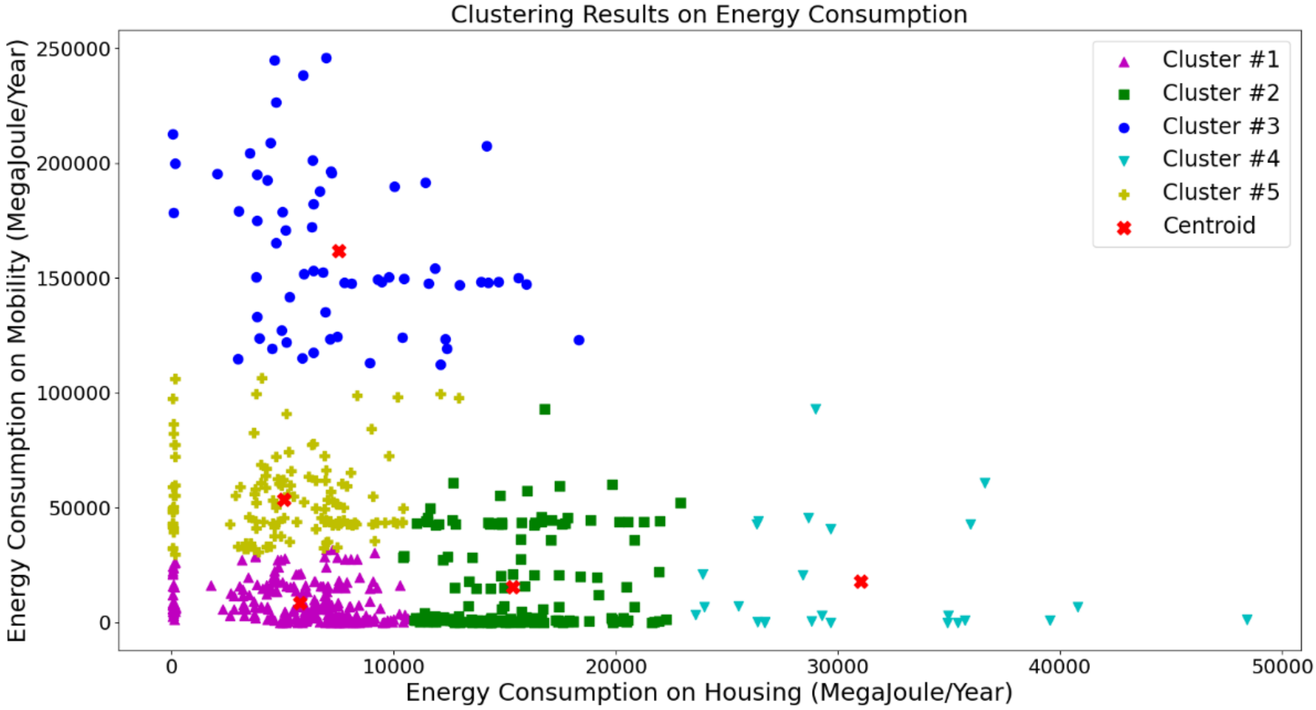

Figure 4 – Clustering results on energy consumption data for housing and mobility sectors in the UK

The applied parameters for the k-means clustering method used in this study are given in **Table 7** and the results for energy consumption clustering are shown in **Figure 4**. Each of the five cluster locations and their associated populations are presented in **Table 8**.

Table 7- k-means clustering method parameters for both energy consumption and CO_2 footprint clustering in the UK.

Parameter	Description	Value/Method
k	The number of clusters	5
Initial centroids	The initial values for the centroids of each cluster	Random
Distance metric	The method used to calculate the distance between data points and cluster centroids	Euclidean distance
Maximum number of iterations	The maximum number of iterations to perform	300
Number of initializations	Number of times the k-means algorithm is run with different centroid seeds	10

Table 8 - Clusters' results for energy consumption data on housing and mobility in the UK

Cluster	Population	Centroid location on housing energy (MJ)	Centroid location on mobility energy (MJ)
1	241	5780.56	8485.24
2	160	15370.63	15371.64
3	62	7520.79	162020.77
4	25	31019.94	17872.91
5	134	5066.83	53734.77

3.2 Clustering based on CO_2 footprint for citizens in the UK

The previous section presented the clustering results based on energy consumption data for UK citizens. The same approach is taken for the CO_2 footprint estimation based on the ECHOES energy consumption calculations, and the results are presented here.

As per before, the optimal number of clusters is determined from the elbow method, the results of which are presented in **Figure 5**. Five clusters are considered as optimal, as after that number of groups, the value of WCSS does not change noticeably.

The clustering results are presented in **Figure 6** and **Table 9**. The clustering results for CO_2 footprint show a similar pattern to that for energy consumption. Clusters A and C have low to medium housing CO_2 footprint, while clusters B and D have more widely spread centroids on housing CO_2 footprint. While the differences in housing CO_2 footprint between the clusters are negligible, the centroids' locations for mobility CO_2 footprint are vastly different, particularly for clusters A, C, and E.

According to **Table 9**, almost 65% of the citizens that are grouped in cluster A and cluster C have a low to medium CO_2 footprint while 30% of the citizens in clusters B and D have low CO_2 footprint on mobility, but high CO_2 footprint on housing. Only 5% of the citizens which are grouped in cluster E have high level of CO_2 footprint on mobility. This observation suggests that the mobility sector might be an important lever to affect behavioural change of citizens due to its high CO_2 footprint and it helps policy makers define mobility-focused roadmaps and incentives to reduce CO_2 footprint.

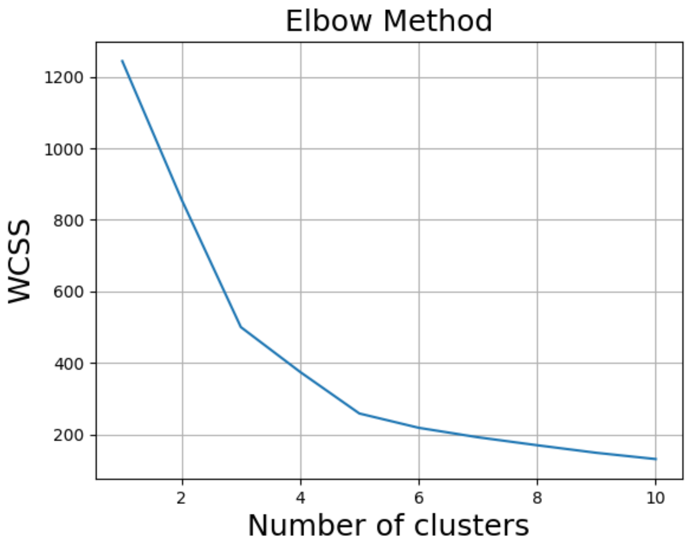

Figure 5- Elbow method results for different number of clusters for CO2 footprint measurement clustering

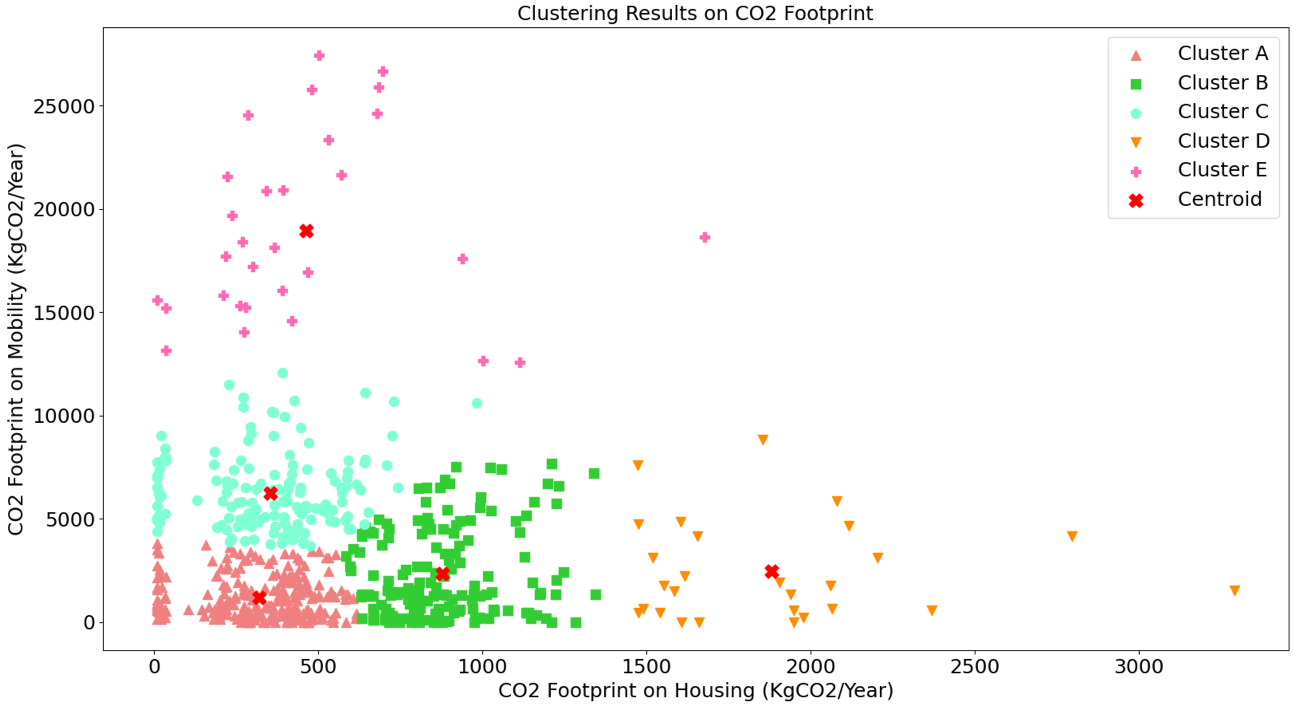

Figure 6- Clustering results on CO2 footprint measurement data for housing and mobility sectors in the UK

Table 9- Clusters results for CO2 footprint measurement data for the housing and mobility sectors in the UK

Cluster	Population	Centroid location on housing CO_2 footprint ($kgCO_2$/year)	Centroid location on mobility CO_2 footprint ($kgCO_2$/year)
A	247	320.63	1180.30
B	161	878.96	2343.67
C	157	352.97	6234.85
D	27	1881.56	2475.36
E	30	463.25	18935.05

Combining energy consumption (section 3.1) and carbon footprint of each citizen can help to assess the carbon intensity of their energy usage by sector. Figure 7 shows the relation between energy consumption and carbon footprint for UK citizens for the mobility sector, for example. It is clear from this figure that there is a semi-linear corelation between energy consumption and carbon footprint data for all UK citizens, expect for those who belong to cluster 3. Further insights into the socio-demographic composition of different clusters with different carbon intensity of their energy consumption will be provided in deliverable D4.3 on bridging clusters for decarbonization with energy models.

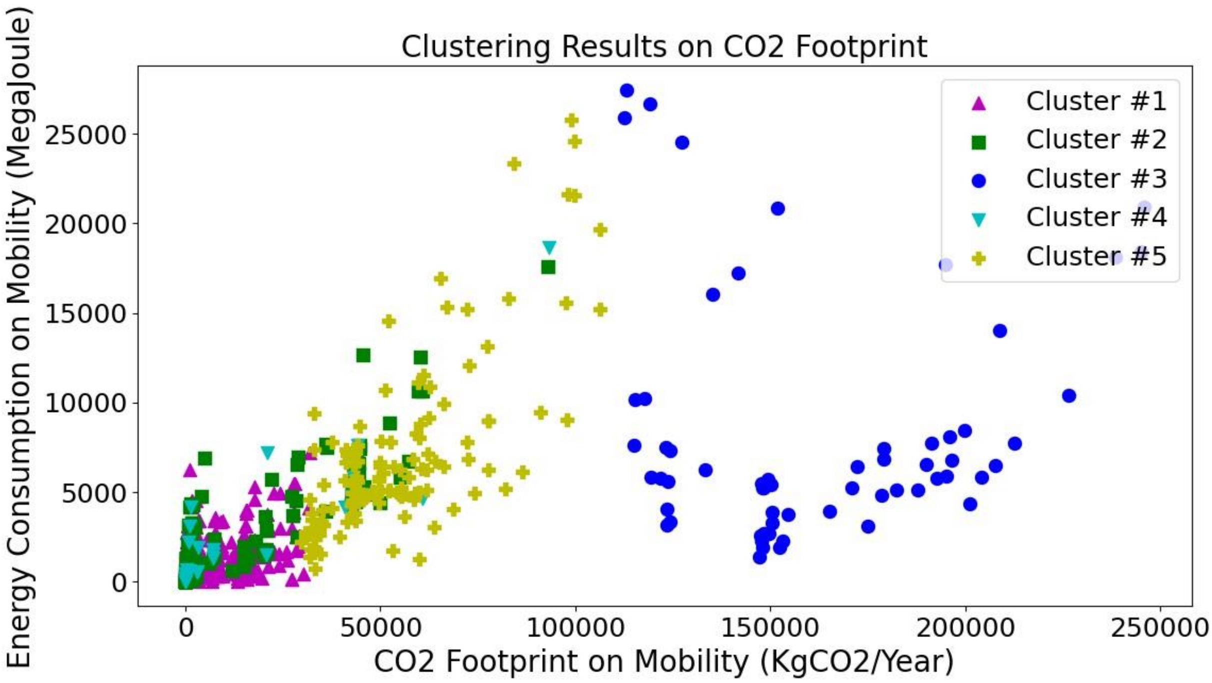

Figure 7 - Correlation between energy consumption and carbon footprint for the UK citizens for the mobility sector

3.3 Clustering based on energy consumption for citizens in Greece

Following the same procedure as per Section 3.2, based on the Elbow method, 3 clusters are considered as an optimum number. The applied parameters for the k-means clustering are given in **Table 7** and the results for energy consumption clustering are shown in **Figure 8**. Each of the three clusters' centroid and their associated populations are presented in **Table 10**.

The clustering results reveal that distinct groups of citizens in Greece can be identified based on both housing and mobility energy consumption, as determined by the number of clusters. Analysis of the centroid data presented in **Table 10** reveals that energy consumption levels among Greek citizens are relatively consistent across the housing and mobility sectors for the majority of the citizens within cluster 1, while significant differences exist in energy consumption for mobility and housing in cluster 2 and cluster 3, respectively.

Table 10 - Clusters results for energy consumption data on housing and mobility in Greece

Cluster	Population	Centroid location on housing energy (MJ)	Centroid location on mobility energy (MJ)
1	316	1106.11	17999.71
2	181	1489.15	86437.60
3	106	4911.83	35735.73

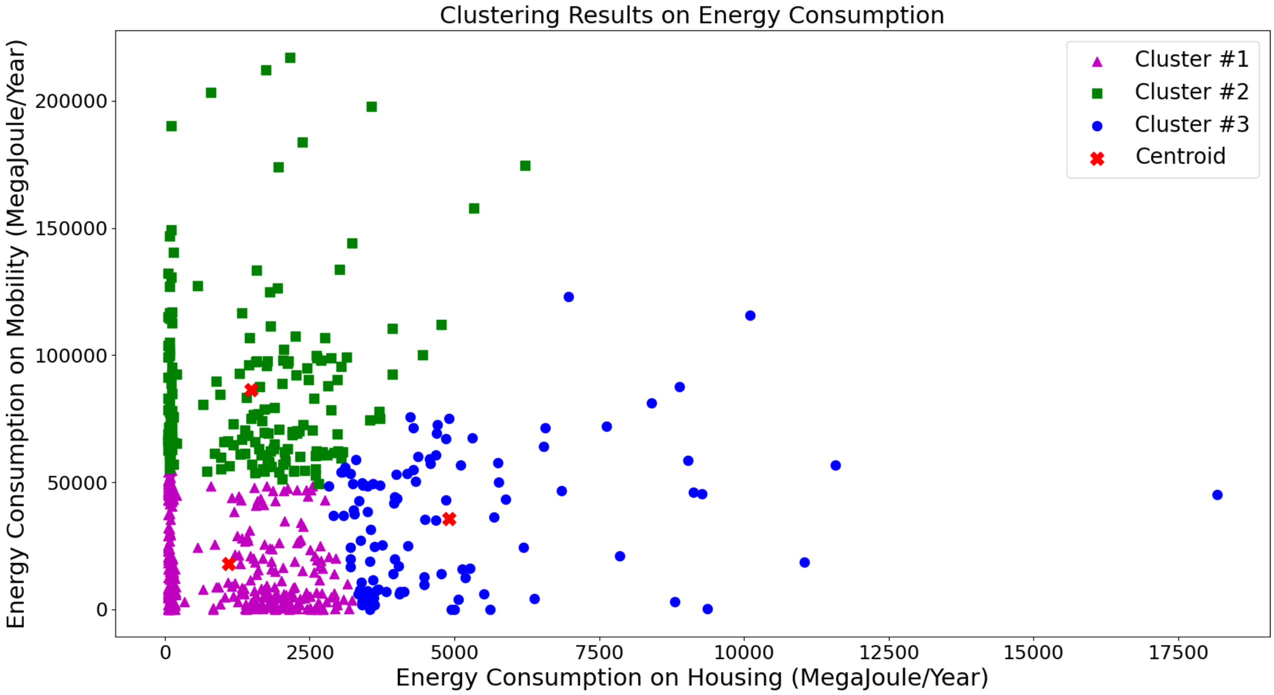

Figure 8 - Clustering results on energy consumption data for housing and mobility sectors in Greece

3.4 Clustering based on CO_2 footprint for citizens in Greece

As per Section 3.3, the optimal number of clusters is determined from the Elbow method and 3 clusters are considered. The clustering results are presented in **Figure 9** and **Table 11**. The results for CO_2 footprint show a similar pattern to that for energy consumption. Cluster 1 has low to medium CO_2 footprint on both sectors, while clusters 2 and 3 have more widely spread centroids for CO_2 footprint on mobility and housing, respectively.

According to **Table 11**, almost 72.6% of the citizens that are grouped in cluster 1 have a low to medium CO_2 footprint, while 14.9% of the citizens in cluster 2 have medium to high CO_2 footprint on mobility. Only 12.4% of the citizens which are grouped in cluster 3 have a high level of CO_2 footprint on housing. This observation suggests that the mobility sector might be an important lever to affect behavioural change in citizens as it helps policy makers define mobility-focused roadmaps and incentives to reduce CO_2 footprint.

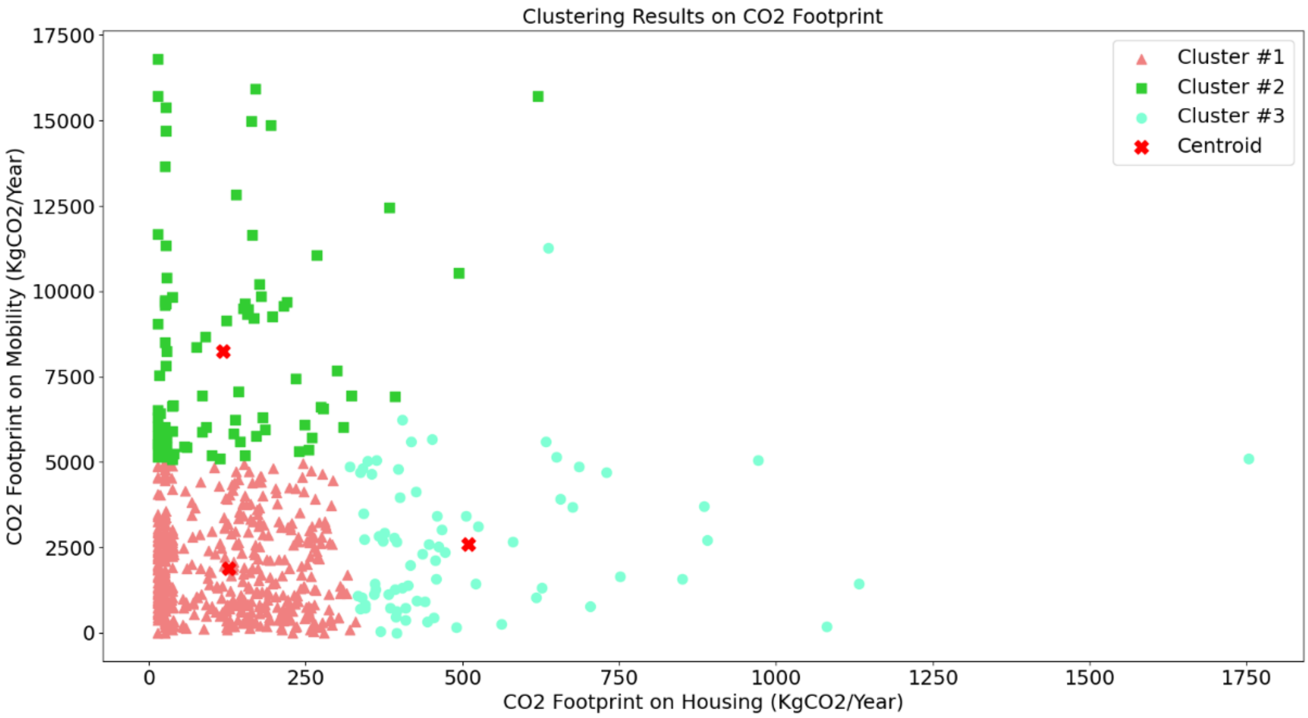

Figure 9 - Clustering results on CO2 footprint measurement data for housing and mobility sectors in Greece

Table 11- Clusters results for CO2 footprint measurement data for the housing and mobility sectors in Greece

Cluster	Population	Centroid location on housing CO_2 footprint ($kgCO_2$/year)	Centroid location on mobility CO_2 footprint ($kgCO_2$/year)
1	438	127.41	1888.73
2	90	118.36	8236.55
3	75	509.80	2603.62

3.5 Clustering based on energy consumption for citizens in The Netherlands

Based on the Elbow method, 4 clusters are considered as an optimum number of clusters. The applied parameters for the k-means clustering method used in this study are given in **Table 7** and the results for energy consumption clustering are shown in **Figure 10**. Each of the four cluster locations and their associated populations are presented in **Table 12** with respective centroids.

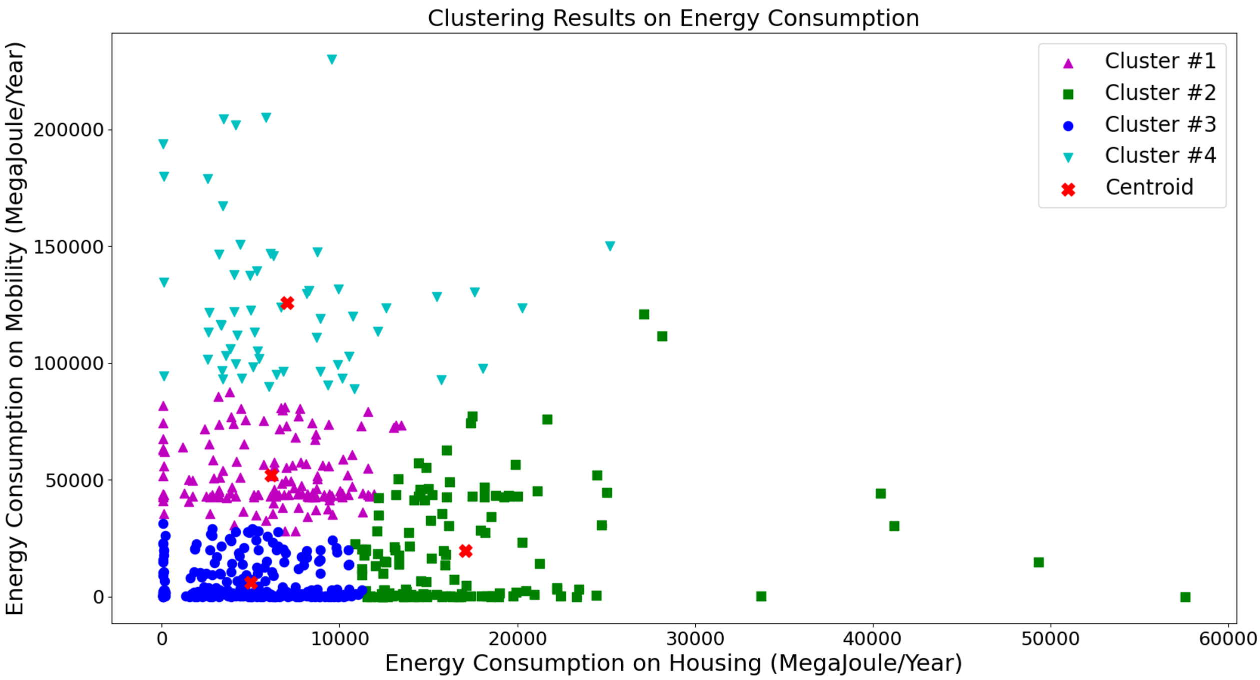

Figure 10 - Clustering results on energy consumption data for housing and mobility sectors in the Netherlands

The clustering results reveal that distinct groups of the Dutch citizens can be identified based on both housing and mobility energy consumption, as determined by the number of clusters. Analysis of the centroid data presented in **Table 12** reveals that energy consumption levels among the Dutch citizens are relatively consistent across the housing and mobility sectors for 44.1% of the citizens located in the cluster 3, while significant differences exist in energy consumption within the mobility sector in cluster 1 and cluster 4 which jointly represent 33.2% of the population. In cluster 2, 21.9% of the Dutch citizens have medium to high energy consumption on housing. This observation suggests that the mobility sector might be an important lever to affect behavioural change in citizens particularly for cluster 4 which pollutes 20 times more than cluster 3 (if centroids are considered for the comparison) and this insight can help policy makers define mobility-focused roadmaps and incentives to reduce energy consumption.

Table 12- Clusters results for energy consumption data on housing and mobility in the Netherlands

Cluster	Population	Centroid location on housing CO_2 footprint ($kgCO_2$/year)	Centroid location on mobility CO_2 footprint ($kgCO_2$/year)
1	144	6139.33	52062.427
2	132	17105.75	19557.54
3	266	4995.31	6117.30
4	60	7036.23	125976.65

3.6 Clustering based on CO_2 footprint for citizens in The Netherlands

The clustering results are presented in **Figure 11** and **Table 13**. The results for CO_2 footprint show a similar pattern to that for energy consumption. Cluster 1 has low to medium CO_2 footprint on the housing and mobility sectors, while clusters 3 and 4 have more widely spread centroids on housing CO_2 footprint. Although, the differences in housing CO_2 footprint between clusters 1 and 2 are negligible, the centroids' locations for mobility CO_2 footprint are vastly different for cluster 2.

According to **Table 13**, almost 40% of the citizens that are grouped in cluster 1 have a low to medium CO_2 footprint on housing and mobility, while 40.8% of the citizens in clusters 3 and 4 have medium to high CO_2 footprint on housing. Only 20% of the citizens which are grouped in cluster 2 have high level of CO_2 footprint on mobility. The population of cluster 3 represents the majority of the Dutch citizens, so this observation suggests that the housing sector might be an important lever to affect behavioural change in the Dutch citizens as it helps policy makers define housing-focused roadmaps and incentives to reduce CO_2 footprint.

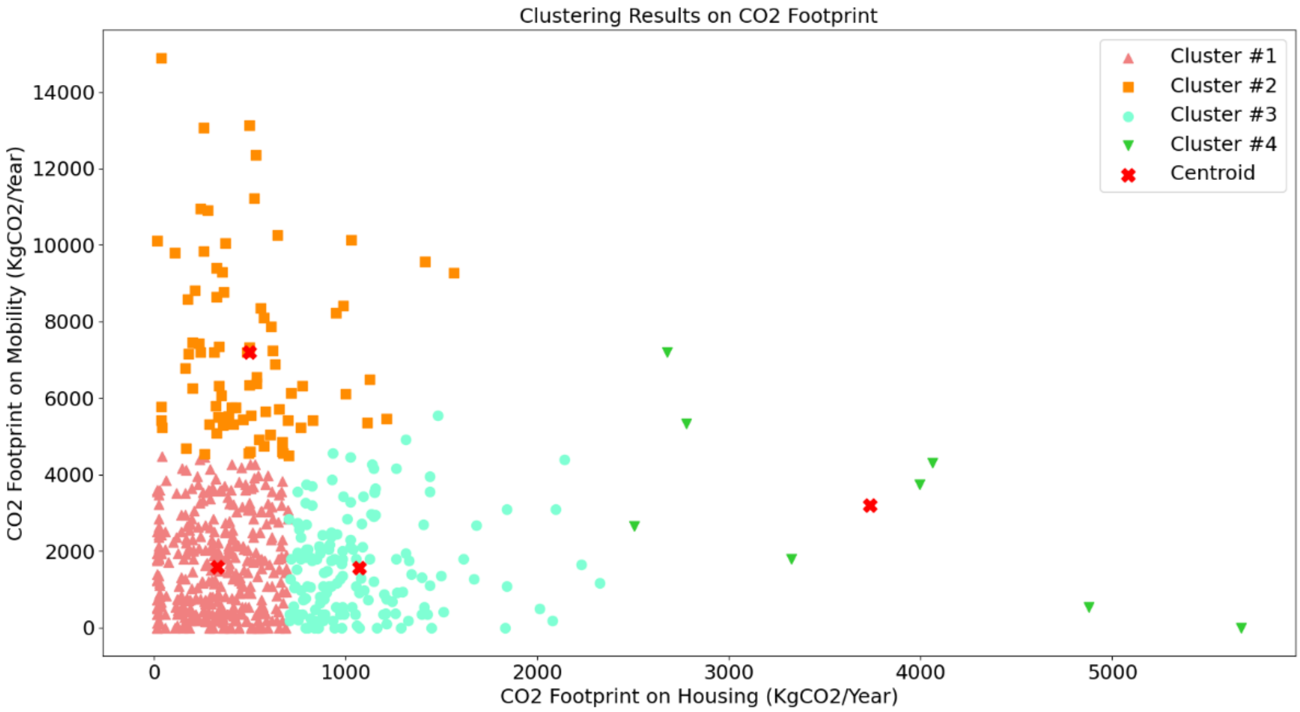

Figure 11 - Clustering results on CO2 footprint measurement data for housing and mobility sectors in the Netherlands

Table 13- Clusters results for CO_2 footprint measurement data for the housing and mobility sectors in the Netherlands

Cluster	Population	Centroid location on housing CO_2 footprint ($kgCO_2$/year)	Centroid location on mobility CO_2 footprint ($kgCO_2$/year)
1	160	331.66	1594.54
2	79	1068.59	1570.25
3	157	496.39	7202.04
4	8	3736.15	3202.97

4 Conclusions

This deliverable is part of the ENCLUDE project, which aims to promote energy citizenship and sustainable practices across the EU. The specific objective of D4.2 was to identify clusters of citizens based on their energy consumption patterns and carbon footprint for decarbonization efforts.

The deliverable first provided background on the importance of understanding citizen energy usage and emissions for energy transitions. It then outlined the methodology used to estimate energy consumption in the housing sector (based on residency characteristics) and mobility sector (based on transportation modes) for citizens in the UK, Greece, and the Netherlands using data from the earlier ECHOES project.

The K-means clustering algorithm was applied to the energy consumption data and CO_2 footprint to group citizens into distinct clusters. For the UK, five clusters were identified based on housing and mobility energy usage patterns. For Greece and the Netherlands, three and four clusters emerged, respectively.

Most of the UK citizens were grouped as low to medium energy consumers on both mobility and housing, while the minority of them consume medium to high energy on mobility, which is a key insight for policy makers because similarities appeared in the CO_2 footprint clustering results, thus reducing energy consumption on mobility by those citizens could lead to their reduced CO_2 footprint.

Similar to the UK, the majority of the Greek citizens were grouped as low to medium energy consumers on both mobility and housing, with only a minority consuming medium to high energy on mobility. Again, reducing the energy consumption on mobility of this minority of citizens would lead to a reduced CO_2 footprint.

For the Dutch citizens, the clustering results are slightly different. While most of the citizens still consumes low to medium energy on housing and mobility, there is noticeable group of citizens that consumes medium to high energy on housing. They represent double the population of the citizens with high energy consumption on mobility. This key insight could be very informative for policymakers, suggesting that they should prioritise the decarbonisation of the mobility sector in the UK and Greece, and that of the housing sector in the Netherlands.

In summary, this deliverable presents a data-driven approach to group citizens based on energy consumption and carbon footprint, offering a basis for developing tailored strategies to engage diverse citizen groups in the clean energy transition. The next step is to further investigate the socio-demographic characteristics of the clustered energy citizens, to better understand the relations between citizens' attributes, their current energy behaviours, and their attitudes to the energy transition. This will be discussed in D4.3.

5 References

Bird, D. N., Schwarzinger, S., Kortschak, D., & Strohmaier, M. (2020). *Report: A Detailed Methodology for the Calculation of Cumulative Energy Demand per Survey Respondent Catastrophic Risk Modelling View project CLEANcultures-An approach for innovative Climate Learning, Evaluation and Action in Neighbourhoods View project*. https://doi.org/10.13140/RG.2.2.36003.40488

Bird, D. N., Schwarzinger, S., Kortschak, D., Strohmaier, M., & Lettmayer, G. (2019). Report: A detailed methodology for the calculation of cumulative energy demand per survey respondent. *The ECHOES Consortium: Brussels, Belgium*.

Blasch, J., Barnes, J. P., & Palm, J. (2022). *Putting People at the Heart of Energy Transitions Social Innovation in Energy: four projects shine a light on the path forward*. https://www.researchgate.net/publication/360218763

Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R., Vanderplas, J., Joly, A., Holt, B., & Varoquaux, G. (2013). *API design for machine learning software: experiences from the scikit-learn project*.

Department for Business, E. & I. S. (2021). *2021 UK greenhouse gas emissions, provisional figures*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1064923/2021-provisional-emissions-statistics-report.pdf

Department for Environment, F. & R. A. (2023, June 27). *Greenhouse gas reporting: conversion factors 2023*. https://www.gov.uk/government/publications/greenhouse-gas-reporting-conversion-factors-2023

ECHOES. (2019). *ECHOES data set*.

European Environment Agency. (2022). *European Environment Agency Datahub*. https://www.eea.europa.eu/en/datahub/datahubitem-view/5d252092-d328-40d8-bca2-c0734bd6143b

Eurostat. (2018). *Complete energy balances - annual data*. http://ec.europa.eu/eurostat/web/products-datasets/-/nrg_110a

Frédéric Magoules, & Hai-Xiang Zhao. (2016). *Data Mining and Machine Learning in Building Energy Analysis*. Wiley-ISTE.

IEA. (2022). *Buildings*. https://www.iea.org/reports/buildings

J MacQueen. (1967). Classification and analysis of multivariate observations. *5th Berkeley Symp. Math. Statist. Probability*.

Lennon, B., & Dunphy, N. P. (2024). Energy and Citizenship in a Time of Transition and Socio-technical Change. In *The Palgrave Handbook of Global Social*

Change (pp. 1–15). Springer International Publishing. https://doi.org/10.1007/978-3-030-87624-1_223-2

Loga, T., Stein, B., & Diefenbach, N. (2016). TABULA building typologies in 20 European countries—Making energy-related features of residential building stocks comparable. *Energy and Buildings*, *132*, 4–12. https://doi.org/10.1016/j.enbuild.2016.06.094

National Oceanic and Atmospheric Administration (NOAA). (2018). *Annual 2018 Global Climate Report*. https://www.ncei.noaa.gov/cdo-web/

Notter, D. A., Meyer, R., & Althaus, H.-J. (2013). The Western Lifestyle and Its Long Way to Sustainability. *Environmental Science & Technology*, *47*(9), 4014–4021. https://doi.org/10.1021/es3037548

Reichl, J., Cohen, J., Kollmann, A., Azarova, V., Klöckner, C., Royrvik, J., Vesely, S., Carrus, G., Panno, A., Tiberio, L., Fritsche, I., Masson, T., Chokrai, P., Lettmayer, G., Schwarzinger, S., & Bird, N. (2019, November 1). *International survey of the ECHOES project*.

Schwarzinger, S., Bird, D. N., Lettmayer, G., Henriksen, I. M., Skjølsvold, T. M., Olaeta, X. U., Alvarez, L. P., Velte, D., Iturriza, I. J., Biresselioglu, M. E., Demir, M. H., Dimitrova, E., Tasheva, M., Tiberio, L., Panno, A., Carrus, G., & Costa, S. (2019). *Comparative assessment of European energy lifestyles. Deliverable 5.1, H2020 project ECHOES*.

Sobhan Naderian, Gioia Falcone, & Anastasia Ioannou. (2022). *D4.1 – Report on qualified clustering input attributes*.

Stanciu, C., Șoriga, I., Gheorghian, A. T., & Stanciu, D. (2016). Comfort air temperature influence on heating and cooling loads of a residential building. *IOP Conference Series: Materials Science and Engineering*, *147*, 012145. https://doi.org/10.1088/1757-899X/147/1/012145

Thorndike, R. L. (1953). Who belongs in the family? *Psychometrika*, *18*(4), 267–276. https://doi.org/10.1007/BF02289263

PARTICIPANTS

TU Delft
TEESlab University of Piraeus Research Center
ETH zürich
UCC University College Cork, Ireland Coláiste na hOllscoile Corcaigh
University of Glasgow
JOANNEUM RESEARCH LIFE
Th!nk E
Utrecht University
Green Partners environmental consulting
Еко-свест
MISSIONS PUBLIQUES Bringing citizens into policy
HOLISTIC
University of Victoria

ENCLUDE project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 101022791

ENCLUDE
Energy Citizens for Inclusive Decarbonization

enclude

@encludeproject

www.encludeproject.eu