Supplementary Material

BERTUG OZARISOY

School of Architecture, Computing and Engineering

University of East London

FEBRUARY 2022

Table of Contents

List of Figures

List of Tables

Supplementary Material 1: Guidance for the Development Stage of a Statistically Representative Sample Size

Table S.1.1: Step-by-Step Development of the Study Sample.

Introduction

This appendix details the problems with invalid data and how invalid data can be assessed based upon research. This appendix provides a summary of the data preparation stages, as well as explaining the concept of developing a statistically representative sample size.

Stage 1: Checking for Invalid Data

Checking a dataset for invalid cases or values is a critical part of data preparation. Invalid datapoints are observations that reflect inaccurate, inattentive or careless response values. Cases that exhibit these types of responses can seriously bias a study. The appropriate procedure for handling invalid data is typically removal of the data in the invalid case, which often results in a reduction of sample size. This appendix is a tool to help future scholars write an appropriate discussion about case removal in their own methodological framework.

Reasons for Removal of Data

Cases and/or observations are flagged as invalid if they meet any of several criteria:

Duplicate cases: Duplicate cases can occur when respondents take a survey more than once. Avoiding this requires the use of unique identifiers (e.g., participant ID, IP address, or email address) in order to isolate whether respondents may have been duplicated in the data. Duplicated cases bias results and should be removed (Johnson, 2005). In the present study, there was no unique identifier was used.

Participant did not consent: Respondents who did not consent are excluded from the study for ethical reasons. In the present study, no data was excluded on this basis.

Met exclusion criteria: Cases that do not meet the inclusion criteria, outlined in the sampling strategy, need to be excluded in order to meet the objectives of the study. There were no data was removed based on this criterion.

Dropped out without responding: Cases that may have begun but dropped off with no response on some/all of the items of interest. These cases should not be counted as part of the final sample and should be excluded from the rest of the analysis to avoid any research bias in the interpretation of the results. In the present study, cases with only two responses entered were removed during the dataset preparation stage.

Impossible values: Impossible values refer to values in a variable that lie outside the theoretical range for that particular variable, for example: a case with a negative value for body mass index (BMI) would be impossible. Unless the convention recommends that the researcher go back to the original source (e.g., the questionnaire survey) and confirm the correct value, it is recommended that these values be set as missing values. In the present study, an impossible *In_WET* temperature of 48.50 was corrected to 25.60.

Summary of the Identification of Invalid Data

Table 1 below shows each of the reasons for removal due to duplicates, non-consent, cases that did not meet the inclusion criteria, and cases that dropped off at onset of the questionnaire survey. These categories are considered critical removal reasons. The column titled 'Met Removal Criteria' presents the frequencies and percentages of each individual case removal reason found in the sample population. The frequencies do not reflect the total number of invalid cases that were removed because some cases had more than one removal reason (e.g., an observation/case may not have met the inclusion criteria and was a duplicate). In the present study, there were zero duplicate cases and there were zero impossible values.

Table 1: Invalid Data Summary for Critical Removal and Removal with Two or More Removal Reasons

Valid categorisation	***n***	***%***
Critical invalid cases	0	0.0
Potential invalid cases with two or more removal reasons	0	0.0
Valid cases	100	100.0
Total	100	100.0

Note: This Table presents the data included into the dataset after the completion of necessary data-mining process.

Supplementary Material 1: Guidance for the Development Stages of the Concept of Statistically Representative Sample Size

Table S.1.2: Step-by-Step Development of the Study Sample. (Continued)

Introduction

This section details the problem of missing data and how it is assessed. In this study, the researcher also prepared a summary of the missing data to avoid any research bias while interpretating the statistical analysis, and here presents a list of options for this process to guide future readers of the study.

Stage 2: Missing Data Report

Missing data can have a profound effect on a study's validity depending on how severe the missing data is within a dataset (Little & Rubin, 2002). Therefore, missing data should not be ignored and best practices should be followed to handle it appropriately. Fortunately, within the last few decades, there have been advances in statistical methodology to properly identify and correct missing data issues. The impact of missing data on an analysis depends on three factors: sample size, the proportion of missing values in the data and the pattern of missing values in the data.

Sample size

Sample size matters in missing data for two reasons: 1) small to moderate samples with missing data are prone to a higher likelihood of biased estimates and larger standard errors, and 2) as sample size increases the standard errors for the estimates decrease, leading to more efficient estimates regardless of the presence of missing data (Cheema, 2014; Dong & Peng, 2013; Little & Rubin, 2002). In other words, the larger the sample size, the less of an effect missing data will have on a study.

Proportion of Missing Data

The percentage of missing data present in the data matrix has implications for researchers in terms of how the missing data must be treated to resolve potential biased estimates in the analysis. Complete-case analysis is only reliable when the percentage of incomplete values in the data matrix is less than 5% (Little & Rubin, 2002; Schafer, 1999). Beyond the 5% threshold, the data should be treated in some manner (e.g., multiple imputation or maximum-likelihood replacement).

Pattern of Missing Data

Little and Rubin (2002) suggest that the pattern of missing data is a more important factor than the actual amount of missing data. Missing values that are randomly scattered throughout the data matrix (i.e., data is missing completely at random; MCAR) is a less serious problem than missing values that are not random. If the missing values on a variable are not random but are associated with other variables in the data, this is referred to as missing at random (MAR) and can be alleviated with imputation methods. If the missing values on a variable are associated with the variable itself (i.e., missing not at random; MNAR), this has serious implications for inferences, regardless of imputation procedures (Dong & Peng, 2013). Recently, Schouten *et al.* (2009) proposed two definitions of representativeness with respect to survey response: strong (given in Definition 1.1) and weak (given in Definition 1.2).

Definition 1.1 (strong). A response subset is representative with respect to the sample if response propensities are the same for all units in the population:

$$\forall_i \; E(R_i) = \rho_i = P(R_i = 1 \mid I_i = 1) = \rho \tag{1}$$

The response of a unit is independent of the response of all other units, where denotes the response of unit *i* and is an indicator showing whether a unit took part in the survey. Schouten *et al.* (2009) notes that strong representativeness corresponds to the Missing Completely at Random (MCAR) pattern for every target variable *y*. This means that non-response does not cause estimators to be biased. Although this definition is appealing, its validity can never be tested in practice. To solve this problem, a weaker definition of representativeness was introduced by Schouten *et al.* (2009).

$$\bar{\rho}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} \rho_{ih} = \rho, for\ h = 1, 2, \ldots, H, \tag{2}$$

Definition 1.2 (weak). A response subset is representative of categorical variable **x** with H categories if the average response propensity over the categories is constant, where is the population size of category *h,* is the response propensity of unit i in category *h*, and summation is over all units in this category.

Supplementary Material 1: Guidance for the Development Stages of the Concept of Statistically Representative Sample Size

Table S.1.3: Step-by-Step Development of the Study Sample. (Continued)

Summary of the Proportion of Missing Data

An extensive analysis of missing data was conducted for the present study. The results and recommendations are presented below.

The three pie charts below summarise the frequency and percentage of missing data in the dataset by variable, case/observation and individual values. The third pie chart represents the full data matrix and was used to evaluate the 5% threshold of the proportion of missing values in the data matrix that was discussed above.

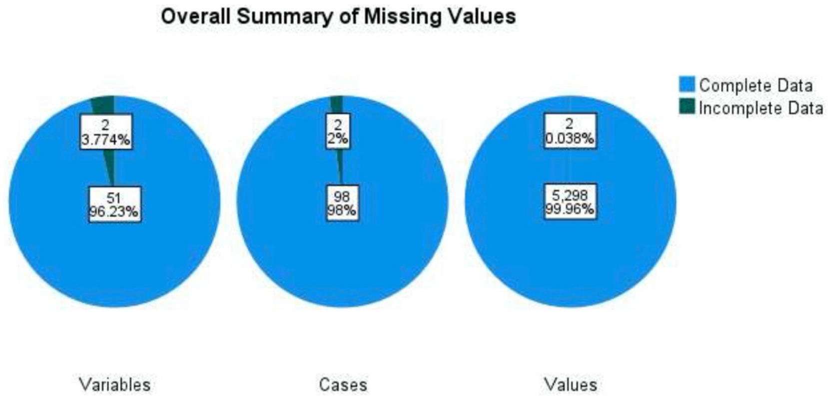

Figure 1: Overall summary of missing data values.

Note: *There is a distinction being drawn between missing data removed for critical reasons and data was not removed. This graph presents the detection of missing data prior to the data mining process of the dataset before conducting the statistical analysis.*

Summary of the Pattern of Missing Data

It was found that the data was missing completely at random (MCAR). After preparing the data for analysis, it was observed that out of 100 recorded cases, 98 cases contained missing data (98.0%) and out of 53 variables, 2 variables contained missing data (2.8%), which amounted to a total of 0.04% missing information in the dataset. To assess whether the pattern of missing values was MCAR, Little's MCAR test (Little, 1988) was conducted. The null hypothesis of Little's MCAR test is that the pattern of the data is MCAR and follows a chi-square distribution. Using an expectation-maximisation algorithm, the MCAR test estimates the univariate means and correlations for each of the variables. The results revealed that the pattern of missing values in the data was MCAR: χ^2 (104) = 121,645, p = 0,114. Even though the proportion of the total missing data is less than 5% and the data is MCAR, the final sample size may still be affected by listwise or pairwise deletion when the analysis is run. Listwise deletion removes a case if a case has any missing value for any of the variables used in an analysis. This is also known as complete-case analysis. Pairwise deletion better maximises all the data available in the data analysis and is preferred over listwise deletion for increasing the statistical power of the study (Newman, 2014).

Table 2: Comparison of Minimum Pairwise and Listwise N for the Final 53 Variables of Interest

Deletion Method	Minimum N
Listwise	98
Pairwise	98

Note: Pairwise and listwise *Ns* are the same because the same two participants had the same missing values. The final *N* may be higher upon analysis.

Note: This Table presents the data at the time of detecting the missingness data in the data-mining process.

Recommendations

The missing values of data allow flexibility when addressing missing data because the proportion of missing data in the sample is less than 5% and the pattern of missing data is MCAR. Based upon these two findings, the data should be fine using either pairwise or listwise deletion methods. Listwise and pairwise deletion are unbiased techniques when data is MCAR; however, pairwise deletion increases power (Newman, 2014). Given the sufficient sample size for the present study, the proportion of missing data and the pattern of missing data, listwise or pairwise deletion is recommended.

Supplementary Material 1: Guidance for the Development Stages of the Concept of Statistically Representative Sample Size

Table S.1.4: Step-by-Step Development of Study Sample. (Continued)

Stage 3: Basic Assumption Testing Report

Introduction

This section details the importance of basic assumption testing in inferential analyses, as well as how basic assumptions were assessed, the summary of the data preparation process, and options for coding data prior to conducting analysis.

Basic Assumptions and Inferential Analyses

Before any inferential analyses are conducted, basic assumptions must be met to avoid bias in a study's findings. The validity of conclusions drawn from a statistical analysis depend on the validity of any assumptions made. Where data is lacking, assumption testing may have to be restricted to simply making a judgment about whether an assumption is reasonable. In addition, scholars may have to judge what effect the violation of an assumption might have on the findings. The effect of violating any of the assumptions is a change in the probability of making a Type I or a Type II error, and the researchers won't usually know whether the change has made it more or less likely to commit an inferential error. Basic assumptions are also accompanied by analysis-specific assumptions. Analysis-specific assumptions are tested during the analysis phase of the project.

How are Basic Assumptions Tested?

In the present study, the researcher used available descriptive statistics in order to fully understand the data. When it comes to making inferences, both parametric and nonparametric analysis can be applied. Nonparametric analyses are considered when there is reason to worry about parametric assumptions, or when the measurements being considered demand them.

Below are the assumptions tested and the recommended guidelines for how to handle assumption violations. Keep in mind that for some assumptions there are no hard and fast rules on cut-offs. It is up to the researcher to account for them and report any biases that may result from not addressing assumption violations.

Sample Sizes of Comparison Groups

A general rule for sample sizes is that group sizes are approximately equal if *n* of the largest group is no more than about twice *n* of the smallest group. Another general rule for sample sizes is at least 10% of the sample should be in each group. Categorical variables with very uneven splits between categories present problems for several multivariate analyses (Tabachnick & Fidell, 2007). The following variables were recoded to reflect the conceptualisation of statistically representative findings in accordance with the research hypotheses.

Table 3: Coded Variables During the Data Preparation Stage.

Age band	Cooling energy consumption in August 2015	Floor level
Cooling consumption on weekdays	Cooling energy consumption in summer 2015	Health condition
Clothing insulation levels of participants	Heating energy consumption in winter 2015	Occupation
Type of cooling control in home	Cooling energy consumption in August 2016	Heating consumption on the weekend
Ethnicity	Cooling energy consumption in summer 2016	Household density
Thermal preference	Heating energy consumption in winter 2016	Income
Interviewed room condition	Metabolic rates of participants	Length of residency
Orientation	Reasons for thermal discomfort	Space conditioning
Overall thermal satisfaction in summer	Thermal sensation in bedrooms 1, 2, 3 and living room	Type of cooling system
Window closing reasons	Window opening patterns in winter	Type of heating system

Note: Additionally, all categorical variables were recoded in ordinal sequence where possible (e.g., metabolic rate)
Variables related to occupants' thermal preferences were recoded from very cold to very hot
All variables were recoded from smallest value to largest value
All dichotomous variables were recoded to 1 = yes, 0 = no

Supplementary Material 1: Guidance for the Development Stages of the Concept of Statistically Representative Sample Size

Table S.1.5: Step-by-Step Development of Study Sample. (Continued)

Stage 3: Basic Assumption Testing Report (Continued)
Normality
Virtually all parametric statistics have an assumption that the data comes from a population that follows a known distribution. This assumption of normality is often erroneous applied, however, because many populations are not normally distributed. Therefore, researchers need to understand what their samples consist of. It is standard practice to assume that the sample mean from a random sample is normal because of the central-limit theorem. However, almost all variables have a slight departure from normality. If researchers have a large enough sample, then any statistical test will reject the null hypothesis. In other words, the data will never be normally distributed if the sample size is large enough. To assess normality, skewness and kurtosis statistics are assessed. Skewness refers to the symmetry of the distribution and kurtosis refers to the peakedness. Variables that have distributions that are very asymmetrical, flat, or peaked could bias any test that assumes a normal (i.e., bell-shaped) distribution. Generally, skewness and kurtosis values (converted as z-scores) that fall outside ±4 should be further inspected for potential outlier removal, nonparametric testing, or transformation. However, researchers may have flexibility in larger samples (Field, 2013). Some normality tests are done for sample sizes smaller than 100 (i.e., Shapiro–Wilks and Kolmogrov–Smirnov tests). If these tests are significant beyond $p < 0.001$, these variables should be further inspected (Gamst, Meyers & Guarino, 2008). Graphing methods are also employed for assessing normality. These graphs include histograms, normal quantile–quantile (Q-Q) plots and box plots. Histograms should look fairly bell-shaped. Q-Q plots should follow a straight line when plotting the expected values against the observed values. Box plots show the overall interquartile range and whether extreme values exist in the variable (see section on outliers below). If the data contains outliers, graphic displays both with and without the outliers should be examined to see how the graphs changed. If a continuous variable has serious deviations from normality, it must be addressed through transformation (log, inverse, Box-Cox, etc.), recoding into an ordinal variable, or assessment for whether nonparametric analysis needs to be conducted. Regardless, researchers should run analysis both with and without outliers to see whether the pattern of results changes.
Outliers
Univariate and multivariate outliers are also known as extreme values and can significantly bias any parametric test. We have checked our variables for univariate outliers. SPSS identifies these values as being three times the interquartile range beyond the 25th and 75th percentile values. In the present study, multivariate outliers were tested before the primary analyses were conducted where appropriate.
Summary of Basic Assumptions - Normality
All continuous variables were constructed horizontally in the dataset with the following information: statistics summary table, histogram, normal Q-Q plot, box plot, and relevant notes about outliers (if applicable). The statistics summary, histogram, Q-Q plot and boxplots were reviewed together to determine if variables were significantly skewed, flat or peaked. A variable that violates normality will present as being non-normal across most, if not all, of the graphs and summary information. In **Appendix F,** the representative findings titled "Test of Normality" can be found: this displays the results from the Kolmogrov–Smirnov and Shapiro–Wilk tests. If the sample size is smaller than 100, we highlight the variables that violated one or both of these tests. If the sample size is larger than 100, this tab can be reviewed, but reviewing these tests is not necessary. Note that if data contains variables with outliers then the "outliers removed" variables were tested as well. Skewness, time-of-day/kurtosis issues, indoor DEW, and data mining methods were used to resolve the issues detected in these variables before undertaking the relevant statistical analyses. There are three remedies for violations of normality: mathematical transformation, categorisation and non-parametric testing. Because the eventual analysis will be a fairly simple bivariate test, in the present study we are recommending non-parametric testing as an alternative.

Supplementary Material 2: Descriptive Analysis

Information on the Convention Applied to Designing the Concept of the Thermal Comfort Assessment

In this study, a dataset related to occupants' thermal sensation votes (TSVs) was designed in accordance with the thermal comfort assessment convention recommended by Wang *et al.* (2018). As presented in Chapter 3, Subsection 3.3.4, in accordance with Wang *et al.* (2018), the thermal sensation scale was set out in two conceptual assessment criteria in order to undertake the statistical analysis accurately.

First, Wang *et al.* (2018) recommend a 7-point discrete thermal sensation scale that can be applied to assess occupants' TSVs (see Chapter 3, Subsection 3.3.4, Figure 3.20(a)). In this case, the TSV is set as an ordinal variable, thus enabling researchers to undertake Cramér's *V* test for the statistical analysis and apply the statistical findings whenever it is appropriate at the time of developing an evidence-based energy policy design. In the present study, the dataset was coded as follows: 0 = -3, 1 = -2, 2 = -1, 3 = 0, 4 = +1, 5 =+2, 6 = +3. Notably, the [-3, +3] scale band represents the outcome of occupants' thermal sensation as an ordinal measure used to accurately conduct the Cramér's *V* test. This type of coding was applied to determine households' TSVs gathered through a questionnaire survey. In the questionnaire survey, questions related to households' thermal sensation were ranked on 7-point Likert scale that could be used as an ordinal measure. In this dataset, to provide consistency of the interpretation of households' TSVs [0 to 6], a coding range representation of thermal sensation scale band at [-3, +3] was used, which was developed by Fanger in the 1970s and was commonly used by thermal comfort scholars between 1990 and 2000.

Second, Wang *et al*. (2018) recommend a 7-point continuous thermal sensation scale that can be applied to assess occupants' TSVs (see Chapter 3, Subsection 3.3.4, Figure 3.20(e)). With the TSV set as a continuous variable, researchers are able to undertake Pearson's correlation analysis. This method of design is commonly applied by thermal comfort researchers to identify "neutral" adaptive thermal comfort thresholds. Using occupants' TSVs as continuous variables is the most well-known method for reporting field survey findings concurrently with *in-situ* measurements or *on-site* environmental monitoring findings. This is an essential method of design that was developed by a team of experts at the University of California at Berkeley to contribute to the ASHRAE Global Thermal Comfort Database II. Further to this on-going method of analysis in thermal comfort studies, in the present study, the dataset was coded as follows: -3 = Cold, -2 = Cool, -1 = Slightly cool, 0 = Comfortable, +1 = Slightly warm, +2 = Warm, +3 = Hot. Notably, the [-3, +3] scale band enables thermal comfort researchers to identify "neutral" adaptive thermal comfort thresholds for benchmarking. In this dataset, to provide consistency of the interpretation of households' TSVs, the [-3, +3] coding range represents the [Cold to Hot] thermal sensation scale, which was recommended by Fanger in the 1970s and further developed by de Dear in 1998 and 2001.

Supplementary Material 2: Descriptive Analysis

Table S.2.1: Descriptive analysis of the variables related to identification of 'neutral' adaptive thermal comfort thresholds for benchmarking.

Variable Name	Mean	Median	Mode	Std. Deviation	Minimum	Maximum	Percentiles 25	Percentiles 50	Percentiles 75
Overall thermal satisfaction in summer [0 to 6] - Discrete	1.64	2.00	3	1.453	-2	3	1.00	2.00	3.00
Overall thermal satisfaction in summer [-3, +3] - Continuous	1.68	2.00	3.00	1.36241	-1.00	3.00	1.00	2.00	3.00
Thermal sensation in bedroom 1 [0 to 6] - Discrete	0.52	1.00	1	1.396	-3	3	0.00	1.00	2.00
Thermal sensation in bedroom 2 [0 to 6] – Discrete	0.46	1.00	1	1.290	-3	3	0.00	1.00	1.00
Thermal sensation in bedroom 3 [0 to 6] – Discrete	0.55	1.00	1	1.234	-2	3	0.00	1.00	1.00
Thermal sensation in kitchen [0 to 6] - Discrete	-0.35	-1.00	-2[a]	1.533	-2	3	-2.00	-1.00	1.00
Thermal sensation in living room [0 to 6] - Discrete	0.20	0.00	-1	1.595	-2	3	-1.00	0.00	1.75
Thermal sensation in bedroom 1 [-3, +3] - Continuous	0.51	1.00	1.00	1.30651	-2.00	2.00	0.00	1.00	2.00
Thermal sensation in bedroom 2 [-3, +3] - Continuous	0.45	1.00	1.00	1.23399	-2.00	2.00	0.00	1.00	1.00
Thermal sensation in bedroom 3 [-3, +3] - Continuous	0.53	1.00	1.00	1.20147	-2.00	2.00	0.00	1.00	1.00
Thermal sensation in kitchen [-3, +3] - Continuous	-0.42	-1.00	-2.00[a]	1.39393	-2.00	2.00	-2.00	-1.00	1.00
Thermal preference	1.50	1.00	1.00	1.07778	0.00	3.00	1.00	1.00	3.00

a. Multiple mode exists. The smallest value is shown.
All these variables had scales with smaller scores indicating colder whereas larger scores indicating warmer

Supplementary Material 2: Descriptive Analysis

Table S.2.2: Descriptive analysis of the variables related to households' socio-demographic characteristics.

Variable Name	Mean	Median	Mode	Std. Deviation	Minimum	Maximum	Percentiles 25th	50th	75th
Age	51.37	53.50	60.00	13.92524	23.00	84.00	42.00	53.50	60.00
Energy efficiency awareness	1.81	2.00	2	0.961	0	3	1.00	2.00	2.00
Health condition	2.68	3.00	3.00	0.93073	1.00	4.00	2.00	3.00	3.00
Metabolic rates of participants (*met*)	4.10	4.00	2.00	1.87757	1.00	7.00	2.00	4.00	6.00
Household density	1.88	2.00	3.00	0.99778	0.00	4.00	1.00	2.00	3.00

a. Multiple mode exists. The smallest value is shown.

Table S.2.3: Descriptive analysis of the variables related to households' habitual adaptive behaviour on home energy use.

Variable Name	Mean	Median	Mode	Std. Deviation	Minimum	Maximum	Percentiles 25th	50th	75th
Cooling consumption patterns on weekdays	1.20	1.00	2.00	0.76541	0.00	2.00	1.00	1.00	2.00
Cooling consumption patterns on the weekend	1.58	2.00	2.00	1.04621	0.00	3.00	1.00	2.00	2.00
Cooling energy consumption in summer of August 2015	1.25	1.00	1.00	1.02863	0.00	3.00	0.00	1.00	2.00
Cooling energy consumption in summer of 2015	2.33	2.00	3.00	1.28751	0.00	5.00	2.00	2.00	3.00
Cooling energy consumption in summer of August 2016	1.33	1.00	1.00	1.02548	0.00	3.00	1.00	1.00	2.00
Cooling energy consumption in summer of 2016	1.79	2.00	1.00	1.28153	0.00	4.00	1.00	2.00	3.00
Heating consumption patterns on weekdays	0.80	1.00	1.00	0.69631	0.00	2.00	0.00	1.00	1.00
Heating consumption patterns on the weekend	0.96	1.00	1.00	0.77746	0.00	2.00	0.00	1.00	2.00
Heating energy consumption in winter of 2015	3.00	4.00	4.00	1.47710	1.00	4.00	1.00	4.00	4.00
Heating energy consumption in winter of 2016	3.00	4.00	4.00	1.47710	1.00	4.00	1.00	4.00	4.00

a. Multiple mode exists. The smallest value is shown.

Supplementary Material 2: Descriptive Analysis

Table S.2.4: Descriptive analysis of the variables related to both *on-site* monitored and *in-situ* recorded environmental parameters.

Variable Name	Mean	Median	Mode	Std. Deviation	Minimum	Maximum	Percentiles		
							25th	50th	75th
Indoor DEW (°C)	21.48	21.90	20.20[a]	3.36021	11.40	32.40	20.20	21.90	23.40
Indoor relative humidity (%)	57.83	59.95	56.10	8.75611	31.10	75.00	52.15	59.95	63.27
Operative air temperature (°C)	30.59	31.10	31.50	1.76860	25.40	34.10	29.52	31.10	31.80
Solar radiation (°C)	33.64	32.90	32.90	2.35445	29.10	39.80	32.10	32.90	34.80
Indoor WET (°C)	24.12	24.60	18.70[a]	2.18689	18.70	31.00	23.00	24.60	25.57
Indoor wet bulb ground temperature (°C)	26.12	26.60	26.80	2.03956	21.00	30.70	25.02	26.60	27.40
Outdoor heat stress index (°C)	36.70	36.00	36.00	2.33766	33.00	43.00	35.00	36.00	38.00
Outdoor relative humidity (%)	59.16	59.00	57.00[a]	11.76264	19.60	78.00	54.00	59.00	67.00
Outdoor air temperature (°C)	32.11	32.00	34.00	2.17015	23.70	36.00	30.25	32.00	34.00
Outdoor DEW (°C)	22.82	23.00	23.00	2.21538	13.00	26.00	22.00	23.00	24.00
Indoor temperature ground (°C)	31.23	31.35	31.80	3.42791	24.70	60.20	29.80	31.35	32.20

a. Multiple mode exists. The smallest value is shown.

Supplementary Material 3: Type of Measures for each Variable

Table S.3.1: List of the type of measures for each variable.

Variable Name	Measures
Residential tower block (RTB) number	Nominal
Age	Scale
Age bands	Ordinal
Energy efficiency awareness	Ordinal
Energy conservation	Nominal
Location of subject respondent	Nominal
Thermal preference	Nominal
Doors opening patterns in summer	Nominal
Doors opening patterns in winter	Nominal
Overall thermal satisfaction in summer (Discrete thermal sensation scale)	Ordinal
Overall thermal satisfaction in summer (Continuous thermal sensation scale)	Scale
Thermal sensation in bedroom 1 (Discrete thermal sensation scale)	Ordinal
Thermal sensation in bedroom 2 (Discrete thermal sensation scale)	Ordinal
Type of heating control at home	Nominal
Thermal sensation in bedroom 3 (Discrete thermal sensation scale)	Ordinal
Thermal sensation in kitchen (Discrete thermal sensation scale)	Ordinal
Thermal sensation in living-room (Discrete thermal sensation scale)	Ordinal
Thermal sensation in living-room (Continuous thermal sensation scale)	Scale
Thermal sensation in bedroom 1 (Continuous thermal sensation scale)	Scale
Thermal sensation in bedroom 2 (Continuous thermal sensation scale)	Scale
Thermal sensation in bedroom 3 (Continuous thermal sensation scale)	Scale
Thermal sensation in kitchen (Continuous thermal sensation scale)	Scale
Type of cooling control at home	Nominal
Clothing insulation level of participants	Nominal
Reasons for thermal discomfort	Nominal
Length of residency	Nominal
Floor level	Nominal
Ethnicity	Nominal
Orientation	Nominal
Interviewed room condition	Nominal
Cooling consumption patterns on weekdays	Ordinal
Cooling consumption patterns on the weekend	Ordinal
Cooling energy consumption in summer of August 2015	Ordinal
Cooling energy consumption in summer of 2015	Ordinal
Cooling energy consumption in summer of August 2016	Ordinal
Cooling energy consumption in summer of 2016	Ordinal

Supplementary Material 3: Type of Measures for each Variable

Table S.3.2: List of the type of measures for each variable. (Continued)

Variable Name	Measures
Energy consumption in April of 2015	Scale
Energy consumption in August of 2015	Scale
Energy consumption in December of 2015	Scale
Energy consumption in February of 2015	Scale
Energy consumption in January of 2015	Scale
Energy consumption in July of 2015	Scale
Energy consumption in June of 2015	Scale
Energy consumption in March of 2015	Scale
Energy consumption in May of 2015	Scale
Energy consumption in November of 2015	Scale
Energy consumption in October of 2015	Scale
Energy consumption in September of 2015	Scale
Energy consumption in April of 2016	Scale
Energy consumption in August of 2016	Scale
Energy consumption in December of 2016	Scale
Energy consumption in February of 2016	Scale
Energy consumption in January of 2016	Scale
Energy consumption in July of 2016	Scale
Energy consumption in June of 2016	Scale
Energy consumption in March of 2016	Scale
Energy consumption in May of 2016	Scale
Energy consumption in November of 2016	Scale
Energy consumption in October of 2016	Scale
Energy consumption in September of 2016	Scale
Energy consumption in January of 2017	Scale
Indoor temperature ground (°C)	Scale
Overall cooling energy consumption in summer of 2015	Scale
Overall heating energy consumption in winter of 2015	Scale
Overall cooling energy consumption in summer of 2016	Scale
Overall heating energy consumption in winter of 2016	Scale

Supplementary Material 4: Sample of A Correlation Analysis

Information on the Applied Convention for Statistical Analysis

In this study, Fisher's exact test was applied to determine the accuracy of the Chi-squared test that were applied throughout the statistical analysis. This appendix presents the stage-by-stage development of the Fisher's exact test that was conducted for Tables 4.4(a) and (b), which are presented in Chapter 4. To conduct accurate analysis according to statistical conventions, the findings of the Chi-squared tests are presented in the appendix. As can be seen, some tables present that the numbers of cells have expected values of less than 5. In the present study, this convention was strictly used to determine whether the Fisher's exact test analysis would be required or not. This is the reason for conducting Chi-squared tests and presenting them in the appendix.

Schmill *et al.* (2014) highlight that the chi-squared test is useful for computing representativeness, which gives the degree of extent for the global representativeness of case studies included in the dataset. Schmill *et al.* also indicate that the chi-squared analysis is the standard and most applicable method of analysis in many disciplines as it allows researchers to provide a reliable result without requiring research limitations at the time of addressing research hypotheses. However, Schmill *et al.* recommend that the chi-squared test is not an applicable test for sample sizes of less than 50, or when the expected frequency for more than one category is less than 5. Following these statistical criteria provides a reasonably representative sample size that should be considered to prevent any research bias. Schmill *et al.* (2014) also indicate that Fisher's exact test can help in cases where there is an expected frequency of zero to avoid any research bias in relatively small sample sizes.

According to Schmill *et al.* (2014), Fisher's exact test should be applied before conducting Cramér's *V* test to avoid any research bias and demonstrate statistically representative research outcomes. Further to this statistical convention, Supplementary Material 6 presents the Chi-squared and Cramér's *V* tests that are presented in Tables S.6.1–4, S.6.5(a) and (b), and S.6.6(a) and (b). In these tables, Fisher's exact test was applied before undertaking Cramér's *V* test to provide reliable outcomes on households' socio-demographic characteristics and their habitual adaptive behaviours regarding energy use.

Supplementary Material 4: Sample of A Correlation Analysis

Supplementary Material 4 presents the step-by-step statistical analysis was conducted for the Tables 4.4 (a) and (b)

Thermal sensation in livingroom (recoded) * Floor level (4 groups)

Crosstab

			Floor level (4 groups)				
			Ground	First	Second	Third-fourth	Total
Thermal sensation in livingroom (recoded)	Cool	Count	2a	6a	3a	7a	18
		% within Floor level (4 groups)	11.1%	21.4%	15.8%	20.0%	18.0%
	Slightly cool	Count	4a	4a	5a	10a	23
		% within Floor level (4 groups)	22.2%	14.3%	26.3%	28.6%	23.0%
	Comfortable	Count	3a	3a	2a	5a	13
		% within Floor level (4 groups)	16.7%	10.7%	10.5%	14.3%	13.0%
	Slightly warm	Count	6a	9a	5a, b	1b	21
		% within Floor level (4 groups)	33.3%	32.1%	26.3%	2.9%	21.0%
	Warm	Count	1a	4a	4a	8a	17
		% within Floor level (4 groups)	5.6%	14.3%	21.1%	22.9%	17.0%
	Hot	Count	2a	2a	0a	4a	8
		% within Floor level (4 groups)	11.1%	7.1%	0.0%	11.4%	8.0%
Total		Count	18	28	19	35	100
		% within Floor level (4 groups)	100.0%	100.0%	100.0%	100.0%	100.0%

Each subscript letter denotes a subset of Floor level (4 groups) categories whose column proportions do not differ significantly from each other at the .05 level.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)	Point Probability
Pearson Chi-Square	16.089[a]	15	.376	.[b]		
Likelihood Ratio	20.887	15	.140	.[b]		
Fisher-Freeman-Halton Exact Test	18.153			.220		
Linear-by-Linear Association	.209[c]	1	.648	.657	.335	.020
N of Valid Cases	100					

a. 17 cells (70.8%) have expected count less than 5. The minimum expected count is 1.44.

b. Cannot be computed because there is insufficient memory.

c. The standardized statistic is -.457.

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	.401	.376
	Cramer's V	.232	.376
N of Valid Cases		100	

Figure S.4.1: Occupants' TSVs in living-room by floor level.

Supplementary Material 4: Sample of A Correlation Analysis

Thermal sensation in kitchen (recoded) * Floor level (4 groups)

Crosstab

			Floor level (4 groups)				
			Ground	First	Second	Third-fourth	Total
Thermal sensation in kitchen (recoded)	Cool	Count	5_a	5_a	4_a	14_a	28
		% within Floor level (4 groups)	27.8%	17.9%	21.1%	40.0%	28.0%
	Slightly cool	Count	4_a	9_a	4_a	11_a	28
		% within Floor level (4 groups)	22.2%	32.1%	21.1%	31.4%	28.0%
	Comfortable	Count	4_a	4_a	4_a	4_a	16
		% within Floor level (4 groups)	22.2%	14.3%	21.1%	11.4%	16.0%
	Slightly warm	Count	2_a	7_a	4_a	1_a	14
		% within Floor level (4 groups)	11.1%	25.0%	21.1%	2.9%	14.0%
	Warm	Count	2_a	2_a	2_a	1_a	7
		% within Floor level (4 groups)	11.1%	7.1%	10.5%	2.9%	7.0%
	Hot	Count	1_a	1_a	1_a	4_a	7
		% within Floor level (4 groups)	5.6%	3.6%	5.3%	11.4%	7.0%
Total		Count	18	28	19	35	100
		% within Floor level (4 groups)	100.0%	100.0%	100.0%	100.0%	100.0%

Each subscript letter denotes a subset of Floor level (4 groups) categories whose column proportions do not differ significantly from each other at the .05 level.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	14.788[a]	15	.467	.[b]	
Likelihood Ratio	15.838	15	.393	.533	
Fisher-Freeman-Halton Exact Test	15.347			.384	
Linear-by-Linear Association	1.232	1	.267	.[b]	.[b]
N of Valid Cases	100				

a. 15 cells (62.5%) have expected count less than 5. The minimum expected count is 1.26.

b. Cannot be computed because there is insufficient memory.

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	.385	.467
	Cramer's V	.222	.467
N of Valid Cases		100	

Figure S.4.2: Occupants' TSVs in kitchen by floor level.

Supplementary Material 4: Sample of A Correlation Analysis

Thermal sensation in bedroom 1 (recoded) * Floor level (4 groups)

Crosstab

			Floor level (4 groups)				
			Ground	First	Second	Third-fourth	Total
Thermal sensation in bedroom 1 (recoded)	Cold	Count	0a	0a	1a	1a	2
		% within Floor level (4 groups)	0.0%	0.0%	5.3%	2.9%	2.0%
	Cool	Count	2a	3a	2a	2a	9
		% within Floor level (4 groups)	11.1%	10.7%	10.5%	5.7%	9.0%
	Slightly cool	Count	2a	4a	1a	6a	13
		% within Floor level (4 groups)	11.1%	14.3%	5.3%	17.1%	13.0%
	Comfortable	Count	2a	2a	5a	7a	16
		% within Floor level (4 groups)	11.1%	7.1%	26.3%	20.0%	16.0%
	Slightly warm	Count	7a	11a	5a	11a	34
		% within Floor level (4 groups)	38.9%	39.3%	26.3%	31.4%	34.0%
	Warm	Count	4a	7a	5a	7a	23
		% within Floor level (4 groups)	22.2%	25.0%	26.3%	20.0%	23.0%
	Hot	Count	1a	1a	0a	1a	3
		% within Floor level (4 groups)	5.6%	3.6%	0.0%	2.9%	3.0%
Total		Count	18	28	19	35	100
		% within Floor level (4 groups)	100.0%	100.0%	100.0%	100.0%	100.0%

Each subscript letter denotes a subset of Floor level (4 groups) categories whose column proportions do not differ significantly from each other at the .05 level.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)	Point Probability
Pearson Chi-Square	9.418[a]	18	.949	.[b]		
Likelihood Ratio	10.934	18	.897	.970		
Fisher-Freeman-Halton Exact Test	10.226			.952		
Linear-by-Linear Association	.576[c]	1	.448	.467	.235	.019
N of Valid Cases	100					

a. 21 cells (75.0%) have expected count less than 5. The minimum expected count is .36.

b. Cannot be computed because there is insufficient memory.

c. The standardized statistic is -.759.

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	.307	.949
	Cramer's V	.177	.949
N of Valid Cases		100	

Figure S.4.3: Occupants' TSVs in bedroom 1 by floor level.

Supplementary Material 4: Sample of A Correlation Analysis

Thermal sensation in bedroom 2 (recoded) * Floor level (4 groups)

Crosstab

			Floor level (4 groups)				
			Ground	First	Second	Third-fourth	Total
Thermal sensation in bedroom 2 (recoded)	Cold	Count	0a	0a	0a	1a	1
		% within Floor level (4 groups)	0.0%	0.0%	0.0%	2.9%	1.0%
	Cool	Count	2a	1a	2a	3a	8
		% within Floor level (4 groups)	11.1%	3.6%	10.5%	8.6%	8.0%
	Slightly cool	Count	2a	5a	2a	5a	14
		% within Floor level (4 groups)	11.1%	17.9%	10.5%	14.3%	14.0%
	Comfortable	Count	3a	4a	5a	10a	22
		% within Floor level (4 groups)	16.7%	14.3%	26.3%	28.6%	22.0%
	Slightly warm	Count	7a	11a	4a	11a	33
		% within Floor level (4 groups)	38.9%	39.3%	21.1%	31.4%	33.0%
	Warm	Count	3a	6a	6a	5a	20
		% within Floor level (4 groups)	16.7%	21.4%	31.6%	14.3%	20.0%
	Hot	Count	1a	1a	0a	0a	2
		% within Floor level (4 groups)	5.6%	3.6%	0.0%	0.0%	2.0%
Total		Count	18	28	19	35	100
		% within Floor level (4 groups)	100.0%	100.0%	100.0%	100.0%	100.0%

Each subscript letter denotes a subset of Floor level (4 groups) categories whose column proportions do not differ significantly from each other at the .05 level.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	11.245[a]	18	.884	.[b]	
Likelihood Ratio	12.300	18	.831	.920	
Fisher-Freeman-Halton Exact Test	12.088			.891	
Linear-by-Linear Association	1.652	1	.199	.[b]	.[b]
N of Valid Cases	100				

a. 20 cells (71.4%) have expected count less than 5. The minimum expected count is .18.

b. Cannot be computed because there is insufficient memory.

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	.335	.884
	Cramer's V	.194	.884
N of Valid Cases		100	

Figure S.4.4: Occupants' TSVs in bedroom 2 by floor level.

Supplementary Material 4: Sample of A Correlation Analysis

Thermal sensation in bedroom 3 (recoded) * Floor level (4 groups)

Crosstab

			Floor level (4 groups)				
			Ground	First	Second	Third-fourth	Total
Thermal sensation in bedroom 3 (recoded)	Cool	Count	2a	1a	2a	3a	8
		% within Floor level (4 groups)	11.1%	3.6%	10.5%	8.6%	8.0%
	Slightly cool	Count	2a	6a	1a	2a	11
		% within Floor level (4 groups)	11.1%	21.4%	5.3%	5.7%	11.0%
	Comfortable	Count	3a	5a	4a	13a	25
		% within Floor level (4 groups)	16.7%	17.9%	21.1%	37.1%	25.0%
	Slightly warm	Count	6a	11a	5a	10a	32
		% within Floor level (4 groups)	33.3%	39.3%	26.3%	28.6%	32.0%
	Warm	Count	4a	4a	7a	7a	22
		% within Floor level (4 groups)	22.2%	14.3%	36.8%	20.0%	22.0%
	Hot	Count	1a	1a	0a	0a	2
		% within Floor level (4 groups)	5.6%	3.6%	0.0%	0.0%	2.0%
Total		Count	18	28	19	35	100
		% within Floor level (4 groups)	100.0%	100.0%	100.0%	100.0%	100.0%

Each subscript letter denotes a subset of Floor level (4 groups) categories whose column proportions do not differ significantly from each other at the .05 level.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)	Point Probability
Pearson Chi-Square	14.632[a]	15	.478	.[b]		
Likelihood Ratio	14.775	15	.468	.598		
Fisher-Freeman-Halton Exact Test	13.857			.489		
Linear-by-Linear Association	.085[c]	1	.771	.775	.400	.027
N of Valid Cases	100					

a. 16 cells (66.7%) have expected count less than 5. The minimum expected count is .36.

b. Cannot be computed because there is insufficient memory.

c. The standardized statistic is -.292.

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	.383	.478
	Cramer's V	.221	.478
N of Valid Cases		100	

Figure S.4.5: Occupants' TSVs in bedroom 3 by floor level.

Supplementary Material 4: Sample of A Correlation Analysis

Thermal sensation in livingroom (recoded) * Orientation (4 groups)

Crosstab

			Orientation (4 groups)				
			South	North East or North West	South West	South East	Total
Thermal sensation in livingroom (recoded)	Cool	Count	6a	4a	6a	2a	18
		% within Orientation (4 groups)	16.7%	11.4%	33.3%	18.2%	18.0%
	Slightly cool	Count	8a	5a	8a	2a	23
		% within Orientation (4 groups)	22.2%	14.3%	44.4%	18.2%	23.0%
	Comfortable	Count	5a	5a	1a	2a	13
		% within Orientation (4 groups)	13.9%	14.3%	5.6%	18.2%	13.0%
	Slightly warm	Count	8a	9a	2a	2a	21
		% within Orientation (4 groups)	22.2%	25.7%	11.1%	18.2%	21.0%
	Warm	Count	7a	8a	0a	2a	17
		% within Orientation (4 groups)	19.4%	22.9%	0.0%	18.2%	17.0%
	Hot	Count	2a	4a	1a	1a	8
		% within Orientation (4 groups)	5.6%	11.4%	5.6%	9.1%	8.0%
Total		Count	36	35	18	11	100
		% within Orientation (4 groups)	100.0%	100.0%	100.0%	100.0%	100.0%

Each subscript letter denotes a subset of Orientation (4 groups) categories whose column proportions do not differ significantly from each other at the .05 level.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)	Point Probability
Pearson Chi-Square	15.274[a]	15	.432	.[b]		
Likelihood Ratio	17.645	15	.282	.[b]		
Fisher-Freeman-Halton Exact Test	15.403			.379		
Linear-by-Linear Association	1.272[c]	1	.259	.269	.137	.014
N of Valid Cases	100					

a. 16 cells (66.7%) have expected count less than 5. The minimum expected count is .88.

b. Cannot be computed because there is insufficient memory.

c. The standardized statistic is -1.128.

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	.391	.432
	Cramer's V	.226	.432
N of Valid Cases		100	

Figure S.4.6: Occupants' TSVs in living-room by orientation.

Supplementary Material 4: Sample of A Correlation Analysis

Thermal sensation in kitchen (recoded) * Orientation (4 groups)

Crosstab

			Orientation (4 groups) South	North East or North West	South West	South East	Total
Thermal sensation in kitchen (recoded)	Cool	Count	11a, b	5b	10a	2a, b	28
		% within Orientation (4 groups)	30.6%	14.3%	55.6%	18.2%	28.0%
	Slightly cool	Count	12a	9a	4a	3a	28
		% within Orientation (4 groups)	33.3%	25.7%	22.2%	27.3%	28.0%
	Comfortable	Count	7a	7a	1a	1a	16
		% within Orientation (4 groups)	19.4%	20.0%	5.6%	9.1%	16.0%
	Slightly warm	Count	4a	5a	2a	3a	14
		% within Orientation (4 groups)	11.1%	14.3%	11.1%	27.3%	14.0%
	Warm	Count	1a	3a	1a	2a	7
		% within Orientation (4 groups)	2.8%	8.6%	5.6%	18.2%	7.0%
	Hot	Count	1a	6a	0a	0a	7
		% within Orientation (4 groups)	2.8%	17.1%	0.0%	0.0%	7.0%
Total		Count	36	35	18	11	100
		% within Orientation (4 groups)	100.0%	100.0%	100.0%	100.0%	100.0%

Each subscript letter denotes a subset of Orientation (4 groups) categories whose column proportions do not differ significantly from each other at the .05 level.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)	Point Probability
Pearson Chi-Square	23.373[a]	15	.077	.[b]		
Likelihood Ratio	23.472	15	.075	.134		
Fisher-Freeman-Halton Exact Test	19.719			.118		
Linear-by-Linear Association	.178[c]	1	.673	.694	.347	.024
N of Valid Cases	100					

a. 15 cells (62.5%) have expected count less than 5. The minimum expected count is .77.

b. Cannot be computed because there is insufficient memory.

c. The standardized statistic is .422.

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	.483	.077
	Cramer's V	.279	.077
N of Valid Cases		100	

Figure S.4.7: Occupants' TSVs in kitchen by orientation.

Supplementary Material 4: Sample of A Correlation Analysis

Thermal sensation in bedroom 1 (recoded) * Orientation (4 groups)

Crosstab

			Orientation (4 groups)				
			South	North East or North West	South West	South East	Total
Thermal sensation in bedroom 1 (recoded)	Cold	Count	2a	0a	0a	0a	2
		% within Orientation (4 groups)	5.6%	0.0%	0.0%	0.0%	2.0%
	Cool	Count	3a	2a	2a	2a	9
		% within Orientation (4 groups)	8.3%	5.7%	11.1%	18.2%	9.0%
	Slightly cool	Count	5a	4a	3a	1a	13
		% within Orientation (4 groups)	13.9%	11.4%	16.7%	9.1%	13.0%
	Comfortable	Count	7a	3a	5a	1a	16
		% within Orientation (4 groups)	19.4%	8.6%	27.8%	9.1%	16.0%
	Slightly warm	Count	15a	9a	6a	4a	34
		% within Orientation (4 groups)	41.7%	25.7%	33.3%	36.4%	34.0%
	Warm	Count	3a	15b	2a, b	3a, b	23
		% within Orientation (4 groups)	8.3%	42.9%	11.1%	27.3%	23.0%
	Hot	Count	1a	2a	0a	0a	3
		% within Orientation (4 groups)	2.8%	5.7%	0.0%	0.0%	3.0%
Total		Count	36	35	18	11	100
		% within Orientation (4 groups)	100.0%	100.0%	100.0%	100.0%	100.0%

Each subscript letter denotes a subset of Orientation (4 groups) categories whose column proportions do not differ significantly from each other at the .05 level.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)	Point Probability
Pearson Chi-Square	22.525[a]	18	.210	.[b]		
Likelihood Ratio	23.728	18	.164	.248		
Fisher-Freeman-Halton Exact Test	20.812			.176		
Linear-by-Linear Association	.081[c]	1	.777	.801	.405	.028
N of Valid Cases	100					

a. 21 cells (75.0%) have expected count less than 5. The minimum expected count is .22.

b. Cannot be computed because there is insufficient memory.

c. The standardized statistic is .284.

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	.475	.210
	Cramer's V	.274	.210
N of Valid Cases		100	

Figure S.4.8: Occupants' TSVs in bedroom 1 by orientation.

Supplementary Material 4: Sample of A Correlation Analysis

Thermal sensation in bedroom 2 (recoded) * Orientation (4 groups)

Crosstab

			Orientation (4 groups)				
			South	North East or North West	South West	South East	Total
Thermal sensation in bedroom 2 (recoded)	Cold	Count	1a	0a	0a	0a	1
		% within Orientation (4 groups)	2.8%	0.0%	0.0%	0.0%	1.0%
	Cool	Count	2a	3a	2a	1a	8
		% within Orientation (4 groups)	5.6%	8.6%	11.1%	9.1%	8.0%
	Slightly cool	Count	6a	4a	4a	0a	14
		% within Orientation (4 groups)	16.7%	11.4%	22.2%	0.0%	14.0%
	Comfortable	Count	10a	4a	5a	3a	22
		% within Orientation (4 groups)	27.8%	11.4%	27.8%	27.3%	22.0%
	Slightly warm	Count	14a	10a	6a	3a	33
		% within Orientation (4 groups)	38.9%	28.6%	33.3%	27.3%	33.0%
	Warm	Count	3a	13b	1a, b	3a, b	20
		% within Orientation (4 groups)	8.3%	37.1%	5.6%	27.3%	20.0%
	Hot	Count	0a	1a	0a	1a	2
		% within Orientation (4 groups)	0.0%	2.9%	0.0%	9.1%	2.0%
Total		Count	36	35	18	11	100
		% within Orientation (4 groups)	100.0%	100.0%	100.0%	100.0%	100.0%

Each subscript letter denotes a subset of Orientation (4 groups) categories whose column proportions do not differ significantly from each other at the .05 level.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)	Point Probability
Pearson Chi-Square	22.221[a]	18	.222	.[b]		
Likelihood Ratio	24.354	18	.144	.[b]		
Fisher-Freeman-Halton Exact Test	22.536			.121		
Linear-by-Linear Association	.765[c]	1	.382	.392	.203	.022
N of Valid Cases	100					

a. 20 cells (71.4%) have expected count less than 5. The minimum expected count is .11.

b. Cannot be computed because there is insufficient memory.

c. The standardized statistic is .874.

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	.471	.222
	Cramer's V	.272	.222
N of Valid Cases		100	

Figure S.4.9: Occupants' TSVs in bedroom 2 by orientation.

Supplementary Material 4: Sample of A Correlation Analysis

Thermal sensation in bedroom 3 (recoded) * Orientation (4 groups)

Crosstab

			Orientation (4 groups) South	North East or North West	South West	South East	Total
Thermal sensation in bedroom 3 (recoded)	Cool	Count	2a	3a	2a	1a	8
		% within Orientation (4 groups)	5.6%	8.6%	11.1%	9.1%	8.0%
	Slightly cool	Count	4a	2a	4a	1a	11
		% within Orientation (4 groups)	11.1%	5.7%	22.2%	9.1%	11.0%
	Comfortable	Count	14a	3b	5a, b	3a, b	25
		% within Orientation (4 groups)	38.9%	8.6%	27.8%	27.3%	25.0%
	Slightly warm	Count	12a	12a	5a	3a	32
		% within Orientation (4 groups)	33.3%	34.3%	27.8%	27.3%	32.0%
	Warm	Count	4a	13a	2a	3a	22
		% within Orientation (4 groups)	11.1%	37.1%	11.1%	27.3%	22.0%
	Hot	Count	0a	2a	0a	0a	2
		% within Orientation (4 groups)	0.0%	5.7%	0.0%	0.0%	2.0%
Total		Count	36	35	18	11	100
		% within Orientation (4 groups)	100.0%	100.0%	100.0%	100.0%	100.0%

Each subscript letter denotes a subset of Orientation (4 groups) categories whose column proportions do not differ significantly from each other at the .05 level.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)	Point Probability
Pearson Chi-Square	20.804[a]	15	.143	.[b]		
Likelihood Ratio	21.924	15	.110	.162		
Fisher-Freeman-Halton Exact Test	20.186			.094		
Linear-by-Linear Association	.010[c]	1	.922	.935	.476	.032
N of Valid Cases	100					

a. 17 cells (70.8%) have expected count less than 5. The minimum expected count is .22.

b. Cannot be computed because there is insufficient memory.

c. The standardized statistic is -.098.

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	.456	.143
	Cramer's V	.263	.143
N of Valid Cases		100	

Figure S.4.10: Occupants' TSVs in bedroom 3 by orientation.

Supplementary Material 4: Sample of A Correlation Analysis

Correlations

		Thermal sensation in livingroom (recoded)	Thermal sensation in kitchen (recoded)	Thermal sensation in bedroom 1 (recoded)	Thermal sensation in bedroom 2 (recoded)	Thermal sensation in bedroom 3 (recoded)
Thermal sensation in livingroom (recoded)	Pearson Correlation	1	.462**	.302**	.146	.200*
	Sig. (2-tailed)		<.001	.002	.147	.046
	N	100	100	100	100	100
Thermal sensation in kitchen (recoded)	Pearson Correlation	.462**	1	.133	.205*	.220*
	Sig. (2-tailed)	<.001		.187	.041	.028
	N	100	100	100	100	100
Thermal sensation in bedroom 1 (recoded)	Pearson Correlation	.302**	.133	1	.763**	.724**
	Sig. (2-tailed)	.002	.187		<.001	<.001
	N	100	100	100	100	100
Thermal sensation in bedroom 2 (recoded)	Pearson Correlation	.146	.205*	.763**	1	.829**
	Sig. (2-tailed)	.147	.041	<.001		<.001
	N	100	100	100	100	100
Thermal sensation in bedroom 3 (recoded)	Pearson Correlation	.200*	.220*	.724**	.829**	1
	Sig. (2-tailed)	.046	.028	<.001	<.001	
	N	100	100	100	100	100

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Figure S.4.11: Pearson's correlation analysis by exploring occupants' TSVs.

Supplementary Material 5: Contingency Tables

Information on the Presentation of Contingency Tables

In this study, whilst correlations are indicative of association, there is scope with the data to perform hypothesis testing of significant differences between variables that would add weight to the results. To provide a clear representation of the study findings and report research outcomes in accordance with the research questions, which were set out to develop a novel methodological framework for the universal applicability of the Energy Performance of Buildings Directives (EPBD) in the residential sector, the relevant concepts of statistical convention were presented in Chapter 3 Subsection 3.3.1 (The Concept of Statistical Representativeness), Subsection 3.3.2 (References to the Works of Other Scholars on Representativeness) and Subsection 3.3.3 (Sample Size Calculation Criteria).

This appendix presents the contingency tables that support the findings of the statistical analysis presented in Tables 4.3 and 4.4(a) and (b) in Chapter 4. These findings are presented in the appendix to provide guidance on the applied statistical method in order to comply with the convention of supporting research outcomes. It should be noted that Chapter 4 presents the identification of a “neutral” adaptive thermal comfort threshold by conducting statistical analysis with *in-situ* measurements, *on-site* environmental monitoring, and a thermal comfort assessment questionnaire survey to develop benchmarking criteria for the South-eastern Mediterranean climate of Cyprus. These contingency tables are presented in the appendix because, according to the conventions of thermal comfort studies, representation of households’ thermal sensation by using descriptive statistics, frequencies, Cramér’s *V* test, Pearson’s correlations and further ordinal logistic regression analysis methods could provide a valid background for the development of reliable thermal comfort thresholds. In Chapter 4, the convention of the thermal comfort assessment method was applied in accordance with the concept of a statistically representative sampling size, which was achieved by undertaking a longitudinal field survey. This field study enables researchers to present households’ *in-vivo* experiences on thermal satisfaction.

It must be stressed that the contingency tables were also produced for the S.6.1–4, S.6.5(a) and (b) and S.6.6(a) and (b) at the time of conducting Cramér’s *V* test to understand households’ socio-demographic characteristics and their habitual adaptive behaviour on energy use. The relevant

Cramer's V tests are presented in Appendix L, but, in order to avoid lengthy documentation, only the Cramer's V tests of the contingency tables for Tables 4.3(1) and (2) and Tables 4.4(b-1)–(b-5) are presented. In this appendix, the researcher decided to demonstrate the below-listed contingency tables because they are noteworthy contributions to the building and environment field where researchers could apply and adopt the statistical conventions presented in Chapter 4. At the same time, in Chapter 3, Subsections 3.3.1 and 3.3.2, scholars in the literature review recommend that reliable representativeness of sampling size within the variables identified to develop the concept of statistical convention plays an important role at the time of developing an evidence-based energy policy design. This is the reason that contingency tables for thermal comfort studies are not the primary factor used for identifying "neutral" adaptive thermal comfort thresholds, but the contingency tables are still presented. In this present study, Chapter 4 aims to demonstrate the longitudinal field survey findings with occupants' TSVs and, because of this, contingency tables are not presented in Chapter 4 but instead in the appendix to provide useful guidance for future scholars.

Important note about Tables 4.4(b-1)–(b-5): These are the contingency tables that support the statistical analysis in Table 4.3 in Chapter 4. In this statistical analysis, occupants' TSVs were identified as ordinal variables to conduct the Cramér's *V* test accurately. In the contingency tables presented in this appendix, it can be seen that household thermal sensation is represented by the terminology of "thermal feeling" indicators to provide a clear understanding to readers about household thermal sensation. It must be stressed that, in the dataset, the TSV code was set to [0 to 6] which represents the [-3, +3] thermal sensation band according to thermal comfort convention. Hence, the researcher decided to report the findings by using the terminology of each thermal feeling at the time of undertaking the statistical analysis for the contingency tables (Tables 4.4(b-1)–(b-5)).

Supplementary Material 5: Contingency Tables

Table S.5.1 [S.6.2]: Relationships Between Occupation, Age Bands, Economic Status and Education.

Occupation	Work outside the home		Work at home		Household activities		Retired				
	n	%	*n*	%	*n*	%	*n*	%	Fisher's Exact	*p*	Cramer's *V*
Age bands									44,81	<0,001	0,399
Less than 35	10	23,8 [a]	2	15,4 [a]	0	0,0 [a]	0	0,0 [a]			
35–45	10	23,8 [a]	2	15,4 [a]	4	23,5 [a]	0	0,0 [a]			
45–55	11	26,2 [a]	4	30,8 [a]	3	17,6 [a]	2	9,1 [a]			
55–65	9	21,4 [a]	4	30,8 [a]	9	52,9 [a]	8	36,4 [a]			
65 or over	2	4,8 [a]	1	7,7 [a]	1	5,9 [a]	12	54,5 [b]			
Economic status									8,69	0,743	0,171
Full time	18	42,9 [a]	4	30,8 [a]	7	41,2 [a]	6	27,3 [a]			
Part time	6	14,3 [a]	2	15,4 [a]	2	11,8 [a]	2	9,1 [a]			
Self employed	3	7,1 [a]	2	15,4 [a]	4	23,5 [a]	6	27,3 [a]			
Unemployed	5	11,9 [a]	1	7,7 [a]	2	11,8 [a]	4	18,2 [a]			
Pensioner	10	23,8 [a]	4	30,8 [a]	2	11,8 [a]	4	18,2 [a]			
Education									7,49	0,591	0,162
Elementary school	9	24,3 [a]	2	16,7 [a]	0	0,0 [a]	4	18,2 [a]			
Secondary school	12	32,4 [a]	4	33,3 [a]	6	35,3 [a]	8	36,4 [a]			
High school	13	35,1 [a]	5	41,7 [a]	7	41,2 [a]	7	31,8 [a]			
Undergraduate/Post graduate	3	8,1 [a]	1	8,3 [a]	4	23,5 [a]	3	13,6 [a]			

Note. For each row category, pairs of column proportions with different superscripts differed significantly, $p < 0,05$.

Supplementary Material 5: Contingency Tables

Table S.5.2 [S.6.3]: Relationships Between Household Age Bands, Tenure Type and Length of Residency.

Residency	Less than 5 years		5-10 years		More than 10 years				
	n	%	*n*	%	*n*	%	Fisher's Exact	*p*	Cramer's *V*
Age bands							56,03	<0,001	0,582
Less than 35	11	64,7 [a]	2	20,0 [a, b]	3	4,1 [b]			
35–45	5	29,4 [a]	6	60,0 [a]	5	6,8 [b]			
45–55	0	0,0 [a]	1	10,0 [a]	19	26,0 [a]			
55–65	1	5,9 [a]	1	10,0 [a, b]	30	41,1 [b]			
65 or over	0	0,0 [a]	0	0,0 [a]	16	21,9 [a]			
Tenure type							32	<0,001	0,595
Owner occupied	8	47,1 [a]	5	50,0 [a]	71	97,3 [b]			
Rented	9	52,9 [a]	5	50,0 [a]	2	2,7 [b]			

Note. For each row category, pairs of column proportions with different superscripts differed significantly, $p < 0,05$.

Supplementary Material 5: Contingency Tables

Table S.5.3 [S.6.4]: Relationships Between Household Income, Energy Advice and Energy Savings.

Income	Less than 2500 TL		2500−5000 TL		5000−7000 TL		More than 7000 TL		Fisher's Exact	*p*	Cramer's *V*
	n	%	*n*	%	*n*	%	*n*	%			
Energy advice									11,55	0,194	0,199
Famagusta Municipality	2	8,7 [a]	3	11,5 [a]	5	17,9 [a]	2	8,7 [a]			
The Electricity Authority	1	4,3 [a]	6	23,1 [a]	1	3,6 [a]	7	30,4 [a]			
None	18	78,3 [a]	16	61,5 [a]	20	71,4 [a]	13	56,5 [a]			
Other	2	8,7 [a]	1	3,8 [a]	2	7,1 [a]	1	4,3 [a]			
Energy savings									16,24	0,049	0,233
Nothing	4	17,4 [a, b]	8	30,8 [b]	1	3,6 [a]	2	8,7 [a, b]			
A little	3	13,0 [a]	5	19,2 [a]	4	14,3 [a]	0	0,0 [a]			
Some	9	39,1 [a]	9	34,6 [a]	17	60,7 [a]	15	65,2 [a]			
A lot	7	30,4 [a]	4	15,4 [a]	6	21,4 [a]	6	26,1 [a]			

Note. For each row category, pairs of column proportions with different superscripts differed significantly, $p < 0,05$.

Supplementary Material 5: Contingency Tables

Table S.5.4. [S.6.5-(a)]: Relationships Between Household Occupation and Cooling and Heating Energy Consumption Patterns.

Heating consumption patterns on the weekdays	0–4 hours		5–9 hours		More than 10 hours		Fisher's Exact	p	Cramer's V
	n	%	n	%	n	%			
Occupation							12,49	0,042	0,253
Work outside the hom	15	42,9 [a]	23	52,3 [a]	4	26,7 [a]			
Work at home	4	11,4 [a]	7	15,9 [a]	2	13,3 [a]			
Household activities	9	25,7 [a]	2	4,5 [b]	6	40,0 [a]			
Retired	7	20,0 [a]	12	27,3 [a]	3	20,0 [a]			
Cooling consumption patterns on weekdays							74,57*	<0,001	0,611
0–4 hours	21	58,3 [a]	0	0,0 [b]	0	0,0 [b]			
5–9 hours	15	41,7 [a]	23	47,9 [a]	0	0,0 [b]			
More than 10 hours	0	0,0 [a]	25	52,1 [b]	16	100,0 [c]			
Cooling consumption patterns on the weekend							49,70	<0,001	0,504
0–4 hours	20	55,6 [a]	0	0,0 [b]	0	0,0 [b]			
5–9 hours	6	16,7 [a]	15	31,3 [a]	3	18,8 [a]			
10–12 hours	8	22,2 [a]	21	43,8 [a]	5	31,3 [a]			
More than 12 hours	2	5,6 [a]	12	25,0 [a, b]	8	50,0 [b]			

Note. For each row category, pairs of column proportions with different superscripts differed significantly, $p < 0,05$. * indicates the chi-square test

Supplementary Material 5: Contingency Tables

Table S.5.5 [S.6.6-(a)]: Relationships Between Household Occupation, Window-Opening Patterns in the Winter, Window-Opening Patterns in the Summer and Heating Control.

Windows closing reasons	Against the warm air		To block outside disturbances/annoya		For safety reasons		Dust		Open always		Fisher's Exact	p	Cramer's V
	n	%	n	%	n	%	n	%	n	%			
Occupation											16,23	0,142	0,260
Work outside the home	16	59,3 [a]	7	41,2 [a]	4	36,4 [a]	11	39,3 [a]	4	36,4 [a]			
Work at home	4	14,8 [a]	4	23,5 [a]	1	9,1 [a]	4	14,3 [a]	0	0,0 [a]			
Household activities	4	14,8 [a]	1	5,9 [a]	5	45,5 [a]	6	21,4 [a]	1	9,1 [a]			
Retired	3	11,1 [a]	5	29,4 [a, b]	1	9,1 [a, b]	7	25,0 [a, b]	6	54,5 [b]			
Windows opening patterns in winter											15,02	0,197	0,230
0–2 hours	3	10,3 [a]	0	0,0 [a]	3	23,1 [a]	3	10,7 [a]	0	0,0 [a]			
2–4 hours	6	20,7 [a]	6	33,3 [a]	7	53,8 [a]	13	46,4 [a]	6	50,0 [a]			
4–6 hours	7	24,1 [a]	4	22,2 [a]	1	7,7 [a]	6	21,4 [a]	3	25,0 [a]			
More than 6 hours	13	44,8 [a]	8	44,4 [a]	2	15,4 [a]	6	21,4 [a]	3	25,0 [a]			
Windows opening patterns in summer											11,44	0,153	0,234
2–6 hours	6	20,7 [a]	0	0,0 [a]	3	23,1 [a]	4	14,3 [a]	1	8,3 [a]			
6–8 hours	5	17,2 [a]	5	27,8 [a]	6	46,2 [a]	11	39,3 [a]	4	33,3 [a]			
More than 8 hours	18	62,1 [a]	13	72,2 [a]	4	30,8 [a]	13	46,4 [a]	7	58,3 [a]			
Type of heating control at home											6,22	0,994	0,118
Wall mounted thermostat	4	13,8 [a]	3	16,7 [a]	2	15,4 [a]	5	17,9 [a]	3	25,0 [a]			
Other types	1	3,4 [a]	3	16,7 [a]	1	7,7 [a]	2	7,1 [a]	2	16,7 [a]			
Remote controller	4	13,8 [a]	1	5,6 [a]	1	7,7 [a]	2	7,1 [a]	1	8,3 [a]			
Automatic thermostat and	7	24,1 [a]	4	22,2 [a]	3	23,1 [a]	6	21,4 [a]	2	16,7 [a]			
None	13	44,8 [a]	7	38,9 [a]	6	46,2 [a]	13	46,4 [a]	4	33,3 [a]			

Note. For each row category, pairs of column proportions with different superscripts differed significantly, $p < 0,05$.

Supplementary Material 6: Cramer’s *V* tests

Table S.6.1: Relationship Between Respondents, Orientation and Floor Level.

Variables		Participants	Orientation	Floor Level
Participants	Partial η^2	1	0,781	0,046
	Significance	—	<0,001	0,205
Orientation	Partial η^2/ Cramer's *V*	0,781	1	0,197
	Significance	<0,001	—	0,234
Floor Level	Partial η^2/ Cramer's *V*	0,046	0,197	1
	Significance	0,205	0,188	—

Participants - Floor Level, $F(3, 96) = 1{,}56$, $p = 0{,}205$, *Partial* $\eta^2 = 0{,}046$
Orientation - Participants, $F(3, 96) = 113{,}90$, $p < 0{,}001$, *Partial* $\eta^2 = 0{,}781$
Orientation - Floor Level, *Fisher’s exact* = 12,11, $p = 0{,}188$, *Cramer*’s $V = 0{,}197$

Outcome: The relationship between orientation and floor level was examined by crosstabulation using chi-square test. As seen in Table S.6.1, no significant relationship was found, *Fisher’s exact* = 12,11, $p = 0{,}188$, *Cramer*’s $V = 0{,}197$.

Supplementary Material 6: Cramer's *V* Tests

Table S.6.2: Relationship Between Age Band, Occupation, Income and Health Conditions.

Research Questions	Age Band	Economic Status	Education Level	Occupa-tion	Income	Health
Q 1.1: What is your age?	1	0,201	0,114	0,399*	0,213	0,496*
	—	0,447	0,991	<0,001	0,245	0,001
Q 1.2: What is your economic status?	0,201	1	0,416*	0,171	0,136	0,178
	0,447	—	<0,001	0,743	0,950	0,708
Q 1.3: What is your highest level of education?	0,114	0,416*	1	0,162	0,190	0,196
	0,991	<0,001	—	0,591	0,386	0,315
Q 1.4: What is your occupation?	0,399*	0,171	0,162	1	0,174	0,342*
	<0,001	0,743	0,591	—	0,488	<0,001
Q 29: What is your monthly income?	0,213	0,136	0,190	0,174	1	0,174
	0,245	0,950	0,386	0,488	—	0,480
Q 28: How is your health in general?	0,496*	0,178	0,196	0,342*	0,174	1
	0,001	0,708	0,315	<0,001	0,480	—

Health condition - Age, *Fisher's exact* = 73,74, p < 0,001, *Cramer*'s V = 0,496
Health condition – Employment status, *Fisher's exact* = 9,06, p = 0,708, *Cramer*'s V = 0,178
Health condition – Education level, *Fisher's exact* = 10,20, p = 0,315, *Cramer*'s V = 0,196
Health condition – Occupation, *Fisher's exact* = 33,81, p < 0,001, *Cramer*'s V = 0,342
Health condition – Income, *Fisher's exact* = 8,63, p = 0,472, *Cramer*'s V = 0,170
Income - Age, *Fisher's exact* = 14.71, p = 0,245, *Cramer*'s V = 0,213
Income – Employment status, *Fisher's exact* = 5,64, p = 0,950, *Cramer*'s V = 0,136
Income – Education level, *Fisher's exact* = 9.53, p = 0,386, *Cramer*'s V = 0,190
Income – Occupation, *Fisher's exact* = 8,49, p = 0,488, *Cramer*'s V = 0,174
Occupation – Age, *Fisher's exact* = 44,81, p < 0,001, *Cramer*'s V = 0,399
Occupation – Employment status, *Fisher's exact* = 8.69, p = 0,743, *Cramer*'s V = 0,171
Occupation – Education, *Fisher's exact* = 7,49, p = 0,591, *Cramer*'s V = 0,162
Age – Education, *Fisher's exact* = 4,08, p = 0,991, *Cramer*'s V = 0,114
Age – Employment status, *Fisher's exact* = 16.09, p = 0,447, *Cramer*'s V = 0,201
Employment status – Education, *Fisher's exact* = 48.81, p < 0,001, *Cramer*'s V = 0,416

Outcome: Age bands were significantly related to the health conditions, and this relationship was strong (*Fisher's exact* = 73,74, p < 0,001, *Cramer*'s V = 0,496). Younger age appeared to report better health condition (good or very good) than older age. Household occupation was significantly related to health conditions, and the relationship was moderate (*Fisher's exact* = 33,81, p < 0,001, *Cramer*'s V = 0,342). A greater proportion of participants with very good condition worked outside the home than those with mediocre health conditions, whereas none of the participants who retired had very good conditions. Household occupations were also significantly associated with age with a moderate-strong relationship (*Fisher's exact* = 44,81, p < 0,001, *Cramer*'s V = 0,399). Economic status was significantly related to educational level (*Fisher's exact* = 48.81, p < 0,001, *Cramer*'s V = 0,416).

Note: The relationships between occupant age, economic status, education level, occupation, income, and health conditions were examined using crosstabulations with chi-square tests or Fisher's Exact tests if over 25% of cells had less than 5 expected counts, as shown in Table S.6.2.

Supplementary Material 6: Cramer's *V* Tests

Table S.6.3: Relationships Between Household Age Band, Tenancy Status and Length of Residency.

Research Questions	Age Band	Tenancy Status	Length of Residency
Q 1.1: What is your age?	1	0,595*	0,582*
	—	<0,001	<0,001
Q 3: Do you own or rent your dwelling?	0,595*	1	0,494*
	<0,001	—	<0,001
Q 2: How many years have you lived in this flat?	0,582*	0,494*	1
	<0,001	<0,001	—

Age – Length of Residency, *Fisher's exact* = 56.03, $p < 0,001$, *Cramer*'s $V = 0,582$

Age – Tenure Type, *Fisher's exact* = 21.28, $p < 0,001$, *Cramer*'s $V = 0,494$

Tenure Type – Length of Residency, *Fisher's exact* = 32.00, $p < 0,001$, *Cramer*'s $V = 0,595$

Outcome: Age band and tenancy status were significantly and strongly related to each other (*Fisher's exact* = 21.28, $p < 0,001$, Cramer's $V = 0,494$). Most participants who rented were 45 years old and younger. Age band and length of residency were also correlated (*Fisher's exact* = 56.03, $p < 0,001$, Cramer's $V = 0,582$). It appeared that younger participants had shorter length of residency. A strong relationship was discerned between tenancy status and length of residency (*Fisher's exact* = 32.00, $p < 0,001$, Cramer's $V = 0,595$). A greater proportion of participants with more than 10 years of residency were owner-occupied than those had 10 years or less residency.

Supplementary Material 6: Cramer's *V* Tests

Table S.6.4: Relationships Between Household Income, Energy Consumption, Energy Advice, Energy Usage and Energy Savings.

Research Questions	Income	Energy Consump-tion	Energy Advice	Energy Usage	Energy Savings
Q 29: What is your monthly income?	1	0,247	0,199	0,176	0,233*
	—	0,108	0,194	0,340	0,049
Q 30: How much electricity (in kWh) did you consume in May through September according to this last overview?	0,247	1	0,116	0,080	0,106
	0,108	—	0,702	0,416	0,754
Q 31: Have you received advice on how to reduce your energy bills?	0,199	0,116	1	0,117	0,182
	0,194	0,702	—	0,826	0,364
Q 4: Do you check your use of electricity by taking frequent meter readings?	0,176	0,080	0,117	1	0,291*
	0,340	0,416	0,826	—	0,040
Q 6: Do you know anything about energy-saving methods?	0,233*	0,106	0,182	0,291*	1
	0,049	0,754	0,364	0,040	—

Energy Advise – Income, *Fisher's exact* = 11,55, p = 0,194, *Cramer*'s V = 0,199
Energy Advise – Energy Savings, *Fisher's exact* = 9,07, p = 0,364, *Cramer*'s V = 0,182
Energy Advise – Energy Usage, *Fisher's exact* = 1,09, p = 0,826, *Cramer*'s V = 0,117
Energy Advise – Energy Consumption, *Fisher's exact* = 1,42, p = 0,702, *Cramer*'s V = 0,116
Energy Savings – Income, *Fisher's exact* = 16,24, p = 0,049, *Cramer*'s V = 0,233
Energy Savings – Energy Usage, *Fisher's exact* = 7,87, p = 0,040, *Cramer*'s V = 0,291
Energy Savings – Energy Consumption, *Fisher's exact* = 1,27, p = 0,754, *Cramer*'s V = 0,106
Income – Energy Usage, *Fisher's exact* = 3,26, p = 0,340, *Cramer*'s V = 0,176
Income – Energy Consumption, $\chi^2(3)$ = 6,09, p = 0,107, *Cramer*'s V = 0,247
Energy Consumption – Energy Usage, $\chi^2(1)$ = 0,64, p = 0,424, *Cramer*'s V = 0,080

Outcome: A moderate relationship was found between energy efficiency awareness and energy conservation (*Fisher's exact* = 7,87, p = 0,040, *Cramer's V* = 0,291). A greater proportion of participants with energy conservation had a lot of energy efficiency awareness while a greater proportion of participants without energy conservation did not have efficiency awareness at all. Income was moderately associated with energy savings (*Fisher's exact* = 16,24, p = 0,049, *Cramer*'s V = 0,233), but income was not significantly associated with energy consumption, energy efficiency awareness, and energy advice.

Note: Table S.6.4 demonstrates the crosstabulation using chi-square analysis or or Fisher's Exact tests if over 25% of cells had less than 5 expected counts that revealed relationships between household income level and occupant awareness of energy consumption.

Supplementary Material 6: Cramer's *V* Tests

Table S.6.5(a): Relationships Between Household Occupation and Cooling and Heating Energy Consumption Patterns.

Research Questions	Occupation	Weekday Heating-Consumption Patterns	Weekend Heating-Consumption Patterns	Weekday Cooling-Consumption Patterns	Weekend Cooling-Consumption Patterns
Q 1.4: What is your occupation?	1	0,253*	0,109	0,098	0,167
	—	0,042	0,896	0,955	0,579
Q 16: When do you turn on heating device(s) on weekdays?	0,253*	1	0,373*	0,611*	0,504*
	0,042	—	<0,001	<0,001	<0,001
Q 17: When do you turn on heating device(s) on the weekend?	0,109	0,373*	1	0,522*	0,706*
	0,896	<0,001	—	<0,001	<0,001
Q 12: When do you turn on cooling device(s) on weekdays?	0,098	0,611*	0,522*	1	0,774*
	0,955	<0,001	<0,001	—	<0,001
Q 13: When do you turn on cooling device(s) on the weekend?	0,167	0,504*	0,706*	0,774*	—
	0,579	<0,001	<0,001	<0,001	1

Occupation – Weekend heating consumption, *Fisher's exact* = 2,41, p = 0,896, *Cramer*'s V = 0,109

Occupation – Weekday heating consumption, *Fisher's exact* = 12,49, p = 0,042, *Cramer*'s V = 0,253

Occupation – Weekend cooling consumption, *Fisher's exact* = 7,63, p = 0,579, *Cramer*'s V = 0,167

Occupation – Weekday cooling consumption, *Fisher's exact* = 1,76, p = 0,955, *Cramer*'s V = 0,098

Weekday cooling consumption – Weekend heating consumption, $\chi^2(4)$ = 54,59, p < 0,001, *Cramer's V* = 0,522

Weekday cooling consumption – Weekday heating consumption, $\chi^2(4)$ = 74,57, p < 0,001, *Cramer's V* = 0,611

Weekday cooling consumption – Weekend cooling consumption, $\chi^2(6)$ = 119,77, p < 0,001, *Cramer's V* = 0,774

Weekend cooling consumption– Weekend heating consumption, $\chi^2(6)$ = 99,69, p < 0,001, *Cramer*'s V = 0,706

Weekend cooling consumption– Weekday heating consumption, *Fisher's exact* = 49,70, p < 0,001, *Cramer*'s V = 0,504

Weekday heating consumption– Weekend heating consumption, $\chi^2(4)$ = 27,89, p < 0,001, *Cramer*'s V = 0,373

Supplementary Material 6: Cramer's *V* Tests

Table S.6.5(b): Relationships Between Household Occupation, and Cooling and Heating Energy Consumption Patterns. (Continued).

Outcome: Weekday cooling consumption patterns were significantly and strongly related to weekend heating consumption patterns on weekend (χ^2 = 54,59, *p* < 0,001, *Cramer's V* = 0,522). Specifically, longer duration of heating consumption was related to longer duration of cooling consumption. Similar result was also found between weekend cooling consumption patterns and weekend heating consumption patterns (χ^2 = 99,69, *p* < 0,001, *Cramer's V* = 0,706), between weekday cooling consumption patterns and weekday heating consumption patterns (χ^2 = 74,57, *p* < 0,001, *Cramer's V* = 0,611), and weekend cooling consumption patterns and weekday heating consumption patterns (*Fisher's exact* = 49,70, *p* < 0,001, *Cramer's V* = 0,504). For heating patterns, weekday consumption was moderately associated with weekend consumption (χ^2 = 27,89, *p* < 0,001, *Cramer's V* = 0,373). For cooling patterns, weekday consumption was strongly associated with weekend consumption (χ^2 = 119,77, *p* < 0,001, *Cramer's V* = 0,774). Occupation was only significantly and moderately related to weekly heating consumption (*Fisher's exact* = 12,49, *p* = 0,042, *Cramer's V* = 0,253), but it was not significant related to any other cooling or heating consumption patterns.

Note: The relationships between household occupation and cooling and heating energy consumption patterns were examined using crosstabulations with chi-square tests or Fisher's Exact tests if over 25% of cells had less than 5 expected counts, as shown in Table S.6.5(a).

Supplementary Material 6: Cramer's *V* Tests

Table S.6.6(a): Relationships Between Household Occupation, Window-Opening Patterns in the Winter, Window-Opening Patterns in the Summer and Heating Control.

Research Questions	Occupation	Winter	Summer	Heating Control
Q 1.4: What is your occupation?	1	0,164	0,167	0,170
	—	0,594	0,525	0,709
Q 23: When do you open your windows in the winter?	0,164	1	0,459*	0,139
	0,594	—	<0,001	0,961
Q 18: When do you open your windows in the summer?	0,167	0,459*	1	0,283
	0,525	<0,001	—	0,053
Q 24: Do you keep room doors open when you do not have heating on?	0,114	0,246	0,030	1
	0,747	0,109	0,956	—
Q 19: Do you keep room doors open when you do not have cooling on?	0,114	0,262	0,168	0,229
	0,746	0,113	0,245	0,262
Q 20: Why do you open the windows?	0,201	0,124	0,185	0,200
	0,291	0,676	0,182	0,410
Q 21: Why do you close the windows?	0,260	0,230	0,234	0,118
	0,142	0,197	0,153	0,994

Supplementary Material 6: Cramer's *V* Tests

Table S.6.6(b): Relationships Between Household Occupation, Window-Opening Patterns in the Winter, Window-Opening Patterns in the Summer and Heating Control (Continued).

Windows opening patterns winter – Doors opening preference in summer, *Fisher's exact* = 5,82, p = 0,113, *Cramer*'s V = 0,262

Windows opening patterns winter – Windows opening pattern in summer, *Fisher's exact* = 36.80, p < 0,001, *Cramer*'s V = 0,459

Windows opening patterns winter – Doors opening preference in winter, $\chi^2(3)$ = 6,05, p = 0,109, *Cramer*'s V = 0,246

Windows opening patterns winter – Windows opening reason, $\chi^2(3)$ = 1,53, p = 0,676, *Cramer*'s V = 0,124

Windows opening patterns winter – Windows closing reason, *Fisher's exact* = 15,02, p = 0,197, *Cramer*'s V = 0,230

Windows opening patterns summer – Doors opening preference in summer, $\chi^2(2)$ = 2,81, p = 0,245, *Cramer*'s V = 0,168

Windows opening patterns summer – Doors opening preference in winter, $\chi^2(2)$ = 0,09, p = 0,956, *Cramer*'s V = 0,030

Windows opening patterns summer – Windows opening reason, $\chi^2(2)$ = 3,41, p = 0,182, *Cramer*'s V = 0,185

Windows opening patterns summer – Windows closing reason, *Fisher's exact* = 11,44, p = 0,153, *Cramer*'s V = 0,234

Occupation – Windows opening pattern in summer, *Fisher's exact* = 5,18, p = 0,525, *Cramer*'s V = 0,167

Occupation – Windows opening pattern in winter, *Fisher's exact* = 7,35, p = 0,594, *Cramer*'s V = 0,164

Occupation – Windows opening reasons, $\chi^2(3)$ = 3,74, p = 0,291, *Cramer*'s V = 0,201

Occupation – Windows closing reasons, *Fisher's exact* = 16,23, p = 0,142, *Cramer*'s V = 0,260

Occupation – Heating control, *Fisher's exact* = 8,94, p = 0,709, *Cramer*'s V = 0,170

Heating control – Windows opening pattern in winter, *Fisher's exact* = 5,41, p = 0,961, *Cramer*'s V = 0,139

Heating control – Windows opening pattern in summer, *Fisher's exact* = 14,14, p = 0,053, *Cramer*'s V = 0,283

Heating control – Windows opening reason, $\chi^2(4)$ = 3,93, p = 0,410, *Cramer*'s V = 0,200

Heating control – Windows closing reason, *Fisher's exact* = 6,22, p = 0,994, *Cramer*'s V = 0,118

Outcome: As seen in Table L.6(b), windows opening patterns in winter was significantly associated with that in summer (*Fisher's exact* = 36.80, p < 0,001, Cramer's V = 0,459), and this relationship was moderate-strong. It appears that longer opening duration in winter was related to longer opening duration in summer. Heating control type at home and occupation were not significantly associated with any opening patterns or opening reasons.

Note: The relationships between household occupation, window-opening patterns in the winter, window-opening patterns in the summer and heating control were examined using crosstabulations with chi-square tests or Fisher's Exact tests if over 25% of cells had less than 5 expected counts, as shown in Table S.6.6(a).

Supplementary Material 7: Worldwide Statistically Representative Archetypes

Table S.7.1: The Literature on Building Stock Aggregation through Archetype Buildings – Energy Modelling.

References	A. Study Location	B. Building Type	C. Sampling Size	D. Primary Aim of Model	E. Methodology	F. Main Findings
Streicher *et al.* (2018)	Sweden	SFHs and MFHs located in urban, suburban and rural regions	10,400 Cantonal Building Energy Certificates; 54 archetypes were identified; SFHs represent 14.4% and 14.1% of the total housing stock; MFHs represent 55% of the total housing stock	To provide the current thermal performance level of the Swiss residential building stock in 2015 based on an analysis of the data available from energy certificates; to estimate a thermal performance level of archetype buildings and their respective building elements	Cantonal Building Energy Performance Certificate data was used; archetype housing stock data was analysed; the statistics of the Federal Register of Buildings and Dwellings were examined; current state of the Swiss residential building stock was assessed; potential improvements in the building stock were proposed	Approximately 75% of all building elements do not yet reach the thermal performance of buildings constructed in the last 15 years; the *U*-values of building elements would have to be reduced by an additional 0.3-0.7 W/m^2K to reach low-energy standards
Yi and Peng (2019)	Seoul, South Korea	High-rise apartment stock, MFHs	The total number of households was 2,830,857, of which about 58% (1,641,383) lived in apartments	To introduce an 'archetype-in-neighbourhood' framework to develop cooling energy supply planning; to demonstrate estimated increases of the maximal month cooling energy demand	A bottom-up hybrid approach was adopted; empirical urban data modelling was used; EnergyPlus was used for model calibration; electricity use data of 659 apartment buildings (51,351 households) sampled from 18 city districts; characteristics of residential energy use during the hottest month (August) was examined	The coefficient of determination from the scatter plot between the observed and the predicted was 0.969, representing about 97% variance in observed peak cooling energy use
Li *et al.* (2019)	Chongqing, China	Urban residential blocks; MFHs in urban residential districts	Households with one elderly retiree and elderly retired couples accounted for 4.86% and 4.40%, respectively, of all households	To develop a localised residential building stock-space heating and cooling modelling approach to estimate energy consumption and related carbon emissions	A bottom-up engineering approach was used; development of residential archetypes; space heating and cooling energy consumption simulation and aggregation; stock total floor area calculation and construction age distribution; EnergyPlus was used for model calibration	The total energy consumption for space heating and cooling can be significantly reduced, with estimated reductions of 57.6%–60.70% in 2020 and 55.3%–57.2% in 2050
Pittau *et al.* (2019)	EU 28	SFHs; MFHs	Post-war buildings built between 1945 and 1969, representing 30% of housing stock	To investigate the effect of massively storing carbon in bio-based construction products	A statistic-based geo-cluster model was developed; a dynamic life-cycle assessment was performed; Eurostat data was used for the archetype housing identification	Up to 3% of the total GHG annually emitted by all sectors in 2015 can be removed by 2050

Supplementary Material 7: Worldwide Statistically Representative Archetypes

Table S.7.2: The Literature on Building Stock Aggregation through Archetype Buildings – Energy Modelling. (Continued)

References	A. Study Location	B. Building Type	C. Sampling Size	D. Primary Aim of Model	E. Methodology	F. Main Findings
Wang *et al.* (2020)	Netherlands, Amsterdam	Residential buildings	- 2178 buildings were included then this data was aggregated at the post code level rather than an individual building - 84 residential postcodes fulfil simulation data requirements	To present a data-driven urban scale energy modelling framework from open-source data harmonization, sensitivity analysis, heating demand simulation at the postcode level to develop energy modelling phase	- City Sim energy modelling software was used to undertake dynamic thermal simulations - Data inputs; weather data, building geometry, construction data, system data, operation data, energy consumption data - Energy model was validated with six years of measured consumption data	- Generalizability of the methodology was developed - Data availability and levels of detail integrated into the BEM. - Accelerating building stock retrofit was presented in urban scale
Molina *et al.* (2019)	South America, Chile	Residential buildings	- 83,887 households and 266,968 people living in 15 regions of the country is considered nationally representative - 15,312 dwellings were selected a representative and randomly sampled group - 496 archetypes were used to classify the housing stock	- To develop archetypal Chilean houses with statistically significant representative values for design optimization parameters related to energy use and indoor air quality - To analyse data and develop housing archetypes can be applied to other building stocks in comparable economically developed countries	- National census data was gathered - Building Permit's dataset was gathered -National socioeconomic characterization survey was used - The Use of Time National Survey database was used to demonstrate a nationwide cross-sectional study - National housing monitoring network data was used - Dwelling quantity, year of construction, occupancy, cooking and heating, construction and finishing materials were used	- Models of the archetypes require calibration after the simulation process to minimize the predictive error - A stochastic approach is required to capture both stock variability and parametric uncertainty - Categorical descriptors and occupant behaviors and poor granularity of physical data were identified as knowledge gap
Molina *et al.* (2019)	Santiago, Chile	Residential buildings	- In their database set, 7,19 and 46 archetypes were found to represent 32%, 58% and 83% of the national stock - A detached single storey house was identified as nationally representative archetype	To demonstrate the simulation of indoor environment conditions using representative models of a housing stock, is a more common method of investigation for policy making decisions	-A single multi-zone archetype is modelled in CONTAM where model inputs are randomly selected - The archetypes were identified by geometry, building size, dwelling type and construction period, values for the floor area and the number of storeys and number of occupants assigned into a simulation model	The outputs can be used to inform future standards and guidelines for Chilean houses that simultaneously focus on energy demand reduction

Supplementary Material 7: Worldwide Statistically Representative Archetypes

Table S.7.3: The Literature on Building Stock Aggregation through Archetype Buildings – Energy Modelling. (Continued)

References	A. Study Location	B. Building Type	C. Sampling Size	D. Primary Aim of Model	E. Methodology	F. Main Findings
Molina *et al.* (2019)	Santiago, Chile	Residential buildings	- 496 common Chilean archetypes were used - A detached single storey house was identified as nationally representative archetype - 8 probabilistic inputs were used	- To demonstrate the Chilean Housing Archetypes Air quality model and a stochastic framework for predicting uncertainties in indoor pollutant concentrations - To predict uncertainties in indoor pollutant concentrations, ventilation and infiltration rates and their associated energy demand	- CONTAM – open-source freely available multi-zone indoor ventilation and pollutant transport tool was used - Dwelling parameters – airflow paths and envelope air permeability were used - Both indoor and outdoor environment parameters were used -Pollutants – emission rates from cooking, emission rates from heating were used - Occupancy and activity data were assigned - Sensitivity analysis and statistical tests were conducted	- 34% of Chilean dwellings are predicted to have unacceptable daily $PM_{2.5}$ concentrations if their windows are closed at all times. - Many houses require remediation measures to improve airtightness and to reduce their annual space heating demand
Li *et al.* (2018)	Yuzhong District, Chongqing (China)	Residential buildings	- A population of around 649.500 was identified in the district - 334 of residential buildings accounting for 60% of the total as a sampling size	- To test and demonstrate a regional district retrofitting scheme by comparing the Energy Use Intensity (EUI) of the nationally representative archetypes	- Building shape, the glazing ratio, building envelope properties, occupancy pattern and heating/cooling equipment were used in the BES model - Correlation analysis for cluster variable selection was conducted - Energy consumption simulations – building-by-building approach were undertaken	A very small variation was found in the estimated stock (+0.03% in heating energy consumption and +2.97% in cooling energy consumption)
De Lemos Martins *et al.* (2019)	Toulouse, France	Residential buildings	Three representative urban blocks were identified	- To identify the statistical sensitivity of multi-scale urban design factors regarding buildings' energy demand	- Urban archetypes characterization and parametrization - Sensitivity analysis was conducted - Statistical sensitivity analysis with a variance-based method was applied	The courtyard aspect ratio and the standard-deviation of the built height account for almost 50% of the overall impact on heating demand
Tardioli *et al.* (2018)	Geneva, Switzerland	Residential buildings	A total of 67 representative buildings were identified in the urban dataset	- To present a novel methodology for identifying building groups and associated representative buildings in urban datasets	- Clustering algorithms are applied to the datasets - Data-cleaning, skew distributions and the presence of outliers were conducted - Datasets are classified using building typology information	Clustering techniques achieved an average accuracy of 89.6% across many building typologies and up to 97% in some instances

Supplementary Material 7: Worldwide Statistically Representative Archetypes

Table S.7.4: The Literature on Building Stock Aggregation through Archetype Buildings – Energy Modelling. (Continued)

References	A. Study Location	B. Building Type	C. Sampling Size	D. Primary Aim of Model	E. Methodology	F. Main Findings
Bianco and Marmori (2021)	Italy	Single and Multi-family houses	Typology 6 was selected as an archetype to represent housing stock built after 2005 – represents 8,000 new buildings per year	- To estimate the energy savings obtained when specific energy efficiency measures are applied - To bridge the identified gap and introduce a novel calculation tool	- Geometric and thermal features of buildings were used - I-REM energy modelling framework was used - Energy consumption data was extracted from the Eurostat and Odysee databases for the validation study	- A saving of 76.8 kWh is fixed for 2030, the double with respect to EU Policy scenario, and 100 kWh for 2040
McKenna *et al.* (2013)	Germany	- Single- and two-family houses - Multi-family houses	-10,000 objects related to energy use was used - 4,575 single- and two-family houses - 5,491 multi-family houses were selected as an archetype	-To analyse the role of refurbishment measures on the reaching, or not of these energy political goals by developing an aggregated building-stock model	- Building stock projections data 2011 to 2050 was gathered - Micro-census survey data was used - Age categorization of housing stock was applied - Renovation measures were predicted by using statistical analysis methods - Sensitivity analysis was conducted to identify effective renovation measures	-The renovation probability of the SFH is increased by 2020 from 1% to 4%. -The model results regarding total final energy demand are significantly higher than in other studies
Famuyibo *et al.* (2012)	Ireland	Residential buildings	13 representative archetypes were identified for the statistical analysis	-To present a methodology for the development of archetypes based on information from literature and a sample of detailed energy-related housing data	- Multilinear regression analysis, clustering and descriptive statistics were used - A housing database was used - The Energy Performance Survey of Irish Housing and the Irish National Survey of Housing Quality databases were used	The linear regression indicates a coefficient of determination R^2 0.391, indicating that 39.1% of the variance in household total energy use
Ahern and Norton (2020)	Ireland	Residential buildings	35 reference dwellings were selected to appropriately characterize 406,918 dwellings	-To present a methodology for the derivation of simplified default-free inputs to a bottom-up energy modelling	- Energy Use Intensity of housing stock was calculated - Housing typology was identified - Analysis of single field empirical data was applied -Heat loss throughout the building fabric was investigated -Orientation factor was considered	The recommended default U-values for walls and roofs for a different dwelling typology correlate with those recommended for the building typology examined

Supplementary Material 7: Worldwide Statistically Representative Archetypes

Table S.7.5: The Literature on Building Stock Aggregation through Archetype Buildings – Energy Modelling. (Continued)

References	A. Study Location	B. Building Type	C. Sampling Size	D. Primary Aim of Model	E. Methodology	F. Main Findings
Pittam *et al.* (2014)	Cork City, Ireland	Residential buildings	- 20 house typologies represent a total of 10,449 housing units - 4 representative archetypes were selected	-To develop representative archetypes using a bottom-up approach for stock modelling	-The GIS web-based mapping application was used -Geometric properties and thermal characteristics for each typology were used -Statistical analysis was used to aggregate bottom-up housing model	-A 6% variation is recorded when compared to an averaged 121 archetypes
Kazas *et al.* (2017)	Turin, Italy	Residential buildings	-Representative urban district was selected to develop energy model -300 scenarios were developed and simulated in a parametric analysis	-To generate detailed thermal energy profiles, at an urban district scale has been developed -To developed engineering bottom-up housing model for policy making decisions in energy use	-A parametric analysis of the variables of building energy performance was carried at an urban scale -Thermal energy demand profile database was developed -The application of the aggregation method to generate the average thermal energy demand profile was developed -Parametric analysis was conducted to develop cost-effective and energy efficient design scenarios	The development of an aggregation method, which could cover the uncertainties of a parametric analysis and would be sufficiently precise and general to be suitable for the existing housing stock
Allacker *et al.* (2018)	EU-27 Member states	Residential buildings	-24 representative residential buildings were selected as an archetype within the EU	- To develop a methodological framework that bridges a gap by using the outcomes from the BES and LCCA studies - To evaluate both energy- and cost- effectiveness of optimization scenarios	- Dynamic thermal simulations were undertaken - Life-cycle-cost-assessment method was applied - Archetypes were selected from the Typology Approach for Building Stock Energy Assessment database	The use of archetypes is useful for analyzing the effects of scenarios acting at the European level but implies also a certain degree of approximation at the building level
Ortiz and Bluyssen (2018)	Netherlands, Amsterdam	Residential buildings	-316 emails were sent out inviting participants to complete the questionnaire -Six groups were represented with the 25 final predicting variables	-To demonstrate the effectiveness of the TwoStep cluster analysis and the development -To present the results of a new questionnaire survey for measuring comfort, health and energy habits	-The questionnaire survey was developed -The questionnaire was developed with the Qualtrics online platform -SPSS software suite was used to conduct the relevant statistical analysis	The archetype analysis represents the greatest social challenge and thus should be considered a 'high priority' method of design to explore subject respondents' behaviour on energy use

Supplementary Material 7: Worldwide Statistically Representative Archetypes

Table S.7.6: The Literature on Building Stock Aggregation through Archetype Buildings – Energy Modelling. (Continued)

References	A. Study Location	B. Building Type	C. Sampling Size	D. Primary Aim of Model	E. Methodology	F. Main Findings
Zhang *et al.* (2019)	Singapore	High-rise residential buildings	30 urban blocks in six typologies that represent a diverse range of urban forms were examined	- To identify correlations between solar harvesting potential and energy consumption - To develop optimization scenarios for planning of appropriate building geometry	- Building energy modelling approach was adopted - EnergyPlus software suite was used - Urban planning layout and geometric parameters were constructed - Sky Exposure Factor and Sky View Factor were calculated concurrently - Solar energy harvesting potential was calculated - Regression analysis was conducted for the energy optimization scenarios	- The building cooling load intensity of the proposed hybrid urban block design was 4.7% lower than that of the current linear slab building typology, which can be attributed to the combined effects of the reduction of solar gain (13.2%)
Guevara-Sanchez *et al.* (2016)	Spain	Linear type medium-rise residential blocks	ASHRAE RP-884 database	- To develop a methodological framework to identify acceptable temperature thresholds	- Spanish climate data was used for Avila, Seville and Madrid - Vulnerable household population was selected by using census data - EnergyPlus software suite was used - Spanish Technical Code criteria were used for the BPE studies - Adaptive temperature thresholds for selected climates were investigated	-This study can set the basis for fuel poverty definitions for all Mediterranean and Southern European countries developing their own methodologies
Calama-Gonzalez *et al.* (2021)	Spain	H-typology of social housing estates	H-typology represents around 13,890 social housing dwellings, in contrast to approximately 10,715 linear-typology buildings	- To assess the current performance of representative building typologies in social housing stock	- Open-access energy simulation tool was used - Statistical analysis was conducted - The Andalusian Agency for Housing and Retrofitting dataset was used - *On-site* environmental monitoring was conducted - Sensitivity analysis was conducted - Statistical analysis was undertaken by using R v.3.5.3 software suite	An average annual discomfort hours of around 68% with higher percentage of annual undercooling discomfort hours

Supplementary Material 7: Worldwide Statistically Representative Archetypes

Table S.7.7: The Literature on Building Stock Aggregation through Archetype Buildings – Energy Modelling. (Continued)

References	A. Study Location	B. Building Type	C. Sampling Size	D. Primary Aim of Model	E. Methodology	F. Main Findings
Pasichnyi *et al.* (2019)	Stockholm, Sweden	Residential buildings	5,532 buildings from seven retrofitting packages were selected	- To present a methodological framework by using rich datasets to develop different building archetype for bottom-up energy modelling approach	- Urban building energy modelling approach was adopted by using archetyping approaches - The case study method was applied - Energy performance certificates data were used - Climate data and reference data on standardized use and building envelope were used - Mean Absolute Scaled Error was applied for the statistical analysis	Archetype subsets would be not only sufficiently large, but also sufficiently homogenous to ensure feasibility of the proposed retrofitting measures
Wang and Holmberg (2014)	Sweden, Switzerland	Residential buildings	1,400 existing Swedish building stock statistics from derived field studies were employed to validate the base-line scenario	-To present a methodological framework combines energy demand modelling and retrofit option rankings with life-cycle cost analysis	- The archetypes and the representative housing stock were developed - The ranking of implementing different retrofit alternatives to the energy saving potential is identified - Sensitivity analysis was conducted - Cost-effectiveness model was conducted - Life-cycle-cost-assessment method was applied	- Energy saving potential of retrofitting is 36-54% in the archetypes - If larger contingents of similar archetypes magnitude the retrofitting on a municipal level, it would contribute to energy-policy design
Nageli *et al.* (2018)	Sweden, Switzerland	Residential buildings	The building stock size is limited to a representative sample stock of 10,000 synthetic buildings	-To present a new method of building stock modelling based on the generation of synthetic building stocks	- Data available from national statistics was used - Building typology analysis was conducted through Monte Carlo sampling from a distribution - Building characteristics, age, structural type were used for the statistical analysis	-The variation within common classifications of building type and construction period can be much larger than the average differences between construction periods or building types

Supplementary Material 7: Worldwide Statistically Representative Archetypes

Table S.7.8: The Literature on Building Stock Aggregation through Archetype Buildings – Energy Modelling. (Continued)

References	A. Study Location	B. Building Type	C. Sampling Size	D. Primary Aim of Model	E. Methodology	F. Main Findings
Dodoo and Gustavsson (2016)	Sweden, Switzerland	Residential buildings	**T.1** – a prefab concrete-frame **T.2** – a massive timber frame **T.3** – a light timber frame	-To determine the impacts of climate change on the performance of different building configurations	-Dynamic hour-by-hour energy balance modelling and systems analysis were conducted -Climate datasets were used -Hourly downscaling of future climate dataset was applied -Overheating risk assessment was conducted	- Overheating risk was found to be slightly higher for the massive-frame building and slightly lower for the light-frame building
Froemelt *et al.* (2018)	Zurich, Switzerland	Residential buildings	Detailed information on the characteristics and consumption behaviour of 9,734 households were identified	-To provide an appropriate basis in support of effective environmental policymaking decisions	- Two-tiered clustering method was applied - Wald-clustering method was applied for the statistical analysis - Swiss Household Budget Survey was used - Swiss energy consumption data was used - Pattern recognition and clustering of households were explored - Individual archetypes and their interrelations were explored	- Delivering tailor-made insights into households' consumption behaviour for policymakers to derive and prioritize targeted measures
Heeren *et al.* (2012)	Zurich, Switzerland	Residential buildings	The building stock is clustered into 13 construction periods according to their year of construction	- To demonstrate an innovative assessment methodology in the form of a life-cycle-based building stock model - To investigate the feasibility of a sustainable energy vision	- Bottom-up energy framework was adopted - Archetype analysis was applied for retrofitting of building envelope - LCCA analysis was applied - Swiss building stock was analyzed - Algorithms were developed for the statistical analysis	- Developed model will be enhanced by means of GIS-integration of Zurich, the city will be clustered in supply and demand regions in order to increase the model's resolution
Lopez-Moreno *et al.* (2021)	Madrid, Spain	Residential buildings	Eight residential urban classes selected for an archetype analysis	-To present a systematic methodology to identify and classify residential urban areas according to representative homogenous urban zones	-Data detailing urban classifications and indicators were collected from previous studies -Cluster analysis was conducted to define and validate archetypes -Statistical Institute of the Community of Madrid dataset was used	The correlation is higher for the open mid-rise, with values greater than 55% for both low- and mid-rise urban blocks

Supplementary Material 7: Worldwide Statistically Representative Archetypes

Table S.7.9: The Literature on Building Stock Aggregation through Archetype Buildings – Energy Modelling. (Continued)

References	A. Study Location	B. Building Type	C. Sampling Size	D. Primary Aim of Model	E. Methodology	F. Main Findings
Brogger and Wittchen (2017)	Denmark	Residential buildings	Using data from the Danish Energy Performance Certificate (EPC) scheme – 27 archetypes were identified as nationally representative sample size	-To identify characteristics of building stock models that are essential for determining the energy-saving potential in national building stocks accurately	-Example building approach was adopted -Statistical approach was adopted to determine bottom-up energy policy design -Heat and electricity supply models were investigated to provide a valid background for the BES analysis -Building energy simulation models were built at a disaggregated level	It was found that the most cost-effective to reduce the energy demand in buildings by only 12%-17%, even though an energy-saving potential of around 40% was identified
Perera *et al.* (2018)	Nablus, Palestine	Residential buildings	The study is limited the scope to a single urban archetype focusing more on the urban density	-To quantify the impact of urban climate on energy system design and assessing the consequences of neglecting this specific aspect on energy system performance	-Urban microclimatic conditions, urban design, energy demand and optimization of energy system modelling were conducted -BES analysis method was applied -CitySim – open-source software suite was used - Pareto optimization of selected urban blocks parameters were conducted	-Energy use increase by 5-8% while grid dependency increases by up to 57% -More fluctuations in demand profile are observed when moving from standalone buildings to dense urban areas
Ben and Steemers (2018)	Cambridge, United Kingdom	Residential buildings	A total 78 households participated in the surveys, including 55 usable cases (response rate 28%) from face-to-face surveys and 23 usable cases (response rate 12%) from postal surveys	-To identify household archetypes and behavioral patterns in order to allow a targeted approach in energy-saving policy and retrofit improvement	- A statistical approach to cluster households based on empirical data collected from a household survey - Non-parametric correlation analysis was carried out in order to determine the relationship between behavioral factors and the households or dwellings characteristics - The questionnaire was paired with data on building characteristics obtained from the Domestic Energy Performance Certificates dataset - Factor analysis method was applied by using SPSS software suite	-The research has identified five different household archetypes to serve as a basis for targeted policy interventions tailored to specific socio-demographic groups regarding domestic energy demand reduction

Supplementary Material 7: Worldwide Statistically Representative Archetypes

Table S.7.10: The Literature on Building Stock Aggregation through Archetype Buildings – Energy Modelling. (Continued)

References	A. Study Location	B. Building Type	C. Sampling Size	D. Primary Aim of Model	E. Methodology	F. Main Findings
Taylor *et al.* (2014)	London, United Kingdom	Residential buildings	3,456 dwelling variants identified for an archetype analysis, identified through six different Design Summer Year (DSY) climate files from locations within the UK	- To examine the overheating risk in London dwelling archetypes when simulated under different UK climates, both in the present and under 'hot future' conditions	- Dynamic thermal simulations were undertaken - EnergyPlus – open-source energy simulation tool was used - 15 dwelling archetypes (27 variants including ground-, mid- and top-floor flats) - Risk of overheating risk in UK cities was predicted - Regression analysis was conducted - The Kendall's computation criterion was applied to determine statistically representative findings	- The weather files can influence the ranking of relative overheating risk between dwelling types, with significant variations in the relative ranking criterion - The calculated overheating metrics of dwelling archetypes could be found within a central range 2 -3 °C
Oikonomou *et al.* (2012)	London, United Kingdom	Residential buildings	Amongst 92 different built form and dwelling age combinations identified, the 15 most common were selected for simulation	-To assess variations in indoor temperatures in London dwellings during periods of hot weather, and the degree to which those dwelling-to-dwelling variations influences the BEM	- EnergyPlus – open-source energy simulation tool was used - Geometry and structure of analytical energy model were constructed - English House Condition Survey data was used - Occupancy schedules were assigned into the BEM - Weather files were assigned into the BEM - Housing typology analysis was conducted	- The effects of built form and other dwelling characteristics appear to be more important determinants of variation in high indoor temperatures than the location of a dwelling within London's urban heat island effect
Nishimwe and Reiter (2021)	Wallonia Region, Belgium	Building stock in general	- Using cadastral data of more than 1,700,000 Walloon buildings - 9,876 statistical sectors were identified as an archetype for the statistical analysis	- To assess the Energy Performance Certificates of the whole building stock and test to what extent different types of variables (building factors and socio-demographics) explain annual domestic energy use	- Cadastral map data was used - GIS software suite was used to develop building geometry into the model - Data mining process was conducted - Ordinal logistic regression analysis method was used to test the parameters for an archetype analysis - Lasso regression method was applied - Descriptive statistics of predictors were conducted	- Considering residential buildings, the building usages explained 66.46% of EPCs, whereas tertiary building usages explained 50.53%.

Supplementary Material 7: Worldwide Statistically Representative Archetypes

Table S.7.11: Pilot studies that evaluated on the literature on infrared thermography for the energy audit of buildings.

References	A. Study Location	B. Building Type	C. Primary Aim of Model	C. Methodology	D. Main Findings
Tejedor *et al.* (2021)	Italy, all the regions including the Mediterranean island of Sicily	Buildings (i.e., residential, offices)	- To propose a critical review on the employment of the quantitative IRT survey for the assessment of the *U*-value of the building envelope. - To demonstrate a novel methodological framework for the IRT technique and its impact on building performance evaluation studies. - To highlight the necessity for specialized thermographs who deal with an evolving methodology.	- A systematic literature review was conducted. - Laboratory test, *in-situ* measurements and infrared radiometer thermography approaches were selected main keyword for the bibliographic analysis. - Common approaches to the U-value assessment were discussed as follows; *(i)* analogies with coeval buildings; *(ii)* calculation method; *(iii)* heat flow meter measurements; *(iv)* laboratory testing; *(v)* IRT survey.	- The *U*-value can be calculated by using IRT. - Further research is required between simulation and experimental data in order to provide reliable results for the development of new techniques based on IRT. - Experimental field-testing studies demonstrate more reliable findings than steady-state analysis of *U*-values of building envelopes. - Sensitivity analysis is required to validate discrepancies between the effects of radiations and boundary conditions.
Cardani *et al.* (2021)	Worldwide	Buildings (i.e., residential, offices)	- To provide an analytical framework for energy auditors and thermographers. - To present a critical review of the use of the IRT survey in the building energy audit.	- Bibliographic analysis was conducted. - Current energy audit approaches in energy audit were conducted. - Both passive and active thermography measures and its implications on building performance evaluation were conducted.	- Passive approach was found to be the most common driver to detect thermally significant defects. - Significance of integration of different non-destructive testing of building envelopes could contribute to the IRT survey development framework. - Further research is required to represent archetype housing stock analysis for the development of benchmarking criterion in residential sector.
Habaibeh *et al.* (2021)	San Siro – Milan, Italy	Social housing stock	To identify an innovative and up-to-date methodological didactic approach for defining the most appropriate solutions for the refurbishment of a social housing stock.	Exploratory case study approach was conducted. Design driven approach was adopted with employing the historical research and survey, on-site visit, hands-on-training and on-site exposition.	To improve the awareness of the students on the possibilities of building renovation and design applications for social housing stock.

Supplementary Material 7: Worldwide Statistically Representative Archetypes

Table S.7.12: Pilot studies that evaluated on the literature on infrared thermography for the energy audit of buildings. (Continued)

References	A. Study Location	B. Building Type	C. Primary Aim of Model	C. Methodology	D. Main Findings
Sbrogio *et al.* (2021)	Worldwide	Buildings (i.e., residential, offices)	To present a methodological framework for evaluating energy and environmental performance of building stock by the use of non-invasive techniques.	A review of instrumental analysis was conducted as follows *(i)* visual testing; *(ii)* thermographic inspection; *(iii)* thermal comfort; *(iv)* post-occupancy evaluation.	Only the use of coring showed the presence of moisture and water percolation on thermal insulation. Sonic trial proved the presence of some mechanical anomalies revealed by different velocities of the sound propagation in the masonry.
Gupta and Gregg (2021)	Victoria, Canada	*On-site* experimental structure	To develop an external IRT method to determine clear wall *U*-values. To determine the viability of an external thermographic survey technique for use in energy audits.	The IRT measures were conducted on a conditioned at-scale insulated wood-frame wall structure. FLIR A65 IR camera was used. A 3D thermal modelling the Nx software package was used to validate IRT survey findings.	*U*-value measurement with IRT in the best-case scenario deviated between 6.25%-25.00%. The *U*-value results with IRT were validated and ranged between 11.53% - %10.00 in the best-case scenario.
Abrahao *et al.* (2021)	Porto, Portugal	*On-site* experimental structure	To develop a thermographic 2D *U*-value map for the characterization of heavy walls in stationary regime. To assess the temperature distribution of each transition phase between each defect and its undisturbed surroundings.	Measurements were conducted in a walk-in climatic chamber – FITOCLIMA 1000. 2D *U*-value map is created. 2D colour map was developed to identify the distribution of the thermal transmittance of the walls. In-situ QIRT test was conducted.	Optimisation of a TWALL mesh comprised of 1600 elements of 8 x 6 pixels. Image quality losses were estimated at 6.65%. 2D correlation coefficient, R was equal to 0.287 which means that only 8.23% of the processed thermal image can be attributed to the original thermogram.
Bartesaghi-Koc *et al.* (2021)	Brescia, Italy	Residential Tower Block	To verify the applicability of the energy rating system which was newly drawn up by the Green Building Council in Italy. To design a methodological framework for low-energy design and retrofitting.	On-site building diagnostic method was used. EnergyPlus software suite was used to undertake dynamic thermal simulations. IRT survey was carried out. The PAN software used to process the IRT survey findings.	The rock wool insulation under the ventilated façade with a density of 70 kg/m^3 and a thickness of 12 cm which is highly breathable allows surface temperatures of over 18° C to be reached and guarantees good hygrometric behaviour.

Supplementary Material 7: Worldwide Statistically Representative Archetypes

Table S.7.13: Pilot studies that evaluated on the literature on infrared thermography for the energy audit of buildings. (Continued)

References	A. Study Location	B. Building Type	C. Primary Aim of Model	C. Methodology	D. Main Findings
De Angelis *et al.* (2020)	Nottingham, United Kingdom	19^{th} century detached house (renovated cottage house)	To develop a novel methodological framework where infrared thermography of a deep retrofitted building is combined with deep learning neural networks. To predict the future effectiveness and economic viability of wall insulation in terms of energy savings.	Exploratory case study approach was adopted. A mathematical model was developed to predict the accuracy of life-long monitoring of buildings. Infrared thermography and temperature sensors were used to assess building fabric thermal performance. FLIR E25 thermal camera was used as building diagnostic tool. The Matlab was used to validate temperature recordings.	High accuracy of predicting the actual energy savings with success rate of about 82% when compared with the calculated values. The Artificial Neural Networks (ANN) predicted heat losses are slightly higher than the calculated ones in 14 out 21 cases for each wall type. The range of error for the insulated wall is -13% to +15% and for the uninsulated wall is -14% to +17.5%.
Lei *et al.* (2021)	Mestre-Venice, Northern Italy	20^{th} century multi-family medium-rise apartment building	To assess the current condition and propose cost effective and energy-efficient retrofit design interventions. To develop a methodological workflow to provide a guidance on the development of retrofit interventions in order to improve structural resistance of existing housing stock.	The housing typology classification was conducted to identify nationally representative housing type for archetype analysis. IRT survey was conducted. 3D analytical model was developed to perform structural seismic analysis. The 3Muri software suite was used to perform overall seismic behaviour of building.	In respect to out-of-plane mechanisms, the weakest panel was n.1 in wall 15 in the north wing and n. 1 in the west wing. Building fabric thermal performance analysis confirmed that low temperatures on internal surfaces (10-14° C) close to the dew point temperatures, especially in the junctions between floor slabs and walls and between the roof and the walls.
Sakiyama *et al.* (2021)	The city of York – England, United Kingdom	Low-energy dwellings which were built according to Passivhaus standard	To present the methodology and results of in-situ testing of building fabric thermal performance to calibrate as-built energy models.	Integrated Environmental Solutions (IES) software suite was used. The in-situ tests included repeat testing of air permeability integrated with thermal imaging survey and heat flux measurements of the building fabric elements.	Validation of the model by altering the wall U-value to 0.26 W/ m^2 K made sense as this brought the external wall U-values closer to the BRUKL limiting parameter of 0.30 W/ m^2 K.

Supplementary Material 8: Step-by-Step Development of Statistically Representative Archetypes in South-eastern Europe - Cyprus

Step 1: Benchmarking indicators for the development of representative archetypes

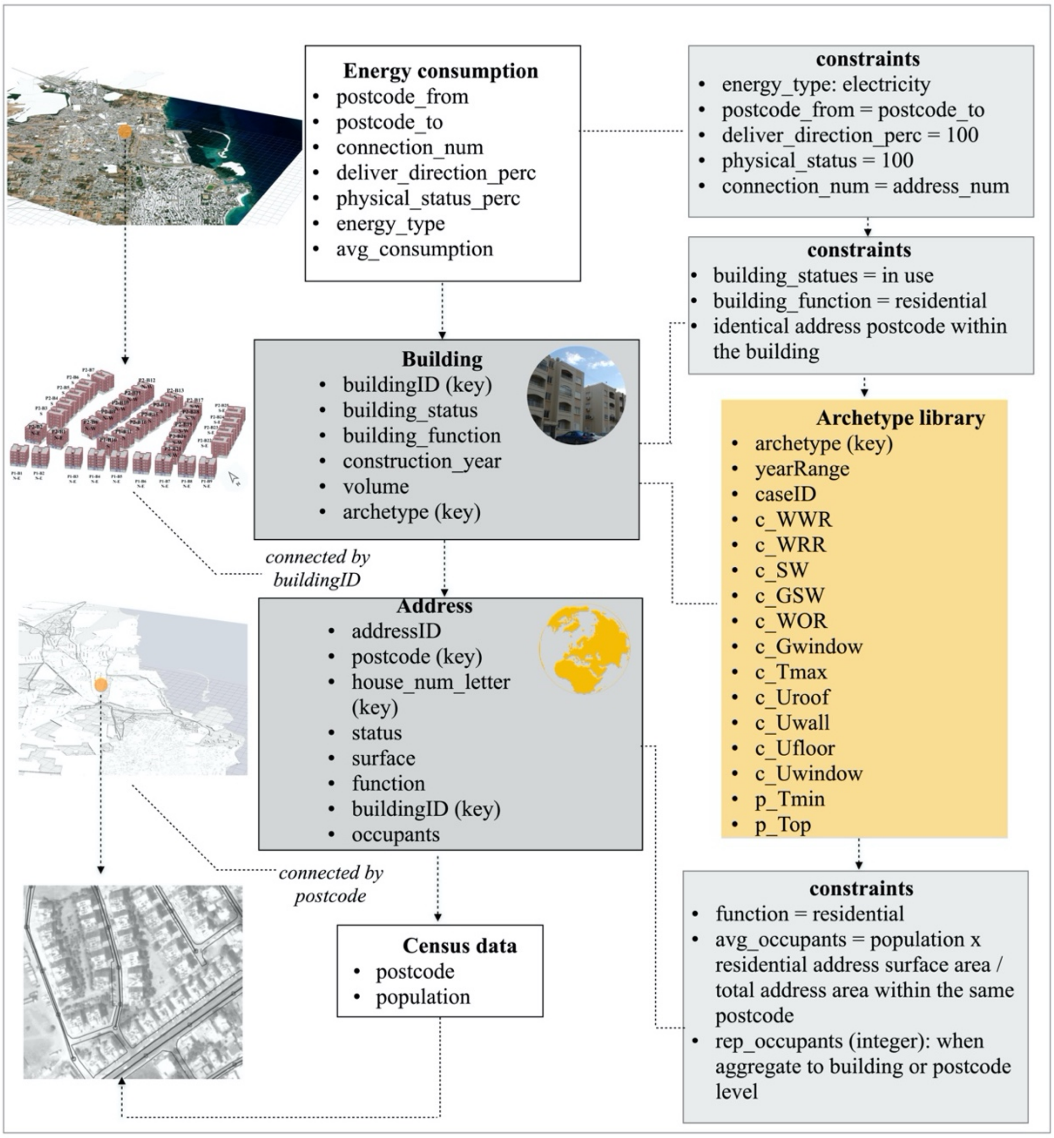

Figure S.8.1: Data model and identification criteria for the representativeness of archetypes in this study.

Figure S.8.2: 3D model of the coastal city of Famagusta.

Supplementary Material 8: Step-by-Step Development of Statistically Representative Archetypes in South-eastern Europe - Cyprus

Step 2: Census data

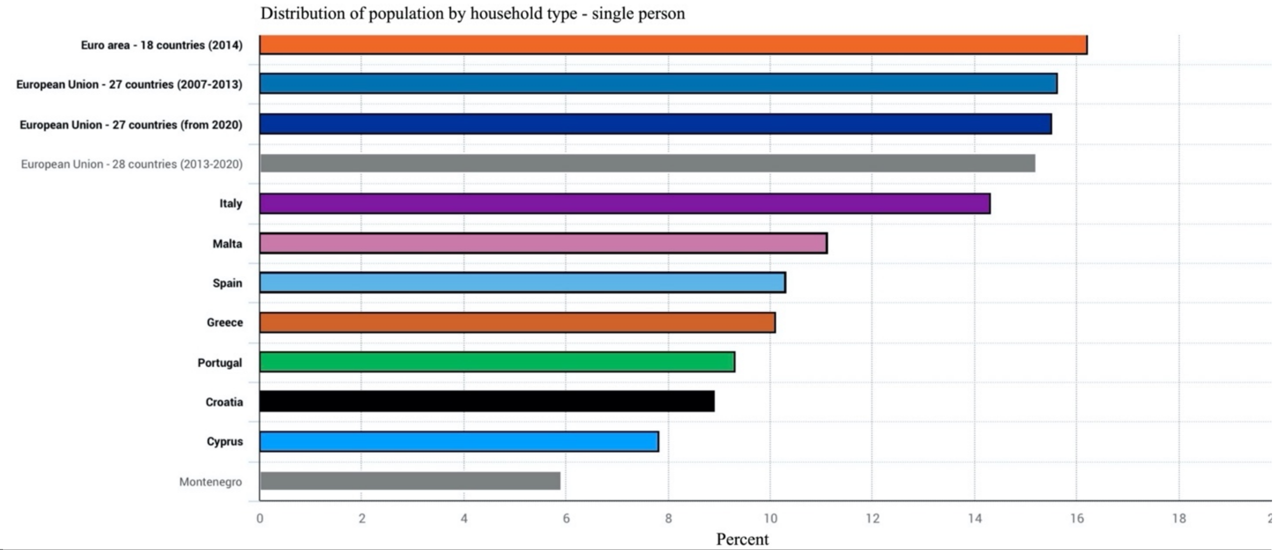

Figure S.8.3: Distribution of population by household type – single occupant. ***Source:*** Eurostat, 2018

Information: Figure S.8.3 shows the population of single-person occupancy in both the EU-27 and South-eastern Mediterranean countries. This figure indicates that 7% of the population lives as a sole occupant in Cyprus.

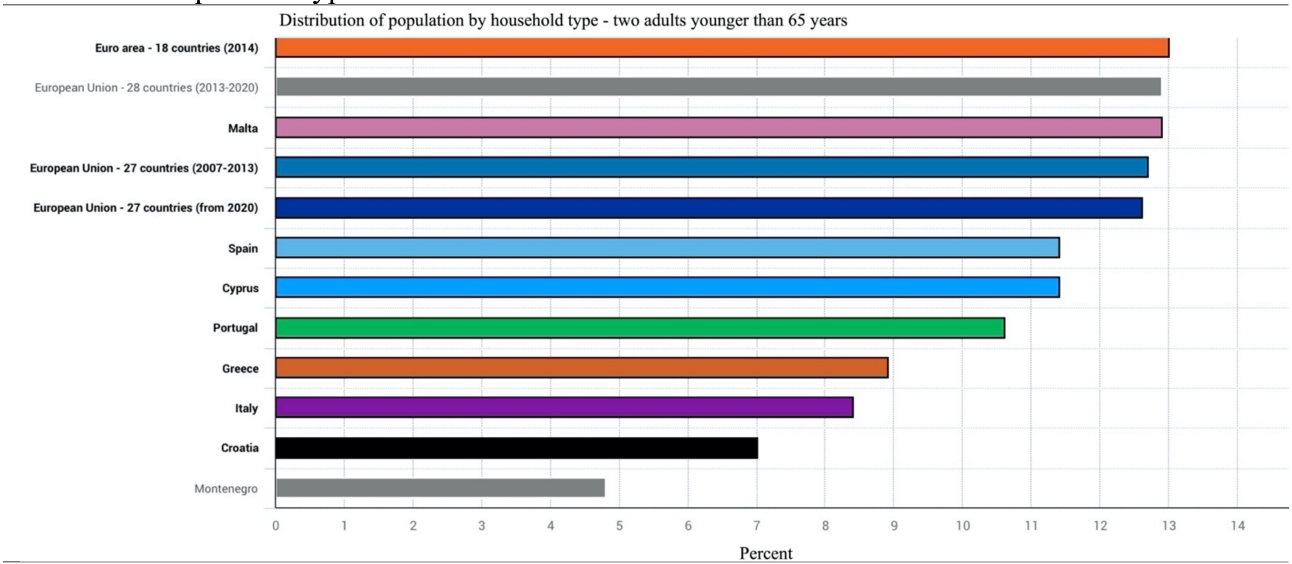

Figure S.8.4: Distribution of population by household type – two adults younger than 65 years. ***Source:*** Eurostat, 2018

Information: Figure S.8.4 shows the population distribution of the household type of two adults younger than 65. In this analysis, the data predominantly examines the census data in the South-eastern Mediterranean region that has been applied to identify aggregate housing stock analysis for the energy modelling of this study. The data shows that 11.5% of Cypriot households are comprised of two adults younger than 65 years. Household type is the determinant variable for exploring the correlations between occupancy type and energy consumption in the energy modelling phase of this study. As shown here, 11.5% for this type of occupancy is very close to the EU-27 average of 12.5%.

Supplementary Material 8: Step-by-Step Development of Statistically Representative Archetypes in South-eastern Europe - Cyprus

Step 2: Census data (Continued)

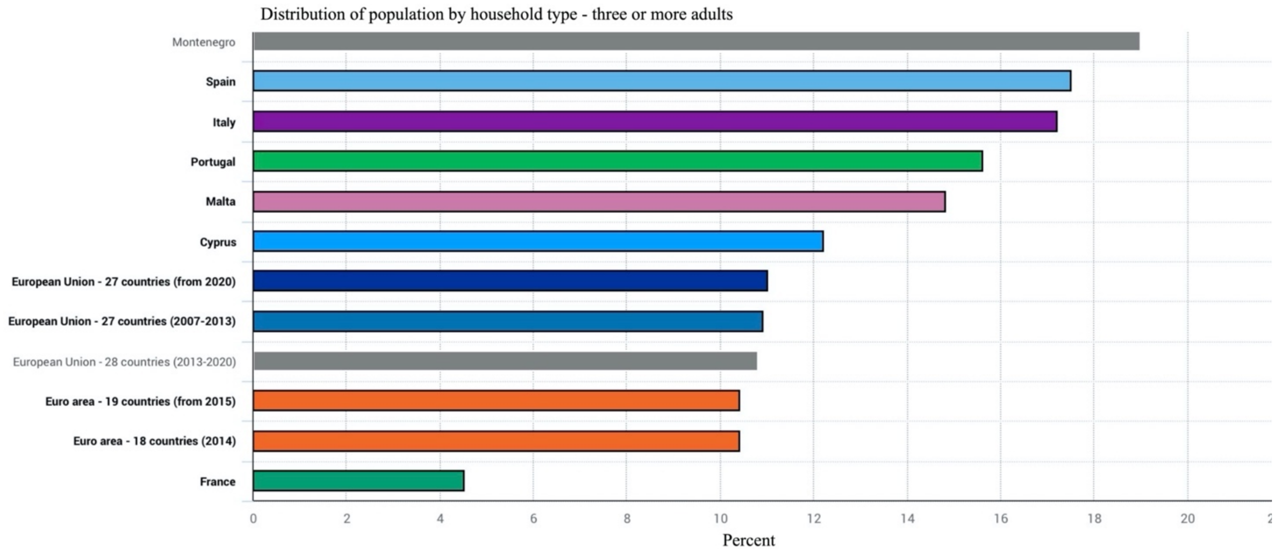

Figure S.8.5: Distribution of population by household type – three or more adults. ***Source:*** Eurostat, 2018

Information: Figure S.8.5 shows the distribution of household occupancy of three or more adults, which represents 12.1% of Cypriot households. This data demonstrates the moderate occupancy type that was applied to determine representative occupancy profiles for the energy modelling phase of the study. It should be noted that identification of representative occupancy profiles though use of census data could lead to obtaining more reliable results to validate the data from the building energy modelling phase of the study.

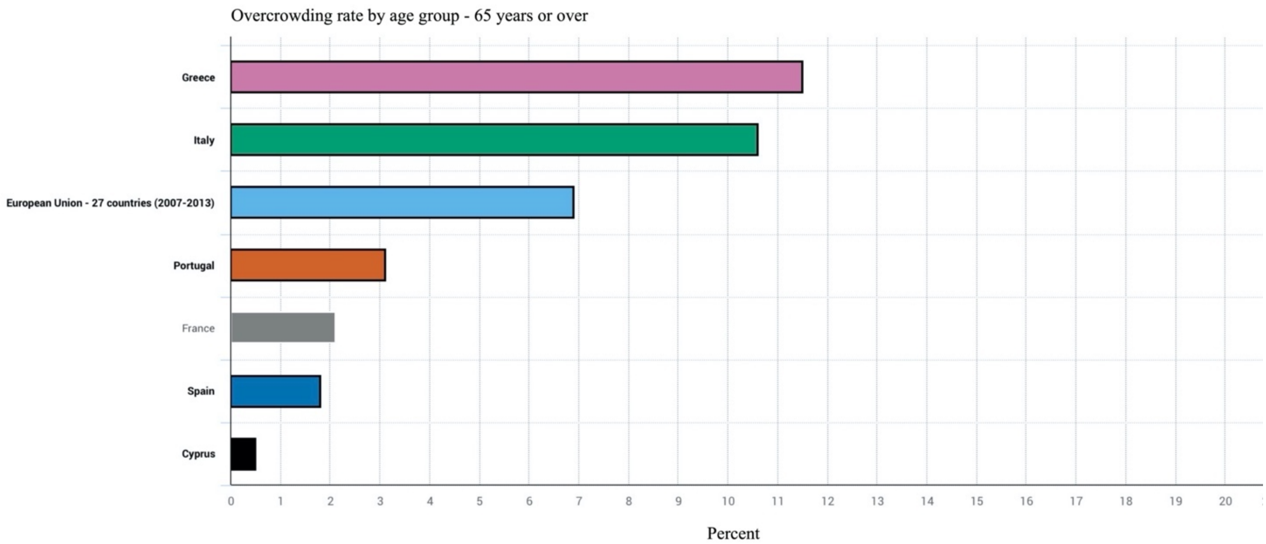

Figure S.8.6: Overcrowding rate by age group – 65 years or over. ***Source:*** Eurostat, 2018

Information: Figure S.8.6 shows the distribution of age group in both the EU-27 and in South-eastern Mediterranean countries. As can be seen from the graph, the ageing population accounts for less than 1% in Cyprus while approximately 7% of EU-27 population is 65 years or over. This data indicates that the younger population is the dominant population in Cyprus and this is the reason that there is a demand for affordable housing schemes on this South-eastern Mediterranean island.

Supplementary Material 8: Step-by-Step Development of Statistically Representative Archetypes in South-eastern Europe - Cyprus

Step 2: Census data (Continued)

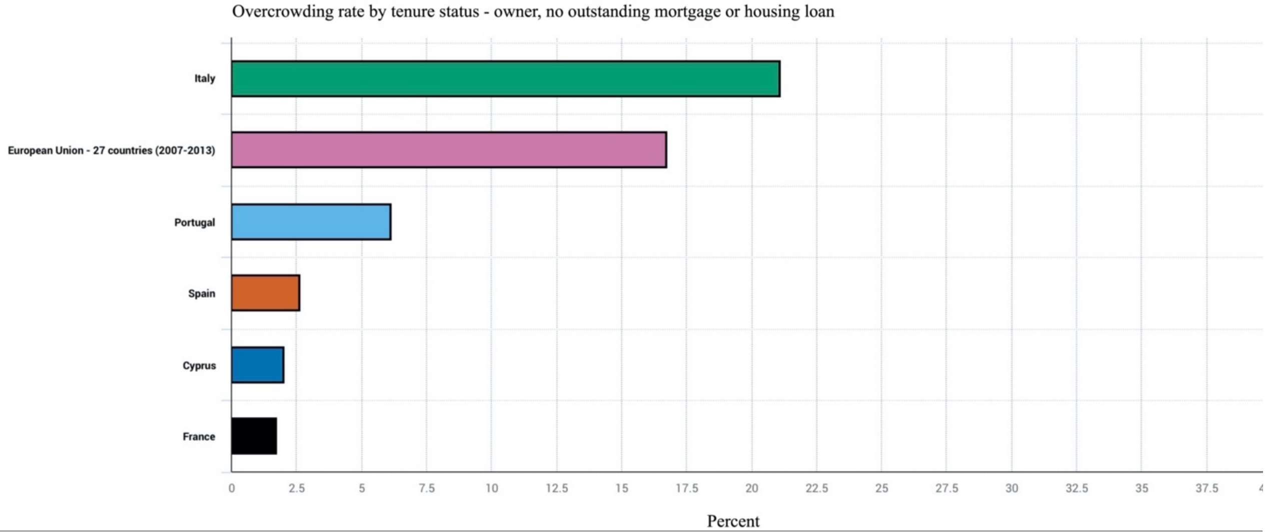

Figure S.8.7: Overcrowding rate by tenure status – owners with no outstanding mortgage or housing loan. ***Source:*** Eurostat, 2018

Information: Figure S.8.7 shows the distribution of tenure type. This graph shows that only 2% of Cypriot households are owner occupiers with no outstanding mortgage or housing loan, while approximately 17.5% of households are owner occupiers with no outstanding mortgage or housing loan in the EU-27 countries. The results show that fewer households were not having to pay any loans when they were first-time property buyers in Cyprus in comparison to other South-eastern Mediterranean countries.

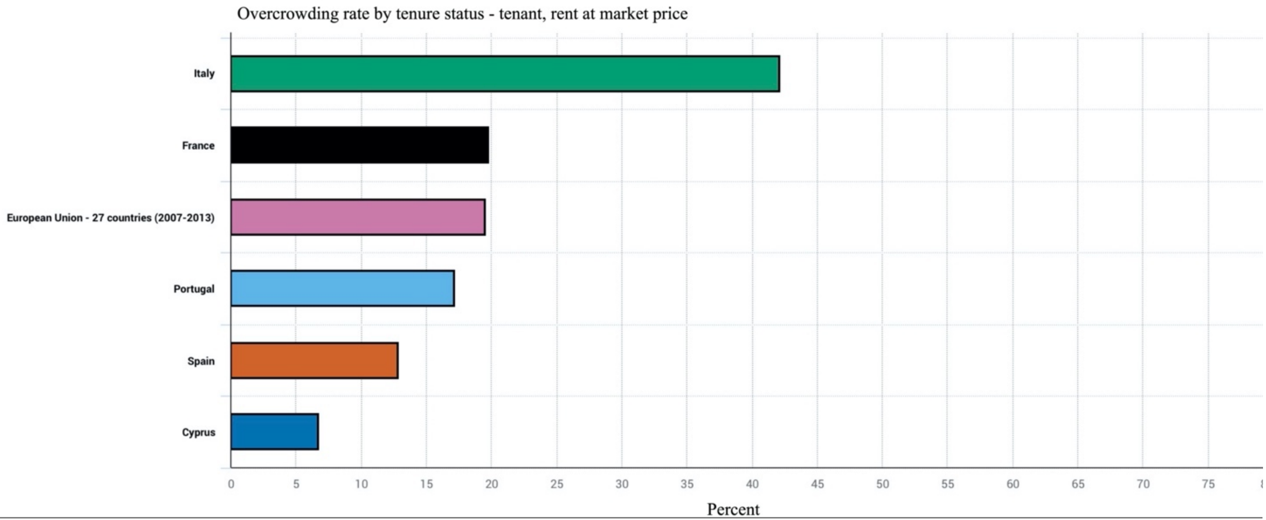

Figure S.8.8: Overcrowding rate by tenure status – tenants renting at market price. ***Source:*** Eurostat, 2018

Information: Figure S.8.8 shows the distribution of tenure type. It can be seen that 6% of Cypriot households were renters, while renters accounted for approximately 20% of households in the EU-27 region. The data demonstrates that private renters should be considered when evaluating the development of energy performance certificates (EPCs) in the South-eastern Mediterranean region.

Supplementary Material 8: Step-by-Step Development of Statistically Representative Archetypes in South-eastern Europe - Cyprus

Step 3: Household size

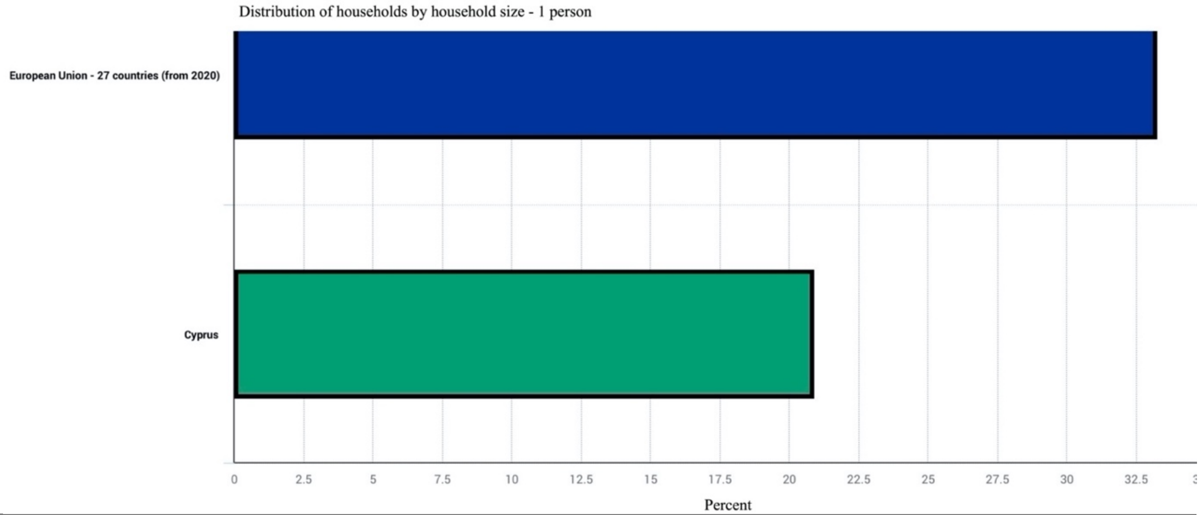

Figure S.8.9: Distribution of households by household size – solo occupancy. ***Source:*** Eurostat, 2018

Information: Figure S.8.9 shows the distribution of households for solo occupants. It shows that 21% of households were comprised of sole occupants in Cyprus, while more than 32.5% of households were comprised of sole occupants in the EU-27 countries.

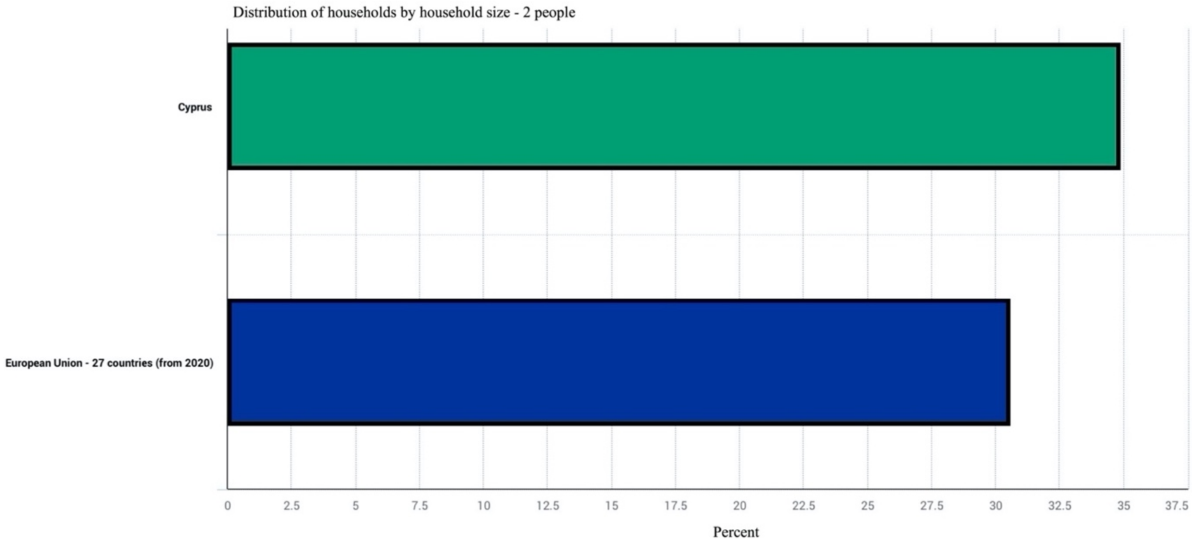

Figure S.8.10: Distribution of households by household size – two people. ***Source:*** Eurostat, 2018

Information: Figure S.8.10 shows the distribution of households that are comprised of two people. It was recorded that Cypriot households that consist of two people are mainly comprised of married couples or professionals. Therefore, two-person occupancy correlates with the younger population recorded in Cyprus. This type of household could have a direct effect on the identification of energy consumption profiles in the South-eastern Mediterranean region. In this study, the energy modelling approach adopted this indicator to validate energy simulation results with households' actual energy bills.

By contrast, more than 30% of EU-27 households are comprised of two people in residential buildings. This data demonstrates that the Cypriot household profile of two residents is relatively higher than the EU-27 average, which proves that two-person households should be considered in the building energy modelling phase of this study.

Supplementary Material 8: Step-by-Step Development of Statistically Representative Archetypes in South-eastern Europe - Cyprus

Step 3: Household size (Continued)

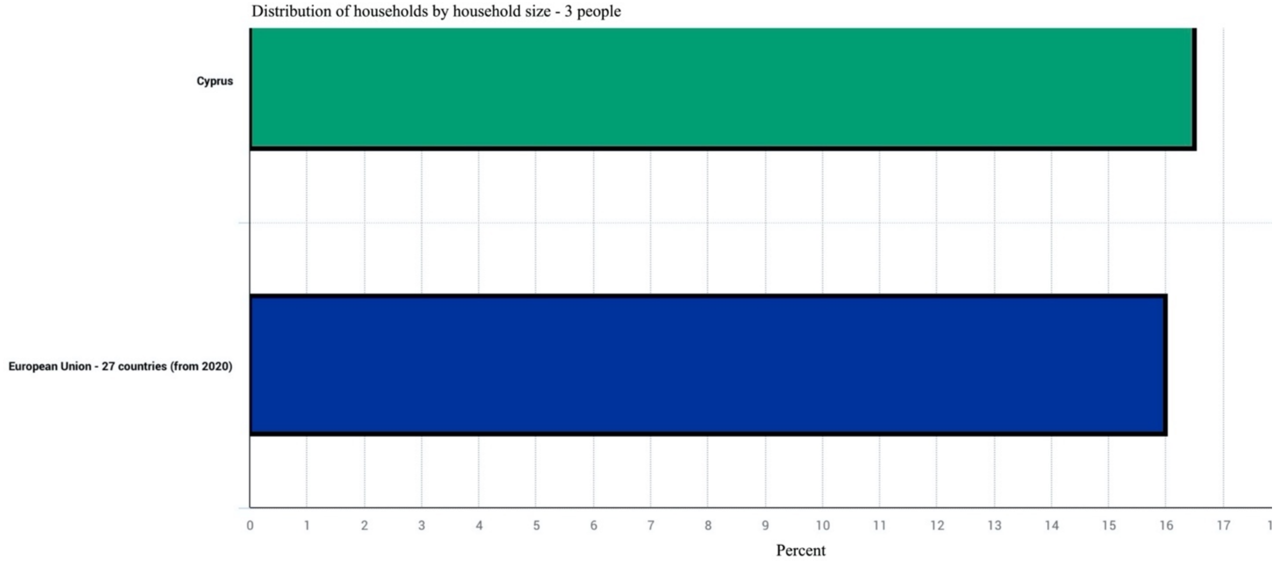

Figure S.8.11: Distribution of households by household size – three people. ***Source:*** Eurostat, 2018

Information: Figure S.8.11 shows the percentage of households with three people. In Cyprus, approximately 17% of households are comprised of three people in residential buildings, while the average for EU-27 countries is 16%.

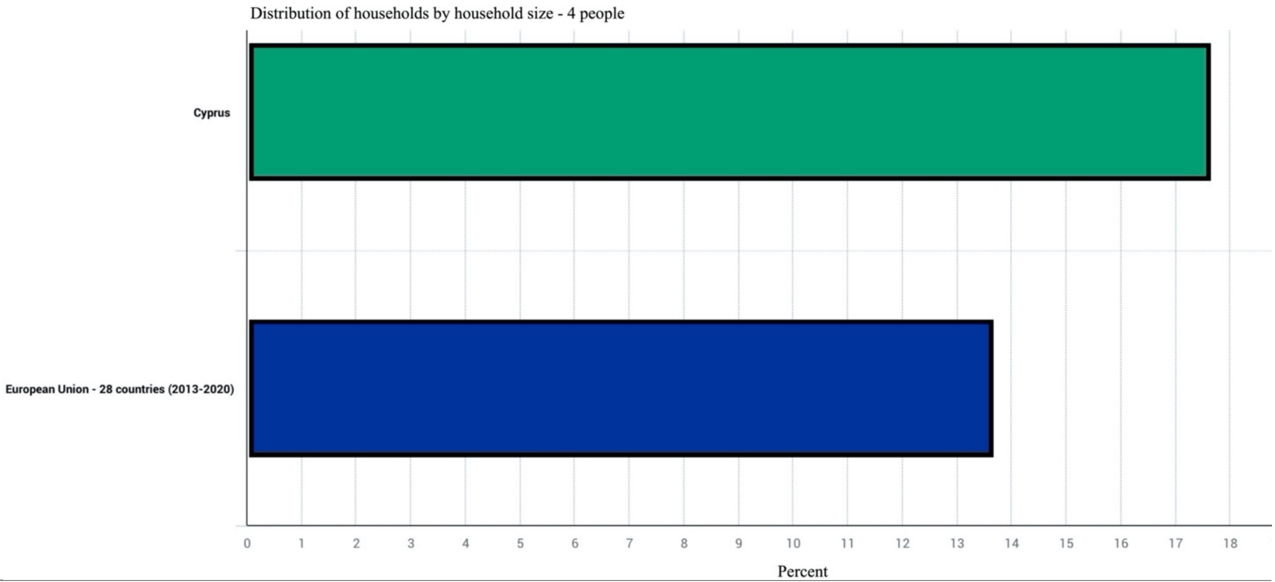

Figure S.8.12: Distribution of households by household size – four people. ***Source:*** Eurostat, 2018

Information: Figure S.8.12 shows the percentage of households with four people. This household type is approximately 18% for Cyprus while for the EU-27 countries it is less at nearly 14%.

Supplementary Material 8: Step-by-Step Development of Statistically Representative Archetypes in South-eastern Europe - Cyprus

Step 3: Household size (Continued)

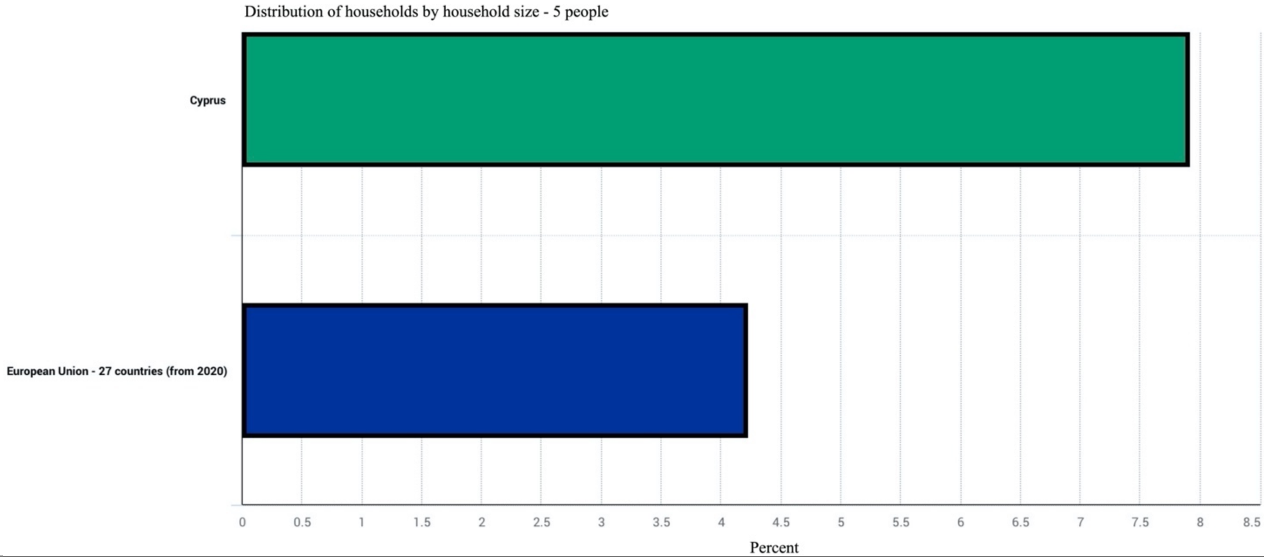

Figure S.8.13: Distribution of households by household size – five people. ***Source:*** Eurostat, 2018

Information: Figure S.8.13 shows the percentage of households comprised of five people. For Cyprus this is approximately 8%, while for the EU-27 countries this is 4.3%. This difference is due to the presence of different age groups in given households. This data correlates with household energy consumption by considering the high-occupancy pattern type in the building energy model to undertake dynamic thermal simulations.

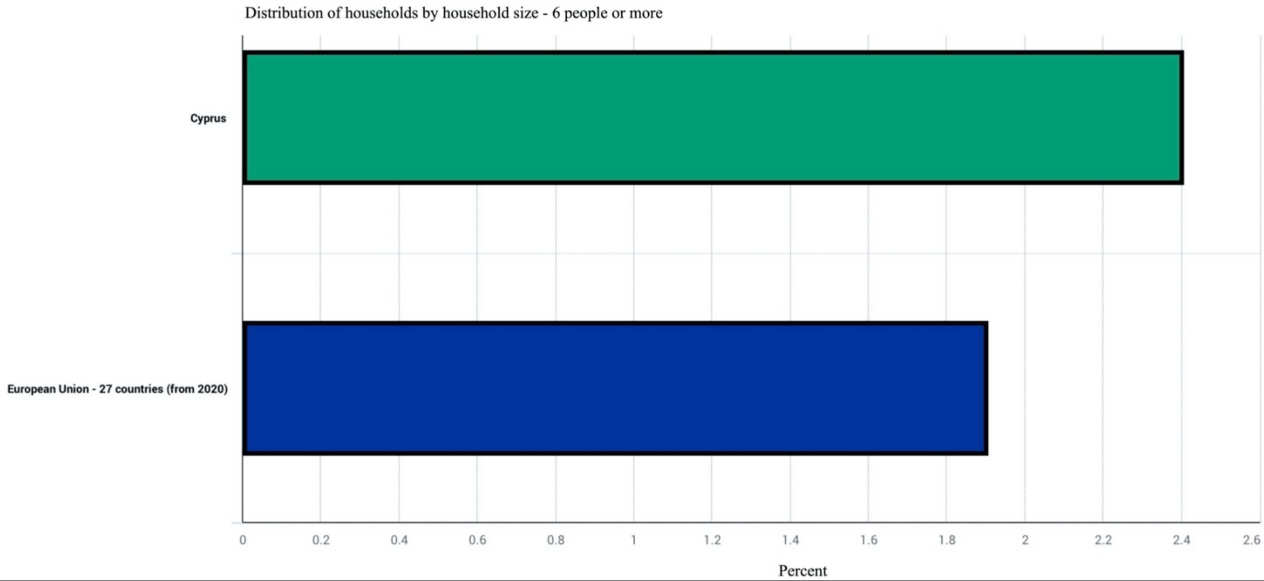

Figure S.8.14: Distribution of households by household size – six or more people. ***Source:*** Eurostat, 2018

Information: Figure S.8.14 shows the distribution of household size for six people. In Cyprus, 2.4% of households consist of six or more people, while 2% of this household type is recorded for the EU-27 region. The data demonstrates that the high-occupancy profile type is relatively higher for Cyprus than for the EU-27 average; this data should be considered when evaluating the reasons for high energy bills and to validate the data in conjunction with building energy simulation findings.

Supplementary Material 8: Step-by-Step Development of Statistically Representative Archetypes in South-eastern Europe - Cyprus

Step 4: Household data

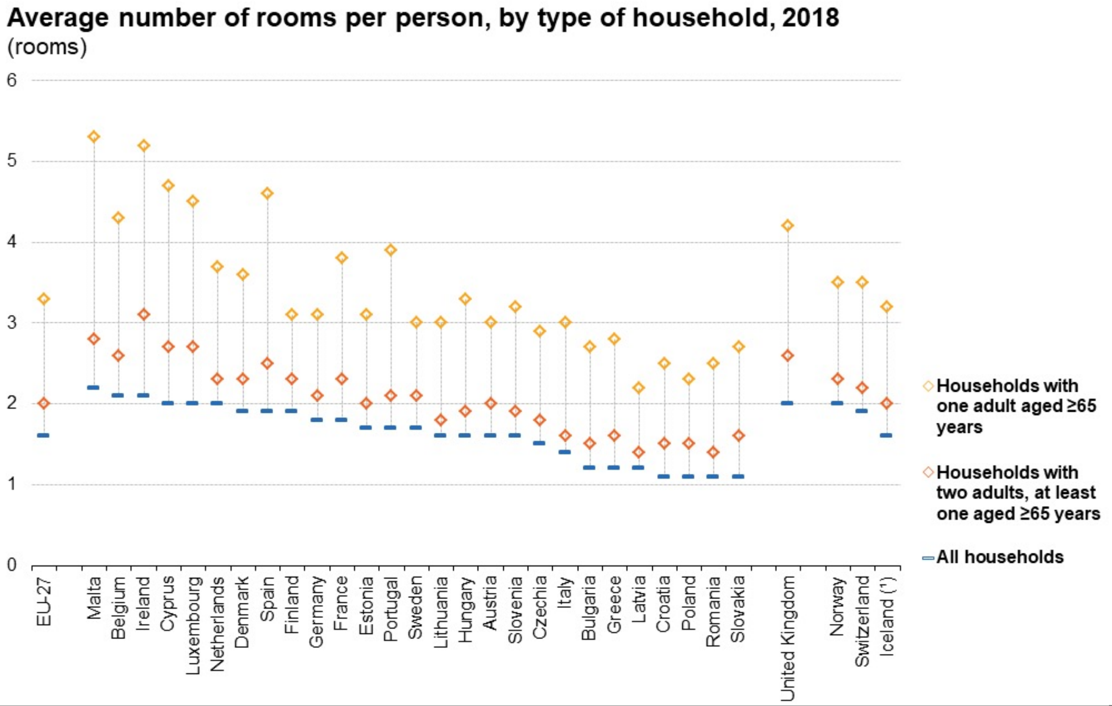

Figure S.8.15: Average number of bedrooms per person by type of household. ***Source:*** Eurostat, 2018

Information: Figure S.8.15 shows the average number of occupied rooms by household type. For a Cypriot household with one adult aged over 65 years, the household owned an average of a 5-bedroom property, while households with two adults with at least one adult aged over 65 owned an average of a 2.5-bedroom property. This data demonstrates that, in Cyprus, ageing households occupy dwellings with a high number of bedrooms; relatively higher than the EU-27 average.

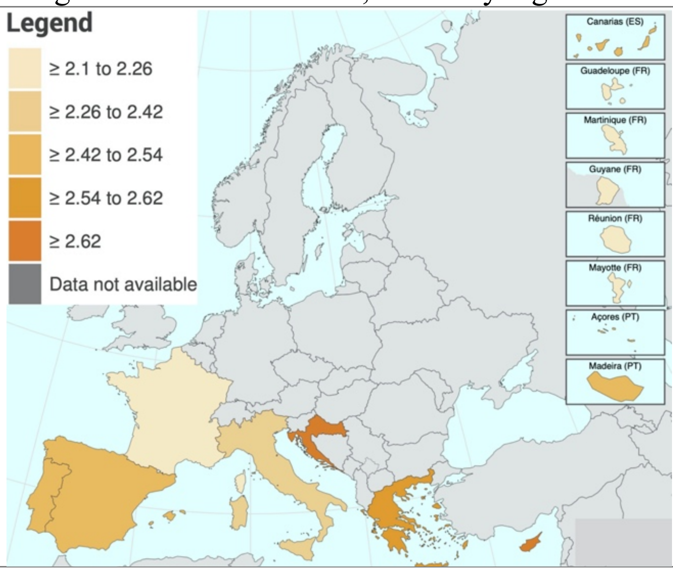

Figure S.8.16: Mapping of household type in the Mediterranean basin. ***Source:*** Eurostat, 2018

Information: Figure S.8.16 demonstrates occupancy density in residential buildings in the South-eastern Mediterranean basin. An average of over 2.62% people per household was recorded both in Cyprus and Croatia; this data should be considered when identifying representative occupancy profiles in building energy models. In this present study, the representative occupancy profiles were identified using both the Eurostat data and the data gathered from the questionnaire survey findings on post-war social housing estates in order to validate the data concurrently.

Supplementary Material 8: Step-by-Step Development of Statistically Representative Archetypes in South-eastern Europe - Cyprus

Step 4: Household data (Continued)

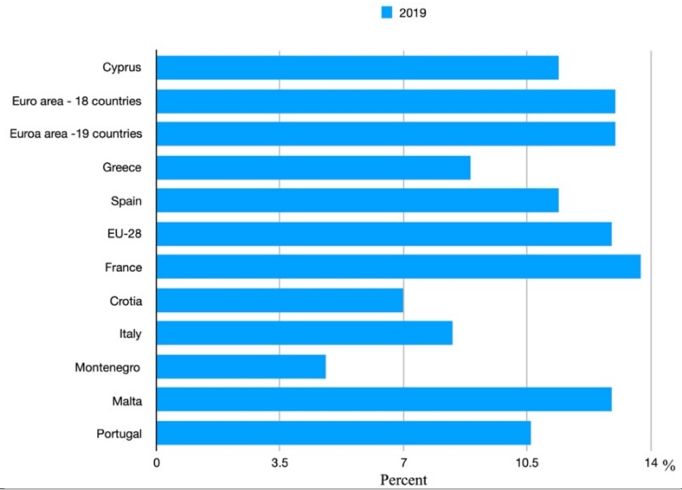

Figure S.8.17: Distribution of household type – two adults under 65. ***Source:*** Eurostat, 2018

Information: Figure S.8.17 shows the distribution of the household type of two adults younger than 65 years old. For Cyprus, this household type is 11%, while for the EU-28 the average was found to be 13%. The data demonstrates that this younger age group is the dominant household type for the Cypriot context; this variable should be taken into consideration while developing representative aggregate energy models.

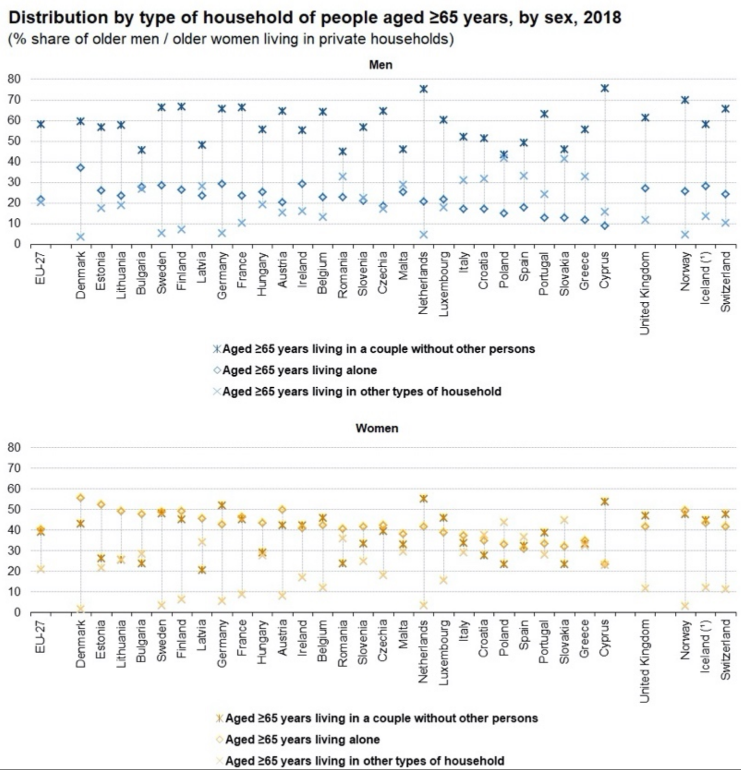

Figure S.8.18(a) and (b): Distribution of household type considering age and sex. ***Source:*** Eurostat, 2018

Information: Figure S.8.18 (a) and (b) demonstrate the distribution of household type by age (over 65 years) and sex. In Cyprus, 75% of this age group of men lives in a couple without an additional (third) person, while the EU-27 average for this type is 60%. Over 50% of this age group of women lives in a couple without an additional (third)person, while the EU-27 average for this type is 40%. This data shows that an elderly couple is a representative dominant occupancy profile and should be considered when developing aggregate energy models for building energy models.

Supplementary Material 8: Step-by-Step Development of Statistically Representative Archetypes in South-eastern Europe - Cyprus

Step 4: Household data (Continued)

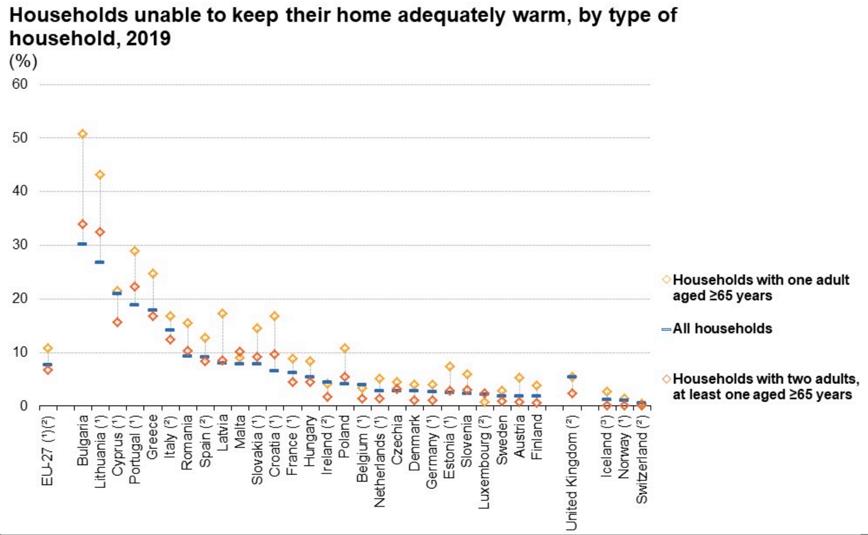

Figure S.8.19: Thermal conditions by household type. ***Source:*** Eurostat, 2019

Information: Figure S.8.19 shows the thermal condition of households unable to keep their homes adequately warm. It can be seen that 15% of households with two adults with at least one occupant aged over 65 years reported that they were unable to keep their home thermally comfortable in the winter in Cyprus, while the EU-27 average was recorded at 8%. While many EU countries found the thermal comfort of their dwellings to be below the EU-27 average, Cyprus was slightly higher than the EU-27 average. Considering each EU country, it was found that the 15% rate was higher than each EU country categorisation. The reason for this is that Cyprus joined the EU in 2004 while the first EPBD mandates were recommended in 2010, the implementation wasn't applied until 2016. Since then, the Republic of Cyprus (RoC) has not been able to achieve the common implementation measures to improve the energy efficiency of its housing stock.

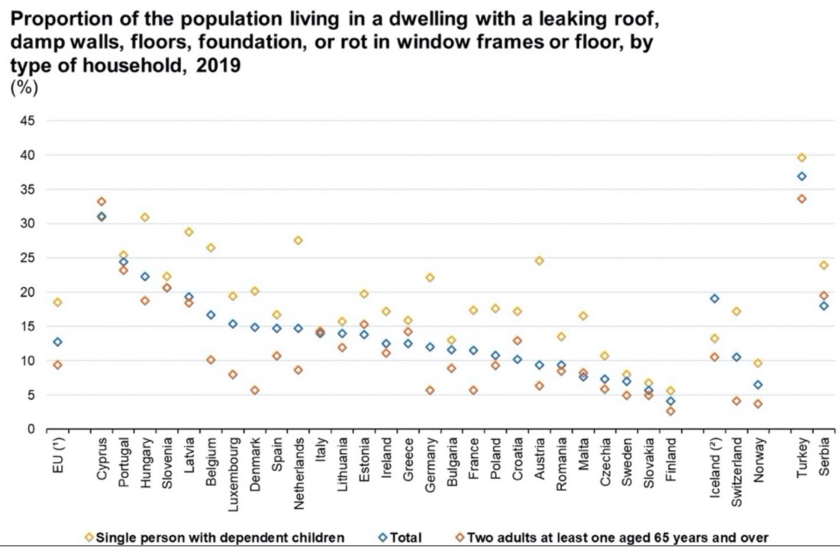

Figure S.8.20: Physical conditions of EU buildings. ***Source:*** Eurostat, 2019

Information: Figure S.8.20 shows the condition of housing stock in the EU. The data shows that, in Cyprus, approximately 35% of houses belonging to households with two adults at least one of whom is aged 65 years or over had structural issues, while the EU-27 average was recorded as just less than 10% for the same household type. It can be seen that worse housing conditions were recorded in Cyprus than in most other EU-27 countries. This is due to the late implementation of EPCs for buildings in the RoC and a lack of awareness around the energy efficiency measures needed to improve the physical quality of the existing housing stock.

Supplementary Material 8: Step-by-Step Development of Statistically Representative Archetypes in South-eastern Europe - Cyprus

Step 4: Household data (Continued)

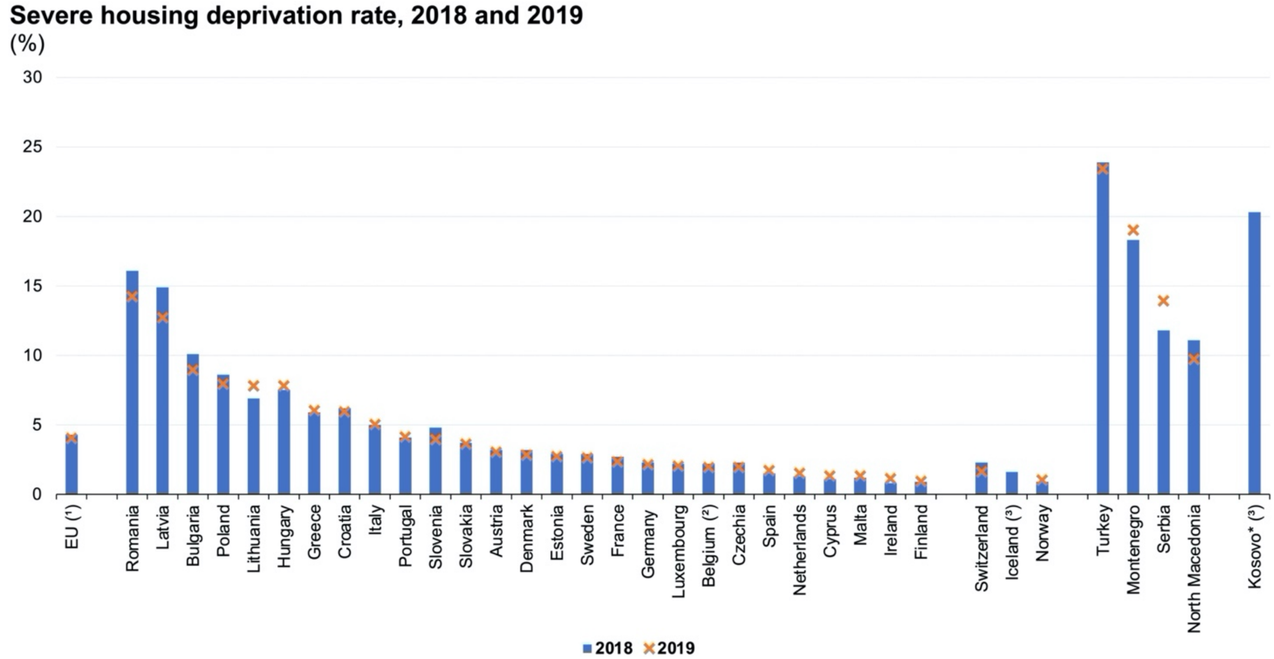

Figure S.8.21: Physical conditions of housing stock in general in the EU. ***Source:*** Eurostat, 2019

Information: Figure S.8.21 demonstrates the physical condition of housing stock in the EU between 2018 and 2019. The data shows that only 2% of Cypriot housing stock is of inadequate quality while the EU-27 average was less than 5%. This data highlights that the EPBD implementation schemes both in 2010 and 2016 and the mandates of issuing EPCs in 2016 had an impact on the overall quality of housing stock in the RoC.

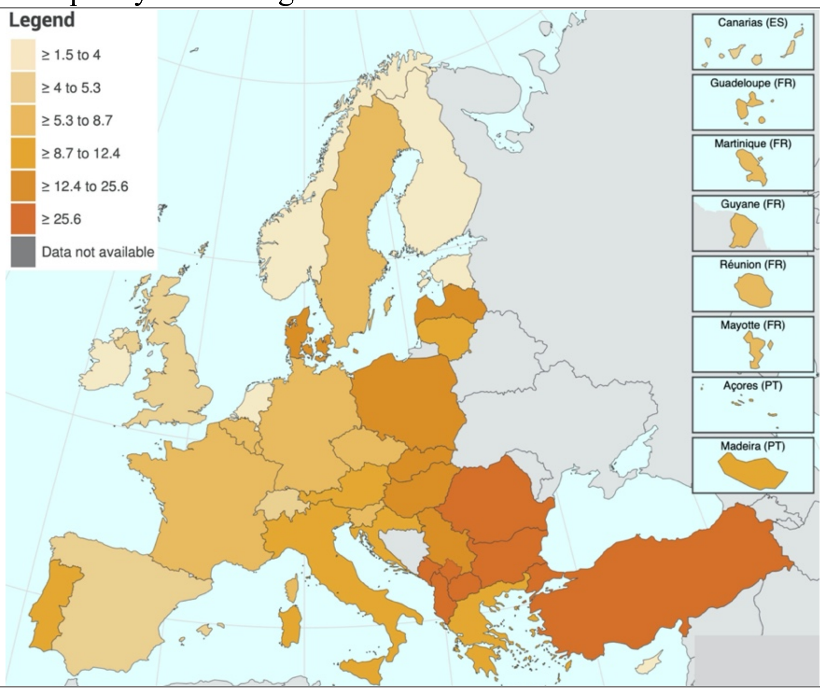

Figure S.8.22: Map of the physical conditions of EU housing stock in general. ***Source:*** Eurostat, 2019

Information: Figure S.8.22 maps the physical conditions of housing stock. It can be seen that Cypriot housing stock falls in the range of 1.5 to 4, and this in the lowest ranking category across Europe. This factor is due to the presence of mostly newly built housing stock across the island, while, in other EU countries, the housing stock dates back to the post–World War II era.

Supplementary Material 8: Step-by-Step Development of Statistically Representative Archetypes in South-eastern Europe - Cyprus

Step 5: Building materials

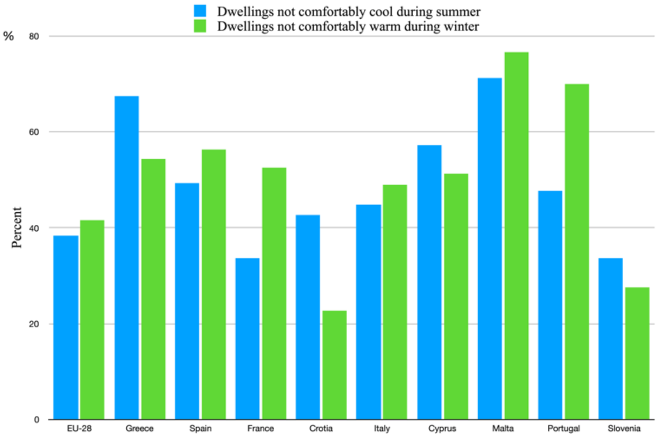

Figure S.8.23: Thermal conditions of Mediterranean housing stock. ***Source:*** Eurostat, 2018

Information: Figure S.8.23 shows the thermal condition of housing stock. It can be seen that in Cyprus only 1% of the population who lives in cities report that they have poor quality homes in comparison to other EU-27 regions. This is because in the dataset a relatively high population sample was extracted for the EU-27 countries but due to the small population of Cyprus, the extracted data has no shown effect size on the thermal quality of housing stock in Cyprus. This confirms that there is a research gap when comparing small cities with small population sizes like those in Cyprus with any other European country with a relatively large population size. Most importantly, in this comparison, the data illustrates that population size and housing stock figures are determinant factors that can make a reliable analysis for benchmarking. Despite the RoC having a relatively small population size in comparison to other EU countries, the data findings reveal that inadequate housing conditions have led to an overheating risk in the summer. This data was used to evaluate the aggregation of housing stock during the building performance evaluation phase of this study.

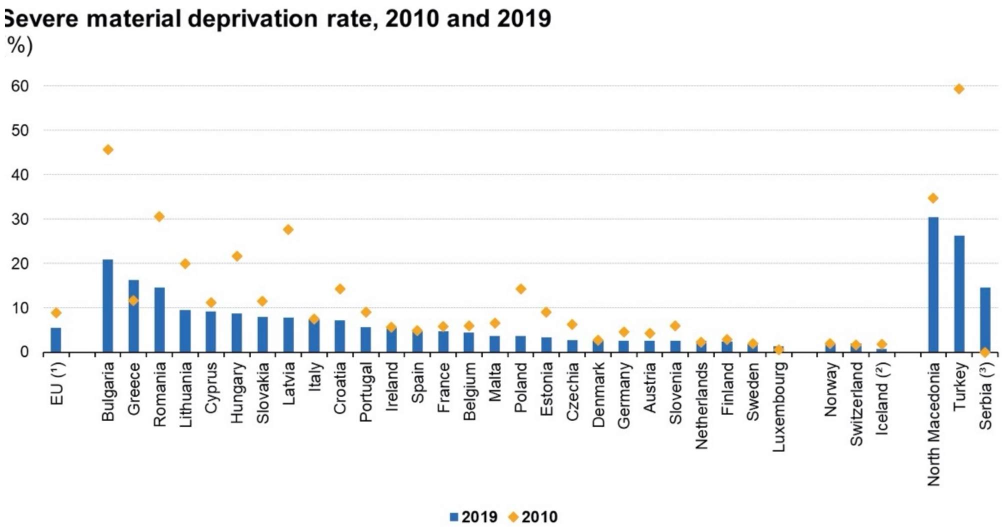

Figure S.8.24: Physical condition of housing stock. ***Source:*** Eurostat, 2019

Information: Figure S.8.24 shows the physical condition of housing stock in 2010 and 2019. In Cyprus, approximately 10% of buildings were inadequately built, while in the EU-27, this was 5%. The results reveal that Cypriot housing stock is twice as likely to be inadequately built as housing stock in other EU-27 countries. This is due to the absence of insulation materials during the construction phase of housing projects and a lack of implementation of the EPBD mandates. This data proves that in Cyprus, most homes are vulnerable to overheating risks and this condition has also led households to not be able to keep their homes warm in the winter.

Supplementary Material 8: Step-by-Step Development of Statistically Representative Archetypes in South-eastern Europe - Cyprus

Step 5: Building materials (Continued)

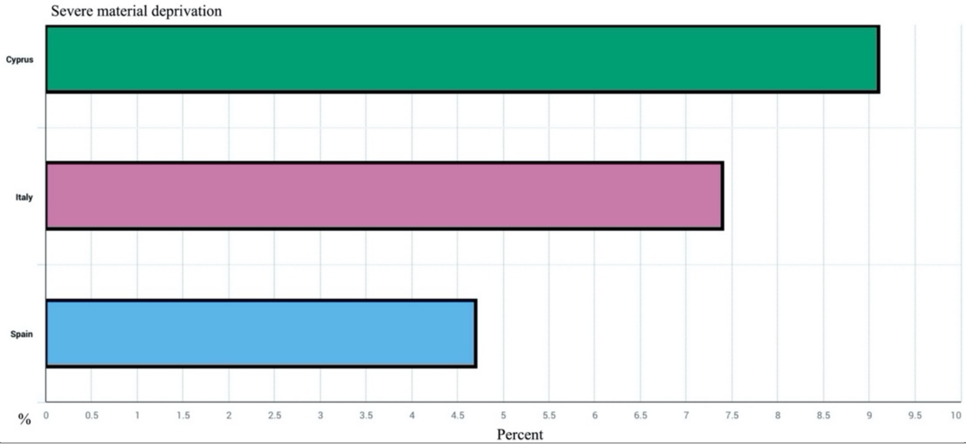

Figure S.8.25: Physical condition of housing stock in the South-eastern Mediterranean basin. ***Source:*** Eurostat, 2019

Information: Figure S.8.25 shows the physical condition of housing stock in Cyprus and two other two major Mediterranean countries, namely Italy and Spain. It can be seen that in Cyprus, over 9% of housing stock was physically inadequate, and this was the highest rate amongst these three countries. This is due to the absence of thermal insulation material during the construction phase of housing projects and the lack of implementation of the EPBD in the RoC.

Step 6: Construction period of European housing stock

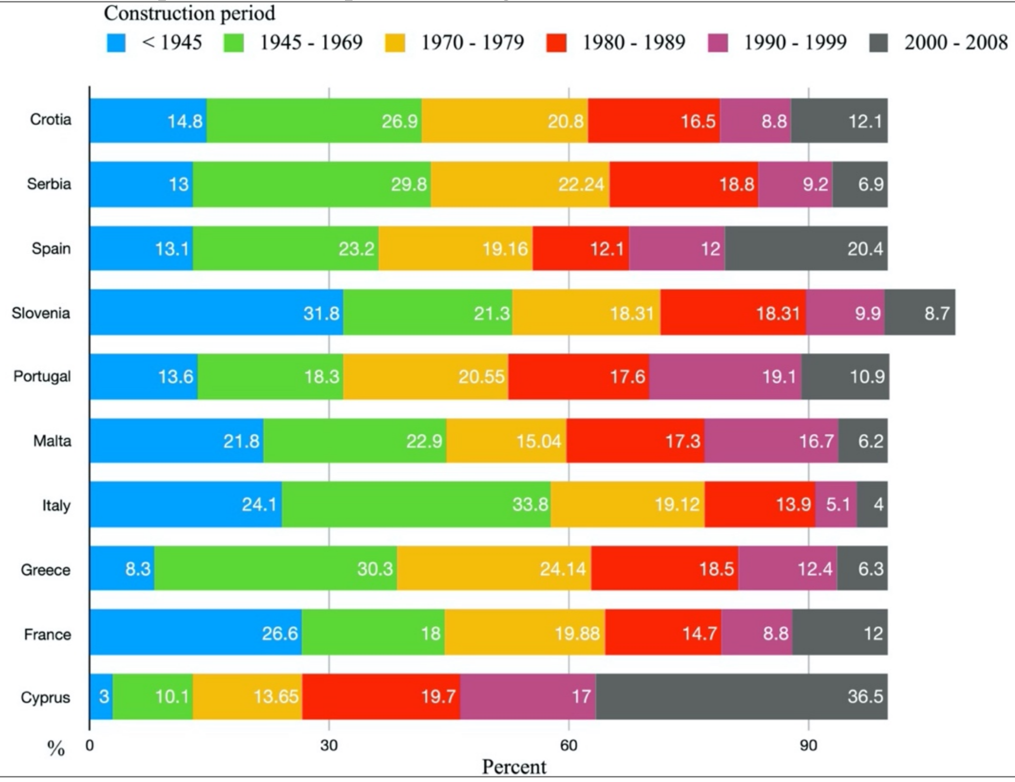

Figure S.8.26: Construction period of housing stock in Europe. ***Source:*** Eurostat, 2018

Information: Figure S.8.26 shows classification of housing stock by construction period. In Cyprus, 36.5% of housing stock was built between 2000 and 2008. Cypriot housing stock is relatively newly built-in comparison to other EU countries. Despite the high amount of newly built housing, the lack of regulatory bodies to implement the EPBD has led to thermally uncomfortable indoor conditions in many of its households.

Supplementary Material 8: Step-by-Step Development of Statistically Representative Archetypes in South-eastern Europe - Cyprus

Step 7: Living conditions

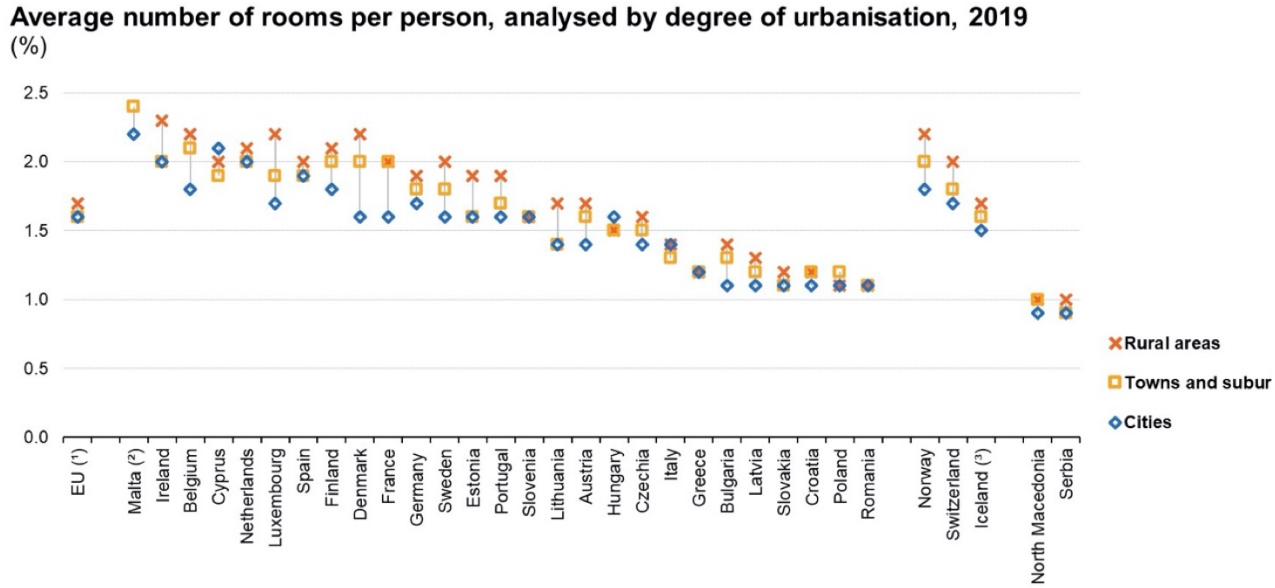

Figure S.8.27: Living conditions by number of rooms per person and urbanisation. ***Source:*** Eurostat, 2019

Information: Figure S.8.27 demonstrates the distribution of average number of occupied rooms by urbanisation type in EU countries in 2019. In Cyprus, households living in cities averaged just over two rooms per person, while households living in rural areas averaged just below two rooms per person. The EU average was just over 1.5 rooms per person in both cities and rural areas.

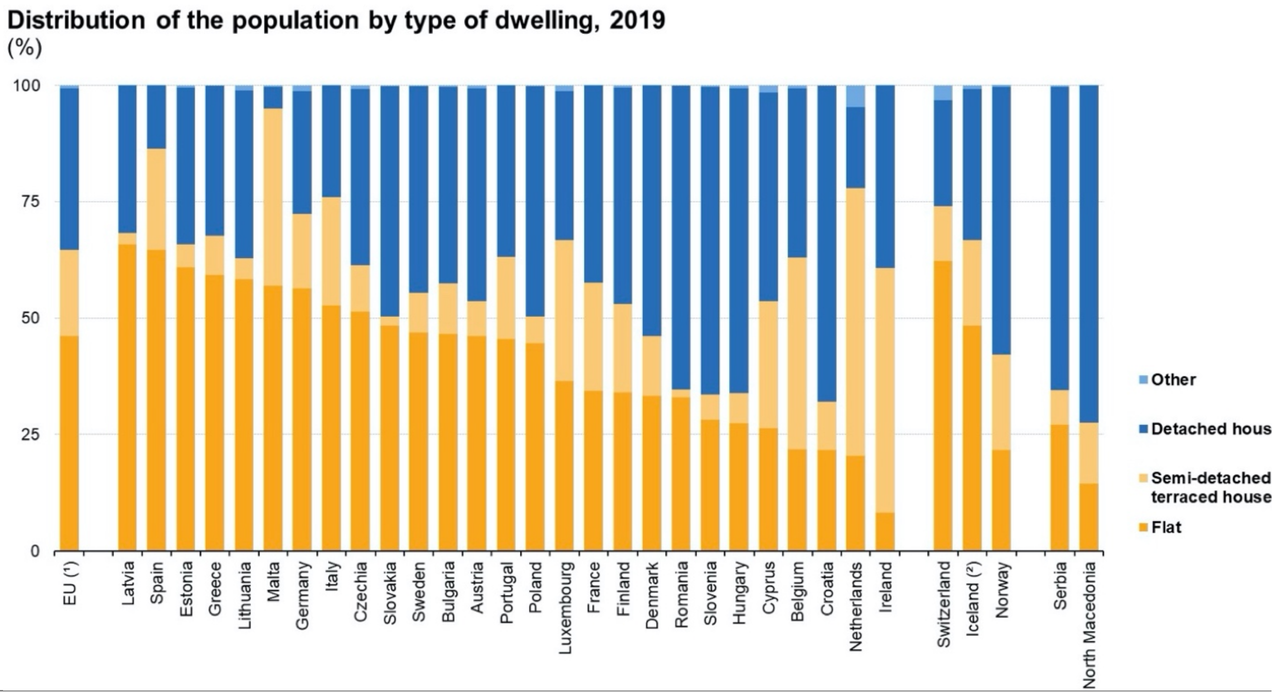

Figure S.8.28: Housing typology classification in Europe. ***Source:*** Eurostat, 2019

Information: Figure S.8.28 demonstrates housing typology classification by using census data for the EU. In Cyprus, in 2019, 26% of the population lived in flats and 18% lived in semi-detached or terraced houses. An average of 47% of the EU-27 population lived in flats. This data highlights that flats (apartments) are the most representative housing archetype for the Cyprus which is noteworthy in developing aggregate building energy models for energy policy design.

Supplementary Material 8: Step-by-Step Development of Statistically Representative Archetypes in South-eastern Europe - Cyprus

Step 8: Urbanisation

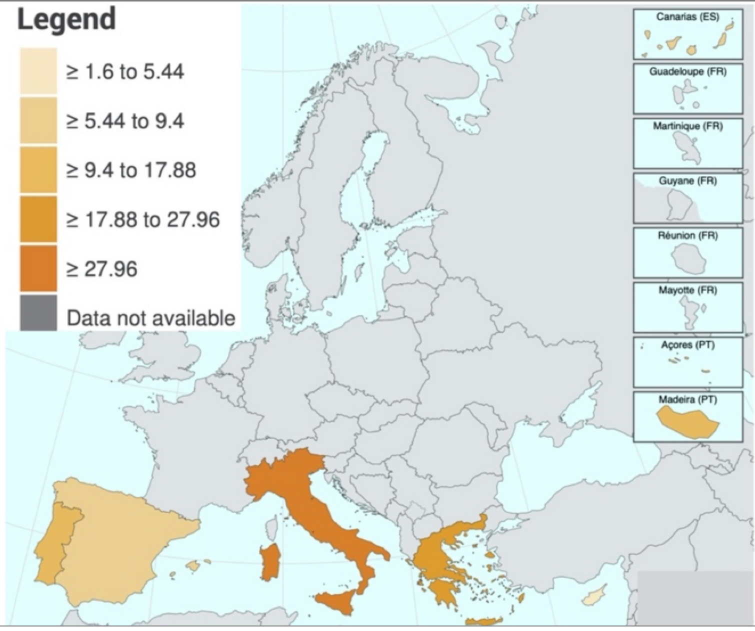

Figure S.8.29: Map of urbanisation in Mediterranean cities. ***Source:*** Eurostat, 2019

Information: Figure S.8.29 maps the degree of urbanisation in countries in the Mediterranean basin. Cyprus shows the smallest degree of urbanisation, between 1.6 to 5.44 – the data represents percent of the population that lives in urban versus rural environments, in comparison to the other EU countries. This is due to the relatively small population size of Cyprus compared to other European countries with larger populations and high numbers of housing stock.

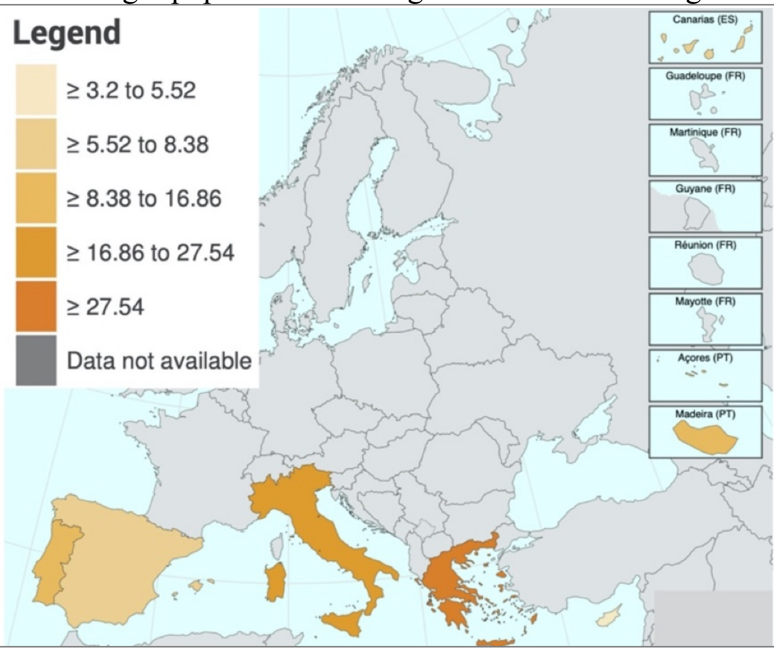

Figure S.8.30: Map of urbanisation in towns and suburban areas. ***Source:*** Eurostat, 2019

Information: Figure S.8.30 shows a map of degree of urbanisation in towns and suburban areas in Mediterranean countries. In Cyprus, 3.2% to 5.52% of households live in towns and suburban areas. In comparison to data presented in Figure S.28, this data proves that very low Cypriot households are located in towns due to the geographical conditions of the island. This is the reason that the identified post-war social housing stock were built the same in both cities and suburban areas without distinction to form. This representative archetype typology can therefore be applied to either urban or suburban regions for developing energy-policy design.

Supplementary Material 8: Step-by-Step Development of Statistically Representative Archetypes in South-eastern Europe - Cyprus

Step 8: Urbanisation (Continued)

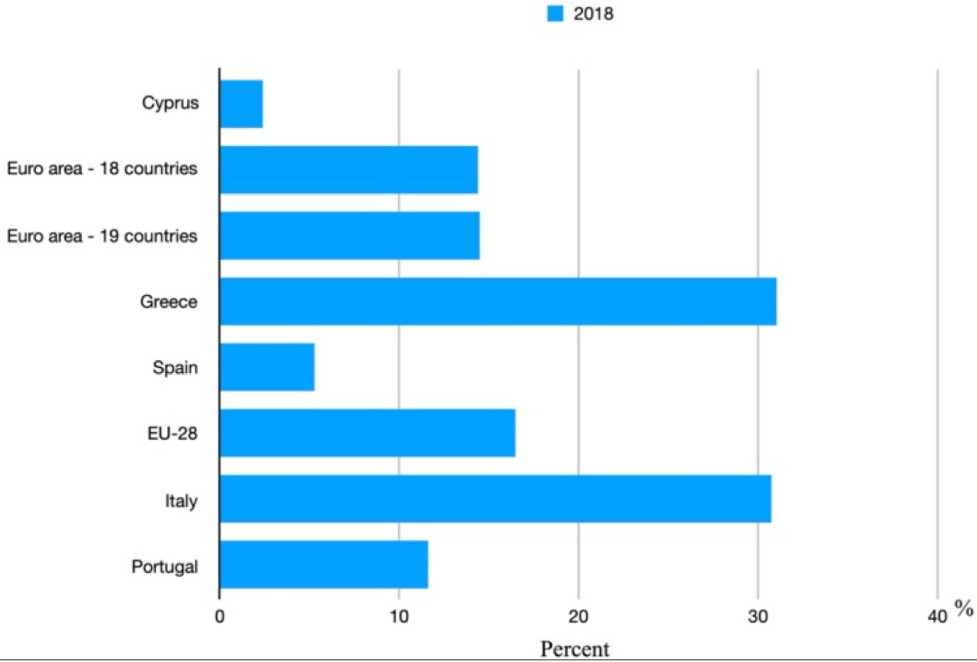

Figure S.8.31: Population living in cities. ***Source:*** Eurostat, 2018

Information: Figure S.8.31 shows the degree of urbanisation in cities. In Cyprus, 5% of the population lives in cities, while the EU-28 average was recorded at 16%. This finding demonstrates that the urbanisation rate is relatively small in Cyprus in comparison to other EU-27 countries. This is due to the smaller population of Cyprus.

Step 9: Identification of nationally representative archetypes

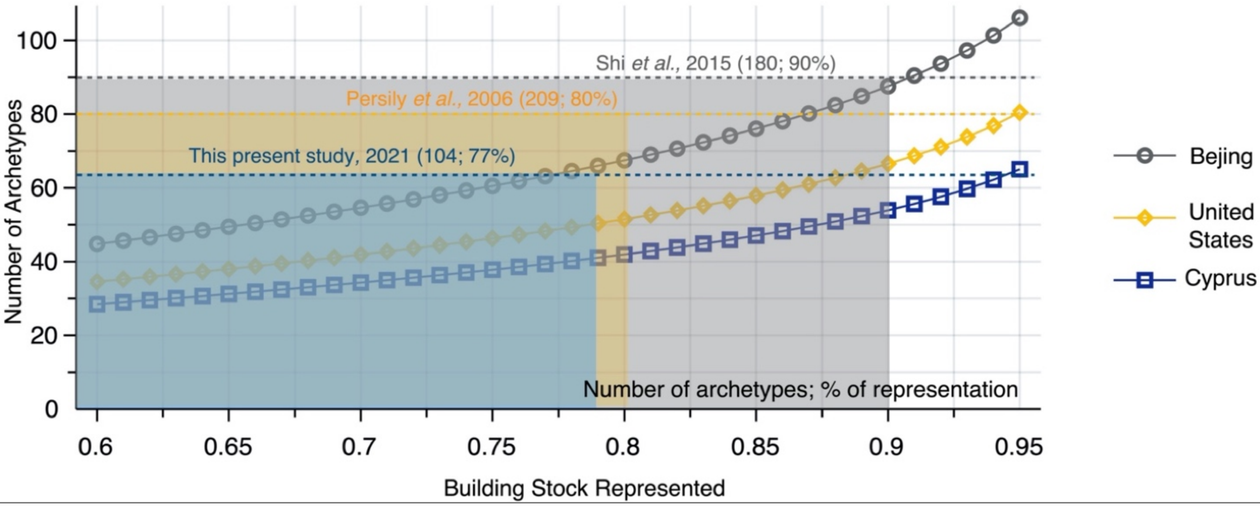

Figure S.8.32: Representativeness versus benchmarking criterion.

Information: Figure S.8.32 shows the representation of archetypes by analysing other scholars' work. After thorough evaluation of census data, household size, buildings' physical condition, building materials, construction period of the housing stock, living conditions and degree of urbanisation, this study set out to conceptualise the dataset to identify nationally representative archetypes. Representativeness of archetypes can help to determine the effect size, which is useful for reporting statistical findings.

In this present study, sample size was calculated using an online calculator and power estimator tool to identify the error margin for benchmarking. Statistical data was constructed to identify the effect size of the selected archetypes and to explore the generalisability of the housing stock, which had a direct impact on the research outcome.

In Supplementary Material 8, all the secondary data resources that were necessary to establish the theoretical foundation developed by Persily *et al.* in 2006 are followed and step-by-step development stages are presented.

Figure S.8.32 illustrates that the post-war social housing estates selected to undertake the building energy modelling analysis in this study represent 77% of the housing stock in Cyprus. This is the benchmarking criterion that was applied to interpret the statistical findings and demonstrate a reliable energy policy design in the residential sector. To provide a reliable assessment method for the analysis of building energy simulations, this study also considered the nationally representative archetype percentage identified by Persily *et al.* in 2006. The methodological framework developed by Persily *et al.* in 2006 demonstrates that 80% representativeness is a reasonable fraction to represent the U.S. housing stock, while the pilot study conducted by Shi *et al.* in 2015 demonstrated that 90% representativeness is required to represent housing stock in Beijing.

In comparison to both the U.S. and Beijing pilot studies, population size is relatively small in Cyprus. Thus, the identified 77% representativeness could result in a reliable assessment criterion for benchmarking in the South-eastern Mediterranean climate of Cyprus.

Note: The statistical data presented in Figure S.8.32 represents secondary data resources gathered for the Republic of Cyprus and does not represent the housing stock in Northern Cyprus.

In this study, the necessary secondary data sources were collected to identify the representativeness of housing stock in Northern Cyprus. It was found that, in Northern Cyprus, residential buildings represent 36% of the entire building stock, and low-, medium- and high- residential tower blocks represent another 56% of the building stock. All of this data was considered to develop a nationally representative archetype analysis criterion for energy modelling in this study.