CEESEN-BENDER

Building intErventions in vulNerable Districts against Energy poveRty

Deliverable 4.1

Digital tool to prioritise buildings for renovations

Dissemination Level: Public

WP4 Set-up and coordination of support tools and capacity-building of energy professionals

Tartu 2026

Co-funded by the European Union

CENTRAL & EASTERN EUROPEAN SUSTAINABLE ENERGY NETWORK

CEESEN-BENDER

Grant Agreement No.	LIFE 101120994
Project's full title	Building intErventions in vulNerable Districts against Energy poveRty
Deliverable number and title	Deliverable 4.1. Digital tool to prioritise buildings for renovations
Type	R – Document, report
Dissemination level	PU – Public
Work Package number	WP4
Work Package Leader	Climate Alliance
Authors	Mariia Chebotareva (UTARTU), Luca Alfieri (UTARTU), Tomislav Cik (DOOR), Gregor Nepste (UTARTU), Indrek Kuusk (UTARTU)
Keywords	Digital tool, multiapartment buildings (MABs), CEESEN-BENDER

Background of the CEESEN-BENDER project

The main goal of "Building intErventions in vulNerable Districts against Energy poveRty" (i.e. CEESEN-BENDER) project, launched on September 1 2023, is **to empower and support vulnerable homeowners and renters living in Soviet-era multiapartment buildings in 5 CEE countries**: Croatia, Slovenia, Estonia, Poland, and Romania. The project will help them through the renovation process by identifying the main obstacles and creating trustworthy support services that include homeowners, their associations, and building managers.

Coordinated by Society for Sustainable Development Design (DOOR), the CEESEN-BENDER project brings together leading European researchers and experts in field from six countries: **Croatia** (Society for Sustainable Development Design / DOOR, Medjimurje Energy Agency Ltd. / MENEA, EUROLAND Ltd. / Euroland, GP STANORAD Ltd. / GP STANORAD), **Estonia** (University of Tartu / UTARTU, Tartu Regional Energy Agency / TREA, The Estonian Union of Co-operative Housing Associations / EKYL), **Slovenia** (Local Energy Agency Spodnje Podravje / LEASP), **Romania** (Alba Local Energy Agency / ALEA, Municipality of Alba Iulia / ALBA IULIA), **Poland** (Mazovian Energy Agency / MAE, Housing Cooperative "Marysin Wawerski" / SM Marysin Waw), **Germany** (Climate Alliance) in addition to **Central Eastern European** Sustainable Energy Network (CEESEN).

As stated, the **main objective** of CEESEN-BENDER is to empower and support vulnerable homeowners and renters living in multiapartment buildings through the renovation process by identifying the main obstacles, and creating trustworthy support services that include homeowners, their associations, and building managers.

Therefore, the **detailed objectives** for CEESEN-BENDER are stated below:

- The project will analyse the ownership structure and physical characteristics of buildings in the 5 pilot sites in targeted regions to comprehensively understand the underlying obstacles that impede or halt homeowner associations, landlords, and property managers from pursuing energy renovations.
- Project partners will identify both legislation and financial, and technical administrative obstacles for the renovation in pilot countries. The identification of obstacles from the homeowners' perspective will help the creation of tailor-made solutions not only for homeowners but also for building managers, landlords, municipalities and other relevant stakeholders involved in the renovation process.
- Through the project methods and tools that can be used to address different aspects of energy poverty will be developed. This includes:
    - Data gathering on energy poverty in the pilot sites;
    - A digital tool identifying buildings with high levels of energy poor households in greatest need of renovation;

  - A model of potential savings in buildings undergoing renovation, and a tool for calculating the return on investment for energy renovations.

- 5 Pilot area roadmaps will be developed that prioritise building renovation based on their potential for maximising emissions reduction via energy savings as well as an increase of quality of life and wellbeing for vulnerable homeowners.

- Within the 5 pilot areas, at least 30 building-level roadmaps will be created that specifies the technical details for renovations. These pilot buildings will be supported in the entire pre- construction phase, drawing of plans, applying for permits, audits or other requirements and for financing. Plans will call for the decarbonization of the heating and cooling supply and integration of renewable energy sources (RES), to produce energy to cover its own consumption.

- Also, a support system for homeowners, municipalities, and other large owners of multiapartment buildings (MABs) in targeted regions will be created to speed up the renovation process, by
  - Training at least 3500 homeowners, landlords and building managers on legal, financial, technical and other aspects of energy renovations.
  - Advocating for changes of regulatory requirements and policies to lower the costs and time needed for the preparatory phase of projects.
  - Train at least 30 energy professionals on energy poverty and related topics.

The CEESEN-BENDER project is carried out from September 2023 until August 2026 and has a total budget of €1,85 million, of which €1,75 million is funded from the European Union's Programme for the Environment and Climate Action (LIFE 2021-2027) under grant agreement n° LIFE 101120994. The project is being carried out by internationally recognized researchers and partners from Society for Sustainable Development Design, University of Tartu, Local Energy Agency Spodnje Podravje, Alba Local Energy Agency, Climate Alliance, Medjimurje Energy Agency Ltd., Mazovian Energy Agency, Tartu Regional Energy Agency, Municipality of Alba Iulia, Central Eastern European Sustainable Energy Network. In addition, associated partners of the project include Housing Cooperative "Marysin Wawerski", EUROLAND Ltd., GP STANORAD Ltd. and The Estonian Union of Co-operative Housing Associations.

Table of Contents

1. Introduction and Relevance of this Deliverable

Energy poverty is a pressing concern in the European Union, characterized by households' inability to access essential energy services that ensure basic living conditions and health standards. This phenomenon is particularly acute in the context of multiapartment buildings (MABs), where structural inefficiencies and financial constraints exacerbate the challenges faced by residents.

In line with the requirements of Energy Performance of Buildings Directive (EPBD) (Directive 2024/1275), the CEESEN-BENDER project aims to enhance renovation of existing building stock in five Eastern European countries by creating statistical data driven tool designed to prioritize buildings for renovation based on energy efficiency and the socio-economic characteristics of their residents. This tool utilizes secondary socioeconomic, technical and energy consumption data to generate rankings of buildings that are least energy-efficient and most likely having house occupants experiencing high levels of energy poverty.

The tool employs a definition of energy poverty that aligns with the European Commission's guidelines, emphasizing the multifaceted nature of the issue. Specifically, energy poverty is defined as a situation where a household must reduce its energy consumption to a degree that negatively impacts the health and well-being of its inhabitants, driven by factors such as high energy expenditures, low income, and poor energy performance of buildings (European Commission, n.d.).

The methodology of the tool development is driven by the available data in Estonia, including information on buildings' residents and energy consumption data. The process of merging these datasets presents significant challenges, particularly in terms of ensuring data consistency and accuracy. The methodology includes regression analysis (multiple linear regression models) combined with LASSO technique.

The alternative ranking used for producing a ranking system that maintained certain intersections with the foundational logic of the original tool. This alternative model placed increased emphasis on factors such as social vulnerability, the condition of buildings, and the perceived necessity for renovations. Consequently, the validation process was structured to assess whether the integration of a more pronounced social perspective, alongside practical, field-based knowledge, would result in significant changes in the prioritization of buildings. The validation utilized a brief online survey administered to buildings that had previously undergone evaluation via the digital tool. The alternative ranking was established through the application of a weighted composite index. This framework prioritized indicators associated with energy poverty, energy vulnerability, and the necessity for building-level renovations, while also integrating pertinent technical and structural information.

The stages of tool development and validation can be visually summarised in Figure 1. The process includes several phases starting from the literature review dedicated to already existing models and tools used on different stages of renovation and ending with validation of the generated code on the data provided by project partners and developing an alternative ranking.

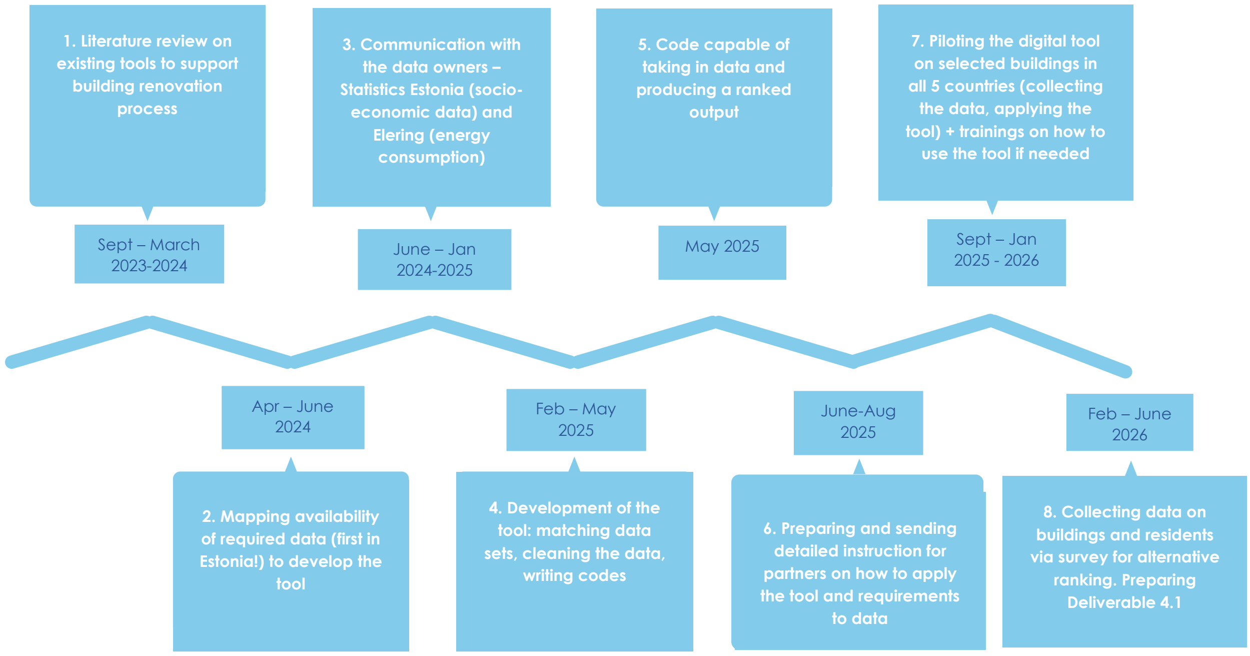


Figure 1. The process of Task 4.1 development and data collection

Source: compiled by the authors

The primary potential users of this statistical tool include first of all data owners, given its data-driven nature. Depending on the specific legal and regulatory context of each country, local authorities may utilize this tool, particularly in the context of municipal building renovations. Additionally, financial intermediaries, such as the former KredEx in Estonia, may employ the tool to inform their decisions regarding grants for the renovation of multi-apartment buildings. Furthermore, energy agencies, including the Tartu Regional Energy Agency and other partners involved in the CEESEN-BENDER project, may leverage this tool to offer data-driven recommendations.

2. Literature review on existing tools used to prioritise buildings for renovation

Literature analysis has revealed that existing tools support decision-making on different stages of renovation process. Some of the tools address the question which buildings should be renovated (this is also the phase where CEESEN-BENDER project contributes), some other tools assist to design the renovation scenarios and help to choose between alternatives (see Figure 2). There are tools which are used during construction, operation and usage of the building. Finally, whether the building should be demolished or reused can also be decided with the help of different digital tools. At least 43 decision support tools applicable in Sustainable Building Renovation (SBR) process have been discovered (Nielsen et al., 2016).

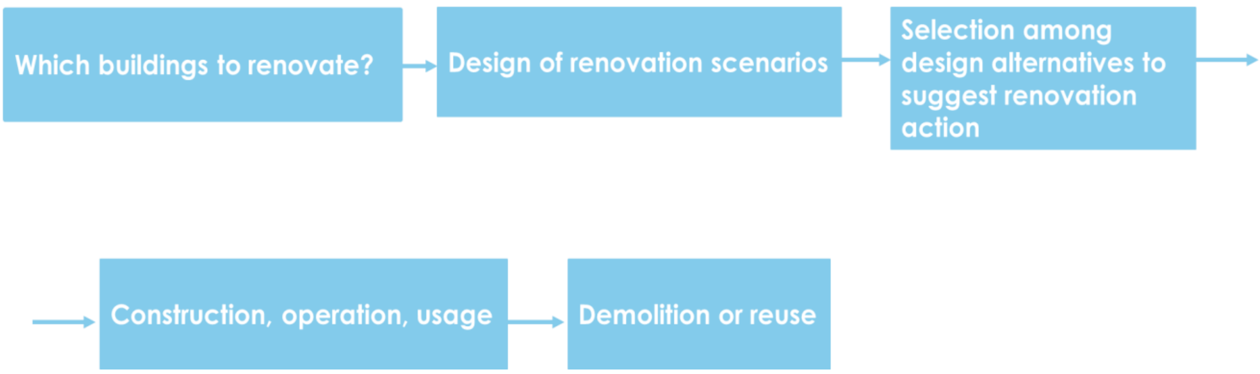


Figure 2. Stages of building renovation process where digital tools can be applied

Source: compiled by the authors based on Nielsen et al. (2016)

Besides diverse digital tools applied on different stages of decision-making process discussed in Nielsen et al. (2016), the previous studies also cover country specific rating systems for building sustainability evaluation (e.g. certification systems LEED in US, BREEAM in UK and DGNB in Germany, HK-BEAM in Hong Kong, Green Mark in Singapore, Green Building Index in Malaysia, Greenship in Indonesia, CASBEE Japan, etc.) (Zhang et al., 2019). However, they are mainly environmentally driven, based on indicators of building performance.

Some other tools introduce methodology for developing a globally working rating tool using a multilevel weighting scheme especially for rating sustainability in existing buildings, such as the one designed by Mahmoud et al. (2019). Moreover, assessment tools addressing sustainability aspects have been developed under the funded research projects (e.g. SuPer-Buildings; OPEN HOUSE). Among them, the Decision Support Tool developed in PrioritEE project which aims at helping local and regional authorities to evaluate the possibility for energy (and financial) savings by applying energy efficiency measures in public buildings (Salvia et al., 2021). Only a few of the tools consider the building stock's resilient aspects (e.g. building condition evaluation) (Jiménez-Pulido et al., 2021).

Digital tools for building renovation prioritization employ diverse methodological approaches including multi-criteria decision analysis (Dell'Anna, 2023; García-Fuentes et al., 2019), bi-level optimization combining Energy Hub and Knapsack Problem algorithms (Wang, Shi, & Niffeler, 2025), Bayesian Networks for uncertainty handling (Giretti et al., 2017), and statistical sensitivity analysis using Sobol' indices (Galimshina et al., 2020). These tools have been applied across scales ranging from small portfolios of 6-56 buildings (Gade et al., 2018) to regional analyses of 888,484 households (Stegnar, 2025), targeting educational (Gade et al., 2018; Piredda et al., 2025), residential (Dell'Anna, 2023; Wang, Shi, & Niffeler, 2025; Galimshina et al., 2020; Stegnar, 2025), and municipal building stocks (Giretti et al., 2017; Salvia et al., 2021).

Prioritization criteria integrate energy-related factors (consumption, efficiency potential, emissions) (Wang, Shi, & Niffeler, 2025; Stegnar, 2025; Piredda et al., 2025), economic considerations (costs, budget constraints) (Dell'Anna, 2023; Wang, Shi, & Niffeler, 2025), and socio-economic indicators including energy poverty and municipal income (Stegnar, 2025). Tool recommendations sometimes diverge from conventional renovation practices (Galimshina et al., 2020), with prioritization results varying based on whether energy-centric or value-based approaches are employed and how criteria are weighted (Stegnar, 2025). Implementation barriers include incomplete data availability (Giretti et al., 2017; Stegnar, 2025), computational complexity limiting optimization accuracy (Wang, Shi, & Niffeler, 2025), and software architecture constraints preventing certain renovation combinations (Wang, Shi, & Niffeler, 2025). The evidence indicates that structured, data-driven prioritization tools can support renovation decision-making, but optimal tool selection depends critically on matching technical sophistication to available data resources, organizational capabilities, and stakeholder priorities. The complete scope of reviewed tools is organised in Table 1.

Table 1. Overview of different tools to prioritise buildings for renovation

Study	**Tool name/type**	**Building category**	**Geographic scope**	**Methodology**
Gade et al. (2018)	REDIS (value-based decision support tool)	Educational (schools)	European countries	Value-based decision support calculating Renovation Value Factor
Stegnar (2025)	Hybrid MultiCriteria Decision-Making framework	Residential (single-family and multi-family houses)	Slovenia (12 regions)	MCDM with Building Priority Factor and Municipal Energy Poverty Factor
Piredda et al. (2025)	Energy performance-based prioritization framework	Educational (schools)	Northern Italy	Data collection, energy simulation, and ranking
Wang, Shi, & Niffeler (2025)	Optimization-based prioritization mechanism (CEA plug-in)	Residential	Altstetten, Zurich	Bi-level optimization combining Energy Hub and Knapsack Problem
Giretti et al. (2017)	Multi-criteria decision support tool	Public/municipal buildings	Not specified	Multi-criteria analysis with Bayesian Networks
Galimshina et al. (2020)	Robust assessment framework	Residential	Not specified	LCA/LCC integration with Sobol' indices
Salvia et al. (2021)	PrioritEE toolbox	Municipal public buildings	European cities	Decision support tool with key performance indicators

Study	Tool name/type	Building category	Geographic scope	Methodology
Dell'Anna (2023)	ELECTRE TRI-B model	Residential (apartment and multifamily buildings)	Turin, Italy	Multi-criteria decision analysis with building categorization
García-Fuentes et al. (2019)	OptEEmAL (web-based tool)	Residential	EU	MCDA with optimization framework and BIM integration

Source: compiled by the authors

Having all the diversity of existing tools which support renovation of both public and private buildings, there is still a lack of studies dedicated to digital solutions which consider not only technical features of the building and energy consumption metrics but also social and economic characteristics of people living in buildings. In particular, energy poverty component. Therefore, CEESEN-BENDER project attempts to contribute to this direction of studies.

3. Description of methodology and data

3.1. Estonian Dataset Merging Process

Task 4.1 tool was initially developed based on the Estonian data. The reason is different data accessibly and comparability among project countries. Being overall more accessible for the developers, Estonian dataset went through an extensive process of acquisition and preparation.

Two types of datasets were provided for the project: (1) information on residential buildings, households, and population statistics from Statistics of Estonia, and (2) electricity consumption data from Elering.

A key challenge was that the datasets were not provided at the same level of detail. Elering data was collected at precise address level, while Statistics Estonia applied pseudonymisation by publishing their data on a 100m x 100m grid (GRD_INSPIR), meaning that the location of an address was represented only within a grid square rather than as an exact building coordinate. To enable merging, Elering also supplied a grid attribute, linking their address-level data to the same 100m x 100m framework used by Statistics Estonia. This approach allowed the datasets to be merged, but it also meant that the project could no longer be conducted at individual building level. Instead, the analysis was performed at the grid level, where multiple buildings and households may be aggregated within a single square. The first objective, therefore, was to create a consistent and analyzable dataset that combined demographic, housing, and energy consumption information at this common spatial resolution.

The workflow for merging the datasets was carried out in R, using the packages *readxl*, *dplyr*, and *openxlsx* to manage input, transformation, and output of the data. File paths were defined for one "first" file and several "other" files.

A custom function *read_tidy_sheet()* was developed to handle the Excel files with two-row headers. The first row often contained merged month labels (for example: "NOV.21"), while the second row specified "IN" or "OUT". The function was designed to:

- Read both header rows and fill missing values in the first row.
- Combine the two rows into standardized column names (for example: NOV.21_IN, NOV.21_OUT).
- Skip the header rows and read the actual data.

The datasets were joined using the column GRD_INSPIR, which served as the unique identifier. Since this column appeared under different name variations across files, the function standardized the name to ensure consistency.

From the first file, two sheets were read and combined into a single dataset. Each of the other files was then processed in the same way, and their data was merged with the first dataset through an inner join on GRD_INSPIR. The final outputs were written back into Excel format, each saved under the original filename with the *suffix _merged*.

The header-cleaning and renaming strategy proved effective in producing usable and uniform datasets. The automation ensured that the numerous monthly energy consumption columns were correctly aligned without manual editing. The inner join reliably linked data across domains such as housing, households, population, and energy use.

The main difficulties arose from inconsistent identifier naming and irregularities in the header rows. In early attempts, GRD_INSPIR was not recognized in some files, which prevented merging. This issue was resolved by adding flexible renaming logic. Blank or partially missing headers also required a fill-forward approach to maintain continuity in the time series columns.

Due to the pseudonymisation process, the merged data cannot be considered strictly address-level. While the initial Elering dataset was provided at building precision, the final merge was only possible through the shared 100m x 100m grid attribute. This shift reduced spatial accuracy but ensured consistency between data sources.

As a result, the dataset reflects grid-level conditions, combining information from all buildings and households located within the same square. This provides a reliable basis for aggregated analysis of housing conditions, household structures, population composition, and energy consumption, but it does not allow conclusions to be drawn about individual addresses or buildings.

The challenges encountered during the process were both technical and methodological. On the technical side, issues included complex headers, inconsistent identifier naming, and ensuring that joins were performed correctly. On the methodological side, the shift from address-level to grid-level analysis represented a trade-off between privacy protection and spatial precision.

Once the identifier and header issues were resolved, the merging process proceeded smoothly. The final dataset is internally consistent and well-structured, although its precision is inherently limited to the grid level.

3.2. Data description and methodology of the tool

In alignment with the methodologies established by prior research, notably Giretti et al. (2017) and Stegnar (2025), the CEESEN-BENDER project has initiated the development of a data-driven tool. This endeavor necessitated a comprehensive mapping of data, along with an assessment of its availability and accessibility. Existing literature indicates that metrics for Sustainable Building Renovation (SBR) are evaluated through a multifaceted approach, incorporating a variety of parameters derived from environmental, social, and economic dimensions of sustainability. Specifically, environmental parameters encompass metrics such as energy consumption, water usage, waste generation, and emissions of carbon dioxide and other pollutants. Social parameters are assessed through both subjective evaluations—pertaining to user satisfaction—and objective evaluations, which consider indoor climate and health metrics before and after renovation interventions. Economic parameters are reflected in considerations of operational costs, rental costs, and asset valuation. (Jensen et al., 2018)

Andersen et al. (2021) aligned SBR indicators with ESG groups of indicators – environmental, social and governance/economic indicators (see Table 2). From the list of indicators presented in Table 2, only some of the indicators have been included into the development of the tool. For example, energy consumptions and the share of energy expenditures in income. This is due to the fact that some data is not available or available but not accessible. Therefore, at first it had to be clarified with the data owners which data is possible to be shared with the project.

Table 2. Sustainable Building Renovation (SBR) indicators

Groups of sustainability indicators	**Indicators categories**	**Indicators**
Environmental indicators		
	Energy and renewable energy	Energy consumption [kWh/m2]
		Existent energy class of building [Letter]
	Material used, waste and durability	Components for re-use [kg]
		Materials for recycling [kg]
		Hazardous waste disposal [kg]
Social indicators		
	Indoor environmental quality	Concentration of substances [number of harmful substances]
		CO2 concentration [ppm CO2]

Groups of sustainability indicators	Indicators categories	Indicators
		Ventilation rate [l/s/m3]
	Energy poverty indicators	Arrears on utility bills
		Low absolute energy expenditure
		High share of energy expenditure in income
		Inability to keep home adequately warm
Economic indicators		
	Economic performance & affordability	House price to income/earnings ratio [house price and average income]
	Flexibility & adaptability (FA)	Space efficiency [usable floor area (UA) / gross floor area]
		Ceiling height [meters], Building depth [meters]

Source: compiled by the authors based on Andersen et al. (2021)

On the next stage of the tool development process, the data owners (in case of Estonia, Statistics Estonia[1] for socio-economic data and Elering[2] – for energy consumption data) have been contacted. It has to be noted that negotiation with these institutions took long time and even after that the Statistics of Estonia provided the data only on aggregated level of 100x100 meters in order to avoid identification of individuals and their personal data, such as income level.

By the end of January 2025, the data from both Statistics Estonia and Elering were received and the processes of merging two datasets together took place. The challenge of merging the data 1) on the level of buildings (energy consumption from Elering) and 2) more aggregated 100x100 meters (which might include several buildings) socio-economic data from Statistics Estonia has been discussed in subchapter 3.1. After two datasets were merged, the final sample was 'cleaned', variables were coded and the final code developed (see step 4 in Figure 1).

As an outcome, the statistical data driven tool in Task 4.1 has been developed in R programme based on the collected data from Estonia (N=194) and piloted to generate scores for 122 selected buildings across 4 other CEESEN-BENDER project partner-countries (namely Romania, Croatia, Slovenia and Poland).

[1] The Statistical Office of Estonia is the Estonian state agency under the Ministry of Finance that produces national statistics (for more information, see https://www.stat.ee/en).
[2] Elering AS is an Estonian state-owned company that manages electricity transmission networks (for more information, see https://www.elering.ee/en).

The statistical tool has been developed utilizing data from Estonia as of the first quarter of 2024. This analysis incorporates two distinct types of data:

Socio-economic data (source: Statistics Estonia)

- District of Tartu (Karlova/ Annelinn)
- Average living area per building (m2)
- Average living area per person (m2)
- Ownership profile (Share of owners, %, Share of renters %, Other type, i.e. close relative or spouse of the owner, who lives in the dwelling, %, Unknown, %)
- Median salary[3] (gross)
- Age profile (the number of people in age groups: 0–4-year-old, 5–9-year-old, 10-14… 85 and older)
- Household profile (Number of households with 1 member, 2, 3 etc. 6 members)

Energy data (source: Elering)

- Electricity in kWh (amount fed into the grid, i.e. production, and the amount taken from the grid, i.e. consumption)
- Gas data in kWh and cubic meters (only the amounts consumed from the grid (no production)

In the model, dependent variable is average difference between consumption and production of electricity for 3 months in kWh:

$$net\_AVERAGE = \frac{net\_jan24 + net\_feb24 + net\_mar24}{3} \quad (1)$$

where net_jan24, net_feb24 and net_mar24 are difference between electricity consumption and production of buildings in January, February and March 2024.

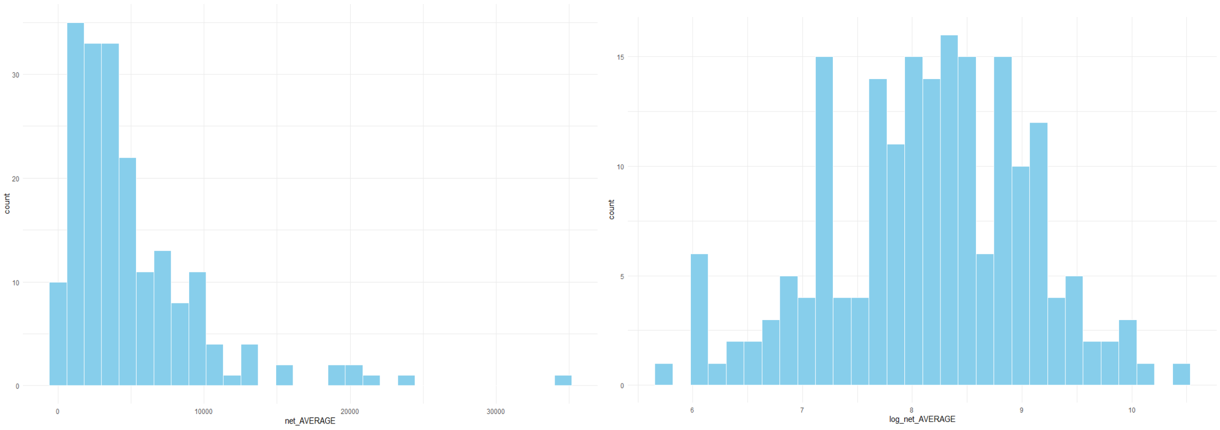
Some of the buildings in dataset were producing enough electricity to feed into the grid. Since it is desired that a bigger part of self-production covers consumption, then decreasing dependent variable is preferred. This would indicate higher energy efficiency of the building.

Since dependent was positively skewed, log transformation was applied to make *y*-variable normally distributed (see Figure 3).

Among independent variables are:

- District of Tartu
- Average living area per building
- Average living area per person
- Ownership structure
- Income (measured by median salary)
- Age (Under 15, 15-65, Over 65)
- Household composition
- Average costs of electricity in euros (for 3 months)
- Average share of electricity costs in income (energy poverty component)

[3] The median salary is calculated as of the first quarter of 2024, and the remaining socio-economic data is as of January 1, 2024.



a) Before transformation b) After transformation

Figure 3. Log transformation of dependent variable (average difference between consumption and production of electricity)

Source: compiled by the authors

Energy poverty component (share of energy expenditure in income) has also been included into the analysis. It is calculated as price[4] per kwh for the first quarter of 2024 multiplied by consumed kwh in each month. Electricity prices per kWh are taken from Nord Pool for Estonia. Costs of electricity for the whole building in euros were averaged for three months – this variable was included as one of independent variables.

Since the initial R code was developed on aggregated 100x100 meter data, it was decided to pilot the tool on the building level data collected by the project partners (ALEA, MENEA, LEASP and MAE). Under UTARTU guidelines, cross-sectional annual data was gathered in 2025 in four countries – Croatia, Romania, Slovenia and Poland – to pilot the R code developed initially on the Estonian data. Table 3 describes three groups of data applied for the analysis.

Table 3. Overview of the dataset for Slovenia, Romania, Poland and Croatia

Socio-economic data of MAB residents	**Technical characteristics of buildings**	**Energy data**
Number of (mostly) empty flats with 0 residents	GFA - Gross Floor Area (m2)	MAB electricity consumption [MWh]
Number of flats with 1 resident	Conditioned area (m2)	MAB natural gas consumption [MWh]
Number of flats with 2 residents	Conditioned area % from GFA	MAB heating costs [€]
Number of flats with 3 residents or more	Year of construction	
Average living area per flat (m2)	Year of last renovation (if available)	

[4] Electricity prices per kWh are from Nord Pool monthly for Estonia: https://data.nordpoolgroup.com/auction/day-ahead/prices?deliveryDate=2024-01-01&currency=EUR&aggregation=MonthlyAggregate&deliveryAreas=EE

Socio-economic data of MAB residents	Technical characteristics of buildings	Energy data
Number of residents living in the MAB		
Average living area per person (m2)		
Percentage of owners living in the MAB (%)		
Percentage of tenants living in the MAB (%)		
Working adults age 18-60 (%)		
Unemployed adults age18-60 (%)		
Pensioners above age 60 (%)		
Children below age 18 (%)		
Average salary per dwelling (€ gross)		
Flats with residents receiving social assistance (%)		

Note: MAB – multi-apartment building

Source: compiled by the authors

From Table 3 it can be seen that the tool was extended by including not only socio-economic characteristics of residents and energy consumption in 2025, but also technical characteristics of the buildings included into sample.

The dependent variable for this analysis was designated as the average annual heating costs per building, measured in euros. This choice was made due to the lack of available information regarding prior renovations and the self-production of electricity or heat, as well as the comprehensive nature of the data presented. The remaining variables outlined in Table 3 were utilized as independent variables. The methodology and procedural steps of the initial analysis, which was based on Estonian data, were replicated.

Correlation Analysis, Ordinary Least Squares (OLS), and LASSO Regression

For continues variables, Pearson and for categorical variables, Spearman correlation tests have been applied. This analysis aims to identify relationships between the independent variables, with the objective of determining which independent variables correlate with each other. Some of the independent variables correlated, therefore, they cannot be included into the model simultaneously. This is the case, for example, with correlation between average living area per person and average living area per building as well as correlation between average building cost of electricity for three months and average proportion of electricity costs in income.

Similar to Stegnar (2025), who studied "how average energy poverty, building age, income, and population density influence the RFPF, a measure of regional prioritization" (p. 26), the Ordinary Least Squares (OLS) technique is applied in this analysis as well.

OLS is used in Multiple Linear Regression (MLR) model with many predictors (independent variables) to explain a single outcome variable (dependent variable). The objective of MLR is to estimate the coefficients that minimize the difference between predicted and actual values of the dependent variable.

The general form of multiple linear equation is:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon \quad (2)$$

where:

- y - dependent variable
- $x_1, x_2, \ldots, x_n$ - independent variables (predictors)
- β_0 - intercept
- $\beta 1, \beta_2, \ldots, \beta_n$ - regression coefficients
- ε - random error term

A positive β coefficient means that an independent variable is positively associated with dependent variable, a negative sign shows reverse relations.

Subsequently, thirteen Multiple Linear Regression (MLR) models based on the Estonian data and sixteen models based on the data from Slovenia, Romania, Poland and Croatia were fitted to ascertain which predictors consistently demonstrated statistical significance across all models.

To further validate the findings obtained from the OLS regression, the Least Absolute Shrinkage and Selection Operator (LASSO) model was employed. This technique is particularly useful when multiple predictors are present, as it effectively identifies the most influential variables among them.

Tests for Multicollinearity and Heteroscedasticity

To evaluate the presence of multicollinearity within the models, the Variance Inflation Factor (VIF) was utilized. A VIF value ranging from 1 to 5 is generally considered acceptable, indicating that multicollinearity is not a significant concern. Additionally, the analysis included an examination of residuals via Q–Q plots (Quantile–Quantile plot) and Residuals vs. Fitted Values plots to assess linearity and homoscedasticity, as well as to verify the normality of the residuals. Finally, Robust Standard Errors were calculated to test for heteroscedasticity, ensuring that the assumptions underlying the regression analyses were adequately met.

In contrast to ranking approaches that depend on subjective evaluations and data collected via survey or overly simplistic weighting mechanisms, this statistical methodology significantly improves transparency and accuracy of the results relying on objective statistical data.

3.3. Alternative ranking: methodology and data collection process

To validate the effectiveness of the tool, an additional ranking exercise was developed to compare the prioritisation produced by the digital model with information collected directly from the field. The aim was not to create a completely separate model, but to generate an alternative ranking that retained some overlap with the original logic of the tool while placing greater emphasis on social vulnerability, building condition, and perceived renovation need. In this way, the validation was designed to test whether a stronger social perspective, combined with practical field-based knowledge, would lead to meaningful shifts in the prioritisation of buildings.

This validation approach reflects the intended role of the task: to assess the effectiveness of the digital tool by comparing its results with additional information not included in the original dataset and to use that comparison as a basis for considering possible refinements of the tool. More specifically, the exercise relied on the perceptions and practical knowledge of building-level stakeholders, like building managers or tenant representatives, in order to cross-reference the validity of the results produced by the digital tool and to assess whether socially relevant aspects could meaningfully affect building renovation priorities.

The main intended output of this exercise was an alternative ranking list for each pilot site, together with a structured comparison with the original ranking. This made it possible to examine whether individual buildings changed position once greater weight was given to social conditions and field-based knowledge, and whether the validation produced findings relevant for interpreting the original results and, where appropriate, informing future improvements of the digital tool.

The validation was designed as a quantitative exercise based on a short online survey completed for buildings already assessed through the digital tool. The survey was intentionally kept concise and operationally feasible, while still producing structured data that could be used to generate an alternative ranking for each pilot site.

The alternative ranking was generated through a weighted composite index. At the conceptual level, the weighting scheme was designed to reflect both the logic of the task and the wider objectives of the project, namely to place the greatest emphasis on indicators linked to energy poverty, energy vulnerability, and building-level renovation need, while still incorporating supporting technical and structural information. In this framework, the validation model was organised into three thematic blocks:

- **Theme 1: Vulnerability of Households and Energy Efficiency** – the determining layer of the model
- **Theme 2: Heating System and Residents' Renovation Readiness** – the supporting layer
- **Theme 3: Structural Characteristics of the Building** – the secondary differentiating layer

To allow variables from different question formats and scales to be combined in a single index, all scored items were transformed to a common 0–1 scale using min-max normalisation:

$$x_{norm} = \frac{x - x_{min}}{x_{min} - x_{max}} \tag{3}$$

This allowed variables measured through different response formats to be combined within a single index. The final composite score was calculated as a weighted sum of the three thematic scores:

$$TOTAL_{score} = 0.60 * T1_{score} + 0.25 * T2_{score} + 0.15 * T3_{score} \tag{4}$$

The resulting score ranges from 0 to 1, where 0 indicates lower renovation priority and 1 indicates higher renovation priority. The score was then used to create an alternative ranking list for each pilot site.

The weights reflect the intended hierarchy of the validation model. Theme 1 receives the highest weight because it captures the core validation logic: social vulnerability, building condition, complaints related to energy hardship, EPC-related information and the building manager's overall priority judgement. Theme 2 adds information on heating systems and residents' renovation readiness. Theme 3 has the lowest weight and mainly provides additional structural differentiation between buildings with otherwise similar scores.

Theme 1: Vulnerability of Households and Energy Efficiency

Theme 1 combines five components: the building's overall energy condition, the estimated share of socially or financially vulnerable households, complaints related to energy hardship, the EPC-related indicator, and the manager's overall judgement of the building's need for renovation.

The full-information version of the Theme 1 score is:

$$T1_{score} = \frac{Cond_n + Vulner_n + Compl_n + Class_n + 2 * Overall_n}{6} \tag{5}$$

where:

- $Cond_n$ is the normalised score for the building's overall energy condition;
- $Vulner_n$ is the normalised score for the estimated share of vulnerable households;
- $Compl_n$ is the normalised score for complaints related to energy hardship;
- $Class_n$ is the EPC-related indicator;
- $Overall_n$ is the normalised assessment of the building's overall renovation need.

The overall renovation-need item was weighted twice because it functions as a summary judgement after the respondent has reflected on the previous Theme 1 indicators. The denominator is therefore 6, reflecting four indicators with weight 1 and one indicator with weight 2.

The EPC-related component was handled conditionally. If a valid EPC existed, the reported energy class was converted into a 0–1 score, where A = 0 and G = 1. If no valid EPC existed, the EPC-related component was assigned a value of 0.67, equivalent to energy class E. If the respondent did not know whether a valid EPC existed, the EPC-related component was treated as unavailable factual information and excluded from the Theme 1 calculation.

When the EPC-related component is unavailable, Theme 1 is calculated using the remaining available components:

$$T1_{score} = \frac{Cond_n + Vulner_n + Compl_n + 2 * Overall_n}{5} \quad (6)$$

Theme 2: Heating System and Residents' Renovation Readiness

Theme 2 combines information on the building's main heating system and the manager's assessment of residents' overall interest in major energy renovation. The full-information version of the Theme 2 score is:

$$T2_{score} = \frac{Heat_n + Ready_n}{2} \quad (7)$$

The heating-system variable was recoded into broader priority categories before normalisation. Oil boilers and coal or other solid fossil fuels received the highest priority score. Gas-based systems, fossil-based district heating[5] and electric heating were treated as high-priority systems. Biomass was placed in the medium-priority category, while heat pumps and solar thermal were treated as the lowest-priority group.

If the respondent selected "Don't know" for the heating system, the item was treated as unavailable factual information and excluded from the Theme 2 calculation. In that case, Theme 2 was based only on the residents' readiness variable:

$$T2_{score} = Ready_n \quad (8)$$

The readiness item was treated as an experience-based judgement and captured the respondent's overall assessment of residents' interest in renovation.

Theme 3: Structural Characteristics of the Building

Theme 3 captures three structural characteristics: building age, number of flats and the building's position in relation to neighbouring buildings. The full-information version of the Theme 3 score is:

$$T3_{score} = \frac{Age_n + Flat_n + Pos_n}{3} \quad (9)$$

Older buildings, larger buildings and more exposed buildings received higher scores. Since these variables are factual building characteristics, "Don't know" responses were treated as unavailable information and excluded from the within-block average. If one or two items were unavailable, the Theme 3 score was calculated using the remaining available item(s). If all three structural items were unknown, Theme 3 was assigned a neutral value of 0.5 to preserve comparability across buildings.

Across the model, "Don't know" responses were included only for factual items where respondents could reasonably lack reliable information. For perception-based questions, the survey collected the respondent's informed assessment. The frequency of "Don't know" responses will be monitored by pilot site as part of the assessment of data completeness.

Data collection

[5] For Estonia, this answer category has been adjusted to just „district heating" and was recoded into the category of lowest priority as in Estonian example this is one of the most sustainable heating systems.

Data collection was organised through the project partners responsible for the pilot sites. Before implementation, partners participated in the WP4 – Tool validation workshop with Q&A, held in Ptuj on 17 March 2026. The training explained the purpose of the validation task, the structure of the survey, the scoring logic, and the practical steps needed to obtain building-level information from relevant stakeholders.

The training also focused on how to encourage participation of building-level stakeholders, particularly building managers and, where relevant, tenant or residents' representatives. Partners were advised to clearly explain the purpose of the survey, present the questions in advance where useful, and support respondents during completion if needed. This was especially important because the validation relied partly on practical knowledge and informed assessments that may not be available in administrative datasets.

Following the training, pilot partners were responsible for contacting relevant stakeholders, disseminating the EUSurvey link (see Annex 1), and supporting the completion of the questionnaire for buildings already ranked by the digital tool. The aim was to collect one completed survey per targeted building, while allowing partners to adapt the outreach process to their local context and stakeholder structure. Data collection was conducted between 31 March and 22 May 2026.

To document the implementation process, partners were also asked to complete a short reporting survey on stakeholder outreach and data collection (see Annex 2). This survey collects information on whether outreach was directed towards the buildings originally ranked by the digital tool, when and how the first contact was made, which stakeholder types were contacted, whether follow-up or additional outreach actions were used, and whether any supporting evidence of outreach was available. It also records the main challenges and external constraints encountered during data collection, as well as partners' reflections and recommendations for similar future activities.

The reporting survey shows that all pilot partners carried out outreach activities, but the intensity, channels and stakeholder structures differed considerably between pilot sites. In most cases, outreach was directed towards the buildings or addresses included in the original digital tool ranking, although the extent to which this resulted in usable building-level validation data varied and is addressed in more detail in the pilot-specific result sections below. First contact was made through different channels, including email, phone calls and in-person communication, depending on the local context and the availability of relevant stakeholders.

The types of stakeholders contacted also differed between pilot sites. In some cases, partners primarily approached building managers, while in others they relied more strongly on residents' representatives, homeowners' associations, municipal actors or other locally informed stakeholders. This was particularly important where building managers were difficult to reach, where not all buildings had a formal building manager, or where partners already had relevant local knowledge from previous project activities. In Alba Iulia, for example, data collection was supported through the technical department of the municipality and its contacts with homeowners' associations, while in Slovenia additional inputs were sought from building representatives and a municipal councillor familiar with the pilot area.

The main challenges reported by partners were linked to limited stakeholder availability, low responsiveness, lack of prior contact with building representatives, language barriers, and the difficulty of collecting reliable building-level information within a short timeframe. Several partners also noted that some building managers were more familiar with administrative or accounting aspects of building management than with energy-related topics, and that information on issues such as energy class, technical building characteristics, or residents' subjective experiences was not always readily available. These implementation differences are important for interpreting the validation results, as they help explain why some pilot sites achieved full coverage while others produced only partial validation datasets or, in the case of Croatia, no usable survey responses.

4. Results and discussion

4.1. Results of the statistical tool development

The results of MLR for Estonia are presented in Table 4. All together 13 models were fitted based on the Estonian data. Positive and negative signs in front of the coefficients indicate whether factors are positively or negatively associated with the dependent variable, which is in this case average difference between electricity consumption and production of buildings in January, February and March 2024. For the future studies, this variable can be made more complex. For example, by including not only electricity consumption but also heating and gas consumption (if the data is available). The magnitude of coefficients in MLR can also be interpreted.

Additionally, Table 4 includes Multiple R-squared and Adjusted R-squared for each of the models and used to characterised the goodness of fit, such as how well the variations in dependent variable can be explained by suggested independent variables. *P*-values for each factor are represented on 0.1, 1, 5 and 10% significance levels. Standard errors are included in parentheses.

LASSO (Least Absolute Shrinkage and Selection Operator) technique further tunes the results of OLS by removing coefficients that become equal to zero, this technique also reduces overfitting (Ranstam & Cook, 2018). In this way, only the most important variables are selected. LASSO is used when there are quite many predictors and not so many observations (Ranstam & Cook, 2018) as it is in our dataset. The factors from 13 MLR models were further included into LASSO model to verify the outcome. Table 5 displays the factors which stayed statistically significant and which coefficients are different from zero. Average share of electricity costs in income (energy poverty component) was quite substantially correlating with the dependent variable and, therefore, it was removed from the model.

Such factors as the number of households with 1, 3 and 5 members and the number of people 15–65 and over 65-year-old are statistically significant and stayed after LASSO has been applied. However, among these factors only one with the largest magnitude from each group was selected for calculation of weighted scores (see formula 10).

Table 4. The results of regression analysis for Estonia

Factors	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12	Model 13
Intercept	9.2604*** (0.4257)	9.2294*** (0.4023)	8.7142*** (0.4046)	8.3955*** (0.4140)	8.4762*** (0.4261)	8.4646*** (0.4115)	8.5873*** (0.4241)	8.7685*** (0.4308)	8.9909*** (0.4407)	8.9507*** (0.4688)	8.8060*** (0.4322)	7.800*** (0.2643)	7.514*** (0.3463)
Median salary	-0.0005* (0.0002)	-0.0005* (0.0002)	-0.0005* (0.0002)	-0.0004· (0.0002)	-0.0003 (0.0002)	-0.0002 (0.0002)	-0.0003 (0.0002)	-0.0006** (0.0002)	-0.0004· (0.0002)	-0.0008** (0.0002)	-0.0007** (0.0002)	-0.0002 (0.0001)	0.00002 (0.0002)
Living area per person	0.1027 (0.0965)	0.1012 (0.0943)	0.0588 (0.0908)	0.0627 (0.0888)	0.0122 (0.0927)	0.0097 (0.0923)	0.0509 (0.0919)	0.0603 (0.0973)	-0.0123 (0.0976)	0.2016· (0.1064)	0.2061· (0.1097)	0.0523 (0.0586)	0.0634 (0.0663)
Part of Tartu	-0.8665*** (0.1454)	-0.8673*** (0.1386)	-0.6388*** (0.1438)	-0.5520*** (0.1451)	-0.4995** (0.1600)	-0.6393*** (0.1552)	-0.6338*** (0.1477)	-0.4654** (0.1514)	-0.7120*** (0.1472)	-0.5666*** (0.1496)	-0.4783** (0.1764)	-0.2809** (0.0931)	-0.3313** (0.1026)
Owner	-0.0004 (0.0042)												
Renter		0.0010 (0.0054)	0.0024 (0.0051)	0.0023 (0.0050)	0.0033 (0.0052)	-0.0011 (0.0051)	0.0026 (0.0052)	0.0042 (0.0042)	0.0050 (0.0057)	0.0029 (0.0076)	0.0046 (0.0057)	0.0017 (0.0033)	0.0029 (0.0037)
Age under 15			0.0154*** (0.0038)										
Age 15-65				0.0049*** (0.0010)									
Age over 65					0.0105*** (0.0026)								
Households with 1 member						0.0155*** (0.0034)							

Factors	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12	Model 13
Households with 2 members							0.0178*** (0.0047)						
Households with 3 members								0.0500*** (0.0109)					
Households with 4 members									0.0328** (0.0105)				
Households with 5 members										0.0697** (0.0247)			
Households with 6 members											0.1088** (0.0383)		
Average cost of electricity												0.0015*** (0.0001)	
Average share of electricity costs in income													1.958*** (0.1376)
Multiple R-squared	0.3486	0.3487	0.4078	0.4298	0.4079	0.429	0.4025	0.4521	0.3685	0.4144	0.4747	0.7505	0.7045
Adjusted R-squared	0.3333	0.3334	0.3903	0.413	0.3903	0.4119	0.3848	0.435	0.3479	0.3908	0.4445	0.7432	0.6958

Note: ***p < 0.001, **p < 0.01, *p < 0.05, ·p < 0.1; Standard Error in parentheses; part (district of the city) is coded as 0 – Annelinn and 1 – Karlova

Source: compiled by the authors

Table 5. LASSO model for Estonia

Significant factors	**LASSO coefficient**	**Weights**
Median salary (gross)	-0.0004512478	0.1
District of Tartu	-0.2222928523	0.4
Number of households with 1 member	0.0024145407	-
Number of households with 3 members	0.0432284071	0.3
Number of households with 5 members	0.0177024871	-
Number of people 15–65-year-old	0.0005288869	-
Number of people over 65-year-old	0.0007886124	0.2

Note: Weights are selected based on the magnitude of LASSO coefficients

Source: compiled by the authors

Calculation of a weighted score:

$$\text{Weighted score}_{\text{Estonia}} = -part * 0.4 + hhold3\_std * 0.3 + Over65\_std * 0.2 - msalary\_std * 0.1 \quad (10)$$

where

- part - district of Tartu
- hhold3_std – standardised number of households with 3 members
- Over65_std – standardised number of people over 65-year-old
- msalary_std - standardised median gross salary

Standardisation of variables is needed since they have different scales and variables with large values may dominate and affect the final score.

The interpretation of the final scores indicates that buildings with the lowest weighted scores should be prioritized for renovation. Figure 4 illustrates the ranking results generated by the R programming language, wherein the buildings depicted in red at the bottom of the figure represent a more urgent need for renovation compared to those shown in orange at the top.

The findings derived from both Multiple Linear Regression (MLR) and Least Absolute Shrinkage and Selection Operator (LASSO) analyses indicate that an increase in median salary is associated with a reduction in the average difference between electricity consumption and production. This observation suggests that renovated residences equipped with electricity-generating systems, such as installed photovoltaic panels, are likely to be inhabited by individuals with higher income levels. This is in line with Harvard Joint Center for Housing Studies (2022) as well as Streimikiene and Balezentis (2020) confirming that income can be a substantial barrier to renovation readiness. Household income is also positively associated with the likelihood of having performed renovations according to Plaut and Plaut (2010).

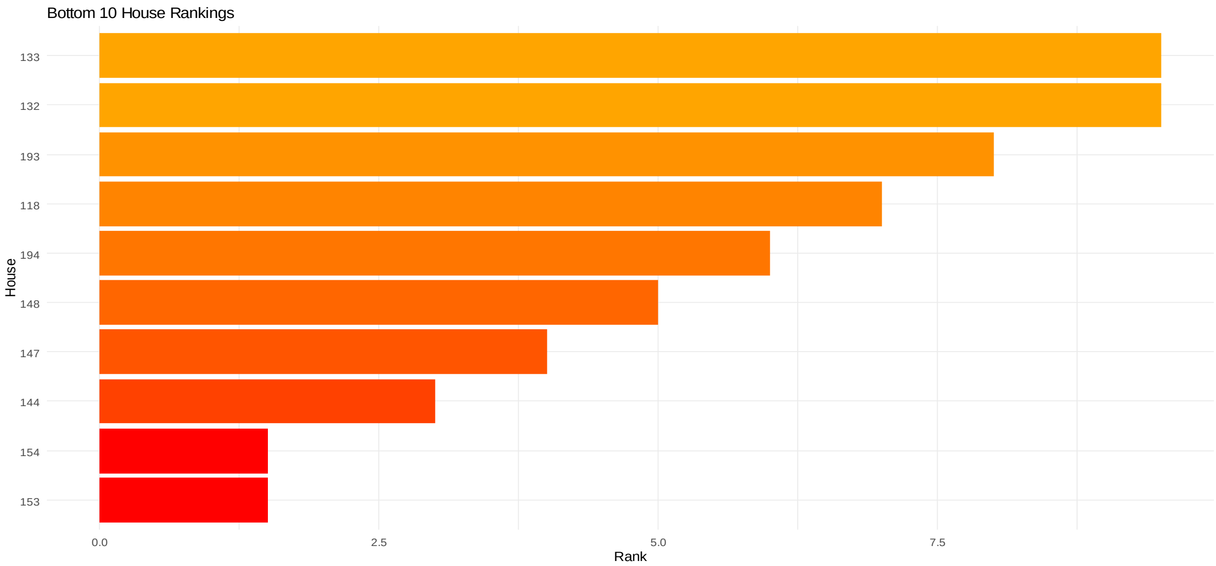


Figure 4. Exemplary outcome of houses' ranking in R programme

Source: compiled by the authors

District of Tartu is another statistically significant factor. Comparing with *Annelinn*, in *Karlova* the average difference between electricity consumption and production is decreasing. This might indicate, for example, that in *Karlova* there are more renovated houses or houses with a higher energy efficiency compared to *Annelinn*. Similar disparities can exist not only on municipal, but also on a regional level as, for example, Stegnar (2025) identified in Slovenia.

Number of households with one, three and five member(s) is positively associated with the average difference between electricity consumption and production. With one, three and five household members this difference is increasing meaning that most probably these households live in unrenovated houses without electricity self-production. Some of the previous studies also include household composition as a factor for renovation analysis. For example, Baker and Kaul (2000) found out that changes in household composition, in particular adding a child, positively impact renovation expenditures. Similarly, Curtis, Grilli, and Lynch (2024) identified associations between changes in family composition—such as childbirth, the arrival or departure of adult family members, or the death of a family member—and households' renovation decisions. Moreover, Drescher and Janzen (2021) have concluded that "being a single parent or living alone significantly increases the chances of facing energy poverty" (p. 10). In the context of our analysis, households consisting of a single member may face insufficient financial resources to undertake renovation projects. Furthermore, the elevated risk associated with securing a loan for such renovations poses a significant barrier to initiating these improvements for single individuals. At the same time, larger households with three and five members might face different obstacles, for example, directing financial resources to other non-housing purposes (e.g. growing needs of children and elderly family members).

The findings of this study have also indicated that as age increases, the average difference between electricity consumption and production also increases. This observation is consistent with the conclusions drawn by Plaut and Plaut (2010), who suggest that older households exhibit a lower preparedness for renovations. Based on these results, renovation policies in Estonia should be focused on specific demographic groups that exhibit greater vulnerability, such as retirees, single individuals, and large families.

The outcomes of the regression analysis, which involved the evaluation of sixteen multiple linear regression (MLR) models alongside the LASSO technique, are delineated in Table 6. This analysis encompasses not only the socio-economic characteristics of the building occupants but also the technical attributes of the structures themselves. Consequently, a greater number of statistically significant factors can be incorporated into the overall findings.

Table 6. LASSO model for Croatia, Poland and Romania

Significant factors	**LASSO coefficient**	**Weights***
Average monthly salary per building	0.000002101748	0.05
Country	0.5015637	0.4
Percent of unemployed adults age 18-60	0.02386472	0.15
Number of (mostly) empty flats with 0 residents	0.06683785	0.3
Year of construction	0.006892765	0.1

Note: Weights are selected based on the magnitude of LASSO coefficients; countries are coded as 1 – Croatia, 2 – Poland, 3 – Romania, 4 - Slovenia

Source: compiled by the authors

The analysis identifies several statistically significant factors influencing the average annual heating costs per building, as evidenced by data collected from Croatia, Poland, Slovenia and Romania. These factors include the average monthly salary, the country of residence, the percentage of unemployed adults aged 18 to 60, the number of predominantly vacant dwellings, and the year of construction. Notably, all identified factors exhibit a positive correlation with heating costs.

The variable "country" presents a complex interpretative challenge due to its categorical nature; however, it is evident that there are substantial disparities in annual heating costs across different countries. This observation aligns with regional variations noted by Stegnar (2025) in the context of a specific country.

Furthermore, a positive relationship is observed between average monthly salary and heating costs. This finding corroborates previous research, such as that of Zarco-Periñán et al. (2021), which established that higher income levels among residents correlate with increased energy consumption for heating and, consequently, greater emissions within urban environments. Similarly, Meier and Rehdanz (2010) identified income as a significant factor influencing heating expenditures in households across Great Britain. This trend can be attributed to the tendency of individuals with higher incomes to utilize more heating services and inhabit larger or more intensively heated living spaces, thereby elevating their heating costs.

The analysis also reveals that a higher unemployment rate within a building correlates with increased heating costs per building. This may suggest that unemployed residents possess limited financial resources to invest in energy-efficient improvements (Drescher & Janzen, 2021) or that they tend to spend more time at home, resulting in higher heating demands (Kearns, Whitley, & Curl, 2019).

Additionally, the positive coefficient associated with the number of vacant apartments indicates that buildings with a higher proportion of unoccupied dwellings tend to incur greater annual heating costs. This phenomenon can be explained by the reduced internal heat gain from vacant units, which may be maintained at lower temperatures, thereby increasing heat transfer from adjacent occupied spaces. Previous studies have highlighted the significant impact of occupancy patterns and the presence of unoccupied flats on the energy performance and heating demand of multi-residential buildings (Oliveira et al., 2020).

Moreover, the results of the multiple linear regression analysis indicate that more recently constructed buildings (as indicated by a higher year of construction) exhibit higher annual heating costs, while controlling for other variables. Although it is generally anticipated that newer buildings demonstrate greater energy efficiency, this finding may reflect variances in building size, occupancy patterns, thermal comfort expectations, and actual operational performance. Prior research has indicated that the relationship between building age and energy consumption is multifaceted, influenced by both the physical characteristics of the building and occupant behavior (Aksoezen et al., 2015). It is also plausible that newer buildings consume more energy in absolute terms if they offer significantly larger heated floor areas or enhanced levels of thermal comfort (Balaras et al., 2005). Consequently, the year of construction should not be interpreted solely as an indicator of energy efficiency; rather, it should be viewed as a characteristic associated with various building attributes that collectively influence heating expenditures.

Calculation of a weighted score for Croatia, Poland and Romania are reflected by formula (11):

$$\text{Weighted score}_{\text{Croatia, Poland and Romania}} = \text{country} * 0.4 + \text{resident0\_std} * 0.3 + \text{unemployed\_std} * 0.15 + \text{year\_std} * 0.1 + \text{bavsalary\_std} * 0.05 \quad (11)$$

where

- country – code for Croatia, Poland, Slovenia and Romania,
- resident0_std – standardised number of (mostly) empty dwellings with 0 residents,
- unemployed_std – standardised number of unemployed adults 18-60 of age (%),
- year_std – standardised year of building construction,
- bavsalary_std – standardised average monthly salary per building (€ gross).

For Slovenia, factor "unemployed" has to be excluded from the analysis, since no observations for this factor are available. Table 7 shows the results of LASSO simulation and distribution of weights for each factor based on the magnitude of LASSO coefficients.

Table 7. LASSO model for Slovenia

Significant factors	LASSO coefficient	Weights*
Average monthly salary per building	0.000002101748	0.1
Country	0.5015637	0.4
Number of (mostly) empty dwellings with 0 residents	0.06683785	0.3
Year of construction	0.006892765	0.2

Note: Weights are selected based on the magnitude of LASSO coefficients

Source: compiled by the authors

Calculation of a weighted score:

$$\text{Weighted score}_{\text{Slovenia}} = \text{country\_std} * 0.4 + \text{resident0\_std} * 0.3 + \text{year\_std} * 0.2 + \text{bavsalary\_std} * 0.1 \quad (12)$$

where

- country – code for Croatia, Poland, Slovenia and Romania,
- resident0_std – standardised number of (mostly) empty dwellings with 0 residents,
- year_std – standardised year of building construction,
- bavsalary_std – standardised average monthly salary per building (€ gross).

Each building collects points based on formula (11) and (12) and the building with smallest number of points should be renovated first/earlier. The outcome of application of the code to the building level data from four countries is presented in Table 8.

Table 8. Ranking of the buildings in Romania, Poland, Croatia and Slovenia presented in the order of needed to be renovated first

Order in the model	Order for renovation	Address of the building
Romania		
75	1	Ampoiului 9-17
89	2	Orizontului 1-13
74	3	V.Goldiş 26
68	4	Livezii 46 Alba Iulia
83	5	Transilvaniei 27
67	6	Orizontului 8-16 Alba Iulia
76	7	Gh. Sincai 17-21

Order in the model	Order for renovation	Address of the building
79	8	Targului 1
87	9	V.Goldiş 14-14a
86	10	V.Goldiş 12-12a
88	11	V.Goldiş 16-16a
73	12	Transilvaniei 23b
92	13	Livezii 49 Alba Iulia
80	14	Cloșca 8
85	15	V.Goldiş 10-10a
63	16	A.I. Cuza 16 Alba Iulia
81	17	Cloșca 1
66	18	Gh. Şincai 27-31 Alba Iulia
78	19	Orizontului 4-6
82	20	Cloșca 5-7
65	21	Transilvaniei 8-10 Alba Iulia
77	22	Gh. Şincai 23-25
84	23	T.Vladimirescu 34
69	24	Transilvaniei 2a
70	25	Transilvaniei 4
71	26	Transilvaniei 6
90	27	V.Alecsandri 76
91	28	Craviei 2-4
64	29	Closca 10 alba iulia
72	30	Transilvaniei 14
Poland		
57	1	Szekspira 2
50	2	Dantego 7
59	3	Tołstoja 1

Order in the model	Order for renovation	Address of the building
43	4	Balzaka 2
55	5	Reymonta 21
34	6	Wolumen 6, 01-912 Warsaw
39	7	Andersena 2
53	8	Petofiego 6
36	9	Szekspira 4, Warsaw
33	10	Sokratesa 2b, 01-909 Warsaw
44	11	Czechowa 2
35	12	Wolumen 4, 01-911 Warsaw
37	13	Aleja Władysława Reymonta 23, 01-840 Warsaw
42	14	Andersena 6
62	15	Hansa Christiana Andersena 8
60	16	Tołstoja 3
61	17	Tołstoja 4
40	18	Andersena 3
41	19	Andersena 5
49	20	Dantego 5
46	21	Dantego 1a
52	22	Petofiego 4
51	23	Petofiego 2
54	24	Petofiego 8
45	25	Dantego 1
47	26	Dantego 1b
48	27	Dantego 3
56	28	Szekspira 1
58	29	Szekspira 3
38	30	Sándora Petöfiego 1, 01-917 Warsaw

Order in the model	Order for renovation	Address of the building
Croatia		
21	1	Vukovarska 11, 40000 Čakovec
32	2	Ivana Zajca 11, 40000 Čakovec
29	3	Vukovarska 13, 40000 Čakovec
22	4	Jakova Gotovca 1, 40000 Čakovec
24	5	Jakova Gotovca 1/b, 40000 Čakovec
31	6	Otokara Keršovanija 6, 40000 Čakovec
12	7	Kolodvorska 4, 40000 Čakovec
18	8	Vukovarska 5, 40000 Čakovec
20	9	Vukovarska 9, 40000 Čakovec
19	10	Vukovarska 7, 40000 Čakovec
30	11	Otokara Keršovanija 4, 40000 Čakovec
28	12	Miroslava Magdalenića 1/b, 40000 Čakovec
14	13	Preloška 68, 40000 Čakovec
27	14	Miroslava Magdalenića 1, 40000 Čakovec
25	15	Katarine Zrinski 4, 40000 Čakovec
5	16	Janka Slogara 4, 40000 Čakovec
2	17	Travnik 12, 40000 Čakovec
11	18	Valenta Morandinija 17, 40000 Čakovec
6	19	Istarska 14, 40000 Čakovec
23	20	Jakova Gotovca 1/a, 40000 Čakovec"
26	21	Kolodvorska 1, 40000 Čakovec
17	22	Vladimira Nazora 26a, 40000 Čakovec
15	23	Tome Masaryka 13, 40000 Čakovec
7	24	Kralja Tomislava 3, 40000 Čakovec
10	25	J.J. Strossmayera 5, 40000 Čakovec
4	26	Josipa Jurja Strossmayera 7b, 40000 Čakovec

Order in the model	Order for renovation	Address of the building
3	27	Istarska 16, 40000 Čakovec
8	28	Kralja Tomislava 29, 40000 Čakovec
1	29	Vladimira Nazora 32, 40000 Čakovec
16	30	Vladimira Nazora 8a, 40000 Čakovec
9	31	J.J. Strossmayera 1, 40000 Čakovec
13	32	Dr. Ivana Novaka 6, 40000 Čakovec
Slovenia		
98	1	Ul 25 Maja 9
107	2	Kajuhova ulica 7
94	3	Kajuhova ul.11
97	4	Arbajterjeva 3
122	5	Mladinska ulica 3,4,5
96	6	Ul. Borisa Kraigherja 20,22,24,26
109	7	Kajuhova ulica 13
93	8	Čučkova 7
99	9	Lackova 7
113	10	Čučkova 9
117	11	Vlahovičeva ulica 9
118	12	Vlahovičeva ulica 7
108	13	Kajuhova ulica 12
110	14	Lackova 5
114	15	Čučkova 11
116	16	Čučkova 5
104	17	Ul. Borisa Kraigherja 13, 15, 17
111	18	Kajuhova ulica 14
120	19	Vlahovičeva ulica 3
121	20	Vlahovičeva ulica 1

Order in the model	Order for renovation	Address of the building
112	21	Kajuhova ulica 15
115	22	Čučkova 13
105	23	Ul. Borisa Kraigherja 19, 21, 23
101	24	Ul. Borisa Kraigherja 8,10,12
95	25	Ul. Borisa Kraigherja 7,9,11
100	26	Ul. Borisa Kraigherja 14 16,18
103	27	Ul. Borisa Kraigherja 1, 3,5
119	28	Vlahovičeva ulica 5
102	29	Ul. Borisa Kraigherja 2,4,6
106	30	Ul. Borisa Kraigherja 25, Kajuhova ulica 4

Source: compiled by the authors

To assess the validity and reliability of the fitted regression models, diagnostic analyses were conducted, including Q-Q plots of the residuals and plots of residuals against fitted values. These analyses were performed for all models utilizing data from Estonia and four partner countries. A Q–Q plot (Quantile–Quantile plot) of residuals has been used to check whether the residuals from regression models are approximately normally distributed. Observed quantiles have been quite closely located to the theoretical ones (a straight diagonal line), therefore, normality assumption is reasonably satisfied.

Another diagnostic test that has been conducted is Residuals vs. Fitted Values plot. It is used to check the assumptions of linearity and homoscedasticity (constant variance of residuals). Our results have shown that residuals have been randomly scattered around zero line, not demonstrating a particular pattern.

Additionally, to detect multicollinearity among predictor variables Variance Inflation Factor (VIF) has been applied. In all the models based on the Estonian data VIF values for all predictors were below 2, which indicates no concern regarding correlation of independent factors. Among sixteen models tested on the data from Slovenia, Poland, Romania and Croatia, there have been VIF values higher than 5 and even 10 which is a sign of severe multicollinearity. This issue was resolved by removing one of the highly correlated predictors.

Limitations of the Tool and Recommendations for Improvement

The current analysis is constrained by several limitations that should be considered for future improvements and current usage of the tool. Firstly, the dataset utilized is relatively small, comprising only 194 observations for Estonia and 122 observations for Romania, Slovenia, Croatia, and Poland together. This limited sample size may affect the robustness and generalizability of the findings. Secondly, the incorporation of longitudinal data spanning a period of at least five years would be advantageous.

Longitudinal data would enable a more comprehensive understanding of trends and changes over time, thereby enhancing the analytical depth of the tool.

In the case of the Estonian dataset, there is an absence of additional variables that could provide valuable insights. Specifically, factors such as household social support and the technical characteristics of buildings are not included, because some data was not available, and technical characteristic of the buildings cannot be merged with 100x100 aggregated dataset, even though energy class and buildings' construction year are publicly available data which can be accessed via the Estonian Building Register.

Moreover, the current dataset fails to account for seasonality, as it only encompasses the first quarter of 2024 (January, February, and March) for Estonia, along with a single year of data for the other four countries. This temporal limitation restricts the ability to analyse seasonal variations that may impact the results. Additionally, data on gas consumption was not available for all buildings within the dataset, and was not used in the tool.

In the context of Estonia, there exists a discrepancy in the levels of aggregation between energy data and socio-economic data. This misalignment necessitates specialized expertise to effectively merge the datasets, which limits usability of the tool. Lastly, the Slovenian sample is incomplete, which may undermine the validity of the findings related to that specific context.

Addressing these limitations through the acquisition of a more extensive and diverse dataset, the inclusion of longitudinal data, and the integration of additional relevant variables would significantly improve the analytical rigor and applicability of the tool. Furthermore, the application of more sophisticated statistical models could be facilitated by a richer dataset, ultimately leading to more nuanced insights.

4.2. Results of the alternative ranking

Across the four pilot sites (Romania, Poland, Croatia and Slovenia), the digital tool initially prioritised 122 buildings or address/building units, with approximately 30 units included in each pilot ranking. For the Estonian case, alternative ranking is possible with some limitations due to specificities of more aggregated 100x100 dimensional data which can contain more than one building. Therefore, it has been decided not to include Estonian pilot site into the alternative ranking because the results would be not accurate.

Field-based validation data suitable for comparison with the digital tool ranking were obtained for 76 units, corresponding to 62.3% of the originally prioritised sample. Coverage differed substantially between pilot sites. The Polish pilot site in Warsaw and the Romanian pilot site in Alba Iulia both achieved full coverage, with 30 of 30 prioritised buildings validated in each case (100%). In the Slovenian pilot site, covering the Municipality of Kidričevo and the Municipality of Ptuj, 16 of 30 original digital tool units remained available for comparison after data cleaning (53.3%). In the Croatian pilot site in Čakovec, despite documented outreach efforts and the engagement of project partners, no completed validation surveys were collected for any of the prioritised buildings. The following sub-sections present the validation results for the three pilot sites where comparative analysis between the validation ranking and the original digital tool ranking could be conducted.

Warsaw (PL)

For the Polish pilot site in Warsaw, the validation dataset covers all 30 buildings included in the original digital tool ranking. Compared with some other pilot sites, the dataset shows a moderate level of variation, although several indicators are relatively uniform across the sample. Most buildings were assessed as being in fair or good overall energy condition, while only four were assessed as poor. Social vulnerability was assessed as either low or medium, with no buildings placed in the high vulnerability category. Complaints related to energy hardship were mostly reported as isolated or occasional cases, while seven buildings had no reported complaints. All buildings use district heating as the main heating system, and most were built between 1971 and 1990. Structural characteristics are also relatively similar, particularly building age and position, while more variation is visible in building size, EPC class, residents' renovation readiness and the general assessment of renovation need. As a result, the validation ranking (see Table 9) includes some tied ranks, but still provides sufficient differentiation between buildings for comparison with the digital tool ranking.

Table 9. Comparison of validation („valid.“) ranking and digital tool ranking for the Polish pilot site of Warsaw

WARSAW	**Score (valid.) (0-1)**	**Rank (valid.)**	**Rank (tool)**	**Rank shift**	**Absolute diff.**
Szekspira 4, Warsaw	0,668	1,5	4	2,5	2,5
Andersena 5	0,668	1,5	9	7,5	7,5
Dantego 1A	0,622	3	14	11	11
Andersena 2	0,604	5	7	2	2
Reymonta 21	0,604	5	23	18	18
Tołstoja 1	0,604	5	27	22	22
Andersena 3	0,593	7	8	1	1
Szekspira 3	0,592	8	26	18	18
Wolumen 4, 01-911 Warsaw	0,591	9	3	-6	6
Szekspira 1	0,569	10	24	14	14
Petofiego 6	0,558	11	21	10	10
Balzaka 2	0,554	12	11	-1	1
Dantego 7	0,549	13,5	18	4,5	4,5
Szekspira 2	0,549	13,5	25	11,5	11,5
Andersena 6	0,538	15	10	-5	5
Czechowa 2	0,53	16	12	-4	4
Dantego 1	0,525	17	13	-4	4
Wolumen 6, 01-912 Warsaw	0,524	18	2	-16	16
Petofiego 2	0,513	19	19	0	0
Sokratesa 2b, 01-909 Warsaw	0,486	20	1	-19	19
Aleja Władysława Reymonta 23, 01-840 Warsaw	0,465	21	5	-16	16
Dantego 1b	0,463	22	15	-7	7
Sándora Petőfiego 1, 01-917 Warsaw	0,461	23	6	-17	17
Hansa Christiana Andersena 8	0,454	24	30	6	6
Tołstoja 4	0,433	25	29	4	4

Dantego 5	0,424	26	17	-9	9
Petofiego 4	0,399	27	20	-7	7
Tołstoja 3	0,368	28	28	0	0
Dantego 3	0,365	29,5	16	-13,5	13,5
Petofiego 8	0,365	29,5	22	-7,5	7,5

Source: compiled by the authors

For the Polish pilot site, the comparison between the digital tool ranking and the validation ranking shows partial overlap, but also several substantial shifts in building prioritisation. The mean absolute rank difference was 8.8 positions, while the median difference was 7.25 positions. Individual absolute differences ranged from 0 to 22 positions. The overall rank-order association between the two rankings was weak and not statistically significant (Spearman's $\rho = 0.222$, $p = 0.238$), suggesting that the validation exercise did not simply reproduce the original tool ranking, but introduced a meaningfully different prioritisation pattern. Table 10 illustrates these results.

Table 10. Categorised agreement and direction of priority shifts for the Polish pilot site of Warsaw

WARSAW	Rank (valid.)	Rank (tool)	Agreement level	Valid. priority shift
Szekspira 4, Warsaw	1,5	4	High agreement / small difference	↖
Andersena 5	1,5	9	Moderate difference	↖
Dantego 1A	3	14	Large difference	↖
Andersena 2	5	7	Very high agreement	↖
Reymonta 21	5	23	Very large difference	↖
Tołstoja 1	5	27	Very large difference	↖
Andersena 3	7	8	Very high agreement	↖
Szekspira 3	8	26	Very large difference	↖
Wolumen 4, 01-911 Warsaw	9	3	Moderate difference	↘
Szekspira 1	10	24	Large difference	↖
Petofiego 6	11	21	Moderate difference	↖
Balzaka 2	12	11	Very high agreement	↘
Dantego 7	13,5	18	High agreement / small difference	↖
Szekspira 2	13,5	25	Large difference	↖
Andersena 6	15	10	High agreement / small difference	↘
Czechowa 2	16	12	High agreement / small difference	↘
Dantego 1	17	13	High agreement / small difference	↘
Wolumen 6, 01-912 Warsaw	18	2	Very large difference	↘
Petofiego 2	19	19	Very high agreement	=
Sokratesa 2b, 01-909 Warsaw	20	1	Very large difference	↘
Aleja Władysława Reymonta 23, 01-840 Warsaw	21	5	Very large difference	↘

Dantego 1b	22	15	Moderate difference	↘
Sándora Petöfiego 1, 01-917 Warsaw	23	6	Very large difference	↘
Hansa Christiana Andersena 8	24	30	Moderate difference	↖
Tołstoja 4	25	29	High agreement / small difference	↖
Dantego 5	26	17	Moderate difference	↘
Petofiego 4	27	20	Moderate difference	↘
Tołstoja 3	28	28	Very high agreement	=
Dantego 3	29,5	16	Large difference	↘
Petofiego 8	29,5	22	Moderate difference	↘

*Colours indicate the size of the absolute difference between rankings: dark green = 0–2 positions, light green = >2–5, lightest red = >5–10, darker red = >10–15, and darkest red = >15. The same colour scheme is used in all categorised agreement tables. In the far-right column ↖ indicates higher priority in the validation ranking, ↘ lower priority in the validation ranking, and = no change in rank.

Source: compiled by the authors

The categorised comparison further shows that five buildings remained within a two-position difference between the two rankings, while six buildings showed a small difference of more than two and up to five positions. Moderate differences of more than five and up to ten positions were observed for eight buildings. At the same time, eleven buildings changed by more than ten positions, including seven buildings with very large differences of more than fifteen positions. The direction of change was balanced. Fourteen buildings were ranked as higher priorities in the validation exercise than in the digital tool ranking, fourteen were ranked lower, and two retained the same rank. This indicates that the inclusion of social vulnerability, perceived renovation need and practical building-level knowledge can substantially affect the relative position of individual buildings in the priority list.

Alba Iulia (RO)

For the Alba Iulia pilot site, the validation dataset covers the same 30 buildings included in the digital tool ranking. The data show a high degree of uniformity across several variables relevant for the validation score. All buildings have the same main heating system and high resident interest in major energy renovation. In terms of key vulnerability indicators, almost all buildings were assessed as having a medium share of socially or financially vulnerable households, while complaints related to energy hardship were generally reported only as isolated cases. Limited variation is also visible in building age and building position, while more differentiation is found in building size, EPC class, overall energy condition and the general assessment of renovation need. Table 11 represents alternative ranking scores for Romania.

Table 11. Comparison of validation („valid.“) ranking and digital tool ranking for the Romanian pilot site of Alba Iulia

ALBA IULIA	Score (valid.) (0-1)	Rank (valid.)	Rank (tool)	Rank shift	Absolute diff.
LIVEZII 46 ALBA IULIA	0,790	1	6	5	5
A.I. CUZA 16 ALBA IULIA	0,680	2	1	-1	1

V.GOLDIȘ 10-10A	0,660	4,5	23	18,5	18,5
CLOȘCA 8	0,660	4,5	18	13,5	13,5
V.GOLDIȘ 16-16A	0,660	4,5	26	21,5	21,5
V.GOLDIȘ 12-12A	0,660	4,5	24	19,5	19,5
CLOȘCA 1	0,660	7,5	19	11,5	11,5
TRANSILVANIEI 27	0,660	7,5	21	13,5	13,5
CLOSCA 10 ALBA IULIA	0,650	9	2	-7	7
TRANSILVANIEI 8-10 ALBA IULIA	0,650	11	3	-8	8
GH. ȘINCAI 27-31 ALBA IULIA	0,650	11	4	-7	7
TARGULUI 1	0,650	11	17	6	6
V.GOLDIȘ 14-14A	0,640	13	25	12	12
ORIZONTULUI 8-16 ALBA IULIA	0,640	14	5	-9	9
CLOȘCA 5-7	0,630	15,5	20	4,5	4,5
ORIZONTULUI 1-13	0,630	15,5	27	11,5	11,5
V.GOLDIȘ 26	0,610	17	12	-5	5
TRANSILVANIEI 14	0,610	18,5	10	-8,5	8,5
T.VLADIMIRESCU 34	0,610	18,5	22	3,5	3,5
TRANSILVANIEI 23B	0,600	20,5	11	-9,5	9,5
GH. ȘINCAI 17-21	0,600	20,5	14	-6,5	6,5
CRAVIEI 2-4	0,590	22,5	29	6,5	6,5
V.ALECSANDRI 76	0,590	22,5	28	5,5	5,5
GH. ȘINCAI 23-25	0,590	24	15	-9	9
LIVEZII 49 ALBA IULIA	0,580	25	30	5	5
ORIZONTULUI 4-6	0,580	26	16	-10	10
AMPOIULUI 9-17	0,550	27	13	-14	14
TRANSILVANIEI 6	0,540	29	9	-20	20
TRANSILVANIEI 4	0,540	29	8	-21	21
TRANSILVANIEI 2A	0,540	29	7	-22	22

Source: compiled by the authors

For the Alba Iulia pilot site, the comparison between the digital tool ranking and the validation ranking shows considerable changes in the relative prioritisation of buildings. The mean absolute rank difference was 10.5 positions, while the median difference was 9 positions. Individual differences ranged from 1 to 22 positions. The validation ranking also includes several tied ranks, reflecting the relatively limited variation in several input variables used for the validation score. The overall rank-order association between the two rankings was very weak and not statistically significant (Spearman's $\rho = 0.035$, $p = 0.852$), indicating that the validation exercise produced a substantially different prioritisation pattern from the original digital tool ranking. Table 12 illustrates these patterns.

Table 12. Categorised agreement and direction of priority shifts for the Romanian pilot site of Alba Iulia

ALBA IULIA	**Rank (valid.)**	**Rank (tool)**	**Agreement level**	**Valid. priority shift**

LIVEZII 46 ALBA IULIA	1	6	High agreement / small difference	↖
A.I. CUZA 16 ALBA IULIA	2	1	Very high agreement	↘
V.GOLDIŞ 10-10A	4,5	23	Very large difference	↖
CLOŞCA 8	4,5	18	Large difference	↖
V.GOLDIŞ 16-16A	4,5	26	Very large difference	↖
V.GOLDIŞ 12-12A	4,5	24	Very large difference	↖
CLOŞCA 1	7,5	19	Large difference	↖
TRANSILVANIEI 27	7,5	21	Large difference	↖
CLOSCA 10 ALBA IULIA	9	2	Moderate difference	↘
TRANSILVANIEI 8-10 ALBA IULIA	11	3	Moderate difference	↘
GH. ŞINCAI 27-31 ALBA IULIA	11	4	Moderate difference	↘
TARGULUI 1	11	17	Moderate difference	↖
V.GOLDIŞ 14-14A	13	25	Large difference	↖
ORIZONTULUI 8-16 ALBA IULIA	14	5	Moderate difference	↘
CLOŞCA 5-7	15,5	20	High agreement / small difference	↖
ORIZONTULUI 1-13	15,5	27	Large difference	↖
V.GOLDIŞ 26	17	12	High agreement / small difference	↘
TRANSILVANIEI 14	18,5	10	Moderate difference	↘
T.VLADIMIRESCU 34	18,5	22	High agreement / small difference	↖
TRANSILVANIEI 23B	20,5	11	Moderate difference	↘
GH. ŞINCAI 17-21	20,5	14	Moderate difference	↘
CRAVIEI 2-4	22,5	29	Moderate difference	↖
V.ALECSANDRI 76	22,5	28	Moderate difference	↖
GH. ŞINCAI 23-25	24	15	Moderate difference	↘
LIVEZII 49 ALBA IULIA	25	30	High agreement / small difference	↖
ORIZONTULUI 4-6	26	16	Moderate difference	↘
AMPOIULUI 9-17	27	13	Large difference	↘
TRANSILVANIEI 6	29	9	Very large difference	↘
TRANSILVANIEI 4	29	8	Very large difference	↘
TRANSILVANIEI 2A	29	7	Very large difference	↘

Source: compiled by the authors

The categorised comparison confirms that only one building remained within a two-position difference between the two rankings, while five buildings showed a small difference of more than two and up to five positions. Moderate differences of more than five and up to ten positions were observed for twelve buildings. Another twelve buildings changed by more than ten positions, including six buildings with very large differences of more than fifteen positions. The direction of change was evenly distributed. Fifteen buildings were ranked as higher priorities in the validation exercise than in the digital tool ranking, while fifteen were ranked lower. This suggests that the validation exercise did not systematically shift buildings in only one direction, but substantially reshuffled their relative priority positions.

Ptuj and Kidričevo (SI)

For the Slovenian pilot site, covering the Municipality of Kidričevo and the Municipality of Ptuj, the validation dataset only partly overlaps with the original digital tool ranking. The digital tool included 30 ranked address/building units, while 16 corresponding units remained available for comparison after data cleaning. Database adjustment was necessary because some buildings were ranked in the digital tool as grouped address units, while the validation data also included individual entrances or repeated entries. In these cases, grouped address units were retained where they matched the original digital tool unit, and duplicate individual entries were removed. The comparison therefore refers only to the validated subset of the original digital tool ranking. Fourteen original digital tool units[6] could not be included in the validation comparison.

The analysed Slovenian subset shows moderate variation across the indicators used for the validation score. Most buildings were assessed as being in poor or fair overall energy condition, with one building assessed as very poor. Social vulnerability was assessed as either medium or high across the sample, while complaints related to energy hardship ranged from none to frequent or recurring. All buildings were reported as having no valid EPC. Heating systems are relatively uniform, with most buildings using district heating and two using biomass. The buildings are also relatively similar in age, mostly constructed between 1946 and 1970, while more variation is visible in residents' renovation readiness, building size and some condition-related assessments. The outcome of alternative ranking in comparison with the original tool is shown in Table 13.

Table 13. Comparison of validation („valid.") ranking and digital tool ranking for the Slovenian pilot sites Ptuj and Kidričevo

PTUJ / KIDRIČEVO	Score (valid.) (0-1)	Rank (valid.)	Rank (tool)	Rank shift	Absolute diff.
Ul. Borisa Kraigherja 19, 21, 23	0,850	1	13	12	12
Ul. Borisa Kraigherja 8,10,12	0,766	2	9	7	7
Čučkova 13	0,749	3	23	20	20
Ul. Borisa Kraigherja 20,22,24,26	0,711	4	4	0	0
Kajuhova ul.11	0,702	5	2	-3	3
Ul. Borisa Kraigherja 1, 3,5	0,660	6	11	5	5
Ul. Borisa Kraigherja 2,4,6	0,627	7	10	3	3
Vlahovičeva ulica 5	0,624	8	27	19	19
Mladinska ulica 3,4,5	0,611	9	30	21	21
Ul. Borisa Kraigherja 14 16,18	0,605	10	8	-2	2
Vlahovičeva ulica 7	0,591	11	26	15	15
Čučkova 11	0,578	12	22	10	10
Kajuhova ulica 13	0,570	13	17	4	4

[6] The original digital tool units not included in the validation comparison were: Kajuhova ulica 7; Čučkova 7; Lackova 7; Čučkova 9; Vlahovičeva ulica 9; Kajuhova ulica 12; Lackova 5; Čučkova 5; Ul. Borisa Kraigherja 13, 15, 17; Kajuhova ulica 14; Vlahovičeva ulica 3; Vlahovičeva ulica 1; Kajuhova ulica 15; Ul. Borisa Kraigherja 25, Kajuhova ulica 4.

Ul 25 Maja 9	0,549	14	6	-8	8
Ul. Borisa Kraigherja 7,9,11	0,498	15	3	-12	12
Arbajterjeva 3	0,398	16	5	-11	11

Source: compiled by the authors

For the Slovenian pilot site, covering the Municipality of Kidričevo and the Municipality of Ptuj, the comparison is based on the cleaned subset of 16 validated building units that could be matched with the original digital tool ranking. Within this subset, the validation ranking produced several substantial changes in building prioritisation. The mean absolute rank difference was 9.5 positions, while the median difference was 9 positions. Individual absolute differences ranged from 0 to 21 positions. The overall rank-order association between the two rankings was weak, negative and not statistically significant (Spearman's $\rho = -0.109$, $p = 0.688$), indicating that the validation ranking did not closely follow the original digital tool prioritisation. The differences in outcomes of the digital tool and alternative ranking are organised in Table 14.

Table 14. Categorised agreement and direction of priority shifts for the Slovenian pilot sites of Ptuj and Kidričevo

PTUJ / KIDRIČEVO	**Rank (valid.)**	**Rank (tool)**	**Agreement level**	**Valid. priority shift**
Ul. Borisa Kraigherja 19, 21, 23	1	13	Large difference	↖
Ul. Borisa Kraigherja 8,10,12	2	9	Moderate difference	↖
Čučkova 13	3	23	Very large difference	↖
Ul. Borisa Kraigherja 20,22,24,26	4	4	Very high agreement	=
Kajuhova ul.11	5	2	High agreement / small difference	↘
Ul. Borisa Kraigherja 1, 3,5	6	11	High agreement / small difference	↖
Ul. Borisa Kraigherja 2,4,6	7	10	High agreement / small difference	↖
Vlahovičeva ulica 5	8	27	Very large difference	↖
Mladinska ulica 3,4,5	9	30	Very large difference	↖
Ul. Borisa Kraigherja 14 16,18	10	8	Very high agreement	↘
Vlahovičeva ulica 7	11	26	Large difference	↖
Čučkova 11	12	22	Moderate difference	↖
Kajuhova ulica 13	13	17	High agreement / small difference	↖
Ul 25 Maja 9	14	6	Moderate difference	↘
Ul. Borisa Kraigherja 7,9,11	15	3	Large difference	↘
Arbajterjeva 3	16	5	Large difference	↘

Source: compiled by the authors

The categorised comparison shows that two buildings remained within a two-position difference between the two rankings, while four buildings showed a small difference of more than two and up to five positions. Moderate differences of more than five and up to ten positions were observed for three buildings. Seven buildings changed by more than ten positions, including three buildings with very large differences of more than fifteen positions. The direction of change was not evenly distributed: ten buildings were ranked as higher priorities in the validation exercise than in the digital tool ranking, five were ranked lower, and one retained the same rank. This suggests that, within the validated subset, the field-based validation exercise placed stronger priority on several buildings that were not among the highest-ranked units in the original digital tool output.

It is important to acknowledge that the results derived from two distinct methodologies—the initial tool and the alternative ranking—can be compared with caution. The observed discrepancies in outcomes may primarily stem from the differing methodologies employed and the varying nature of the data collected. In the case of the initial tool, the data utilized is objective, sourced from statistical offices or manually gathered by partners from their respective data records. Conversely, the alternative ranking relies on data obtained through surveys administered to building managers and tenants. Surveys often present inherent limitations, such as reliance on self-reported data and the anonymity of respondents, which precludes the ability to identify the individuals who provided the responses.

While it is generally advisable to base decisions on objective data, such data is not always readily available or easily accessible, as evidenced by the experiences of the CEESEN-BENDER project. In instances where timely decision-making is essential and field-based knowledge is required, the survey approach serves as a flexible and viable alternative.

Another factor contributing to the disparities between the two rankings is the differing weights assigned to the various factors that constitute the final score. In the regression analysis, only those factors that are statistically significant influence the final score. In contrast, the project team members involved in the alternative ranking determined which factors would contribute to the final score. When methodologies are replicated, different weights may be assigned to the factors, thereby reflecting the perceived importance and contribution of each factor to the overall score. These considerations should be taken into account for future implementations by diverse user groups.

List of references

Aksoezen, M., Daniel, M., Hassler, U., Kohler N. (2015). Building age as an indicator for energy consumption. *Energy and Buildings*, 87, 74-86. 10.1016/j.enbuild.2014.10.074

Andersen, R., Jensen, L.B., & Ryberg, M. (2021). Using digitized public accessible building data to assess the renovation potential of existing building stock in a sustainable urban perspective. *Sustainable Cities and Society*, 75, 103303. https://doi.org/10.1016/j.scs.2021.103303

Baker, K., & Kaul, B. (2000). Impact of Changes in Household Composition on Home Improvement Decisions. *Joint Center for Housing Studies Harvard University*. Retrieved from https://www.jchs.harvard.edu/research-areas/working-papers/impact-changes-household-composition-home-improvement-decisions

Balaras, C. A., Droutsa, K., Dascalaki, E., Kontoyiannidis, S. (2005). Heating energy consumption and resulting environmental impact of European apartment buildings. Energy and Buildings, 37(5), 429-442. 10.1016/j.enbuild.2004.08.003

Curtis, J., Grilli, G., & Lynch, M. (2024). Residential renovations: Understanding cost-disruption trade-offs. *Energy Policy*, 192, 114207. https://doi.org/10.1016/j.enpol.2024.114207

Dell'Anna, F. (2023). An ELECTRE TRI B-Based Decision Framework to Support the Energy Project Manager in Dealing with Retrofit Processes at District Scale. *Sustainability*, *15*(2), 1250. https://doi.org/10.3390/su15021250

Drescher, K., & Janzen, B. (2021). Determinants, persistence, and dynamics of energy poverty: An empirical assessment using German household survey data. *Energy Economics*, 102, 105433. 10.1016/j.eneco.2021.105433

European Commission (n.d.). *Energy poverty*. Retrieved from https://energy.ec.europa.eu/topics/markets-and-consumers/energy-consumer-rights/energy-poverty_en

Gade, A. N., Larsen, T., Nissen, S., Jensen, R. L. (2018). REDIS: A value-based decision support tool for renovation of building portfolios. *Building and Environment*, 142, 107-118. https://doi.org/10.1016/J.BUILDENV.2018.06.016

Galimshina, A., Moustapha, M., Hollberg, A., Padey, P., Lasvaux, S., Sudret, B., & Habert, G. (2020). Statistical method to identify robust building renovation choices for environmental and economic performance. *Building and Environment*, 183, 107143. https://doi.org/10.1016/j.buildenv.2020.107143

García-Fuentes, M. Á., Álvarez, S., Serna, V., Pousse, M., & Meiss, A. (2019). Integration of Prioritisation Criteria in the Design of Energy Efficient Retrofitting Projects at District Scale: A Case Study. *Sustainability*, *11*(14), 3861. https://doi.org/10.3390/su11143861

Giretti, A., Carbonari, A., Corneli, A., Villa, V., & Di Giuda, G. (2017). Decision Support Tool for Multi-Criteria Analyses of the Quality of Large Building Stock. *Proceedings of the 34rd ISARC*, 22-29. https://doi.org/10.22260/ISARC2017/0003

Harvard Joint Center for Housing Studies (2022). Home Repairs and Updates Pose Considerable Burdens for Lower-Income Homeowners. Retrieved from https://www.jchs.harvard.edu/blog/home-repairs-and-updates-pose-considerable-burdens-lower-income-homeowners

Jensen, P.A., Maslesa, E., Berg, J.B., & Thuesen, Ch. (2018). 10 questions concerning sustainable building renovation. *Building and Environment*, 143, 130-137. https://doi.org/10.1016/j.buildenv.2018.06.051

Jiménez-Pulido, C., Jiménez-Rivero, A., & García-Navarro, J. (2021). Strategies to Promote Deep Renovation in Existing Buildings. In: Andreucci, M.B., Marvuglia, A., Baltov, M., Hansen, P. (eds) *Rethinking Sustainability Towards a Regenerative Economy*. Future City, vol 15. Springer, Cham. https://doi.org/10.1007/978-3-030-71819-0_2

Kearns, A., Whitley, E., Curl, A. (2019). Occupant behaviour as a fourth driver of fuel poverty (aka warmth & energy deprivation). *Energy Policy*, 129, 1143-1155. https://doi.org/10.1016/j.enpol.2019.03.023

Nielsen, A. N., Jensen, R. L., Larsen, T.S., & Nissen, S.B. (2016). Early stage decision support for sustainable building renovation – A review. *Building and Environment*, 103, 165-181. https://doi.org/10.1016/j.buildenv.2016.04.009

Mahmoud, S., Zayed, T., & Fahmy, M. (2019). Development of sustainability assessment tool for existing buildings. *Sustainable Cities and Society*, 44, 99-119. https://doi.org/10.1016/j.scs.2018.09.024

Meier, H., Rehdanz, K. (2010). Determinants of residential space heating expenditures in Great Britain. *Energy Economics*, 32(5), 949–959. 10.1016/j.eneco.2009.11.008

Oliveira, R., Figueiredo, A., Vicente, R., & Almeida, R. M.S.F. (2020). Impact of unoccupied flats on the thermal discomfort and energy demand: Case of a multi-residential building. *Energy and Buildings*, 209, 109704. 10.1016/j.enbuild.2019.109704

Piredda, G., Prataviera, E., Carnieletto, L., De Carli, M., & da Porto, F. (2025) Strategic framework for sustainable energy renovation policies in educational buildings: a priority-based approach. *Energy and Buildings*, 338, 115702. https://doi.org/10.1016/j.enbuild.2025.115702

Plaut, P., & Plaut, S. (2010). Decisions to Renovate and to Move. *Journal of Real Estate Research*, *32*(4), 461–484. https://doi.org/10.1080/10835547.2010.12091286

Ranstam, J., & Cook, J. A. (2018). LASSO regression. *BJS*, 105(10), 1348. https://doi.org/10.1002/bjs.10895

Salvia, M., Simoes, S. G., Herrando, M., Čavar, M., Cosmi, C., Pietrapertosa, F., Pedro Gouveia, J., Fueyo, N., Gómez, A., Papadopoulou, K., Taxeri, E., Rajić, K., Di Leo, S. (2021). Improving policy making and strategic planning competencies of public authorities in the energy management of municipal public buildings: The PrioritEE toolbox and its application in five mediterranean areas. *Renewable and Sustainable Energy Reviews*, 135,110106. https://doi.org/10.1016/j.rser.2020.110106

Stegnar, G. (2025). Strategic Prioritization of Residential Buildings for Equitable and Sustainable Renovation. *Sustainability*, *17*(5), 2203. https://doi.org/10.3390/su17052203

Streimikiene, D., & Balezentis, T. (2020). Willingness to Pay for Renovation of Multi-Flat Buildings and to Share the Costs of Renovation. *Energies*, *13*(11), 2721. https://doi.org/10.3390/en13112721

Wang Y., Shi, Z., & Niffeler, M. (2025). Prioritization of Urban Energy System Renovation through Bi-level Energy Hub Optimization. *IOP Conference Series: Earth and Environmental Science*. https://doi.org/10.1088/1755-1315/1554/1/012054

Zarco-Periñán, P. J., Zarco-Soto, I. M., Zarco-Soto, F. J., & Sánchez-Durán, R. (2021). Influence of Population Income on Energy Consumption for Heating and Its CO_2 Emissions in Cities. *Energies*, *14*(15), 4531. https://doi.org/10.3390/en14154531

Zhang, Y., Wang, H., Gao, W., Wang, F., Zhou, N., Kammen, D. M., & Ying, X. (2019). A Survey of the Status and Challenges of Green Building Development in Various Countries. *Sustainability*, 11(19), 5385. https://doi.org/10.3390/su11195385

Annexes

Annex 1. Validation Survey for Prioritizing Building Renovation

Introduction

This short survey supports the validation of a digital tool used to prioritise buildings for energy renovation. It is designed to complement the existing tool with building managers' knowledge and perceptions of the ranked buildings, with stronger emphasis on social vulnerability, building condition, and renovation need.

This survey is part of the CEESEN-BENDER project, which supports vulnerable households living in multi-apartment buildings built after World War II and before the 1990s in Croatia, Slovenia, Estonia, Poland, and Romania. The project aims to improve support for renovation by identifying key barriers and strengthening trusted support services.

Participation is completely voluntary and anonymous. No personal data will be collected, and the results will be used only in aggregated form to improve and interpret the existing digital tool.

We are asking only for your experience and your best-informed assessment of the buildings you have been asked to evaluate.

We greatly appreciate your time and effort.

Estimated time: approximately 5 minutes per building.

The CEESEN-BENDER project has received funding from the European Union's Programme for the Environment and Climate Action (LIFE 2021–2027) under grant agreement no. LIFE 101120994. The information and views set out in this material are those of the author(s) and do not necessarily reflect the official opinion of the European Union or CINEA; neither the European Union nor the granting authority can be held responsible for them.

Building address (e.g. Street name 42):

____

Vulnerability of Households and Energy Efficiency

How would you rate the building's overall energy condition?

(Please consider the main features affecting energy use, such as insulation, windows, roof, and heating system)

a) Excellent (Like a new building, no issues)

b) Good

c) Fair

d) Poor

e) Very poor (Extremely inefficient, constant issues)

What share of households in this building would you consider socially or financially vulnerable?

(For example, households struggling with low income, debt, poverty, social exclusion, or broader financial hardship)

a) Low (rouglhy less than 10%)

b) Medium (10-20%)

c) High (roughly more than 20%)

In the past 12 months, how common were complaints from residents or building representatives about cold homes in winter, overheating in summer, damp or mould, or energy bills being too high?

a) None

b) Isolated cases

c) Occasional complaints

d) Frequent or recurring

Does this building currently have a valid Energy Performance Certificate (EPC)?

a) Yes

b) No

c) I do not know

What is the energy class of the building according to the EPC?

a) A

b) B

c) C

d) D

e) E

f) F

g) G

On the scale of 1 to 10, 1 being the LOWEST PRIORITY, and 10 the HIGHEST PRIORITY, how would you rate this specific building need for renovation?

Lowest prioritiy (1) (2) (3) (4) (5) (6) (7) (8) (9) (10) Highest priority

Heating System and Residents' Renovation Readiness

What is the building's main space-heating system?

(If more than one system is used, please select the one used by most dwellings or providing most of the heating.)

a) Oil boilers

b) Coal or other solid fossil fuel boiler

c) Individual gas boilers in flats

d) Central gas boiler for the whole building

e) Distrit heating (mainly fossil-fuel based, if known)

f) Electric heating (e.g. electric radiators or electric boilers)

g) Biomass

h) Heat pump

i) Solar thermal

j) I don't know

How would you rate the residents' overall interest in a major energy renovation of this building?

a) Low

b) Medium

c) High

Structural Characteristics of the Building

Approximately, when was the building originally constructed?

a) Before 1945
b) 1946-1970
c) 1971-1990
d) 1991-2000
e) 2001 or later
f) I don't know

Approximately, how many residential units (flats) are in this building?

a) 10 or less
b) 11-20
c) 21-40
d) 41-100
e) More than 100
f) I don't know

Which of the following best describes how this building is attached to neighbouring buildings?

a) Terraced (mid-row building)
b) Semi-detached (end-of-row building)
c) Detached building
d) I don't know

Annex 2. Data Collection Process Documentation Form – Validation Survey for the Digital Tool

*Fields marked with * are mandatory.*

About this survey

This survey is used to document how your organisation carried out stakeholder outreach and data collection for the T4.1 digital tool validation survey. The purpose is to collect comparable information across partners on the outreach process, follow-up actions, major challenges, and lessons learned.

This survey should be completed by one person per organisation, ideally the person who participated in the training "WP4 – Tool validation workshop with Q&A" held in Ptuj on 17 March 2026 and who was most involved in contacting stakeholders. Other colleagues may support the completion of the survey, but only one final response per organisation should be submitted.

Section 1. Respondent Information

*** Name and surname:**

___

*** Organisation:**

___

*** Pilot site (where the ranking of the buildings through task T4.1 took place):**

___

Section 2. Scope of Outreach & Stakeholder Support

*** Did your organisation attempt to collect survey data for the same ranked buildings / addresses identified through the digital tool?**

☐ Yes, for all targeted buildings

☐ Yes, for most targeted buildings

☐ Only partly

☐ No

Please explain why the initial outreach did not cover all the buildings / addresses originally ranked by the digital tool.

Complete this field if applicable.

*** When was the first outreach attempt sent or made?**

Date: ____ / ____ / ______

*** How was the first outreach made? Please select all that apply.**

☐ Email

☐ Phone call

☐ In-person meeting

☐ Workshop / group meeting

☐ Other

Please specify:

____

*** Who was contacted in the first outreach? Please select all that apply.**

☐ Building managers

☐ Residents' representatives / building representatives

☐ Homeowners' associations

☐ Other

Please specify:

*** Approximately how many unique contacts were approached in the first outreach?**

____

*** Were any follow-up, reminder, or additional outreach actions used after the first contact?**

☐ Yes

☐ No

*** Please briefly describe any follow-up, reminder, or additional outreach actions used after the first contact. Please indicate when these actions took place, how they were carried out, and whether they targeted the same or different stakeholder types.**

*** Please describe any additional methods used to support stakeholder engagement or survey completion, if any.**

Section 3. Challenges, Lessons Learned and Recommendations

*** What were the main challenges in collecting the survey data in your pilot site, including any factors outside your organisation's control?**

*** Based on your experience, what is your overall assessment of the outreach and data collection process, and what would you recommend for future rounds of similar work?**

Section 4. Upload of Outreach / Supporting Evidence

*** Please upload any supporting evidence related to the first outreach and any follow-up, reminder, or additional outreach actions.**

In EUSurvey, this field is completed using the file upload option. Supporting evidence may include emails, screenshots, call notes, meeting notes, invitations, supporting materials, or other relevant documents.

Uploaded file(s): ____

Co-funded by the European Union

The CEESEN-BENDER project has received funding from the European Union's Programme for the Environment and Climate Action (LIFE 2021-2027) under grant agreement no LIFE 101120994. The information and views set out in this material are those of the author(s) and do not necessarily reflect the official opinion of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them.