Master Thesis on Intelligent Interactive Systems

Universitat Pompeu Fabra

The Price Impact of Sustainability on Housing Prices in Barcelona

A Multidimensional Data-Driven Approach

Author: Niels Box

Supervisor: Manuel Portela

5 July 2023

Universitat Pompeu Fabra
Barcelona

Master Thesis on Intelligent Interactive Systems

Universitat Pompeu Fabra

The Price Impact of Sustainability on Housing Prices in Barcelona

A Multidimensional Data-Driven Approach

Author: Niels Box

Supervisor: Manuel Portela

5 July 2023

upf. Universitat Pompeu Fabra Barcelona

Table of Contents

Abstract:

This research applies a data-driven multidimensional approach to study the price impact of sustainability on housing prices in Barcelona, Spain. The price impact is studied with sustainable pricing factors related to five dimensions: ecological, environmental, social, cultural, and economic-financial. In total 23 location-bounded sustainable pricing factors are proposed. Most of the sustainable pricing factors are PCA components constructed on sustainable features with a- high correlation to avoid multicollinearity in pricing models. The pricing models are estimated as semi-log hedonic pricing models. The hedonic pricing model assumes that coefficients are equal to the utility buyers are expected to extract from a one-unit increase in the pricing factor. The semi-log model implies that the pricing factors can be interpreted as price elasticities, where each unit increase in the pricing factor is equal to the percentage change in the housing price. The sample of the model consists of respectively 13.500 and 10.500 observations dependent on the inclusion/exclusion of properties with missing energy labels. Results provide evidence that an increase in every sustainable dimension increases the willingness to pay for housing. Furthermore, evidence is provided that the strength of the increase or decrease of the willingness to pay for housing by sustainable factors will be overestimated if sustainability is only studied from one dimension. Hereby, prior literature is extended which often studied sustainability only from one dimension or variable. Additionally, prior work is extended by constructing demonstrative maps that visualize the total price impact of sustainability on housing prices. These maps show that houses with a high/low total price impact of sustainability are locally clustered in Barcelona caused by a wide variety of sustainable factors. These findings suggest that the price impact of sustainability on house prices can be increased and made fairer by local policy interventions. Furthermore, the code, which offers a high flexibility to visualize the results, to construct demonstrative maps is shared on GitHub.

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

Section 1: Introduction

The City Council of Barcelona has the goal to reach a more sustainable living environment for the inhabitants of the city with the Barcelona Agenda 2030. Barcelona currently ranks as the 49th most sustainable city in the world out of 100 global cities (Arcadis, 2022). The ranking is based on the people, planet, and nature aspects following the sustainable development goals of the UN. 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 increase of greenness in the neighborhood Eixample with the Eixos Verds Plan increases the mental health of 30.000 inhabitants. The same conclusions were stated by the work of Triguero-Mas et al. (2015) in Catalonia finding a relationship between the self-perceived general and mental health of people and the surrounding by and access to green spaces.

Sustainability is also becoming a topic of increasing importance in the real estate market. For example, in the U.K., 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 (Banco Santander, 2023; BBVA, 2022). It is driven by a lower expected risk and costs of living for borrowers after sustainable investments (World Green Building Council, 2022). Sustainability is also stimulated by local policies. For instance in Barcelona with Pla de Barris that includes plans for all of the 73 subdistricts to improve social, economic, and urban conditions (Ajuntament de Barcelona, 2023a). With the higher attention to sustainability, a "brown" discount is developing for real estate that does not meet the "green" market expectations according to Sam Carson head of sustainability, valuations, and advisory services at CBRE U.K. (Funds Europe, 2023). This discount is driven by the demand side. Tenants are willing to pay higher prices for more sustainable properties (D. Worford, 2022).

Home buyers that are more sustainable aware are willing to pay more for housing (Mandell & Wilhelmsson, 2011). In Barcelona, for instance, related to the condition of properties itself, prior research already provided evidence that higher energy labels are positively correlated with housing prices (Chen & Marmolejo Duarte, 2018; Dell'Anna et al., 2019). For locational sustainable factors is shown in Barcelona for instance that better access to public services and amenities, higher perceived security, shorter distance to the seashore, shorter distance to a highway, and closeness to central business districts increases significantly the 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, a significant negative relationship is found between better access to parks and gardens and shorter commuting times and the willingness to pay for housing (Dell'Anna et al., 2019; Graells-Garrido et al., 2021; Marmolejo-Duarte & Chen, 2022). Although, this is mainly caused by the city structure of Barcelona where parks and gardens are mostly located at the periphery and the low commuting times compared 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 our research the price impact of sustainability on housing prices is studied from five different dimensions: ecological, environmental, social, cultural, and economic-financial as proposed by Kauko (2019). Semi-log hedonic pricing models are estimated that include a high number of sustainable pricing factors. The hedonic pricing model assumes that the price paid for housing is equal to the utility buyers are expected to extract from it. The semi-log hedonic model ensures that pricing factors can be interpreted as semi-elastic. The coefficients will represent the effect in percentages of the change in the housing prices of a one-unit increase in the pricing factor rather than monetary values. In the model, potential multicollinearity is prevented by the construction of PCA components if a high correlation between sustainable variables is observed. To provide insight into which houses are positively and negatively impacted by sustainable pricing factors in the city of Barcelona the results are visualized by demonstrative maps. These maps show the total price impact of selected sustainable factors on the predicted housing price. Hereby, the maps make it possible to identify the difference in the price impact of sustainability between different areas in Barcelona. Thereby local clustering tendencies of sustainability pricing can provide opportunities for local policy intervention, which could stimulate a more equal/fairer pricing of sustainability for housing in Barcelona. Additionally, to stimulate further research the code to construct the maps is shared on GitHub including a wide range of options to visualize the results of the valuation models.

The remainder of this research is structured as followed: section 2 includes the literature review. Section 3 includes the data and methodology. In respectively sections 4 and section 5, the results and their robustness are discussed. The conclusion and discussion of the research are included in section 5. Lastly, in sections 6 and 7 the reference list and appendix are included.

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 economic-financial. The five dimensions will be analyzed in our empirical analysis of the 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 property by different transportation options (Kauko, 2019). The price effect of energy labels on housing prices is the most studied. The certification of properties with an energy label has been mandatory for residential housing for member states of the European Union since 2009 (The European Parliament & The City Council of the European Union, 2002). In most research, a positive relationship between an increase in the energy label and the willingness to pay by homebuyers is reported. For example, in international work evidence is provided in the Netherlands, Germany, and England by respectively Brounen & Kok (2011), Cajias et al (2016), and Fuerst et al. (2015). Specifically to Spain, Ayala et al. (2016) found this premium in different regions with energy labels which were estimated by surveys filled in by property owners. Also, La Paz et al. (2019) found this positive premium in Alicante. Although, 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). However, a lot of observations were excluded from the sample caused by missing energy labels. To address this issue, Chen & Marmolejo Duarte (2018) estimated a Heckman selection model. The results of the models that corrected for the sample selection bias increased the found willingness to pay for housing from 0.9% to 2% for each level increase in the energy label compared to a model that did not correct for the sample selection bias. But in future work, Marmolejo-Duarte & Chen (2022) reconsidered these findings by showing that the premium for energy labels became an insignificant pricing factor if additional variables related to the architectural quality of properties were included in the model. Hereby, Marmolejo-Duarte & Chen (2022) argue that earlier findings are caused by the presence of an information asymmetry in the pricing models.

The accessibility of an area by both short and long-commuting 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 they also found a significant negative relationship between the closeness to the bus and the housing price. 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. However, in Barcelona, Graells-Garrido et al. (2021) did not find evidence for a relationship between the accessibility to the bus, metro, and shared bicycles and housing rents.

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. The same results were also found in Spain by Taltavull de La Paz et al. (2019). In line, Zhang et al. (2016) found in China that an improvement in the rail network results in higher housing prices. They found a 0.023% increase in the housing price with each 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 and housing prices. This is mainly caused by a tradeoff between the higher accessibility to the highway and the negative externalities of living close to 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 effect of the closeness to a highway on the housing price even as Del Anna et al. in Barcelona (2019).

Section 2.2: Environmental Dimension of Sustainability

Research that focused on the environmental dimensions covered the accessibility, amount, and view of natural areas, such as the beach or parks. For the beach, a higher willingness to pay for housing is found for properties closer to the beach. For example, Dell'Anna et al. (2019) found in Barcelona a significant negative relationship between the willingness to pay for housing and distance to the seashore.

Also, Marmolejo-Duarte & Chen (2022) observed a positive price premium for residential real estate located within 200 meters of the seashore.

This higher willingness to pay for housing is also often found in prior literature for the accessibility to parks and gardens. 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 housing prices. These same conclusions were stated in China by Cui et al. (2018) and in Germany by Brandt et al. (2014). However, for Barcelona, Dell'Anna et al. (2019) found contrary results with a significant positive relationship between the distance to parks and housing prices. They explained that it is caused by the structural design of Barcelona where parks are mainly located at the periphery. Instead of the closeness, another environmental factor is the view on nature. This is also found in prior work to increase the 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. The visual basin was 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 focussed mainly 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. These often reported a positive relationship between better accessibility and willingness to pay for housing. 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 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, educational facilities, financial services, government services, professional services, and recreational and healthcare facilities by Graells-Garrido et al. (2021). Although, Graells-Garrido et al. (2021) only studied the Spearman rank correlation between these factors and housing rents. Hereby they did not correct for the price effect that other variables have on the willingness to pay for housing.

Concerning demographic factors, an often-studied price factor is the population density of a neighborhood. Hereby are found mixed results depending on the country in prior research. 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). In Spain, this positive relationship between population density and the willingness to pay for housing was also found for cities in the north, south, and center by Ayala et al. (2016) and in Alicante by Taltavull de La Paz et al. (2019).

The safety of the neighborhood is another often considered pricing factor. Prior work mostly reported a higher willingness to pay for housing when neighborhoods have a (considered) higher safety. For example, Ceccato et al. (2020) found in Sweden that the closeness to crime hotspots had a significant impact on the prices paid for single-family houses. In Barcelona, Buonanno et al. (2013) found that the perceived security had a significant positive and the crime perception rate had a significantly negative relationship with the willingness to pay for housing.

The last often-studied demographic factor is the population growth of an area. Population growth is a combination of natural population growth and the net immigration rate. In theory, higher population growth increases the demand for housing. In the U.S., Jeanty et al. (2010) found that this demand increased housing prices. However, this increase in housing prices will in the longer term result in lower population growth. In Barcelona, Buonanno et al. (2013) only studied the immigration rate. They found that an increase in the immigration rate significantly increases housing prices.

Section 2.4: Cultural Dimension of Sustainability

Empirical work that studied the impact of the cultural dimension of sustainability on the willingness to pay for housing focused mostly on the premium paid for buildings that are monumental or are located close to monuments and other places of recognition of religion and culture. The research often 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 with a monumental status were 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%.

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 the housing price. In Barcelona, also a significant positive correlation is found by Graells-Garrido et al. (2021). Additionally, Graells-Garrido et al. (2021) found a significant positive correlation between the willingness to pay for housing and the accessibility to entertainment places. In these entertainment places, often current and past cultural values are expressed and reflected.

Section 2.5: Economic-Financial Dimension of Sustainability

The economic-financial dimension of sustainability as proposed by Kauko (2019) is often studied by the effect of income equality and welfare of a neighborhood on housing prices. Higher-income equality in an area has often been associated with higher housing prices in prior research. For example, Chen & Marmolejo Duarte (2018) found in Barcelona that the cumulative percentage of people in higher socioeconomic classes in a subdistrict is significantly positively correlated with the willingness to pay for housing. The same conclusion was stated in future work by Marmolejo-Duarte & Chen (2022) when evaluating the cumulative percentage of people with high occupational positions in a neighborhood as an income equality measure.
Subject to the welfare of an area, higher welfare is often found to result in increasing housing prices. For example, Bruyne et al. (2013) found in Belgium that a lower unemployment rate had a positive relationship with housing prices. This is supported in later work by findings of Eicholtz et al. (2013) and Cajias et al. (2016) in respectively the U.S. and Germany. In agreement, a higher average income was found by Brandt et al.(2014) and Taltavull de La Paz (2019) to result in significant increases in housing prices in respectively Germany and Alicante, Spain. Similarly, higher economic activity in an area was found to be positively correlated with housing prices by Mandell et al. (2011) in Sweden. In line, in Barcelona, Marmolejo-Duarte et al. (2022) found that closeness to a place with high economic activity (CBD) increases the willingness to pay for housing.

In summary, prior literature almost always finds evidence for a positive relationship between housing prices and sustainable variables. This is found for all five dimensions of sustainability as proposed by Kauko (2019): ecological, environmental, social, cultural, and economic-financial.

Hereby for each dimension of sustainability, an increase in the willingness to pay for housing can be expected when the factors associated with the respective sustainable dimension increase. This is defined by hypothesis 1-5, stating that for each hypothesis an increase in the studied sustainable dimension results in a higher willingness to pay for housing. For the ecological dimensions, the price effect of the energy label is excluded from the hypothesis because it is the only sustainable variable that is not location-bounded. Additionally, related to the economic-financial dimension found prior work a negative relationship between fairer income equality, and a positive relation of welfare on housing prices. In line, hypothesis 5 proposes both a positive effect of higher welfare and a negative effect of higher income equality on housing prices. All of the in-depth studied sustainable variables are location-bounded. Thereby, a hypothesis is defined that the pricing impact of sustainability on housing prices shows local clustering tendencies between properties.

> **H1:** An increase in the ecological dimension of sustainability, excluding the effect of energy labels, increases the willingness to pay for housing.
>
> **H2:** An increase in the environmental dimension of sustainability increases the willingness to pay for housing.
>
> **H3:** An increase in the social dimension of sustainability increases the willingness to pay for housing.
>
> **H4:** An increase in the cultural dimension of sustainability increases the willingness to pay for housing.
>
> **H5:** A higher in welfare and income equality for the economic-financial dimension of sustainability respectively increases and decreases the willingness to pay for housing.
>
> **H6:** The total price impact of sustainability on housing prices 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 in respectively sections 3a and 3b. Two samples are retrieved. One that includes the observations with the missing energy labels and one that excludes the observations with missing energy labels. The valuation models are described in section 3c and the construction of the demonstrative maps to visualize the results is described in section 3d.

Section 3.1: 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 retrieved dataset used in this research includes all the housing advertisements in Barcelona on Idealista on 17 April 2023. The housing advertisements 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 markets. The residential real estate market for Barcelona fulfills this requirement by being stable over the last years with a relatively consistent increase in housing prices (Idealista, 2023b). Besides, the housing prices also 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.
1^{st} Floor	Dummy variable, is equal to 1 if the property is located on the 1^{st} floor, else 0.
2^{nd}-5^{th}Floor	Dummy variable, is equal to 1 if the property is located on the 2^{nd} - 5^{th} floor, else 0.
6^{th} Floor or higher	Dummy variable, is equal to 1 if the property is located on the 6^{th} 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 A, else 0.
Energy Label Consumption B	Dummy variable, is equal to 1 if the energy label consumption is B, else 0.
Energy Label Consumption C	Dummy variable, is equal to 1 if the energy label consumption is C, else 0.
Energy Label Consumption D	Dummy variable, is equal to 1 if the energy label consumption is D, else 0.
Energy Label Consumption E	Dummy variable, is equal to 1 if the energy label consumption is E, else 0.
Energy Label Consumption F	Dummy variable, is equal to 1 if the energy label consumption is F, else 0.
Energy Label Consumption G	Dummy variable, is equal to 1 if the energy label consumption is 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 in our research are treated by both assuming similarity between properties in a neighborhood and the exclusion of observations. The building year was missing in approximately 4900 (30%) of the advertisements. If the building year is missing it is in our sample set equal to the median of the subdistrict in which the residential property was located. Floor-level data was missing in approximately 2100 (13%) of the housing advertisements. The observations with missing floor-level data are excluded from the original sample. The last variable that was often missing was the consumption energy label. The energy label was not reported for approximately 3500 (25%) of the residential properties 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. The sample excluding observations with missing energy labels is used to test the robustness of the reported results. But as argued by Chen & Marmolejo Duarte (2018) missing values for the energy labels are often the result of a sample selection bias. The energy label might be not reported because it is expected to have a negative effect on the housing price. Thereby in the pricing models, which do not include the observations with missing energy label data, the strength of the positive/negative paid premium for higher/lower energy labels will be underestimated compared to the reference category. The use of the Heckman selection model solves the sample selection problem by estimating two regressions (J.J. Heckman, 1976). Firstly, the probability of the presence of an energy label in housing advertisements is estimated with a Probit model using energy consumption-related housing-specific variables as predictors. The predicted probability of the presence of an energy label by the Probit model is converted to an Inverse-mill ratio. Secondly, the valuation model is estimated with the log housing price as the outcome variable and the housing-specific, included sustainable variables and the inverse mill ratio as explanatory variables. With this method, the observations with missing energy label data can be included in the sample. So the sample selection bias is avoided. However, for completeness results without the observation of the missing energy labels are discussed in the robustness section to identify potential errors in the pricing models. The only downside of the Heckman selection model in this research is the low pseudo r-squared of 2%. This can be caused either by the absence of a sample selection bias or missing data related to relevant indicators for the decision to report the energy label in the housing advertisement.

Besides observations with missing data, a negative building age of minus 1 is observed for four observations of residential properties that are still under development. Therefore, we excluded these observations from the sample. Furthermore, the building surface m² has two observations with a value of zero. However, in the housing advertisements, the specified building surface m² is 1.000. We replaced these values for the building surface m² with the correct values.

We identify outliers for the continuous housing-specific characteristic variables for the log housing price, building surface in m², building age, and housing price per m². The outliers are for both samples identified before the exclusion of the observation with missing energy labels. The variable housing price per m² is not selected as both predicting or explaining variables in any of the models. However, this variable causes a highly non-linear relationship between the log housing price and building surface in m². The relationship cannot be captured in the pricing models and thereby creates an incorrect model if the observations are not excluded. This makes it necessary to include it as the variable for identifying outliers. 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

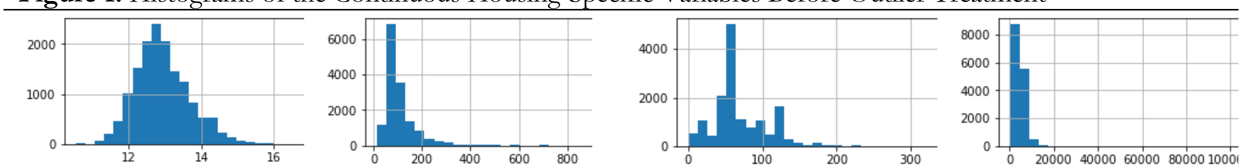


Fig 1a: Histogram of the Log Price

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

We excluded the observations that have a value for one of these variables shown in the histograms that is 2 standard deviations below or above the mean to deal with outliers. This is equal to a confidence interval of 95% under the Gaussian normal distribution. 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 we excluded approximately 400-600 (2.5%) of the observations. The continuous housing-specific variables after the outlier treatment represent a more normal distribution as shown in figure 2.

Figure 2: Histograms of the Continuous Housing-Specific Variables After Outlier Treatment

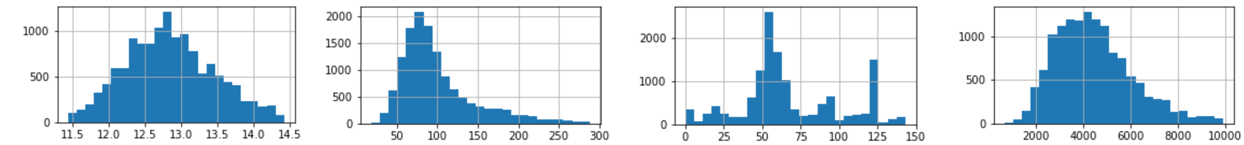


Fig 2a: Histogram of the Log Price	Fig 2b: Histogram of the Building surface m^2	Fig 2c: Histogram of the Building age	Fig 2d: Histogram of the Price per Building surface m^2

The final sample for the housing-specific characteristic sample including observations with missing energy labels consists of 13.358 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 including observations with missing energy labels are shown in table 2. 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. Hereby, the continuous variable shows a high similarity for most of the 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 labels.

Table 2: Summary Statistics of the Continuous Housing-Specific Variables for the Sample Including Observations with Missing Energy Labels

Variable	**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 including observations with missing energy labels.

Table 4 shows that the housing-specific dummy variables for the sample including observations with missing energy labels are having a good variety. The house dummy variable and energy consumptions label A have the lowest presence. However, the 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 variable (81.82%) for the sample including observations with missing energy labels. Table 5 included in the appendix for the housing-specific dummy variables of the sample excluding observations with missing energy labels shows the same characteristics. The only 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 Including Observations with Missing Energy Labels

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 including observations with missing energy labels.

Section 3.2: Sustainability Data

The data for the creation of the sustainability variables, except the energy labels, are both retrieved from the City Council of Barcelona (Ajuntament de Barcelona, 2023) and OpenStreetMap (Openstreetmap Contributers, 2023). The City 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. First of all, the dataset should include at least subdistrict-specific information. Secondly, the dataset should not have a too high correlation with other used sustainable data. Altogether, the sustainable information retrieved from the City Council of Barcelona covers a wide range of aspects of sustainability consisting of statistical data such as the unemployment rates, population density, vulnerability to a heat impact, social cohesion rate, and locational data of public amenities/services. The only important sustainable variables, which are not reported in any of the datasets are the location of the bus stops, highways, train stations, and beaches.

These variables are found in earlier research to have a significant impact on the price of residential estate in Barcelona (Chen & Marmolejo Duarte, 2018; Dell'Anna et al., 2019; Marmolejo-Duarte & Chen, 2022). Therefore this information is retrieved by the Overpass API (Overpass Turbo, 2023). The Overpass API reports information on the location of these variables 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 from which the variables are retrieved, and a short description of the data source are included in table 6 included in the appendix.

Our retrieved sustainable information from the City Council of Barcelona and OpenStreetMap consists of three types of data. Dependent on the type the data is converted into variables containing sustainable information. First of all, we retrieved data that contained statistics on a subdistrict level, such as the unemployment rate, and p80/p20 income distribution. If for these subdistrict level statistical data multiple countings are done during the last reported year, we used the average value. Additionally on Idealista, a lower number of subdistricts are specified compared to the dataset of the City Council of Barcelona. Idealista merges some of the neighboring subdistricts of which information is individually reported by the City Council of Barcelona[1]. For the subdistricts which are presented as one subdistrict on Idealista, we either calculated the weighted average of the statistic by the number of inhabitants or we used the sum of the statistic. The statistical subdistrict-level variables contain no outliers since the reported numbers are the average numbers of the subdistricts. Secondly, we retrieved data that contains geographic information with benchmark scores of an area compared to other areas in Barcelona.

[1] Idealista specifies 69 subdistrict in Barcelona and the City Council of Barcelona specifies 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 on 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

The location of the areas are included as multiline strings in the datasets. We matched these with the location of the properties. The location of the properties is on Idealista specified by the latitude and longitude. We observed missing values for the observations where the longitude and latitude of a property did not match with any of the multiline strings. We replaced these missing values with the median value of the subdistrict. No outliers are identified in the datasets with geographic information with benchmark scores since the data is scaled in pre-defined levels. Thirdly, we retrieved data that contains information about the location of sustainable variables, such as the location of bus stops, parks & gardens, or universities. We calculated the distance of the property to each of these locations of the sustainable variables included in the dataset. This distance 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 sustainable variable. With this data we could calculate the distance to the nearest location of the variables and when relevant the number of locations of the variable within a specified range. The specified range depends on the expected distance within the presence of a location of the variable that is expected to add utility for a homeowner. For example, related to a hospital, we reasoned that is expected to provide utility to a homeowner if it is present within a range of 1 kilometer of the property. However, we reasoned that for a bar or restaurant, it 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. With the truncation, we make sure that the calculated variables retrieve a better normal distribution. Furthermore, does it support our reasoning that the presence of an amenity does not provide additional utility for a homeowner 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.

We constructed PCA components for the sustainable variables that have a high correlation with each other. This is done to avoid high multicollinearity in the valuation models. These sustainable variables often contain information on similar types of amenities. Oladunni & Sharma (2016) found that this method is suitable 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. To ensure comparability we used a constant scale during the creation of the PCA components with a distribution from 0 to 1.

The created PCA components are mostly a mix of the distance to the nearest amenity and the number of amenities within a certain range of one or multiple amenities that belong to a similar category. Thereby, most of the created PCA components represent an accessibility/access to indicator for the included variables. We ensured that the minimum distance to coefficients has negative signs and that the number within the prespecified range has positive coefficients. So a higher score for the PCA complements represents higher accessibility. This offers interpretation benefits for the results of the pricing models. The only PCA components, which we did not create based on features with distance information are the income distribution and income & unemployment PCA components. We created these PCA components based on the economic statistical data of the subdistrict. We scaled these economic statistics used to construct the PCA components beforehand. So they retrieve similar coefficients during the creation of the PCA component[2]. Additionally, we included the squared value of the statistic when creating the PCA components. With this, we capture the non-normal distribution of income equality and welfare statistic. Our results for the economical statistical PCA components have the interpretation that a higher value implies higher income equality or welfare for a subdistrict. An overview of the created PCA components, the explained variance of and weight given to the input features, and the truncation of these features are included in table 7. As shown in the table, the lowest share of the variance that is explained by a PCA component is approximately 72.25% of the input features. We found this lowest share of explained variance for the PCA components that represent the highest number of variables. We cannot split these PCA components into multiple components. Splitting the PCA component introduces multicollinearity in the pricing models.

Table 7: Information on the Construction of the Sustainable Pricing Factors

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]

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

Variable	E. Var %	Feature	Coef.	Trun.
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 restaurants within 0.25 km	0.072	[0,4]
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 hospitals 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

Variable	E. Var %	Feature	Coef.	Trun.
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 pricing factor (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 of the sample that includes the observations with missing energy labels is included in table 8. The summary statistics for the sustainable variables show a high variety based on the standard deviation, 25%, and 75% quantile. This suggests that we can observe high differences between properties and the price impact of the sustainability factors on housing prices. Both the results for the sample with the inclusion and the exclusion of observations with missing energy labels, as shown in table 8 and table 9 in the appendix, are similar.

Table 8: Summary Statistics of the Sustainable Variables for the Sample Including Observations with Missing Energy Labels

Variable	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	1	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 including observations with missing energy labels.

Section 3.3: Valuation Models

The housing-specific and sustainable variables used in the valuation models are included in table 6. The housing-specific characteristics, $\beta_1 - \beta_{22}$, contain variables specifically bounded to the residential properties. These are for example the floor area, building year, and condition of the residential property. The energy consumption label was chosen over the energy emission label because it is more directly associated with the living costs for home buyers. Moreover, the energy consumption label is more often reported in housing advertisements and it 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 economic-financial as proposed by Kauko (2019). The pricing factors for the ecological dimension, $Z_1 - Z_2$ include information about the access to transport options from a property for short (bus & metro) and long (highway & train) commutes. We included access to short commuting options as one sustainable variable since there is a high correlation between bus and metro accessibility. We included the access to long commuting options as one variable since there is a trade-off between the distance to the highway and train in Barcelona. We excluded the energy consumption label from the ecological dimensions of sustainability and included it in the housing-specific characteristic. It is the only sustainable variable that is not location but property-bounded. Also, it is impossible to increase the sustainability score of the energy label by local policy intervention in contrast to other sustainability factors. Policy intervention for energy labels will probably consist of support programs available to every property 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, 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, $Z_8 - Z_{18}$, includes information related to the safety, and accessibility of public amenities and demographics of the neighborhood.

The cultural dimension $Z_{19} - Z_{21}$ contains information concerning the recognition of religion and culture in the area. This is captured by 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 economic-financial dimension, $Z_{22} - Z_{23}$ includes pricing factors containing information about the income distribution and the level of welfare of an area.

Additionally, are dummy variables included, to capture the effect of other geographical pricing factors on the district level that are not included in the model. The inclusion of the dummy district variables decreases the probability that our estimated valuation models will be subject to an omitted variable bias. Hereby, we decrease the probability of potentially wrongly specified pricing models and an overestimated strength of the coefficients of the sustainable variables.

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

Housing-specific Pricing Factors	**Sustainable Pricing Factors**
$\beta_1 Building\ surface\ m^2$	$Z_1\ Bus\ \&\ Metro\ PCA$
$\beta_2\ sq(Building\ surface\ m^2)$	$Z_2\ Distance\ to\ Highway/Train\ (km)$
$\beta_3\ Building\ age$	$Z_3\ Distance\ to\ Beach\ (km)$
$\beta_4\ sq(Building\ age)$	$Z_4\ Park\ \&\ Garden\ PCA$
$\beta_5\ House\ (YES = 1, ELSE\ 0)\ (REF = Ground\ Floor)$	$Z_5\ Viewpoint\ PCA$
$\beta_6\ Mezannine\ (YES = 1, ELSE\ 0)\ (REF = Ground\ Floor)$	$Z_6\ Neighboordhood\ size\ (10\ ha)$
$\beta_7\ 1^{st} Floor\ (YES = 1, ELSE\ 0)\ (REF = Ground\ Floor)$	$Z_7\ Vulnerable\ to\ heat\ impact\ (1-5)$
$\beta_8\ 2^{nd} - 5^{th}\ Floor\ (YES = 1, ELSE\ 0)\ (REF = Ground\ Floor)$	$Z_8\ Police\ PCA$
$\beta_9\ 6^{th}\ Floor\ or\ higher\ (YES = 1, ELSE\ 0)\ (REF = Ground\ Floor)$	$Z_9\ Bar\ \&\ Restaurant\ PCA$
$\beta_{10}\ New\ housing\ Development\ (YES = 1, ELSE\ 0)\ (REF = Good\ condition)$	$Z_{10}\ Secundary\ \&\ Lower\ School\ PCA$
$\beta_{11}\ Needs\ renovation\ (YES = 1, ELSE\ 0)\ (REF = Good\ condition)$	$Z_{11}\ University\ PCA$
$\beta_{12} Elevator\ (YES = 1, ELSE\ 0)$	$Z_{12}\ Pharmacy\ PCA$
$\beta_{13}\ Terrace\ (YES = 1, ELSE\ 0)$	$Z_{13}\ Hospital\ \&\ Clinique\ PCA$
$\beta_{14}\ Heating\ (YES = 1, ELSE\ 0)$	$Z_{14}\ Big\ Shooping\ Place\ PCA$
$\beta_{15}\ Outdoor\ Facilities\ (YES = 1, ELSE\ 0)$	$Z_{15}\ Social\ Cohesion\ Score$
$\beta_{16}\ Parking\ space\ included (YES = 1, ELSE\ 0)$	$Z_{16}\ Natural\ population\ growth$ ‰
$\sum_{i=17}^{22} B_i Dummy\ Consumption\ Energy\ Label (= 1\ if\ label = energy\ label\ with\ \ Label\ D\ as\ reference\ category)$	$Z_{17}\ Net\ Immagration\ rate$ ‰
	$Z_{18}\ Density\ net\ (hab/1000\ ha)$
	$Z_{19}\ Performing\ Arts\ PCA$
	$Z_{20}\ Religious\ Institution\ PCA$
	$Z_{21}\ Museum, Library\ \&\ POI\ Cult. PCA$
	$Z_{22} Income\ Distribution\ PCA$
	$Z_{23} Income\ \&\ Unemployment\ PCA$
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 we 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 the marginal utilities of the housing-specific, sustainability, and dummy-district variables.

The hedonic pricing model or a similar models are often used in prior literature. This is supported by the general conclusion of researchers ten years ago that machine learning methods do at best equal the performance of the hedonic pricing model (Kauko, 2019). We estimated the hedonic pricing model as a semi-log model. The semi-log model offers some advantages over a linear model. The model helps to normalize the housing prices by taking the logs and thereby decreasing the impact of outliers. Furthermore, 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 our results and compare them with earlier work in this research field (Marmolejo Duarte & González Tamez, 2009). Furthermore, this model is also applied in earlier work on the price effect of sustainable factors on housing prices in Barcelona (Chen & Marmolejo Duarte, 2018; Dell'Anna et al., 2019; Marmolejo-Duarte & Chen, 2022). The semi-log hedonic pricing model is estimated concerning the different sustainable dimensions. We specified the models as followed:

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

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

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

Model 4: $ln(y) = \alpha + \sum_{i=1}^{22} \beta_i\ Housing - Specific\ Var + \sum_{i=8}^{18} Z_i\ Social\ Var + \sum_{i=1}^{10} X_i\ District\ Dummy + \varepsilon$ (4)

Model 5: $ln(y) = \alpha + \sum_{i=1}^{22} \beta_i\ Housing - Specific\ Var + \sum_{i=19}^{21} Z_i Cultural\ Var + \sum_{i=1}^{10} X_i\ District\ Dummy + \varepsilon$ (5)

Model 6: $ln(y) = \alpha + \sum_{i=1}^{22} \beta_i\ Housing - Specific\ Var + \sum_{i=22}^{23} Z_i\ Econ/Fin\ Var + \sum_{i=1}^{10} X_i\ District\ Dummy + \varepsilon$ (6)

Model 7: $ln(y) = \alpha + \sum_{i=1}^{22} \beta_i\ Housing - Specific\ Var + \sum_{i=1}^{23} Z_i\ Sustainable\ Var + \sum_{i=1}^{10} X_i\ District\ Dummy + \varepsilon$ (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 ε is the error term. A description of the variable is included in table 10. 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.

During the estimation for each model homoskedasticity of the residuals for our models was rejected by the Breusch-Pagan test at the 1% significance level. Thereby, to allow for heteroskedasticity we use robust standard (Huber-White) standard errors in the estimated pricing models (Guggisberg, 2019).

Assumptions for a correct specification of the pricing models are verified by calculating a variety of statistics both before and after the model estimation. These tests are included for each model in the shared code on GitHub for the estimation of the pricing models. Firstly, before the model estimation, the correlation matrix is calculated for the included predictors in the models to check for variables that have a too high correlation with each other. Variables included in the correlation matrix are only shown if they have a correlation coefficient of over 0.25 with another variable in the pricing model. As an overview, we included the correlation matrix for all the variables of both the sample that includes and excludes observations with missing energy labels in respectively figures 3 and 4 in the appendix. These correlation matrixes show no warningly high values for any of the variables and high similarity with each other. Secondly, before the model estimation, the variable inflation factor (VIF) is calculated for each variable. A high VIF value for a variable can represent the potential existence of multicollinearity in a model. Therefore if the VIF is too high there will be chosen to either exclude variables in the model or create additional PCA components. The results for the VIF statistics are discussed before the discussion of the model results. In line also after the model estimation, some statistics are reported. First of all, the skewness, kurtosis, and histogram are reported for the residuals of the models. The residuals are calculated by subtracting the actual log housing price from the predicted log housing price. The statistics are reported to verify the assumption of a normal distribution of the error term in the semi-log hedonic pricing model. Secondly, the Ramsey Reset test is calculated for the results of the model. The Ramsey Reset checks by the inclusion of the squared values of the predictors in the model for a correct model specification (Ramsey, 1969). The reported tests-statistic of the Ramsey Reset is a p-value for an F-test these squared values are not significantly different from zero. If this is true, these additional variables do not significantly increase the predictive power of the model. This Ramsey Reset test statistic for each model is discussed before the model results. Thirdly, is homoskedasticity of the error term tested by the Breusch-Pagan test (Breusch & Pagan, 1979). Homoskedasticity appears when the variety of the pricing errors is constant across all levels of the predictors. Heteroskedasticity is present when homoskedasticity does not hold.

As mentioned above, is in our research for all the valuation models homoscedasticity rejected at the 1% significance level. Thereby robust (Huber-White) standard errors are applied to the valuation models. The robust standard errors allow for heteroskedasticity in the pricing errors in the semi-log hedonic pricing models. The robust standard errors are also used in the Ramsey Reset test discussed above. Additionally is for completeness a correlation matrix shown between the squared residuals and the predictors in the model. This is also to check for misspecification of the pricing models. But in none of these correlation matrices, a high correlation is reported between the squared residuals and any of the predictors.

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

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

where α is the constant in the model, the $\beta's$ are the housing-specific pricing factors that could influence the decision to publish the energy label in the housing advertisement, and ε is the error term. Homoskedasticity by the Breusch-Pagan test is rejected for the Probit model. Therefore robust (Huber-White) standard errors are used in line with the valuation models. Additionally, the same tests for the model assumptions are conducted for the valuation models.

Section 3.4: Visualization of the Results by a Geographical Map

To display the results of the pricing models above we developed a map to visualize the price impact of sustainability. The map uses the latitude and longitude of the properties to visualize the results either on a subdistrict, neighborhood, or housing-specific scale. The map provides helpful insight given that the sustainable variables used in this research are mostly location-bounded.

Therefore is expected that the price impact of the sustainable pricing factors on a local scale are similar to each other.

We built the map with the use of Leaflet, which allows interaction with 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). We shared the code to build the map, even as the results of the estimated pricing models, on GitHub. In our shared code different parameters can be changed to build different types of maps. These parameters are described below. We construct in this research two demonstrative maps, which will be discussed in the results section. Furthermore are two demonstrative maps discussed in the robustness section. We visualize with these demonstrative maps the price impact of sustainability on housing prices both on an area-specific scale for two different sets of sustainable variables.

Our provided code to construct maps 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 of a variable that can be chosen to color the map. However, also general information on the residential property is provided when clicking on the icon of the observations. The general information includes information about the predicted price, listing price, and impact of housing-specific, sustainable, and district dummy pricing factors on the predicted price of the property. Besides, the option to display also general information on the total price impact of a list of selected sustainable pricing factors is included for two reasons. Firstly, given that some sustainable variables in the pricing models show non-positive relationships between higher sustainability and housing prices. Secondly, to exclude some sustainable variables that cannot be changed by policy interventions. An example of such a non-sustainable relationship between sustainability and housing prices is found for income equality and housing prices in prior research. An example of a variable that cannot be influenced by local policy is the distance to the beach. These variables can have therefore a lower relevance when visualizing the total price impact of sustainability factors on the housing prices in Barcelona.

More specific also information is provided for the individual housing-specific, sustainable, and dummy district variables. All the observations cannot be displayed at once given the large sample size. Hence we included three different optional methods to cluster the observations. The first option is automatic clustering of the observations by Leaflet.

With this method, we display information regarding the individual residential properties only when clicking on and opening clusters. With the second option (SVM_cluster), we cluster multiple observations into one observation. We do this by the use of the support vector machine algorithm (Scikit Learn, 2023). The algorithm clusters observations based on location by the latitude and longitude values of the properties. These clusters will be used to take the average of the sustainable variables of properties inside the cluster to calculate the price impact of sustainability. The downside of this method is that the information for the housing-specific and dummy district variables will be suppressed. With the third option (Subdistrict_cluster), the results can be clustered according to the 68 subdistricts in Barcelona on Idealista (Idealista, 2023a). The data for the area that is included in the subdistrict is retrieved from the City Council of Barcelona (Ajuntament de Barcelona, 2023b). These clusters will then contain information on the average value of the sustainable variables of all the properties in the sample that are located in the subdistrict. This makes it easier to conclude on a subdistrict scale in which neighborhood housing prices are higher/lower impacted by sustainable pricing factors. The downside of this method is that housing-specific information about the properties is suppressed.

The parameters, with a description, used to construct the maps are included in table 11. The parameters to construct the map offer high flexibility. We included this flexibility for the selected valuation models, the selected shown observations, the variable to color the map, the number of categories to color the map, and the clustering method that is used. The selected shown observations can be chosen by filtering the data based on the options equal to, not equal to, higher than, and lower than for the included parameters in the pricing models. The variables to color the observations/clusters are all the continuous variables included in the pricing models, in addition to the total price impact of housing-specific, selected sustainable, or sustainable variables. The number of categories can be based on the quantiles of the continuous variable used to color the map. We will always visualize the categories 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 our GitHub.

Table 11 The Parameters Used to Construct Maps to Visualize the Results of Pricing Models

Parameters	**Description**
Selected Sustainable Features	List including the 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 variables can be chosen that are included in the pricing model.
Map_save_name	The name of the map used in 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 was 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 used to estimate the pricing models.
Model_result	The estimated results of the pricing model.
Color_var	The variable that 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", and "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 observations 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. For 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.
Ref_group_dic	Dictionary, which contains information on the reference category for dummy variables. This dictionary is provided in the shared Github with the same name as the parameter.
N_clusters	The number of clusters that will be used to cluster the properties. Only relevant if the parameters: cluster = True
Lat_col	The name of the column in the dataframe contains information on the latitude of the properties.
Long_col	The name of the column in the dataframe contains information on 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
SVM_Cluster (True/False)	If specified true the residential properties will be clustered based on location by the Support Vector Machine.
Subdistrict_Cluster (True/False)	If specified true the residential properties will be clustered based on the 68 subdistricts specified by Idealista.
Save (True/False)	Save the demonstrative map on the computer.

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

Our first demonstrative map shows the total price impact of sustainability when the properties are clustered by the subdistrict in which properties are located. A screenshot of the demonstrative map is provided in the results section (4.2.1) and the HTML code is shared on GitHub.

Hereby the clusters which are in the highest total 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 total sustainability price impact will be displayed in black. The parameters used to construct demonstrative map 1 are included in table 12.

Our second demonstrative map discussed in this research will show the total price impact of selected sustainable features by clustering the properties into clusters based on the subdistricts. A screenshot of the demonstrative map is provided in the results section 4.2.2 and the HTML code is shared on GitHub. 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 the 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 Demonstrative Maps 1 and 2

Parameters	Demonstrative Map 1	Demonstrative Map 2
Selected Sustainable Features	Sustainable_predictors	[e for e in sustainable_predictors if e not in ['Distance to Beach (km)',' Neighborhood size (10 ha)',' Income Distribution PCA']]
Map_save_name	'Demonstrative_map_1'	'Demonstrative_map_2'
Title	'Total Price Impact Sustainability Variables Model 7 on Residential Properties'	'Total Price Impact Selected Sustainability Variables Model 7 on Residential Properties'
Subtitle	"Heckman Selection Model Barcelona"	"Heckman Selection Model Barcelona"
Legend_title	'Quantile Total Price Impact Sustainable Variables'	'Quantile Total Price Impact Selected Sustainable Variables'
Circle_Multiplier	N/A	N/A
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	{}	{}
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	N/A	N/A
Lat_col	'latitude'	'latitude'
Long_col	'longitude'	'longitude'

Parameters	Demonstrative Map 1	Demonstrative Map 2
Show_all (True/False)	True	True
SVM_Cluster (True/False)	False	False
Subdistrict_Cluster (True/False)	True	True

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

Additionally, we constructed two demonstrative maps that show the total price impact of sustainability on the predicted housing prices by clustering the houses in 100 clusters based on location with the support vector machine (by latitude and longitude of the properties). This is to test the robustness of the findings for hypothesis 6, which staties that the price impact of sustainability on housing prices in Barcelona shows local clustering tendencies. The only difference between these demonstrative maps is the clustering method. The sustainable variables to color the map are equal. The advantage of clustering by SVM is that the size of the clusters varies less compared to clustering by the subdistricts. The disadvantage of these demonstrative maps is that in practice also policies are implemented on a subdistrict level. Hereby, implementation of the results can be harder. The parameters for the demonstrative maps included in the robustness section are included in the appendix in table 13.

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, section 4.1 includes a subsection for the discussion of the findings of the Probit model in to predict the probability of the presence of an energy label in the housing advertisement.

Section 4.1: Results of the pricing models

Section 4.1 includes the Probit model to predict the probability of the presence of an energy label in housing advertisements. Besides that, includes it the results for the pricing models (models 1-7) estimating the log housing prices.

Section 4.1.1: Probit Model Heckman Selection:

The results for the Probit model are included in Table 14a. The results for the VIF test statistics, included in table 15 in the appendix, show that only the air conditioning dummy has a high VIF test statistic (12.90.). This can imply potential multicollinearity in the model. However, this 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 test (p-value = 0.24). The estimated model has a low pseudo R-squared (2%). This suggests that the included variables only have a small predictive power when explaining the presence of an energy consumption label. Indicating that the sample selection bias is smaller than in prior work. This is supported by the relatively low number of observations that are missing energy labels (25%) in comparison to earlier work in Barcelona (Chen & Marmolejo Duarte, 2018; Dell'Anna et al., 2019; Marmolejo-Duarte & Chen, 2022). That a low number of observations are missing is caused by more time that has passed since the obligation to report the energy label in housing advertisements.

The results of the Probit model show that both the building surface m^2 and building age have a positive relationship with the probability of the presence of an energy label. Although the coefficients are not significantly different from zero. The state of the property has a significant effect on the probability that the energy label is present. 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 the energy label is not reported in 60% of the housing advertisements for new housing projects. The significant negative coefficient for the needs renovation variable is likely caused by an expected low energy label. A property that needs renovation might have for instance thinner glass in the windows or an older door. This could be a reason to not report the energy label even though it is obligated. We observe coefficients in contrast to the expectations for the facilities of a property. 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 homeowners 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 predictors when estimating the probability of a missing energy label.

Altogether are the results different compared to those found in the work of Chen & Marmolejo Duarte (2018). Related to non-energy consuming related variables, Chen & Marmolejo Duarte (2018) found no significant coefficients for the building surface m², terrace, and property state variables, and a significant negative coefficient for outdoor facilities. Related to the energy-consuming features Chen & Marmolejo Duarte (2018) found a negative significant coefficient for the heating dummy variable and no significant coefficient for the air conditioning dummy variable. Additionally, Chen & Marmolejo Duarte (2018) included overall more features in the model. However, in our case, additional variables would not increase the explained variance by the model. Furthermore, we could not reason/found other housing-specific variables that have a high correlation with the presence of the energy label in the housing advertisement based on the retrieved housing-specific data from Idealista.

Section 4.1.2: Model 1: Housing-Specific Variables

Model 1, shown in table 14a, 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 statistic, included in the appendix in table 15, show that for none of the variables, the test statistic is higher than 6.5 except for the building surface m² and building age. The high VIF test statistic for the building surface m² and building age is mostly caused by the inclusion of the squares of the variables. The air conditioning dummy variable is excluded from the sample given the high observed VIF statistic (65). The Ramsey Reset test (p-value = 0.023) rejects the misspecification of the model at the 1% significance level. This p-value would be lower if the squared values of the building surface m² and building age were not included in the pricing model.

Overall, 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 as indicated by the significant positive coefficient. However, the strength of this relationship decreases when the building surface becomes higher as shown by the significant negative 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 caused by the higher probability of the presence of some outdated facilities/characteristics of a property that should be taken care of 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. When an apartment is exponentially older it has a higher probability of having historical value. Evidence for the pricing of this historical value in the residential market is for example found by Debrezaion et al. (2011) and Lazrak et al. (2014). Furthermore, show the results that a significantly higher price is paid for a house (15.9%) compared to an apartment on the ground floor. Additionally, the variables for the floor level show that the willingness to pay for housing in Barcelona increases when apartments are located on a higher floor. We observe an increasing pattern in the paid premium compared to the ground floor for the mezzanine (4.5%), 1st floor (11%) 2nd -5th floor (15.6%), and 6th floor or higher (21.4%). The observed increasing pattern is in line with earlier research about the housing market in Barcelona (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. 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. Hereby the probability that the model will overfit on a dummy variable with a low presence. Subject to the property state, no evidence is provided for a significantly higher willingness to pay for newly developed properties compared to the reference category (good condition). However, when a property needs to be renovated the results of the model show the presence of a significant discount (-17.2%). This discount is likely correlated with the costs that are expected for a renovation of the property. Moreover, 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%, 8.9%, 5.9%, and -4% for respectively the elevator, heating, parking space included, terrace, and outdoor facilities. The significant negative coefficient for the outdoor facilities coefficients is in contrast to the research on the Barcelona housing market (Chen & Marmolejo Duarte, 2018; Dell'Anna et al., 2019; Marmolejo-Duarte & Chen, 2022). They all found a significant coefficient for the presence of a swimming pool. In our sample, the outdoor facilities variable has a correlation coefficient of 94% with the swimming pool variable. The contrary findings are likely the consequences of recent dryness in Barcelona. This has resulted in policies to prevent high water consumption by inhabitants such as the prohibition to fill the swimming pools. Hereby the presence of a swimming pool does not add utility to home buyers. The policies to prevent high water consumption were not implemented and/or relevant at the time of earlier research (Ajuntament de Barcelona, 2023c).

However, another reason could be the higher cost of living when maintaining a garden, green area, and/or swimming pool. This could be higher than the associated monetary value of the expected utility of the facilities.

The energy consumption label dummy variables are the only housing-specific variables that are part of a sustainable dimension (ecological) (Kauko, 2019). The results show an increasing premium for properties with higher energy labels compared to lower energy labels. These findings are in line with earlier research in Barcelona (Chen & Marmolejo Duarte, 2018; Dell'Anna et al., 2019; Marmolejo-Duarte & Chen, 2022). However, the trend is not perfectly linearly increasing. But it shows a higher willingness to pay for properties with higher energy labels (A, B, C) compared to properties with lower energy labels (E, F, G) when using properties with energy label D as the reference category. 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%). Additionally, the IMR ratio is significant and negative. This indicates that the price of a property is negatively related to the probability that an energy label is shown. The findings are in line with the expectations of the sample selection bias by showing that energy labels are not reported in housing prices for a reason. Furthermore are the results and also the strength of the coefficients of the IMR ratio in-line with earlier work that addressed 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 14a, includes the variables related to the ecological dimensions of sustainability. The variables represent the accessibility to short and long commuting options. As proposed by Kauko (2019) the energy labels would also belong to the ecological dimension. However, we included the energy label in the housing-specific instead of sustainable pricing factors. Because the energy labels are not influenced by political decisions on the structure/quality 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%. That the inclusion of ecological variables is not introducing multicollinearity is shown by no newly found high VIF-test statistics as included in table 14 in the appendix. The Ramsey Reset test for misspecification is rejected at the 1% significance level for model 2 (p-value = 0.018). 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 coefficient for the bus & metro PCA component is insignificant. This result provides no evidence that an increase in access to short-distance commuting options increases the willingness to pay for housing. The insignificance is likely caused by the high accessibility of residential properties in Barcelona. For example, 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 in the sample. Also, for the metro station the accessibility is high with 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. They found an insignificant correlation between the housing rents and accessibility to the bus, metro, and bike pike-up places. However, Graells-Garrido et al. (2021) did not correct the impact of other variables on the willingness to pay for housing. This could make their findings 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 are in contrast with earlier work about Barcelona. 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 this variable is combined with the distance to the train. Moreover, uses Dell'Anna et al. (2019) the location of highway ramps, but this data is not provided by OpenStreetMap. Thereby in this research, the distance to any point on the highway is used.

In conclusion, we find that the result provides no evidence to reject hypothesis 1, an increase in the ecological dimension of sustainability, excluding the effect of energy labels, increases the willingness to pay for housing. The relationship between the access to and presence of bus and metro stations and willingness to pay for housing is insignificant. However, better access to the highway or train does increase the willingness to pay for housing.

Section 4.1.3: Model 3: Environmental Dimension of Sustainability

Model 3, shown in table 14a, includes the variables related to the environmental dimensions of sustainability. This dimension is captured by the distance to the beach, the accessibility to parks & gardens, the accessibility to viewpoints, the 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, included in table 15 in the appendix, show that the distance to the beach and vulnerability to heat impact environmental variables have high VIF values. These high values are caused by the correlation with the district dummies in the model. However, the standard errors of the coefficients are still relatively low. Thereby decreasing the likelihood that the observed coefficients are subject to a high degree of multicollinearity. Misspecification of the model is rejected by the Ramsey Reset test at the 5% significance level by the inclusion of the environmental variables (p-value: 0.068). 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 higher distance to the beach in kilometers significantly decreases the willingness to pay for housing. This is in line with earlier work that found evidence that housing prices in Barcelona are negatively correlated with the distance to the seashore (Dell'Anna et al., 2019; Marmolejo-Duarte & Chen, 2022).

In contrast, a significant negative relationship between the accessibility to green space and viewpoints and the willingness to pay for housing is found by the PCA components for parks & gardens, and viewpoints. This is in contradiction to earlier findings 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 by Dell'Anna et al. (2019). Dell'Anna et al. (2019) argue that this is caused by the structure of Barcelona. In Barcelona is chosen to locate the parks at the periphery. However, it does not imply that access to parks, gardens, and viewpoints is not important in Barcelona. For example, Triguero-Mas et al. (2015) found in Catalonia that better access to parks and green spaces has a positive influence on the (self-perceived) general and mental health. Additionally, show the results of a significant negative relationship between neighborhood size and housing prices. The results report evidence of a preference for housing in the smaller subdistricts.

The same negative relationship is found for the vulnerability to a heat impact. Residential properties in areas that are more vulnerable to a heat impact have lower housing prices. The asking price for a residential property decreases by 10.2% for each level that the vulnerability to heat impact increases. The price impact of these factors, the size of the neighborhood, and vulnerability to a heat impact are not addressed in research on housing prices in Barcelona.

Altogether, the significant negative impact of a higher distance to the beach, neighborhood size, and higher vulnerability to a heat impact does not provide evidence to reject hypothesis 2: an increase in the environmental dimension of sustainability increases the willingness to pay for housing. There is found a significant negative relationship between housing prices and access to and presence of parks & gardens and viewpoints. However, this effect is mainly caused by the structure of the city (Dell'Anna et al., 2019). There is chosen to locate parks and green spaces mainly at the periphery. Furthermore, the access to and presence of parks & gardens, and viewpoints has a positive effect on the mental of the inhabitants of Barcelona (Triguero-Mas et al., 2015; Vidal Yañez et al., 2023).

Table 14A: Semi-Log Pricing Model Results of the Sample Including Observations with Missing Energy Labels

	Probit		**Model 1**		**Model 2**		**Model 3**	
	Coef.	**Stderr.**	**Coef.**	**Stderr.**	**Coef.**	**Stderr.**	**Coef.**	**Stderr.**
Constant	0.352***	0.09	12.202**	0.12	12.404***	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
2nd -5th 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..115***	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	0.268***	0.08						
Parking space included			0.089***	0.01	0.097***	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
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		

	Probit		Model 1		Model 2		Model 3	
	Coef.	**Stderr.**	**Coef.**	**Stderr.**	**Coef.**	**Stderr.**	**Coef.**	**Stderr.**
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 14 includes the results for semi-log hedonic priding models of the sample including observations with missing energy labels. The results report both the coefficients (Coef.) and the standard errors (Stderr.) of the variables. *** denotes significance at the 1% level, *** denotes significance at the 5% level, and * denotes significance at the 10% level.

Section 4.1.4: Model 4: Social Dimension of Sustainability

Model 4, shown in table 14b, includes variables related to the social dimension of sustainability. The social dimension of sustainability consists of among others: the closeness to the police stations, the 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% in comparison 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, no variables with problematic VIF statistics are shown in table 15 in the appendix. The high VIF test statistic for the neighborhood density is caused by correlation with the dummy district variables. Additionally, overall higher observed VIF statistics for most social-related variables in contrast to the variables of other earlier sustainable dimensions are observed. This is mainly caused by the correlation between the access to the amenities/services variables. For example, bars and restaurants are in practice often located close to big shopping places. These VIF test statistics are not problematically high. Misspecification of the model is rejected in the Ramsey Reset test (p-value: 0.61). The results for the housing-specific variables are mostly in line with the findings of 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 low presence of the dummy variable in the sample (2.1%). Hereby the dummy variable can overfit on a low number of observations.

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. That access to healthcare has no impact on housing prices 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 rank correlation between variables and the housing rent.
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). The police PCA component might not perfectly capture the safety of a neighborhood but no other relevant data is provided by the City Council of Barcelona. The bar & restaurant PCA component has the largest positive coefficients (0.125) of the social PCA components. This shows that the presence of social connection places has a relatively large influence on the willingness to pay for housing in Barcelona. This significant positive correlation was also found by Graells-Garrido et al. (2021).
Mixed results are found for the educational PCA components. A significant negative relationship with the housing price is found for the secondary and lower education PCA component. In contradiction, a positive relationship with the housing price is found for the university PCA component. Earlier work on accessibility to education only reported significant positive price effects in Barcelona. However, a distinction was never made between the different types of education (Graells-Garrido et al., 2021).
Related to the accessibility of big shopping places a significant 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.

Related to the demographic social sustainable variables a significant negative relationship is found between an increase in the social cohesion, natural population growth, and density of the residential area on the willingness to pay for housing.

The social cohesion variable is based on the number of local shops, street markets, and fairs, and the number of neighborhood activities that could potentially cause social cohesion (Department of Urban Resilience, 2020). The findings for the negative relationship with the housing price might indicate that inhabitants of Barcelona prefer to live in neighborhoods where it is less crowded as measured by a lower degree of community activities. The social cohesion on housing prices in not analyzed in earlier work in Barcelona.

The preference for living in less crowded areas could also explain 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. Prior work reports that it both increases/decreases the willingness to pay for housing (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 is in line with the findings of Jeanty et al. (2010). They reported that after a high population growth in an area, people tend to move to other areas creating negative pressure on housing prices. The natural population of this research is represented by past data from 2019. Furthermore, could it further support the finding that inhabitants of Barcelona prefer to live in less crowded areas with a decrease in willingness to pay for housing if the neighborhood has been subject to a high(er) natural population growth. However, in contrast to the natural population growth, the relationship between the net immigration rate and housing prices is significant and positive. This is in line with the significant positive correlation found by Graells-Garrido et al. (2021) in Barcelona. However, the strength of the coefficient for the immigration rate is lower compared to the natural population growth. Overall, this suggests that an increase in population growth has resulted in a lower willingness to pay for housing in Barcelona.

In conclusion, we found that higher access to public services and amenities in general increases the willingness to pay for housing. Related to the demographic statistics is found that an increase in the crowdedness of an area and the growth of the population decreases the willingness to pay for housing. This provides no evidence to reject hypothesis 3, an increase in the social dimension of sustainability increases the willingness to pay for housing.

Although, we could argue that the significant negative price effect of better access to second and lower education, higher social cohesions score, and net immigration rate, reflect more unsustainable preferences in the willingness to pay for housing in Barcelona. On the other side, might it correctly reflect the preferences of home buyers, who are looking to live in more "quiet neighborhoods given the already busy life in Barcelona.

Section 4.1.5: Model 5: Cultural Dimension of Sustainability

Model 5, as shown in table 14b, 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 15 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 = 0.23). 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, which is even as for model 4 an insignificant positive price effect.

For the cultural-related sustainable variables, both positive and negative relationships with the willingness to pay for housing are found. The coefficient of the PCA components of performing arts is significant and negative. It implies that the accessibility to places of cultural expression by cinemas, theatres, and concerts provides negative utility to homeowners. This finding is in contradiction to the significant positive correlation with housing rents in Barcelona found by Graells-Garrido et al. (2021) when using a more general definition for entertainment paces. The negative relationship between access to performing arts and housing prices in our research might be related to the possible nuisance during big events. For example, traffic and noise were found to be negative pricing factors for housing prices in the Netherlands (Ossokina & Verweij, 2015).
In contrast, a significant positive relationship is found between the religious institutions PCA component and willingness to pay for housing. This finding is in line with the significant positive correlation with housing rents in Barcelona found in earlier work by Graells-Garrido et al. (2021).

A larger increase in the willingness to pay for housing is observed for the museum, library & POI cult. PCA component. The positive results for access to cultural places are in line with the provided evidence in the work by Lazrak et al. (2014). Lazrak et al. (2014) found in the Netherlands evidence of a spillover effect of monumental buildings on housing prices.

To sum up, in general an increase in the willingness to pay for housing when the culturally sustainable variables increase is reported in our model. This provides no evidence to reject hypothesis 4, an increase in the cultural dimension of sustainability increases the willingness to pay for housing. There is found a significant negative relationship between the PCA component for performing arts and willingness to pay for housing. But on the other hand, there is found a significant positive relationship between the PCA components for religious institutions and museums, libraries, and POI cult. and the willingness to pay for housing in Barcelona.

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

Model 6, as shown in table 14b, includes the variables related to the economic-financial dimension of sustainability. To capture the economic-financial dimension, variables are included for the income distribution and welfare of the subdistricts. The inclusion of the variables increases the R-squared by 2.2 percentage points to 84.8% compared to the pricing model that only included housing-specific variables (model 1). Table 15 in the appendix shows that the economic-financial variables have high VIF-test statistics. This is mainly caused by the correlation between welfare and income distribution variable and the correlation of the variables with the district dummies. However, the standard errors of the economic-financial variables remain low in comparison to the coefficients. Hereby it implies that there is 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. This is already tried to be captured by the inclusion of the squared variables in the construction of the PCA components. The results for the housing-specific variables are in-line with the findings of the earlier pricing models. The only difference is in agreement with models 4 and 5 that the coefficient of the energy consumption label A is insignificant in the pricing model.

In specific to 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 the income equality and the willingness to pay for housing. The relationship implies that home buyers who can afford more expensive houses tend to cluster in certain subdistricts in Barcelona driving up housing prices. This has an increasing effect on income equality measured by the p80/20 income distribution and Gini index. The findings are in line with the work of Chen & Marmolejo Duarte (2018) and Marmolejo-Duarte & Chen (2022) in Barcelona, who found respectively that the cumulative of people in high socioeconomic classes and high occupational positions have a significant positive impact 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. The income & unemployment PCA component is based on the average household income and unemployment rate. The findings are in line with reported findings in other countries by Mandell et al. (2011), Eicholtz et al. (2013), Cajias et al. (2016) internationally, and in Spain, Alicante, by Taltavull de La Paz et al. (2019).

As a result, we find no evidence to reject hypothesis 5, a higher welfare and income equality for the economic-financial dimension of sustainability respectively increases and decreases the willingness to pay for housing. We acknowledge that 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.

Section 4.1.7: Model 7: Every Dimension of Sustainability

Model 7, shown in table 14b, includes every dimension of sustainability: ecological, environmental, social, cultural, and economic-financial in the pricing model. The inclusion of every dimension increases the R-squared 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 15 in the appendix, show that there are an additional number of variables introduced with high VIF test statistics.

This is mainly caused by the inclusion of a high number of predictors a relation to some local/area characteristics. Thereby some coefficients for sustainable variables could potentially to some degree be impacted by multicollinearity. It makes it 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 results for the Ramsey Reset test provide evidence that the model is correctly specified at the 5% significance level (p-value 0.051). The findings for the housing-specific variables are mostly in line with those of the other discussed models. In contrast to the findings for models 4, 5, and 6 the coefficient for the energy consumption label A is significant in model 7 even as in models 1, 2, and 3. The non-significance of the energy consumption label A in models 4, 5, and 6 was in contrast to other models and earlier research(Chen & Marmolejo Duarte, 2018; Dell'Anna et al., 2019). Furthermore, the housing-development dummy variable becomes significant at the 5% confidence level in contrast to the findings of all the other models.

The findings for the sustainable variables of each dimension of sustainability are compared for model 7 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 in model 7 are showing evidence of a negative relationship between better access to short-distance commuting options and housing prices. The findings for the minimum distance to the highway/train are in agreement with the results of model 2 with a significant and negative relationship. However, 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. Although the coefficient of the park & garden PCA component has become insignificant. This coefficient of the park & garden PCA component was significant at the 10% level in model 3. In addition the strength of the price effect of the viewpoint PCA component and vulnerability of heat impact level almost halved.

Thirdly, concerning the social dimension of sustainability, the coefficients for the strength of the price impact for most of the socially sustainable variables have become less or even insignificant compared to 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 with a positive effect on the housing price. Related to the other social variables the significance of the coefficients remained the same. But the strength of the coefficients at least halved for the bar & restaurant PCA component, big shopping place PCA component, social cohesion, natural population growth, and residential population density.
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 relationships are similar to model 5. The relationship between the cultural sustainability variables and willingness to pay for housing is equal. However, the price effect of the religious institution PCA component has become insignificant. Moreover, the strength of the other cultural sustainable variables, performing arts and museum, library, and POI culture PCA components, have at least halved. This is likely caused by an omitted variable bias in model 5.
Lastly, the economic-financial sustainable variables also have in model 7 a significant relationship with housing prices. However, the strength of the relationships decreased with the inclusion of other sustainable dimensions. This decrease in strength is approximately 200% for the income equality PCA component. This decrease is only approximately 33% for the income & unemployment PCA component compared to the findings for model 6.

In conclusion, the one-dimensional sustainability pricing models and the pricing models that include all dimension show both that higher sustainability increases the willingness to pay for housing. This finding holds for all five dimensions: ecological, environmental, social, cultural, and economic-financial as proposed by Kauko (2019). However, there is observed a change in the strength coefficients and the significance for some sustainable variables in the multi-dimensional sustainability model compared to the one-dimensional sustainability model. It suggests that models, which only include one dimension of sustainability will suffer from an omitted variable bias by not including pricing factors related to other dimensions. On the other side, the inclusion of a high number of sustainable variables makes the model likely to some degree subject to multicollinearity. This is for example shown by higher VIF test-statistic values for the variables in model 7. However, the found relationships between the housing prices and the sustainable variables are mostly similar to those for the individual models.

The lower strength of the coefficients in model 7 is compared to the one-dimensional sustainable models showing that the model does not overestimate the price effect. Thereby model 7 is the most conservative model concerning the strength of the price effect of the different sustainable factors by reporting lower coefficients compared to the one-dimensional models.

Table 14B: Semi-Log Hedonic Pricing Model Results of the Sample Including Observations with Missing Energy Labels

	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	12.049***	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	0.208***	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	0.101***	0.01	0.099***	0.01
Outdoor facilities	-0.033***	0.01	-0.039***	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	0.022***	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
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

	Model 4		Model 5		Model 6		Model 7	
	Coef.	**Stderr.**	**Coef.**	**Stderr.**	**Coef.**	**Stderr.**	**Coef.**	**Stderr.**
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 14 includes the results for semi-log hedonic priding models of the sample including observations with missing energy labels. The results report both the coefficients (Coef.) and the standard errors (Stderr.) of the variables. *** denotes significance at the 1% level, *** denotes 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 the demonstrative maps are shared on GitHub.

Section 4.2.1: Demonstrative Map 1

Note: It is recommendable when reading the discussions of the visualizations of the map to open the map, which is shared on GitHub.

Demonstrative map 1, of which a screenshot is shown in figure 5, includes 68 clusters as provided on Idealista for Barcelona. Every sustainable pricing as specified in pricing model 7 is used as a variable to color the total price impact of sustainability in the map for the sample including the observations with missing energy label data.

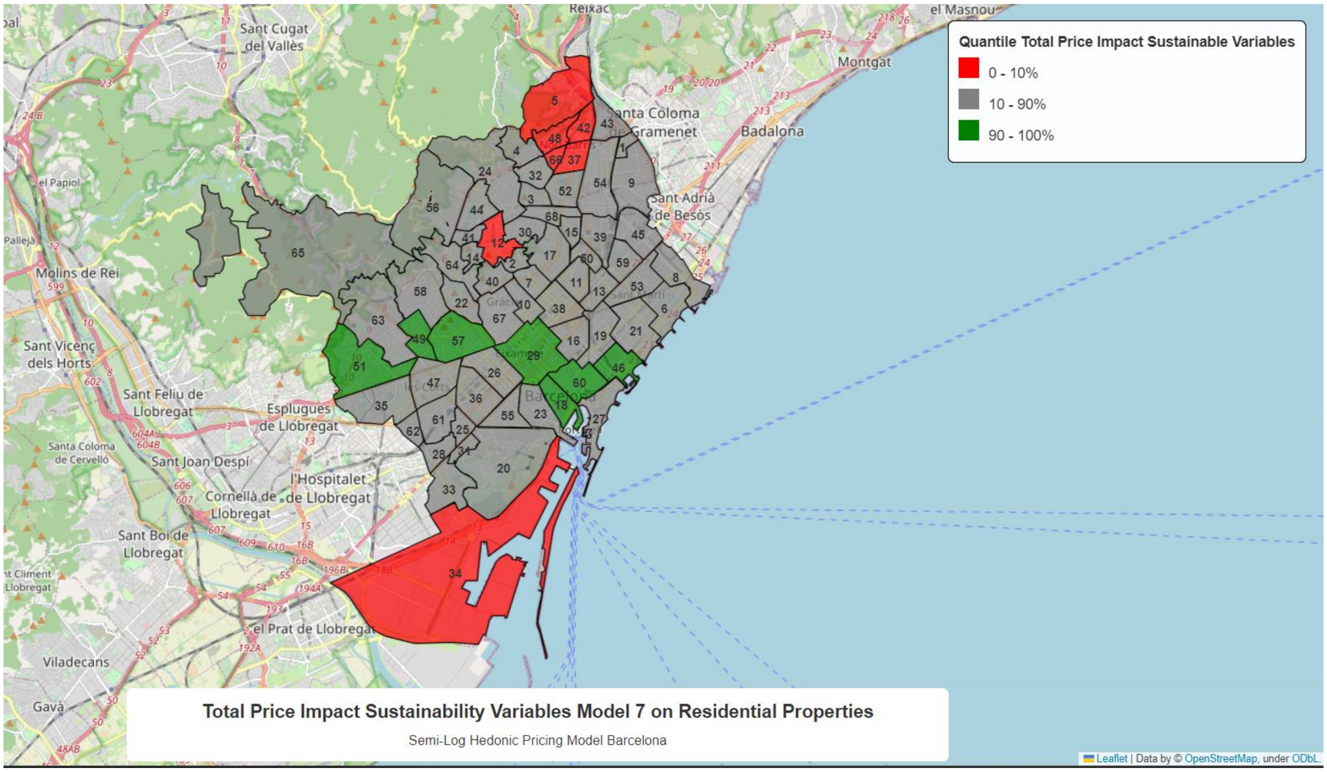
The demonstrative map shows that the houses in the subdistrict with the 10% lowest quantile of total sustainable price impact by the sustainable variables in model 7 are located in the districts: Nou Barris, Horta Guinardo, and Sants Montuic. The houses in the subdistrict in the highest 10% quantile of the total price impact of sustainability are located in the districts: Eixample, Ciutat Vella, Sant Marti, Sarrià Sant Gervasi, and Les Corts. An index including the names of the subdistricts, which are displayed by the numbers on the map, is included in table 16 in the appendix.

Interaction with demonstrative map 1 shows that the low price impact of sustainability for the houses in the lowest 10% quantile is mainly caused by a high average distance to the beach, high distance to the highway/train, high vulnerability to heat impact, high income equality[3], and low welfare. Additionally, the subdistricts within the lowest price impact of sustainability have mostly more than 50% of the values for the higher accessibility to public amenities and services that are than at least 20% of the other subdistricts. Although, there is observed that they can have high values relative to the other subdistricts on some of these accessibility PCA components.
In contrast, the sustainable factor that mostly impacted the negative price impact of sustainability for the lowest 10% quantile, the distance to the beach, distance to the highway/train, vulnerability to heat impact, income equality, and welfare, have mostly opposite values for the properties in the subdistricts with the highest quantile of the total price impact of sustainability. Especially for the subdistricts located closer to the beach. Furthermore, the observations in the subdistricts in the highest category have in general a higher accessibility to public amenities and services compared to other subdistricts if the acceability factors are positively correlated with housing prices. These values for accessibility to public amenities and services are mostly higher than 80% of the other subdistricts.

In conclusion, shows demonstrative map 1 that sustainable pricing in Barcelona is bounded to specific areas/regions. The subdistricts with a high and low total impact of sustainability in the housing prices are in general 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. These sustainable factors are hard to change through policy interventions. The distance to the beach is location bounded and the relationship between housing prices and higher income equality is negative. Although additionally is shown in the map that also the values for the other sustainable variables for the properties with a high/low price impact of sustainability are relatively high/low. This suggests that the difference in the price impact of sustainability between properties in Barcelona is not only driven by the variables with the highest weight, such as income equality and distance to the beach.

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

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



Section 4.2.2: Demonstrative Map 2

Note: It is recommendable when reading the discussions of the visualizations of the map to open the map, which is shared on GitHub.

Demonstrative map 2 clusters the observations on the 68 subdistricts of Barcelona as defined by Idealista for the sample including observations with missing energy labels. The map is colored using the total price impact of the selected sustainable variables. These are all the sustainable variables as included in pricing model 7 excluding the distance to the beach (km), neighborhood density (10 ha), and income distribution PCA component. A screenshot of demonstrative map 2 is provided in figure 6 below. An index for the specific names of the subdistricts, displayed by the number of the colored subdistrict is included in table 16 in the appendix.

The screenshot shows that the cluster of houses in the quantile of the 10% of the subdistrict with the lowest total price impact of the selected sustainable variables are located in almost the same subdistrict as compared to demonstrative map number 1. The difference is that the subdistrict in Sants-Monjuic is not anymore in the lowest 10% quantile for the total price impact of the selected sustainable variables. And there is shown a subdistrict in Sant Andreu in the 10% quantile with the lowest total price impact of the selected sustainable variables. The subdistrict with the highest total price impact for the selected sustainable variables in demonstrative map 2 are shown in the districts Sarria Sant Gervasi, Les Corts, Eixample, and Sant Martí. The number of subdistricts with the highest total price impact of sustainability around Eixample and Sant Martí has lowered, and the number of subdistricts has increased around the districts Les Corts and Sarria Sant Gervasi.

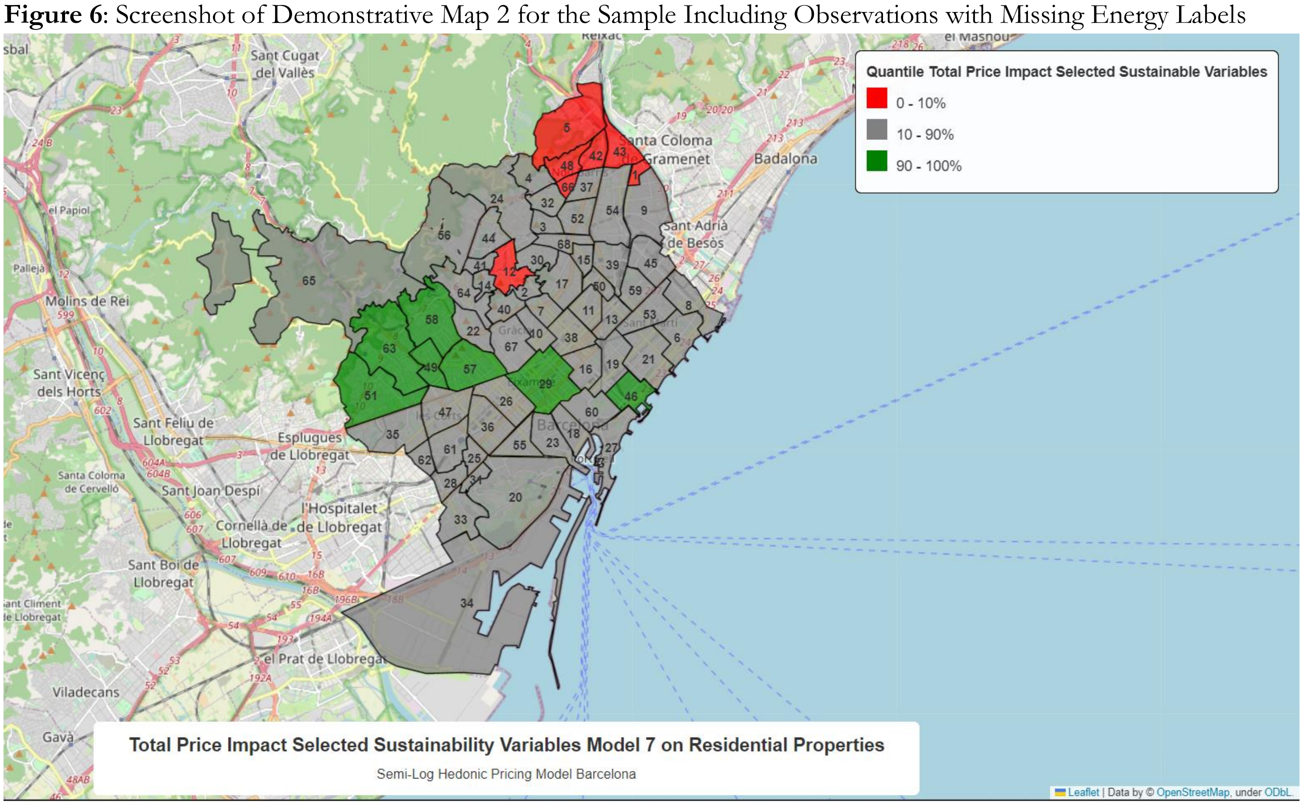
Interaction with demonstrative map 2 shows that the low total price impact of the selected sustainable variables of the subdistricts in Nou Barris and Sants Andreu is mainly caused by a high distance to the nearest highway/train, high vulnerability to a heat impact, and low welfare. In addition, do these subdistricts often have low values for accessibility PCA components that have a positive correlation with the housing price, and do the subdistricts have a high past natural population growth which has a negative correlation with housing prices. In line, the subdistrict in Gracia has the same reasons for the low total price impact of sustainability. Although the past natural population growth has not been that high in the subdistrict. The population density which is also negatively related to housing prices is high in this subdistrict.

The high total price impact of the selected sustainable variables in the highest quantile is mainly caused by high welfare. This welfare is measured by the PCA component of income and unemployment. Every subdistrict in the quantile of subdistrict with the highest total price impact of the selected sustainable variable has a value for the welfare that is at least higher than 85% of the other subdistricts. Besides that, the subdistricts also have high values for other selected sustainable variables. For instance, for none of the subdistricts is the vulnerability to heat impact higher than 15% of the other subdistrict. Moreover, the houses are in most of the subdistricts located nearby a highway or train and the subdistricts have a relatively low population density.

In addition, the subdistricts have in general high values for the accessibility components that have a positive relationship with the housing prices: park & garden, bar & restaurant, university, pharmacy, big shopping place, and museum, library & POI. cult. The values for these PCA components are for most of the subdistricts at least higher than 75% of the other subdistricts.

To sum up, demonstrative map 2 shows that the low/high total price of sustainability is not only caused by sustainable variables that are hard to change by political interventions. The shown subdistricts that have a high/low price impact of sustainability are similar compared to demonstrative map 1. However, demonstrative map 2 shows that these differences in the price impact of sustainability are caused by a wider range of reasons. The subdistricts with the lowest price impact of the selected sustainable variable often have a high distance to the nearest highway or train, high vulnerability to a heat impact, and low access to public services/amenities that are positively correlated with housing prices. The subdistricts with a high total price impact of the selected sustainable variables had oppositive values for these factors. Furthermore, it are often subdistricts in which the average welfare of the inhabitants is high. These results suggest that more fair housing prices subject to the total impact of sustainability should be treated by policy interventions that are locally implemented. There is space for flexibility in political interventions since the high/low price impact of sustainability is caused by a wide range of sustainable aspects. In conclusion, local clustering tendencies of the price impact of sustainability in housing prices are observed in demonstrative maps 1 and 2. This provides no evidence to reject H6, the total price impact of sustainability on housing prices in Barcelona shows local clustering tendencies.

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



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 excludes observations with missing energy labels. Furthermore, demonstrative maps 3 and 4 are discussed. These demonstrative maps cluster the observations based on the location by the support vector machine instead of the subdistrict. Besides that, the same results are visualized in maps 3 and 4 as in respectively demonstrative maps 1 and 2.

Section 5.1: Robustness Pricing Models

The focus in this subsection will be on the differences for the sample that excludes observations with missing energy labels compared to the sample that includes observations with missing energy labels. The results for pricing models 1 to 4 are included in table 17a in the appendix and the results for pricing models 5 to 7 are included in table 17b of the appendix. The VIF test statistic for the variables included in the pricing models is included in table 18 in the appendix.

The findings for model 1, which only includes the housing-specific variables, are similar to the results found for the sample that includes observations with missing energy labels. The reported differences are related to the variables that were correlated with the probability of the presence of an energy label in the housing advertisement. Firstly, the new housing development variable has become higher (0.168) and significant at the 1% level. Secondly, the coefficient of the heating variable has halved. Thirdly, the strength of the dummy energy consumption labels variables is lower in the sample that excludes observations with missing energy label data. This is in line with the 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 (models 2-7). Additionally, in models 4, 5, and 6 the coefficient of the energy consumption label A has become significant in contrast to the findings for the sample that includes the observations with missing energy labels.

The results for model 2, including the ecological dimensions of sustainability, show no differences compared to the results for the sample that includes observations with missing energy labels. The same findings are found for model 3 including the environmental dimension of sustainability. For model 4, including the social dimension of sustainability, only small differences are observed in comparison to the sample that includes observations with missing energy labels. No more evidence is provided for a relationship between the policy PCA component and the secondary & lower education PCA component and the housing prices. 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 labels. 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, show that the coefficient for the religious institution and museum, library & POI culture has decreased in strength in comparison to the results for the sample that includes observations with missing energy labels. This could be caused by the exclusion of observations in more cultural and historical neighborhoods. In these places, there could be a higher probability that a building has no energy label. Although, the results for the Probit model for the presence of an energy label in the housing advertisement did not find evidence that there was a general relationship between the presence of an energy label and the building age.

The results of model 6, including the economic-financial dimensions, are equal in terms of strength and significance compared to the coefficients for the sample that includes observations with missing energy labels.

Model 7, in which every dimension of sustainability is included, shows similar results to those reported for the sample including observation with missing energy labels. This suggests a high degree of robustness. The strength of the coefficients is for all the sustainable variables are 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 labels.

In conclusion, the findings for the sustainable variables are almost completely similar for the sample that excludes observations with missing energy labels compared to the sample that includes observations with missing energy labels. Only big differences are observed in the housing-specific variables which correlate with the presence/absence of an energy label in the housing advertisement.

Section 5.2: Robustness Demonstrative Maps

Section 5.2 discusses the results for demonstrative maps 3 and 4, which are similar to respectively demonstrative maps 1 and 2. The only difference between demonstrative maps 1 and 2 is the applied clustering method. Demonstrative maps 3 and 4 are clustered by the support vector machine (100 clusters) and demonstrative maps 1 and 2 are clustered by the subdistricts (68 subdistricts). The sample which includes observations with missing energy labels is used to visualize the results since in section 5.1 is found that the results for the valuation methods are robust for both samples. A screenshot of demonstrative maps 3 and 4 are respectively included in figures 7 and 8 of the appendix as well as a link to the HTML code of the maps on GitHub.

The advantage of demonstrative maps 3 and 4 is that clustering by the support vector machine makes the clusters more similar in size to each other by using the latitudes and longitude as input. However, the screenshots show that the visualization of the results by demonstrative maps 3 and 4 are almost identical to respectively demonstrative maps 1 and 2. The screenshots show that in both demonstrative maps, the clusters with the highest and lowest price impact of the selected sustainable variables are more or less located in the same districts. There are only small differences found in the exact districts they are located. Furthermore interacting with the demonstrative maps shows that the low/high total price impact of sustainability is caused by the same sustainable variables as for demonstrative maps 1 and 2 that visualize the results by clustering the observation by the subdistricts. The high/low total price impact of sustainability in demonstrative map 3 for the clusters in the highest/lowest quantile is mainly caused by a low/high distance to the beach, low/high distance to the highway, low/high vulnerability to heat impact, high/low income inequality, and high/low welfare. In addition, demonstrative map 4 shows, by excluding the distance to the beach, neighborhood size, and income equality sustainable variables as sustainable variables, that the areas with a high and low total price impact of sustainability in the housing prices remain almost similar to demonstrative map 3. In addition, demonstrative map 4 shows that the clusters with a high/low price impact of sustainability often have relatively high/low price impact of accessibility PCA components that are positively correlated with housing prices.

In conclusion, shows the results that the same areas are found that have a high/price impact on sustainability compared to clustering observations on the subdistrict. Furthermore are the reasons for this high impact of sustainability on the housing price similar.

Thereby, can be concluded that the conclusion to not reject the hypothesis 6 is robust with different visualization methods for the results.

Section 6: Conclusion and Discussion

This research investigates the price impact of sustainability on housing prices in Barcelona. A data-driven approach is applied to study sustainability 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. Earlier state-of-the-art work studied the effect of an increase in sustainability such as closer distance to the coast, higher perceived neighborhood security, and better access to public services/amenities on housing, only from one perspective or dimensions (Buonanno et al., 2013; Chen & Marmolejo Duarte, 2018; Dell'Anna et al., 2019; Graells-Garrido et al., 2021; Marmolejo-Duarte & Chen, 2022). Furthermore, our work elaborates on earlier research by visualizing the total price impact of sustainability on housing prices for neighborhoods/subdistricts.

The price impact of the five sustainable dimensions on housing prices is estimated by semi-log hedonic pricing models. The pricing models in which the five sustainable dimensions are included on an individual basis show that higher values for each sustainable dimension increase the willingness to pay for housing. This same conclusion is found for a pricing model that includes every dimension of sustainability. Thereby the first five hypotheses, which stated for each sustainable dimension separately that an increase in the sustainable variables related to the dimension increases the willingness to pay for housing, are not rejected in our research. Additionally, results from the model that included every dimension of sustainability show often lower strength for the relationship between the housing price and sustainable variables. This illustrates the importance of a multidimensional approach when studying the price impact of sustainability. By not including variables related to other sustainable dimensions, the reported results are likely to suffer from an omitted variable bias. Moreover, sustainable variables with a low (in)significance can become (in)significant when controlling for other dimensions of sustainability.

To sum up, the model including every dimension of sustainability shows 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. Evidence for a significant negative price effect on housing prices is found for the distance to the highway/train, distance to the beach, neighborhood size, vulnerability to heat waves, natural population growth, density, and access to bus & metro, park & gardens, viewpoints, and performing arts variables. Robustness of the results is provided by stating the same conclusions for the sustainable variables for both the sample in- and excluding observations with missing energy labels.

The results are visualized by constructing two demonstrative maps clustering the observations in the sample based on the subdistrict in which a property is located. The demonstrative maps show that the properties with a high/low total price impact of sustainability are clustered in specific areas. This implies that we cannot reject hypothesis 6 stating that the pricing of sustainability in Barcelona shows local clustering tendencies. In more detail, the first demonstrative map, which included every sustainable dimension, shows that a high/low total price impact of sustainability for a cluster is mainly caused by a wide variety of sustainable variables. These variables include the value for the distance to the beach, distance to the highway or train, vulnerability to a heat wave, welfare, and income equality in the area of the property. The second demonstrative map excluded sustainable variables that are strictly location-bounded or are observed to have a negative relationship between social fairness and housing prices. These sustainable factors, the distance to the beach, neighborhood size, and income distribution are not likely to be changed by policy intervention. The second demonstrative map showed that the areas in which a high/low total price impact of sustainability was observed remained almost the same. However, the map better highlighted that these areas with a high/low price impact of sustainability also have in general low values for the access to public services/amenities sustainability factors that are positively correlated with housing prices. This implies that policy intervention to address the unfair pricing of sustainability is best suited for a localized approach. These policies can be flexible given the wide range of sustainable factors by which the unfair pricing of sustainability is caused. The robustness of the conclusion for demonstrative maps 1 and 2 is verified by clustering the properties based on the location with the support vector machine in demonstrative maps 3 and 4.

Additionally, the code to construct the maps is shared on our GitHub. Hereby, offering the opportunity for future work to visualize the results of the pricing models by focusing on other aspects.

Overall, with the results of the thesis, I exceeded my expectations. Before starting my research, I expected to extend state-of-the-art research by the identification of a wider scope of sustainable pricing factors and providing deeper insight into how they influence housing prices in Barcelona. This given my familiarity with both data-driven research and real estate. However, I was not expecting to be able to visualize the results of valuation models in an interactive map that is easy to analyze and also includes in-depth information. I became more familiar with linking and combining data while remaining interpretability. For instance by linking the geographic data with the location of properties and the creation of PCA components. In addition, I was not familiar with any methods for web scraping and constructing geographical maps, and for both, I had to learn HTML coding. Related to the results, I was expecting to find that the impact of sustainability in housing prices is locally clustered caused by a few key factors. But, I was not expecting to find that this is driven by overall bad scores for almost all sustainable factors. Related to the research process, I learned the importance to share findings with people that are less familiar with the topic. This helped me to get insights from new viewpoints and better describe and interpret results.

This research has limitations and provides opportunities for future work. Firstly, the pricing models measure sustainability from a wide view. However, additional sustainable factors can be introduced that do not have a high correlation with the included variables. Secondly, the visualized results by the demonstrative maps are only limited studied in terms of interpretation. The findings are not connected to the costs to improve the score of the sustainable factors. Thirdly, similar visualization maps are not reported for other cities in earlier research. Therefore cannot be verified if the results are in line with the price impact of sustainability on housing prices found in earlier work. This is in contrast to the findings for the pricing models. Fourthly, this research applies in contrast to most work in the field of some degree of machine learning by the construction of PCA components for highly correlated variables. However, in future work machine learning could also be used to predict housing prices. This allows more non-linear relationships between housing prices and the predicting variables. But the challenge that exists when applying machine learning methods is a visualization of the results given the allowance for non-linearity in the models.

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 7). *BarcelonaCiutat_Barris.csv*. Https://Opendata-Ajuntament.Barcelona.Cat/Data/En/Dataset/20170706-Districtes-Barris/Resource/B21fa550-56ea-4f4c-9adc-B8009381896e.

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

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

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

Breusch, T. S., & Pagan, A. R. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica*, *47*(5), 1287. https://doi.org/10.2307/1911963

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

Department of Urban Resilience. (2020). *What impact do programmed activities held in public space have on relationships between the people living in the neighbourhood?* .

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

Guggisberg, M. (2019). Misspecified Discrete Choice Models and Huber-White Standard Errors. *Journal of Econometric Methods*, *8*(1). https://doi.org/10.1515/jem-2016-0002

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

J.J. Heckman. (1976). The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models. *Annals of Economic and Social Measurement*, *5*(4).

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

Overpass Turbo. (2023). *overpass turbo*. Https://Overpass-Turbo.Eu/.

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

Ramsey, J. B. (1969). Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis. *Journal of the Royal Statistical Society: Series B (Methodological)*, *31*(2), 350–371. https://doi.org/10.1111/j.2517-6161.1969.tb00796.x

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

Scikit Learn. (2023). *Support Vector Machines*. Https://Scikit-Learn.Org/Stable/Modules/Svm.Html.

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

The European Parliament, & The Council of the European Union. (2002). *Directive 2002/91/EC of the European Parliament and of the Council of 16 December 2002 on the energy performance of buildings*.

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.

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/.

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 Sample Excluding Observations with Missing Energy Labels

Variable	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 3 includes the summary statistics for the continuous housing-specific variables for the sample excluding observations with missing energy labels.

Table 5: Summary Statistic for the Housing-Specific Dummy Variables for the Sample Excluding Observations with Missing Energy Labels

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 excluding observations with missing energy labels.

Table 6: Description of the Sustainable Variables

Variable(s)	Sus. Dim.	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 G, Energy Label Consumption N_A	Ecological	https://www.idealista.com/	Energy Labels in the housing advertisement of the property
Distance to nearest bus (km), Number of bus stations within 0.25 km	Ecological	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=estacions-bus	Bus stops of the city of Barcelona
Distance to nearest highway (km)	Ecological	https://www.openstreetmap.org/	Keyword: [way["highway"]["maxspeed"="value"]; value = [100,105,110,115,120,125, 130]
Distance to nearest metro (km), Number of metro stations within 0.25 km	Ecological	https://www.openstreetmap.org/	Keyword: node["public_transport"="station"]["station"="subway"]
Distance to nearest train (km)	Ecological	https://www.openstreetmap.org/	Keyword: node["public_transport"="station"]['train'='yes']
Distance to nearest park or garden (km), Number of parks and gardens within 0.25 km	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
Distance to nearest viewpoint (km), Number of viewpoints within 1 km	Environmental	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=np-nasia-miradors	Viewpoints in 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 Barcelona to heat exposure
Distance to nearest beach (km)	Environmental	https://www.openstreetmap.org/	Keyword:
Distance to nearest police station (km), Number of police stations within 1 km	Social	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=comissaries-policia	Police stations
Distance to nearest bar (km), Number of bars within 0.25 km	Social	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=culturailleure-espaismusicacopes	Music and drinks spaces
Distance to nearest restaurant (km), Number of restaurants within 0.25 km	Social	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=equipament-restaurants	List of restaurant equipment
Distance to nearest under three-years-old school (km), Number of under 3 years-old schools within 0.5 km, Distance to nearest 3-6 years-old school (km), Number of 3-6 years-old schools within 0.5 km, Distance to nearest primary school (km), Number of primary schools within 0.5 km, Distance to nearest secondary school (km), Number of secondary schools within 0.5 km, Distance to nearest university (km), Number of universities within 0.5 km	Social	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=equipament-educacio	List of education equipments
Distance to nearest pharmacy (km), Number of pharmacies within 0.25 km, Distance to nearest hospital or clinique (km), Number of hospitals or clinics within 0.5 km	Social	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=equipament-sanitat	List of health equipment
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

Variable(s)	Sus. Dim.	Resource	Description Dataset
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
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
Distance to nearest shopping gallery (km), Number of shopping galleries within 1 km	Social	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=galeries-comercials	Shopping galleries
Distance to nearest shopping center (km), Number of shopping centers within 1 km	Social	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=grans-centres-comercials	Large shopping centers
Distance to nearest large establishment (km), Number of large establishments within 1 km	Social	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=grans-establiments	Large establishments
Social Cohesion	Social	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=cohesio-social	Social cohesion
Distance to nearest cinema (km), Number of cinemas within 0.5 km, Distance to nearest theatre (km), Number of theatres within 0.5 km, Distance to nearest concert place (km)	Cultural	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=culturailleure-cinemesteatresauditoris	Spaces where cinema, theater, and concerts take place
Distance to nearest library (km), Number of libraries within 0.5 km, Distance to nearest museum (km), Number of museums within 0.5 km, Distance to nearest point of interest culture (km), Number of point of interest culture within 0.5 km	Cultural	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=culturailleure-bibliotequesimuseus	Spaces with a library or study room and museum spaces
Distance to nearest religious institution (km), Number of religious institution within 0.5 km	Cultural	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=equipament-serveis-religiosos	List of religious services equipment
Distance to nearest point of interest culture (km), Number of point of interest culture within 0.5 km	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
Gini Index	Economic-Financial	https://opendata-ajuntament.barcelona.cat/data/en/dataset/atles-renda-index-gini	Gini (%) índex
AVG Gross taxable household income by subdistrict	Economic-Financial	https://opendata-ajuntament.barcelona.cat/data/en/dataset?q=&name=atles-renda-bruta-per-llar	The 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 (Sus. Dim.), the source (Resource), and a description of the dataset of the variables (Description Dataset).

Table 9: Summary Statistics of the Sustainable Variables for the Sample Excluding Observations with Missing Energy Labels

Variable	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 excluding observations with missing energy labels.

Figure 3: Correlation Matrix of the Variables for the Sample Including Observations with Missing Energy Labels Data That Have an Absolute Correlation Coefficient Higher Than 0.25

Figure 4: Correlation Matrix of the Variables for the Sample Excluding Observations with Missing Energy Labels Data That Have an Absolute Correlation Coefficient Higher Than 0.25

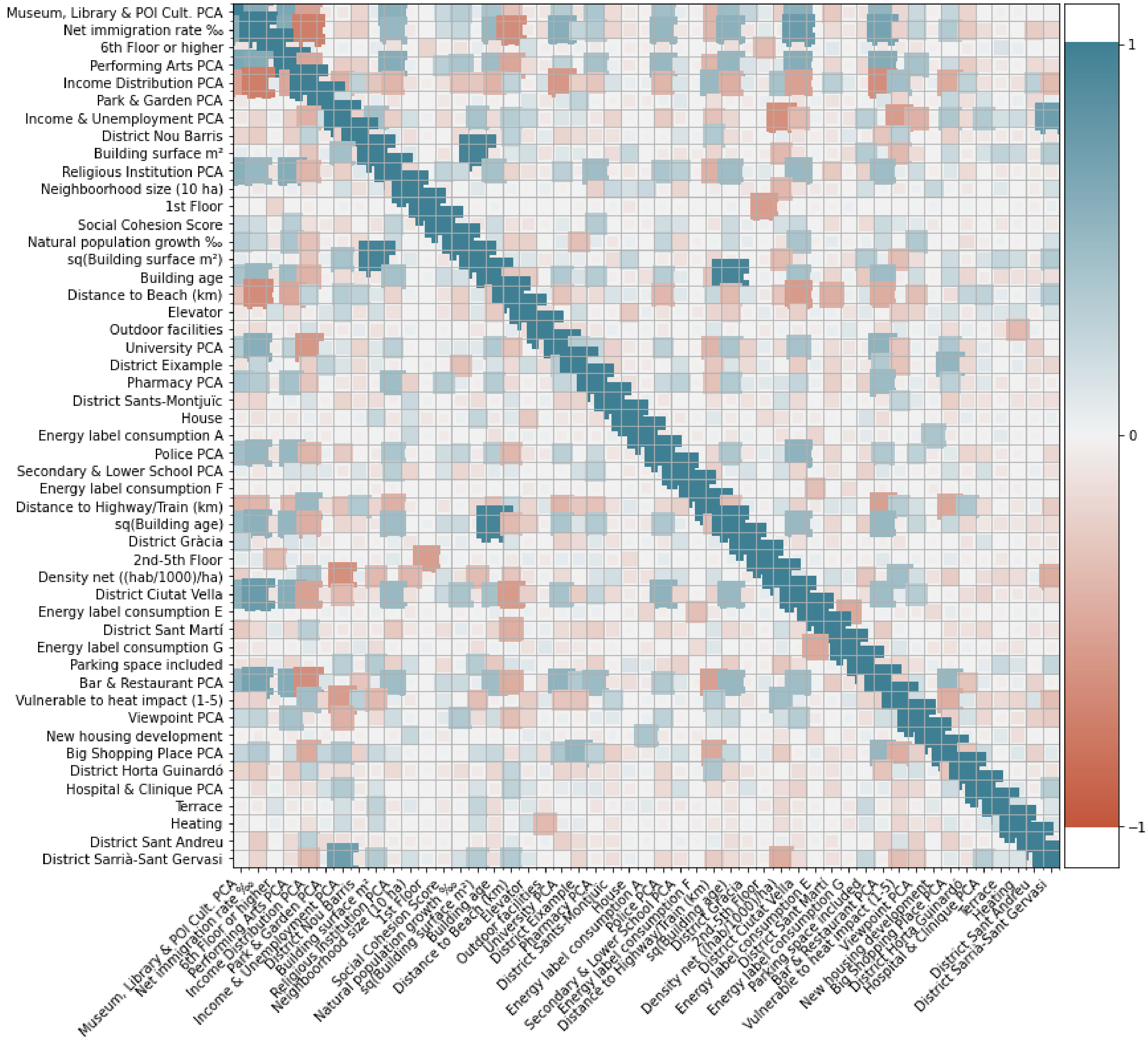


Table 13: Parameters to Construct Demonstrative Maps 3 and 4

Parameters	**Demonstrative Map 3**	**Demonstrative Map 4**
Selected Sustainable Features	Sustainable_predictors	[e for e in sustainable_predictors if e not in ['Distance to Beach (km)','Neighborhood size (10 ha)',' Income Distribution PCA']]
Map_save_name	'Demonstrative_map_1'	'Demonstrative_map_2'
Title	'Total Price Impact Sustainability Variables Model 7 on Residential Properties'	'Total Price Impact Selected Sustainability Variables Model 7 on Residential Properties'
Subtitle	"Heckman Selection Model Barcelona"	"Heckman Selection Model Barcelona"
Legend_title	'Quantile Total Price Impact Sustainable Variables'	'Quantile Total 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	{}	{}
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
SVM_Cluster (True/False)	True	True
Subdistrict_Cluster (True/False)	False	False

Table 13 includes the parameter values to construct demonstrative map 3 and demonstrative map 4

Table 15: VIF Test Statistics of the Sample Including Observations with Missing Energy Labels

Variable	**Probit Model**	**VIF Model 1**	**VIF Model 2**	**VIF Model 3**	**VIF Model 4**	**VIF Model 5**	**VIF Model 6**	**VIF Model 7**
Building surface m²	6.01	116.26	117.07	117.10	119.31	117.99	116.81	120.37
sq(Building surface m²)		37.48	37.60	37.57	37.95	37.76	37.48	38.35
Building age	5.28	107.48	107.51	111.10	111.86	108.63	107.59	116.96
sq(Building age)		47.91	48.01	49.67	50.88	49.55	48.30	53.15
House		1.38	1.38	1.43	1.40	1.38	1.39	1.45
Mezzanine		1.56	1.56	1.56	1.56	1.56	1.56	1.57
1st Floor		3.38	3.38	3.38	3.39	3.38	3.38	3.40
2nd -5th Floor		6.21	6.22	6.23	6.24	6.23	6.21	6.28
6th Floor or higher		2.25	2.25	2.25	2.25	2.25	2.25	2.27
New housing development	1.14	2.17	2.28	2.86	2.29	2.18	2.80	5.15
Needs renovation	1.24	1.35	1.35	1.41	1.37	1.35	1.41	1.64
Elevator	4.58	6.37	6.52	6.56	6.48	6.40	6.40	6.85
Terrace		1.74	1.75	1.79	1.75	1.75	1.77	1.89
Heating	2.55	3.46	3.54	3.91	3.58	3.46	3.85	5.51
Outdoor facilities	2.49	2.90	2.90	2.91	2.92	2.90	2.90	2.93
Air Conditioning	12.90							

Variable	Probit Model	VIF Model 1	VIF Model 2	VIF Model 3	VIF Model 4	VIF Model 5	VIF Model 6	VIF Model 7
Parking space included		1.42	1.43	1.45	1.48	1.44	1.47	1.52
Energy label consumption A		1.17	1.17	1.22	1.17	1.17	1.17	1.24
Energy label consumption B		1.17	1.17	1.17	1.17	1.17	1.17	1.17
Energy label consumption C		1.14	1.14	1.14	1.15	1.14	1.14	1.15
Energy label consumption E		2.25	2.26	2.26	2.26	2.26	2.25	2.27
Energy label consumption F		1.23	1.23	1.23	1.23	1.23	1.23	1.23
Energy label consumption G		1.45	1.46	1.46	1.46	1.46	1.46	1.46
Bus & metro PCA			5.75					6.53
Distance to Highway/Train (km)			6.71					12.17
Distance to Beach (km)				32.20				47.05
Park & garden PCA				2.46				2.63
Viewpoint PCA				5.29				7.95
Neighborhood size (10 ha)				2.99				4.24
Vulnerable to heat impact (1-5)				24.16				28.73
Police PCA					5.94			6.60
Bar & restaurant PCA					9.02			10.30
Secondary & lower educ. PCA					9.32			11.41
University PCA					3.61			4.04
Pharmacy PCA					10.05			11.00
Hospital & Clinique PCA					3.60			3.81
Big Shopping Place PCA					5.23			6.34
Social Cohesion Score					2.86			2.94
Natural population growth ‰					3.38			5.20
Density net (hab/1000/ha)					17.32			38.49
Performing Arts PCA						5.43		43.59
Religious Institution PCA						8.72		8.21
Museum, Library & POI Cult. PCA						6.02		13.59
Income Distribution PCA							29.93	7.81
Income & Unemployment PCA							13.82	78.70
District Eixample		6.34	6.57	8.51	8.43	6.62	8.22	25.20
District Ciutat Vella		5.66	5.81	10.62	13.32	6.98	12.24	11.54
District Sant Martí		3.57	3.59	6.87	4.45	3.67	4.62	18.55
District Sants-Montjuïc		3.81	3.94	4.71	4.71	4.06	5.61	9.24
District Horta Guinardó		3.25	3.34	3.95	3.80	3.32	4.45	7.25
District Gràcia		2.71	2.71	3.11	3.11	3.00	3.24	5.71
District Nou Barris		2.56	2.66	3.07	3.02	2.86	3.68	4.10
District Sarrià-Sant Gervasi		3.34	3.54	3.43	3.66	3.42	3.73	4.87
District Sant Andreu		2.56	2.57	3.01	3.00	2.63	3.43	4.57
IMR		110.94	128.71	201.19	139.35	111.32	202.46	4.48

Table 15 includes the VIF test statistics for the variables included in the semi-log hedonic pricing models of the sample including observations with missing energy labels.

Table 16: The Index for the Codes of the Subdistricts in Demonstrative Maps 1 and 2

Code	Subdistrict	Code	Subdistrict
1	Baró de Viver	35	La Maternitat i Sant Ramon
2	Can Baró	36	La Nova Esquerra de l'Eixample
3	Can Peguera - El Turó de la Peira	37	La Prosperitat
4	Canyelles	38	La Sagrada Família
5	Ciutat Meridiana - Torre Baró - Vallbona	39	La Sagrera
6	Diagonal Mar i el Front Marítim del Poblenou	40	La Salut
7	El Baix Guinardó	41	La Teixonera
8	El Besòs	42	La Trinitat Nova
9	El Bon Pastor	43	La Trinitat Vella
10	El Camp d'En Grassot i Gràcia Nova	44	La Vall d'Hebron - La Clota
11	El Camp de l'Arpa del Clot	45	La Verneda i la Pau
12	El Carmel	46	La Vila Olímpica del Poblenou
13	El Clot	47	Les Corts
14	El Coll	48	Les Roquetes
15	El Congrés i els Indians	49	Les Tres Torres
16	El Fort Pienc	50	Navas
17	El Guinardó	51	Pedralbes
18	El Gòtic	52	Porta
19	El Parc i la Llacuna del Poblenou	53	Provençals del Poblenou
20	El Poble Sec - Parc de Montjuïc	54	Sant Andreu
21	El Poblenou	55	Sant Antoni
22	El Putxet i el Farró	56	Sant Genís Dels Agudells - Montbau
23	El Raval	57	Sant Gervasi - Galvany
24	Horta	58	Sant Gervasi - La Bonanova
25	Hostafrancs	59	Sant Martí de Provençals
26	L'Antiga Esquerra de l'Eixample	60	Sant Pere - Santa Caterina i la Ribera
27	La Barceloneta	61	Sants
28	La Bordeta	62	Sants - Badal
29	La Dreta de l'Eixample	63	Sarrià
30	La Font d'En Fargues	64	Vallcarca i els Penitents
31	La Font de la Guatlla	65	Vallvidrera - El Tibidabo i les Planes
32	La Guineueta	66	Verdun
33	La Marina del Port	67	Vila de Gràcia
34	La Marina del Prat Vermell	68	Vilapicina i la Torre Llobeta

Table 16 includes the code and the name of the subdistrict shown in demonstrative maps 1 and 2.

Table 17A: Semi-Log Hedonic Pricing Model Results of the Sample Excluding Observations with Missing Energy Labels

	Model 1		**Model 2**		**Model 3**		**Model 4**	
	Coef.	**Stderr.**	**Coef.**	**Stderr.**	**Coef.**	**Stderr.**	**Coef.**	**Stderr.**
Constant	11.594***	0.03	11.786***	0.03	12.445***	0.03	11.568***	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

	Model 1		Model 2		Model 3		Model 4	
	Coef.	**Stderr.**	**Coef.**	**Stderr.**	**Coef.**	**Stderr.**	**Coef.**	**Stderr.**
New housing development	0.168***	0.02	0.179***	0.02	0.213***	0.02	0.175***	0.02
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 17 includes the results for semi-log hedonic priding models of the sample excluding observations with missing energy labels.. The results report both the coefficients (Coef.) and the standard errors (Stderr.) of the variables. *** denotes significance at the 1% level, *** denotes significance at the 5% level, and * denotes significance at the 10% level.

Table 17B: Semi-Log Hedonic Pricing Model Results of the Sample Excluding Observations with Missing Energy Labels

	Model 4		Model 5		Model 6		Model 7	
	Coef.	**Stderr.**	**Coef.**	**Stderr.**	**Coef.**	**Stderr.**	**Coef.**	**Stderr.**
Constant	11.568***	0.03	11.563***	0.03	11.548***	0.04	12.110***	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
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

	Model 4		Model 5		Model 6		Model 7	
	Coef.	Stderr.	Coef.	Stderr.	Coef.	Stderr.	Coef.	Stderr.
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**		**0.875**	

Table 17 includes the results for semi-log hedonic priding models of the sample excluding observations with missing energy labels. The results report both the coefficients (Coef.) and the standard errors (Stderr.) of the variables. *** denotes significance at the 1% level, *** denotes significance at the 5% level, and * denotes significance at the 10% level.

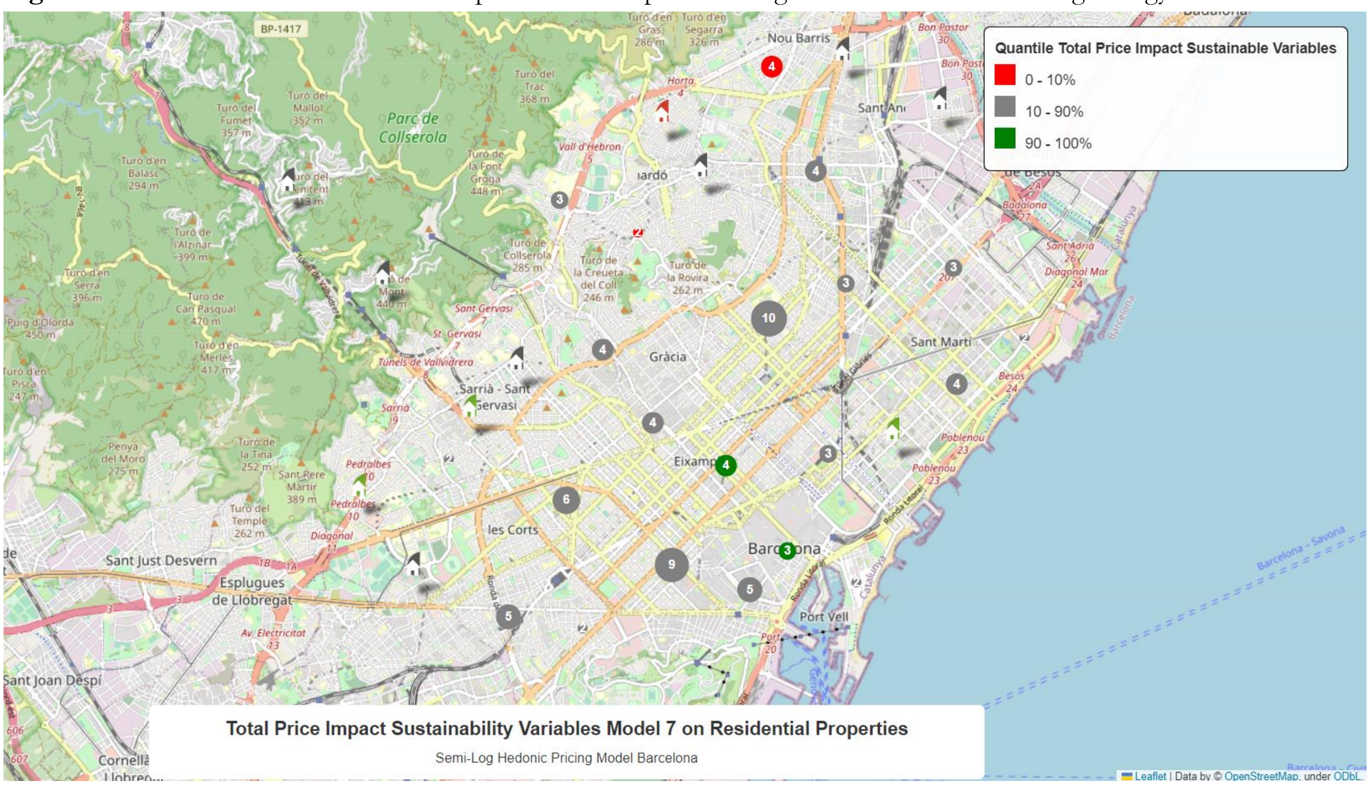
Table 18: VIF Test-Statistics of the Sample Excluding Observations with Missing Energy Label Data

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

Variable	VIF Model 1	VIF Model 2	VIF Model 3	VIF Model 4	VIF Model 5	VIF Model 6	VIF Model 7
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

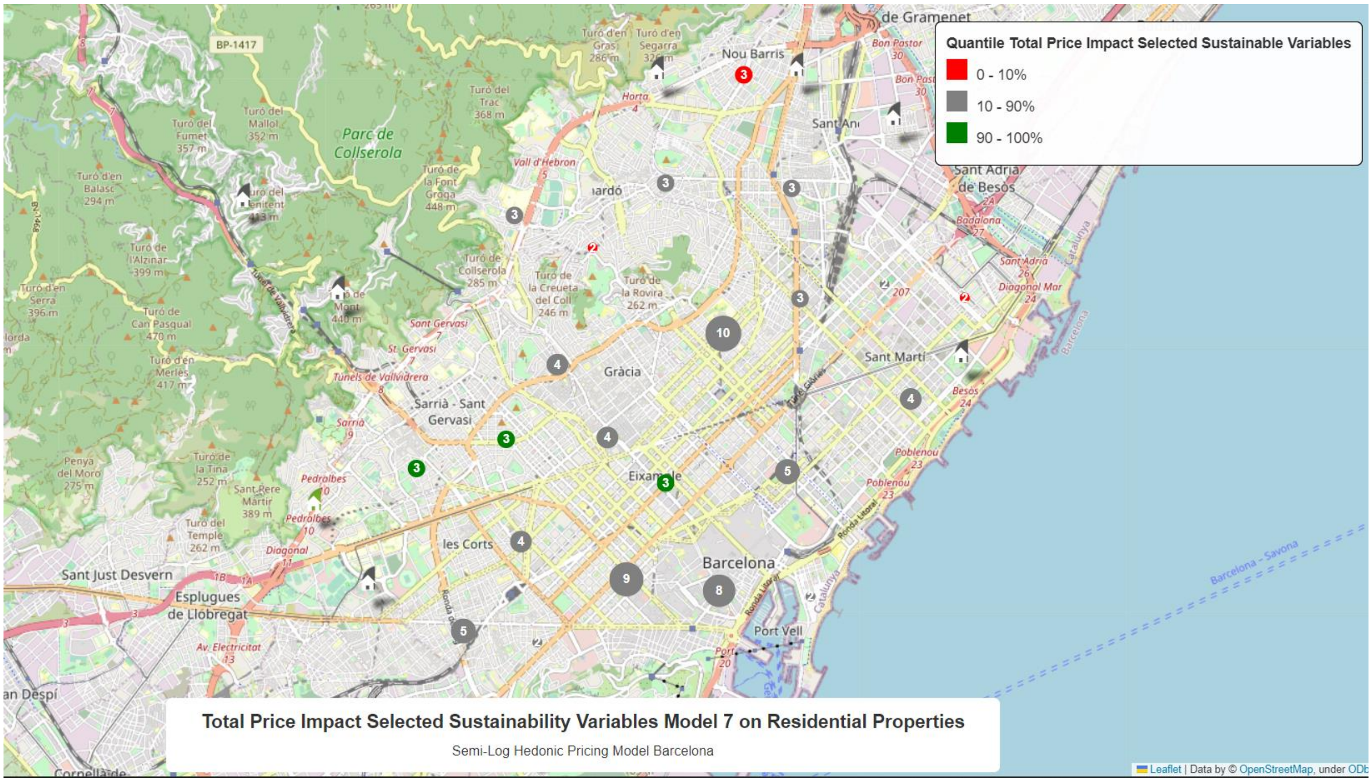
Table 18 includes the VIF test statistics for the variables included in the semi-log hedonic pricing models of the sample excluding observations with missing energy labels

Figure 7: Screenshot of Demonstrative Map 3 for the Sample Including Observations with Missing Energy Labels



GitHub Link

Figure 8: Screenshot of Demonstrative Map 4 for the Sample Including Observations with Missing Energy Labels



GitHub Link