The Price Impact of Sustainability on Housing Prices in Barcelona

A Multidimensional Data-Driven Approach

Pompeu Fabra Barcelona

Master of Intelligent Interactive Systems

Abstract:

This study applies a data-driven multidimensional approach to study the price impact of sustainability on housing prices in Barcelona, Spain. The sustainable pricing factors are studied with pricing factors related to five dimensions: ecological, environmental, social, cultural, and financial-economic. In total 22 location-bounded sustainable pricing factors are proposed, based on a high number of sustainable variables, which are combined into PCA components if a high correlation is observed to avoid multicollinearity in pricing models. Hereby, prior literature is extended which often studied the price impact of sustainability only from one dimension or variable. The results of semi-log hedonic pricing models, estimated on respectively 13.500 and 10.500 dependent on the inclusion/exclusion of observations with missing energy labels data, provide evidence that the sustainable dimensions all increase the willingness to pay for housing. Additionally, demonstrative maps, which visualize the total price impact of sustainability on housing prices, show that houses with a high/low total price impact of sustainability are locally clustered. These results suggest that the price impact of sustainability on house prices can be increased and made fairer by local policy interventions in Barcelona. Furthermore, the code, which offers a high flexibility to visualize the results, to construct the demonstrative map is shared on GitHub.

Keywords: Sustainability; Data-Driven; Barcelona; Housing Prices

Universitat Pompeu Fabra Barcelona

Author:

Niels Box

Supervisor:

Manuel Portela

15 June, 2023

Section 1: Introduction

The Council of Barcelona has a goal to achieve a more sustainable living environment for the inhabitants of the city the Barcelona Agenda 2030 (Arcadis, 2022). Barcelona currently ranks as the 49th most sustainable city in the world out of 100 global cities based on the people, planet and nature aspects derived from the sustainable development goals of the UN (Arcadis, 2022). Higher scores on these sustainability aspects will increase the quality of life of people (Eurostat Statistics Explained, 2022). For instance, in Barcelona, Yanez et al. (2023) found that the development of the Eixos Verds Plan in Eixample, which creates more greenness in the neighborhood, is expected to increase the mental health of 30.000 inhabitants. This conclusion is supported by work of Triguero-Mas et al. (2015) for Catalonia studying the relationship between mental health and the presence of green spaces.

Sustainability has become a topic of increasing importance in the real estate market. For example, in the UK 77% of the people said they consider buying a more green home as their next property (N. Gosling, 2022). This is encouraged by financial institutions offering lower interest rates or a higher loan amount for green mortgages, driven by a lower expected risk and costs of living for borrowers after sustainable investments (World Green Building Council, 2022). Sustainability is also simulated by local policies. In Barcelona, a neighborhood plan is developed (Pla de Barris), including plans for each of the 73 subdistricts, that implements plans to improve social, economic, and urban actions conditions (Ajuntament de Barcelona, 2023a). As a result there is developing a "brown" discount for real estate that does not meet the "green" market expectations Tom Carson (Funds Europe, 2023). This green premium, or brown discount, is driven by the demand side, tenants are willing to pay higher prices for more sustainable properties (D. Worford, 2022).

Home buyers are willing to pay more for housing if they are more sustainable aware (Mandell & Wilhelmsson, 2011). For instance related to the condition of properties itself, prior research in Barcelona already provided evidence that higher energy labels are positively correlated with housing prices (Chen & Marmolejo Duarte, 2018; Dell'Anna et al., 2019). More specifically locational sustainable factors in Barcelona, for instance, is showed that the higher values for access to public services and amenities, perceived security, shorter distance to the seashore, a highway, and central business districts increase significantly willingness to pay for housing (Buonanno et al., 2013; Chen & Marmolejo Duarte, 2018; Dell'Anna et al., 2019; Graells-Garrido et al., 2021; Marmolejo-Duarte & Chen, 2022) On the other side, is a significant negative relationship found for better access to parks and gardens and shorter commuting times (Dell'Anna et al., 2019; Graells-Garrido et al., 2021; Marmolejo-Duarte & Chen, 2022). Although these relationship are found given the city structure of Barcelona where park and gardens are mainly located at the periphery and Barcelona has low commuting times in contrast to other cities.

These studies were mainly conducted by studying sustainability from one dimension, but sustainability in itself has many dimensions (United Nations, 2023). To address this issue, in this research the price impact of sustainability on housing prices is studied from five different dimensions: ecological, environmental, social, cultural, and financial-economic as proposed by Kauko (2019). This results in pricing models, in which potential multicollinearity is avoided by the construction of PCA components if a high correlation is observed, that include a more diverse and higher number of sustainable factors. To provide insight in which properties are positively and negatively impacted by sustainable pricing factors in the city of Barcelona the results are visualized by the construction of two demonstrative maps, showing the pricing impact of the sustainable factors. These maps, of which the code is shared on GitHub and include a wide range of options to visualize the results of the valuation models, make it possible to identify the difference in the price impact of sustainability between different areas in Barcelona.

Table of Contents

Section 2: Literature review

The literature review discusses the main findings concerning the impact of sustainability on housing prices. The price effect of sustainability is reviewed based on five dimensions as proposed by Kauko (2019): ecological, environmental, social, cultural, and financial-economic, which will also be analyzed in the empirical analysis of this research. The research reviewed consists of international, national, and regional studies.

Section 2.1: Ecological Dimension of Sustainability

The ecological dimension of sustainability ranges from energy efficiency to the accessibility of a property (Kauko, 2019). The price effect of an energy label, of which certification of have been mandatory for residential housing for member states of the European Union since 2009, is the most studied (EUR-lex, 2023). Almost always a positive relationship between an increase in the energy label and the willingness to pay by homebuyers is reported. In international work for example in the Netherlands, Germany, and England found by respectively Brounen & Kok (2011), Cajias et al (2016), and Fuerst et al.(2015). In specific to Spain, Ayala et al. (2016) found this premium in different regions with energy labels produced by surveys even as La Paz et al. (2019) in Alicante. Although La Paz et al. (2019) found that the exact premium depended on the climate area where the property was located. In Barcelona, evidence was found for an increase in the housing price of 1.89% for each level increase in the energy label by Dell'Anna et al. (2019), although a lot of observations were excluded from the sample by missing energy labels. To address this issue, Chen et al. (2018) estimated in later work a Heckman selection model showing an increase of the housing price of 0.9% to 2.0% for each level increase in the energy label. However, in future work, Chen et al. (2022) reconsidered these findings by showing that energy labels can become insignificant pricing if other variables are included related to the architectural quality of properties arguing that earlier findings are mainly caused by the presence of an information asymmetry in the models. The accessibility of an area by both short, by bus and metro, and long-commuting, train, and highway, transport options is another often-studied ecological pricing factor. Research on short-commuting transport options often reported mixed results. For example, Cui et al. (2018) found in China a significant positive relationship between the closeness to the metro and the housing price, but also found a significant negative relationship for the closeness to the bus. Contradictory, Eichholtz et al. (2013) reported in the U.S. a significant and positive price premium for houses within a range of 0.25 miles of a public transport option. Specific to Spain, Taltavull de La Paz et al. (2019) found a positive relationship between the closeness to the bus and housing prices, but Graells-Garrido et al. (2021) did not find evidence for this relationship for the accessibility to the bus, metro, and shared bicycles. Related to longer transport commuting options a positive relationship is often reported for the closeness to train stations and housing prices. For example, Debrezion et al. (2011) found for a majority of cities in the Netherlands a significant positive relationship between the closeness of train stations and housing prices. These results were also found in Spain by Taltavull de La Paz et al. (2019). In line with Zhang et al. (2016) found in China that an improvement in the rail network results in higher housing prices showing a 0.023% increase in price with a 1% increase in the mileage of the network in the area close to the property. In contrast to this mixed results are found to the closeness to highways caused by a tradeoff between the higher accessibility and the negative externalities of the highway such as air and noise pollution (Tillema et al., 2012). For example, Debrezion et al. (2011) found in the Netherlands significantly lower housing prices if a highway was located within 100 meters. However, Ayala et al. (2016) found in Alicante, Spain, a positive price effect even as Del Anna et al. in Barcelona (2019).

Section 2.2: Environmental Dimension of Sustainability

The research that focused on the environmental dimensions covered the accessibility, amount, and view of different types of natural areas, such as the beach or parks. For the distance to the beach, a higher willingness to pay is mainly found for properties closer to the beach. For example, Dell'Anna et al. (2019) reported in Barcelona a significant negative relationship between the housing price and distance to the seashore and Chen et al. (2022) observed a positive price premium for residential real estate located within 200 meters of the seashore. This higher willingness to pay is also found often found in prior literature for the accessibility to parks and garden. For example, Park et al. (2017) and Kim et al. (2020) found in Seoul a significant positive relationship between the accessibility and closeness to parks, and in China the same conclusions were stated by Cui et al. (2018) and in Germany by Brandt et al. (2014). Respectively to Barcelona, Dell'Anna et al. (2019) found contrary results with a significant positive relationship between the distance to parks and housing prices. Dell'Anna et al. (2019) it was likely caused by the structural design of the city, where the parks are often located at the periphery. Instead of the closeness, another environmental factor is the view on nature, which also is found in prior work to cause a higher willingness to pay for housing. For example, Lee et al. (2020) found in China a positive price premium for both houses with a river and/or green view. In line with this, a positive premium paid for a sea view in Sweden was found by Wilhelmsonn et al. (2020). These findings are supported for Spain in Malaga by Castro Noblejas et al. (2022) for properties with a higher quality visual basin based on areas with vegetation and a sea view.

Section 2.3: Social Dimension of Sustainability

Prior literature that studied the social dimension of sustainability is mainly focused on the accessibility of public amenities/services and the impact of demographic statistics on the willingness to pay for housing. Related to the accessibility to public amenities a wide different range of facilities are studied, which often reported a relationship between a higher willingness to pay and willingness to pay. For example, Cui et al (2018) found in China a significant positive relationship between the closeness to public common goods, such as the hospital, educational facilities or work, and housing prices. In Spain, Alicante, this significant price effect was also observed by Taltavull de La Paz et al. (2019) for retail areas, but not for the closeness to healthcare facilities. In specific to Barcelona a significant positive correlation was found between willingness to pay for housing and access to food places, shops, education, financial services, government services, professional services, and recreational and healthcare facilities was found by Graells-Garrido et al. (2021). Concerning demographic factors, an often-studied price factor is the population density of a neighborhood, for which mixed results are found in different countries. For example, a negative relationship with the housing price is found by Eicholtz et al. (2013) and Lazrak et al. (2014) in respectively the U.S and Netherlands. However, opposite results were found in Germany by Cijas et al. (2016) and in Spain, this positive price effect is also found in different regions by Ayala et al. (2016) and in Alicante by Taltavull de La Paz et al. (2019). The safeness of the neighborhood is another often considered pricing factor. Prior work reported mostly a higher willingness to pay for housing when neighborhoods have a (considered) higher safety. For example, Ceccato et al. (2020) found in Sweden the closeness to crime hotspots had a significant impact on the prices paid for single-family houses. In Barcelona, Buonanno et al. (2013) found for perceived security a significant positive and for the crime perception rate a significantly negative relationship and the willingness to pay for housing. The last often-studied demographic factor is population growth, caused by a combination of the natural population growth and the net immigration rate, where a higher population growth increases the demand for housing. That this increase in housing price causes higher housing prices is for example found by Jeanty et al. (2010) in the U.S. With only studying the immigration rate Buonanno et al. (2013) found in Barcelona that a higher immigration rate has a positive effect on housing prices.

Section 2.4: Cultural Dimension of Sustainability

Empirical work that studied the impact of the cultural dimension of sustainability investigated mainly the premium paid for buildings that were classified as monumental or were located close to monuments or recognition of religion and culture. In general, this research found that the cultural value of a building or neighborhood is positively related to higher housing prices. For example, Debrezaion et al. (2011) and Lazrak et al. (2014) found respectively that properties that have a monumental status are sold for a premium of 17.6% and 26.9% in the Netherlands. In addition, Lazrak et al. (2014) observed a significant spillover effect for properties within a 50-meter radius of 0.28%. Instead of the classification of buildings of monumental also the recognition of culture in neighborhoods can have significant positive relationships with housing prices. For example for the recognition of religion, Brandt et al. (2014) observed in Germany a significant positive relationship between the closeness to the place of worship and housing the housing price. In Barcelona, this significant correlation is found by Graells-Garrido et al. (2021). Additionally did Graells-Garrido et al. (2021) found also a significant positive correlation between the willingness to pay for housing and the accessibility to entertainment places, where often cultural values are expressed and reflected.

Section 2.5: Economic/Financial Dimension of Sustainability

The fifth sustainable dimension as proposed by Kauko (2019) is the financial-economic aspect. Income equality and the welfare of a neighborhood are often included pricing factors representing this dimension of sustainability in empirical work. In specific to income equality often positively a relationship with housing prices is found in prior work. For example, Chen et al. (2018) found in Barcelona the cumulative percentage of people in higher socioeconomic classes in a subdistrict is significantly positively correlated with housing prices. This was also found in later work by Chen et al. (2022) by including the cumulative percentage of people with high occupational positions in a neighborhood as an income equality measure. Different results are found for welfare statistics, which are mainly reported to have a positive effect on housing prices. For example, Bruyne et al. (2013) found in Belgium that a lower unemployment rate had a positive relationship with housing prices supported by findings of Eicholtz et al. (2013) and Cajias et al. (2016) in respectively the U.S. and Germany. And a higher average income was found by Brandt et al.(2014), and Taltavull de La Paz (2019) found to have a significant positive effect on housing prices in respectively Germany and Alicante, Spain. Even more, higher economic activity was found to be positively correlated with housing prices in Sweden by Mandell et al. (2011), and in Barcelona was found that closeness to a place with high economic activity (CBD) was found to have a significant positive relationship with housing prices Marmolejo-Duarte et al. (2022).

In summary, prior literature almost always finds evidence for a positive relationship between housing prices and sustainable variables. This is found for all the five dimensions of sustainability that were analyzed: ecological, environmental, social, cultural, and economic-financial as proposed by Kauko (2019). Hereby for each dimension of sustainability, a higher willingness to pay is defined by the hypothesis when the sustainability factors increase. Relative to the ecological dimensions, the price effect of the energy label is excluded from the hypothesis, given the fact that the energy label is the only non-location bounded variable. Furthermore, only for the economic-financial dimension there is often observed a negative relationship between income equality, but a positive relation with relation to welfare the housing prices. Therefore the hypothesis proposes both a positive effect on welfare and a negative effect for income equality on the housing price. Additionally, since most sustainable variables, except the energy labels, are location-bounded a hypothesis is defined that a map can be constructed that defines the areas which are more/less impacted by the sustainable pricing factor for Barcelona.

H1: An increase in the ecological dimension of sustainability, excluding the effect of energy labels, results in higher housing prices.

H2: An increase in the environmental dimension of sustainability results in higher housing prices.

H3: An increase in the social dimension of sustainability causes results in higher housing prices.

H4: An increase in the cultural dimension of sustainability causes results in higher housing prices.

H5: An increase in welfare results in higher and an increase in income equality results in lower housing prices for the economic-financial dimension of sustainability.

H6: The observed price impact of sustainability in Barcelona shows local clustering tendencies.

Section 3: Data and Methodology

The data and methodology section consists of a description of both the collection of the housing-specific and sustainable data included in respectively sections 3a and 3b. Two samples are retrieved, one which includes the observations with the missing energy labels and one that excludes the observations with missing energy labels. Subject to the methodology are the applied valuation models described in section 3c and the construction of the demonstrative maps described to visualize the results in section 3d.

Section 3a: Housing-specific Data

The housing prices are collected from Idealista (2023a) by the use of web scraping. Idealista is the most popular and biggest housing platform in Spain. For Barcelona, it has listed more than 16.000 residential properties for sale. The dataset used in this research includes all the housing advertisements in Barcelona listed on Idealista on 17 April 2023. Although these prices on Idealista only show the asking price and thereby represent only the supply side, McGreal et al. (2010) showed that the asking prices tend to be close to transaction prices in the non-rapidly rising boom or bust housing market. The residential real estate market for Barcelona has been stable over the last year with a relatively consistent increase in prices (Idealista, 2023b).

Besides the housing prices also the housing-specific characteristics are collected from the housing advertisement on Idealista. The housing-specific variables that are created based on the information shown in the housing advertisements are included in table 1 below.

Table 1: An Overview of the Housing-Specific Variables

Variable	Description
Building Surface m^2	Continuous variable, that represents the building surface in m^2 of the property
Building age	Continuous variable, that represents the age of the property in years
House	Dummy variable, is equal to 1 if the property is a house, else 0.
Mezzanine	Dummy variable, is equal to 1 if the property is located on a mezzanine, else 0.
Ground Floor	Dummy variable, is equal to 1 if the property is located on the ground floor, else 0.
1st Floor	Dummy variable, is equal to 1 if the property is located on the 1st floor, else 0.
2nd-$5^{th}Floor$	Dummy variable, is equal to 1 if the property is located on the 2nd - 5th floor, else 0.
6th Floor or higher	Dummy variable, is equal to 1 if the property is located on the 6th floor or higher, else 0.
New Housing Development	Dummy variable, is equal to 1 if the property is newly constructed, else 0
Good condition	Dummy variable, is equal to 1 if the property is second-hand/ in good condition, else 0.
Needs renovation	Dummy variable, is equal to 1 if the property is second-hand/ needs renovation, else 0.
Elevator	Dummy variable, is equal to 1 if the property has an elevator, else 0.
Terrace	Dummy variable, is equal to 1 if the property has a terrace, else 0.
Balcony	Dummy variable, is equal to 1 if the property has an elevator, else 0.
Heating	Dummy variable, is equal to 1 if the property has a heating system, else 0.
Air Conditioning	Dummy variable, is equal to 1 if the property has air conditioning, else 0.
Outdoor Facilities	Dummy variable, is equal to 1 if the property has at least one of the following amenities: green area, garden, or swimming pool, else 0.
Parking Space Included	Dummy variable, is equal to 1 if a parking space is included in the house price, else 0.
Energy Label Consumption A	Dummy variable, is equal to 1 if the energy label consumption is label A, else 0.
Energy Label Consumption B	Dummy variable, is equal to 1 if the energy label consumption is label B, else 0.
Energy Label Consumption C	Dummy variable, is equal to 1 if the energy label consumption is label C, else 0.
Energy Label Consumption D	Dummy variable, is equal to 1 if the energy label consumption is label D, else 0.
Energy Label Consumption E	Dummy variable, is equal to 1 if the energy label consumption is label E, else 0.
Energy Label Consumption F	Dummy variable, is equal to 1 if the energy label consumption is label F, else 0.
Energy Label Consumption G	Dummy variable, is equal to 1 if the energy label consumption is label G, else 0.
Energy Label Consumption NA	Dummy variable, is equal to 1 if the energy label consumption is missing, else 0.

Table 1 includes an overview of the housing-specific characteristic variables with a description

Missing values in the housing-specific variables were dealt with by both assuming similarity between houses in the neighborhood and exclusion of observations. For the building year, which is missing in approximately 4900 (30%) of the advertisements an assumption is made that the building year is equal to the median of the subdistrict in which the residential property was located. The observations with missing floor-level data in housing advertisements are excluded from the sample. This results in the exclusion of approximately 2100 (13%) observations from the original sample. The last variable that was often missing was the consumption energy label of the residential property, which was missing in approximately 3500 (25%) of the observations in the remaining sample. These observations are included in the sample for the reported results of this study with the estimation of a Heckman selection model. As argued by Chen et al (2018), the missing observations for the energy labels are often the result of a sample selection bias, since the energy label might be not reported because it is expected to have a negative effect on the housing price. This would result in the pricing models, in which the strength of the price effect of energy labels on housing prices underestimates both the positive and negative price premiums compared to the reference category.

With the use of the Heckman selection model, this problem is solved is used by estimating the probability of the presence of an energy label in housing advertisements with a Probit model. The predictions of the Probit models are converted to an Inverse-mill ratio, which will include in the valuation model. This avoids the potential sample-selection bias, but for completeness results without the observation of the missing energy labels are discussed in the robustness section to identify potential errors. The only downside of the Heckman-selection model in this research is the low pseudo r-squared of the Heckman solution model with 2%, which can be caused either by the absence of a sample selection bias or missing data related to relevant indicators for the sample selection bias. Besides observations with missing data, a negative building age of minus 1 is observed for four observations, since the residential properties are still under development. The observations with negative buildings year are excluded from the sample. Furthermore are for the building surface m² variable two observations detected with a value of zero, whereas is the housing advertisement the specified building surface m² is 1.000. Therefore these values for the building surface m² are replaced with the correct values.

Outliers for the continuous housing-specific characteristic variables are identified for the log housing price, building surface in m^2, age of the building, and the housing price per m^2 for both samples before the exclusion of the observation with missing energy labels. The variable housing price per m^2 is not selected as both predicting or explaining variables in any of the models. However, it is included as the variable for identifying outliers, since the inclusion of the observations of outliers for this variable causes a highly non-linear relationship between log housing price and building surface in m^2. This relationship cannot be captured in the pricing models and thereby creates an incorrect model if the observations are not excluded. As shown in figure 1 below the distribution of the variables shows that before the treatment, outliers are mainly detected in the upper range of the distribution of the variables.

Figure 1: Histograms of the continuous housing-specific variables before outlier treatment

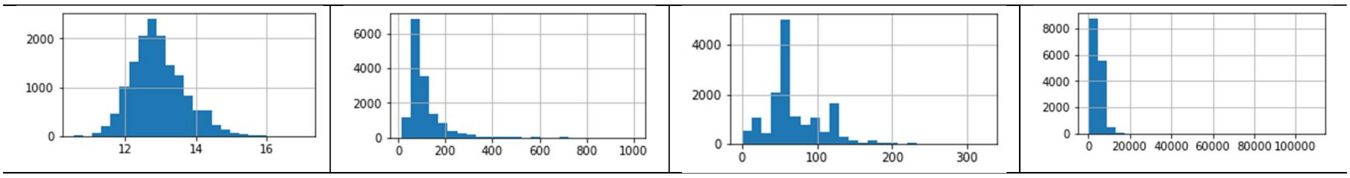


Fig 1a: Histogram of the Log Price	Fig 1b: Histogram of the Building surface m^2	Fig 1c: Histogram of the Building age	Fig 1d: Histogram of the Price per Building surface m^2

The deal with these outliers, the observations that have a value for one of these variables that is 2 standard deviations below or above the mean for one of these variables, a confidence interval of 95% under the gaussian normal distribution, are excluded from the sample. The treatment results in an exclusion of 1538 (11%) observations of the sample, which is more than under the normal distribution. However, for each continuous variable selected for the outlier treatment approximately 400-600 (2.5%) observations which is lower than under the gaussian normal distribution. The continuous housing-specific characteristic variables represent better a more normal distribution as shown in figure 2 below.

Figure 2: Histograms of the continuous housing-specific variables after outlier treatment

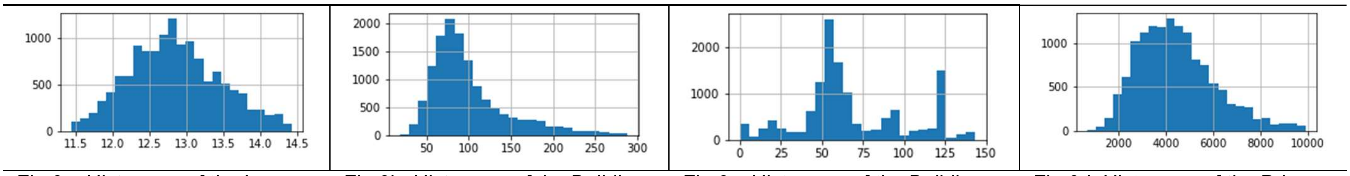


Fig 2a: Histogram of the Log Price

Fig 2b: Histogram of the Building surface m^2

Fig 2c: Histogram of the Building age

Fig 2d: Histogram of the Price per Building surface m^2

The final sample for the housing-specific characteristic sample that includes the observation with missing energy labels consists of 13358 observations. The final sample for the housing-specific characteristic sample, which excludes data points with missing energy labels consists of 10.104 observations. The summary statistics for the continuous housing-specific variables for the sample that includes observations with missing energy label data are shown in table 2 below. The summary statistics of the continuous housing-specific shows that 50% of the observations for the log price, building surface m², building age, sq(building surface m²), and sq(building age) are observed within a close interval around the mean showing similarity between most residential properties. Table 3 in the appendix shows that this distribution also holds for the continuous housing-specific features in the sample table excluding observations with missing energy label data.

Table 2: Summary Statistics of the Continuous Housing-Specific Variables for the sample that includes observations with missing energy label data.

	count	mean	std	min	25%	50%	75%	max	skew	kurt
Log price	13358	12.84	0.61	11.44	12.41	12.79	13.24	14.42	0.22	-0.41
Building surface m²	13358	99.73	46.13	18	69	87	116	288	1.44	2.03
Building age	13358	67.49	32.3	0	50.5	58	88	143	0.42	-0.37
sq(Building surface m²)	13358	12073.3	12629	324	4761	7569	13456	82944	2.49	6.98
sq(Building age)	13358	5597.99	4963	0	2550	3364	7744	20449	1.18	0.17

Table 2 includes the summary statistics for the continuous housing-specific variables for the sample that excludes the observations with missing energy label data.

Table 4 below shows that the housing-specific dummy variables in the sample that includes the observations with missing energy label data are having a good variety. The house dummy variable and energy consumptions label A have the lowest presence. However, these variables are still present in respectively 2.13% and 2.14% of the observations. The housing-specific dummy variable with the highest presence is the good condition dummy variable, which is present in approximately 81.82% of the observations in the sample that includes observations with missing energy label data. Table 5 included in the appendix for the sample that excludes the observations with missing energy label data show the same results. The only remarkable difference is the decrease in the presence of the new housing development dummy variable from 3.7% to 2.2%. This suggests that housing advertisements of newly constructed residential properties often exclude information on the energy consumption label.

Table 4: Summary Statistic of the Housing-Specific Dummy Variables for the sample that includes observations with missing energy label data

Variable	count	mean	Variable	count	mean
House	13358	0.02	Energy label consumption B	13358	0.02
Mezzanine	13358	0.05	Energy label consumption C	13358	0.04
Ground Floor	13358	0.1	Energy label consumption D	13358	0.09
1st Floor	13358	0.23	Energy label consumption E	13358	0.39
2nd – 5th Floor	13358	0.49	Energy label consumption F	13358	0.07
6th Floor or higher	13358	0.1	Energy label consumption G	13358	0.13
New housing development	13358	0.04	District Eixample	13358	0.21
Good condition	13358	0.82	District Ciutat Vella	13358	0.15
Needs renovation	13358	0.14	District Sant Martí	13358	0.1
Elevator	13358	0.77	District Sants-Montjuïc	13358	0.11
Terrace	13358	0.32	District Horta Guinardó	13358	0.09
Heating	13358	0.51	District Gràcia	13358	0.07
Outdoor facilities	13358	0.57	District Les Corts	13358	0.04
Air conditioning	13358	0.98	District Nou Barris	13358	0.06
Parking space included	13358	0.1	District Sarrià-Sant Gervasi	13358	0.1
Energy label consumption A	13358	0.02	District Sant Andreu	13358	0.06

Table 4 includes the summary statistics for the housing-specific dummy variables for the sample that includes the observations with missing energy label data.

Section 3b: Sustainability Data

The data for the creation of the sustainability variables, except the energy labels, are both retrieved from the council of Barcelona (Ajuntament de Barcelona, 2023) and OpenStreetMap (Openstreetmap Contributers, 2023). The Council of Barcelona provides a total of 564 open datasets on its website to stimulate research and innovation. Relevant datasets, which include sustainable information, are used in this research if they satisfy two conditions. Firstly, the datasets should include at least subdistrict-specific information. Secondly, the data should not have a too high correlation with other used sustainable data. Altogether, the sustainable information retrieved from the council of Barcelona covers a wide range of aspects of sustainability consisting of statistical data such as the unemployment rates, population density, geographical data with detailed scores for areas, such as vulnerability to a heat impact or social cohesion and data with the location of public amenities, such as the location of bars, hospitals or parks and gardens. The only important sustainable variable, which are not reported in any of the datasets is the location of the bus stops, train stations, beaches, and highways. These variables are in earlier research on the pricing of residential real estate in Barcelona and found to have significant predictive power (Chen & Marmolejo Duarte, 2018; Dell'Anna et al., 2019; Marmolejo-Duarte & Chen, 2022). Therefore this information is retrieved by the overpass API which reports information on the location of these properties retrieved from OpenstreetMap. OpenStreetMap is a free wiki world map, which is managed and hosted by volunteers (Openstreetmap contributers, 2023). An overview of the retrieved and created sustainable variables, the sustainable dimension of the variables, a link to the data source of which the variables are retrieved and a short description of the data source is included in table 6 included in the appendix.

The sustainable information retrieved from the Council of Barcelona and Openstreetmap consists of three types of variables. The type of variables is differently converted into variables containing sustainable information. First of all, data is retrieved that contains statistics on a subdistrict level, such as the unemployment rate, p80/p20 income distribution. If for these subdistrict level statistical data multiple countings are done during the last reported year, the average value is taken. Additionally on Idealista, a lower number of subdistricts are specified since Idealista merges some of the neighboring subdistricts of which information is individually reported by the council of Barcelona[1]. Therefore for the subdistricts which are presented as one subdistrict on Idealista, based on the variable the weighted average concerning the number of inhabitants or the sum is calculated from these statistics. The statistical subdistrict-level variables contain no outliers since the reported numbers are already an average number of each subdistrict. Secondly, data is retrieved that contains geographic information with benchmark scores specific to an area compared to other areas in Barcelona. The location of these areas is given as multiline strings in the datasets, which can be matched with the location of the properties specified by the latitude and longitude. For missing values, where the longitude and latitude of a property do not match with any of the multi-string included for the map, the median value for the subdistrict is used as a score. No outliers are identified in the datasets with geographic information with benchmark scores since the data is scaled in pre-defined levels. Thirdly, data is retrieved that contains information about the location of sustainable variables, such as the location of bus stops, parks & gardens, or universities. The distance of the property to each of these locations of the variables included in the dataset is calculated by the haversine formula by using both the latitude and longitude of the property and the latitude and longitude of the locations of the variables. Hereby, the distance to the nearest location of the variables and when relevant the number of locations of the variable within a specified range is calculated. The applied range depends on the expected distance within the presence of a location of the variable is expected to add utility for a home owner. For example, related to a hospital value is expected to provide utility to a home owner if it is present within a range of 1 kilometer of the property, whereas a bar or restaurant will only provide utility within a range of 0.25 kilometers. The outliers of the data that contain information about the distance to the nearest location and the number of locations within a prespecified for the locational sustainable variables are truncated. The truncation makes sure that the calculated variables have more normal distributions. Furthermore, does it support the reasoning that the presence of an amenity does not add any utility for a home owner if the distance is higher than a certain threshold or the number of amenities within the prespecified range is already sufficient. The ranges for which the truncation is applied are presented in table 7 which includes the relevant information related to the creation of the PCA components and distance to variables.

[1] Idealista specifies 69 subdistrict in Barcelona and the council of Barcelona specified 74 subdistrict in Barcelona. The following subdistrict are merged together by Idealista: Subdistrict Torre Baró, Ciutat Meridiana, and Vallbona are on Idealista presented as Subdistrict Ciutat Meridiana - Torre Baró – Vallbona, Subdistrict Can Peguera and El Turó de la Peira are on Idealista presented as Subdistrict Can Peguera - El Turó de la Peira, Subdistrict La Vell d’Hebron and La Clota are in idealista presented as Subdistrict La Vall d'Hebron - La Clota, subdistrict Sants Genís Del Agudells and Montbau are on idealista presented as subdistrict Sant Genís Dels Agudells - Montbau

To avoid high multicollinearity in the valuation models PCA components are constructed for the sustainable variables that have a high correlation with each other and contain information on similar types of amenity. Oladunni & Sharma (2016) found that this method is suitable for features when predicting housing prices with traditional hedonic pricing models. They found that PCA explains more variance of house prices than other comparable methods such as SVM and KNN. The PCA components will present a large share of the variance of the variables used to create the components if these individual variables have a high correlation with each other. To ensure comparability a constant scale during the creation of the PCA components is ensured with a distribution from 0 to 1. Most PCA components are a mix of the distance to the nearest amenity and the number of amenities within a certain range for one or multiple amenities that have a high correlation. Thereby for these PCA components the produced results represent an accessibility indicator for the included features, where is made sure that the minimum distance to coefficients has negative signs and the number within the prespecified range all have positive coefficients. The only PCA components that are not created on features with distance information are the income distribution and income & unemployment PCA components. These PCA components are created based on the economic statistics of the subdistrict. These individual variables to create the economic statistics are scaled beforehand, such that they retrieve similar coefficients during the creation of the PCA component[2]. Additionally, is squared of the statistic included in the PCA components given that inequalities or incomes often have a non-normal distribution. The results for the economical statistical PCA components have the interpretation that a higher value will be equal to higher income equality or welfare for a subdistrict. An overview of the created PCA components, their explained variance of the input features, the weight given to the input features, and the truncation of the variables on which the components are created are included in table 7 shown below. As shown in the table, the lowest share of the variance that is explained by a PCA component is 72.5% of the variance observed in the variables that are used an input. This lowest share of explained variance is found for the PCA components that represent the highest number of variables. The PCA components cannot be split into multiple components, since it introduces multicollinearity in the pricing models.

Table 7: Information on the construction of the sustainable variables

Variable	**E. Var %**	**Feature**	**Coef.**	**Trun.**
Bus & Metro PCA	98.11%	Distance to nearest bus (km)	-0.004	[0,0.4]
		Number of bus stops within 0.25 km	0.399	[0,25]
		Distance to nearest metro (km)	$-2*10^4$	[0,1.25]
		Number of metro stations within 0.25 km	0.001	[0,2]
Distance to Highway/Train (km)	NA	Distance to nearest highway (km)	N/A	[0,10]
		Distance to nearest train (km)	N/A	[0,3.5]
Distance to Beach (km)	100%	Distance to nearest beach (km)	1	[0,5]
Park & Garden PCA	99.15%	Distance to nearest park or garden (km)	-0.061	[0,0.5]
		Number of parks and gardens within 0.25 km	0.485	[0,2]
Viewpoint PCA	87.07%	Distance to nearest viewpoint (km)	-0.197	[0,3]
		Number of viewpoints within 1 km	0.206	[0,2]
Police PCA	95.48%	Distance to nearest police station (km)	-0.042	[0,2.5]
		Number of police stations within 1 km	0.224	[0,4]
Bar & Restaurant PCA	79.40%	Distance to nearest bar (km)	-0.-021	[0,2]
		Number of bars within 0.25 km	0.131	[0,5]
		Distance to nearest restaurant (km)	-0.006	[0,2.5]
		Number of restaurant within 0.25 km	0.072	[0,4]

[2] The average income household is defined in € 100.000 and the p80/p20 income distribution is divided by 10.

Variable	E. Var %	Feature	Coef.	Trun.
Second & Lower school PCA	72.22%	Distance to nearest under three years-old school (km)	-1*10^3	[0,1]
		Number of under 3 years-old schools within 0.5 km	0.027	[0,10]
		Distance to nearest 3-6 years-old school (km)	-8*10^4	[0,1]
		Number of 3-6 years-old schools within 0.5 km	0.031	[0,10]
		Distance to nearest primary school (km)	-8*10^4	[0,1]
		Number of primary schools within 0.5 km	0.031	[0,10]
		Distance to nearest secondary school (km)	-0.001	[0,1]
		Number of secondary schools within 0.5 km	0.023	[0,10]
University PCA	97.55%	Distance to nearest university (km)	-0.028	[0,2.5]
		Number of universities within 0.5 km	0.186	[0,5]
Pharmacy PCA	99.88%	Distance to nearest pharmacy (km)	-0.002	[0,1]
		Number of pharmacies within 0.25 km	0.100	[0,10]
Hospital & Clinique PCA	98.85%	Distance to nearest hospital or clinique (km)	-0.016	[0,1.5]
		Number of hospital or clinics within 0.5 km	0.195	[0,5]
Big Shopping Place PCA	72.82%	Distance to nearest shopping gallery (km)	-0.015	[0,3]
		Number of shopping galleries within 1 km	0.012	[0,3]
		Distance to nearest shopping center (km)	-0.008	[0,3]
		Number of shopping centers within 1 km	0.014	[0,3]
		Distance to nearest large establishment (km)	-0.013	[0,3]
		Number of large establishments within 1 km	0.091	[0,9]
Performing Arts PCA	89.94%	Distance to nearest cinema (km)	-0.007	[0,2.5]
		Number of cinemas within 0.5 km	0.011	[0,2]
		Distance to nearest theatre (km)	-0.007	[0,2.5]
		Number of theatres within 0.5 km	-0.087	[0,10]
		Distance to nearest concert place (km)	-0.013	[0,2.5]
		Number of concert places within 0.5 km	0.013	[0,2]
Religious Institution PCA	99.94%	Distance to nearest religious institution (km)	-9*10^4	[0,1.5]
		Number of religious institution within 0.5 km	0.067	[0,15]
Museum, Library & POI Cult. PCA	93.01%	Distance to nearest library (km)	-1*10^4	[0,1.5]
		Number of libraries within 0.5 km	0.004	[0,10]
		Distance to nearest museum (km)	-3*10^4	[0,1.5]
		Number of museums within 0.5 km	0.003	[0,10]
		Distance to nearest point of interest culture (km)	-1*10^4	[0,1.5]
		Number of point of interest culture within 0.5 km	0.019	[0,50]
Income Distribution PCA	97.65%	Income Distribution P80/P20/10	-2.134	NO
		Gini Index	-2.178	NO
		sq(Income Distribution P80/P20/10)	-1.324	NO
		sq(Gini Index)	-1.498	NO
Income & Unemployment PCA	99.05%	AVG income household in € 100.000	0.365	NO
		AVG unemployment rate %	0.022	NO
		sq(AVG income household in € 100.000)	-0.496	NO
		sq(AVG unemployment rate %)	0.002	NO

Table 7 includes the name of the sustainable feature (variable) , the explained variance of the input variables (E. Var %) , the input variables (Features) and their coefficients (Coef.) together with the applied truncation (Trunc). The value NA indicates that the information is not available for the variable

A final description of the sustainable features for the sample that includes the observations with missing energy label data is included in table 8 shown below. The summary statistics for the sustainable variables show based on the standard deviation, 25%, and 75% quantile that there is a higher variety in the sustainable variable compared to the housing-specific variables. This suggests there could be observed high differences between properties in certain areas to their price impact of sustainability on the residential properties. Both the results for the sample with the inclusion and the exclusion of observations with missing energy labels, as shown in table 8 below and table 9 in the appendix, are very similar compared to each other.

Table 8: Summary Statistics of the Sustainable Variables for the sample that includes the observations with the missing energy label observations

	mean	std	min	25%	50%	75%	max	skew	kurt
Bus & Metro PCA	0.37	0.18	0	0.24	0.36	0.48	1	0.58	0.41
Distance to Highway/Train (km)	1.15	0.72	0.02	0.62	0.99	1.51	4.91	1.13	1.35
Distance to Beach (km)	3.29	1.4	0.01	2.14	3.44	4.61	5	-0.37	-1.02
Park & Garden PCA	0.4	0.39	0	0.01	0.5	0.51	1	0.37	-1.33
Viewpoint PCA	0.41	0.26	0	0.21	0.32	0.64	1	0.67	-0.79
Neighborhood size (10 ha)	13.51	12.2	2.3	8.04	11.1	14.1	142.37	5.95	51.28
Vulnerable to heat impact (1-5)	2.86	0.79	1	2	3	3	5	0.08	-1.03
Police PCA	0.44	0.28	0	0.29	0.32	0.55	1	0.33	-0.78
Bar & Restaurant PCA	0.49	0.35	0	0.19	0.34	0.87	1	0.3	-1.36
Secondary & Lower School PCA	0.47	0.2	0	0.32	0.47	0.62	1	0	-0.6
University PCA	0.35	0.38	0	0.05	0.06	0.62	1	0.79	-1.04
Pharmacy PCA	0.5	0.22	0	0.4	0.5	0.6	1	0	-0.2
Hospital & Clinique PCA	0.36	0.28	0	0.21	0.22	0.6	1	0.67	-0.34
Big Shopping Place PCA	0.33	0.3	0	0.08	0.2	0.54	1	0.91	-0.48
Social Cohesion Score	0.21	0.18	0	0.09	0.15	0.3	1.73	1.81	4.73
Natural population growth ‰	-2.09	1.96	-8.92	-3.3	-2.5	-0.6	3.6	0.14	0.62
Net immigration rate ‰	26.86	22.41	-6.6	9.9	18.3	44.1	91.1	1.18	0.89
Density net ((hab/1000)/ha)	0.74	0.23	0.02	0.63	0.74	0.91	1.37	-0.4	0.1
Performing Arts PCA	0.3	0.3	0	0.06	0.17	0.41	1	1.18	0.11
Religious Institution PCA	0.44	0.25	0	0.27	0.4	0.6	1	0.56	-0.42
Museum, Library & POI Cult. PCA	0.25	0.23	0	0.09	0.16	0.34	0.99	1.35	1.01
Income Distribution PCA	0.46	0.21	0	0.3	0.47	0.62	1	-0.34	-0.59
Income & Unemployment PCA	0.2	0.2	0	0.07	0.15	0.27	1	1.92	3.67

Table 8 includes the summary statistics for the sustainable variable for the sample that includes the observations with missing energy label data.

Section 3c: Valuation Models

The housing-specific and sustainable variables used in the valuation models are included in Table 6. The housing-specific characteristic, $\beta_1 - \beta_{22}$, contain variables specifically bounded to the residential properties, such as the floor area, building year, and condition of the property. The energy consumption label was chosen over the energy emission label because it more directly represents a decrease in living costs for home buyers and the energy consumption label is always reported if the energy emission label is reported in the sample. The sustainable pricing factors are represented by five dimensions, ecological, environmental, social, cultural, and financial-economic, as proposed by Kauko (2019). The pricing factors for the ecological dimension, $Z_1 - Z_2$ include information mainly relating to transportation accessibility options of the neighborhood of a property for short (bus & metro) and long (highway & train) commute. The energy consumption label is excluded from the ecological dimensions of sustainability and included in the housing-specific characteristic given that is the only sustainable variable that is dependent on the location of the residential property. Thereby it is impossible to increase the sustainability score of these variables by local policy intervention. Policy intervention specific to energy labels will be property-specific and probably consists of support programs/grants available to every property in

Barcelona. The sustainable variables for the short and long commutes are each represented as one sustainable feature since there is a high correlation between the bus and metro accessibility and a trade-off between the distance to the highway and train in Barcelona. The pricing factors for the environmental dimension, $Z_3 - Z_7$, includes information related to nature closely: the beaches, parks & gardens, and viewpoints on nature, as well as the geographic structure of the area of the property: the neighborhood size and vulnerability level to heat impact. The social dimension pricing factors, $Z_8 - Z_{17}$, includes information related to the safety, and accessibility of public amenities and demographics of the neighborhood. The cultural dimension $Z_{18} - Z_{20}$ contains information concerning the recognition of religion and culture in the area with pricing factors for the presence of amenities for performing arts, religious institutions, and museums, libraries & cultural points of interest close to the residential properties. The financial-economic economic dimension, $Z_{33} - Z_{34}$ includes pricing factors about the income distribution and the level of welfare of the area. Additionally, are dummy variables included, to capture the effect of other geographical pricing factors of a district that are not included in the model. The inclusion of the dummy district variables decreases the probability that the estimated valuation models will be subject to an omitted variable bias.

Table 10: The Housing-Specific, Sustainable and Dummy District Pricing Factors

Housing-specific Pricing Factors	**Sustainable Pricing Factors**
$\beta_1 Building\ surface\ m^2$	$Z_1\ Bus\ \&\ Metro\ PCA$
$\beta_2\ sq(Building\ surface\ m^2)$	$Z_2\ Distance\ to\ Highway/Train\ (km)$
$\beta_3\ Building\ age$	$Z_3\ Distance\ to\ Beach\ (km)$
$\beta_4\ sq(Building\ age)$	$Z_4\ Park\ \&\ Garden\ PCA$
$\beta_5\ House\ (YES = 1, ELSE\ 0)\ (REF = Ground\ Floor)$	$Z_5\ Viewpoint\ PCA$
$\beta_6\ Mezannine\ (YES = 1, ELSE\ 0)\ (REF = Ground\ Floor)$	$Z_6\ Neighboordhood\ size\ (10\ ha)$
$\beta_7\ 1^{st} Floor\ (YES = 1, ELSE\ 0)\ (REF = Ground\ Floor)$	$Z_7\ Vulnerable\ to\ heat\ impact\ (1-5)$
$\beta_8\ 2^{nd} - 5^{th}\ Floor\ (YES = 1, ELSE\ 0)\ (REF = Ground\ Floor)$	$Z_8\ Police\ PCA$
	$Z_9\ Bar\ \&\ Restaurant\ PCA$
$\beta_9\ 6^{th}\ Floor\ or\ higher\ (YES = 1, ELSE\ 0)\ (REF = Ground\ Floor)$	$Z_{10}\ Secundary\ \&\ Lower\ School\ PCA$
	$Z_{11}\ University\ PCA$
$\beta_{10}\ New\ housing\ Development\ (YES = 1, ELSE\ 0)\ (REF = Good\ condition)$	$Z_{12}\ Pharmacy\ PCA$
	$Z_{13}\ Hospital\ \&\ Clinique\ PCA$
$\beta_{11}\ Needs\ renovation\ (YES = 1, ELSE\ 0)\ (REF = Good\ condition)$	$Z_{14}\ Big\ Shooping\ Place\ PCA$
	$Z_{15}\ Social\ Cohesion\ Score$
$\beta_{12} Elevator\ (YES = 1, ELSE\ 0)$	$Z_{16}\ Natural\ population\ growth$ ‰
$\beta_{13}\ Terrace\ (YES = 1, ELSE\ 0)$	$Z_{17}\ Density\ net\ (hab/1000\ ha)$
$\beta_{14}\ Heating\ (YES = 1, ELSE\ 0)$	$Z_{18}\ Performing\ Arts\ PCA$
$\beta_{15}\ Outdoor\ Facilities\ (YES = 1, ELSE\ 0)$	$Z_{19}\ Religious\ Institution\ PCA$
$\beta_{16}\ Parking\ space\ included (YES = 1, ELSE\ 0)$	$Z_{20}\ Museum, Library\ \&\ POI\ Cult. PCA$
$\sum_{i=17}^{22} B_i Dummy\ Consumption\ Energy\ Label (= 1\ if\ label = energy\ label\ with\ \ Label\ D\ as\ reference\ category)$	$Z_{21} Income\ Distribution\ PCA$
	$Z_{22} Income\ \&\ Unemployment\ PCA$
Dummy District Variables	$\sum_{i=1}^{10} X_i Dummy\ District\ (REF = Les\ Corts)$

Table 10 includes the housing-specific, sustainable, and dummy district pricing factors used to estimate the pricing models.

The valuation model to be applied to estimate housing prices in Barcelona is the hedonic pricing model as proposed by Rosen (1974). The hedonic pricing model assumes that the price that is paid for housing is equal to the total utility buyers are expected to extract from it. The total utility is the sum of all the marginal utilities created by the housing-specific, sustainability, and dummy-district variables. The hedonic pricing model or a more or less similar model is often used in prior literature supported by the general conclusion of researchers ten years ago that ML methods do at best equal the performance of the hedonic pricing model (Kauko, 2019). The hedonic pricing model is estimated as a semi-log model. The semi-log model offers some advantages over the traditional OLS pricing model (Marmolejo Duarte & González Tamez, 2009) The model helps to normalize the housing prices by taking the logs and thereby decreases the impact of outliers. Furthermore, can the pricing factors can be interpreted as semi-elastic. Thereby the coefficients will represent the effect in percentages of the change in the housing prices of a one-unit increase rather than absolute values. This interpretation makes it easier to explain the results and compare them with earlier work in this research field. Furthermore, is this model also applied in earlier work on the price effect of sustainable factors on housing prices in Barcelona. The semi-log hedonic pricing model is estimated concerning different sustainable dimensions, which are specified as followed:

Model 1: $ln(y) = \alpha + \sum_{i=1}^{22} \beta_i Housing - Specific\ Var\ + \sum_{\mathrm{i}=1}^{10} X_i\ District + \varepsilon\ (1)$

Model 2: $ln(y) = \alpha + \sum_{i=1}^{22} \beta_i\ Housing - Specific\ Var + \sum_{i=1}^{2} Z_i Ecological\ Var + \sum_{\mathrm{i}=1}^{10} X_i\ District\ Dummy + \varepsilon\ (2)$

Model 3: $ln(y) = \alpha + \sum_{i=1}^{22} \beta_i\ Housing - Specific\ Var\ + \sum_{i=2}^{7} Z_i\ Environmental\ Var + \sum_{\mathrm{i}=1}^{10} X_i\ District\ Dummy + \varepsilon\ (3)$

Model 4: $ln(y) = \alpha + \sum_{i=1}^{22} \beta_i\ Housing - Specific\ Var + \sum_{i=8}^{17} Z_i\ Social\ Var + \sum_{\mathrm{i}=1}^{10} X_i\ District\ Dummy + \varepsilon\ (4)$

Model 5: $ln(y) = \alpha + \sum_{i=1}^{22} \beta_i\ Housing - Specific\ Var + \sum_{i=18}^{20} Z_i Cultural\ Var + \sum_{\mathrm{i}=1}^{10} X_i\ District\ Dummy + \varepsilon\ (5)$

Model 6: $ln(y) = \alpha + \sum_{i=1}^{22} \beta_i\ Housing - Specific\ Var + \sum_{i=21}^{22} Z_i\ Econ/Fin\ Var + \sum_{\mathrm{i}=1}^{10} X_i\ District\ Dummy + \varepsilon\ (6)$

Model 7: $ln(y) = \alpha + \sum_{i=1}^{22} \beta_i\ Housing - Specific\ Var + \sum_{i=1}^{22} Z_i\ Sustainable\ Var\ + \sum_{\mathrm{i}=1}^{10} X_i\ District\ Dummy + \varepsilon\ (7)$

where α is the constant in the model and the $\beta's$ are the non-sustainable housing-specific pricing factors, the $Z's$ are the sustainable pricing factors, and the $X's$ are the district dummies as included in table 10 and ε is the error term. In each model, the housing-specific pricing factors and dummy district variables are included. In addition, models 2-6 each include one dimension of the sustainable pricing factors. Model 7 is the complete model which includes all the housing-specific characteristics and sustainable pricing factors. The correlation matrix for the pricing factors that have an absolute correlation of over 0.25 with other pricing factors is shown for both the sample that include and exclude the observation with missing energy labels in respectively figure 3 and 4 in the appendix. The correlation matrixes shows a high similarity with each other. Furthermore will after the estimation of the hedonic pricing models the variable inflation factor (VIF) be calculated, which represents the potential existence of multicollinearity in a model. If the VIF is too high there will be chosen to either exclude variables of the model or create additional PCA components.

To deal with the sample selection bias for the sample that includes the observation with the missing energy labels the Heckman selection model is estimated. The Heckman selection model solves the sample selection bias by predicting whether or not the energy label is missing in a Probit model. With these predicted variables for this dummy variable, the inverse Mill ratio (IMR) can be calculated. The inverse Mill ratio (IMR) is added to the above-specified hedonic pricing models (models 1-7). The Probit model, which includes variables related to the state of the property or is likely to correlate with the energy consumption of the property, to predict whether or not an energy label is missing is specified as followed:

PROBIT: $Energy\ Consumption\ Label\ Present\ (y) = \alpha + \beta_1\ Building\ Surface\ m^2 + \beta_2\ Building\ Age + \beta_3\ New\ housing\ development + \beta_4\ Needs\ renovation + \beta_5\ Elevator + \beta_6\ Terrace + \beta_7\ Balcony + \beta_8\ Air\ conditioning + \beta_9\ Outdoor\ Facilities + \varepsilon\ (8)$

where α is the constant in the model and the $\beta's$ are the non-sustainable housing-specific pricing factor that could influence the decision to publish the energy label in the housing advertisement, and ε is the error term.

Section 3d: Visualization of the Results by a Map

To display the results of the pricing models above a map is developed to visualize the pricing of sustainability on a property and neighborhood-specifical scale. The map provides helpful insight given that the sustainable variables in this research are mostly bounded to the location of a property. Therefore it is expected, as defined in hypothesis 6, that the price impact of the sustainable pricing factors on a local scale are similar to each other in absolute terms. The Map is built with Leaflet, which allows interactive of the map. For example, a user, can zoom in and out on the map, where based on the zoom level the properties will be clustered (Leaflet, 2023). The code to build the map, with the estimated pricing models, is shared on Github[3]. The parameters that can be changed to build different types of maps are described below. Furthermore the parameters for the construction of two demonstrative maps are described, which will be discussed in the results section, which visualizes the price impact of sustainability on the housing price both on housing-specific and neighborhood levels.

The map provides different options to visualize the price impact of the variables in the pricing models. The main feature of the map is to display the information for the details of the variables chosen to color the map. However, also general information, which includes information about the predicted price, listing price, impact of housing-specific, selected-sustainable features, and sustainable and district dummy variables in the predicted price of the residential property is provided when clicking on the icon of the observations. Besides the price impact of the sustainable-variables included in the visualized valuation model, also the selected sustainable features price impact option is included since some sustainable variables in the pricing models show non-positive relationships that between a higher sustainability and housing prices and to exclude some sustainable variables that cannot be changed by policy interventions. An example of such a non-sustainable relationship is for example the income equality and housing-prices found in prior research. An example of a variable that cannot be influenced by local policy is the distance to the beach. These variables have therefore a lower relevance when showing the total price impact of sustainability on the housing prices in Barcelona.

More specifically is also the data provided concerning the housing-specific, sustainable, and dummy district on a variable-specific scale. All the observations cannot be displayed at once given the large sample size. Hence observations two different options to cluster the observations are included. The first option is automatic clustering of the observations by Leaflet, where information regarding the individual residential properties is hereby only displayed when clicking on and opening the clusters. The second option is clustering into one observation by the support vector machine based on the latitude and longitude of the residential properties. Hereby the average values for the sustainable variables will be provided, the downside of this method is that the information for the housing-specific and dummy-district variables will be suppressed. There parameters, with a description, used to construct the maps are included in the table 11 shown below. The parameters to construct the map offer high flexibility for the selected valuation models, the selected shown observations, the variable to color the map, the number of categories to color

[3] https://github.com/NielsUPF/Demonstrative_MAP

the map, and the clustering method that is used. The selected shown observations can be chosen by subsetting the data frame based on the options equal to, not equal to, higher than, and lower than for the included parameters in the relevant plotted pricing model. The variables to color the observations are all the continuous variables included in the pricing models, in addition to total the absolute price impact of housing-specific, selected sustainable, or sustainable variables. The number of categories will be based on the quantiles of the continuous variable used to color the map. The categories will always be visualized in three colors, the lowest category will be visualized in red, and the highest category in green. The observations in the middle categories will be visualized in gray. Additionally can be chosen to only display the observations in the lowest and highest category. The notebook to construct the map is shared on GitHub[4].

Table 11 The parameters used to maps to visualize the results of pricing models

Parameters	**Description**
Selected Sustainable Features	List with features that are used to calculate a sustainability price impact that excludes the not included sustainable features. The list makes it for example possible to exclude the price impact of sustainability on factors that cannot be influenced by policymakers, such as the distance to the beach. Only sustainable variable can be chosen that are included in the pricing model.
Map_save_name	The name of the map used to the file.
Title	The title that will be given to the map.
Subtitle	The subtitle that will be given to the map.
Circle_Multiplier	The Multiplier applied to use for the visualization of the circles by the clustering of the observations with the Markercluster. The perfect parameter setting depends on the length of the used (subsetted) dataframe or the number of clusters that are specified. The formula for the size of the circle is as follows: radius = Circle_Multiplier * $\log(n\_obs\_in\_cluster)/\,2$
DF	The dataframe that is used to estimate the pricing models.
Model_result	The estimated results of the pricing model.
Color_var	The variable which is used to color the map. The variables that can be selected include the predicted price, the predictors used in the model and three variables to calculate the total price effect: “Selected Sustainable Features Price Impact”, “Sustainable Features Price Impact”, “Housing-Specific Features Price Impact”.
N_color_cat	The number of categories to color the specified variable to color. The observations highest category will be displayed in green, and the observations in the lowest category would be displayed in red. The remaining observation are displayed in black.
Legend_title	Title of the legend, which should include relevant information about the colors and the scale of the variable which is used to color the map.
Filter_dic	Dictionary which can contain options to only include a subset of the observation in the dataframe in the map. If there is chosen the include all the observations an empty dictionary ({}) should be provided as input. With respect to filtering/using a subset of the dataframe the following keys are expected to contain information: ‘filter_variable’ : variable used to filter, ‘filter_sign’ : possible values are: ‘higher’, ‘lower’, ‘equal to’, and ‘not equal to’. ‘ ‘filter_value’: value/threshold to filter on
Variable_type_dic	Dictionary, which contains information on the variable types used in the pricing model. This dictionary is provided in the shared GitHub file with the same name as the parameter.
Parameters	**Description**

[4] https://github.com/NielsUPF/Demonstrative_MAP

Ref_group_dic	Dictionary, which contains information on the reference category for dummy variable. This dictionary is provided in the shared Github with the same name as the parameter.
N_clusters	Number of clusters in which will be used to cluster the properties. Only relevant if the parameters: cluster = True
Lat_col	The column in the dataframe that contains information of the latitude of the properties.
Long_col	The column in the dataframe that contains information of the longitude of the properties.
Show_all (True/False)	If specified true all the observations are shown in the map, If specified false only the observations in the highest and lowest category are specified in the map
Cluster (True/False)	If specified true support vector machine will be used to clusters the residential properties based on location. If specified false all observations will be shown individually

Table 11 includes the name and a description of the parameters included in the function to build a map to visualize the results of the pricing models.

The first demonstrative map will show the absolute price impact of sustainability when the datasets of the properties are clustered based on location by the support vector machine algorithm with a prespecified set number of clusters of 100. Hereby the clusters which are in the highest absolute sustainability impact, selecting all the sustainable features in the model, will be displayed in green, the 10% lowest will be displayed in red, whereas the clusters in the 10% – 90% quantile of the absolute sustainability price impact will be displayed in black. The parameters used to construct demonstrative map 1 are included in table 12 below.

The second demonstrative map discussed in this research will cover the absolute price impact of selected sustainable features by clustering the houses into 100 clusters. The selected sustainable features will exclude the sustainable features that have a high negative relationship between a higher sustainable score and the housing price and cannot be influenced by political measures/interventions. The excluded variables are Distance to the beach (km), Neighborhood size 10 (ha), and Income Distribution PCA. The applied color scale in the second map is equal to those applied for the first map. The set parameters to construct demonstrative map 2 are included in table 12 below.

Table 12: Parameters to construct the demonstrative maps

Parameters	**Demonstrative Map 1**	**Demonstrative Map 2**
Selected Sustainable Features	Model_7_predictors_order	[e for e in model_7_predictors_order if e not in ['Distance to Beach (km)','Neighboorhood size (10 ha)','Income Distribution PCA']]
Map_save_name	'Demonstrative_map_1'	'Demonstrative_map_2'
Title	'Absolute Price Impact Sustainability Variables Model 7 on Residential Properties'	'Absolute Price Impact Selected Sustainability Variables Model 7 on Residential Properties'
Subtitle	"Heckman Selection Model Barcelona"	"Heckman Selection Model Barcelona"
Legend_title	'Quantile Absolute Price Impact Sustainable Variables'	'Quantile Absolute Price Impact Selected Sustainable Variables'
Circle_Multiplier	15	15
DF	df_ols	df_ols
Model_result	SL_ols_model_7_result	SL_ols_model_7_result
Color_var	"Sustainable Features Price Impact"	"Selected Sustainable Features Price Impact"
N_color_cat	10	10
Model_predictors	Model_7_predictors_order	Model_7_predictors_order
Filter_dic	{}	{}
Parameters	**Demonstrative Map 1**	**Demonstrative Map 2**

Variable_type_dic	Variable_type_predictors (specified in the notebook)	Variable_type_predictors (specified in the notebook)
Ref_group_dic	Ref_group_dic (specified in the notebook)	Ref_group_dic (specified in the notebook)
N_clusters	100	100
Lat_col	‘latitude’	‘latitude’
Long_col	‘longitude’	‘longitude’
Show_all (True/False)	True	True
Cluster (True/False)	True	True

Table 12 includes the parameter values to construct demonstrative map 1 and demonstrative map 2

Section 4: Results

The results section consists of both a section to discuss the findings of the valuation models (section 4.1) and a section that discusses the visualization of the valuation models by the demonstrative maps (section 4.2). Each section consists of multiple subsections related to the discussed valuation model or demonstrative map. Additionally includes section 4.1 a subsection for the discussion of the findings of the Probit model to predict the probability of the presence of an energy label in the housing advertisement.

Section 4.1: Results of the pricing models

Table 13A includes the results for the Probit model in the Heckman selection model to predict the probability that the energy consumption label is present and the results of the valuation models 1-3. The results for valuation models 4-7 are included in table 13B. The results of the models are discussed in increasing order.

Section 4.1.1: Probit Model Heckman:

The results for the Probit selection model are included in Table 13A. The results for the VIF statistics, included in table 14 in the appendix, show that only the air conditioning dummy variables have a high test statistic (12.90), which is likely caused by the high presence of the dummy variable (98%). The model is correctly specified as shown by the rejection of the Ramsey reset (p-value = 0.24). Overall has the estimated model a low pseudo R-squared (2%), showing that the included variables only have a small predictive power when explaining the presence of an energy consumption label. This indicates that the sample selection problem is smaller than in prior work, which is supported by the relatively low number of observations that are missing energy labels (25%) in comparison to earlier work in Barcelona, which is supported by the longer time period that has passed by since the obligation to report the energy label (Chen & Marmolejo Duarte, 2018; Dell’Anna et al., 2019; Marmolejo-Duarte & Chen, 2022).

The results for the model show that both the building surface m² and building age have a positive relationship with the probability of the presence of an energy label in the housing advertisement, although this relationship is not significant. The state of the property has a significant effect on the probability that the energy label is shown in the housing advertisement. Both properties that need renovation or are newly constructed have a significantly higher probability of missing an energy label in the housing advertisement. The high and significant new housing development dummy is likely to be caused by the fact that in 60% of the housing advertisement of the observations, in which the dummy variable is equal to 1, the energy label is not reported. Regarding the dummy variable that needs renovation, the significant negative coefficients is likely caused by an excepted low energy label by the current condition of the property for instance by the presence of thinner glass in the windows. This could incentivize home owners/real estate agents to not report the energy label even though it is obligated. Relatively to most dummy variables for the facilities of a property observed relationships in contrast to the expectations. The results show a significant positive relationship between the presence of a heating system and air conditioning and the presence of an energy label. This shows that the presence of energy-consumptive facilities increases the probability that home owners measure/report the energy label. Concerning the other facilities only evidence is found that residential properties with a terrace are less likely to report the energy label. The other included dummy variables, elevator, and outdoor facilities are insignificant when predicting the likelihood of a missing energy label. Altogether are the results different compared to those found in the work of Chen et al. (2018), who found no significant coefficients for the building surface m², terrace, and property state dummy variables, and a significant negative relationship between the outdoor facilities. More related to the energy-consuming features they found a negative significant relationship for the heating dummy variable and no relationship for the air conditioning dummy variable on the probability of the presence of the energy labels. However, Chen et al. (2018) included overall more features in the model, but the adding additional variables do not increase the explained variance by the model. Furthermore is not shown or cannot be reasoned that the other housing-specific variables have a high correlation with the presence of the energy label in the housing-advertisement.

Section 4.1.2: Model 1: Housing-Specific Variables

Model 1, shown in table 13A below, includes only the housing-specific characteristics in the semi-log hedonic pricing model. The R-squared of the model (82.6%) shows that these variables have high importance when valuing residential properties. The results for the VIF test, included in the appendix, show that for none of the variables the test-statistic, expect the buildings' surface m² and building age caused by the inclusion of the squares of the variables, is higher than 6.5. The air conditioning dummy variable is excluded from the sample given the observed VIF statistic, which was higher than 65. The inclusion of the squares of building surface m² and building age decreased the probability of the Ramsey reset test to a p-value of 0.009, which only rejects a correct specification of the model at the 1% significance level. However, without the exclusion of the squared variables, the Ramsey reset would be rejected with a higher probability.

The findings for the building-specific characteristics are in line with expectations. The building's surface m² has a positive correlation with the property price indicated by the significant positive coefficient. However, the strength of this relationship decreases when the building surface becomes higher, as shown by the significant coefficient for the squared term. The negative significant coefficient for the building age indicates that the price of a property becomes lower when the age of the property increases, likely to be caused by the higher probability of some outdated facilities/characteristics that could be found in the property and should be treated in the short-term. However, the significant coefficient for the squared of the building age, shows that this effect decreases if the property is older, making it more likely to carry historical value. Evidence for the pricing of historical value in the residential market is for example found by Debrezaion et al. (2011) and Lazrak et al. (2014). Furthermore shows the results that a significantly higher price is paid for a house (15.9%) compared to an apartment on the ground floor. Additionally, the dummy variables for the floor level show that the willingness to pay increases once the apartment is located on a higher floor, since the significant implied premium compared to the ground floor for the mezzanine (4.5%), 1st floor (10.9%) 2nd -5th floor (15.5%), and 6th floor or higher (21.4%). This showing increasing pattern is in line with earlier research about the housing market in Barcelona of (Dell'Anna et al., 2019; Marmolejo-Duarte & Chen, 2022). Dummy variables for the 2nd – 5th floor could be combined into one category because an F-test rejects that the individual coefficients for the floors are significantly different from each other. The dummy variables for the 6th floor or higher are combined into one category given the low presence of the individual dummy variables relative to the sample size, decreasing the probability that the model will overfit on dummy variables with low presence. Subject to the property state, no evidence is provided for a significantly higher willingness to pay for a residential property when it is a newly developed property compared to when it is in good condition. However, when a property needs to be renovated the results of the model imply that the is a significant discount (-17.2%), which is likely more or less equal to the average costs that are expected for the renovation of the property. The presence of facilities has a significant effect on the asking price for a property. From high to low the price premiums in the asking price are 20.3%, 11.7%, 9%, 5.9%, and -4% for respectively the elevator, heating, parking space included, terrace, and outdoor facilities dummy variables. The significant negative coefficient for the outdoor facilities coefficients in contrast to the research on the Barcelona housing market of (Chen & Marmolejo Duarte, 2018; Dell'Anna et al., 2019; Marmolejo-Duarte & Chen, 2022), which all found a significant coefficient for the swimming pool dummy variable. The outdoor facilities variable correlates 94% with the swimming pool variable. The contrary findings are likely the reason for the current dryness in Barcelona, which resulted in policies to prevent high water consumption by inhabitants such as the prohibition to fill the swimming pools, which were not implemented/relevant at the time of earlier research (Ajuntament de Barcelona, 2023b). Another reason could be the higher cost of living for, maintaining a garden, green area, and/or swimming pool, which exceeds the expected utility of these outdoor facilities.

The energy consumption label dummy variables (ecological) are the only housing-specific variable, which are part of a sustainable dimension (Kauko, 2019). The results show an increasing premium if the energy label increases in line with the findings of earlier research. However, the trend is not perfectly increasing. It reflects a higher willingness to pay for properties with higher energy labels (A, B, C) compared to properties with lower energy labels (E, F, G). The highest premium compared to a property with energy label D is found for energy label B (13.1%). The highest discount compared to a property with energy label D is found for the lowest energy label G (-4%). The increasing trend for the energy label dummy variables is in line with the reported findings of earlier research about Barcelona (Chen & Marmolejo Duarte, 2018; Dell'Anna et al., 2019; Marmolejo-Duarte & Chen, 2022). Additionally the IMR ratio is significantly and negative, indicating that the price of a property is negatively related with the probability that a energy label is shown. This is in line with the expectations of the sample selection bias by showing that energy labels are not reported in housing prices for a reason and results of earlier research on the sample selection bias in Barcelona (Chen & Marmolejo Duarte, 2018).

Section 4.1:2: Model 2: Ecological Dimension of Sustainability

Model 2, included in table 13A, includes the variables related to the ecological dimensions of sustainability, representing the accessibility to short and long commuting options. As proposed by Kauko (2019) the energy labels would also belong to the ecological dimension. However, they are included in housing-specific instead of sustainable pricing factors since they are not influential by political decisions on the neighborhood of a residential property. The energy efficiency of a property could be stimulated by political intervention/grants, but these policies will like be equal for/affecting every property in the municipality of Barcelona. Hereby it would not directly decreases the difference in the pricing of sustainability between areas in Barcelona. The inclusion of the ecologically sustainable variables increases the R-squared by 0.8 percentage points compared to the pricing model that only includes housing-specific variables (model 1) to 83.4%. The inclusion of ecological variables is not likely to introduce multicollinearity in the model, shown by no found new high VIF-test statistics as shown in table 14 in the appendix. However, still, the Ramsey reset test for misspecification cannot be rejected by model 2 (p-value = 0.008). The coefficients for the housing-specific variables are in line with the findings in model 1.

The results for the ecological-related variables show that the bus & metro PCA component on housing prices is insignificant. This provides no evidence for a relationship between the accessibility of short-distance commuting options and housing prices. The insignificance is likely caused by the high accessibility for residential properties in general in Barcelona. For example in the sample for a property, the highest observed minimum distance to a bus stop is only 700 meters and the average number of bus stops within 250 meters is 9. Moreover, for the metro station, is respectively 3 kilometers and 0.5. These findings for the short commuting distance are in-line with the work of Graells-Garrido et al. (2021) in Barcelona, who also found an insignificant correlation between the housing rents and accessibility to the bus, metro and bike pike-up places was found. However, this it without correcting for the price effect of other variables, making the findings potentially subject to an omitted variable bias. Concerning the accessibility to long-distance commuting options, the observed coefficient for the minimum distance to the highway or train station is significant and negative. These findings contrast earlier work about Barcelona, where Dell'Anna et al. (2019) showed a significantly positive price effect for the distance to the highway. This could be caused that in this research it is combined with the distance to the train and the absolute distance to only the highway is used. In conclusion the results for the significance positive relationship between closeness to the highway/train and housing prices no reason to reject hypothesis 1, an increase in the environmental dimension of sustainability results in higher housing prices. Although, the relationship between the access to and presence of bus and metro stations is insignificant this does not provide evidence to reject the hypothesis.

Section 3.1.3: Model 3: Environmental Dimension of Sustainability

Model 3, shown in table 13 A, includes the variables related to the environmental dimensions of sustainability, which is captured by the distance to the beach, PCA components for accessibility to parks & gardens and viewpoints, size of the neighborhood, and the vulnerability to heat impact. The inclusion of the environmental variables increases the R-squared of the pricing model with only housing-specific variables by 2.5 percentage points to 85.1% compared to the pricing model that only includes housing-specific variables (model 1). The VIF test statistics show that the environmental variables distance to the beach and vulnerability to heat impact have high values caused by the correlation with the district dummies in the model. However, the standard errors of the coefficients are still relatively low, decreasing the likelihood that the observed coefficients are heavily influenced by multicollinearity. Misspecification of the model is rejected by the Ramsey reset test for misspecification rejected at the 1% significance level by the inclusion of the environmental variables (p-value: 0.036). The coefficients for the housing-specific variables are in-line with the earlier reported findings of models 1 and 2.

Subject to the environmental variables, the distance to the beach in kilometers has a significant negative relationship with the housing price in-line with earlier research, which found evidence that housing prices are negatively correlated with the distance to the seashore in Barcelona (Dell'Anna et al., 2019; Marmolejo-Duarte & Chen, 2022). This shows that accessibility/closeness to the beach is valued in housing advertisements. In contrast, a significant negative relationship between is between accessibility to green space and viewpoints and the housing price is found by the PCA components for park & garden and viewpoints. This is a contradiction to earlier founding in other cities (Brandt et al., 2014; Cui et al., 2018; Kim & Kim, 2020; Park et al., 2017). However, it is in line with earlier reported findings for Barcelona, which argue that it is caused by the structure of the city in which parks are chosen to be located at the periphery (Dell'Anna et al., 2019). This does not show that the access to park & garden and viewpoint is not an important sustainable factor in Barcelona since Triguero-Mas et al. (2015) found in Catalonia that it has a significant impact on the mental health of inhabitants. Additionally, show the result of a significant relationship between the neighborhood size and the housing price, a preference for housing in the smaller areas in Barcelona, where there are fewer places to build. The same negative relationship is found for the vulnerability to a heat impact, where residential in properties an area that is more vulnerable to a heat impact are listed for a lower asking price. The asking price for a residential property decreases by 10.2% for each level that the vulnerability to heat impact increases. This price effect of the size of the neighborhood and the vulnerability to a heat impact is not earlier addressed in research about housing prices in Barcelona. Altogether, do the significant negative impact of a higher distance to the beach, neighborhood size, higher vulnerability to a heat impact not provide evidence to reject hypothesis 2, An increase in the environmental dimension of sustainability results in higher housing prices. There is found a significant negative relationship between housing prices and access to and presence of park & gardens and viewpoints, but this is mainly caused by the structure of the city. The access to and presence of park & gardens and viewpoints still has a positive effect on mental of inhabitants of Barcelona (Triguero-Mas et al., 2015; Vidal Yañez et al., 2023).

Table 13A: Semi-Log Pricing Model Results of the Sample Including Observations with Missing Energy Label Data

	PROBIT		Model 1		Model 2		Model 3	
	Coef.	Stderr.	Coef.	Stderr.	Coef.	Stderr.	Coef.	Stderr.
Constant	0.352***	0.09	12.202**	0.12	11.404***	0.12	13.073***	0.06
Building surface m²	$2 * 10^{-4}$	0.00	0.016***	0.00	0.016***	0.00	0.016***	0.00
sq(Building surface m²)			$-3 * 10^{5}$***	0.00	$-3 * 10^{5}$***	0.00	$-3 * 10^{5}$***	0.00
Building age	$1 * 10^{-4}$	0.00	-0.006***	0.00	-0.006***	0.00	-0.005***	0.00
sq(Building age)			$4 * 10^{5}$***	0.00	$4 * 10^{5}$***	0.00	$4 * 10^{5}$***	0.00
House			0.159***	0.02	0.175**	0.02	0.209***	0.02
Mezzanine			0.045***	0.01	0.048***	0.01	0.046***	0.01
1st Floor			0.110***	0.01	0.113***	0.01	0.112***	0.01
2^{nd} -5^{th} Floor			0.156***	0.01	0.157***	0.01	0.157***	0.01
6th Floor or higher			0.214***	0.01	0.212***	0.01	0.208***	0.01
New housing development	-0.759***	0.06	0.029	0.06	0.034	0.05	0.049	0.05
Needs renovation	-0.116***	0.03	-0.172***	0.01	-0.174***	0.01	-0.170***	0.01
Elevator	0.038	0.03	0.203***	0.01	0.182***	0.01	0.172***	0.01
Terrace	-0.045*	0.03	0.059***	0.01	0.062***	0.01	0.066***	0.01
Heating	0.160***	0.03	0.117***	0.01	0.097***	0.01	0.111***	0.01
Outdoor facilities	0.022	0.03	-0.040***	0.01	-0.038***	0.01	-0.037***	0.01
Air conditioning	0.267***	0.08						
Parking space included			0.090***	0.01	0.098***	0.01	0.090***	0.01
Energy label consumption A			0.004	0.02	0.006	0.02	0.071***	0.02
Energy label consumption B			0.131***	0.01	0.124***	0.01	0.126***	0.01
Energy label consumption C			0.068***	0.01	0.070***	0.01	0.063***	0.01
Energy label consumption E			-0.006	0.01	-0.008	0.01	-0.002	0.01
Energy label consumption F			-0.025**	0.01	-0.024**	0.01	-0.021**	0.01

	PROBIT		Model 1		Model 2		Model 3	PROBIT
	Coef.	Stderr.	Coef.	Stderr.	Coef.	Stderr.	Coef.	Stderr.
Energy label consumption G			-0.040***	0.01	-0.034***	0.01	-0.027***	0.01
Bus & metro PCA					-0.001	0.01		
Distance to Highway/Train (km)					-0.105***	0.00		
Distance to Beach (km)							-0.104***	0.00
Park & garden PCA							-0.011*	0.01
Viewpoint PCA							-0.133***	0.01
Neighborhood size (10 ha)							-0.002***	0.00
Vulnerable to heat impact (1-5)							-0.102***	0.00
District Eixample			0.000	0.01	-0.057***	0.01	-0.166***	0.01
District Ciutat Vella			-0.093***	0.01	-0.145***	0.01	-0.341***	0.02
District Sant Martí			-0.218***	0.01	-0.228***	0.01	-0.497***	0.02
District Sants-Montjuïc			-0.276***	0.01	-0.334***	0.01	-0.234***	0.01
District Horta Guinardó			-0.347***	0.01	-0.296***	0.01	-0.281***	0.01
District Gràcia			-0.017	0.01	-0.024*	0.01	-0.077***	0.01
District Nou Barris			-0.528***	0.01	-0.461***	0.01	-0.391***	0.01
District Sarrià-Sant Gervasi			0.094***	0.01	0.014	0.01	0.081***	0.01
District Sant Andreu			-0.419***	0.01	-0.437***	0.01	-0.452***	0.01
IMR			-0.490***	0.10	-0.484***	0.09	-0.491***	0.09
R-squared	0.020		0.826		0.834		0.851	

Table 13 includes the results for semi-log hedonic priding models of the sample including observations with missing energy label data. The results report both the coefficients (Coef.) and the standard errors (Stderr.) of the variables. *** denotes a significance at 1% level, *** denotes a significance at the 5% level, and * denotes significance at the 10% level.

Section 3.1.4: Model 4: Social Dimension of Sustainability

Model 4, included in table in 13b shown below, includes the variables related to the social dimension of sustainability consisting of for instance: closeness to the police stations, demographics score of the subdistrict, and the accessibility to public services and amenities. The inclusion of the social-related sustainable variables increases the R-squared by 2 percentage points to 84.6% compared to a pricing model which only includes housing-specific variables (model 1). The inclusion of the social-related sustainable variables introduces, except the neighborhood density variable caused by correlation with the subdistrict dummy variables, no variables with problematic VIF statistics. The higher observed VIF statistics for most social-related variables in contrast to the variables of other earlier dimensions of sustainable is mainly caused by correlation in the distribution of the amenities. For example, bars and restaurants are in practice often located close to big shopping places. Misspecification of the model is rejected in the Ramsey reset test (p-value: 0.56). The housing-specific variables are mostly in with the findings of the earlier models. The only difference is the change of the energy consumption label A dummy variable to negative and insignificant. This is likely caused by the presence of outliers in combination with the low presence of the dummy variable in the sample (2.1%).

Most of the social-related sustainable coefficients have a significant relationship with housing prices. Only the healthcare variables, the pharmacy, and hospital & clinique PCA components have insignificant coefficients. This is in line with the findings of Taltavull de La Paz et al. (2019) for Alicante, Spain. But in contrast to the findings of Barcelona by Graells-Garrido et al. (2021), although they only studied the spearman correlation making the findings possibly subject to an omitted variable bias. The significant positive coefficient for the police PCA component is in line with findings for the negative price effect of crime perception rate and the positive price effect of perceived security in Barcelona by Buonanno et al (2013). Although the police PCA components might not perfectly capture the safety of a neighborhood other relevant data is not provided by the council of Barcelona. The bar & restaurant PCA components have the largest positive coefficients of the social PCA component, showing that the presence of social connection places is relatively highly valued in the housing prices of Barcelona. This significant positive correlation was also found by Graells-Garrido et al. (2021). Mixed results are found for the educational PCA components, where a significant negative relationship between the housing price is found for the secondary and lower education PCA component and a positive relationship is found for the university PCA component. Earlier work on the accessibility to education only reported significant positive price effects in Barcelona, however never a distinction is made to the different types of education (Graells-Garrido et al., 2021). Related to the accessibility of big shopping places a significance positive price effect is found on the housing prices, in line with the work of Taltavull de La Paz et al. (2019) for Alicante and the work of Graells-Garrido et al. (2021) for Barcelona.

Specifically to the demographic social sustainable variables significant negative effects are found for the social cohesion, natural population growth, and density of the residential area on the willingness to pay for housing. The social cohesions variable is not included in earlier research, however, these findings might indicate that inhabitants of Barcelona like to live in neighborhoods where it is relatively less crowded as reflected by a lower degree of community activities/interaction. The preference for less crowded areas also explains the negative relationship between housing prices and the density of residential areas. The reported results of the price effect of neighborhood density in prior literature are mixed (Cajias et al., 2016; de Ayala et al., 2016; Eichholtz et al., 2013; Lazrak et al., 2014). That a higher natural population growth causes lower housing prices could be argued by Jeanty et al. (2010) that after a high population growth in an area, people tend to move to other areas possibly creating negative pressure on housing prices. However, in contrast to the natural population growth, the relationship between the net immigration rate and housing prices is significantly and positive in-line with the significant positive correlation found by Graells-Garrido et al. (2021) in Barcelona. Overall is found that higher accessibility to public services/amenities causes higher housing and that a better distribution of and growth of the population Barcelona results in higher housing prices. This provides no evidence to reject hypothesis 3, an increase in the social dimension of sustainability causes results in higher housing prices. However, more mixed results are found in contrast to the results for the above discussed dimensions with the significant negative price effect of a higher access and presence of second and lower education, higher social cohesions score, and net immigration rate, which reflect more unsustainable preferences in the willingness to pay for housing in Barcelona.

Section 3.1.5: Model 5: Cultural Dimension of Sustainability

Model 5, as included in table 13b shown below, includes the variables related to the cultural dimension of sustainability: accessibility to performing arts, religious recognition places, and libraries, museums & cultural points of interest. The inclusion of the culturally related variables increases the R-squared by 0.9 percentage points to 83.5% compared to the model that only includes housing-specific variables (model 1). The VIF-tests, included in table 14 in the appendix, show that no new high test-statistics are introduced. Furthermore is a misspecification of the model rejected by the Ramsey reset test (p-value = 0.17). The reported results for the housing-specific variables are in line with the findings of the earlier models. The only difference is for the dummy energy consumption label A, even as for model 4, with an insignificant positive price effect.

For the cultural-related sustainable variable, both positive and negative relationship with the willingness to pay for housing are found. The coefficient of the PCA components of performing arts is significant and negative implying that the accessibility to places of cultural expression by cinemas, theatre, and concert provide negative utility to home owners. This finding is in contradiction to the significant positive correlation with housing rents in Barcelona found by Graells-Garrido et al. (2021). Although Graells-Garrido et al. (2021) used a more general definition for entertainment paces. The negative findings in this research might be related to possible nuisance, measured by traffic and noise, that is caused during big events, which was for example found as a negatively pricing factor in housing prices in the Netherlands (Ossokina & Verweij, 2015). In contradiction, the results show that the closeness to and number of religious institutions and libraries, museums & cultural POI is significantly valued in housing prices of Barcelona. The findings for the religious institutions are in-line with the significant positive correlation with housing rents in Barcelona found by Graells-Garrido et al. (2021). The findings for the positive valuation of the accessibility/closeness of cultural places and points of interest in line with the provided evidence in the work by Lazrak et al. (2014) in the Netherlands finding a spillover effect of monumental buildings on housing prices. To sum up, the positive price effect on housing of the cultural sustainable variables, except the performing arts PCA components given the negative externalities of events hypothesis 4, an increase in the cultural dimension of sustainability causes results in higher housing prices, cannot be rejected.

Section 4.1.6: Model 6: Economic-Financial Dimension of Sustainability

Model 6, as shown in table 13b below, includes the variables related to the economic-financial dimension of sustainability by including variables for the income distribution and welfare of the subdistricts. The inclusion of these variables increases the R-squared by 2.2 percentage points to 84.8% compared to the pricing model that only included housing-specific variables. Table 14 in the appendix shows that the economic/financial variables have high VIF-test statistic, mainly caused by the correlation between welfare and income distribution variable and the correlation of these variables with the district dummies. However, the standard errors of the economic-financial variables remain low in comparison to the coefficient, implying only a limited impact of multicollinearity in the model. The Ramsey reset test for misspecification is not rejected for the model (p-value =0.000) due to the non-linearity which is often measured in income distribution and welfare statistics, although this is captured by the inclusion of the squared variables in the construction of the PCA components. The results for the housing-specific variable are in-line with the results of earlier models. Only in agreement with models 4 and 5 is the coefficient of energy consumption label A is insignificant in the pricing model.

In specific the economic-financial related sustainable variables are the results in line with prior literature. The coefficient for the income distribution PCA components, capturing the degree of income equality, shows a significant negative relationship between income equality and housing prices. The relationship implies that home buyers who can afford more expensive houses tend to cluster in certain subdistricts in Barcelona increasing the p80/20 income distribution and Gini index. These results are in line with the findings Chen et al. (2018) and Chen et al. (2022) in Barcelona, which also found that respectively the cumulative of people in higher socioeconomic classes and high occupational positions had a significant positive effect on housing prices. In contradiction, a significant positive relationship is found between the welfare of the area and housing prices shown by the significant income & unemployment PCA component. This is in-line with reported findings in other countries by Mandell et al. (2011), Eicholtz et al. (2013), Cajias et al. (2016), and in Alicante, Spain by Taltavull de La Paz et al. (2019). Based on these findings hypothesis 5, an increase in welfare results in higher and an increase in income equality results in lower housing prices for the economic-financial dimension of sustainability, is not rejected.

Section 4.1.7: Model 7: Every Dimension of Sustainability

Model 7, shown in table 13b shown below, includes every dimension of sustainability: ecological, environmental, social, cultural, and economic-financial in the pricing model. The inclusion of every dimension increases which included only the housing-specific variables by 4.3 percentage points to 86.9% compared to the pricing model that includes only housing-specific variables (model 1). The results for the VIF-statistics, included in table 14 in the appendix, show that there are an additional number of variables introduced with high VIF test-statistics. This is mainly the result of the inclusion of a high number of predictors with all a relation to some local/area characteristics. This makes it is important, when interpreting the coefficients, to also check the findings for the sustainable variables in the other models and the correlation matrix for the identification of possible multicollinearity. The findings for the housing-specific variables are in line with those of the other discussed models. In contrast to the findings for models 4, 5, and 6 is the coefficient for the energy consumption label A again significant in model 7. The non-significance of the energy consumption label A in model 4, 5, and 6 in contrast to other models and earlier research in Barcelona suggest that the insignificance is caused by the low presence of the dummy variable.

The findings for the sustainable variable will for each dimension of sustainability can be compared with the findings in the individual models. Firstly, for the ecological dimensions of sustainability, the reported findings are different compared to model 2. The bus & metro PCA components is in model 7 significant showing evidence for a negative relationship between higher accessibility to those short-distance commuting options and the willingness to pay for housing. The findings for the minimum distance to the highway/train are in agreement with the results model 2 with a significant and negative relationship, but the strength of the relationship has decreased. This strength was likely overestimated in model 2 by not including variables related to the other sustainable dimensions.
Secondly, related to the environmental dimension of sustainability findings are mostly in line with model 3. The only difference is the insignificance of the park & garden PCA component, which was significant at the 10% level in model 3, and has by the inclusion of more sustainable variables the strength of the price effect of the viewpoint PCA component and vulnerability of heat impact level has almost halved. In contrast, the strength of the price effect of the distance to the beach and neighborhood size has remained constant.
Thirdly, concerning the social dimension of sustainability are most of the results different than those reported for model 4. The strength of the price impact for most of the socially sustainable variables has become less or even insignificant. These findings show a high probability of an omitted variable bias in model 4. The social-related variables for which the significant relationship with the housing price has disappeared are the police PCA component, the secondary & lower education PCA component, and the net immigration rate. In contrast significance, at the 10% level, is reported for the pharmacy PCA component compared to model 4, providing evidence for a positive correlation with the housing price. Related to the other social variables the significance of the coefficients remained the same. But in terms of the strength implied willingness to pay for housing at least halved for the bar & restaurant PCA component, big shopping place PCA component, social cohesion, natural population growth, and residential population density compared to the findings of model 4. The only social variable of which the strength of the price effect has not been halved compared to the model is the university PCA component.
Fourthly, regarding the cultural dimensions of sustainability, the results of model 7 show that the founded relationship between the housing price and the culturally sustainable variables is in line with model 5. However, the price effect of the religious institution PCA component has become insignificant and the strength of the other cultural sustainable variables, the performing arts and museum, library and POI culture PCA components, have at least halved likely caused by an omitted variable bias in model 5.
Fifth and lastly, the economic-financial sustainable variables also in model 7 is found a significant relationship with housing prices. However, while the standard errors of the variables remained constant, the strength of the relationships decreased with the inclusion of the other sustainable dimensions. This decrease in strength is approximately 200% for the income equality PCA

component, but only approximately 33% for the income & unemployment PCA component compared to model 6.

In conclusion, the one-dimensional sustainability and the pricing models that include all dimension shows that in both models sustainability has a significant price effect for all the five dimensions: ecological, environmental, social, cultural, and economic-financial. The change in the strength coefficients and the significance for some sustainable variables in model 7 compared to the other models shows that models related to only one sustainability dimension will likely result from an omitted variable bias related to other sustainable dimensions. However, likely, the high inclusion of the number of the sustainable variable makes the model to some degree subject to multicollinearity as shown by higher test-statistic for the VIF values as also reported in the VIF-test statistics. But the findings that the strength of the coefficients is mostly lower compared to the other models, show that the pricing model does at least not overestimate the price effect, making model 7 the most conservative model concerning the individual sustainable variables.

Table 13B: Semi-Log Pricing Model Results of the Sample Including Observations with Missing Energy Label Data

	Model 4		Model 5		Model 6		Model 7	
	Coef.	Stderr.	Coef.	Stderr.	Coef.	Stderr.	Coef.	Stderr.
Constant	12.129***	0.12	12.137***	0.11	12.09***	0.12	12.662***	0.07
Building surface m²	0.015***	0.00	0.016***	0.00	0.016***	0.00	0.015***	0.00
sq(Building surface m²)	$-3*10^{5}$***	0.00	$-3*10^{5}$***	0.00	$-3*10^{5}$***	0.00	$-3*10^{5}$***	0.00
Building age	-0.006***	0.00	-0.005***	0.00	-0.006***	0.00	-0.005***	0.00
sq(Building age)	$4*10^{5}$***	0.00	$3*10^{5}$***	0.00	$4*10^{5}$***	0.00	$3*10^{5}$***	0.00
House	0.212***	0.02	0.176***	0.02	0.2078***	0.02	0.270***	0.02
Mezzanine	0.041***	0.01	0.047***	0.01	0.045***	0.01	0.045***	0.01
1st Floor	0.108***	0.01	0.110***	0.01	0.104***	0.01	0.108***	0.01
2nd -5th Floor	0.148***	0.01	0.147***	0.01	0.151***	0.01	0.148***	0.01
6th Floor or higher	0.218***	0.01	0.213***	0.01	0.225***	0.01	0.215***	0.01
New housing development	0.053	0.05	0.046	0.05	0.080	0.05	0.087	0.04
Needs renovation	-0.166***	0.01	-0.165***	0.01	-0.163***	0.01	-0.161***	0.01
Elevator	0.189***	0.01	0.193***	0.01	0.187***	0.01	0.162***	0.01
Terrace	0.071***	0.01	0.065***	0.01	0.066***	0.01	0.076***	0.01
Heating	0.105***	0.01	0.113***	0.01	0.100***	0.01	0.099***	0.01
Outdoor facilities	-0.033***	0.01	-0.040***	0.01	-0.040***	0.01	-0.035***	0.00
Parking space included	0.095***	0.01	0.098**	0.01	0.044***	0.01	0.073***	0.01
Energy label consumption A	-0.016	0.02	0.003	0.02	0.029	0.02	0.078***	0.02
Energy label consumption B	0.119***	0.01	0.126***	0.01	0.117***	0.01	0.112***	0.01
Energy label consumption C	0.060***	0.01	0.059***	0.01	0.065***	0.01	0.058***	0.01
Energy label consumption E	-0.005	0.01	-0.006	0.01	-0.005	0.01	-0.001	0.01
Energy label consumption F	-0.027***	0.01	-0.030***	0.01	-0.023***	0.01	-0.024***	0.01
Energy label consumption G	-0.029***	0.01	-0.037***	0.01	-0.033***	0.01	-0.019***	0.01
Bus & metro PCA							-0.033***	0.01
Distance to Highway/Train (km)							-0.015**	0.01
Distance to Beach (km)							-0.090***	0.01
Park & garden PCA							0.007	0.00
Viewpoint PCA							-0.058***	0.01
Neighborhood size (10 ha)							-0.003***	0.01
Vulnerable to heat impact (1-5)							-0.057***	0.00
Police PCA	0.021***	0.01					-0.011	0.00
Bar & restaurant PCA	0.125***	0.01					0.065***	0.01
Secondary & lower educ. PCA	-0.016**	0.01					-0.017	0.01
University PCA	0.079***	0.01					0.056***	0.01
Pharmacy PCA	0.012	0.01					0.020*	0.01
Hospital & Clinique PCA	-0.016	0.01					-0.001	0.01
Big Shopping Place PCA	0.123***	0.01					0.045***	0.01
Social Cohesion Score	-0.054***	0.01					-0.020*	0.01
Natural population growth ‰	-0.018***	0.00					-0.008***	0.00
Net immigration rate ‰	0.004***	0.00					0.001	0.00
Density net (hab/1000/ha)	-0.169***	0.01					-0.051***	0.02

	Model 4		Model 5		Model 6		Model 7	
	Coef.	Stderr.	Coef.	Stderr.	Coef.	Stderr.	Coef.	
Performing Arts PCA			-0.082***	0.01			-0.035**	0.01
Religious Institution PCA			0.046***	0.01			-0.020	0.01
Museum, Library & POI Cult. PCA			0.387***	0.02			0.210***	0.02
Income Distribution PCA					-0.415***	0.03	-0.151***	0.04
Income & Unemployment PCA					0.673***	0.03	0.503***	0.04
District Eixample	-0.106***	0.01	0.013	0.01	0.120***	0.01	-0.095***	0.01
District Ciutat Vella	-0.222***	0.02	-0.182***	0.01	0.103***	0.02	-0.236***	0.02
District Sant Martí	-0.098***	0.01	-0.176***	0.01	0.055***	0.01	-0.242***	0.02
District Sants-Montjuïc	-0.162***	0.01	-0.259***	0.01	-0.008	0.01	-0.045***	0.02
District Horta Guinardó	-0.205***	0.01	-0.334***	0.01	-0.041***	0.01	-0.061***	0.01
District Gràcia	0.061***	0.01	-0.103***	0.01	0.192***	0.01	0.050***	0.01
District Nou Barris	-0.375***	0.01	-0.522***	0.01	-0.177***	0.02	-0.143***	0.02
District Sarrià-Sant Gervasi	0.129***	0.01	0.076***	0.01	-0.087***	0.01	-0.020*	0.01
District Sant Andreu	-0.248***	0.01	-0.456***	0.01	-0.091***	0.02	-0.214***	0.02
IMR	-0.456***	0.09	-0.460***	0.09	-0.403***	0.09	-0.433***	0.08
R-Squared	0.846		0.835		0.848		0.869	

Table 13 includes the results for semi-log hedonic priding models of the sample including observations with missing energy label data. The results report both the coefficients (Coef.) and the standard errors (Stderr.) of the variables. *** denotes a significance at 1% level, *** denotes a significance at the 5% level, and * denotes significance at the 10% level.

Section 4.2: Visualization of the Results by Demonstrative Maps

Section 4.2 discusses the results of the pricing models that are visualized in the demonstrative maps. Section 4.2.1 discusses the results that are visualized in demonstrative map 1. Section 4.2.2 discusses the results that are visualized in demonstrative map 2. The code to construct and of the demonstrative maps can be found on GitHub[5].

Section 4.2.1: Demonstrative map 1

https://github.com/NielsUPF/Demonstrative_MAP/blob/main/Demonstrative_map_1.html

Demonstrative map 1, of which a screenshot is shown in figure 5 below, includes the results of the sample with 100 clusters based on the location of the sample including the observations with missing energy label data using all the sustainable predictors as specified in model 7 to color the map. It is recommendable when reading the discussions of the visualizations of the map to open the map, which can be found in the GitHub repository link specified under the name of the subsection. The visualization by demonstrative map 1 shows that the cluster of houses with the 10% lowest absolute sustainable price impact by the predictors used in model 7 is located in the districts: Nou Barris, Horta Guinardo, and Sants Montuic. The 10% clusters of houses with the highest sustainable price impact are located in the districts: Eixample and Ciutat Vella.

Interaction with demonstrative map 1 shows that the low sustainable score for the houses with 10% lowest absolute price impact of sustainability is mainly caused by a high average distance to the beach, distance to the highway/train, income equality[6], and low average income. In comparison, all these variables are very high for the properties in the clusters with the highest absolute price impact of sustainability. Furthermore the observations in the clusters of houses in the highest category in general a higher accessibility to public amenities and services compared to other clusters. The accessibility scores for PCA components related to the environmental,

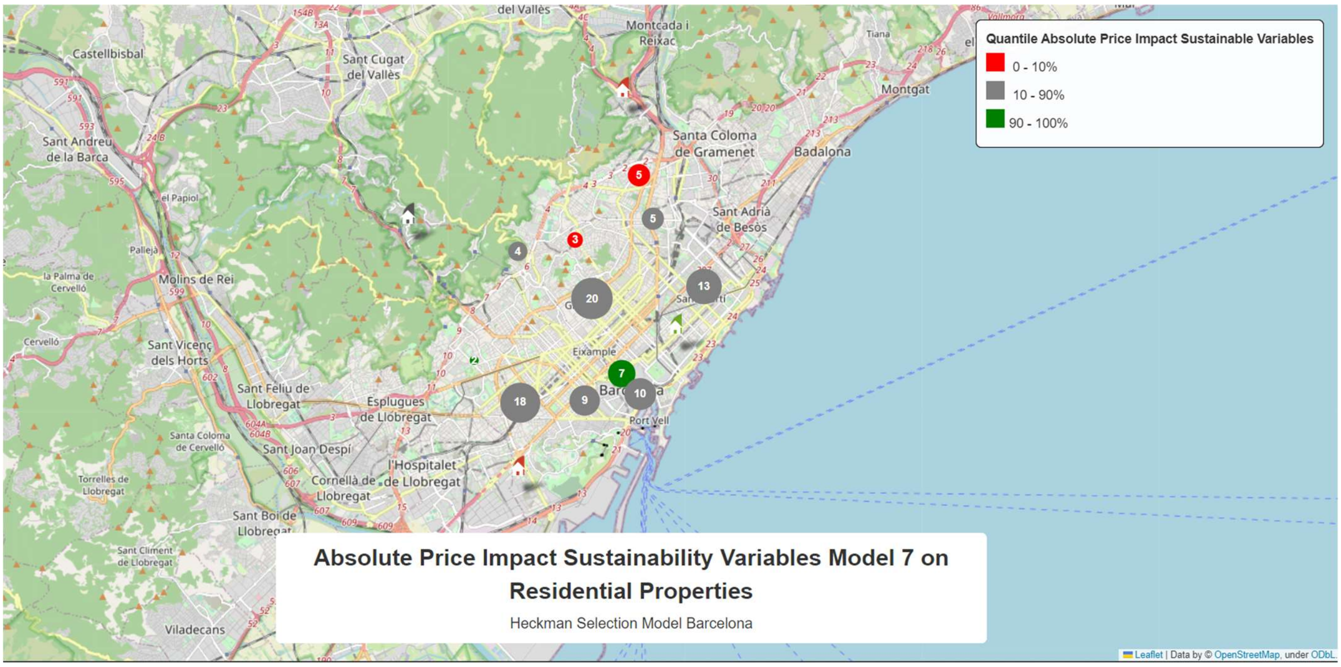
[5] The GitHub repository includes the notebook used to construct the demonstrative maps and the HTML codes of the demonstrative maps. The GitHub repository can be found on: https://github.com/NielsUPF/Demonstrative_MAP

[6] A higher income equality has a negative relationship with the housing price in model 7.

social, and cultural dimension of sustainability for the cluster of houses in the 10% quartile absolute price impact of sustainability have mostly a score that is higher than at least 80% of the other clusters. Although, also for some of the clusters within the lowest absolute price impact of sustainability relative to those sustainable are sometimes high for most of these clusters, at least 50% of the PCA components regarding the environmental, social, and cultural dimensions of sustainability are reported to have a score of lower than at least 20% of the observations.

In conclusion, shows the map clearly that sustainable pricing in Barcelona is bounded to specific areas/regions. The cluster of houses with high and low absolute impact of sustainability in the housing prices are located close to each other. This provides opportunities for political measures and interventions to make the pricing of sustainable factors fairer around the city. However, the difference in the pricing effect of the sustainable pricing factors in model 7 is mainly determined by the distance to the beach and income equality, which are hard to tackle through policy interventions. The distance to the beach is very location bounded and the relationship and the relationship between housing and more fair income equality is negative. However, additionally shows the results as well that the values for the other variables related to sustainability are often on higher-side for the clusters with housing observations in highest quantile of the absolute price impact of sustainability and in the lower-side by the lowest quantile with absolute price impact of sustainability.

Figure 5: Screenshot of Demonstrative Map 1 for the Sample Including Observations with Missing Energy Label Data



Section 4.2.2: Demonstrative Map 2
https://github.com/NielsUPF/Demonstrative_MAP/blob/main/Demonstrative_map_2.html

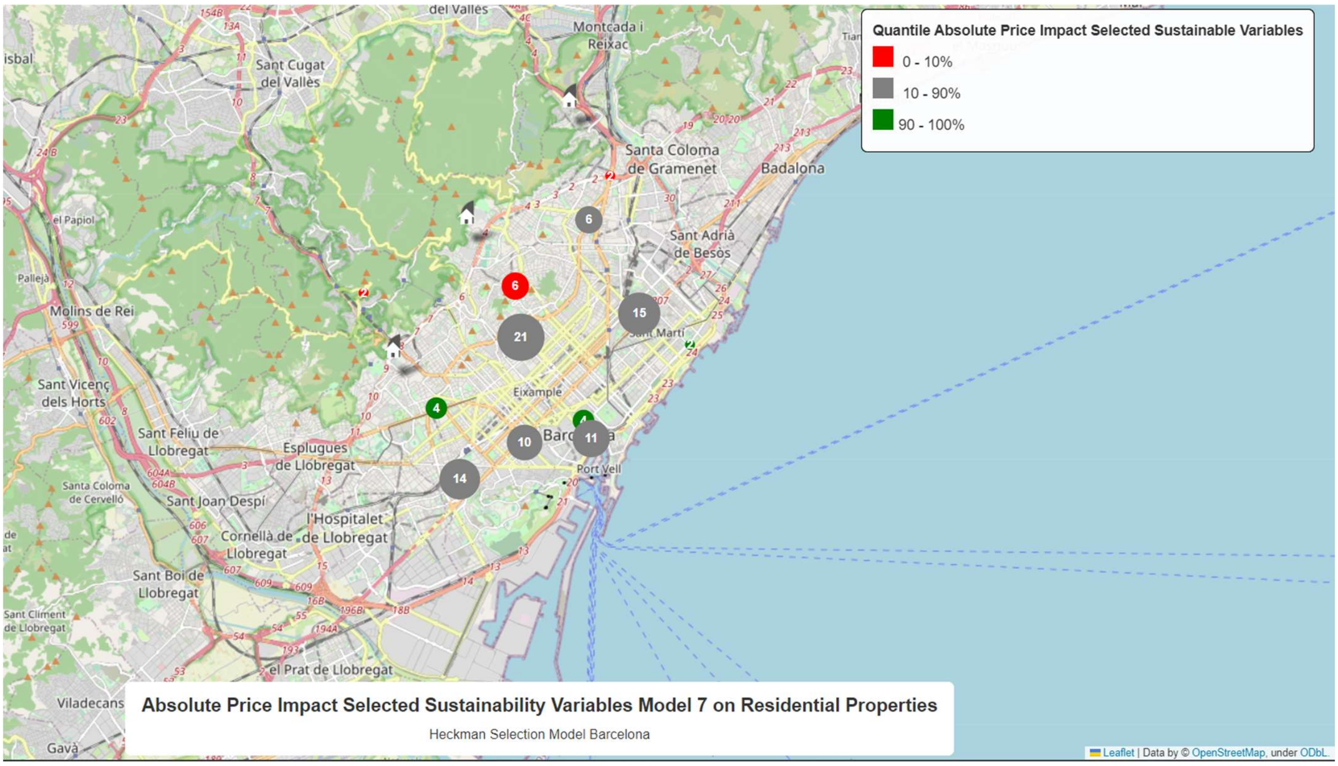
Demonstrative map 2, of which a screenshot is provided in figure 6 below, clusters the observations based on the location of the sample including the missing energy label data in 100 clusters using the absolute price impact of sustainable variables, excluding the distance to beach (km), neighborhood density (10 ha), income distribution PCA component, as estimated in model 7 to color the map. It is recommendable when reading the discussions of the visualizations of the map to open the map, which can be found in the GitHub repository link specified under the name of the subsection.

The screenshots show that the cluster of houses with the 10% lowest absolute selected sustainable variables price impact is located in other subdistricts compared to demonstrative map number 1. The cluster of houses is in demonstrative map 2 mainly located around the districts Gracia, Sarria Sant Gervasi, and Horta Guinardó. The distribution of the 10% clusters of houses with the highest absolute price impact for the selected sustainable variables also changed to the border between the district Sant Gervasi and Les Corts, but also remained present in the districts: Eixample and Ciutat Vella.

Interaction with demonstrative map 2 shows that the low total price impact of the selected sustainable variables in the district Sant Gervasi is mainly caused by low accessibility scores related to most of the variables for the ecological, environmental, social, and cultural dimensions of sustainability. The clusters have very high scores on the welfare aspect of sustainability, achieving at least a score for the income & unemployment PCA component which is higher than 90% of the observations. The low absolute price impact of the selected sustainable variable in the clusters located in the districts Gracia and Horta Guinardó is caused by multiple reasons. This reason is often in-specific related to the cluster of houses. Mostly caused by a combination of a bad score for the following sustainable variables: bus & metro PCA components, distance to the nearest highway/train, population density, cultural variables, and welfare. The high variety in the specific reason implies that policy intervention to increase the absolute price impact of sustainability in these areas should be implemented on a neighborhood scale. The high absolute impact of the selected sustainable variables for the clusters of properties around the border of Sant Gervasi and Les Corts is mainly caused by the relatively high welfare as measured by the value for the income & unemployment PCA component. Additionally, have the clusters not commonly shared sustainable scores for variables shared. They all have high values on different sustainable variables. The high absolute price impact of the selected sustainable variables of the clusters of properties in Eixample and Ciutat Vella is mainly caused by an overall high score for the selected sustainable variables. Especially related to the accessibility PCA components related to the environmental, social, and cultural dimensions of sustainability. Subject to the variable with the highest impact, only the price impact of the bar & restaurant PCA component is for each of these clusters found to have a share that is higher than 20% of the absolute price impact of the selected sustainable variables.

In conclusion shows demonstrative map 2 more variety in the pricing of the selected sustainable variable, by excluding the sustainable variables that are strictly location bounded and are found to significantly negatively correlate with higher fairness as estimated in pricing model 7. These results that more fair housing prices from the pricing of sustainability should be treated by policy interventions that are very locally implemented, specifically developed for the neighborhoods. The observations that most clusters of properties with a high/impact of sustainable factors in both demonstrative map 1 and map 2 provides no evidence to reject H6, the observed price impact of sustainability in Barcelona shows local clustering tendencies.

Figure 6: Screenshot of Demonstrative Map 2 for the Sample Including Observations with Missing Energy Label Data



Section 5: Robustness

The robustness section discusses the results of the semi-log hedonic pricing models for the residential properties in Barcelona for the sample that exclude observations with missing energy label data. The focus will hereby be on the differences compared to the sample that includes the observations with missing energy label data. The results for model 1 to 4 are included in table 15a in the appendix and the results for pricing model 5 to 7 are included in table in 15b the appendix. The VIF test-statistic for the variables included in the pricing models are included in table 16 in the appendix.

Section 5.1: Robustness Pricing Models

The findings for model 1, which only includes the housing-specific variables, are similar to the results found for the sample which includes the observations with missing energy labels. The differences are mainly related to the variables that were correlated/had predictive power for the presence of an energy label in the housing advertisement. The new housing development dummy variable has become higher (0.168) and significant. The coefficient of the heating dummy variable has halved. Related to the energy labels, the strength of the variables has been included in the sample that excludes observations with missing energy label data. This is in line with earlier work of Chen & Marmolejo Duarte (2018), in which the sample selection bias was addressed in

Barcelona. The findings for the housing-specific variables are also observed in the other pricing models (model 2-7). Additionally, in models 4, 5, and 6 the coefficient of the energy consumption label has become significant.

The findings for models 2 and 3, including respectively the ecological and environmental dimensions of sustainability, show no differences compared to the results for the sample that includes observations with missing energy label data. For model 4, which includes the social dimension of sustainability, only small differences are observed in comparison to the sample that includes observations with missing energy label data. No more evidence is provided for a relationship between policy PCA component and secondary & lower education PCA component and housing price. These variables were also insignificant when the other sustainability dimension were included in model 7 for the sample that includes the observations with missing energy label data. Besides that, small differences are found in the strength of the price effect of the socially sustainable variables. The results for model 5, including the cultural dimension of sustainability, the coefficient for the religious institution and museum, library & POI culture have decreased in strength in comparison to the results for the sample that includes observations with missing energy label data This could be caused by the exclusion of observations in a more cultural and historical neighborhood, in which there is a higher probability that a building has no energy label. The results of model 6, including the economic-financial dimensions of sustainability, are equal to those for the sample which includes the observations with missing energy labels in terms of strength and significance of the coefficients. Also, model 7, in which every dimension of sustainability is included, shows similar results to those reported for the sample including observation with missing energy label data suggesting a high degree of robustness. The strength of the coefficients is for all the sustainable variables almost similar. The only difference is reported by the insignificant coefficient of the pharmacy PCA component, but it is only significant at the 10% level in the sample including the observations with missing energy label data.

Section 5.2: Robustness Demonstrative Maps

The visualization of the results by demonstrative maps 1 and 2, of which a screenshot and link are included in figure 7 and 8 in the appendix, are identical to those found for the sample that included the observations with missing energy label data. The screenshots show that in both demonstrative maps, the clusters with the highest and lowest price impact of the selected sustainable variables are located in the same districts. Furthermore interacting with the demonstrative maps shows that the low/high price impact of sustainability is caused by the same sustainable variables as for demonstrative maps that visualize the results of the sample that includes observations with missing energy label data. The similarity in findings is caused by two reasons. First of all, are the properties that are missing energy labels distributed evenly over the city of Barcelona. Therefore the clusters include more or less the same (number of) observations. Secondly, are the observed price effects of the sustainable variables, almost equal to those for the sample that includes observations with missing energy label data, hereby also the total price impact is almost equal.

In conclusion, show the results for the sample excluding the observations that are missing energy labels robustness compared to the results for the sample including those observations. This applies specifically to the non-energy consumption-related variables, which do not correlate with the probability of the presence of an energy label in a housing advertisement. This suggests that observed willingness to pay for housing concerning the sustainable variable is likely representative of expected utility derived from the sustainable factors in Barcelona

Section 6: Conclusion and Discussion

This research investigates the price impact of sustainability on housing prices in Barcelona. A data-driven approach is applied in which sustainability is studied from five different dimensions: ecological, environmental, social, cultural, and economic-financial as proposed by Kauko (2019) The only difference compared to the proposed dimensions is the exclusion of the energy label, which is included in the housing-specific variables since it is the only sustainable factor that is property bounded instead of the location. Specific to the energy label evidence for a higher/lower willingness to pay for housing is found in line with earlier research in Barcelona (Chen & Marmolejo Duarte, 2018; Dell'Anna et al., 2019). The results for sustainable variables related to those dimensions showed both in hedonic semi-log pricing models, in which the dimensions are included on an individual basis, and in a model in which every dimension is included, that higher values for the variables of sustainable dimensions increased overall the willingness to pay for housing. Additionally, results from the model include all the dimensions of sustainability by an often observed lower strength of the relationship between the housing price and sustainable variables the importance of a multidimensional approach. Moreover, sustainable variables with a low (in)significance can become (in)significant when controlling for other dimensions of sustainability. In conclusion, the model including every dimension of sustainability showed evidence for a significantly positive price effect on housing prices of the access to bars & restaurants, universities, pharmacies, big shopping places, museums, libraries & POI culture, and the income & unemployment variables included in the model as PCA components. Evidence for significant negative price effect is found for the distance to highway/train, distance to the beach, neighborhood size, vulnerability to heat waves, natural population growth, density, and access to bus & metro, park & gardens, viewpoints, performing arts, and the income distribution PCA components. Robustness of the results is provided by stating the same conclusions for the sustainable variables for both the sample including and excluding observations with missing energy label data.

The results are visualized by constructing two demonstrative maps consisting of 100 clusters in which the properties in the sample are clustered based on location. The demonstrative maps show that the properties with a high/low price impact of sustainability are clustered in specific areas. The first demonstrative map, which included every sustainable dimension, shows that a high/low total price impact of sustainability for a cluster was mainly caused by a combination of a low/high distance to the beach, low/high distance to the highway or train, high/low welfare and high/low-income equality in the area of the property. The second demonstrative map excluded sustainable variables that are strictly location-bounded and are observed to have a negative relationship between social fairness and housing prices: distance to the beach, neighborhood size, and income distribution PCA component. These sustainable factors are not likely to be changed by policy intervention. With the exclusion of the variables, a higher variety in the reasons for a high/low price impact of sustainability in housing prices is found, which are different for most clusters of properties. This implies that policy intervention to address the unfair pricing of sustainability is best suited for a localized approach. Additionally, the code to construct the maps is shared on GitHub. Hereby, offering the opportunity for future work to visualize the results of the pricing models by focusing on other aspects.

This research provides opportunities for future work. Firstly, the pricing models measure sustainability from a wide view, however, new sustainable factors can be introduced that do not have a high correlation with the included variables. Secondly, are the visualized results by the demonstrative maps only limited in terms of interpretation and not connected to the costs to improve the score of the sustainable factors. Thirdly, are similar visualization maps not reported for different cities in earlier research. Therefore cannot be verified, in contrast to the findings for the pricing models, if results are in line with the pricing of sustainability in earlier work.

Section 7: Reference List

Ajuntament de Barcelona. (2023a). *El plan de los barrios de Barcelona*. Https://Www.Pladebarris.Barcelona/ca/Accions/Habitatge#Eix=.Rehabilitacio.

Ajuntament de Barcelona. (2023b, June 15). *Restrictions in the EXCEPTIONALITY phase*. Https://Ajuntament.Barcelona.Cat/Ecologiaurbana/Es/Sequia-En-Barcelona/Fase-Excepcionalidad.

Arcadis. (2022). *Sustainable cities index*.

Baldominos, A., Blanco, I., Moreno, A., Iturrarte, R., Bernárdez, Ó., & Afonso, C. (2018). Identifying Real Estate Opportunities Using Machine Learning. *Applied Sciences*, *8*(11), 2321. https://doi.org/10.3390/app8112321

Banco Santander. (2023). *Energy Efficiency Loan*. Https://Www.Bancosantander.Es/En/Santander-Sostenible/Prestamo-Verde-Eficiencia-Energetica.

BBVA. (2022, May 19). *BBVA's sustainable solutions for a green future*. Https://Www.Bbva.Com/En/Sustainability/Bbvas-Sustainable-Solutions-for-a-Green-Future.

Brandt, S., Maennig, W., & Richter, F. (2014). Do Houses of Worship Affect Housing Prices? Evidence from Germany. *Growth and Change*, *45*(4), 549–570. https://doi.org/10.1111/grow.12066

Brounen, D., & Kok, N. (2011). On the economics of energy labels in the housing market. *Journal of Environmental Economics and Management*, *62*(2), 166–179. https://doi.org/10.1016/j.jeem.2010.11.006

Buonanno, P., Montolio, D., & Raya-Vílchez, J. M. (2013). Housing prices and crime perception. *Empirical Economics*, *45*(1), 305–321. https://doi.org/10.1007/s00181-012-0624-y

Cajias, M., Fuerst, F., & Bienert, S. (2016). Are Energy Efficiency Ratings Ignored in the German Housing Market? Evidence from a Large-Sample Hedonic Study. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2799206

Castro Noblejas, H., Vías Martínez, J., & Mérida Rodríguez, M. F. (2022). Relation between the Views and the Real Estate Application to a Mediterranean Coastal Area. *ISPRS International Journal of Geo-Information*, *11*(7), 365. https://doi.org/10.3390/ijgi11070365

Ceccato, V., & Wilhelmsson, M. (2020). Do crime hot spots affect housing prices? *Nordic Journal of Criminology*, *21*(1), 84–102. https://doi.org/10.1080/2578983X.2019.1662595

Chen, A., & Marmolejo Duarte, C. R. (2018). The marginal price of housing energy-efficiency in Metropolitan Barcelona: issues of sample selection biases. *Libro de Proceedings, CTV 2018: XII Congreso Internacional Ciudad y Territorio Virtual: "Ciudades y Territorios Inteligentes": UNCuyo, Mendoza, 5-7 Septiembre 2018*, 247–262. https://doi.org/10.5821/ctv.8245

Cui, N., Gu, H., Shen, T., & Feng, C. (2018). The Impact of Micro-Level Influencing Factors on Home Value: A Housing Price-Rent Comparison. *Sustainability*, *10*(12), 4343. https://doi.org/10.3390/su10124343

de Ayala, A., Galarraga, I., & Spadaro, J. V. (2016). The price of energy efficiency in the Spanish housing market. *Energy Policy*, *94*, 16–24. https://doi.org/10.1016/j.enpol.2016.03.032

De Bruyne, K., & Van Hove, J. (2013). Explaining the spatial variation in housing prices: an economic geography approach. *Applied Economics*, *45*(13), 1673–1689. https://doi.org/10.1080/00036846.2011.636021

Debrezion, G., Pels, E., & Rietveld, P. (2011). The Impact of Rail Transport on Real Estate Prices. *Urban Studies*, *48*(5), 997–1015. https://doi.org/10.1177/0042098010371395

Dell'Anna, Bravi, Marmolejo-Duarte, Bottero, & Chen. (2019). EPC Green Premium in Two Different European Climate Zones: A Comparative Study between Barcelona and Turin. *Sustainability*, *11*(20), 5605. https://doi.org/10.3390/su11205605

Eichholtz, P., Kok, N., & Quigley, J. M. (2013). The Economics of Green Building. *Review of Economics and Statistics*, *95*(1), 50–63. https://doi.org/10.1162/REST_a_00291

Eurostat Statistics Explained. (2022, September). *Quality of life indicators - natural and living environment*. Https://Ec.Europa.Eu/Eurostat/Statistics-Explained/Index.Php?Title=Quality_of_life_indicators_-_natural_and_living_environment.

Fuerst, F., McAllister, P., Nanda, A., & Wyatt, P. (2015). Does energy efficiency matter to home-buyers? An investigation of EPC ratings and transaction prices in England. *Energy Economics*, *48*, 145–156. https://doi.org/10.1016/j.eneco.2014.12.012

Funds Europe. (2023). *Real estate sustainable properties at a greenium*. Https://Www.Funds-Europe.Com/Insights/Real-Estate-Sustainable-Properties-at-a-Greenium.

Graells-Garrido, E., Serra-Burriel, F., Rowe, F., Cucchietti, F. M., & Reyes, P. (2021). A city of cities: Measuring how 15-minutes urban accessibility shapes human mobility in Barcelona. *PLOS ONE*, *16*(5), e0250080. https://doi.org/10.1371/journal.pone.0250080

Idealista. (2023a). *Idealista*. Www.Idealista.Com.

Idealista. (2023b). *Idealista Data*. Https://Www.Idealista.Com/En/Data/.

Jeanty, P. W., Partridge, M., & Irwin, E. (2010). Estimation of a spatial simultaneous equation model of population migration and housing price dynamics. *Regional Science and Urban Economics*, *40*(5), 343–352. https://doi.org/10.1016/j.regsciurbeco.2010.01.002

Kauko, T. (2019). Innovation in urban real estate: the role of sustainability. *Property Management*, *37*(2), 197–214. https://doi.org/10.1108/PM-10-2017-0056

Kim, E. J., & Kim, H. (2020). Neighborhood Walkability and Housing Prices: A Correlation Study. *Sustainability*, *12*(2), 593. https://doi.org/10.3390/su12020593

Lazrak, F., Nijkamp, P., Rietveld, P., & Rouwendal, J. (2014). The market value of cultural heritage in urban areas: an application of spatial hedonic pricing. *Journal of Geographical Systems*, *16*(1), 89–114. https://doi.org/10.1007/s10109-013-0188-1

Leaflet. (2023). *Leaflet an open-source JavaScript library for mobile-friendly interactive maps*. Https://Leafletjs.Com/.

Lee, H., Lee, B., & Lee, S. (2020). The Unequal Impact of Natural Landscape Views on Housing Prices: Applying Visual Perception Model and Quantile Regression to Apartments in Seoul. *Sustainability*, *12*(19), 8275. https://doi.org/10.3390/su12198275

Mandell, S., & Wilhelmsson, M. (2011). Willingness to Pay for Sustainable Housing. *Journal of Housing Research*, *20*(1), 35–51. https://doi.org/10.1080/10835547.2011.12092034

Marmolejo Duarte, C., & González Tamez, C. (2009). Does noise have a stationary impact on residential values? *Journal of European Real Estate Research*, *2*(3), 259–279. https://doi.org/10.1108/17539260910999992

Marmolejo-Duarte, C., & Chen, A. (2022). The effect of energy performance ratings over residential prices or how an insufficient control of architectural-quality may render spurious conclusions. *Cities*, *126*, 103674. https://doi.org/10.1016/j.cities.2022.103674

McGreal, S., Brown, L., & Adair, A. (2010). List price and sale price variation across the housing market cycle. *International Journal of Housing Markets and Analysis*, *3*(2), 89–99. https://doi.org/10.1108/17538271011049731

N. Gosling. (2022, August 11). *Buyers want green homes. But the market and government need to catch up*. Https://Www.Insidehousing.Co.Uk/Comment/Buyers-Want-Green-Homes-but-the-Market-and-Government-Need-to-Catch-up-76905.

Oladunni, T., & Sharma, S. (2016). Hedonic Housing Theory — A Machine Learning Investigation. *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 522–527. https://doi.org/10.1109/ICMLA.2016.0092

Openstreetmap contributers. (2023). *Opensteetmap*. Www.Openstreetmap.Org.

Ossokina, I. V., & Verweij, G. (2015). Urban traffic externalities: Quasi-experimental evidence from housing prices. *Regional Science and Urban Economics*, *55*, 1–13. https://doi.org/10.1016/j.regsciurbeco.2015.08.002

Park, J., Lee, D., Park, C., Kim, H., Jung, T., & Kim, S. (2017). Park Accessibility Impacts Housing Prices in Seoul. *Sustainability*, *9*(2), 185. https://doi.org/10.3390/su9020185

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, *82*(1), 34–55. https://doi.org/10.1086/260169

Taltavull de La Paz, P., Perez-Sanchez, V., Mora-Garcia, R.-T., & Perez-Sanchez, J.-C. (2019). Green Premium Evidence from Climatic Areas: A Case in Southern Europe, Alicante (Spain). *Sustainability*, *11*(3), 686. https://doi.org/10.3390/su11030686

Tillema, T., Hamersma, M., Sussman, J. M., & Arts, J. (2012). Extending the Scope of Highway Planning: Accessibility, Negative Externalities and the Residential Context. *Transport Reviews*, *32*(6), 745–759. https://doi.org/10.1080/01441647.2012.724726

Triguero-Mas, M., Dadvand, P., Cirach, M., Martínez, D., Medina, A., Mompart, A., Basagaña, X., Gražulevičienė, R., & Nieuwenhuijsen, M. J. (2015). Natural outdoor environments and mental and physical health: Relationships and mechanisms. *Environment International*, *77*, 35–41. https://doi.org/10.1016/j.envint.2015.01.012

United Nations. (2023). *The 17 Goals*. Https://Sdgs.Un.Org/Goals.

Vargas-Calderón, V., & Camargo, J. E. (2022). Towards robust and speculation-reduction real estate pricing models based on a data-driven strategy. *Journal of the Operational Research Society*, *73*(12), 2794–2807. https://doi.org/10.1080/01605682.2021.2023672

Vidal Yañez, D., Pereira Barboza, E., Cirach, M., Daher, C., Nieuwenhuijsen, M., & Mueller, N. (2023). An urban green space intervention with benefits for mental health: A health impact assessment of the Barcelona “Eixos Verds” Plan. *Environment International*, 107880. https://doi.org/10.1016/j.envint.2023.107880

World Green Building Council. (2022, March 17). *What are green mortgages & how will they revolutionise home energy efficiency?* Https://Worldgbc.Org/Article/What-Are-Green-Mortgages-How-Will-They-Revolutionise-Home-Energy-Efficiency/.

XGBoost Developers. (2023). *XGBoost*. Https://Xgboost.Readthedocs.Io.

Zhang, X., Liu, X., Hang, J., Yao, D., & Shi, G. (2016). Do Urban Rail Transit Facilities Affect Housing Prices? Evidence from China. *Sustainability*, *8*(4), 380. https://doi.org/10.3390/su8040380

Section 8: Appendix

Table 3: Summary Statistics of the Continuous Housing-Specific Variables for the sample that excludes observations with missing energy label data.

	count	mean	std	min	25%	50%	75%	max	skew	kurt
Log price	10104	5.94	0.61	4.53	5.51	5.89	6.35	7.51	0.24	-0.38
Building surface m²	10104	100.12	46.39	18	69	87	117	288	1.44	1.99
Building age	10104	67.4	32.63	0	50	58	88	143	0.4	-0.4
sq(Building surface m²)	10104	12176	12743	324	4761	7569	13689	82944	2.47	6.82
sq(Building age)	10104	5607	4988	0	2500	3364	7744	20449	1.17	0.14

Table 2 includes the summary statistics for the continuous housing-specific variables for the sample that excludes the observations with missing energy label data.

Table 5: Summary Statistic for the Housing-Specific Dummy Variables for the sample that excludes observations with missing energy label data.

Variable	count	mean	Variable	count	mean
House	10104	0.02	Energy label consumption C	10104	0.05
Mezzanine	10104	0.05	Energy label consumption D	10104	0.11
Ground Floor	10104	0.1	Energy label consumption E	10104	0.51
1st Floor	10104	0.23	Energy label consumption F	10104	0.09
2nd – 5th Floor	10104	0.49	Energy label consumption G	10104	0.17
6th Floor or higher	10104	0.11	District Eixample	10104	0.22
New housing development	10104	0.02	District Ciutat Vella	10104	0.15
Good condition	10104	0.84	District Sant Martí	10104	0.1
Needs renovation	10104	0.14	District Sants-Montjuïc	10104	0.11
Elevator	10104	0.77	District Horta Guinardó	10104	0.08
Terrace	10104	0.32	District Gràcia	10104	0.07
Heating	10104	0.54	District Les Corts	10104	0.04
Outdoor facilities	10104	0.56	District Nou Barris	10104	0.06
Parking space included	10104	0.1	District Sarrià-Sant Gervasi	10104	0.1
Energy label consumption A	10104	0.03	District Sant Andreu	10104	0.06
Energy label consumption B	10104	0.03			

Table 5 includes the summary statistics for the housing-specific dummy variables for the sample that excludes the observations with missing energy label data.

Table 6: Description of the Sustainable Variables

Variable(s)	Sustainable Dimension	Resource	Description Dataset
Energy Label Consumption A, Energy Label Consumption B, Energy Label Consumption C, Energy Label Consumption D, Energy Label Consumption E, Energy Label Consumption F, Energy Label Consumption N_A	Ecological	https://www.idealista.com/	Energy Labels in the housing advertisement of the property
Number of Bus stops within … meters, Distance to the nearest Bus stop	Ecological	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=estacions-bus	Bus stops of the city of Barcelona
Distance to the Highway	Ecological	https://www.openstreetmap.org/	Keyword:
Number of Metros Stations within … meters, Distance to nearest Metro Station	Ecological	https://www.openstreetmap.org/	Keyword:
Distance to Train Station	Ecological	https://www.openstreetmap.org/	Keyword:
Numbers of Parks/Gardens within … meters, Distance to nearest Park/Garden	Environmental	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=culturailleure-parcsjardins	Parks and gardens
Neighborhood area size	Environmental	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=est-superficie	Neighborhoods' area size of the city of Barcelona
Number of Viewpoints within … meters, Distance to nearest Viewpoint	Environmental	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=np-nasia-miradors	Viewpoints in teh city of Barcelona
Noise Pollution	Environmental	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=tramer-mapa-estrategic-soroll	Noise maps by street section from Strategic Noise Map of the city of Barcelona
Vulnerability to Heat Exposure	Environmental	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=impacte-de-la-calor	Most vulnerable areas in the city of Barecelona to heat exposure
Distance to the Beach	Environmental	https://www.openstreetmap.org/	Keyword:
Number of Police Stations within .. Meters, Distance to nearest Police Station	Social	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=comissaries-policia	Police stations
Number of Cinemas within … meters, Distance to nearest Cinema, Number of Theatres within … meters, Distance to nearest Theatre, Number of Concert places within … meters, Distance to nearest Concert place	Social	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=culturailleure-cinemesteatresauditoris	Spaces where cinema, theater and concerts take place
Number of Bars within … meters, Distance to nearest Bar	Social	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=culturailleure-espaismusicacopes	Music and drinks' spaces
Number of restaurants within … meters, Distance to nearest Restaurant	Social	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=equipament-restaurants	List of restaurant equipments
Number of education facilities 0-3 year within … meters, Distance to nearest education facility 0-3 years, Number of education facilities 4-6 years within … meters, Distance to nearest education facility 4-6 years, Number of Primary Schools within … meters, Distance to nearest Primary School, Number of Secondary Schools within … meters, Distance to nearest Secondary School, Number of Universities within … meters, Distance to nearest University	Social	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=equipament-educacio	List of education equipments
Number of Pharmacy within … meters, Distance to nearest Pharmacy …, Number of Hospitals/ Clinique's within … meters, Distance to nearest Hospital/Clinique	Social	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=equipament-sanitat	List of health equipments
Immigration rate	Social	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=est-demo-taxa-immigracio	Immigration registration rate (‰ inhabitants) of the city of Barcelona

Emigration rate	Social	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=est-demo-taxa-emigracio	Leave rate due to emigration (‰ inhabitants) from the city of Barcelona
Life Expectancy	Social	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=est-sp-esp-vida	Life expectancy of the city of Barcelona
Mortality rate	Social	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=est-demo-taxa-mortalitat	Mortality rate (‰ inhabitants) of the city of Barcelona
Birth rate	Social	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=est-demo-taxa-natalitat	Birth rate (‰ inhabitants) of the city of Barcelona
Population density	Social	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=est-densitat	Population density (inhabitants / ha) of the city of Barcelona
% population under 18 years old	Social	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=est-padro-domicilis-menors-18-anys	Addresses of the city of Barcelona with people under 18 years of age
% population between 18-64 years olds	Social	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=est-padro-domicilis-18-a-64-anys	Addresses of the city of Barcelona with people from 18 to 64 years old
% population older than 65 years old	Social	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=est-padro-domicilis-65-mes-anys	Addresses with people aged 65 or over
Number of Shopping Galleries within … meters, Distance to nearest Shopping Gallery	Social	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=galeries-comercials	Shopping galleries
Number of Large Shopping Centres within … meters, Distance to nearest Shopping Centre	Social	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=grans-centres-comercials	Large shopping centers
Number of Large Establishments within … meters, Distance to nearest Large Establishment	Social	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=grans-establiments	Large establishments
Number of Street Market and Fairs within … meters. Distance to nearest Street Market and Fair	Social	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=mercats-fires-carrer	Street markets and fairs
Social Cohesion	Social	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=cohesio-social	Social cohesion
Number of Libraries within … meters, Distance to nearest Library, Number of Museums within … meters, Distance to nearest Museum	Cultural	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=culturailleure-bibliotequesimuseus	Spaces with a library or study room and museum spaces
Number of Religious institutions within … meters, Distance to nearest Religious institution	Cultural	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=equipament-serveis-religiosos	List of religious services equipments
Number of Cultural Interest Points within … meters, Distance to nearest viewpoint	Cultural	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=punts-informacio-turistica	Cultural interest points
P80/P20 income distribution	Economic/Financial	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=atles-renda-p80-p20-distribucio	P80/P20 income distribution
AVG Gross taxable household income by subdistrict	Economic/Financial	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=atles-renda-bruta-per-llar	Average gross taxable income per household (€/Year) for the city of Barcelona
Unemployment rate	Economic/Financial	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=est-atur-pes	Weight of the registered unemployment in the population from 16 to 64 years of age of the city of Barcelona

Table 6 includes a description of the sustainable variables. Additionally, is information provided about the sustainable dimension, the source, and a description of the dataset of the variables.

Table 9: Summary Statistics of the Sustainable Variables for the sample that excludes the observations with the missing energy label observations

	mean	std	min	25%	50%	75%	max	skew	kurt
Bus & Metro PCA	0.37	0.18	0	0.24	0.36	0.48	1	0.6	0.47
Distance to Highway/Train (km)	1.14	0.72	0.02	0.61	0.98	1.5	4.91	1.14	1.37
Distance to Beach (km)	3.28	1.4	0.01	2.13	3.43	4.61	5	-0.37	-1.03
Park & Garden PCA	0.4	0.39	0	0.01	0.5	0.51	1	0.36	-1.33
Viewpoint PCA	0.41	0.26	0	0.21	0.32	0.64	1	0.67	-0.79
Neighborhood size (10 ha)	13.66	12.91	2.3	8.08	11.1	14.1	142.37	6.02	50.01
Vulnerable to heat impact (1-5)	2.85	0.79	1	2	3	3	5	0.1	-1.02
Police PCA	0.44	0.28	0	0.29	0.32	0.55	1	0.32	-0.77
Bar & Restaurant PCA	0.5	0.35	0	0.19	0.34	0.99	1	0.26	-1.4
Secondary & Lower School PCA	0.46	0.2	0	0.32	0.46	0.62	1	0	-0.61
University PCA	0.35	0.38	0	0.05	0.24	0.63	1	0.76	-1.08
Pharmacy PCA	0.5	0.22	0	0.4	0.5	0.6	1	-0.02	-0.19
Hospital & Clinique PCA	0.37	0.28	0	0.21	0.22	0.6	1	0.66	-0.36
Big Shopping Place PCA	0.34	0.3	0	0.08	0.2	0.54	1	0.87	-0.56
Social Cohesion Score	0.21	0.18	0	0.09	0.15	0.3	1.73	1.8	4.81
Natural population growth ‰	-2.09	1.96	-8.92	-3.3	-2.5	-0.6	3.6	0.12	0.55
Net immigration rate ‰	27.27	22.42	-6.6	9.9	19.5	46.9	91.1	1.15	0.82
Density net ((hab/1000)/ha)	0.74	0.23	0.02	0.63	0.74	0.91	1.37	-0.41	0.11
Performing Arts PCA	0.31	0.31	0	0.06	0.17	0.42	1	1.15	0.02
Religious Institution PCA	0.44	0.25	0	0.27	0.4	0.6	1	0.54	-0.45
Museum, Library & POI Cult. PCA	0.26	0.23	0	0.09	0.16	0.35	0.99	1.32	0.88
Income Distribution PCA	0.46	0.21	0	0.3	0.47	0.62	1	-0.32	-0.61
Income & Unemployment PCA	0.2	0.2	0	0.07	0.15	0.27	1	1.9	3.57

Table 9 Includes the summary statistics for the sustainable variable for the sample that excludes the observations with the missing energy label data.

Fig 3: Correlation Matrix of the variables for the Sample that Includes Observations with Missing Energy Labels Data that Have an Absolute Correlation Coefficient Higher Than 0.25

Fig 4: Correlation Matrix of the variables for the Sample that Excludes Observations with Missing Energy Labels Data that Have an Absolute Correlation Coefficient Higher Than 0.25

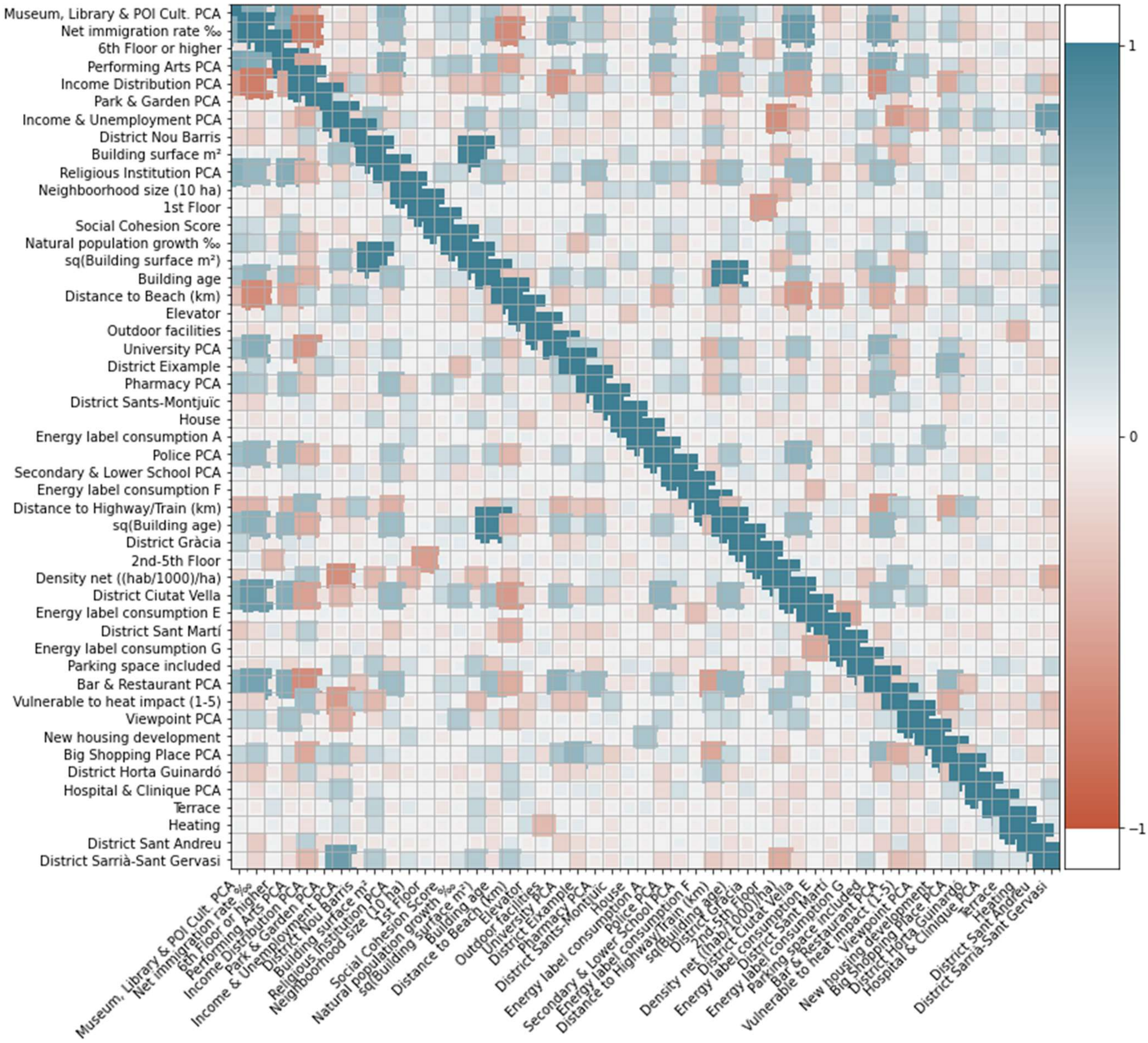


Table 15A: Semi-Log Pricing Model Results of the Sample Excluding Observations with Missing Energy Label Data

	Model 1		Model 2		Model 3		Model 4	
	Coef.	Stderr.	Coef.	Stderr.	Coef.	Stderr.	Coef.	Stderr.
Constant	4.686***	0.03	4.878***	0.03	5.537***	0.03	4.660***	0.03
Building surface m²	0.017***	0.00	0.016***	0.00	0.016***	0.00	0.015***	0.00
sq(Building surface m²)	-$3 * 10^{5}$***	0.00	-$3 * 10^{5}$***	0.00	-$3 * 10^{5}$***	0.00	-$3 * 10^{5}$***	0.00
Building age	-0.006***	0.00	-0.006***	0.00	-0.005***	0.00	-0.005***	0.00
sq(Building age)	$4 * 10^{5}$***	0.00	$4 * 10^{5}$***	0.00	$4 * 10^{5}$***	0.00	$3 * 10^{5}$***	0.00
House	0.173***	0.03	0.189***	0.03	0.223***	0.02	0.224***	0.03
Mezzanine	0.047***	0.01	0.050***	0.01	0.050***	0.01	0.045***	0.01
1st Floor	0.111***	0.01	0.117***	0.01	0.115***	0.01	0.110***	0.01
2^{nd} -5^{th} Floor	0.158***	0.01	0.162***	0.01	0.163***	0.01	0.154***	0.01
6th Floor or higher	0.230***	0.01	0.229***	0.01	0.223***	0.01	0.237***	0.01
New housing development	0.168***	0.02	0.179***	0.02	0.213***	0.02	0.175***	0.02

	Model 1		Model 2		Model 3		Model 4	
	Coef.	Stderr.	Coef.	Stderr.	Coef.	Stderr.	Coef.	Stderr.
Needs renovation	-0.118***	0.01	-0.122***	0.01	-0.117***	0.01	-0.115***	0.01
Elevator	0.174***	0.01	0.153***	0.01	0.144***	0.01	0.164***	0.01
Terrace	0.073***	0.01	0.078***	0.01	0.081***	0.01	0.082***	0.01
Heating	0.054***	0.01	0.053***	0.01	0.048***	0.01	0.048***	0.01
Outdoor facilities	-0.047***	0.01	-0.045***	0.01	-0.043***	0.01	-0.004***	0.01
Parking space included	0.085***	0.01	0.094***	0.01	0.087***	0.01	0.090***	0.01
Energy label consumption A	0.048**	0.02	0.050**	0.02	0.109***	0.02	0.034*	0.02
Energy label consumption B	0.152***	0.02	0.148***	0.02	0.150***	0.02	0.148**	0.02
Energy label consumption C	0.061***	0.02	0.066***	0.02	0.061***	0.01	0.061***	0.01
Energy label consumption E	-0.022**	0.01	-0.019**	0.01	-0.010	0.01	-0.012	0.01
Energy label consumption F	-0.043***	0.01	-0.037***	0.01	-0.032***	0.01	-0.036***	0.01
Energy label consumption G	-0.058***	0.01	-0.048***	0.01	-0.037***	0.01	-0.038***	0.01
Bus & metro PCA			0.005	0.00				
Distance to Highway/Train (km)			-0.100***	0.01				
Distance to Beach (km)					-0.104	0.00		
Park & garden PCA					-0.007***	0.00		
Viewpoint PCA					-0.122***	0.01		
Neighborhood size (10 ha)					-0.002***	0.01		
Vulnerable to heat impact (1-5)					-0.104***	0.00		
Police PCA							0.020	0.01
Bar & restaurant PCA							0.116***	0.01
Secondary & lower educ. PCA							-0.007	0.01
University PCA							0.078***	0.01
Pharmacy PCA							0.008	0.01
Hospital & Clinique PCA							-0.001	0.01
Big Shopping Place PCA							0.108***	0.01
Social Cohesion Score							-0.074***	0.01
Natural population growth ‰							-0.018***	0.00
Net immigration rate ‰							0.004	
Density net (hab/1000/ha)							-0.180***	0.02
Performing Arts PCA								
Religious Institution PCA								
Museum, Library & POI Cult. PCA								
Income Distribution PCA								
Income & Unemployment PCA								
District Eixample	0.001	0.01	-0.053***	0.01	-0.166***	0.01	-0.103	0.01
District Ciutat Vella	-0.087***	0.01	-0.137***	0.01	-0.335***	0.02	-0.208***	0.02
District Sant Martí	-0.208***	0.01	-0.215***	0.01	-0.487***	0.02	-0.089***	0.02
District Sants-Montjuïc	-0.269***	0.01	-0.323***	0.01	-0.231***	0.01	-0.160***	0.01
District Horta Guinardó	-0.345***	0.01	-0.296***	0.01	-0.284***	0.01	-0.208***	0.01
District Gràcia	-0.014	0.01	-0.021	0.01	-0.077***	0.01	0.059***	0.02
District Nou Barris	-0.517***	0.02	-0.456***	0.01	-0.380***	0.02	-0.372***	0.02
District Sarrià-Sant Gervasi	0.104***	0.01	0.027**	0.01	0.088***	0.01	0.136***	0.01
District Sant Andreu	-0.416***	0.02	-0.434***	0.01	-0.445***	0.02	-0.252***	0.02
R-squared	0.831		0.838		0.856		0.851	

Table 15 includes the results for semi-log hedonic priding models of the sample excluding observations with missing energy label data. The results report both the coefficients (Coef.) and the standard errors (Stderr.) of the variables. *** denotes a significance at 1% level, *** denotes a significance at the 5% level, and * denotes significance at the 10% level.

Table 15B: Semi-Log Pricing Model Results of the Sample Excluding Observations with Missing Energy Label Data

	Model 4		Model 5		Model 6		Model 7	
	Coef.	Stderr.	Coef.	Stderr.	Coef.	Stderr.	Coef.	Stderr.
Constant	4.660***	0.03	4.655***	0.03	4.637***	0.04	5.200***	0.06
Building surface m²	0.015***	0.00	0.016***	0.00	0.016***	0.00	0.015***	0.00
sq(Building surface m²)	$-3 * 10^{5}$***	0.00	$-3 * 10^{5}$***	0.00	$-3 * 10^{5}$***	0.00	$-3 * 10^{5}$***	0.00
Building age	-0.005***	0.00	-0.005***	0.00	-0.006***	0.00	-0.004***	0.00
sq(Building age)	$3 * 10^{5}$***	0.00	$3 * 10^{5}$***	0.00	$4 * 10^{5}$***	0.00	$3 * 10^{5}$***	0.00

	Model 4		Model 5		Model 6		Model 7	
	Coef.	Stderr.	Coef.	Stderr.	Coef.	Stderr.	Coef.	Stderr.
House	0.224***	0.03	0.191***	0.03	0.220***	0.02	0.283***	0.02
Mezzanine	0.045***	0.01	0.051***	0.01	0.049***	0.01	0.052***	0.01
1st Floor	0.110***	0.01	0.111***	0.01	0.106***	0.01	0.112***	0.01
2^{nd} -5^{th} Floor	0.154***	0.01	0.150***	0.01	0.157***	0.01	0.156***	0.01
6th Floor or higher	0.237***	0.01	0.229***	0.01	0.243***	0.01	0.233***	0.01
New housing development	0.175***	0.02	0.169***	0.02	0.203***	0.02	0.234***	0.02
Needs renovation	-0.115***	0.01	-0.114***	0.01	-0.117***	0.01	-0.113***	0.01
Elevator	0.164***	0.01	0.165***	0.01	0.163***	0.01	0.136***	0.01
Terrace	0.082***	0.01	0.079***	0.01	0.074***	0.01	0.086***	0.01
Heating	0.048***	0.01	0.052***	0.01	0.046***	0.01	0.043***	0.01
Outdoor facilities	-0.004***	0.01	-0.047***	0.01	-0.046***	0.01	-0.042***	0.01
Parking space included	0.090***	0.01	0.095***	0.01	0.043***	0.01	0.072***	0.01
Energy label consumption A	0.034*	0.02	0.050**	0.02	0.064***	0.02	0.110***	0.02
Energy label consumption B	0.148**	0.02	0.151***	0.02	0.135***	0.02	0.138***	0.02
Energy label consumption C	0.061***	0.01	0.056***	0.02	0.061***	0.01	0.062***	0.01
Energy label consumption E	-0.012	0.01	-0.018**	0.01	-0.016*	0.01	-0.004	0.01
Energy label consumption F	-0.036***	0.01	-0.045***	0.01	-0.037***	0.01	-0.029***	0.01
Energy label consumption G	-0.038***	0.01	-0.052***	0.01	-0.047***	0.01	-0.025**	0.01
Bus & metro PCA							-0.034***	0.01
Distance to Highway/Train (km)							-0.015**	0.01
Distance to Beach (km)							-0.094***	0.00
Park & garden PCA							0.011	0.01
Viewpoint PCA							-0.049***	0.01
Neighborhood size (10 ha)							-0.002***	0.00
Vulnerable to heat impact (1-5)							-0.058***	0.00
Police PCA	0.020	0.01					-0.014***	0.01
Bar & restaurant PCA	0.116***	0.01					0.050***	0.01
Secondary & lower educ. PCA	-0.007	0.01					-0.002	0.01
University PCA	0.078***	0.01					0.060***	0.01
Pharmacy PCA	0.008	0.01					0.022	0.01
Hospital & Clinique PCA	-0.001	0.01					0.001	0.01
Big Shopping Place PCA	0.108***	0.01					0.034	0.01
Social Cohesion Score	-0.074***	0.01					-0.038***	0.01
Natural population growth ‰	-0.018***	0.00					-0.008***	0.00
Net immigration rate ‰	0.004						0.001	0.00
Density net (hab/1000/ha)	-0.180***	0.02					-0.067***	0.02
Performing Arts PCA			-0.083***	0.01			-0.044***	0.02
Religious Institution PCA			0.028*	0.01			-0.018	0.02
Museum, Library & POI Cult. PCA			0.387***	0.02			0.204***	0.03
Income Distribution PCA					-0.402***	0.01	-0.114***	0.04
Income & Unemployment PCA					0.682***	0.01	0.516***	0.04
District Eixample	-0.103	0.01	0.018	0.01	0.120***	0.01	-0.091***	0.02
District Ciutat Vella	-0.208***	0.02	-0.167***	0.02	0.110***	0.02	-0.219***	0.02
District Sant Martí	-0.089***	0.02	-0.163***	0.02	0.063***	0.02	-0.247***	0.02
District Sants-Montjuïc	-0.160***	0.01	-0.246***	0.01	-0.002	0.02	-0.039***	0.02
District Horta Guinardó	-0.208***	0.01	-0.328***	0.01	-0.041***	0.02	-0.068***	0.02
District Gràcia	0.059***	0.02	-0.095***	0.02	0.193***	0.02	0.049***	0.02
District Nou Barris	-0.372***	0.02	-0.504***	0.02	-0.169***	0.02	-0.144***	0.02
District Sarrià-Sant Gervasi	0.136***	0.01	0.091***	0.01	-0.081***	0.01	-0.011	0.01
District Sant Andreu	-0.252***	0.02	-0.449***	0.02	-0.092***	0.02	-0.219***	0.02
R-squared	0.851				0.840		0.853	

Table 15 includes the results for semi-log hedonic priding models of the sample excluding observations with missing energy label data. The results report both the coefficients (Coef.) and the standard errors (Stderr.) of the variables. *** denotes a significance at 1% level, *** denotes a significance at the 5% level, and * denotes significance at the 10% level.

Table 16: VIF Test-Statistics of the Sample Excluding Observations with Missing Energy Label Data							
Variable	**VIF Model 1**	**VIF Model 2**	**VIF Model 3**	**VIF Model 4**	**VIF Model 5**	**VIF Model 6**	**VIF Model 7**
Building surface m²	84.51	87.67	96.83	96.92	87.25	98.67	118.39
sq(Building surface m²)	29.09	30.02	32.82	32.17	29.58	33.09	38.21
Building age	84.14	87.33	101.53	96.16	85.12	95.79	116.76
sq(Building age)	40.18	42.01	46.91	45.39	41.55	45.10	53.38
House	1.36	1.37	1.42	1.38	1.36	1.37	1.45
Mezzanine	1.52	1.53	1.54	1.54	1.53	1.54	1.57
1st Floor	3.25	3.30	3.34	3.28	3.26	3.34	3.42
2^{nd} -5^{th} Floor	6.00	6.09	6.22	6.10	6.03	6.18	6.44
6th Floor or higher	2.26	2.28	2.29	2.28	2.27	2.32	2.36
New housing development	1.42	1.43	1.47	1.47	1.43	1.43	1.52
Needs renovation	1.25	1.25	1.25	1.25	1.25	1.25	1.26
Elevator	6.22	6.28	6.26	6.38	6.26	6.27	6.54
Terrace	1.73	1.74	1.75	1.74	1.74	1.74	1.77
Heating	2.71	2.72	2.73	2.73	2.71	2.74	2.77
Outdoor facilities	2.71	2.73	2.77	2.72	2.71	2.79	2.83
Parking space included	1.43	1.44	1.47	1.47	1.45	1.49	1.55
Energy label consumption A	1.47	1.48	1.56	1.48	1.47	1.50	1.59
Energy label consumption B	1.37	1.38	1.41	1.41	1.37	1.42	1.47
Energy label consumption C	1.44	1.45	1.46	1.45	1.44	1.46	1.48
Energy label consumption E	5.31	5.36	5.54	5.39	5.32	5.50	5.65
Energy label consumption F	1.77	1.79	1.82	1.79	1.77	1.81	1.85
Energy label consumption G	2.52	2.56	2.63	2.56	2.52	2.60	2.68
Bus & metro PCA		5.56					6.34
Distance to Highway/Train (km)		6.05					11.84
Distance to Beach (km)			23.35				38.60
Park & garden PCA			2.39				2.64
Viewpoint PCA			5.25				7.92
Neighborhood size (10 ha)			2.88				3.79
Vulnerable to heat impact (1-5)			21.89				28.24
Police PCA				6.04			6.66
Bar & restaurant PCA				9.29			10.48
Secondary & lower educ. PCA				9.14			11.19
University PCA				3.67			4.15
Pharmacy PCA				10.12			11.08
Hospital & Clinique PCA				3.62			3.83
Big Shopping Place PCA				5.27			6.54
Social Cohesion Score				2.89			2.98
Natural population growth ‰				3.45			5.14
Density net (hab/1000/ha)				17.30			27.45
Performing Arts PCA					5.53		36.09
Religious Institution PCA					8.93		8.25
Museum, Library & POI Cult. PCA					6.16		13.81
Income Distribution PCA						19.48	7.98
Income & Unemployment PCA						9.63	56.43
District Eixample	5.11	5.18	6.18	8.05	5.42	5.46	13.86
District Ciutat Vella	4.12	4.12	7.02	13.05	5.62	5.82	9.76
District Sant Martí	2.65	2.72	4.38	3.76	2.74	3.60	15.62
District Sants-Montjuïc	2.88	2.89	4.00	4.01	3.17	3.77	7.26
District Horta Guinardó	2.33	2.68	3.25	2.87	2.41	3.23	5.56
District Gràcia	2.17	2.23	2.55	2.73	2.53	2.54	4.37
District Nou Barris	2.03	2.31	2.83	2.56	2.37	2.88	3.61
District Sarrià-Sant Gervasi	2.84	2.93	3.08	3.11	2.95	3.50	4.21
District Sant Andreu	1.94	1.97	2.36	2.39	2.03	2.72	4.53

Table 16 includes the VIF test-statistics for the variables included in the Hedonic Semi-Log Pricing Models of the sample that excludes observations with missing energy label data.

Figure 7: Screenshot of Demonstrative Map 1 for the Sample Excluding Observations with Missing Energy Label Data

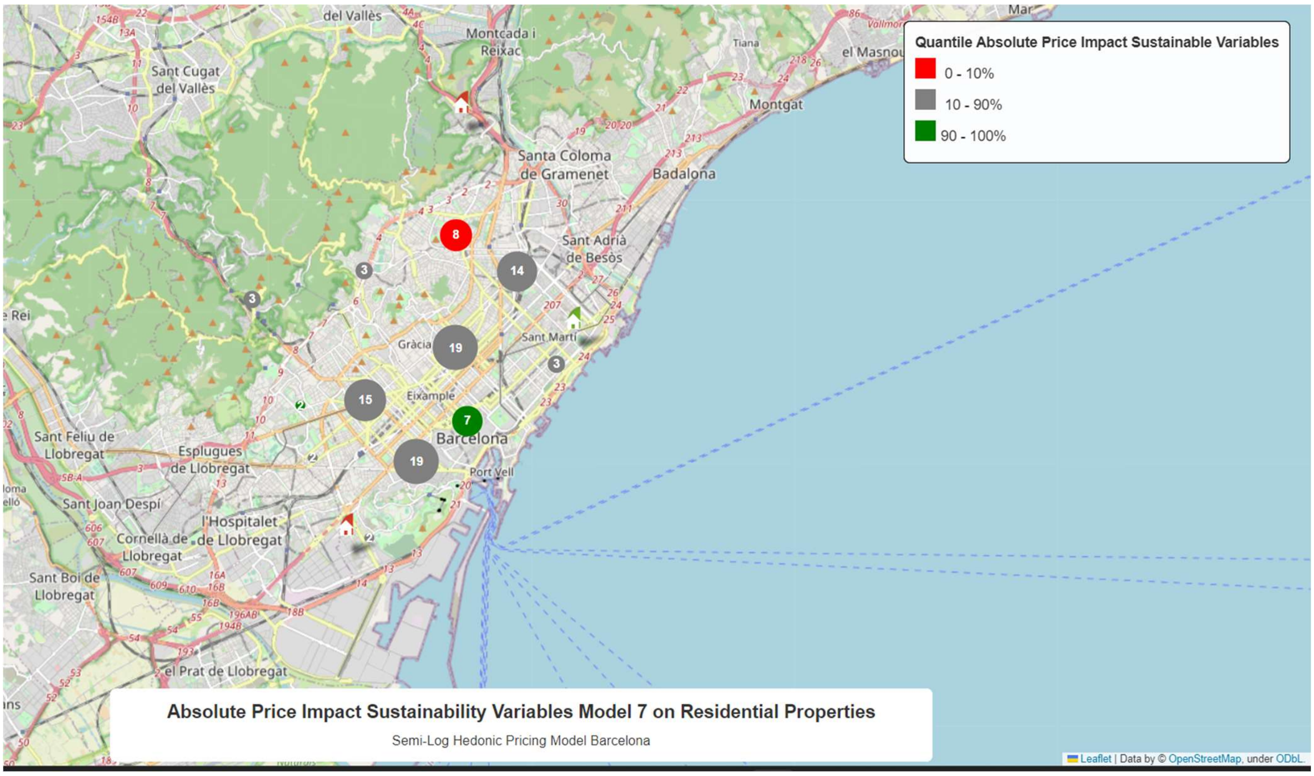


Figure 8: Screenshot of Demonstrative Map 2 for the Sample Excluding Observations with Missing Energy Label Data