WHY

D4.1

R2: WHY Toolkit

LEGAL DISCLAIMER

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 891943.

DOCUMENT INFORMATION

Deliverable title	R2: WHY Toolkit
Dissemination level	Public
Submission deadline	28/02/2023
Submission date	09/02/2024
Version number	1
Authors	Cruz E. Borges (UD) Carlos Quesada Granja (UD) Aritz Bilbao Jayo (UD) Xabier Cantero Lopez (UD) Leire Astigarraga (GOI) Ruth Carbajo Garcia(UD) Armando Aguayo Mendoza (UD)
Internal reviewers	
External peer reviewers	
Document approval	
Scope of the document according to the DoA	The report will contain the detailed description of the methodologies followed to prepare the WHY Toolkit and the results of the test simulations carried out.

TABLE OF CONTENTS

Contenido

TABLE OF FIGURES

LIST OF ACRONYMS AND ABBREVIATIONS

Acronym	Long text
PED	Positive Energy District
TSO	Transmission system operators
MAPE	Mean absolute percentage error
MTLF	Medium term load forecasting
STLF	Short term load forecast
SOM	Self-Organizing Map

1. Introduction

This deliverable presents the main results of applying the methodologies developed in WP2 and WP3 to develop tools of interest. In this sense, we present here the three different tools:

- A medium term load forecasting (MTLF) tool
- A model to assess the impact of a change of tariff
- An improved "standardised" load profiles

Section 2 deals with the explanation of the MTLF tool. This tool works taking into account an improved version of the time series clusterization methodology created in Deliverable D2.1. This methodology takes into consideration clusterization from different datasets and selects at the same time the number of overall clusters as well as to identify the similar clusters between different datasets. 30 clusters were selected following this methodology. Then, the clusters have been used to produce MTLF following different algorithm principles:

- A naïve method.
- A random walk process.
- Standard load profiles.
- Different classification methods using a survey describing the time of use of certain actions.
- Different classification methods using a survey using socio-economic data.
- A comparison with a short term load forecast (STLF) method from the state of the art.

The results of the different methods show promising results of different methods to produce forecasts from time series data and from different types of statistical data. Nevertheless, certain assessment have not been finished and yet and cannot be provided:

- Only training errors are provided for the machine learning models. There is a new batch of answers received from an additional survey that will be used to test the generalisation and forecasting ability of the models.
- The results from the STLF experiment are not completed and thus cannot be provided.
- A proper discussion of the resulting models with respect to the state of the art and the potential use both in the project and by different practitioners is needed to be included.

Section 3 presents the intervention level tool. This tool was designed to produce yearly load profiles after an intervention has been carried out. Several interventions have been considered such as:

- Different change of tariff (a Time of Use and a Price Signal),
- Different types of blackouts,
- Economic interventions.

A natural experiment has been carried out to produce relevant data for the change of tariff and results of the experience will be provided in Deliverable D5.2 (energy cooperative use case). On the other hand, surveys have been prepared to assess the other two and models will be provided for each one of them in Deliverable D5.2 (PED use case) and Deliverable D5.3 (ethics assessment of results).

Finally, Section 4 shows the counterfactual level modelling. Here it is provided a discussion on how to use the clusterization method produced in D2.1 as a relevant tool for desegregating the total energy consumption of a year in the same fashion as standard load profiles from TSOs work. The results show that the standardised load profiles are adjusted to aggregations of consumer loads and not individually datasets, thus rendering them inaccurate for the proposed use of them. A proposal to use a different standardised load profile with less error is proposed.

2. The association level model

In this section, we present an analysis of the prediction error of different medium-term load forecast algorithms on residential time series data using the association level model. Medium-term load forecasting involves predicting hourly energy consumption for the next year and is crucial for various applications, including the deployment of time-of-use tariffs and optimising energy efficiency in residential buildings. The most obvious application is the deployment of time-of-use tariffs, even if there is no access to smart meter readings. However, this kind of forecasting is also very useful for planning and sizing purposes, such as ensuring that the electricity supply meets the expected demand and optimising energy efficiency in residential buildings.

The association level model aims to identify statistical relationships between variables to improve the accuracy of the forecasts. We analyse the relationships between energy consumption patterns and other relevant variables, such as socio-economic indicators and time-of-use survey responses. By identifying these relationships, we can improve the accuracy of our medium-term load forecasts.

We use various algorithms and data from a time-of-use survey, which is a questionnaire designed to gather data on various aspects of energy consumption and user behaviour. The survey included questions on the new tariff system implemented in Spain on June 1, 2021, which unified all previous tariff models into a single model with differentiated prices based on time-of-use. It also collected data on energy consumption patterns, consumer categorization, and user decision-making mechanisms related to energy transition.

Our objective is to study the prediction error of different medium-term forecast algorithms on residential time series data. We compare the performance of these algorithms with state-of-the-art models, such as the FFORMA ensemble training method, to identify the most accurate forecasting method. Ultimately, the insights gained from this analysis will help us improve the accuracy of medium-term load forecasts and better understand the impact of time-of-use tariffs on energy consumption patterns and user behaviour.

For this study, we used a dataset consisting of nearly 84000 time series on energy consumption with a duration of at least one year. The data was collected from publicly available sources (see Table 1) and included contextual information on the time series. This dataset has been expanded from the one used in Deliverable D2.1 by incorporating a few other datasets (EDRP, SGSC, SAVE, and NESEMP). In addition, existing long datasets containing at least the full years of 2019, 2020, and 2021 were split into pre-, during, and post-COVID-19 sub-datasets. This was done because changes in electricity consumption patterns caused by the pandemic lockdowns and the return to the new normal can be considered different enough to warrant separate treatment.

Table 1. Characteristics of the datasets used

Name	No. sites	Location	Sampling period	Collection period	TS length
EDRP	13 071	GB	30 min	2008-2010	2 yr
GoiEner pre-COVID	12 123	ES	1 h	2014-2020	1.5 yr
GoiEner during COVID	15 502	ES	1 h	2020-2021	1 yr
GoiEner post-COVID	17 457	ES	1 h	2021-2022	1 yr
ISSDA	6 084	IE	30 min	2009-2010	1.5 yr
Kaggle	1 380	various	1 h	2016	1 yr
Low Carbon London	5 269	GB	30 min	2011-2014	1.5 yr
NEEA pre-COVID	62	US	15 min	2018-2020	1 yr
NEEA during COVID	183	US	15 min	2020-2021	1 yr
NEEA post-COVID	168	US	15 min	2021-2022	1 yr
Elergone Energia	351	PT	15 min	2012-2014	2 yr
SAVE	2 628	GB	15 min	2017-2018	2 yr
SGSC	10 907	AU	30 min	2012-2014	2 yr

Regarding the survey, it aims to collect data on the time-of-use of electricity in various types of residences and the socio-economic factors that impact it. The survey was conducted on 283 GoiEner customers who willingly agreed to participate and share their identification number in order to match their answers with their corresponding load profiles. The questions cover a range of topics such as the type of residence, the primary heating device used, the daily routine of the occupants, the time of year they live in the residence, and demographic information about the inhabitants. Moreover, socio-economic information is collected, including the type of building, ownership of the residence, and the amount of money spent on heating and transportation fuel:

- **Assets and time of use data**
    - Type of residence: this variable indicates the type of residence for which the energy is being used. Five different types of residence have been considered: main residence, secondary residence, small business or similar, common areas of buildings/external lighting and others.
    - Main heating device: single gas boiler, electric boiler, energy storage, electric heating, heat-pump, butane, oil boiler, biomass boiler or stove, central heating, heat network, passive building, do not have/use boiler, unknown or other.
    - Type of electrical devices used in the residence (more than one option allowed): batteries, electric water heater/heat pump, electric kitchen, electric vehicle, others.
    - The inhabitants of residence maintain on weekends the same consumption pattern as on weekdays (yes/no).
    - When the inhabitants of the residence (in minutes from midnight):
        - On weekdays:
            - Have breakfast.
            - Have lunch.
            - Have dinner.
            - Sleep
        - On weekends:
            - Have breakfast.
            - Have lunch.
            - Have dinner.
            - Sleep

- What time of the year do inhabitants live in the residence (multiple options):
  - Autumn.
  - Winter.
  - Spring.
  - Summer.
  - Long holidays.
  - Every or almost every weekend.
- Is there someone at home all day during working days?
  - No, all inhabitants work outside home.
  - Yes, telecommuting.
  - Yes, unemployed or retired.
  - Yes, doing housework

- **Socio Economic data**
  - Type of building: flat, semi-detached, chalet, other.
  - Residence ownership: this variable indicates if the client owns, rents or the dwelling has been transferred.
  - Number of not dependent adults living in the residence.
  - Average age of the non-dependent adults living in the residence.
  - Number of non-dependent women living in the residence.
  - Number of dependent people living in the residence. We consider dependent people those inhabitants who are still studying and are not receiving a salary.
  - Average age of the dependent people living in the residence.
  - Number of dependent women living in the residence.
  - Total surface: less than 80 m^2, between 80 m^2 and 120 m^2, more than 120 m^2 and unknown.
  - Building age: this variable indicates the age of the building among the following four options: pre 1980, between 1980 and 2006, between 2006 and 2019 or after 2019.
  - Energy certificate of the building: an efficient building (A, B or C), inefficient building (D,E,F or G), it does not have an energy certification or the client does not know if the building has an energy certification.
  - Type of municipality
    - Climatic zone of the residence: arid or semir-arid, atlantic, mediterranean, continental, mountain climate or other.
    - Size of the locality to which the residence belongs: city (more than 10,000 inhabitants), rural town or isolated core.
  - Total salary
    - Maximum educational level achieved by one of the members of the family unit: without studies, high school studies, vocational training or university studies.
    - Approximate annual net income of the household: more than 100,000€, between 50,000€ and 100,000€m, between 30,000€ and 50,000€, between 15,000 € and 30,000€ and less than 15,000€.
    - Family unit saving per year: more than 10,000€, between 1,000€ and 10,000€, less than 1,000€ or they have been using savings from previous years.
  - Amount of euros spent annually on heating.
  - Amount of euros spent annually on transportation fuel
  - Amount of euros spent annually on electricity.
  - Level of awareness about climate change (1-20).

In annex A, a summary of the main findings of this survey can be found.

2.1. Methodology

As stated before, the objective is to forecast the hourly electricity usage pattern of a single household for a year. This is typically known in the literature as the medium term load forecasting problem (MTLF). In order to achieve this goal, we have relied on the outcomes of Deliverable D2.1, which presents a step-by-step process for generating 40 typical consumption templates by means of clustering techniques. The process of clustering time-series data required multiple stages, such as data cleaning, feature engineering, and the use of the Self-Organizing Map (SOM) algorithm for cluster analysis. After cleaning the data and selecting relevant time-series data, we designed five sets of features to capture different aspects of the time-series data. The SOM algorithm was identified as the most efficient approach for identifying clusters, resulting in the identification of 40 clusters that were considered representative of the main patterns of electrical consumption that were common across all analysed datasets.

It is necessary to comment on this paragraph that the approach followed for obtaining the primary clusters has undergone some modifications compared to the one presented in Deliverable D2.1 including a reduction in the total number of clusters to 30. These modifications include a *meta-clustering* process that was performed based on the results obtained in Deliverable D2.1. Specifically, for all datasets listed in Table 1, it was generated all potential clusterization between 2 and 30 clusters using SOM[1]. This resulted in a total of 337 heatmaps that were re-clustered using SOM. This step was taken to capture the specificities of each dataset, and the second clustering is intended to group all significant clusters that are present. Validation indicators for this new clustering suggested that 30 or 32 clusters would be the most appropriate number of clusters to select. Taking the Occam's Razor principle[2], 30 clusters were finally selected.

Several different methodologies have been tested in order to perform this activity. The next subsections explain the details of each one:

2.1.1. Naïve method

In this case, we have used the naïve method to predict the consumption pattern by randomly selecting patterns. This means that we have not used any specific criteria or algorithm to select the patterns, but rather selected them randomly without any bias.

2.1.2. Random walk

In this study, we calculate the electric behaviour cluster for each supply point during each time period using only the pre- and in-datasets. The "pre" dataset refers to the period before the lockdowns, while the "in" dataset covers the period between the lockdown and the change of tariff that occurred in Spain on June 1, 2021. The "pst" dataset refers to the period after this change of tariff. This division into three distinct periods has become standard practice in Spain, and is increasingly being used in other contexts as well.

We then forecast that each supply point will follow the same electric behaviour cluster in the next period. Specifically, if a supply point is classified as belonging to electric behaviour cluster 1 during the "in" period, we predict that it will belong to electric behaviour cluster 2 during the "pst" period. We measure the error between our predictions and the actual observed behaviour, which requires

[1] Except for the Kaggle and NEEA datasets, which generated up to 20 clusters due to their smaller size.
[2] https://en.wikipedia.org/wiki/Occam%27s_razor

denormalization of the electric behaviour clusters. For supplies with more than one year of data, additional steps are required to ensure the accuracy of our predictions.

2.1.3. Standard load profiles

The TSOs of several countries (Germany,[3] Ireland[4] and Spain[5] are the ones used in this analysis) provide consumer load profile templates that are used for prediction activities. Typically, these load profiles are used in a counterfactual manner, namely, the actual energy consumption of a period is used to "distribute" it among each hour of the year. However, some exceptions exist where more complex methodologies are used[6]. However, for this particular use of these methodologies, the most correct comparison would be to forecast the actual yearly consumption instead of following a counterfactual approach. Nevertheless, we have decided that it is not worth the effort as the mean error achieved by practitioners using their algorithm to forecast the annual consumption of a residential home seems to be around 3-5% and the differences will be low (but this is something to be interesting to check at future stages). Moreover, we

2.1.4. Using data from the time of use survey

In this approach we only used those features related to assets and time of use data (a list of these features is available in Section 2). All these variables were extracted from the survey performed by GoiEner to its clients. In total, we were able to match 283 surveys with their corresponding load profiles, and their corresponding distance to each of the 30 clusters defined in the project, selecting the closest cluster in terms of MAPE[7].

In order to evaluate if it is possible to infer the closest cluster to each of the clients using only data from the survey (load profiles have not been used as training data), we followed two different approaches: manual and automatic.

In both approaches, we have followed the same strategy to set the ground truth, i.e. which cluster a user (or load profile) belongs to. The Euclidean distance between each cluster, expressed as a vector of 8904 values (24 hours x 7 days x 53 weeks in a year, see Section 5.3 of Deliverable D2.1 "*Cluster Visualization*" for further details), and each one of the load profiles in the dataset, also expressed as a vector of the same length, has been computed. The cluster with the minimum distance is assigned to the user. The way in which each load profile has been compacted into a 8904-value vector, knowing that each load profile has a different length and starts and ends at different dates and times, has been solved by averaging all values with the same date and time regardless of their year by its median. Therefore, this calculated distance has been used as ground truth for classification tasks presented in this section, assigning to each of the clients the closest cluster according to the calculated distance.

The manual approach consists in creating a decision tree after manually analysing which of the features gathered through the survey can help in classifying each of the users to their closest clusters according to their load profiles. In figure 1, a representation of the developed decision tree can be seen. As it can be seen, we have not been able to create a decision tree capable of differentiating the 30 clusters using

[3] Standardlastprofile Strom | BDEW
[4] Standard Load Profiles | RMDS (rmdservice.com)
[5] https://www.ree.es/es/clientes/generador/gestion-medidas-electricas/consulta-perfiles-de-consumo
[6] https://bscdocs.elexon.co.uk/guidance-notes/load-profiles-and-their-use-in-electricity-settlement
[7] https://en.wikipedia.org/wiki/Mean_absolute_percentage_error

data exclusively extracted from the survey, this is why there are some paths in the tree that lead to several cluster numbers.

On the other hand, in the automatic approach we have used four different machine learning algorithms: Logistic Regression, Decision Trees, Random Forests and Support Vector Machines (SVMs).

Regarding the automatic approach, we have opted for four different algorithms: two of them are tree-like-models (Decision tree and Random Forest) as the three we have created manually, a Logistic Regression model and a SVM. In order to evaluate their performance, we have used stratified cross validation with 3 folds (we have not divided the dataset in more folds due to the low number of samples). This is 3-fold cross-validation, which involves splitting the data into three equal parts, using two of the parts to train the model (and calculate the hyperparameters using Grid Search) and the third part to test it. This process is then repeated three times, with each of the three parts used as the test set once. The results from each of the three runs are averaged to get a more accurate estimate of the model's performance. This technique helps to reduce overfitting and provides a more reliable evaluation of the model's ability to generalise to new data. In figures 2, a representation of the decision tree created by the Decision Tree in one of the foldings can be seen. The tree representation of the hundred of estimators created by the Random Forest have not been provided due to space constraints, however each of the estimators created for the Random Forest follow a similar pattern to the one seen in figure 2. In table 2, a subset of the dataframe used for training can be seen.

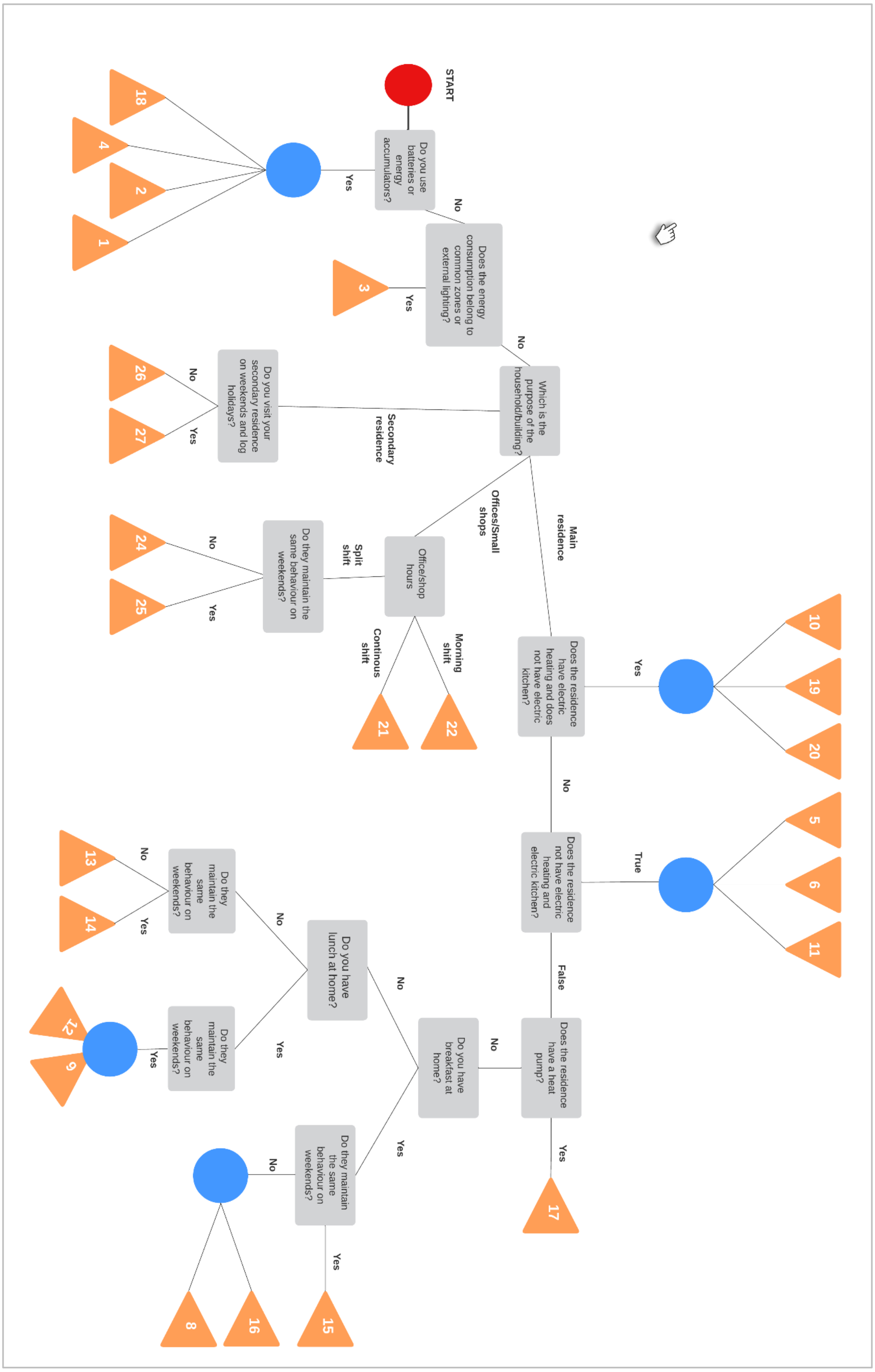

Figure 1. Decision tree for the survey method

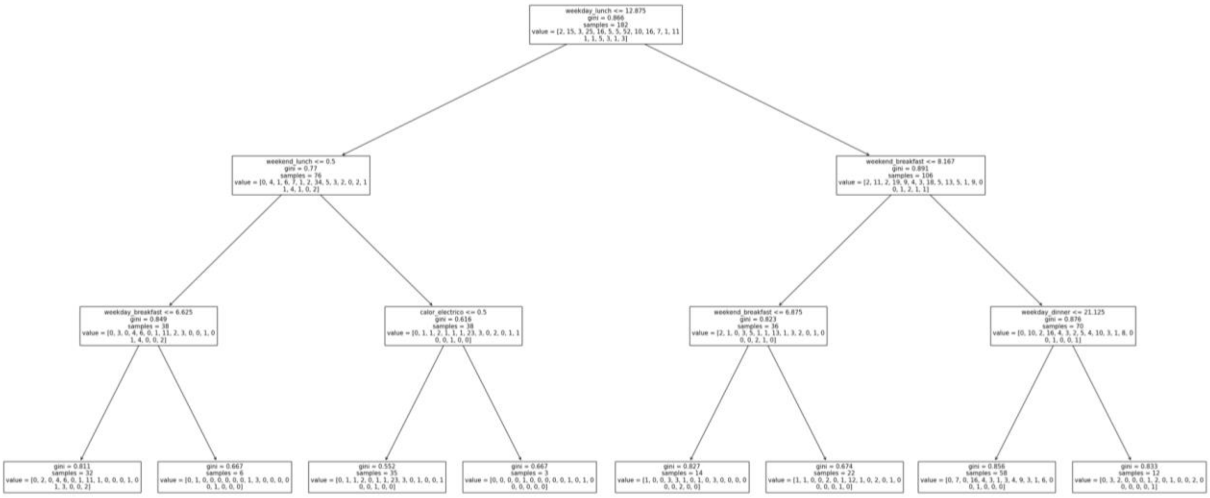

Figure 2. Decision tree created in order to infer the cluster to which a client belongs to using data from the time of use survey

Table 2. Example of a data frame with features extracted from the time of use survey.

main_residence	electric_heat	electric_kitchen	heat_pump	same_pattern_weekends	…	weekday_breakfast	weekday_lunch	weekday_dinner	weekday_sleep	weekend_breakfast
1	0	1	0	0	…	75	130	210	230	100
1	0	1	0	1	…	60	155	85	975	7.67E+15
1	0	1	0	1	…	95	1375	210	2375	105
1	0	1	0	0		0	0	0	0	0
1	0	0	0	0	…	725	0	200	220	95
1	0	1	0	1	…	7.08E+15	145	210	230	80
1	0	0	0	1	…	85	145	215	0	90
1	0	1	0	0		75	140	200	220	90
1	1	1	0	0	…	6.58E+15	130	210	0	9.08E+15
1	0	1	0	1	…	775	140	0	0	90

2.1.5. Using the socio economic information

In this section, instead of using data from the time of use survey, we have focused on the creation of machine learning models using exclusively socio economic information listed in Section 2. In order to be able to compare the performance of the models trained with data from the time of use survey, with the models trained with socio economic data, we have trained the same machine learning model with the same set of folds and hyperparameter optimization process. In figure 3, a representation of the decision tree created by the Decision Tree in one of the foldings can be seen. In table 3, an example of a dataframe with the input features can be seen.

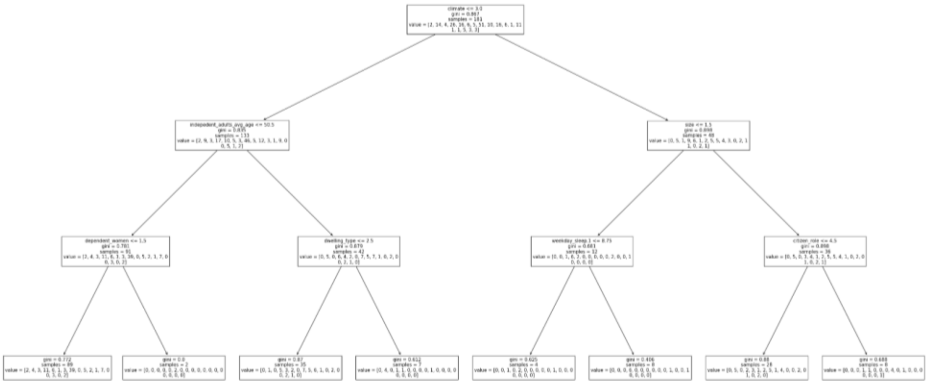

Figure 3. Decision tree created in order to infer the cluster to which a client belongs to using socio economic data

Table 3. Example of a dataframe with socio economic features extracted from the survey

dwelling_type	dwelling_age	certificate	climate	locality_size	…	rent	size	indepedent_adults	indepedent_adults_avg_age	independent_women_adult
1	3	2	4	3	…	2	2	20	450	10
3	4	1	2	2	…	2	2	20	520	10
1	4	4	2	3	…	2	1	10	680	0
1	3	3	4	2		2	2	30	580	20
1	2	4	2	3	…	2	1	10	530	10
1	3	4	2	3	…	2	2	30	400	20
1	4	3	4	3	…	2	1	10	610	0
1	3	3	2	3		2	2	30	500	10
1	4	2	3	3	…	2	1	20	470	10
1	2	3	2	3	…	2	1	30	500	10

2.1.6. Comparison with state-of-the-art models for the short-term load forecast problem

Finally, in order to compare with state of the art models, we implement a FFORMA[8] like methodology to the energy contexts. The methodology works as follows:

Figure 4. presents a graphics representation of the methodology developed for our predictive modelling. This methodology comprises three distinct phases. The first phase involves the training of predictive models, in which we use historical data to develop models that can predict future outcomes with a high degree of accuracy. In the second phase, we train the FFORMA ensemble, which is a combination of the models trained in the previous phase to improve prediction accuracy. This process involves fine-tuning the individual models and combining them in a way that takes advantage of their respective strengths. Finally, in the third phase, we use the FFORMA ensemble to generate forecasts on a testing dataset in order to validate the overall output of this approach.

[8] Montero-Manso, Pablo, et al. "FFORMA: Feature-based forecast model averaging." International Journal of Forecasting 36.1 (2020): 86-92.

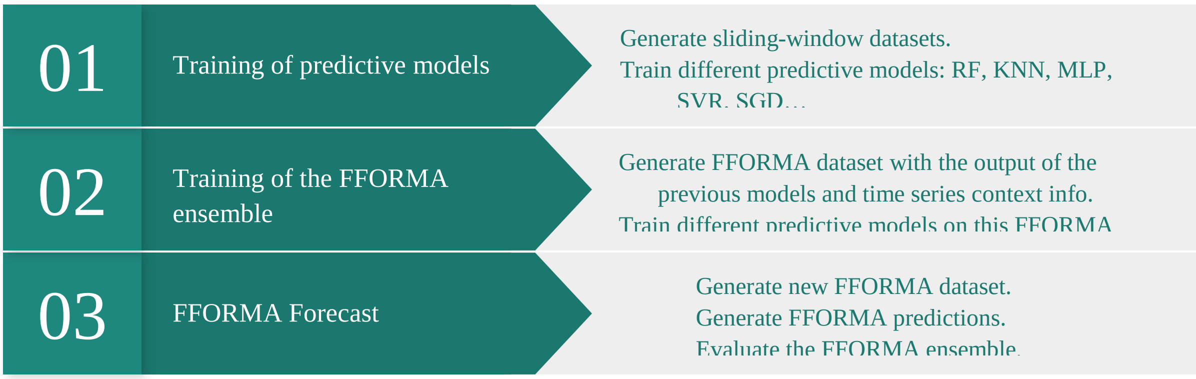

Figure 4. Adaptation of the FFORMA methodology used

2.1.7. Training of predictive models

First, 271 time series are selected from the goi4_pst dataset. The files are loaded in and resampled if necessary (one sampling per hour is done, summing the values when the samples are inferior to one hour). It is preprocessed to generate the required format, using a sliding window (being the window: *N* features + *X* output values. *N* and *X* are parameterizable). For this experiment we use an *N* of 72 (i.e. 72 hours, 3 full days) and an X of 6 (the next 6 hours to predict) are considered.

The resulting dataset will have the following format: a time series column identifying the time series, followed by the *N*+*X* columns of the window (72 + 6). This dataset will be used to train the models. Table 4 contains an extract of the dataset.

Table 4. Time series set dataset

timeseries	col1	col2	col3	col4	col5	...	col73	col74	col75	col76	col77	col78
edrp_16005	0.054	0.064	0.157	0.07	0.152	...	0.088	0.07	0.071	0.07	0.082	0.244
edrp_16005	0.064	0.157	0.07	0.152	0.167	...	0.07	0.071	0.07	0.082	0.244	0.302
edrp_16005	0.157	0.07	0.152	0.167	0.075	...	0.071	0.07	0.082	0.244	0.302	0.093
edrp_16005	0.07	0.152	0.167	0.075	0.158	...	0.07	0.082	0.244	0.302	0.093	0.084
edrp_16005	0.152	0.167	0.075	0.158	0.352	...	0.082	0.244	0.302	0.093	0.084	0.087
						...						
iss_1072	0.654	1.819	1.159	1.702	1.683	...	0.444	0.516	0.43	0.429	0.534	0.457
iss_1072	1.819	1.159	1.702	1.683	1.522	...	0.516	0.43	0.429	0.534	0.457	0.476
iss_1072	1.159	1.702	1.683	1.522	1.255	...	0.43	0.429	0.534	0.457	0.476	0.493
iss_1072	1.702	1.683	1.522	1.255	1.09	...	0.429	0.534	0.457	0.476	0.493	0.435
iss_1072	1.683	1.522	1.255	1.09	1.12	...	0.534	0.457	0.476	0.493	0.435	0.512
iss_1072	1.159	1.702	1.683	1.522	1.255	...	0.43	0.429	0.534	0.457	0.476	0.493
iss_1072	1.702	1.683	1.522	1.255	1.09	...	0.429	0.534	0.457	0.476	0.493	0.435
iss_1072	1.683	1.522	1.255	1.09	1.12	...	0.534	0.457	0.476	0.493	0.435	0.512

From the remaining series and datasets, 100 series are randomly selected and processed following the procedure explained above. The resulting dataset will be used as a test to validate the generated models.

As models we use linear regression, k-nearest neighbour, random forest, multi-layer perceptron regressor, support vector regression and stochastic gradient descent regression. The models are trained, with different combinations of hyper-parameters, using the previous 72 hours as features, seeking to predict the subsequent 6 hours. The error of the distribution (comparing the preceding value with the real value) is calculated using MAPE. Lastly, we obtain the median of the error.

Finally, the best combination of hyperparameters is selected from each model. For this purpose, the error distributions are compared using a Friedman post-hoc test.

2.1.7.1. Training of the FFORMA ensemble

We select each of the models trained in the previous phase with the best combination of hyperparameters obtained. From the rest of the time series, a dataset is obtained with the same processing and format described in the previous phase (resulting in 72 feature values and 6 output values) and the error matrix is validated.

The error matrix is grouped by time series so that the median error is calculated for each algorithm and time series. Finally a new dataframe is generated with the name of the timeseries, the features (which would be the context information) and as many columns as algorithms we have selected, where the previously calculated median error appears. Table 5 contains an extract of the dataset.

Table 5. Input dataset for FFORMA

timeseries	*c1*	*c2*	*c3*	*cN*		*LR*	*KNN*	*DT*	*RF*	*SVR*	*NN*
iss_1006	*X*	*X*	*X*	*…*	*X*	*0.54811*	*0.61777*	*0.32998*	*0.57801*	*0.71003*	*0.40407*
iss_1098	*X*	*X*	*X*	*…*	*X*	*0.83303*	*0.17789*	*0.5375*	*0.65178*	*0.34714*	*0.44878*
iss_1064	*X*	*X*	*X*	*…*	*X*	*0.19695*	*0.65397*	*0.00733*	*0.53527*	*0.48521*	*0.74434*
iss_1049	*X*	*X*	*X*	*…*	*X*	*0.56224*	*0.95642*	*0.82219*	*0.01026*	*0.54687*	*0.91906*
iss_1045	*X*	*X*	*X*	*…*	*X*	*0.57416*	*0.00921*	*0.88677*	*0.58802*	*0.09379*	*0.99203*
iss_1035	*X*	*X*	*X*	*…*	*X*	*0.54934*	*0.00711*	*0.59692*	*0.1981*	*0.60414*	*0.95327*
edrp_1652	*X*	*X*	*X*	*…*	*X*	*0.26332*	*0.50877*	*0.17103*	*0.86269*	*0.503*	*0.53945*
edrp_1639	*X*	*X*	*X*	*…*	*X*	*0.5402*	*0.08202*	*0.13182*	*0.87296*	*0.36103*	*0.52338*
edrp_1613	*X*	*X*	*X*	*…*	*X*	*0.34565*	*0.62834*	*0.62024*	*0.34381*	*0.83402*	*0.63461*
edrp_1611	*X*	*X*	*X*	*…*	*X*	*0.39588*	*0.84569*	*0.52631*	*0.5516*	*0.53319*	*0.43718*
edrp_1600	*X*	*X*	*X*	*…*	*X*	*0.52053*	*0.46178*	*0.68806*	*0.16365*	*0.69213*	*0.40228*
edrp_16010	*X*	*X*	*X*	*…*	*X*	*0.45923*	*0.57891*	*0.47762*	*0.61993*	*0.65552*	*0.38793*

The resulting dataset is divided between train and test. As features would be the context information, while output would be the error of each algorithm. The same algorithms are used as in the previous phase, with different combinations of hyper-parameters, and an error matrix between actual and predicted is obtained. In this way, the best model (with its best combination of hyper-parameters) available is selected.

2.1.7.2. FFORMA Forecast

With a larger set of time series than in the previous section, a FFORMA dataset is generated as described in the previous section. The value predicted by FFORMA is then calculated using the following formula:

$$\frac{(\square\square\square\square\square\square\square\square \cdot \square\square\square\square\square\square\square)}{\sum_{\square}^{\square} \square\square\square\square\square\square\square}$$

That is, with the FFORMA model we predict the error of each of the models for that time series (based on the context information of the timeseries). A matrix is obtained with as many columns as prediction models have been used.

The final prediction, after applying FFORMA, would follow the following formula:

$$\frac{(\square_{\square\square} \cdot \square_{\square\square}) + (\square_{\square\square\square} \cdot \square_{\square\square\square}) + (\square_{\square\square} \cdot \square_{\square\square}) + (\square_{\square\square} \cdot \square_{\square\square}) + (\square_{\square\square\square} \cdot \square_{\square\square\square}) + (\square_{\square\square} \cdot \square_{\square\square})}{\square_{\square\square} + \square_{\square\square\square} + \square_{\square\square} + \square_{\square\square} + \square_{\square\square\square} + \square_{\square\square}}$$

where:

- F: is the value predicted (Forecast) with the corresponding model.
- E: is the error predicted by FFORMA for that model.

Using this algorithm, a matrix of X columns is obtained (in our case, the 6 columns that were being used as output) with the predictions, which will be compared with the real values of the validation dataset and therefore calculate the error, which will validate the result of the FFORMA.

2.2. Results

First, we show the results of the models that could be used over the entire dataset of time series. These models are:

- the naïve method (displayed as `random`),
- the standardised load profiles from the German and Spanish TSO's (displayed below as `de_H0, de_G6 and es_p20td`)[9],
- the use of the clusters as a forecasting tool (displayed as `forecast`),
- the use of the clusters as a counterfactual method both using the upper whisker and the energy consumption (displayed as `min30_u and min30_c` respectively)

Figure 5 shows the distribution of errors using boxplot in terms of MAPE over the entire dataset of these methods. Table 6 provides the most relevant numeric values, namely the first and third quartile and the median. It can be clearly seen that the use of the clusters is far better than the rest of the methods achieving surprisingly strong results with a 46.74 % of MAPE error. Please note that a MAPE of 40 % is considered in the state of the art an average result when producing STLF for similar datasets. When using the clusters as a proper forecasting tool (please note that the clusters need the actual data to produce the clusters, so it is not a valid forecasting tool), this value worsened to 58.30 %. Which is not that far from the state of the art for STLF. Finally, the naïve method is sensibly worsen that the other two but surprisingly better than the best standardised method.

[9] Please note that we have selected just the best proforma ones. For a complete assessment consult Section 4.

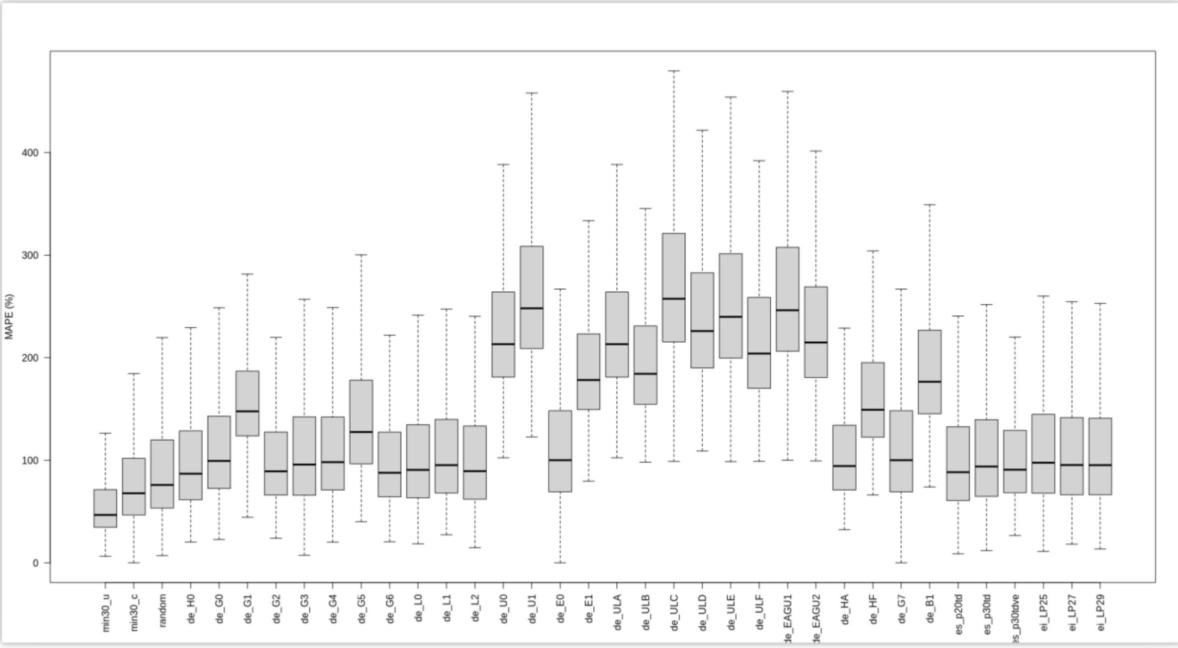

Figure 5. Boxplot MAPE (%) error of MTLF with different methods over the entire dataset

Table 6. MAPE (%) error of MTLF with different methods over the entire dataset

Method	**Q_1**	**Median**	**Q_3**
min30_u	34.73	46.74	71.45
forecast	42.00	58.30	91.10*
min30_c	46.79	67.89	101.91
random	53.50	75.96	119.93
de_H0	61.58	86.93	128.66
de_G6	64.41	87.83	127.49
es_p20td	60.79	88.44	132.75

** Only over Goiener and NEEA datasets*

Next, we provide the results of using the Time of Use and the Socio Economic survey to build models that forecasts the cluster of a household. Please note that in both cases we do not use any information (apart from the clusterization that has been done *a priori*) from the time series but information provided by end users in a survey. Table 7 shows the MAPE error for the training dataset of several machine learning models, including a decision tree manually created using the time of use data (displayed as survey_tou), logistic regression (displayed as lr_tou and lr_se respectively), decision tree (displayed as dt_tou and dt_se respectively), random forest (displayed as rf_tou and rf_se respectively), and SVM (displayed as svm_tou and svm_se respectively). Here the baseline is given by the distance between the load profiles of the user and the nearest clusters (displayed as min30_u and min30_c).

As before, Figure 5 shows the distribution of the training errors using boxplot in terms of MAPE over the answers of the survey. On the other hand, Table 7 provides the most relevant numeric values, namely the first, second and third quartiles for both training and test sets. Please note that the test set was produced with answers to a second iteration of the Time of Use survey carried out a year after the first one. Even as the second survey was answered by more than 700 people, only 44 answers were new (namely, have not provided information to deanonymize the survey and thus was not included in the training set for the models).

The first thing to notice is that all methods work consistently well and with MAPE errors between 50 and 70 %. This is between the best forecasting method and the naïve method as presented before. Special mention has to be provided to the decision tree manually created and both logistic regressions methods as the three methods achieve very good results (MAPE error between 52.09 % achieved by the logistic regression and 59.50 % achieved by the decision tree) and in all cases, the results are interpretable.

Next, we can see that all methods improve significantly their performance under the test set but without changing significantly the conclusions found with the training set. This is not the usual behaviour of forecasting methods, but given that the number of elements in the test set is an order of magnitude less than the training set, this could be just that the population in the test set was highly biassed. Please, take into consideration that due to the nature of the problem assessed, any sample taken in this research is independent and identically distributed.

Finally, it is important to note that Figure 6 includes the results of a Friendman test with Nemenyi post hoc (the letters over the boxplots). Only the results of the training set are provided given the low amount of answers contained in the test set. It could be seen that there are 7 different groups of models (letters a to g) but the most clear groups are the two best ones: the one that only contains the baseline method (`min30_u`) and the one that contains the two logistic regression methods (lr_tou and lr_se). The logistic regression model could be found in Annex B.

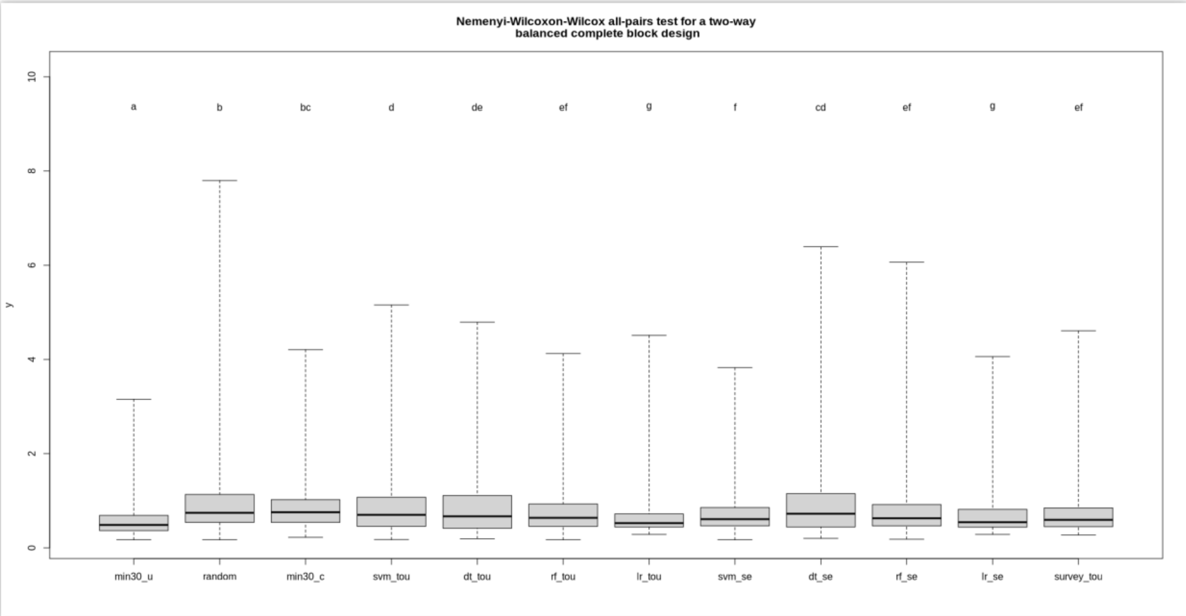

Figure 6. Boxplot MAPE error (given as a fraction of 1) of MTLF with different methods over the answers of the survey

Table 7, MAPE (%) error of MTLF with different methods over the answers of the survey

Method	Train set			Test set		
	Q1	Median	Q3	Q1	Median	Q3
min30_u	36.66	47.91	68.17	30.85	39.85	57.72
lr_tou	43.96	52.09	71.86			
lr_se	43.96	53.75	82.43			
survey_tou	45.08	59.5	84.92	37.61	49.11	77.78
svm_se	46.23	60.65	85.64	40.05	44.58	66.68
rf_se	46.04	62.67	89.7	38.20	51.76	68.30
rf_tou	45.88	64.12	94.32	43.25	52.60	69.44
dt_tou	42.32	68.65	110.78	40.05	45.09	69.97
svm_tou	46.98	70.67	107.25	40.05	44.58	66.68
dt_se	44.82	72.83	114.17	40.73	48.33	68.30
random	53.89	73.99	111.07	46.99	61.05	107.15
min30_c	54.25	75.02	101.98	41.37	64.14	87.36

2.3. Software

The FFORMA predictor is developed in Python and the repository is accessible at the following link:

https://github.com/DeustoTech/why_predictor

In order to run it, it is necessary to install the necessary python dependencies (listed in the requirements.txt file) using pip with the following command:

```
$ pip install -r requirements.txt
```

Before executing the script, it is advisable to generate a config.env configuration file where the different default execution parameters are defined, although these can also be selected (if necessary) with the corresponding command line parameter. For more information, the help of the different parameters can be obtained with the following command:

```
$ python3 -m why_predictor -h
```

Finally, to execute the predictor, the required parameters should be indicated, otherwise, the values defined in the configuration file will be taken. The following command would launch a complete execution (including the three phases described in section 2.1.6) with the parameters of the config.env file:

```
$ python3 -m why_predictor
```

3. The intervention level model

In this section we introduce how we have used the WHY toolkit in order to produce intervention based tools of interest. The main objective is to produce realistic load profiles of a household before and after an intervention. As in the rest of the project, we consider it as "an intervention" to any substantial modification of the initial conditions given by external actions.

Deliverable D1.2 includes a list of interventions considered. The partner of the project first consider interventions on 5 categories:

- Legislative and regulatory instruments
- Economic and fiscal instruments
- Agreement based or cooperative instruments
- Information and communication instruments
- Knowledge and innovation instruments

This list was the base where it was discussed which interventions were going to be relevant for the project. Later, in Deliverable D1.3, each use case manager discussed with the stakeholders to decide which interventions were going to be included in each one of the use cases. The result were that in the:

- **PED use case:** it is going to use two types of blackouts interventions.
- **Energy cooperatives use case:** it is going to use two tariff schemes.
- **Energy communities**: it is going to assess different governance structures.
- **EU and Global**: it is going to use two tariff schemes.

It can be seen that only three interventions were prioritised: blackouts, tariff schemes and governance structures. Given that governance structures do not seem to be relevant to the generation of load profiles (but it is to other relevant topics of the use case like fostering the participation and the management procedures of energy communities), in this section we only provide information about how the other two were implemented.

3.1. Tariff schemes

First, a series of legislative modifications carried out in Spain during 2021[10] and 2022[11] made it possible to perform two large scale natural experiments where the tariff structure of more than 15 000 clients in Goiner cooperative changed from constant energy tariffs to Time of Use tariffs or Price Signals, respectively. This way, we were able to set up pre-post control trials to assess the impact of a Time of Use energy tariff and a Price Signal. A discussion of these interventions can be found in Deliverables D1.3, D5.1 and D5.2.

The results of this intervention can be measured via the load profile of each one of the consumers, which was affected due to the intervention, and results of a survey discussing the impact and the socio-economic description of the household. In order to use this information we have measured the consumption before and after each one of the interventions. In particular, we measure the amount of energy in each one of the three periods of the Time of Use tariff (peak, valley and flat) before the lockdowns and during the tariff scheme. Then a "differences in difference" experimental scheme was

[10] https://www.boe.es/buscar/pdf/2020/BOE-A-2020-1066-consolidado.pdf,
https://www.boe.es/eli/es/res/2021/03/18/(3)/dof/spa/pdf
https://www.boe.es/eli/es/rd/2021/03/09/148/dof/spa/pdf
https://www.boe.es/boe/dias/2021/04/22/pdfs/BOE-A-2021-6390.pdf

[11] https://www.boe.es/boe/dias/2022/05/14/pdfs/BOE-A-2022-7843.pdf

carried out and a model relating the reduction of energy consumption in each of the periods was related to the socioeconomic description of the person. Please note that we have used the same variables and information described in Section 2.1.5. For the Price Signal, we have followed the same approach but instead of "fixed" time periods, each day has between one and three time periods (green, yellow and red) depending on the prices that the energy has each day. Please note that: a) the amount of periods; b) the amount of time each period is present each day and; c) the time each period is present changes every day. In fact, it is not known until 19:00 of the previous day. The results of this assessment are provided in Deliverable D5.2 as part of the Energy Cooperative use case.

Nevertheless, for the EU and Global use case, the model will not be used as we have real data from the three scenarios: the business as usual, the implementation of a time of use tariff, and the implementation of a price signal. A mean electrical profile with all the data will be constructed for each of the three scenarios. Moreover, from all the components that HiSim could simulate, the following configuration was set:

- No distributed generation (Photovoltaic systems) and no battery storage.
- No demand response controls or scheduling loads.
- No electric vehicles.
- For space heating and water heating different options will be tested:
    - Oil
    - Natural Gas
    - Heat Pump[12]
    - District heating

The load consumption profile and the component configuration is then plugged into HiSim and is being used to simulate the behaviour of all standard buildings from the TABULA database[13]. Table 8 includes the translation from the type of building set in TABULA and the one used in PRIMES / TIAM-ECM. Namely, for each TABULA entry we are simulating its three different insulation configurations times the four space heating and water heating options times the three different tariff structures. In total, around 36 000 buildings are simulated.

Table 8. Relation between the TABULA categories and PRIMES / TIAM-ECM ones.

	PRIMES / TIAM	TABULA
Type building/city	Single-Urban	Terraced House
	Single-Rural	Single-Family House
	Multi-Urban	Apartment Block
	Multi-rural	Multi-Family House
Socio-economic	Low	Stage 1
	Medium	Stage 2
	High	Stage 3
Year of construction	1920	Construction Year Class
	1940	Construction Year Class

[12] Working in continuous mode, not following a price signal.

[13] https://webtool.building-typology.eu/#bm

	PRIMES / TIAM	TABULA
	1960	Construction Year Class
	1980	Construction Year Class
	2000	Construction Year Class
	2005	Construction Year Class
	2010	Construction Year Class
	2015	Construction Year Class
	2020	Construction Year Class

3.2. Blackouts

On the other hand, for the blackout it is obvious that the produced load profile would be a constant 0. Moreover, this does not mind that the behaviour of the people was not relevant in this aspect. In particular, the objective of the use case is to size several components that allow it to provide a certain "level of service" during two different types of blackouts. The discussion now is what energy services need to be prioritised in each one of the scenarios / interventions.

To this end, a large-scale survey was conducted. The relevant question were:

> *Assume that the supply of electricity suffers frequent but short interruptions in your neighbourhood due to a bad grid infrastructure. This way, **only a small number of houses are affected and the general communication and service infrastructure is still working**. The utility provider has a partial solution to temporarily provide a limited local electricity supply that cannot support all the normal loads. So they are asking which services you would prefer in order to estimate the resulting loads. Please, rank the following energy services from 0 to 10 stars where 0 stars means it is extremely low priority for you and 10 stars is absolutely needed for you:*
>
> *Assume that full black-outs could occur in your region. This means **the entire region** (and possibly even beyond) **is without electricity supply for at least a day or two**. Please note that in this situation **several services are not working or working** only in a very limited way (like cellular network, internet, television, etc.). The utility provider has a partial solution to provide limited local electricity supply but cannot support the electricity supply all the normal loads so they are asking which loads they should prioritise. Please, rank the following energy services from 0 to 10 stars where 0 stars means it is extremely low priority for you and 10 stars is absolutely needed for you:*

In both cases, a list of 18 energy services where provided:

- *Ensure that drinking water is available in my home.*
- *Allow me to cook.*
- *Allow me to commute long distances.*
- *Allow me to communicate with my family and peers.*
- *Allow me to clean the house.*
- *Allow me to heat my house.*
- *Allow me to cool my house.*

- *Allow me to move goods.*
- *Allow me to work at home.*
- *Provide me and the family with entertainment.*
- *Allow me to light up my house.*
- *Allow me to operate a Home Energy Management system.*
- *Allow me to operate the washing machine, dishwasher, etc.*
- *Allow me to keep the food refrigerated.*
- *Allow me to generate hot water for the appliances.*
- *Allow me to generate hot water for showering and cleaning.*
- *Allow me to operate home security systems.*
- *Allow me to operate my smart home devices (automatic window openers, etc.)*
- *Allow me to operate ventilation.*

This prioritisation is going to be used in the PED use case. More information could be found in Deliverable D5.1 and the results could be consulted in Deliverable D5.2 under the PED use case.

3.3. Energy communities

An "energy community" encompasses various contexts, notably in energy systems, where it involves citizens and select businesses engaging in energy activities like generation, consumption, and sharing. It's often defined as a localised area, like a neighbourhood or microgrid, where energy production closely matches consumption, allowing for energy sharing, storage, and connections to external grids and markets[14,15].

Citizen participation through community energy models is key to accelerating clean energy transition[16]. The governance structure of an energy community addresses human behaviour and people engagement through various means, including: Efforts toward achieving a low-carbon energy transition require a bottom-up governance approach, acknowledging the importance of policy changes and household involvement at various levels and scales[17].

The governance of energy communities focuses on the rules and organisational frameworks that support community involvement in energy activities like generation, consumption, and sharing. This governance includes legal and regulatory elements, decision-making procedures, and the participation of stakeholders such as residents, local authorities, and businesses. Understanding the most common archetypes within a population and tailoring energy policies to these prevalent archetypes can enhance engagement and expedite the desired transformation brought about by changes in energy policy.

The objective of this section is to present how to use the investment profiles defined in Deliverable D2.3 to foster the participation of an energy community. First, it outlines the approach to formulating an intervention strategy to promote the energy transition in EU households (section 3.3.1). Second, a six-step heuristic specifically designed to develop effective interventions is introduced (section 3.3.2).

[14] Gruber, L., Bachhiesl, U., & Wogrin, S. (2021). The current state of research on energy communities. Elektrotechnik Und Informationstechnik, 138(8), 515–524. https://doi.org/10.1007/S00502-021-00943-9/FIGURES/6

[15] Project-Consortium. (n.d.). What are Energy Communities? Retrieved January 10, 2024, from https://uia-initiative.eu/en/news/what-are-energy-communities

[16] Vida Rozite, Matthieu Prin, Silvia Laera, Josh Oxby, & Alexandre Roussel. (2023, August 9). Empowering people – the role of local energy communities in clean energy transitions – Analysis - IEA. International Energy Agency. https://www.iea.org/commentaries/empowering-people-the-role-of-local-energy-communities-in-clean-energy-transitions

[17] Sohre, A., & Schubert, I. (2022). The how and what of bottom-up governance to change household energy consumption behaviour. Energy Research & Social Science, 89, 102570. https://doi.org/10.1016/J.ERSS.2022.102570

This heuristic is a structured guide for developing and implementing practical measures to advance the energy transition at the household level. This heuristic ensures that interventions are impactful and tailored to the people profile in the target audience.

3.3.1. Defining an intervention strategy to foster the energy transition in EU households.

Based on the five categories of interventions outlined in Deliverable D1.2, Table 9 provides an overview of the advantages and disadvantages associated with each category of policy-based interventions.

Table 9. Advantages and disadvantages of the five intervention categories

Intervention categories	Advantages	Disadvantages
Legislative and regulatory instruments	(1) Actors are forced to comply whether they want to or not, (2) all actors are affected equally, and (3) they improve the predictability of the public authority's actions.	(1) Costs to comply are paid by the involved actors, (2) limited coping abilities in complex dynamic situations and (3) change is not coming from the actors themselves but they are forced to comply.
Economic and fiscal instruments	(1) market failures can be corrected and (2) a functioning market could theoretically be created.	(1) in some cases they can generate additional costs due to the subsidies, (2) they cause additional efforts regarding the administrative aspects of loans and taxes.
Agreement based or cooperative instruments	(1) it relies solely on the voluntary contribution of individual actors, which indicates their motivation to comply.	(1) actors outside the network might not have an incentive to cooperate or might even oppose them, (2) not complying will very likely not have any consequences if no sanctions are agreed upon.
Information and communication instruments	(1) the potential to provide the information to a large group of users.	(1) that information does not necessarily lead to compliance and (2) that it will be quite challenging to reach actors, which are not interested in the objectives.
Knowledge and innovation instruments	(1) it allows the involved actors to contribute which lowers the resistance, furthermore it is flexible enough to cope with complex situations, (2) involved actors will very likely be highly motivated to participate.	(1) the group of involved actors might be limited to a small group of frontrunners, (2) requires a high degree of cooperation and identification with the cause of all involved actors.

The willingness of individuals to engage in energy policy changes may be affected by the alignment of these advantages with the interests, goals and values of different archetypes. The willingness of individuals to engage in energy policy changes may be affected by the alignment of these benefits with the interests, goals and values of different archetypes. For example, legislative and regulatory instruments because they tend to impose collaboration through enforced compliance; but not all types of people (archetypes) agree with them.

Table 10 includes the archetypes and the persuasive principles[18], extracted from the authors' previous research titled *Socio-Economic Effect on ICT-Based Persuasive Interventions Towards Energy*

[18] Casado-Mansilla, D., Tsolakis, A. C., Borges, C. E., Kamara-Esteban, O., Krinidis, S., Avila, J. M., Tzovaras, D., & López-De-Ipiña, D. (2020). Socio-Economic Effect on ICT-Based Persuasive Interventions Towards Energy Efficiency in Tertiary Buildings. Energies 2020, Vol. 13, Page 1700, 13(7), 1700. https://doi.org/10.3390/EN13071700

Efficiency in Tertiary Buildings[19], which characterise these archetypes. Additionally, the table presents the strategies for obtaining their collaboration and commitment.

Table 10. Archetypes, Persuasive Principles, and Strategies for Engagement.

Archetypes	Persuasive Principle	Strategy
Early Adopter	Authority, Suggestion	Use technology and environmental experts to recommend new energy technologies. Offer personalised suggestions on the latest innovations.
The Uninterested	Social Proof, Social Recognition	Show how peers and the community are adopting energy-saving technologies. Publicly recognize those who adapt to new standards.
The Homo Economicus	Cause and Effect, Self-Monitoring	Provide detailed analyses of the cost savings and benefits of energy-saving technologies. Tools to monitor personal savings.
The Fearful	Safety, Cause and Effect	Emphasise how energy-saving technologies can provide long-term financial stability and security.
The Stubborn	Reciprocity, Social Recognition	Highlight how their actions contribute to environmental well-being and recognize their commitment to the environment.
The Influencer	Social Proof, Suggestion	Create social media trends about energy-saving technologies and suggest that they promote them in their circles.
The Careful	Authority, Self-monitoring	Provide detailed, expert-validated information. Tools for them to monitor the benefits and safety of technologies.
The Activist	Reciprocity, Social Recognition	Provide platforms for them to share their initiatives and publicly recognize their efforts in promoting sustainability.

An important aspect in defining intervention strategies to promote the energy transition is to identify which archetypes are most prominent in the target audience. Annex C includes a description of the eight archetypes based on their values and expectations.

Finally, investments in energy transition typically do not adhere to a simple yes/no decision model; instead, they tend to follow a staged approach. This approach frames the behavioural process by focusing on the stages of change over time (Transtheoretical Model[20]). Consequently, the 'Stages of Change' are incorporated as a comprehensive dimension within the model, where the various phases are identified as distinct elements or categories of this dimension.

3.3.2. Intervention Strategy Steps

To stay current with potential strategies and policy-based interventions, it is essential to follow these steps to develop effective Intervention Strategies.

1. **Identify the target population**: Use surveys and socio-economic data to identify the dominant archetypes in your target population.

[19] Casado-Mansilla, D., Tsolakis, A. C., Borges, C. E., Kamara-Esteban, O., Krinidis, S., Avila, J. M., Tzovaras, D., & López-De-Ipiña, D. (2020). Socio-Economic Effect on ICT-Based Persuasive Interventions Towards Energy Efficiency in Tertiary Buildings. Energies 2020, Vol. 13, Page 1700, 13(7), 1700. https://doi.org/10.3390/EN13071700

[20] Wayne W. LaMorte. (2022, November 3). The Transtheoretical Model (Stages of Change). Boston University School of Public Health. https://sphweb.bumc.bu.edu/otlt/MPH-Modules/SB/BehavioralChangeTheories/BehavioralChangeTheories6.html

2. **Analyse the current stage of change**: Understand what stage of decision-making they are in. This will help in crafting a more personalised intervention approach.
3. **Choose the right policy-based intervention**: Based on the identified archetypes, select the policy interventions that would be the most effective.
4. **Implement & Communicate**: Roll out the interventions and ensure consistent and clear communication. For instance, if applying economic incentives, make sure all stakeholders are aware of the potential benefits.
5. **Monitor & Iterate**: Post-implementation, monitor the effectiveness of the interventions. If the desired behavioral change is not observed, iterate the strategy by possibly combining multiple interventions or tweaking the existing ones.
6. **Engage with the community**: Especially with influencers and activists, as they can amplify the message and bring about a more significant societal shift.

Table 11 shows the categories of policy-based interventions and matches them with archetypes most likely to adopt each intervention type.

Table 11. Match between policy-based interventions and Archetypes

Policy-Based Interventions	Most Likely to Adopt
Legislative and Regulatory Instruments	– Early Adopter (Values legal compliance, environmental consciousness) – The Homo Economicus (Prioritises legal compliance) – The Careful (Prioritises legal aspects) – The Activist (Aware of regulations and norms) – The Uninterested (Conforms to legal requirements and standards)
Economic and Fiscal/Tax Instruments	– The Homo Economicus (Detail-oriented, assesses costs and benefits) – The Fearful (Concerned about financial matters)
Agreement-Based or Cooperative Instruments	– The Influencer (Values social capital and agreement amongst peers) – The Stubborn (Passionate about environmental causes) – The Activist (Aligns with values and beliefs, believes in the responsible use of technology)
Information and Communication Instruments	– Early Adopter (Keeps up with trends) – The Influencer (Values social capital and popularity) – The Activist (Follows eco-trends, values open knowledge) – The Uninterested (May be reached if information becomes mainstream)
Knowledge and Innovation Instruments	– Early Adopter (Draws from knowledge to make decisions) – The Stubborn (Ambition depends on their competence) – The Activist (Values open knowledge, sharing economy)

The WHY project toolkit offers a comprehensive vision that extends beyond just offering genera strategies[21,22,23]. It enables the tailored design of energy policy changes to suit the prevalent archetypes within the target audience. The WHY toolkit delivers detailed insights for each archetype, including the values that drive their willingness to change, the factors they consider a priority when making decisions, and the level of influence they exert or receive from their surroundings, as well as the behaviours they are likely to exhibit as a result. This information facilitates a quantitative estimation of the impact of policy interventions.

For instance, if "I am a policy maker and interested in how to implement efficient Nudges[24]" and going to "I am interested in different types of energy consumers and how to nudge them", results shows:

> *"People differ in their energy using profiles (including their motivations to use or save energy) and therefore require different approaches to nudge them towards energy efficiency.*
>
> *People's motivation to change behaviour depends on the following six factors:*
>
> - *Perceived behavioural control,*
> - *subjective norm,*
> - *attitude,*
> - *personal moral norms,*
> - *willingness,*
> - *age.*
>
> *Policy design shall make an assessment of the impact of policies on the various types and use a balanced policy mix (including various different types of nudging interventions) to successfully nudge people towards an efficient energy usage."*

Beyond a general strategy, the WHY toolkit, through its strategy design steps (section 3.3.2) allows to (1) Identify the most prominent archetypes within the target population and tailor a set of strategies (Table 10) based on the persuasive principles that resonate best with each archetype; (2) determine the stage of decision-making each archetype is currently in to facilitate the development of a more personalised intervention approach; (3) Select the most effective policy interventions (Table 9) to achieve desired outcomes; (4) ensure consistent and clear communication when implementing the interventions; (5) monitor the effectiveness of the interventions and make adjustments as needed; (6) Engage with the community by fully collaborating with the most receptive archetypes to serve as agents of change for the less engaged archetypes.

> *As a policymaker interested in implementing efficient nudges for different types of energy consumers, understanding your target audience is crucial. If the prominent archetypes within your audience are the Early Adopter and the Stubborn, interventions categorised under Knowledge and Innovation Instruments (table 11) are likely to be successful. A mixed-strategy approach could include:*
>
> - *Utilising technology and environmental experts to endorse new energy technologies.*

[21] Project-Consortium. (n.d.). *I am a policymaker and interested in how to implement efficient nudges • Nudge*. NUDGE Website. Retrieved January 10, 2024, from https://www.nudgeproject.eu/policymakers/

[22] Project-Consortium. (2022). Applying behavioural insights to energy policy - A toolkit for practitioners. Behavioural Insight Toolkit. https://bitoolkit.userstcp.org/?_ga=2.58082026.1722329258.1704819046-1850446277.1703012207&_gl=1*2ue08l*_ga*MTg1MDQ0NjI3Ny4xNzAzMDEyMjA3*_ga_W59Y7X0807*MTcwNDgxOTA0Ni4xLjAuMTcwNDgxOTA0Ni4wLjAuMA

[23] Project-Consortium. (2022). Applying behavioural insights to demand-side energy policies and programmes - Checklist browser. Behavioural Insight Toolkit. https://www.bitoolkit.userstcp.org/checklist.html

[24] https://www.nudgeproject.eu/policymakers/

- *Providing personalised recommendations on the latest innovations.*
- *Emphasising the environmental benefits of their actions.*
- *Acknowledging their contribution to environmental well-being.*

Such strategies can effectively engage these archetypes and encourage the desired behavioural changes.

3.4. Comparison between the EU and LATAM investment archetypes

A review of the most relevant results in the search for implications and possible measures to be implemented sheds light on such significant elements as the existence/absence of representation of investor profiles between the EU and LATAM samples, specifically profiles related to processes of Trust, Security and Autonomy. In this sense, it has been observed that the determinants to justify the absence of these profiles in the case of LATAM countries are deeply related to structural processes of insecurity and legal inequalities, structural injustice related to a lack of definition of frameworks, and randomness in their application, which are not applied in the same way to each of the groups, as well as language barriers, identification, disaffection, levels of poverty and social and geographical diversity.

Archetypes were also discussed, for example the early adopter and the uninterested, and legislation is proposed for these cases. The consequences of remaining in the archetype and the consequences of there being a large percentage of people in these archetypes are observed.

It is important to highlight in this case the possible presence of socio-cultural imaginaries and their evolution, such as shared imaginaries, but also the cultural influence of capitalist models of success, which is very influential in LATAM. The different approaches to quality of life, associated with the transition processes, show a certain consensus around the aspirational model and the model of success that continues to have to do with the capitalist model and the existence of possible generational biases in the data. Another element related to this approach is the evaluation of joint decision-making processes where status parameters appear to be related to the religious-cultural roots of these societies, which is very clear, but which also contrasts with the diversity of ways of thinking and non-linear value matrices that can be observed due to the imprint of the original peoples. There is therefore a great polarisation in the concept of status, so that in the EU, status means contributing to the common good, while in LATAM it is a more individualistic version.

The absence of processes of denialism, not specifically present in the case of energy communities, was also very noteworthy, although not specifically present through its determinants such as mistrust of information sources, institutions, etc.

Regarding the policy proposals, drawing on the model of Directives, such as an EU in which recommendations made based on common objectives where countries then legislate, was considered. Moreover, the necessity of constructing knowledge of the context to make relevant policy recommendations was considered and generate public policies together with social partners. In this sense, also, questioning the type of transitions was considered in terms of a Faster or slower transition, transition that is technically faster but does not affect the creation of a fairer society, or a slower transition that focuses on eliminating structural injustice.

Furthermore, processes of interdependence and globalisation along with the consideration of human rights were considered as part of the issues to take into account in policy proposals. The question of how to include these drivers in the narratives with elements such as justice and accountability was dominant in the experts' reflections. For this purpose, three priorities were considered in terms of design transition processes where consensus horizons were reached with the social partners, generating social inertia in terms of getting a certain group of people to settle into an archetype and generating virtuous/pernicious dynamics and reviewing resistance processes.

In addition, the approach to consider the relations between interdependence and decision-making was considered in terms of arranging a breakdown relations/structures power balance, assuming two options in terms of **Generation of alliances** to alleviate this irregular power balance in terms of **Generation/management of demand** when is necessary for example in the case of vulnerability situations.

This generation/management of demand as an economic analogy involves understanding what your customers/concerned needs are and what steps you need to take to meet that demand. Applied to these eventual vulnerability situations, it can be recommended to predict and plan upcoming requests and make sure that the resources needed to meet them are available. In this sense, the use of historical data and current trends to predict, manage and plan this demand can be considered.

3.5. Investment decisions

Finally, other interventions could be simulated with the system and the data collected. In particular, in the same survey we collect information about the costs of different technologies and the barriers associated with it. The formulation were as follow:

- *How much do you think a "technology" costs for personal use?*
- *How much do you think an "energy efficient technology" costs for personal use?*[25]
- *If your "technology" suddenly brakes tomorrow, will you buy the energy efficient option?*

If the person being surveyed answer not to the previous question, then the following questions were raised:

- *Why not? (please select all that apply to you)*
    - *a. I do not have enough money for an energy efficient one.*
    - *b. I do not think buying a more expensive option will solve any problem with the environment or my pocket.*
    - *c. I do not think it will be cost-effective to me.*
    - *d. I do not trust that energy efficient means high-end, therefore I would rather buy the cheaper option.*
    - *e. I do not want to invest money in energy efficiency.*
    - *f. My peers do not use this kind of appliance.*
- *How much does it have to cost the energy efficient option for you to buy it? (Please answer 0 if you never are willing to pay any extra cost for energy efficiency).*
- *"Energy efficient technologies" allow you to save substantial energy (even more than XXX€ per year in some cases). This means that in a couple of years the efficient option might in some cases pay its extra costs only with the energy saved. In how many years does the efficient option need to pay back its extra cost for you to invest in it?*

[25] In some cases, this was the only question raised as no "alternative" was clear (for example with solar photovoltaic or the energy storage systems).

The “technologies” included in the questionnaire were:

- *A refrigerator*
- *Insulation of the façade*
- *A solar photovoltaic system*
- *An energy storage system*
- *A heat pump (compared with a gas boiler)*
- *Mobility (comparing an electric with a gas one but also including other questions like the commute distance or the use of public transport or other transport means)*

This survey could allow us to build causal relationships about interventions that modify the price (or the ROI) of the different technologies and renovation rates / investment decisions. Moreover, it also provides a list of barriers that need to be overcome above the price. Given that any use case considered these interventions as relevant, it was decided that the assessment of these questions will be provided in Deliverable D5.3 along with the ethics assessment of its impact and policy recommendations. In any case, the RAW data will be published with the rest of the questionnaire when prepared.

To develop the main ideas of the technologies, the responses obtained from the surveys are presented to a panel of experts. As a starting point for presenting the technological cases and their implication on the respondents, the results obtained with the prioritization and investment decision in utopian scenarios related to shared appliances, passive buildings, local flexibility, and 15’ cities are presented (see figure 7).

Among these scenarios, "15’ cities" stood out as the preferred option for a large majority of the respondents. This preference was partly attributed to the relevance of this option for urban residents, who experienced a change in work and mobility dynamics due to telework during the COVID-19 pandemic.

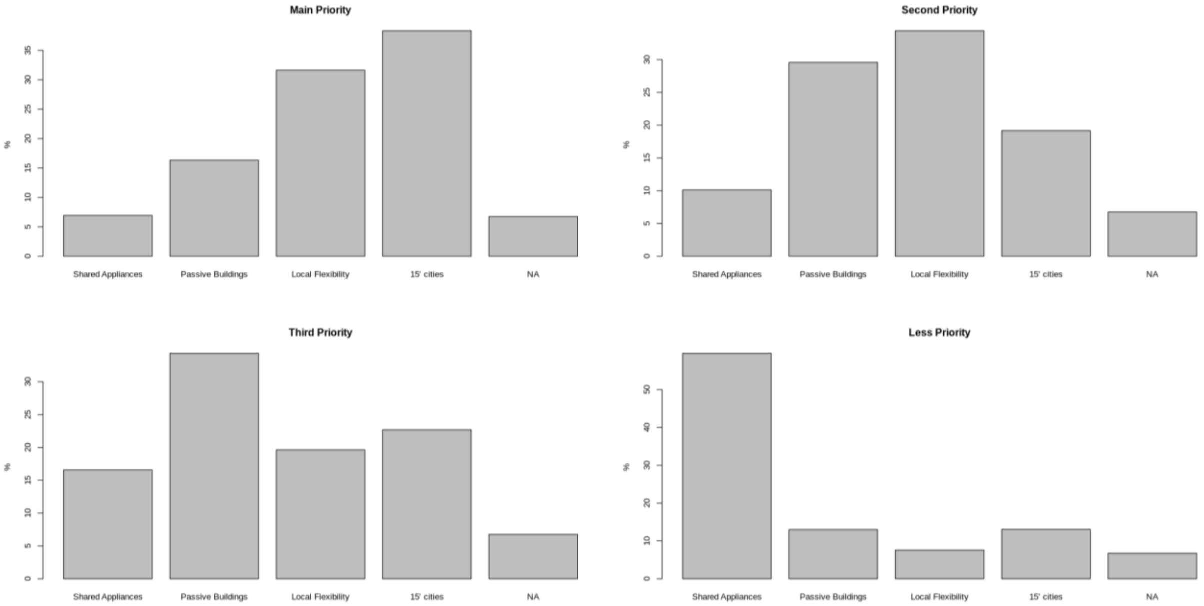

Figure 7. Rank in order of priority

The complexity of the descriptive texts of the scenarios presented in the survey was extensively discussed. Despite the richness of information provided, it was concluded that this complexity did not prevent people from making coherent decisions. In fact, it was noted that there was a clear preference for some options, such as "15’ cities," and a low preference for “shared appliances,” suggesting that the respondents could discern between the options presented.

Regarding the distribution of money among the options (see figure 8), it was found that, while the distribution was relatively uniform across all options (the four to choose from plus the opportunity to save), there was a notable trend towards saving. This could be attributed to the economic uncertainty generated by the COVID-19 pandemic, leading to greater caution in investments. It was also noteworthy that the impact of the respondents' preferences was not reflected in the investment made.

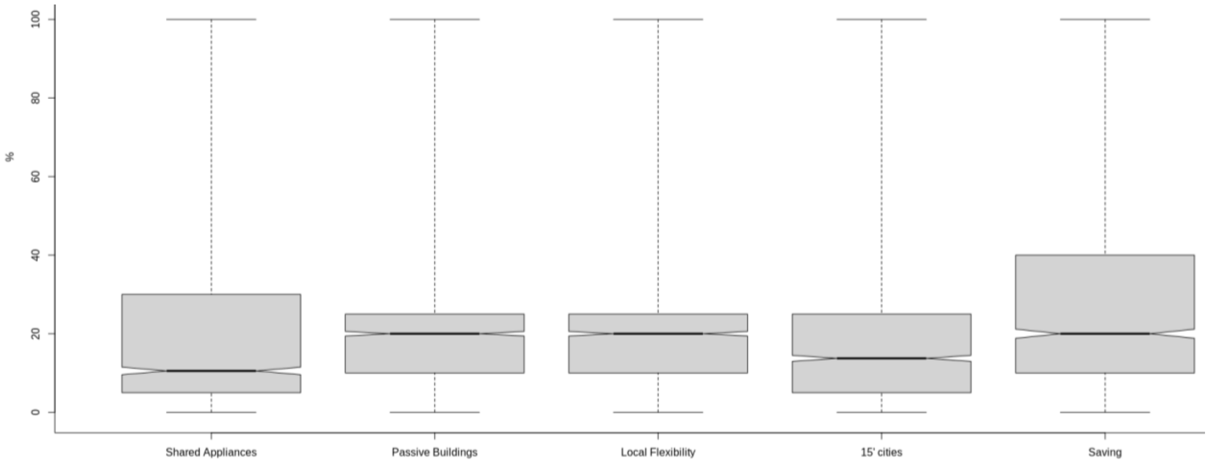

Figure 8. Money (%) willing to be spent on each of the scenarios

The session also addressed regional and socioeconomic differences in preferences. It was highlighted that megacities in Latin America, with their high population density, could explain the greater preference for "15’ cities" in this region. Additionally, how cultural differences and perceptions might influence the willingness to change was explored, especially in the case of electric vehicles, which face resistance due to the cultural association with traditional driving sound and experience. Also, as it is depicted in figure 9, it is found a direct relationship between easiness on getting credit for investment and the amount of invested money.

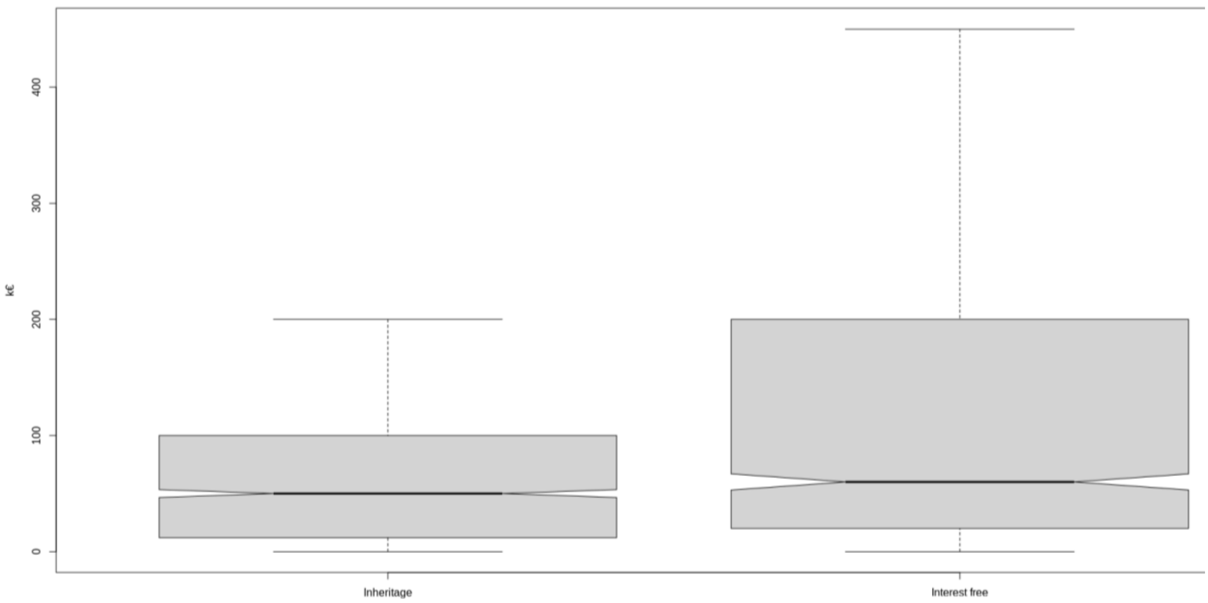

Figure 9. Money (%) willing to be spent on each of the scenarios

After this initial section, we proceeded to analyse the technologies. As is shown in the below Figure 10, there are a set of barriers for engaging people into investment in energy transition to support the current energy transition. The different barriers for each technology, such as credit, trust, cost-effectiveness, high-end, no interested peers and others (NA) are represented. The barrier represented by the red segment, 'Credit' is the most significant barrier across all technologies. Cost-effective and Trust rank second and third, respectively.

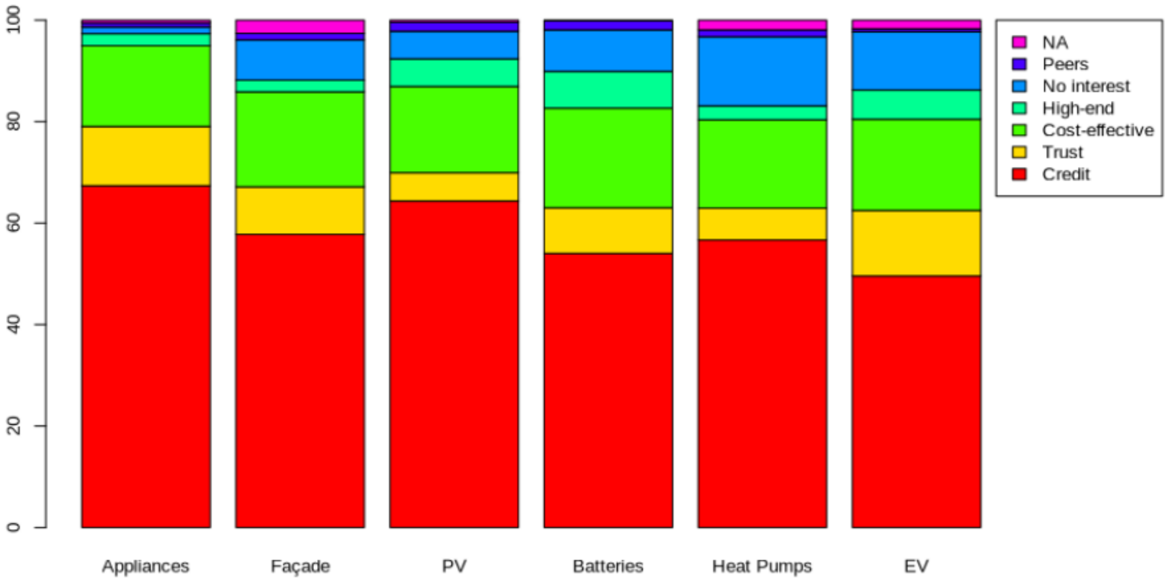

Figure 10. Main barriers across technologies

This observation highlights the importance of the economic factor with respect to the respondents' intention to make any change. These same data and the rest of the factors obtained from the survey were interpreted by a group of experts who sought justifications for some of the most noteworthy results.

Considering the original opinion of the experts, compared to the results shown in Fig. 8, it was surprising that in the facade, photovoltaic panels and batteries the percentage of people willing to install was 50%, as they expected a much lower number (especially in the facade). Also surprising was the interest shown by the respondents in heat pumps, which will be discussed later on. The low presence of electric vehicles was expected.

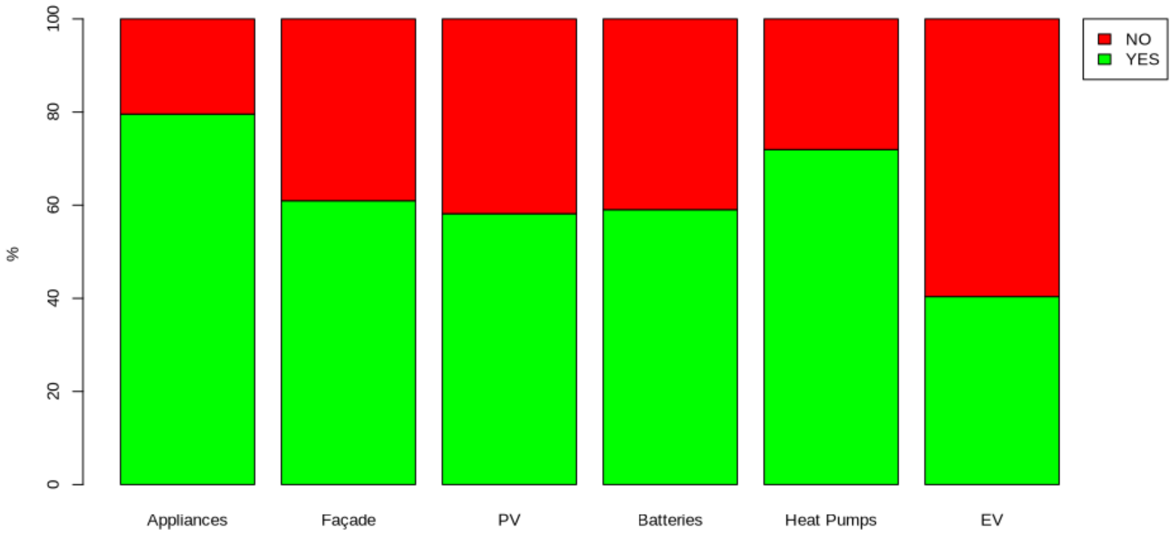

Figure 11. Willingness to buy at current prices

Likewise, the Payback-expected by the experts was of a much more varied range and the homogeneity of the answers was surprising, with a clear generalised trend at 5 years, as shown in Fig. 9.

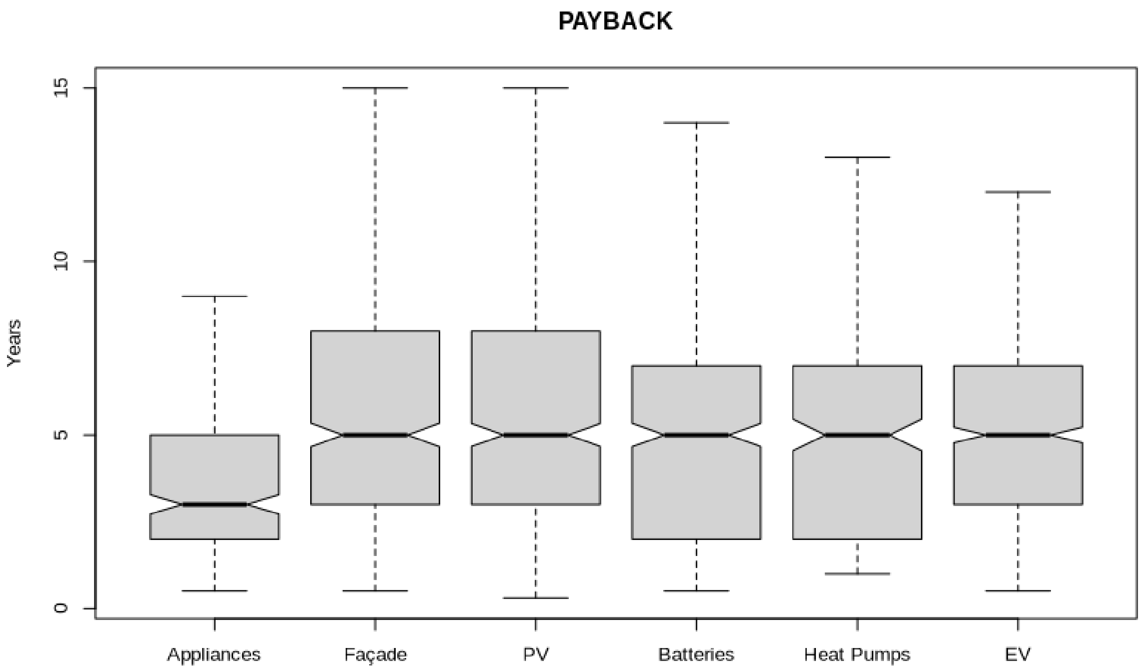

Figure 12. Payback times per technology

A crucial aspect was the willingness to pay more for energy-efficient products. Most respondents would be willing to pay 30-60% more for products such as refrigerators probably because they are familiar with the technology and understand that it will help them to save money. On the other hand, consumers are expecting a reduction of around 60% over current prices for technologies like facades, solar panels and batteries (fig. 11). This represents a paradox, as they are willing to retrieve a ROI in 5 years which is feasible in some of these technologies at current prices. There are two potential explanations for this behaviour. The first one is the lack of culture of assessing the cost over the lifetime of the installation. The second one is the lack of access to credit to overcome the large investment cost that these infrastructures present. Nevertheless, at the moment there are a large collection of financial instruments to support these cases but the market continues to be flat. This could be related to lack of knowledge about these facilities, lack of trust on the technologies that increase the perceived risk on the operation, lack of interest to get into new investment projects or the bureaucratic cost of the actions. Finally, expectations were different for heat pumps and electric vehicles, where respondents were looking for price parity with conventional alternatives (fig. 12). The difference in perceptions could be attributed to the lack of widespread knowledge about heat pumps and cultural factors in the case of electric vehicles.

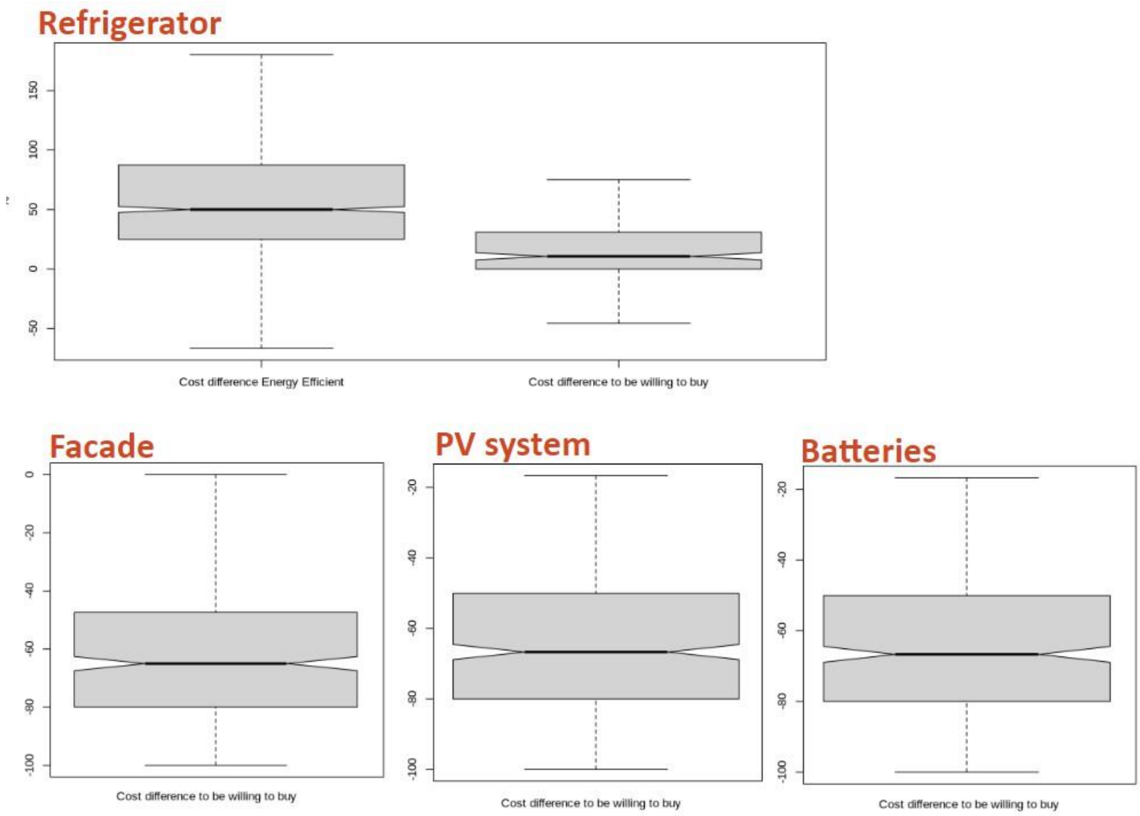

Figure 13. Willingness to pay more for energy efficient products (e.g. refrigerator) compared to other technologies (façade, photovoltaic system, batteries)

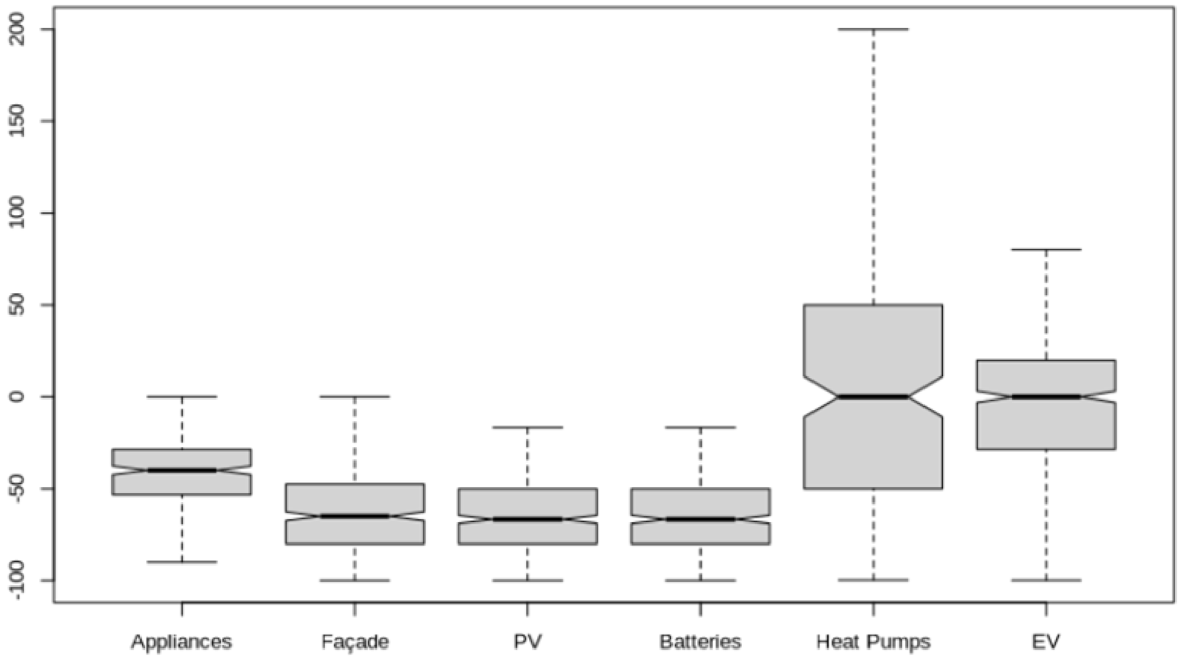

Figure 14. Cost difference to be willing to buy

Considering them together, the experts concluded that Appliances, Heat Pumps and EV technologies were grouped differently, due to the difference in people's willingness to invest in them. This categorization identifies those technologies that, as a general rule, are not sufficiently known by the general public, causing uncertainty at the moment of investing or recognizing their potential for economic savings.

4. The counterfactual level model

In this section, we introduce how we have used the WHY toolkit in order to evaluate counterfactual problems. Counterfactuals are a complex, widely difficult, and poorly understood concept.[26] Basically, we are trying to solve problems that can be stated as "what **would have happened**, given *current knowledge*, if we **had changed** something in the past". Note that counterfactual problems are not related to forecasting (as in association and intervention level models), but rather to "*backcasting*" (predicting / understanding what happened in the past).

There are certain activities where it is necessary to have access to the actual load profile of the consumer and not just the energy consumption. The most obvious example is the use of "price signals"[27] and the associated "energy management optimization".[28] However, until recently, with the mass deployment of smart meters, this information was not widely available. To overcome this problem, TSOs usually conduct campaigns to estimate "standardised load profiles" and provide this information for each year. The usual procedure is as follows:

- At the beginning of the year, the TSO provides a list of weights for each hour of the year for different consumer types (usually for each contract type or similar deterministic classification).
- To reconstruct a load profile for a customer, you simply multiply the annual energy consumption of the customer by the weights for the required category.
- If you do not need to produce the full year of data, you can simply deaggregate any period by accessing the total energy consumption for the period and reweighting the weights for the period (i.e., dividing the weights for the period by the sum of all the weights for the period).

In this section, we describe how the results of the project can be used to create similar counterfactual models for disaggregating energy load profiles.

Our work in Deliverable D2.1 provides a list of electricity consumption archetypes. Each archetype consists of a vector of coefficients that summarise the electrical energy consumption for each hour of the year. To be used, the coefficients must be multiplied by the upper whisker of the energy consumption distribution, which is the minimum between the maximum energy demand for an hour of the year and 1.5 times the interquartile range of the hourly energy demand. Using the upper whisker instead of other metrics allows us to remove outliers in energy consumption, which are quite common in residential load profiles. However, the only way to use these clusters as counterfactual models is to assume that the customer has not produced any outlier energy consumption during the year and use the maximum load as the upper whisker. The main problem with this approach is that while consumers with smart meters have this value, typically only consumers with high energy use have installed a maximeter that provides this value.

To overcome this problem, we simply transformed the coefficients into a weight by dividing them by the total sum of the coefficients. Then, to evaluate the usefulness, we conducted the following experiments:

- Estimate the error of the standardised load profiles used in different countries (as explained in Section 2.1.2). To do this, we simply use the standardised load profile to desegregate each individual time series we have in the dataset and estimate its error.

[26] Robinson, E. J., & Beck, S. (2000). What is difficult about counterfactual reasoning? In P. Mitchell & K. J. Riggs (Eds.), Children's reasoning and the mind (pp. 101–119). Psychology Press/Taylor & Francis (UK).

[27] A tariff scheme where each hour of the day could have a different price. It should trigger demand response actions mainly but it has been seen how it also fosters energy efficiency and energy conservation actions.

[28] Under a price signal tariff scheme, control different appliances to reduce the cost.

- As a contrast method, we will use the error that the electrical archetypes have as used in Section 2. Namely, we assume that we have one year of data to properly segment the time series into its electrical archetype, and then use the upper whisker of the year to desegregate it. This method will be displayed as `min30_u` (representing the best cluster).
- We also provide the error using renormalizing the clusters to be able to work with energy consumption. Namely, we normalise the values of the clusters such as they sum 1. This method will be displayed as `min30_c` (representing and the best cluster) and by `C_`*n* (with *n* from 1 to 30) representing each one of the clusters.

In all cases, we calculate the MAPE error between each of the methods and reality. The following Table 12 show the numerical results obtained by applying these methods to the dataset.

We have gathered the 30 archetypical load profiles obtained from clustering (see Section 2), and some of the “standardised load profiles” estimated by the TSOs of different countries (Germany, Ireland, and Spain) for 2023. The fact that there are different estimations per country is because each of them are targeted or segmented to different sectors of the population, such as households, different kinds of industries, etc.

Then, we have rearranged all the load profiles in the dataset, as well as the TSOs’ “standardised load profiles”, into a 24 x 371 matrix, with rows representing the hours of the day (from 00h to 23h), and columns representing the dates of that ISO year, the same way as carried out in Deliverable D2.1 for visualisation purposes. In the case of the TSO data, this rearrangement has been straightforward, as they provide a 1-year long estimation. However, in the case of the load profiles in the different dataset, this rearrangement has required an *ISO-year alignment* as explained in Deliverable D2.1, where the median time series of the different years has been computed.

After rearranging both the “standardised load profiles” of the TSOs and all the time series in the dataset, we measure its error in terms of MAPE. Table 12 provides the results of such a process. It can be seen that any of the standardised methods is near the proposed methods (by a large margin of almost 20 percentage points).

Table 12. MAPE (%) error of MTLF with different methods over the entire dataset

Method	Q1	Median	Q3
min30_u	34.73	46.74	71.45
min30_c	46.79	67.89	101.91
de_H0	61.58	86.93	128.66
de_G6	64.41	87.83	127.49
es_p20td	60.79	88.44	132.75
de_G2	66.09	89.28	127.57
de_L2	62.07	89.47	133.42
de_L0	63.52	90.67	134.69
es_p30tdve	68.52	90.83	129.17
es_p30td	64.86	93.95	139.68

Method	Q1	Median	Q3
de_HA	71.03	94.42	134.20
ei_LP29	66.51	95.28	141.08
de_L1	68.16	95.30	139.87
ei_LP27	66.42	95.42	141.64
de_G3	65.91	95.85	142.36
ei_LP25	67.92	97.63	144.83
de_G4	71.06	98.21	142.19
de_G0	72.48	99.36	142.99
de_E0	69.20	100.13	148.31
de_G7	69.20	100.13	148.31
de_G5	96.67	127.55	178.11
de_G1	123.76	147.73	186.87
de_HF	122.56	149.24	195.17
de_B1	145.34	176.59	226.87
de_E1	149.54	178.26	223.15
de_ULB	154.66	184.31	230.94
de_ULF	170.04	204.06	258.87
de_U0	181.17	213.24	264.02
de_ULA	181.17	213.24	264.02
de_EAGU2	180.81	214.86	269.06
de_ULD	190.05	226.01	282.75
de_ULE	199.67	239.84	301.40
de_EAGU1	206.31	246.29	307.64
de_U1	209.06	248.17	308.61
de_ULC	215.51	257.45	321.21

Moreover, we have computed which of the TSO profiles plus the 30 archetypes is more similar to each of the time series. Figure 13 provides the results of this assessment. It turns out that Cluster #9, followed by Cluster #10, #13, #11, and #12, is the one that is most similar to over 14% of the time series. All these five clusters belong to the group that we called "*main houses*" in Deliverable D2.1, and they represent the consumption of a typical household. This typical consumption is exemplified in Figure 14, where we can see bands of high electricity consumption in the evening, following the anti-solar pattern, suggesting that household members are at home; a summer band of reduced consumption, suggesting that household members are out, probably on holidays; notable contrasts between weekday and weekend energy consumption, and so on.

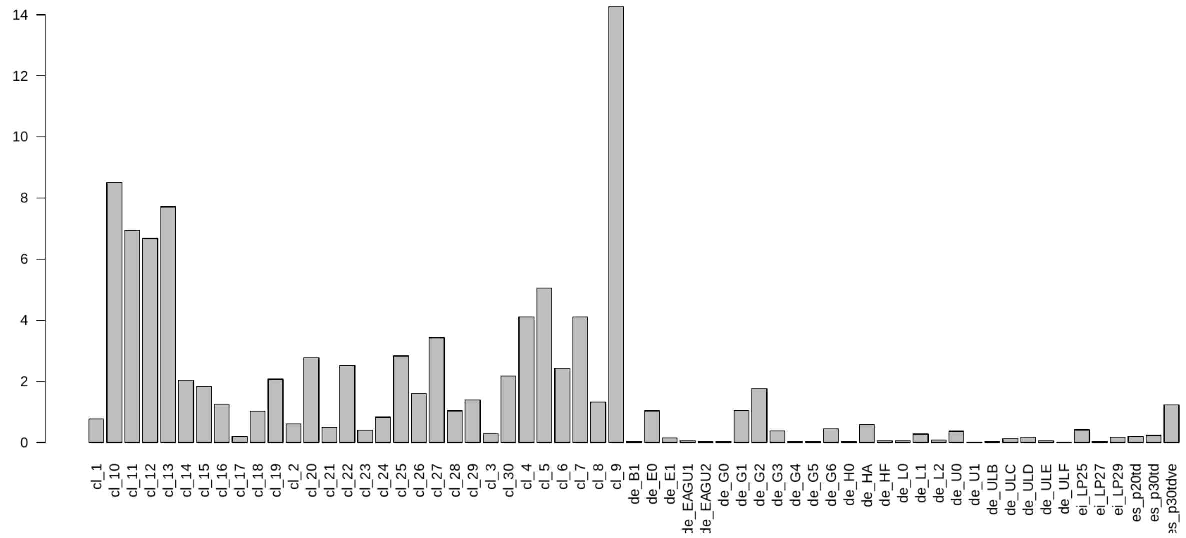

Figure 15. Similarity count (in percent) of the GoiEner time series with respect to the standardised load profiles estimated by the TSOs

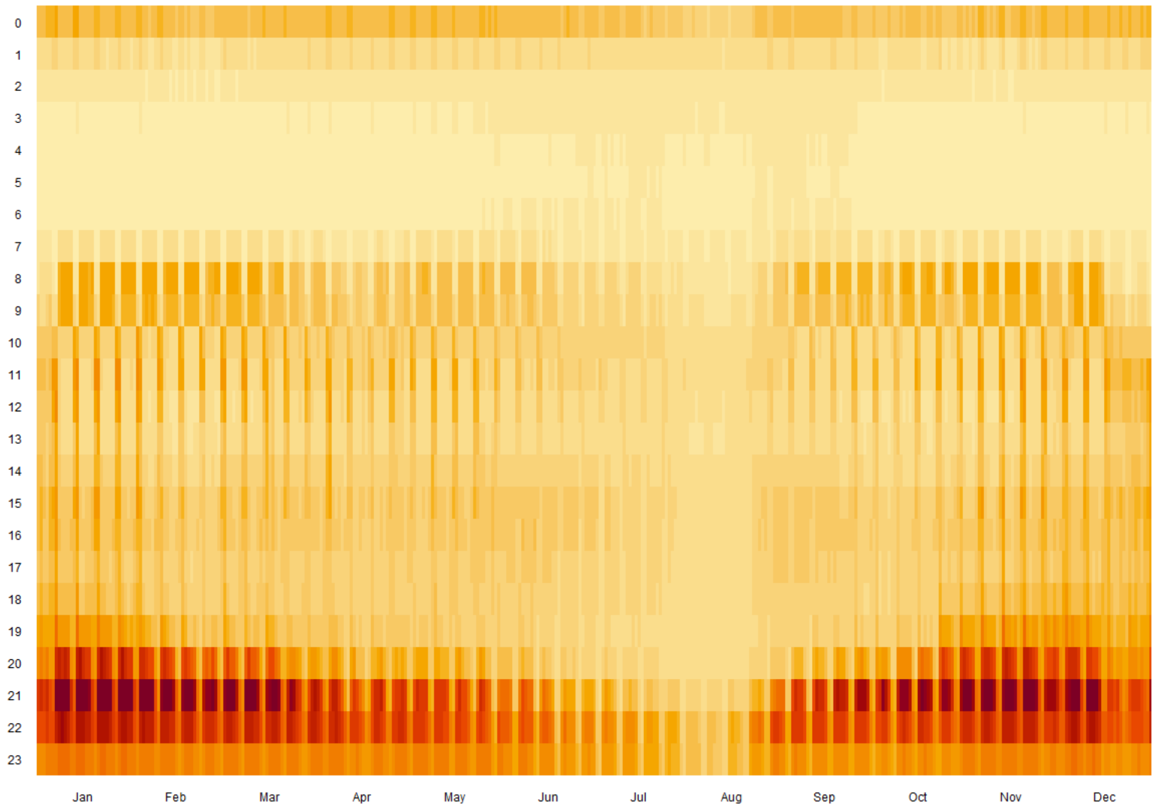

Figure 16. Heatmap of archetype #9, representing the yearly electricity consumption of a typical household

Finally, Fig 15 depicts the distribution of the MAPE for each dataset corresponding to the Spanish "standardised load profile" (`es_p2td`) for household energy consumption. The error behaviour of all datasets is similar, except for the Kaggle and Elergone Energia datasets. Notably, the Kaggle dataset contains aggregated energy consumption data from large buildings worldwide, while the Elergone dataset lacks metadata, making it impossible to determine the type of recorded consumption. However, previous analyses suggest that the Elergone dataset is also an aggregated dataset. This finding confirms that both the Kaggle and Elergone datasets contain aggregated data, as they show the best fit with the "standardised load profiles".

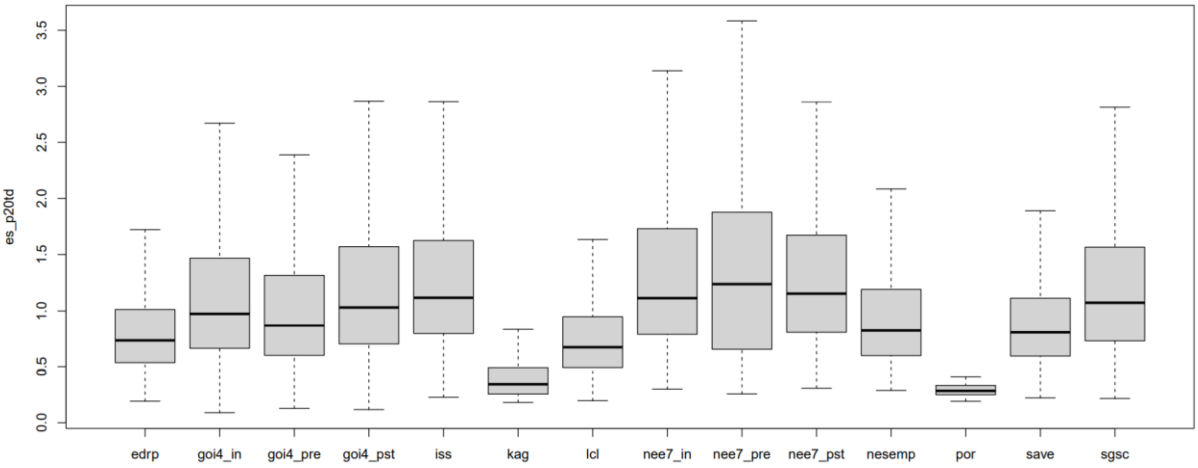

Figure 17. Similarity count (in percent) of the GoiEner time series with respect to the "standardised load profiles" estimated by the TSOs

The primary conclusion is that TSOs often do not use the most appropriate "standardised load profiles" to estimate energy consumption. Instead, they rely on profiles designed for aggregation, which leads to inaccurate estimates of individual consumption patterns.

Annex A: Summary of main findings of the survey about the ToU

In July 2019, the CNMC (National Commission on Markets and Competition) published the first proposal for a new methodology for the calculation of tolls for the transport and distribution of electricity (new tariffs). Charges were planned for 1 January 2020, but the process was delayed and modified. Finally, the new tariffs entered into force on 1 June 2021. This change in tariff has led to a number of changes in the charging of electricity consumption.

To better understand the impact of these modifications, we proposed a large-scale research action that could prove to be beneficial for both the consumer-members of Goiener and for Goiener itself as an electricity retailer.

In order to analyse the qualitative part of this research action, a survey was sent in January 2022. The objective of this document is to analyse the results obtained in this survey. The following points will be assessed:

- Knowledge of new tariffs and new opportunities for the members
- Degree of follow-up of suggested changes and barriers between the different experimental groups
- Type of energy consumer
- Socio-economic profile

The survey was answered by 691 partners. It should be noted that to send the survey it was not necessary to answer all the questions.

1. Opinion about the tariff change

This first part analyses the opinion of the partners about the change of tariff. 36% of the sample considered that this change of rate is not an appropriate measure, since the same ones are always penalised and 26% believe that they will not achieve the stated objectives. The rest considered the measure appropriate.

89% have taken measures to adapt consumption to the distribution of periods due to the change of tariff. The most common measures have been to change the time of use of clothes/dishwashers and other household appliances. The main obstacle to these measures being respected or not taken is that working/school hours are incompatible with these actions. In addition, the prioritisation of home comfort over energy costs has in many cases been an obstacle.

For a better rate adequacy, 27% have considered bringing household consumption to peak hours in the future and reducing the power contracted during peak hours.

2. Opinion about the intervention

A number of recommendations about the new tariffs and electricity consumption were sent to the partners between July and December. 19% said that they were following the advice beforehand. The other 59% stated that they had applied all or some of the advice and the other 21% did not do so for various reasons. For 57% of the sample the advice was helpful.

3. Characterisation of energy consumption

95% of the supply contracts for the partners who have responded to the survey, give electricity to the first homes.

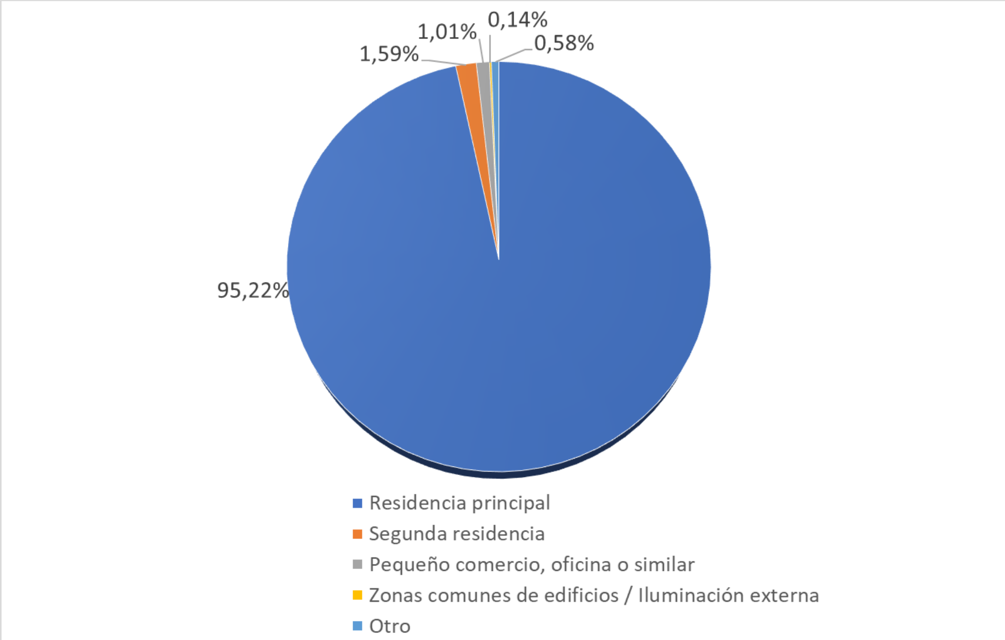

Figure 18. Distribution of supply contracts

100% of the occupants of the second house use the house in summer. The least used time is winter. Only 9% used the second house at this time. As for weekends, 45% go to the second home all or almost all weekends.

In order to analyse small businesses, they were asked about their schedule. Most, 57%, stated that they remained open continuously in the morning and afternoon.

Asking about **the type of main** heating they use, most of them, 54%, use an individual gas boiler. Moreover, the results have varied widely.

It should be noted that a large majority of respondents use firewood to warm the home.

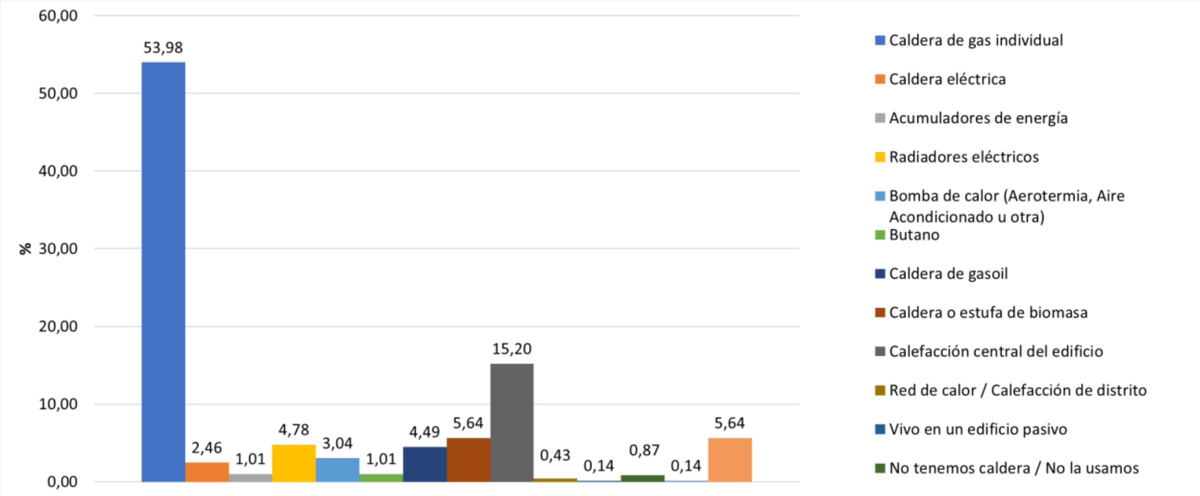

Figure 19. Main types of heating in homes

Focusing on electrical **systems,** 79% of the sample says they have an electric cooker at home. 16% have electric thermal or heat pump for Sanitary Hot Water, 3.2% electric vehicle and 2.6% of the sample has batteries (these results are not mutually exclusive, that is, a person has been able to choose more than one device or technology).

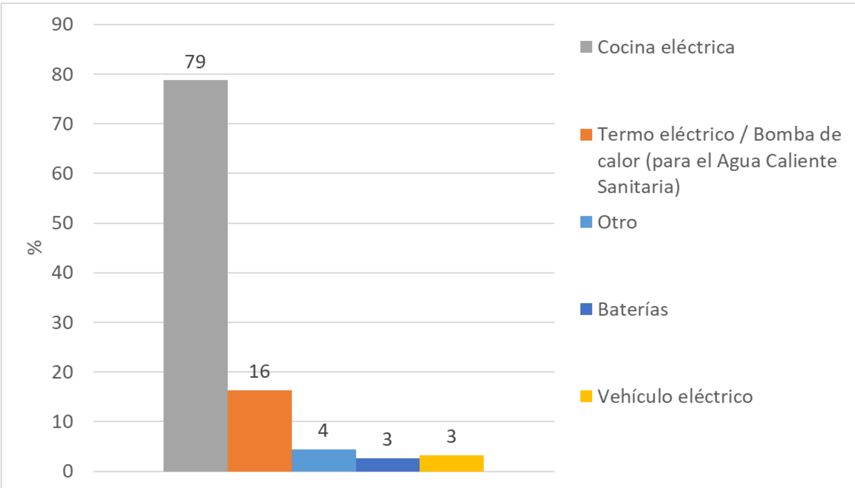

Figure 20. Household electrical systems

It should be noted that 23% of households that use electric thermo/heat pump for Sanitary Hot Water use electric radiators to warm the house. 12% have a biomass boiler or stove, 11.5% heat pump and another 11.5% central heating in the building. To a lesser extent electric boilers (6%), energy accumulators (6%), radiant soils (4%) and wood (4%) are used.

4. Characterization of the societal decision-making profile

The classification created by the European project GreenSoul was used when classifying the profiles. Three types of profiles are distinguished:

1. I prefer not to make decisions. Therefore, I try to do the same, without thinking too much. I don't have time to research, because I prioritise concentration in my daily habits.

2. I prefer to do things with the least physical and cognitive effort. I worry about my social context, although I am generally very busy, tired or impatient with other tasks. I therefore try to act without spending much time reflecting.
3. I prefer to reflect on my decisions. I know what I want and I evaluate my options. In addition, I analyse my daily actions, my performance patterns, and I learn from them.

Among the three options, 74% have been identified with the third profile, 22% with the second and 3% with the first.

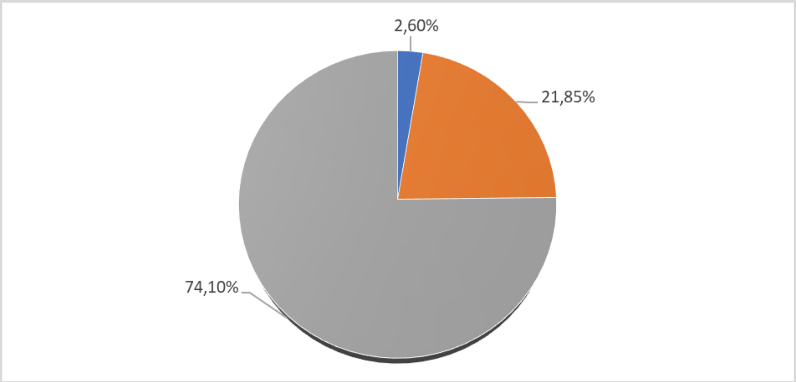

Figure 21. Distribution of the characterization of the decision-making profile

5. Characterization of the societal energy efficiency profile

The profiles were classified within the framework of the European project NUDGE. This classification differentiates 6 groups of energy consumers according to the degree of commitment and environmental knowledge:

1. I am aware of environmental protection, I think I am well informed and I am responsible for my contribution.
2. I am worried about the environment, but I prioritise my family's comfort and short-term spending control.
3. I am concerned about the environment, but I do not know the impact of my actions or what I can do to reduce it.
4. I am worried about my energy expenditure.
5. I am concerned about what people in my environment think (family, friends or co-workers).
6. I am not concerned at all about my energy consumption.

64% of the sample was identified with the first option, 23% with the second, 6% with the third, 4% with the fourth and 0.14% (1 response) with the sixth.

6. Description of the family unit

In order to have a description of the family unit, the data of the housing has been collected. The following figures show the collected data.

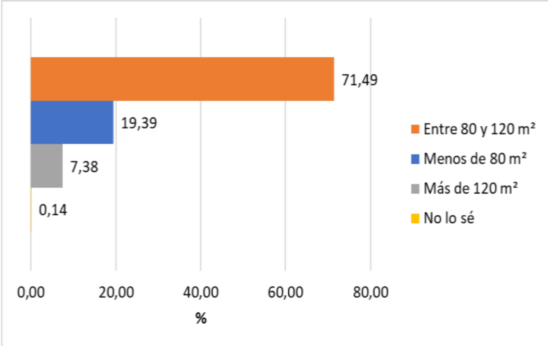

Figure 22. Ownership

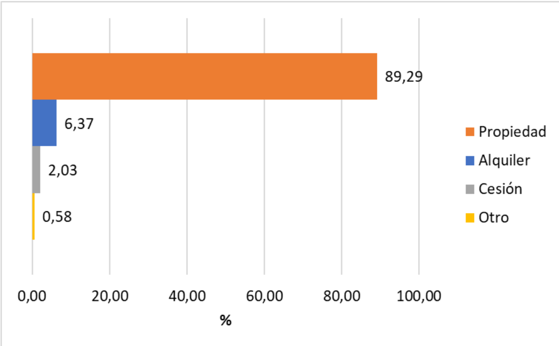

Figure 23. Useful area of housing

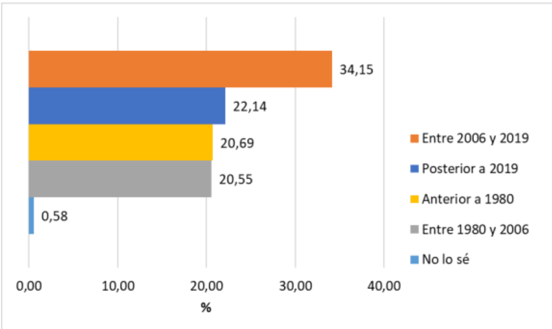

Figure 24. Year of housing/building completion

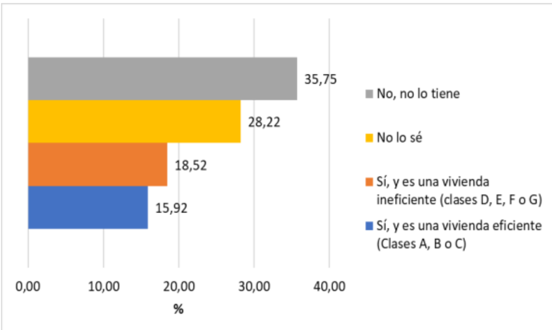

Figure 25. Home/building energy certificate

The economic situation of family units was also analysed. The following two graphs show the net annual income of the household and the annual savings respectively.

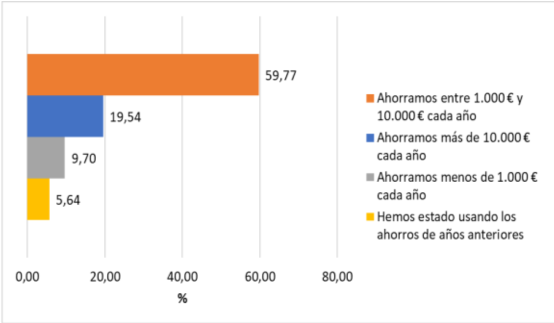

Figure 26. Net annual income

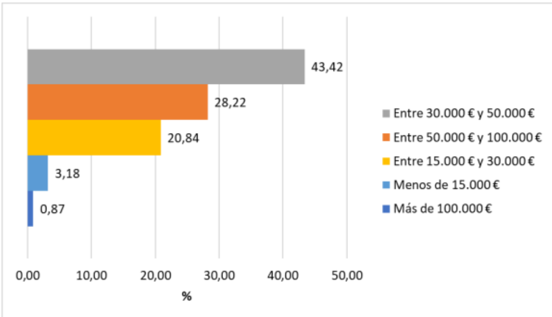

Figure 27. Annual savings

As a follow-up to the above, the average annual costs of heating and sanitary hot water, transport fuel and domestic electricity per household unit are presented.

Table 13. Average annual expenditure

	Average annual expenditure (€)
Heating and ACS	656.80
Transport fuel	1278
Electricity	628

Finally, a study has been carried out on climate change and the possibilities to address it. 96% say they are sufficiently aware (at different levels) about climate change.

42% say they have taken one or more actions to combat climate change and would like to continue doing these or similar actions. 13% said they had taken some action but did not plan to take any further action at the moment. 18% believe that their environmental behaviour should change.

48% of the sample indicates that they know the role of citizens in achieving the decarbonisation objectives of the economy for the years 2030 and 2050.

36% say they have a clear idea of what an energy community is and 42% do not know very well what it is. Among those who have clear, 81% would like to participate in an energy community. Among those who do not know what it is (18%) or do not know it exactly, 61% would like to participate in an energy community.

Looking ahead, 72% would be willing to invest in PV solar panels, 29% in electric cars, 23% in heat pumps and 20% in batteries (these results are not mutually exclusive, that is, a person has been able to choose more than one device or technology).

In order to know the type of scenario in which people are disposed to invest, 5 types of scenarios have been proposed:

1. Maintain the status quo but exclusively using electric vehicles
2. Encourage the mass deployment of the self-driving vehicle (combined with improvements in public transport)
3. Promote measures to reduce the number of journeys (teleworking, decentralisation of public administrations, digitisation of services, etc.)
4. Promote the mass creation of pedestrian areas (combined with improvements in public transport) so that a large part of the public can make a "walking" life
5. Other

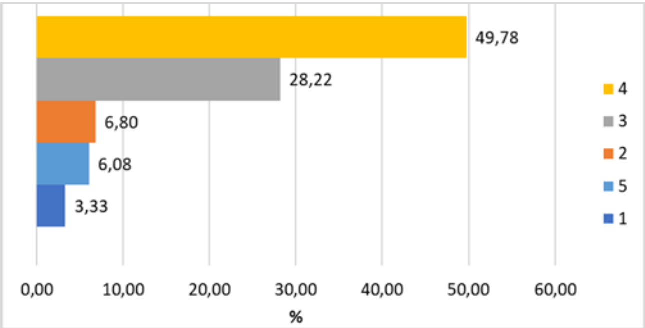

Figure 28. Future scenarios

Annex B: Summary of main findings of the survey about the PS

Over the past two years, various events (particularly harsh winter in Asia, impact of the pandemic on supply chains, war sanctions in Ukraine, etc.) have led to a sudden increase in the price of natural gas. Due to the mechanism by which electricity prices are set on the wholesale market, this increase in gas prices has resulted in a disproportionate increase in electricity prices. In order to mitigate the impact, the government introduced the Gas Price Adjustment Mechanism, which has involved a change in the way electricity is charged to households.

In order to analyse the qualitative impact of this mechanism, a questionnaire was sent in January 2023. The objective of this document is to analyse the results obtained in this questionnaire.

The questionnaire was sent to 16,990 members and answered by 699. It should be noted that to send the questionnaire it was not necessary to answer all questions.

1. Opinion on the Gas Adjustment Mechanism (MAG)

This first part analyses the opinion of the partners about the MAG. 46% of respondents indicated that their knowledge about the MAG was low, i.e. on a scale from 0 to 10, where 0 is not knowledgeable and 10 is understanding the mechanism perfectly, 46% was below 5.

When asked about their views on the mechanism, 40% indicate that it is not a good measure. On the other hand, 30% indicated that it is a good measure to combat price increases.

93% have adapted to avoid the most expensive hours. 70 per cent have indicated that the time of wearing clothes/dishwashers has changed. Other of the most mentioned actions have been using the ECO function, changing the time of other appliances and adjusting the thermostat. 43% indicate that these changes have not affected household coexistence. On the other hand, 37% said they had to prioritise savings versus comfort.

The main obstacle to these measures being respected or not taken is that working/school hours are incompatible with these actions. In addition, the prioritisation of home comfort over energy costs has in many cases been an obstacle.

MAG is what is called a price signal. Smart devices can already be found on the market to adjust their consumption to a price signal in order to reduce the economic cost of their operation generally without affecting the comfort of users. Most would be willing to make the washing machine (55%) and the heating system (54%) smart. However, today, 69% have no smart device.

2. View of the Telegram channel for price information

In August, a Telegram channel was created to report hourly prices. 3.386 partners are currently using the channel. 38% did not know of its existence. 21% looked at it and tried to follow it and 18% looked at it carefully and followed it. The 40% Telegram channel has served as aid at different scales.

3. Blackouts

There were two different situations of blackouts. In the first situation the blackouts are brief but frequent, while in the second situation the blackout lasts several days.

In both cases, drinking water at home has been the most necessary service.

4. Characterisation of energy consumption

This section asked about the energy consumption of households. These questions have only been answered by those who did not answer the previous questionnaire or did not give their CUPS. In total, there have been 528 people.

94% of the supply contracts of the partners who have responded to the questionnaire give electricity to the first homes.

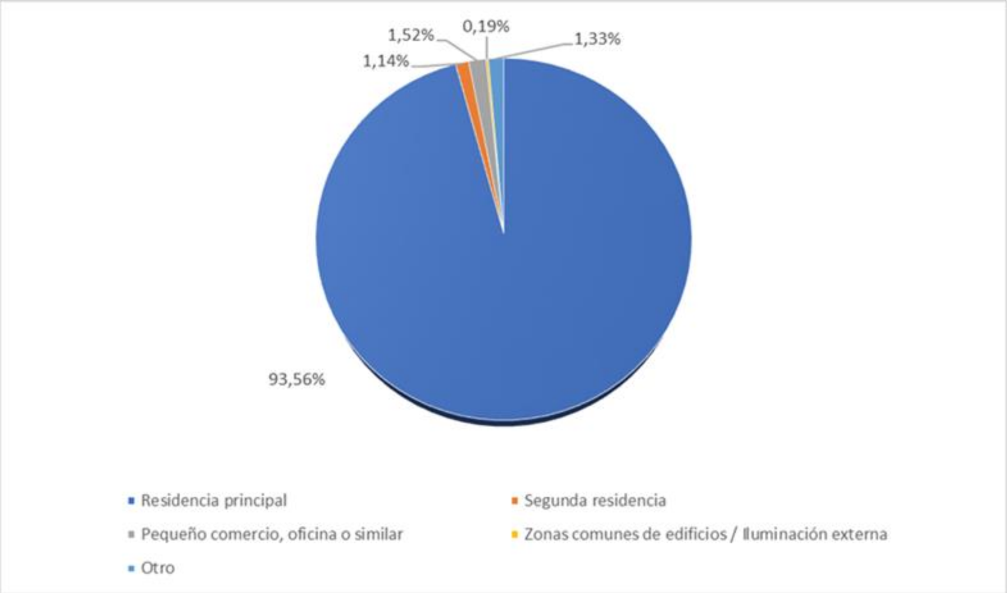

Figure 29. Distribution of supply contracts

50% of the occupants of the second housing use the housing in summer. The time when they least use it is autumn, only 16% use it at this time. As for weekends, 67% go to the second home all or almost all weekends.

In order to analyze small businesses, he asks about their schedule. Most, 50%, stated that they remained open continuously in the morning and afternoon. Also, 50% indicate that they close the premises for a time in summer.

Asking about **the type of main** heating they use, most 52% use an individual gas boiler. Moreover, the results have varied widely.

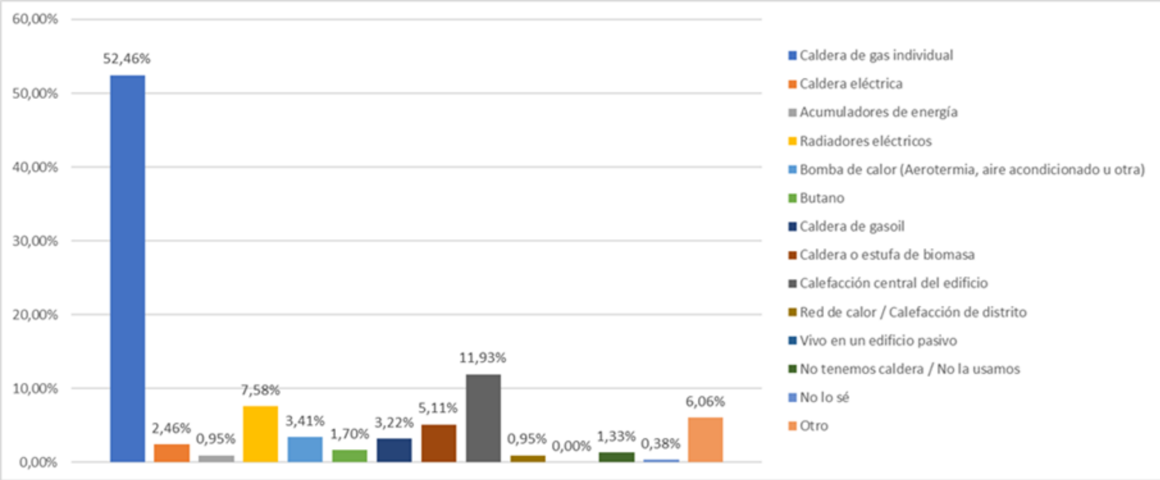

Figure 30. Main types of heating in homes

Focusing on **electrical systems,** 75% of the sample says they have electric cooking at home. 19% have electric thermal or heat pump for Sanitary Hot Water (ACS), 6% batteries, 3% air conditioning and 2% of the electric vehicle (these results are not mutually exclusive, that is, a person has been able to choose more than one device or technology).

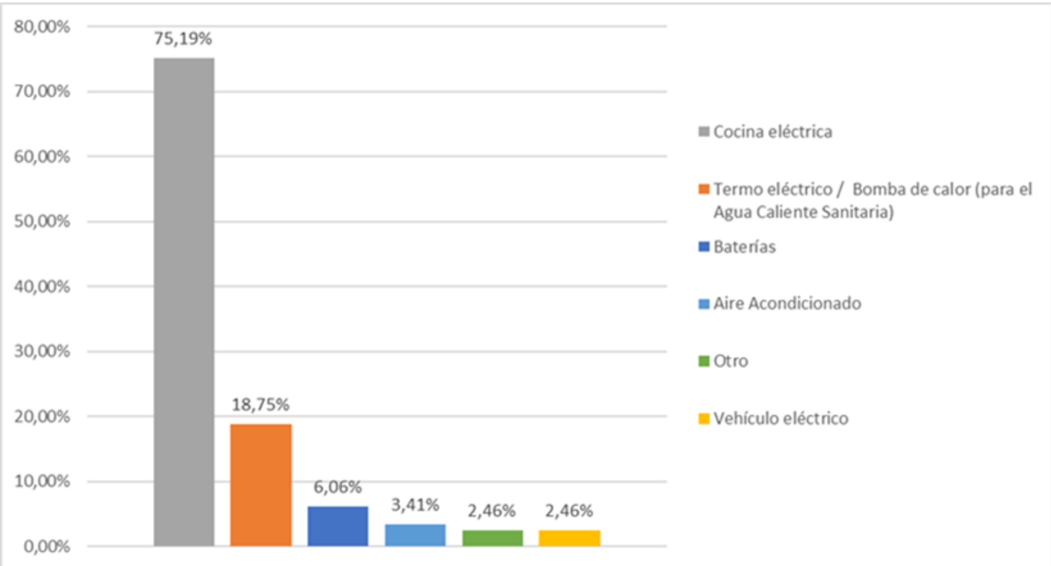

Figure 31. Household electrical systems

5. Characterization of the societal decision-making profile

The classification created by the European project GreenSoul was used when classifying the profiles. Three types of profiles are distinguished:

1. I prefer not to make decisions. I therefore try to do the same, without thinking too much. I do not have time to research, because I prioritise concentration in my daily habits.
2. I prefer to do things with the least physical and cognitive effort. I worry about my social context, although I am generally very busy, tired or impatient with other tasks. I therefore try to act without spending much time reflecting.

3. I prefer to reflect on my decisions. I know what I want and I evaluate my options. In addition, I analyse my daily actions, my performance patterns, and I learn from them.

Among the three options, 72% have been identified with the third profile, 20% with the second and 5% with the first.

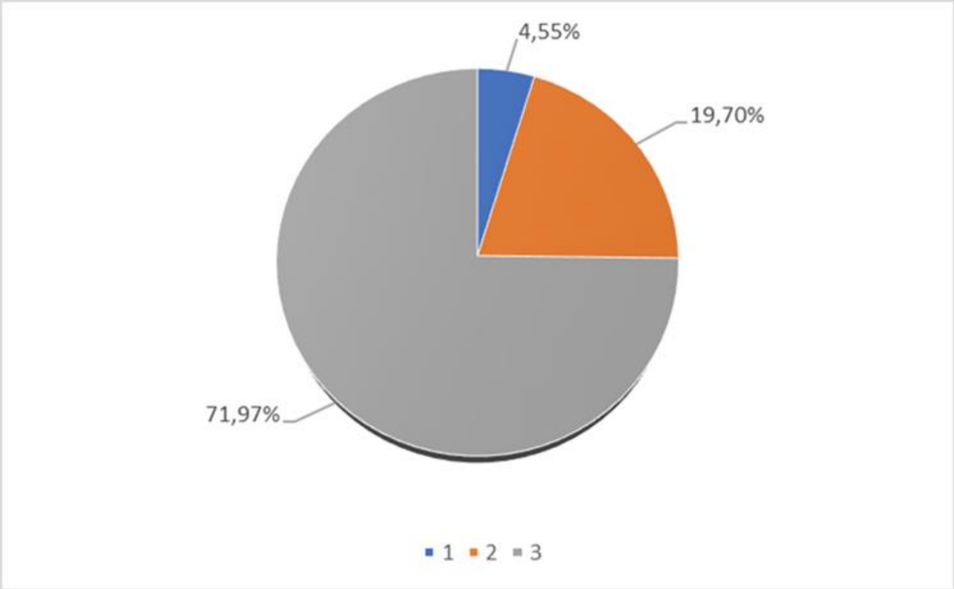

Figure 32. Distribution of the characterization of the decision-making prof

6. Characterization of the societal energy efficiency profile

The profiles were classified within the framework of the European project NUDGE. This classification differentiates 6 groups of energy consumers according to the degree of commitment and environmental knowledge:

1. I am aware of environmental protection, I think I am well informed and I am responsible for my contribution.
2. I'm worried about the environment, but I prioritise my family's comfort and short-term spending control.
3. I am concerned about the environment, but I do not know the impact of my actions or what I can do to reduce it.
4. I'm worried about my energy expenditure.
5. I am concerned about what people in my environment think (family, friends or co-workers).
6. I am not concerned at all about my energy consumption.

55% of the sample was identified with the first option, 25% with the second, 9% with the third and 9% with the fourth.

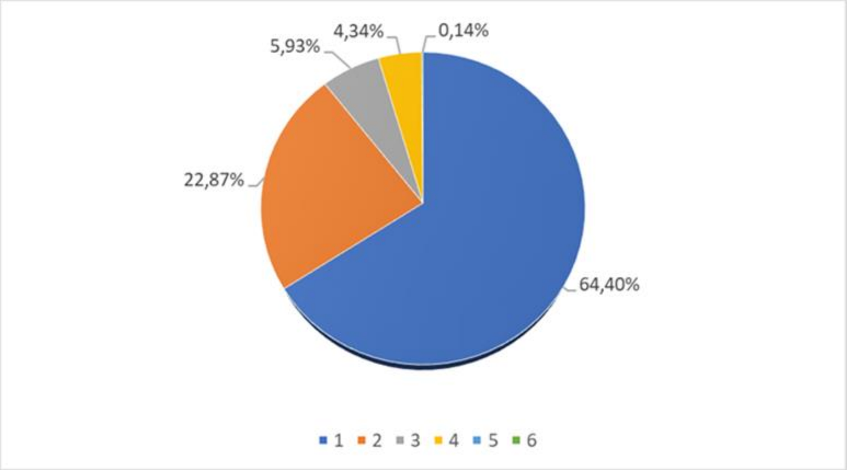

Figure 33. Distribution of the characterization of the energy efficiency profile

7. Description of the family unit

In order to have a description of the family unit, the data of the housing **has been collected first.** The following figures show the collected data.

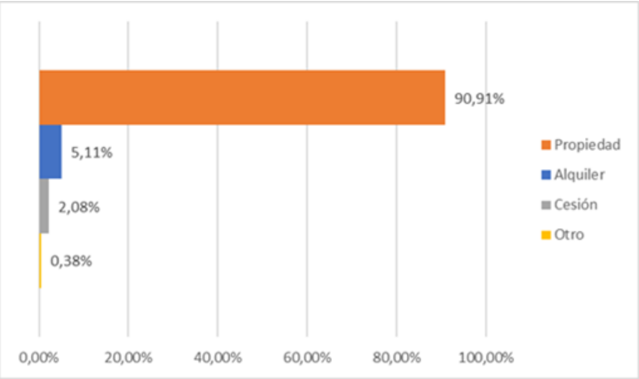

Figure 34. Ownwership

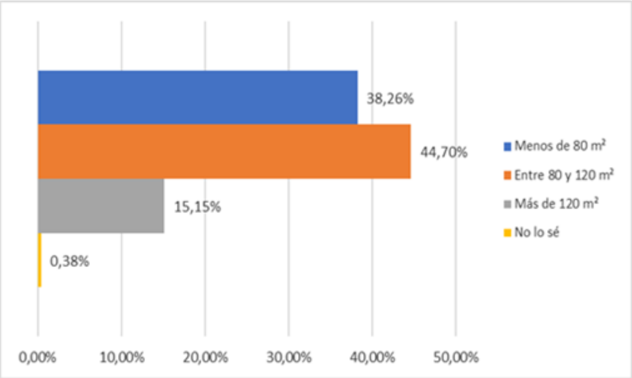

Figure 35. Useful area of housing

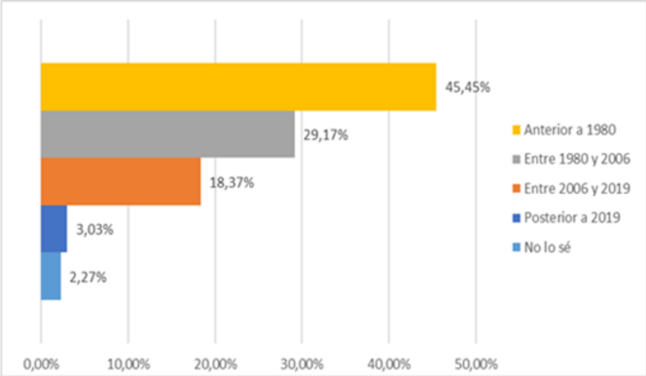

Figure 36. Year of housing/building completion

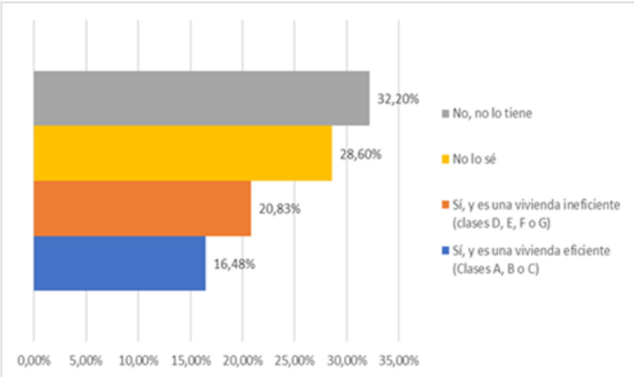

Figure 37. Home/building energy certificate

Once the housing data are analysed, the location of these houses is analysed.

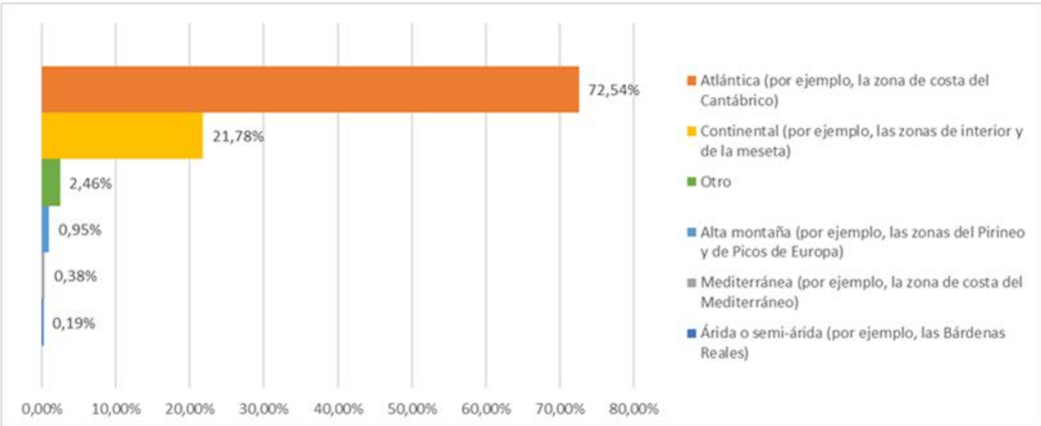

Figure 38. Climate zone

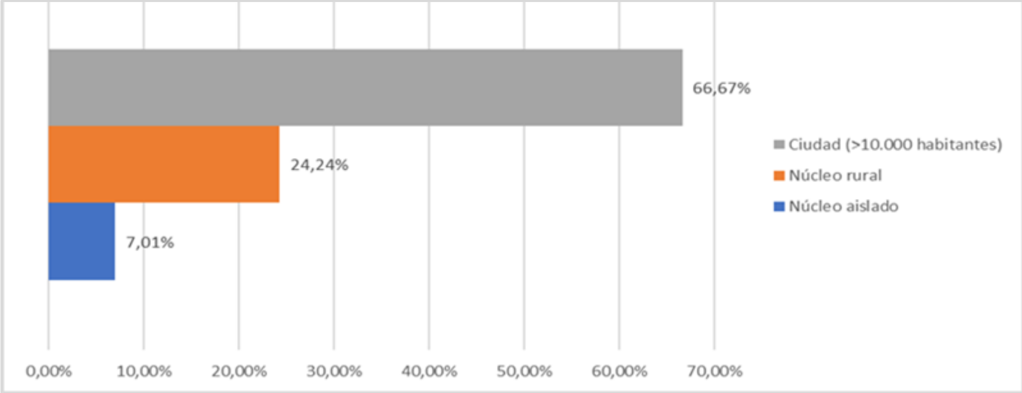

Figure 39. Size of the locality

The economic situation (income and savings) of family units will then be analysed.

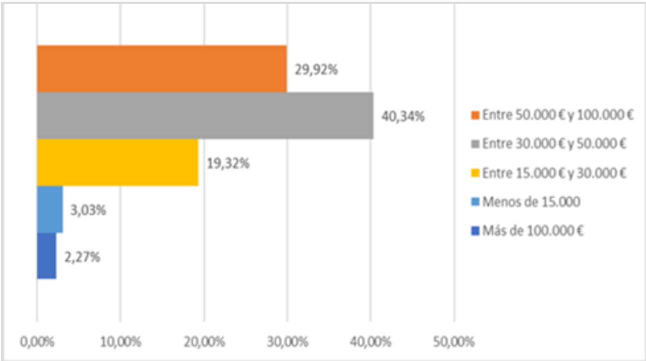

Figure 40. Net annual income

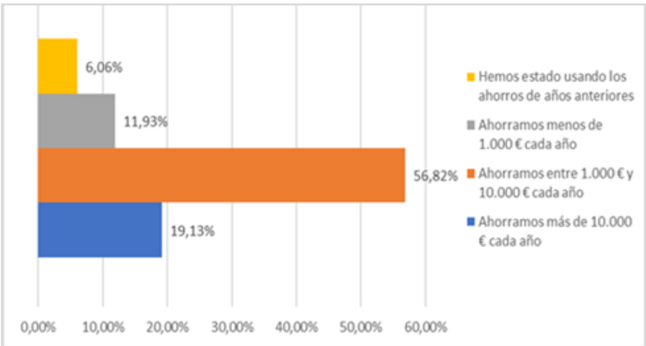

Figure 41. Annual savings

As a follow-up to the above, the average annual costs of heating and sanitary hot water, transport fuel and domestic electricity per household unit are presented.

Table 14. Average annual expenditure

	Average annual expenditure (€)
Heating and ACS	625
Transport fuel	1200
Electricity	700

Finally, a study has been carried out on climate change and the possibilities to address it. 89% say they are sufficiently aware (at different levels) of climate change.

41% say they have carried out one or more actions to combat climate change and would like to continue doing these or similar actions. 14% stated that they had taken any action, but that they did not yet plan to take any further action. 20% believe that their environmental behaviour should change.

49% of the sample indicates that they know the role of citizens in achieving the decarbonisation objectives of the economy for the years 2030 and 2050.

47% say they are clear about what an energy community is and 37% do not know very well what it is. Among those who are clear, 36% would like to participate in an energy community, that is, 17% of the total. Among those who do not know what it is (11%) or do not know it well, 29% would like to participate in an energy community, that is, 3% of the total.

Looking ahead, 79% would be willing to invest in PV solar panels, 30% in electric cars, 27% in heat pumps and 21% in batteries (these results are not mutually exclusive, that is, a person has been able to choose more than one device or technology).

In order to know the type of scenario in which investment is to be invested or in which investment is expected in the future, 5 types of scenarios have been proposed:

1. Maintain the status quo but only using electric vehicles.
2. Promote the mass deployment of the autonomous vehicle (coupled with improvements in public transport).
3. Promote measures to reduce the number of journeys (teleworking, decentralisation of public administrations, digitisation of services, etc.).
4. Promote the mass creation of pedestrian areas (combined with improvements in public transport) so that a large part of the public can make a "walking" life.
5. Another.

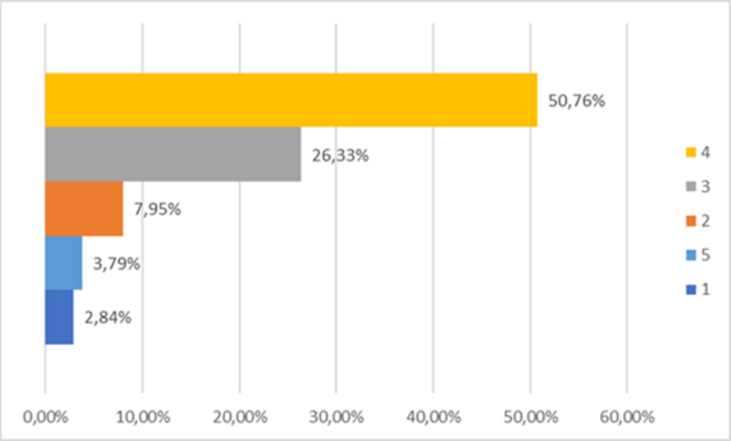

Figure 42. Future scenarios

42% of those who did not respond or did not give their CUPS in the previous questionnaire have given their CUPS in this questionnaire.

33% of those who did not give the CUPS indicated their political spectrum in a range from 0 to 10, where 0 is the extreme right and 10 the extreme left. Those who gave the CUPS were not asked to indicate the political spectrum for privacy reasons.

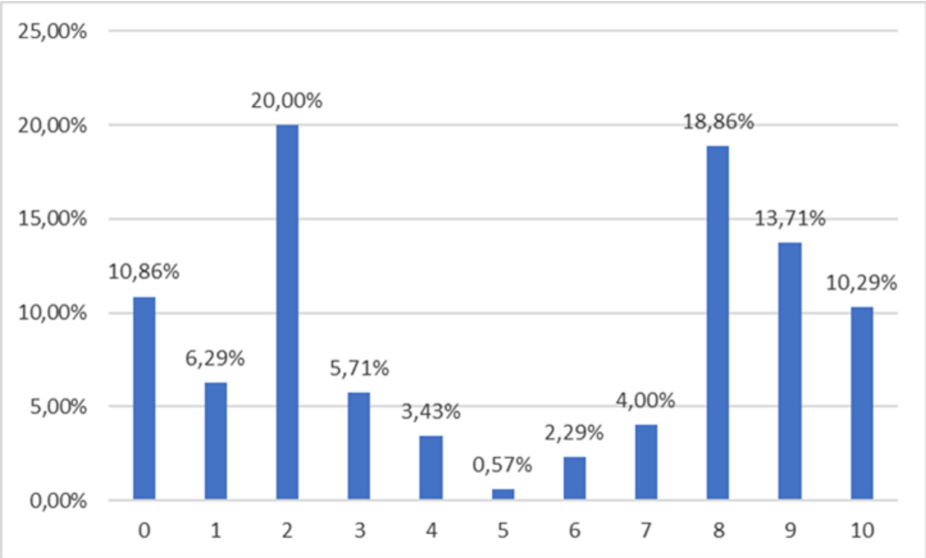

Figure 43. Political spectrum, where 0 is the extreme right and 10 the extreme left

8. Determinants of investments

This section analyses the determinants of investments in different areas: household appliances, insulation, energy generation, energy storage, heating and mobility.

Table 15. Determinants of investments

		Household appliances	***Isolation***	***Energy generation***	***Energy storage***	***Heating***	***Mobility***
Expected price (€)		*650*	*-*	*5,000*	*4,000*	*2,000*	*20,000*
Expected price of the most efficient version (€)		*850*	*20,000*	*-*	*-*	*3,000*	*33,000*
Would you buy the efficient option?	*Yes*	*4.20%*	*32.76%*	*30.19%*	*17.45%*	*33.91%*	*14.31%*
	No	*1.29%*	*5.15%*	*5.29%*	*11.87%*	*1.72%*	*21.46%*
How much does the efficient option have to cost to buy it? (€)		*750*	*5,000*	*2,750*	*500*	*900*	*15,000*
Payback to buy the efficient option (years)		*2*	*5*	*5*	*5*	*4*	*5*

Annex C: Logistic regression model

Table 16. Socioeconomic feature importance.

Feature Name	Feature importance
heating_cost	0.010270545293579906
indepedent_adults_avg_age	0.004319088077902646
fuel_cost	0.0016012686023675281
energy_tansition_knowledge	0.0008620860347852582
climate_awareness	0.00086144309884223
climate_change_actions	0.0008524804441390073
citizen_role	0.0005976269193319933
electrical_cost	0.0005182134698241981
weekday_sleep.1	0.0004987642442644932
dependent_people_avg_age	0.0004470243443017983
dwelling_age	0.0004142894504365344
education_level	0.000383208063039062
climate	0.0003377656332828905
indepedent_adults	0.00027700651399781653
locality_size	0.0002714214880996923
size	0.0002322037409068435
annual_savings	0.00022527481165093252
energy_community	0.00018781398880110707
rent	0.0001866127626836809
dwelling_type	0.00018605198448425541
certificate	0.00016213356192509965
people_at_home	0.0001521147839161015
independent_women_adult	7.931716563230613e-05
dependent_people	4.616146922813915e-05
dependent_women	2.5182197986386692e-05

Table 14. Non-socioeconomic feature importance

Feature Name	Feature Importance
weekday_dinner	0.0018793525440073322
weekday_lunch	0.0018577139734900676
weekend_lunch	0.0015730148519163866
weekday_breakfast	0.0006112045699925024
weekday_sleep.1	0.0005924708292186729
weekend_sleep	0.0005924708292186729
weekend_breakfast	0.0003192212776187516
electric_heat	0.0001989299871953937
electric_kitchen	8.100043967635865e-05
weekend_dinner	3.5313967205140095e-05
main_residence	2.8622216895506412e-05
same_pattern_weekends	1.7282885867286636e-05
heat_pump	1.4793894989832464e-05
office_hours	1.4152426068479874e-05
summer	6.2974863179021024e-06
secondary_residence	4.722026997974134e-06
shop_office	4.716398483912803e-06
long_holidays	3.1533633302811327e-06
weekends	3.150738966725482e-06
energy_accumulators	3.149615102110667e-06
spring	3.1467473511765354e-06
autumn	1.5779040103532554e-06

Annex D: Archetypes descriptions

This annex provides a description of the eight behavioural clusters (archetypes).

Early Adopter
Early Adopter: Represents those who are forward-thinking and eager to embrace new technologies or practices, often driven by a sense of responsibility towards the environment. This group may be among the first to invest in renewable energy technologies, for instance. • **Motivations/Values:** Responsibility, environmental consciousness, personal significance, and knowledge. • **Decision-making:** Evaluates choices based on impact on satisfaction, trends, and authority. Avoids excessive bragging. • **Social Interactions:** Navigates peer pressure and desire for novelty. • **Outcome/Behavior:** Stays true to decisions and shares achievements in a balanced manner. • **Overall:** Characterized by responsibility, knowledge, and commitment to significance.

The Uninterested
The Uninterested: These individuals will make decisions based on existing norms and regulations. They might not be the first to adopt new energy-saving technologies, but they will if it aligns with their values or becomes a standard or required practice. • **Motivations/Values**: Legal conformity, technical compatibility, well-being, and community obligations. • **Decision-making**: Navigates peer pressure, stays updated with trends, respects authority. • **Social Interactions**: Sometimes flaunts intentions, seeks agreement with peers. • **Outcome/Behavior**: Emphasizes agreement after decisions are made. • **Overall**: Thoughtful and considerate; aligns choices with values and responsibilities.

The Homo Economicus
The Homo Economicus: This archetype represents the calculated decision-maker, always analyzing cost vs. benefit. They'll invest in energy-saving if it makes clear financial sense or offers tangible benefits. • **Motivations/Values:** Thorough analysis, legal compliance, environmental sustainability. • **Decision-making:** Assesses risks, prioritizes safety, trust, and cost-benefit analysis. • **Social Interactions:** Prioritizes maintaining a good credit score. • **Outcome/Behavior:** Typically doesn't reconsider decisions. • **Overall:** Balances financial factors, responsibility, and sustainability.

The Fearful
The Fearful: Driven by the desire for security, they might invest in energy-saving technologies if they believe it'll provide financial stability or safeguard their future. • **Motivations/Values:** Well-being, financial security, trust, legality. • **Decision-making:** Risk-averse, prioritizes cost-efficiency. • **Social Interactions:** Relies on acquired knowledge and qualified technicians. • **Outcome/Behavior:** Aims for a secure future for family. • **Overall:** Driven by security and trust.

The Stubborn
The Stubborn: Passionate about environmental causes, they'll likely be advocates for renewable energy, energy-saving technologies, and sustainable practices. ● **Motivations/Values: Motivations/Values:** Environmental causes, combating societal discomfort. ● **Decision-making:** Actions influenced by competence, technology, and mental well-being. ● **Social Interactions:** Dedicated to significant actions. ● **Outcome/Behavior:** Alleviates societal discomfort and eco-anxiety. ● **Overall:** Deep dedication to the environment and society.

The Influencer
The Influencer: Their decisions are socially driven. If energy-saving becomes trendy or popular, they'll likely adopt it and even influence others to do the same. ● **Motivations/Values:** Social capital, novelty, popularity, shaping norms. ● **Decision-making:** Aims for societal influence and awareness. ● **Social Interactions:** Values successful socializing and community agreement. ● **Outcome/Behavior:** Seeks societal influence and connectedness. ● **Overall:** Driven by social capital and influence.

The Careful
The Careful: They need all the facts and assurances before making a decision. They'll invest in energy-saving technologies if they're convinced of the benefits and if it's deemed a safe and informed choice. ● **Motivations/Values:** Environmental concern, safety, autonomy. ● **Decision-making:** Requires legal compliance, financial access, and theoretical knowledge. ● **Social Interactions:** Values cooperation but seeks autonomy. ● **Outcome/Behavior:** Committed and seeks specific activities. ● **Overall:** Motivated by knowledge and control.

The Activist
The Activist: Strongly driven by their beliefs, they'll proactively seek out ways to save energy and reduce environmental impact. They might also rally others to do the same. ● **Motivations/Values:** Environmental consciousness, moral obligation, responsible technology use. ● **Decision-making:** Follows regulations and norms. ● **Social Interactions:** Advocates for open knowledge and the sharing economy. ● **Outcome/Behavior:** Becomes a role model. ● **Overall:** Environmentally conscious and value-driven.