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

31 August, 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

31 August, 2023

Universitat Pompeu Fabra Barcelona

Table of Contents

Abstract:

In this research, I apply a data-driven multidimensional approach to study the price impact of sustainability on housing prices in Barcelona, Spain. The price impact of sustainability is captured by sustainable pricing factors related to five sustainable dimensions: ecological, environmental, social, cultural, and economic-financial. In total, I proposed 23 location-bounded sustainable pricing factors derived from the quality and healthiness of the living environment of a property. I estimated most of these sustainable pricing factors as PCA components based on sustainable features with a high correlation to avoid multicollinearity in pricing models. I used these sustainable pricing factors, alongside housing-specific, and dummy district pricing factors to predict the asking price from housing advertisements on Idealista. Asking prices tend to be close to transaction prices in stable residential real estate markets, such as the Barcelona housing market over the last years with consistently increasing prices. The sample of this research consists of approximately 13.350 or 10.100 observations dependent on the inclusion/exclusion of properties with missing energy labels. I used this sample to estimate 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 term "utility" represents the expected satisfaction/benefit that homebuyers receive from buying a property in monetary value. The semi-log model ensures that the pricing factors can be interpreted as price elasticities, where a unit increase in the pricing factor is equal to the percentage change in the housing price. With the models, I found evidence that an increase in each of the five sustainable dimensions increases the willingness to pay for housing. Furthermore, I found evidence 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. With this finding, I extend prior literature which often only studied sustainability from one dimension or variable. Additionally, I constructed demonstrative maps that visualize the total price impact of sustainability on housing prices across the city. These demonstrative maps show that houses with a high/low total price impact of sustainability calculated by a high/low sum of the value of selected sustainable factors times their coefficients are locally clustered in Barcelona. The differences in the high/low total price impact of sustainability are driven by a wide variety of sustainable factors. This suggests that the price impact of sustainability on house prices can be distributed more equally and made fairer across Barcelona by local policy interventions. Furthermore, I shared the code, which offers high flexibility to visualize the results, to construct demonstrative maps on GitHub to stimulate further research.

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 based on the people, planet, and nature aspects following the sustainable development goals of the UN (Arcadis, 2022). Higher scores on these sustainability aspects will increase the quality of life of people (Eurostat Statistics Explained, 2022). For instance, in Barcelona, Yanez et al. (2023) found that the 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 who found a positive relationship between the self-perceived general and mental health of people and the surrounding by and access to green spaces.

Sustainability is 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 which 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 caused by the demand side since tenants are willing to pay higher prices for more sustainable properties (D. Worford, 2022).

Home buyers who 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 provided evidence that higher energy labels are positively correlated with housing prices (Chen & Marmolejo Duarte, 2018; Dell'Anna et al., 2019). Related to the location of the property in Barcelona, for instance, is provided evidence in earlier work 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 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 lower commuting times in Barcelona in comparison to other cities.

These above-mentioned studies were mainly conducted by studying sustainability from one dimension, but sustainability in itself has many dimensions (United Nations, 2023). To address this issue, in this research the price impact of sustainability, which is defined as the value of the studied sustainable factors times their coefficients, on housing prices is studied from five different dimensions: ecological, environmental, social, cultural, and economic-financial as proposed by Kauko (2019). These sustainable dimensions focus on the quality and healthiness of the living environment and property-specific characteristics. The sample to estimate the price impact of sustainability consists of approximately 13.350 or 10.100 properties in Barcelona depending on the in- or exclusion of observations with missing energy labels. This sample is retrieved from all the housing advertisements that were listed on Idealista on 17 April 2023 for Barcelona. Duplicates in the sample are removed by the house-id specified in the housing advertisement. However, the same property might appear multiple times in the sample if it is listed by multiple agencies. Since the housing advertisements on Idealista only include the asking price, the asking price is assumed to be equal to the transaction price. This transaction price tends to be close to the asking price in stable housing markets McGreal et al. (2010). The residential real estate market for Barcelona has fulfilled this requirement over the last few years by showing a consistent increase in housing prices (Idealista, 2023b). Besides, in the housing advertisement on Idealista, the specified sustainable characteristics of properties in housing advertisements are limited. Therefore in this research, most sustainable factors are location-bounded and relate to the quality and healthiness of the living environment.

The price impact of sustainability is captured by sustainable factors that are included in semi-log hedonic pricing models in addition to housing-specific factors and dummy district variables. The hedonic pricing model assumes that the price paid for housing is equal to the utility buyers are expected to extract from it. The term 'utility' refers to the satisfaction or benefits that individuals are expected to gain from purchasing the property often expressed in terms of monetary value. The semi-log hedonic model ensures that pricing factors can be interpreted as semi-elastic. As a result, 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.

I provide insight into which houses are positively and negatively impacted by sustainable pricing factors in the city of Barcelona by visualizing the results of the pricing models in demonstrative maps. With these maps, I show the total price impact of selected sustainable factors on the predicted housing price. I make it possible with the maps to identify the difference in the total price impact of sustainable factors between different areas in Barcelona. This makes it easier to distinguish areas with a high/positive total price impact of selected sustainable factors, where the sum of the values of selected sustainable variables times their coefficients is high, from areas with a low/negative total price impact of selected sustainable factors. The visualization of the results offers insights into the differences in the price impact of sustainability across the city providing opportunities for policy intervention to distribute the price impact of sustainability more equally in Barcelona. So the fairness of the pricing of sustainability in the city can be increased. Additionally, to stimulate further research I shared the code to construct the maps on GitHub. I included a wide range of options to visualize the results of the valuation models in the code which will be discussed later in this research. Hereby, readers of this report on research report can also retrieve additional insights from the results of this research by constructing their own geographical maps.

I structured the remainder of this report on research as follows: in Section 2 I included the literature review. In Section 3 I included the data and methodology. In Sections 4 and Section 5, I discuss the results and their robustness. I included the conclusion and discussion of the research in Section 5. Lastly, in Sections 6 and Section 7 I included the reference list and the appendix.

Section 2: Literature review

In the literature review, I discuss the main findings of earlier research concerning the price impact of sustainability on housing prices. I discuss the price effect of sustainability from five sustainable dimensions as proposed by Kauko (2019): ecological, environmental, social, cultural, and economic-financial. The sustainable dimensions focus primarily on the quality and healthiness of the living environment around the properties alongside property-specific characteristics. I analyze the five sustainable dimensions also in the empirical analysis of this report of research. The discussed research consists of international, national, and regional studies.

Section 2.1: Ecological Dimension of Sustainability

The ecological dimension of sustainability ranges from the energy efficiency of a property to the accessibility of the property by different transportation options (Kauko, 2019). 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 found. 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 that were estimated on 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 the work of Dell'Anna et al. (2019). However, a lot of observations are excluded from the sample in the work of Dell'Anna et al. (2019) due to missing energy labels in housing advertisements. To address this issue, Chen & Marmolejo Duarte (2018) estimated a Heckman selection model to correct for a sample selection bias. Chen & Marmolejo Duarte (2018) found in the model that corrected for the sample selection bias an increase in the 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. However, in future work, Marmolejo-Duarte & Chen (2022) reconsidered these findings by showing that the premium for energy labels becomes an insignificant pricing factor if additional variables related to the architectural quality of properties are included in the model. Hereby, Marmolejo-Duarte & Chen (2022) argue that earlier findings are the consequences of an information asymmetry. This is by the exclusion of relevant property-specific features in the pricing models that are correlated with the presence of an energy label in the housing advertisement.

The accessibility of an area by both short and long-commuting transport options is another studied ecological pricing factor in earlier work. 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. However, 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 transportation 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 Alicante, 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. Mixed results are found in earlier work between 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 significantly lower housing prices if a highway was located within 100 meters in the Netherlands. However, Ayala et al. (2016) found in Alicante, Spain, a positive effect of the closeness to a highway on the housing price. This positive effect is also found by Dell'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 on natural areas, such as the seashore or parks. For the seashore, a higher willingness to pay for housing is often found for properties closer to the seashore. For example, Dell'Anna et al. (2019) found in their research 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 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, South Korea, a significant positive relationship between the accessibility and closeness to parks and housing prices. The 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 in their work contrary results with a significant positive relationship between the distance to parks and housing prices. Dell'Anna et al. (2019) explained that this is likely caused by the structural design of Barcelona where parks are mainly located at the periphery. Instead of the closeness, another often studied 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 their work in China a positive price premium for both houses with a river and/or green view. In line with this, a positive paid premium for housing for properties with a sea view in Sweden was found by research by Wilhelmsonn et al. (2020). These findings are supported by the work of Castro Noblejas et al. (2022) in Malaga, Spain, for properties with a higher-quality visual basin. This visual basin in the work of Castro Noblejas et al. (2022) was present for properties that were located in an area with vegetation and a sea view.

Section 2.3: Social Dimension of Sustainability

Prior literature that studied the social dimension of sustainability focused 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. Higher accessibility to facilities is often reported to have a positive effect on the 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 hospitals, 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. A significant positive correlation was found for Barcelona 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). However, Graells-Garrido et al. (2021) only studied the Spearman rank correlation between these factors and housing rents. Hereby, Graells-Garrido et al. (2021) did not correct for the price effect that other variables could have on the willingness to pay for housing. Another often-studied pricing factor is the safety of the neighborhood. 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. Buonanno et al. (2013) found in Barcelona evidence 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.

Concerning demographic factors, an often-studied price factor is the population density of a neighborhood. Mixed results are found in prior research depending on the country in which the residential real estate is analyzed. For example, a negative relationship with the housing price is found by Eicholtz et al. (2013) and Lazrak et al. (2014) in the U.S. and Netherlands. However, opposite results are found in Germany by Cijas et al. (2016) with a positive relationship between housing prices and population density. This is also found by Ayala et al. (2016) for Spanish cities in the north, south, and center and by Taltavull de La Paz et al. (2019) for Alicante, Spain. 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 indeed that a higher demand for housing prices increased housing prices. However, Jeanty et al. (2010) also found that this increase in housing prices will lower population growth. Thereby reducing the increase of the housing prices in the long run. In Barcelona, Buonanno et al. (2013) only studied the immigration rate and found evidence 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 on the premium paid for properties that are monumental or on properties that are located close to monuments or places for recognition of religion and culture. This prior 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 in the Netherlands respectively a premium of 17.6% and 26.9% which was paid for properties with a monumental status. In addition, Lazrak et al. (2014) observed a significant spillover effect with a positive premium paid of 0.28% for properties within a 50-meter radius of monuments. Besides, the attention to the recognition of culture in neighborhoods can also have significant positive relationships with housing prices. For example, Brandt et al. (2014) observed in Germany a significant positive relationship between the closeness to the place of worship and the housing prices. Graells-Garrido et al. (2021) found evidence for the same positive correlation in Barcelona. 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, 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 studied in earlier work by the effect of income equality and the 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. Marmolejo-Duarte & Chen (2022) stated the same conclusion in future work using 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 suggested to be positively correlated with higher 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, Brand et al. (2014) and Taltavull de La Paz (2019) found that a higher average income results in significant increases in housing prices in respectively Germany and Alicante, Spain. Similarly, Mandell et al. (2011) found in Sweden that higher economic activity in an area is positively correlated with housing prices(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 higher sustainability and housing prices for all five dimensions of sustainability as proposed by Kauko (2019): ecological, environmental, social, cultural, and economic-financial. Thereby for each dimension of sustainability in this report of research, an increase in the willingness to pay for housing can be expected when the factors associated with the respective sustainable dimension increase. I have defined this by hypothesis 1-5 (H1 – H5), 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, I excluded the price effect of the energy label from the hypothesis (H1) because it is the only sustainable variable that is not location-bounded. Related to the economic-financial dimension prior work did find a negative effect for higher income equality, and a positive effect of higher welfare on housing prices is found in earlier work. Therefore, I expect a positive effect of higher welfare and a negative effect of higher income equality on housing prices in hypothesis 5 (H5). All of the in-depth sustainable variables studied in the hypothesis (H1-H5) in this report of research are location-bounded. Hence the price impact of sustainability is expected to show local clustering tendencies across Barcelona. I defined this in hypothesis (H6) stating that the price impact of sustainability on housing prices shows local clustering tendencies between areas in Barcelona.

> **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

In the data and methodology section, I included a description of both the collection process of the housing-specific and sustainable data in respectively Section 3a and Section 3b. I retrieved two samples. One sample includes the observations with missing energy labels and one excludes the observations with missing energy labels. I describe the valuation models for the empirical analyses of this report of research in Section 3c and the construction of the demonstrative maps to visualize the empirical results in Section 3d.

Section 3.1: Housing-specific Data

I collected the housing prices and housing-specific features from Idealista (2023a). Idealista is the most popular and biggest housing platform in Spain (similarweb, 2023). More than 16.000 residential properties are listed for sale on Idealista in Barcelona at the time of writing this research in April 2023. I retrieved the data used in this research from all the housing advertisements in Barcelona from Idealista on April 17, 2023, by web scraping. The housing advertisements on Idealista only show the asking price and thereby represent only the supply side. However, McGreal et al. (2010) showed that the asking prices tend to be close to transaction prices in a non-rapidly rising boom or bust housing market. 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).

I have done the web scraping with the use of the Undetected Chromedriver library in Python (pypi.org, 2023a). To start, I first visited with the Undetected Chromedriver the list pages for all the districts of Barcelona that are specified on Idealista. I installed an alarm to signal a CAPTCHA, which prevented the content on the webpage from being shown. I could identify the CAPTCHA by the length of the HTML code of the webpage. After the CAPTCHA was solved and the key q was pressed on the keyboard the alarm would turn off and the web scraping could continue. From the list pages, I collected general information on the housing advertisements in Barcelona, such as the URL-link to the housing advertisement, title, house-id, and the price of the property. I collected this data by reading the webpage with Beautiful Soup, which is a Python library used to read HTML documents (pypi.org, 2023b). Since each list page has the same structure on Idealista, I could write a code that contained information where the general information was included to read the HTML-code of the list pages and collect the general information. For some districts, the number of housing advertisements exceeded 1.800. Therefore not all the houses would be shown by visiting only the list pages of the subdistricts since a list page only includes 30 listings and the maximum page number is 60. For those districts that had over 1.800 housing advertisements, I visited the list pages of the subdistricts. In the end, after I removed the duplicates by their house-ids, I collected a sample of 16.641 housing advertisements by web scraping the list pages. I did not take into account that the same property might be listed by multiple agencies. To retrieve more housing-specific information, I used the URL-link to the property-specific advertisement which was mentioned on the list pages to visit the property-specific webpages. Again, by reading the HTML-code with Beautiful Soup, I could collect information from the housing advertisement. The housing advertisements contained more housing-specific information, which was not shown on the list pages. The housing-specific information I collected consisted of among others the list price, the size of the property, the facilities, the location, the agency selling the house, and the house-id. I shared the code of the web scraping on GitHub.

After I collected all the data by web scraping I started the first steps of the data treatment since not all the web scraped from Idealista is in the suitable format. To start, some of the data on Idealista is provided in lists, such as the building features or basic features. Thereby, to identify the specific facilities I created initial dummy variables. An example is the dummy variable elevator which I set in the beginning to zero for all properties. I changed the zero of the dummy variables to one if the facility was present. For instance, for the dummy variables for the presence of an elevator, the list should contain the word "With Lift" to be changed to one. I repeated this process for all the housing facilities or features that are not on an individual basis displayed in the housing advertisements on Idealista. Also, continuous variables are included in lists in the housing advertisement on Idealista, such as the building surface m^2. These variables are displayed as the variable value followed by the variable name. To clean those continuous variables I collected first the element in the list that included the variable value and variable name into an empty list for all the properties in the dataset. If the continuous variable was not present in the housing advertisement I appended an empty string to the list. Afterwards, I removed the variable name for each element in this list. For example, for the variable building surface m^2, I removed the word "built". I appended the list with all the values of the continuous variable to the sample after the removal of the variable name. Afterwards, for the remainder of the features, I made sure the datatype was correctly specified in the sample. I shared this code to treat the data from the sample on GitHub. I included an overview of all the collected housing-specific variables 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.

I treated missing values in the housing-specific variables in this research by both assuming similarity between properties in a neighborhood and the exclusion of observations. I observed missing data for the building year in approximately 4900 (30%) of the housing advertisements. Therefore, if the building year is missing in the sample, I assumed the building year to be the median of the subdistrict in which the residential property was located. I observed missing floor-level data in approximately 2100 (13%) of the housing advertisements. I excluded these observations with missing floor-level data from the original sample. The last variable, I often observed with missing values was the consumption energy label. The energy label was not reported for approximately 3500 (25%) of the residential properties in the remaining sample. 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. I could the sample selection bias with the use of the Heckman selection model by estimating two regressions (J.J. Heckman, 1976). Firstly in the Heckman selection model, 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 transformed 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 (IMR) as explanatory variables. With this method, I can include the observations with missing energy label data in the sample. For completeness, I also discussed the results of the sample excluding observation of the missing energy labels in the robustness section. This helps to identify potential errors in the pricing models. The only downside of the Heckman selection model in this report of research is the low pseudo r-squared of 2%. This is 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, I observed a negative building age of minus 1 for four observations of residential properties that are still under development. Therefore, I 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. I replaced the building surface m² of zero in these four observations with the correct building surface m² of 1.000.

I identified outliers for the continuous housing-specific characteristic variables for the log housing price, building surface in m², building age, and housing price per m². I identified the outliers on the sample that included the observation with missing energy labels. I did not select the variable housing price per m² as both predicting or explaining variables in any of the models. However, I observed that this variable captures the highly non-linear relationship between the log housing price and building surface in m², which I could not capture this relationship in the pricing models by the inclusion of additional predictors. So, without the removal of outliers for the log housing price per building surface in m², I would have an incorrect model by not including all the relationships between the outcome and explanatory variables in the model. It makes it necessary to include it as the variable for identifying outliers. Figure 1 below shows the distribution of the variables. Before the treatment of outliers extreme values 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

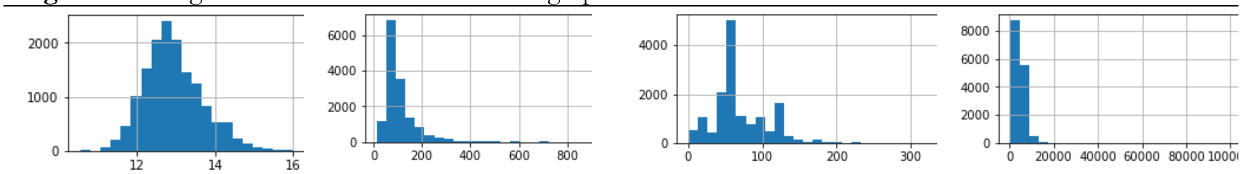


Fig 1a: Histogram of the Log Price | Fig 1b: Histogram of the Building surface m^2 | Fig 1c: Histogram of the Building age | Fig 1d: Histogram of the Price per Building surface m^2

I excluded the observations that have a value for one of these variables shown in the histograms which is two 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. I excluded with this treatment 1538 (11%) observations of the sample, which is more than under the normal distribution. However, for each continuous variable selected for the outlier treatment I excluded approximately 400-600 (2.5%) of the observations. This percentage is lower compared to the Gaussian normal distribution for the variables. Furthermore is it lower compared to earlier research on Barcelona, for example of Chen et al. (2018), who excluded all the observations that had a housing price of one standard deviation above or below the average housing price. A stricter treatment for outliers was not needed in my research since the continuous variables already present a more normal distribution after the exclusion of the observations that were two standard deviations below or above the mean for the continuous variables as shown in Figure 2 below.

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

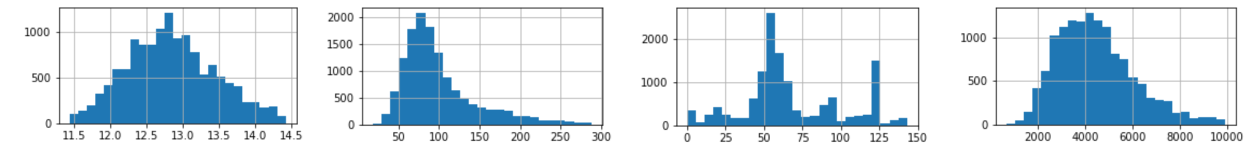


Fig 2a: Histogram of the Log Price | Fig 2b: Histogram of the Building surface m^2 | Fig 2c: Histogram of the Building age | Fig 2d: Histogram of the Price per Building surface m^2

The final sample for the housing-specific characteristic sample including observations with missing energy labels consists of 13.358 observations compared to the sample excluding observations with missing energy labels 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. So, the housing-specific continuous variables show 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 variables in the sample 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 have a good variety. The house and energy consumption label A housing-specific dummy variable dummy variables have the lowest presence. However, these dummy 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 notable difference that we can observe is the decrease in the presence of the new housing development housing-specific 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	**mean**	**Variable**	**mean**
House	0.02	Energy label consumption B	0.02
Mezzanine	0.05	Energy label consumption C	0.04
Ground Floor	0.10	Energy label consumption D	0.09
1st Floor	0.23	Energy label consumption E	0.39
2nd – 5th Floor	0.49	Energy label consumption F	0.07
6th Floor or higher	0.10	Energy label consumption G	0.13
New housing development	0.04	District Eixample	0.21
Good condition	0.82	District Ciutat Vella	0.15
Needs renovation	0.14	District Sant Martí	0.10
Elevator	0.77	District Sants-Montjuïc	0.11
Terrace	0.32	District Horta Guinardó	0.09
Heating	0.51	District Gràcia	0.07
Outdoor facilities	0.57	District Les Corts	0.04
Air conditioning	0.98	District Nou Barris	0.06
Parking space included	0.10	District Sarrià-Sant Gervasi	0.10
Energy label consumption A	0.02	District Sant Andreu	0.06

Table 4 includes the summary statistics for the housing-specific dummy variables for the sample including observations with missing energy labels. The sample consists of 13.358 observations.

Section 3.2: Sustainability Data

I retrieved the data for the creation of the sustainability variables except for the energy labels 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. I used the datasets that include sustainable information in my research if they satisfy two conditions. First of all, the dataset should include at least subdistrict-specific information. Secondly, the data in the dataset should not have a too high correlation with other used sustainable data. Altogether, the sustainable information that I retrieved from the City Council of Barcelona covers a wide range of aspects of sustainability. It includes among others 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, I retrieved this information by the Overpass API (Overpass Turbo, 2023). The Overpass API reports information on the location of these variables from OpenStreetMap, which is a free wiki world map managed and hosted by volunteers (OpenStreetMap contributers, 2023). In Table 6 in the appendix, I included 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.

The retrieved sustainable information from the datasets of the City Council of Barcelona and OpenStreetMap consists of three types of data. Depending on the type the data, I transformed the content of the datasets into variables containing sustainable information. The first type of data that I retrieved 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, I used the average value. Additionally on Idealista, a lower number of subdistricts are specified compared to the datasets 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, I either calculated the weighted average of the statistic by the number of inhabitants or used the sum of the statistic. I did not observe outliers for the statistical subdistrict-level variables since the reported numbers are the average numbers of the subdistricts. The second type of data I retrieved contained geographic information with benchmark scores of an area compared to other areas in Barcelona. The location of the areas with a specified score is included as multiline strings in the datasets with the geographical information. I matched these multi-line strings with the location of the properties, which is specified by the latitude and longitude on Idealista, to find the correct score for the properties. I replaced missing values with the median value of the subdistrict where a

[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

property was located if the longitude and latitude of a property did not match with any of the multiline strings in the dataset. I observed no outliers in the datasets with geographic information with benchmark scores since the data is scaled in pre-defined levels. The third type of data I retrieved contained information about the location of sustainable variables, such as the location of bus stops, parks & gardens, or universities. I calculated the distance of the property to each of these locations of the sustainable variables included in the dataset by the haversine formula by using both the latitude and longitude of the property and the latitudes and longitudes of the locations of the sustainable variable. The haversine formula calculates the great circle distance between two locations on a sphere. The calculated distance by the haversine formula shows a high correlation with the actual travel distance. For example, Phibbs et al. (1995) showed that for the distance to hospitals in upstate New York, the United States, the correlation between the calculated distance by the haversine formula and Google Maps was 0.826 for distances less than 15 miles. I could not calculate the exact travel distance/travel time since APIs to calculate the travel/distance travel time, such as Distance Matrix API from Google or BING allow no or only a limited amount of search queries for free in a given period. This number of queries is too low for my research given the high number of observations in the sample in combination with the high number of sustainable variables. With the calculated distances by the haversine formula, I created two sustainable variables: the distance to the nearest location of the variables and when relevant the number of locations of the variable within a specified range. I made this specified range dependent on the expected distance where the presence of a location of the variable is expected to add utility for a homeowner. For example, related to a hospital, I reasoned that it is expected to provide utility to a homeowner if it is present within a range of 1 kilometer of the property. However, for example, related to a bar or restaurant, I reasoned that only is expected to provide utility within a range of 0.25 kilometers to a homeowner. I truncated outliers of the sustainable variables that contain information about the distance to the nearest location and the number of locations within a prespecified scale for the locational sustainable variables on the upper range. This is to retrieve a more normal distribution for the sustainable variables. Furthermore, I reasoned 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. In Table 7 shown below, I included information about the applied truncation levels for the sustainable variables.

To avoid high multicollinearity between variables, I constructed PCA components for the sustainable variables that have a high correlation with each other. Since the sustainable variables often contain information on similar types of amenities. Oladunni & Sharma (2016) found that PCA components are suitable for predicting housing prices with traditional hedonic pricing models. They found that PCA explains more variance of housing prices compared to other methods, such as the support vector machine (SVM) and K-nearest neighbors (KNN). In my research, 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, in my research, most of the created PCA components represent an accessibility/access to indicator for the included variables. I ensured that the coefficients for the minimum distance variables have negative signs and that the coefficients for the number within the prespecified range variables have positive coefficients. So a higher score for the PCA components in my research represents a higher accessibility. This offers interpretation benefits for the results of the pricing models. In addition, I ensured comparability between the different PCA components by using a constant scale (0-1) during the creation of the PCA components. So, all the values of the PCA components are distributed within a range of 0 to 1. The only PCA components, which I did not create based on features with distance information are the income distribution PCA component and the income & unemployment PCA component. I created these PCA components based on the economic statistical data of the subdistricts. Before the construction of the PCA components, I scaled these economic statistics to similar a scale to retrieve similar coefficients for each of the input variables during the creation of the PCA component[2]. Additionally, I included the squared value of the statistic when creating the PCA components to capture the non-linearity of the income equality and welfare statistic. The economic statistical PCA components have the interpretation that a higher value implies higher income equality for the income distribution PCA components and higher welfare for the income & unemployment PCA components. In Table 7, I included an overview of the created PCA components, the explained variance by the PCA components, the weight given to the input sustainable features, and the applied truncation levels of the sustainable features. As shown in Table 7, the lowest share of the variance that is explained by a PCA component is approximately 72.25% of the input features for the Second & Lower School PCA component. This is the PCA component that represents the

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

highest number of variables. However, I cannot split these PCA components into multiple components. By splitting the PCA component into multiple PCA components I would introduce 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]
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]

Variable	E. Var %	Feature	Coef.	Trun.
Performing Arts PCA	89.94%	Distance to nearest cinema (km)	-0.007	[0,2.5]
		Number of cinemas within 0.5 km	0.011	[0,2]
		Distance to nearest theatre (km)	-0.007	[0,2.5]
		Number of theatres within 0.5 km	-0.087	[0,10]
		Distance to nearest concert place (km)	-0.013	[0,2.5]
		Number of concert places within 0.5 km	0.013	[0,2]
Religious Institution PCA	99.94%	Distance to nearest religious institution (km)	$-9*10^{4}$	[0,1.5]
		Number of religious institution within 0.5 km	0.067	[0,15]
Museum, Library & POI Cult. PCA	93.01%	Distance to nearest library (km)	$-1*10^{4}$	[0,1.5]
		Number of libraries within 0.5 km	0.004	[0,10]
		Distance to nearest museum (km)	$-3*10^{4}$	[0,1.5]
		Number of museums within 0.5 km	0.003	[0,10]
		Distance to nearest point of interest culture (km)	$-1*10^{4}$	[0,1.5]
		Number of point of interest culture within 0.5 km	0.019	[0,50]
Income Distribution PCA	97.65%	Income Distribution P80/P20/10	-2.134	NO
		Gini Index	-2.178	NO
		sq(Income Distribution P80/P20/10)	-1.324	NO
		sq(Gini Index)	-1.498	NO
Income & Unemployment PCA	99.05%	AVG income household in € 100.000	0.365	NO
		AVG unemployment rate %	0.022	NO
		sq(AVG income household in € 100.000)	-0.496	NO
		sq(AVG unemployment rate %)	0.002	NO

Table 7 includes the name of the sustainable 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

I included in Table 8 a final description of the sustainable features of the sample that includes the observations with missing energy labels. The summary statistics for the sustainable variables show a high variety based on the standard deviation, 25%, and 75% quantile. This suggests that the results of my research will show high differences between properties regarding 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 are similar as shown in Table 8 and Table 9 in the appendix.

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	0.00	0.24	0.36	0.48	1.00	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	1.40	0.01	2.14	3.44	4.61	5.00	-0.37	-1.02
Park & Garden PCA	0.40	0.39	0.00	0.01	0.50	0.51	1.00	0.37	-1.33
Viewpoint PCA	0.41	0.26	0.00	0.21	0.32	0.64	1.00	0.67	-0.79
Neighborhood size (10 ha)	13.51	12.20	2.30	8.04	11.10	14.1	142.37	5.95	51.28
Vulnerable to heat impact (1-5)	2.86	0.79	1.00	2.00	3.00	3.00	5.00	0.08	-1.03
Police PCA	0.44	0.28	0.00	0.29	0.32	0.55	1.00	0.33	-0.78
Bar & Restaurant PCA	0.49	0.35	0.00	0.19	0.34	0.87	1.00	0.30	-1.36
Secondary & Lower School PCA	0.47	0.20	0.00	0.32	0.47	0.62	1.00	0.00	-0.60
University PCA	0.35	0.38	0.00	0.05	0.06	0.62	1.00	0.79	-1.04
Pharmacy PCA	0.50	0.22	0.00	0.40	0.50	0.60	1.00	0.00	-0.20
Hospital & Clinique PCA	0.36	0.28	0.00	0.21	0.22	0.60	1.00	0.67	-0.34

Variable	mean	std	min	25%	50%	75%	Max	skew	Kurt
Big Shopping Place PCA	0.33	0.30	0.00	0.08	0.20	0.54	1.00	0.91	-0.48
Social Cohesion Score	0.21	0.18	0.00	0.09	0.15	0.30	1.73	1.81	4.73
Natural population growth ‰	-2.09	1.96	-8.92	-3.30	-2.50	-0.60	3.60	0.14	0.62
Net immigration rate ‰	26.86	22.41	-6.60	9.90	18.30	44.1	91.10	1.18	0.89
Density net ((hab/1000)/ha)	0.74	0.23	0.02	0.63	0.74	0.91	1.37	-0.40	0.10
Performing Arts PCA	0.30	0.30	0.00	0.06	0.17	0.41	1.00	1.18	0.11
Religious Institution PCA	0.44	0.25	0.00	0.27	0.40	0.60	1.00	0.56	-0.42
Museum, Library & POI Cult. PCA	0.25	0.23	0.00	0.09	0.16	0.34	1.00	1.35	1.01
Income Distribution PCA	0.46	0.21	0.00	0.30	0.47	0.62	1.00	-0.34	-0.59
Income & Unemployment PCA	0.20	0.20	0.00	0.07	0.15	0.27	1.00	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

I included in Table 6 the housing-specific, sustainable, and dummy district variables used to estimate the housing prices in this report of research. 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. I chose to include the energy consumption label in the sample over the energy emission label since it is more directly associated with the living costs for home buyers. Moreover, I observed that the energy consumption label is more often reported in housing advertisements and is always reported if the energy emission label is reported.

I represented the sustainable pricing factors in this research 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. I included the access to short commuting options as one sustainable variable since there is a high correlation between bus and metro accessibility. I included the access to long commuting options as one variable because I observed a trade-off between the distance to the highway and train in Barcelona. I excluded the energy consumption label from the ecological dimensions of sustainability and included it in the housing-specific characteristic. Since it is the only sustainable variable in my research that is not location but property-bounded. Also, it is impossible to increase the sustainability score of the energy label by local policy intervention that only is implemented in certain neighborhoods in Barcelona. Policy intervention for energy labels will probably consist of support programs available to every neighborhood 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 environmental characteristics of the area of the property: the neighborhood size and vulnerability level to heat impact. The pricing factors for the social dimension, $Z_8 - Z_{18}$, 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.

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

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

Additionally, I included dummy districts district variables in the valuation model. The dummy variable captures the effects of geographical pricing factors on a district level that are not included in the model. This decreases the probability of potentially wrongly specified pricing models and an overestimated strength of the coefficients of the sustainable variables caused by an omitted variable bias.

The valuation model I used to estimate the housing prices in Barcelona is the hedonic pricing model as proposed by Rosen (1974). The hedonic pricing model assumes that the price paid for housing is equal to the total utility buyers are expected to extract from it. The term "utility" refers to the satisfaction/benefit expressed in a monetary value individuals are expected to retrieve from buying the property. The total utility is the sum of the marginal utilities of the housing-specific, sustainability, and dummy-district variables. The hedonic pricing model or similar models are often used in prior literature in this field. Ten years ago the general conclusion of the research was that machine learning methods do at best equal the performance of the hedonic pricing model (Kauko, 2019). Although, lately, there is provided more evidence that machine learning models outperform traditional economic models, such as the hedonic pricing model, by achieving a higher accuracy (Foryś, 2022; Kok et al., 2017; Mora-Garcia et al., 2022; Rico-Juan & Taltavull de La Paz, 2021). This is caused by the advantages that machine learning models can offer over traditional economic models. For example, most machine learning models can cope well with big amounts of data, deal with bad hierarchization of data, measurement errors of data, and allow for non-linear variables (Mora-Garcia et al., 2022). A higher accuracy reflected by lower pricing errors increases the fairness when valuing houses. (Vargas-Calderón & Camargo, 2022) However, the goal of this study is to provide deep insight into the price impact of sustainability on housing prices. For instance to suggest policy recommendations for a more equal pricing of sustainability across Barcelona. Machine learning algorithms are mostly not developed to serve this purpose or other political purposes (Ho et al., 2021). The weakness of most machine learning models is that they are at times weak at explaining the rationale behind price levels (Rico-Juan & Taltavull de La Paz, 2021). Only the more advanced state-of-the-art models can identify the characteristics that contribute mostly to the housing prices. This in combination with the wish to maintain comparability with prior research, which also mostly used traditional economic models, resulted in the decision to estimate the housing prices in my research by a traditional economic model.

The hedonic pricing model applied in my research is the semi-log hedonic pricing 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 the results of this research and compare them with earlier work. (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). I estimated the semi-log hedonic pricing model for only the housing-specific variables, for each of the sustainable dimensions individually, and for all the sustainable dimensions. I specified the pricing models as follows:

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, the $X's$ are the district dummies, and ε is the error term. I included a description of the variable in the models in Table 10. In each model, I included the housing-specific pricing factors and dummy district variables. In Models 2-6, in addition, I included one dimension of sustainability by the sustainable pricing factors. In Model 7, the complete model, I included all the housing-specific characteristics, sustainable pricing factors, and dummy district variables. Homoscedasticity of the residuals of all the valuation models was rejected by the Breusch-Pagan test at the 1% significance level. Therefore, I used robust (Huber-

White) standard errors in the estimated pricing models to allow for heteroskedasticity in the estimated pricing models (Guggisberg, 2019).

I verified the assumptions for a correct specification of the pricing models 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 containing the code for the estimation of the pricing models. Firstly, before the model estimation, I calculated the correlation matrix for the included predictors in the models to identify and treat variables that have a too high correlation with each other. Variables included in this 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, I included the correlation matrix for all the variables of both the sample that includes and excludes observations with missing energy labels in respectively Figure 3 and Figure 4 in the appendix. We can observe that the correlation matrixes show no warningly high correlations between any of the variables. Besides, we can see that the correlation matrixes for both samples have a high similarity with each other. Secondly, before the model estimation, I calculated the variable inflation factor (VIF) 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 I will choose to either exclude variables in the model or create additional PCA components. I will discuss the results for the VIF statistics before the discussion of the model results in Section 4.

In line, also after the model estimation, I reported some statistics to verify the model assumptions. These statistics can also be found on the shared code for the estimation of the pricing models on GitHub. First of all, I reported the skewness, kurtosis, and histogram for the residuals of the models. I calculated the residuals by subtracting the actual log housing price from the predicted log housing price. I report these statistics to verify the assumption of a normal distribution of the error term in the semi-log hedonic pricing model. Secondly, I calculated the Ramsey Reset test for the results of the model. The Ramsey Reset verifies by the inclusion of the squared values of the predictors in the model a correct model specification (Ramsey, 1969). The reported tests-statistic of the Ramsey Reset is a p-value for an F-test to test that the coefficients of the squared values are not significantly different from zero. If the F-test is rejected these additional squared variables do not significantly increase the predictive power of the model. I will discuss the results of the Ramsey Reset test statistic for each pricing model in the result section. Thirdly, I test the homoskedasticity of the error term 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 in the distribution of the pricing errors does not hold. As mentioned above, homoscedasticity is for all the valuation models rejected at the 1% significance level. Hence, I appended robust (Huber-White) standard errors to the valuation models. The robust standard errors allow for heteroskedasticity in the pricing errors of the semi-log hedonic pricing models. I used the robust standard errors also in the Ramsey Reset test discussed above. Fourthly, I made correlation matrixes between the squared residuals and the predictors in the model to verify once more a more correct specification of the pricing models. I observed in none of these correlation matrices a high correlation between the squared residuals and any of the predictors.

I estimated the Heckman selection model to include the observations with missing energy labels in the valuation models for the sample containing observations with missing energy labels. The Heckman selection model solves the sample selection bias by predicting whether or not the energy label is present in a housing advertisement with a Probit model. With these predicted variables for this dummy variable, I calculated the Inverse Mill ratio (IMR). If the Inverse Mill ratio (IMR) is included in the above-specified hedonic pricing models (Models 1-7) observations with missing energy labels can be included in the sample. The Probit model, on which the Inverse Mill ratio (IMR) is based, 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. I specified the Probit model as follows:

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 be correlated with the presence of the energy label in the housing advertisement, and ε is the error term. The homoskedasticity of the residuals is rejected by the Breusch-Pagan for the Probit model. To allow for heteroskedasticity, I used robust (Huber-White) standard errors for the Probit model. Additionally, I conducted the same tests to verify the model assumptions for the valuation models.

Section 3.4: Visualization of the Results by a Geographical Map

I developed geographical maps to display the results of the pricing models. I built the map with the use of Leaflet, which allows interaction by users with the geographical map. For example, a user can zoom in and out on the map. Based on the zoom level the properties will be clustered (Leaflet, 2023). I shared the code to build the map, even as the results of the estimated pricing models, on GitHub. In my shared code different parameters can be changed to build different types of maps, which I describe later in this section. With the code, I constructed two demonstrative maps (Demonstrative Map 1 and Demonstrative Map 2), which I will discuss in the results section. Furthermore, I constructed two demonstratives (Demonstrative Map 3 and Demonstrative Map 4), which I will discuss in the robustness section. The difference between the demonstrative maps in the result sections and the robustness section in the clustering method. In both sections, I visualize with the two demonstrative maps the price impact of sustainability on housing prices on an area-specific scale for two different sets of sustainable variables The total price impact of sustainability is calculated by multiplying the coefficients of a group of selected sustainable variables by choice with their values for the sustainable variables for a pricing model of choice. In the geographical maps, a high price impact of the selected sustainable variables on the housing price is assumed to be positive, and a low price impact of the selected sustainable variables is assumed to be negative. The geographical maps provide helpful insights since the sustainable variables used in my research are mostly location-bounded.

The code to construct maps provides different options to visualize the price impact of the variables in the pricing models. I have included examples of these methods in a shared slide show on GitHub. 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, I included the option to display also general information on the total price impact of a list of selected sustainable pricing factors for two reasons. Firstly, given that some sustainable variables in the pricing models in this research show non-positive relationships between higher sustainability and housing prices. Secondly, to exclude some sustainable variables that cannot be changed by policy interventions, which are included in the pricing models of this research. 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 have a lower relevance when visualizing the total price impact of sustainability factors on the housing prices in Barcelona reported in this research. More specifically, I also provide in the demonstrative maps information for the individual housing-specific, sustainable, and dummy district variables.

All the observations cannot be displayed at once given the large sample size. Hence, I included three options to visualize the results either on a housing-specific, neighborhood, or subdistrict, level in the geographical maps with the latitude and longitude of the properties. The first option I have included involves automatically clustering the observations by Leaflet. This method allows to display of information regarding the individual residential properties only when clicking on and opening clusters. The second option (SVM_cluster) I have included clusters of multiple observations into one observation 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. The created clusters 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 is suppressed. The third option (Subdistrict_cluster) I have included clusters of the properties in the sample according to the 68 subdistricts in Barcelona on Idealista (Idealista, 2023a). I retrieved the coordinates of the subdistrict from the City Council of Barcelona (Ajuntament de Barcelona, 2023b). The subdistrict clusters contain information on the average value of the sustainable variables of all the properties in a subdistrict making 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.

In Table 11, I included the parameters to construct the geographical maps with a description. The parameters to construct the map offer high flexibility. I included this flexibility for the selected valuation models, the selected shown observations, the variable(s) to color the map, the number of categories to color the map, and the clustering method. The shown observations can be filtered 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 specified quantiles of the continuous variable used to color the map. I decided to always visualize the categories in three colors. The lowest category will be visualized in red, and the highest category will be visualized in green. The observations in the other categories will be visualized in gray. Additionally, I included an option to only display the observations in the lowest and highest categories (Show_all). I shared the code to construct the geographical maps on 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 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 will 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 that 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.

Parameters	Description
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.

Related to the demonstrative maps, in Demonstrative map 1 I show the total price impact of sustainability as calculated by pricing Model 7 when the properties are clustered by their subdistricts. A screenshot of the demonstrative map is provided in the results (Section 4.2.1) and the HTML code is shared on GitHub. The clusters in the highest 10% quantile for the total price impact of sustainability, calculated by multiplying the values of all sustainable features in Model 7 times their coefficients, are displayed in green. The cluster in the lowest 10% quantile of the total price impact sustainability is displayed in red. The clusters in the 10% – 90% quantile of the total price impact of sustainability are displayed in grey. In Table 12, I included the parameters used to construct Demonstrative Map 1.

In Demonstrative Map 2, I show the total price impact of selected sustainable features by clustering the properties into clusters by the subdistricts. These selected sustainable features are all the sustainable features included in Model 7 except the sustainable features that have a high negative relationship between a higher sustainable score and the housing price and/or cannot be influenced by political measures/interventions. The excluded sustainable features are the distance to the beach (km), neighborhood size 10 (ha), and income distribution PCA. I included a screenshot of the demonstrative map in the results Section 4.2.2 and shared the HTML code on GitHub. The applied color scale in Demonstrative Map 2 is equal to those applied for Demonstrative Map 1. In Table 12, I included the parameters to construct Demonstrative Map 2.

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'

Parameters	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'
Show_all (True/False)	True	True
SVM_Cluster (True/False)	False	False
Subdistrict_Cluster (True/False)	True	True
Save (True/False)	True	True

Table 12 includes the parameter values to construct Demonstrative Map 1 and Demonstrative Map 2

Additionally, I constructed two demonstrative maps (Demonstrative Map 3 and Demonstrative Map 4) to visualize the total price impact of sustainability on the predicted housing prices by clustering the properties in 100 clusters based on their location with the support vector machine (SVM) (by the latitude and longitude of the properties). I applied a different visualization method in these demonstrative maps to test the robustness of the findings for hypothesis 6 (H6), which states that the price impact of sustainability on housing prices in Barcelona shows local clustering tendencies. The clustering method is the only difference between Demonstrative Maps 3 and 4 compared to Demonstrative Maps 1 and 2. In Demonstrative Map 3 I used the same sustainable variables to color the map as for Demonstrative Map 1. In Demonstrative Map 4 I used the same sustainable variables to color the map as for Demonstrative Map 2. The advantage of clustering by the support vector machine (SVM) is that the size of the clusters varies less compared to clustering by the subdistricts. The disadvantage of clustering by the support vector machine (SVM) is that in practice also policies are implemented on a subdistrict level. Hereby, implementation of the results can be harder. In Table 13 in the appendix, I included the parameters used to construct Demonstrative Map 3 and Demonstrative Map 4.

Section 4: Results

I discuss in the results section the findings of the valuation models (Section 4.1) and the visualization of the results of Model 7 by the demonstrative maps (Section 4.2). Each section consists of multiple subsections related to the discussed valuation model or demonstrative map. Additionally, in Section 4.1, I included a subsection for the discussion of the findings of the Probit model to predict the probability of the presence of an energy label in the housing advertisement.

Section 4.1: Results of the pricing models

In Section 4.1, I discuss the Probit model to predict the probability of the presence of an energy label in housing advertisements. Besides that, in this section, I discuss the results for the pricing models (Models 1-7) that estimate 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 us that only the air conditioning dummy has a high VIF test statistic (12.90.). This can imply potential multicollinearity in the model. However, the high VIF test statistic is likely only 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. It indicates that the sample selection bias is smaller than in prior work. This is also 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 m² 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 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, the are results different compared to those found in the work of Chen & Marmolejo Duarte (2018), which estimated a similar to predict the probability of the presence of an energy in housing advertisements in Barcelona to address the sample selection bias. 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, I could not identify additional variables in the housing advertisements that increased the explained variance by the model. Furthermore, there are no other housing-specific variables that have a high correlation with the presence of the energy label in the housing advertisements on 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 m^2 and building age. The high VIF test statistic for the building surface m^2 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. I observed that the p-value of the Ramsey Reset test would be lower if the squared values of the building surface m^2 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 surface m^2 has a positive correlation with the property price as shown 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. This is 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, the results show 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 can 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 to decrease the probability that the model will overfit on a dummy variable with a low presence given the low presence of the individual dummy variables relative to the sample size. 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 evidence of a significant discount (-17.2%). The 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 dummy variables. The significant negative coefficient for the outdoor facilities coefficients is in contrast to earlier 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 the sample of this research, 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 and is therefore not a positive pricing factor when buying a property. The policies to prevent high water consumption were not implemented and/or relevant at the time of earlier research (Ajuntament de Barcelona, 2023c). Another reason I suggest could be the higher cost of living when maintaining a garden, green area, and/or swimming pool, which can be higher than the associated monetary value of the expected utility of the outdoor facilities.

The energy consumption label dummy variables are the only housing-specific variables that are part of a sustainable dimension in this research (ecological) (Kauko, 2019). The results show that we can observe 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). The trend is not perfectly linearly increasing. However, 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. It suggests that the price of a property is negatively related to the probability that an energy label is shown. These findings are in line with the expectations of the sample selection bias. Furthermore are the results and 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 ecological 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, I 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 likely be equal for and affect every property in the municipality of Barcelona. Hereby it would not directly decrease the difference in the pricing of sustainability between areas in Barcelona. The inclusion of the ecologically sustainable variables increases the R-squared by 0.8 percentage points compared to the pricing model that only includes housing-specific variables (Model 1) to 83.4%. The inclusion of ecological variables is not introducing multicollinearity as 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 a maximum distance of 3 kilometers and an average number of 0.5 within 250 meters. 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, in contrast to this research, 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, Dell'Anna et al. (2019) use 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, I found that the result provides no evidence to reject hypothesis 1 (H1), 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 dimension 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 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. Hence the likelihood that the observed coefficients are subject to a high degree of multicollinearity is low. Misspecification of Model 3 is rejected by the Ramsey Reset test at the 5% significance level when the environmental variables are included in the model (p-value: 0.068). The coefficients for the housing-specific variables are in line with the earlier reported findings of Model 1 and Model 2.

Subject to the environmental variables, a 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, we can observe a significant negative relationship between the accessibility to green space and viewpoints and the willingness to pay for housing shown 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 in the research of Dell'Anna et al. (2019). Dell'Anna et al. (2019) argue that this is caused by the structure of Barcelona since it 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. 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, we can observe that the results show 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 increase in the vulnerability to heat impact. The price impact of these factors, the size of the neighborhood, and vulnerability to a heat impact are not addressed in earlier research on housing prices in Barcelona.

Altogether, I found a significant negative price impact of a higher distance to the beach, neighborhood size, and higher vulnerability to a heat impact on housing prices. This does not provide evidence to reject hypothesis 2 (H2): an increase in the environmental dimension of sustainability increases the willingness to pay for housing. Although, I also 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
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		
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 that 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 as 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. Besides, overall higher VIF statistics for most of the social-related variables 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. However, these VIF test statistics for access to public amenities/services 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%). Therefore the dummy variable can overfit on a low number of observations, where the value of the variable is equal to 1.

The results show that 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, Graells-Garrido et al. (2021) only studied the Spearman rank correlation between variables and housing rent and did not take into account the effect of other pricing factors. The observed significant positive coefficient for the police PCA component is in line with findings for the negative price effect on housing prices of the crime perception rate and perceived security in Barcelona by Buonanno et al (2013). Only a note has to be made that the police PCA component might not perfectly capture the safety of a neighborhood. However, no other relevant data is provided by the

City Council of Barcelona or related sources. 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 components. However, 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 but did not make a distinction 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, Spain, and the work of Graells-Garrido et al. (2021) for Barcelona.

Concerning the demographic social sustainable variables we can observe a significant negative relationship 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 is not analyzed in earlier work in Barcelona. I suggest that the preference for living in less crowded areas could also explain the negative relationship between housing prices and the density of residential areas. However, 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). Jeanty et al. (2010) reported that after a high population growth in an area, people tend to move to other areas causing negative pressure on housing prices. The natural population of this research is represented by past data from 2019. Therefore it could represent a past high/low population growth for an area. Furthermore, I think it could further support my earlier suggestion that the inhabitants of

Barcelona prefer to live in less crowded areas resulting in 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. Although, 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, I found that higher access to public services and amenities, in general, increases the willingness to pay for housing. Related to the demographic statistics, I found that an increase in the density of an area and population growth decreases the willingness to pay for housing. This provides no evidence to reject hypothesis 3 (H3), an increase in the social dimension of sustainability increases the willingness to pay for housing. I found a significant negative price effect of better access to second and lower education, higher social cohesions score, and net immigration rate that reflect more unsustainable preferences in the willingness to pay for housing in Barcelona. On the other side, it might correctly reflect the preferences of home buyers, who are looking to live in more "quiet neighborhoods" 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, accessibility to religious recognition places, and accessibility to 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 has even as for Model 4 an insignificant positive effect on housing prices.

For the cultural-related sustainable variables, we can observe both positive and negative relationships with the willingness to pay for housing. The coefficient of the PCA components of performing arts is significant and negative. It implies that an increase in access 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 places. I suggest that this negative relationship between access to performing arts and housing prices in my research could be related to the possible nuisance during big events. Ossokina & Verweij (2015) found that traffic and noise in streets were found to be negative pricing factors for housing prices in the Netherlands. In contrast, a significant positive relationship is found between the religious institutions PCA component and the 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, by the coefficient, 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 positive spillover effect of monumental buildings on housing prices.

To sum up, in general, I found an increase in the willingness to pay for housing when the culturally sustainable variables increase. This provides no evidence to reject hypothesis 4 (H4), an increase in the cultural dimension of sustainability increases the willingness to pay for housing. I found a significant negative relationship between the PCA component for performing arts and willingness to pay for housing. On the other hand, the results show a significant positive relationship between the PCA components for religious institutions and museums, libraries, and POI culture 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, I included variables 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 variables 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. 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. I already tried to capture this 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 Model 4 and Model 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 we can observe that the results are in line with prior literature. The coefficient for the income distribution PCA components, capturing the degree of income equality, shows a significant negative relationship between income equality and the willingness to pay for housing. The relationship implies that home buyers who can afford more expensive houses tend to cluster in certain neighborhoods in Barcelona driving up housing prices in these neighborhoods. This has a decreasing effect on the income equality measured by the p80/20 income distribution and Gini index in my research resulting in a negative relationship between housing prices and income equality. 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 number of people in high socioeconomic classes and high occupational positions in a neighborhood have a significant positive impact on housing prices. We can observe a significant positive relationship 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).

To sum up, I found no evidence to reject hypothesis 5 (H5), a higher welfare and income equality for the economic-financial dimension of sustainability respectively increases and decreases the willingness to pay for housing. We can acknowledge that an increase in income equality results in lower housing prices and an increase in welfare results in higher housing prices in Barcelona.

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 we interpret 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 10% 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 Model 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 is significant at the 5% confidence level in contrast to the findings of the other models (Model 1 – 6).

The findings for the sustainable variables of each dimension of sustainability are compared for Model 7 with the findings of models containing only one dimension of sustainability. Firstly, for the ecological dimensions of sustainability, the reported findings are different compared to Model 2. The bus & metro PCA component in Model 7 shows 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 variable are in agreement with the results of Model 2 with a significant and negative relationship. However, the results show that the results show that 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 the results 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 the vulnerability of heat impact level almost halved. Thirdly, concerning the social dimension of sustainability, we can observe that the coefficients for the strength of the price impact for most of the socially sustainable variables have become less or sometimes even insignificant compared to Model 4. The social-related variables for which the significant relationship with the housing price has disappeared in Model 7 compared to Model 4 are the police PCA component, the secondary & lower education PCA component, and the net immigration rate. In contrast, the pharmacy PCA components have become significant at the 10% level with a positive effect on the housing price in Model 7. Related to the other social variables the significance of the coefficients remained the same compared to Model 4, 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 in which the strength of the price effect has not been halved in Model 7 compared to Model 4 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 by not including the sustainable variables of other dimensions which are both correlated with the culturally sustainable variables and the housing price. Lastly, the economic-financial sustainable variables have in both Model 7 and Model 6 the same 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, I found that higher sustainability increases the willingness to pay for housing for both the one-dimensional sustainability pricing models and the pricing models that include all dimensions. The results show that higher sustainability increases the willingness to pay for housing. The results show that this holds for all five dimensions: ecological, environmental, social, cultural, and economic-financial as proposed by Kauko (2019). However, there is a change in the strength coefficients for most sustainable variables and a change in the significance for some sustainable variables in the multi-dimensional sustainability model compared to the one-dimensional sustainability model. I argue that this 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 and the significance of the 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 compared to the one-dimensional sustainable models shows us that the model does not overestimate the price effect. Therefore Model 7 is the most conservative model when studying 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

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

In Section 4.2 I discuss the demonstrative maps which visualize the results of the pricing models. The main feature of the demonstrative maps is to visualize the total price impact of selected sustainable variables on housing prices. The price impact is assumed to be positive and high if the values of the sustainable variables times their coefficients are relatively high. For a negative and low price impact, the opposite holds. In Section 4.2.1 I discuss the results that are visualized in Demonstrative Map 1. In Section 4.2.2 I discuss the results that are visualized in Demonstrative Map 2. I shared the code to construct and the demonstrative maps 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 I shared on GitHub.

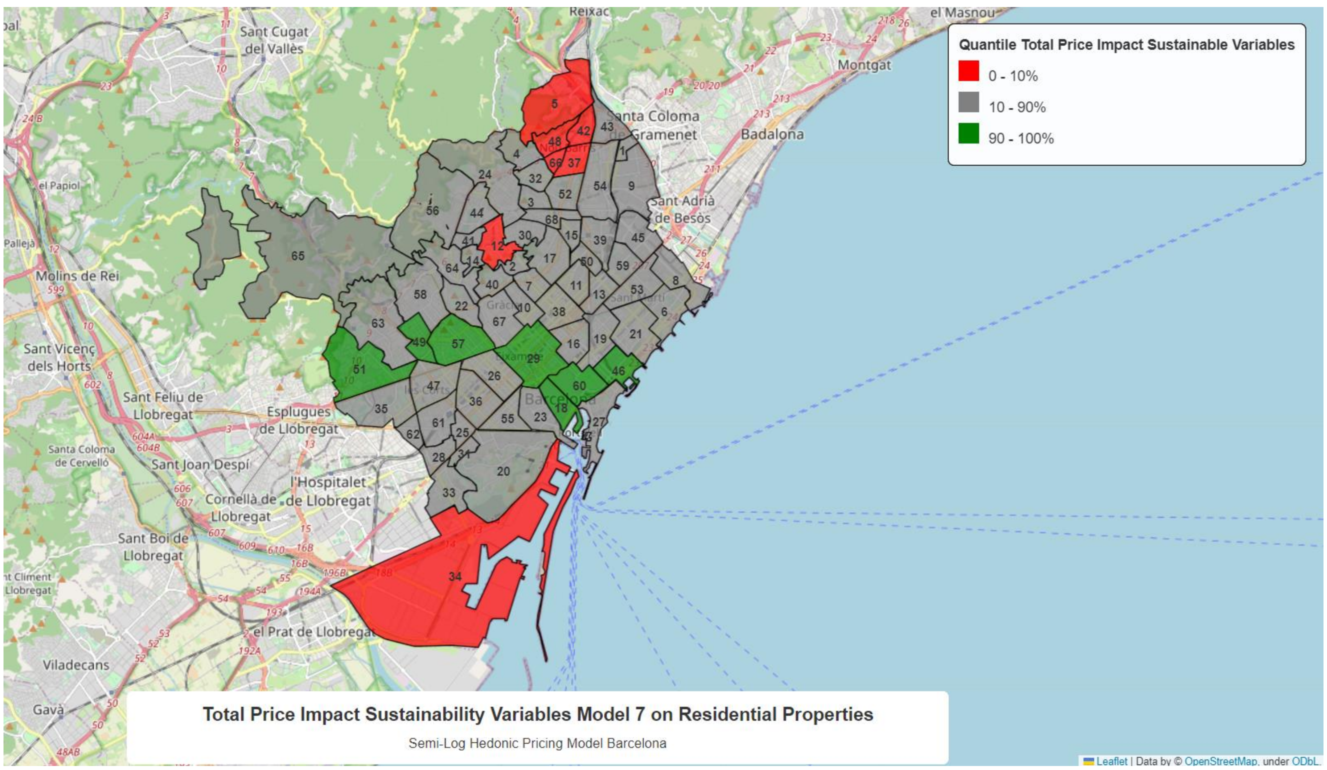
Demonstrative Map 1, of which a screenshot is shown in Figure 5, includes 68 clusters where each cluster is similar to a subdistrict as included on Idealista for Barcelona. Every sustainable pricing factor as specified in Pricing Model 7 is used as a variable to color the total price impact of sustainability 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. I included an index of the names of the subdistricts, which are displayed by the numbers on the map, in Table 16.

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 demonstrative maps show that the subdistricts within the lowest price impact of sustainability have mostly for more than 50% of the accessibility to public amenities and services PCA components values that are lower than at least 80% of the other subdistricts. However, they also sometimes have high values relative to the other subdistricts on some of these accessibility PCA components. In contrast, the demonstrative map shows mostly opposite values for the properties in the subdistricts with the highest quantile of the total price impact of sustainability for the sustainable factors that caused the low price impact of sustainability for the

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

properties in the subdistrict in the lowest quantile. These variables were the distance to the beach, distance to the highway/train, vulnerability to heat impact, income equality, and welfare. Furthermore, Demonstrative Map 1 shows that the observations in the subdistricts in the highest category have in general a higher accessibility to public amenities and services compared to other subdistrict if they are positively correlated with housing prices. The values of these accessibility factors to public amenities and services PCA components are mostly higher than those of 80% of the other subdistricts.

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



In conclusion, I found by interaction with 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 more equal/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, distance to the highway/train, vulnerability to heat impact, income equality, and welfare. The distance to the beach and income equality are hard to change through policy interventions. Since the distance to the beach is strictly geographical and a higher income equality has a negative correlation with housing prices Besides, it is shown in the map that also the values of 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 that have a high weight in the calculated total price impact of sustainability.

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 I shared on GitHub.

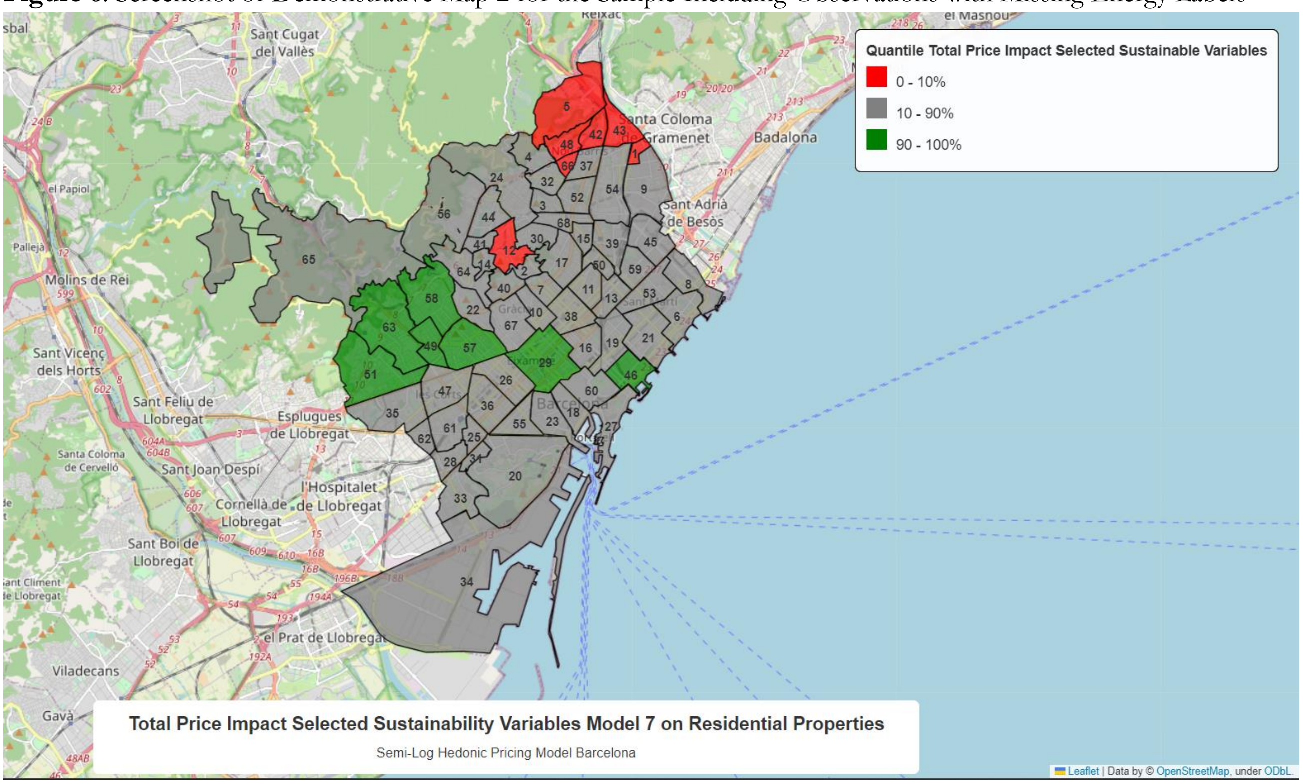
Demonstrative Map 2 clusters the observations on the 68 subdistricts of Barcelona as defined by Idealista for the sample including observations with missing energy labels. The map is colored using the total price impact of the selected sustainable variables. These are all the sustainable variables as included in pricing Model 7 excluding the distance to the beach (km), neighborhood density (10 ha), and income distribution PCA component. I included a screenshot of Demonstrative Map 2 in Figure 6 below. I included an index for the specific names of the subdistricts, displayed by the number of the colored subdistrict 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 Demonstrative Map 1. A small difference is that the subdistrict in Sants-Montjuic is no longer shown in the lowest 10% quantile for the total price impact of the selected sustainable variables. Instead, there is shown a subdistrict in Sant Andreu in the 10% quantile with the lowest total price impact of the selected sustainable variables. The other subdistricts with the lowest price impact of the selected sustainable variables are similar to Demonstrative Map 1. The subdistricts 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, Demonstrative Map 2 shows that these subdistricts have often low values for accessibility PCA components that have a positive correlation with the housing price and that the subdistricts had a high past natural population growth, which has a negative correlation with housing prices. The demonstrative map shows that the subdistrict in Gracia has the same reasons for the low total price impact of sustainability. Only, the past natural population growth has not been that high in the subdistrict, but instead, the population density which is also negatively related to housing prices is high in this subdistrict. For the subdistrict in the highest quantile of the total price impact of sustainability shows Demonstrative Map 2 that the high total price impact of the selected sustainable variables in the highest quantile is mainly caused by high welfare measured by the PCA component of income and unemployment. Every subdistrict in the highest quantile of the 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 with a high price impact of sustainability also have high/positive values for other selected sustainable variables. For instance, for the subdistricts is the vulnerability to heat impact lower than 85% of the other subdistricts in the map. Moreover, the houses are in most of the subdistricts located near 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 PCA components that have a positive relationship with the housing prices: park & garden, bar & restaurant, university, pharmacy, big shopping place, and museum, library & POI. culture. The values for these PCA components are for most of the subdistricts at least higher than 75% of the other subdistricts.

To sum up, I found that 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 opposite values for these sustainable factors. Furthermore, the inhabitants of these subdistricts often have a high average welfare. The finding that areas with a high/low price impact of sustainability are located close to each other implies that a more equal distribution of the price impact of sustainability on housing prices could be stimulated 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. The local clustering tendencies of the price impact of sustainability in housing prices in Demonstrative Maps 1 and 2 provide no evidence to reject hypothesis 6 (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

In the robustness section, I discuss 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 in Section 5.1. Furthermore, Demonstrative Maps 3 and 4 are discussed in Section 5.2. These demonstrative maps cluster the observations based on the location by the support vector machine

(SVM) instead of the subdistrict. Besides that, the same results are visualized in Demonstrative Maps 3 and 4 as in respectively Demonstrative Maps 1 and 2.

Section 5.1: Robustness Pricing Models

I will focus in this subsection on the differences for the sample that excludes observations with missing energy labels compared to the sample that includes observations with missing energy labels. I included the results for pricing Model 1 to Model 4 in Table 17a in the appendix and the results for pricing Model 5 to Model 7 in Table 17b of the appendix. I included the VIF test statistics for the variables in the pricing models 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 are correlated with the probability of the presence of an energy label in the housing advertisement. Firstly, the new housing development variable is higher (0.168) and has become significant at the 1% level. Secondly, the coefficient of the heating variable has halved. Thirdly, the strength of the dummy energy consumption label variables is lower in the sample that excludes observations with missing energy label data. This is in line with the findings in the findings in 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 Model 4, Model 5, and Model 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. I also found no differences in the results for Model 3 including the environmental dimension of sustainability. For Model 4, including the social dimension of sustainability, we can only observe small differences in comparison to the sample that includes observations with missing energy labels. Since there is no more evidence provided for a significant relationship between the policy PCA component and the secondary & lower education PCA component and the housing prices. However, these variables were also insignificant when the other sustainability dimensions were included in Model 7 for the sample that includes the observations with missing energy labels. Besides that, we can observe for

Model 4 small differences 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. I argue that this could be caused by the exclusion of observations in more cultural and historical neighborhoods in the sample excluding observations with missing energy labels. In these places, there could be a higher probability that a building has no energy label due to the age of the property and traditional building features. However, 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 showing a high degree of robustness of the results. The strength of the coefficients is that 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, I found in the results for the sustainable variables only small differences for the sample that excludes observations with missing energy labels compared to the sample that includes observations with missing energy labels. Thereby, I found evidence that the results show a high degree of robustness and found further support for the conclusion to not reject hypotheses 1 to 5 (H1 – H5). I observed only notable differences in the housing-specific variables that correlate with the presence/absence of an energy label in the housing advertisement.

Section 5.2: Robustness Demonstrative Maps

In Section 5.2 I will discuss the results for Demonstrative Maps 3 and 4, which are similar to respectively Demonstrative Maps 1 and 2. The only difference compared to 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). Both maps make a clear distinction between areas with a high and low price impact of sustainability for the selected sustainable variables, which is calculated by the sum of the values of the selected sustainable variables times their coefficients. I used the sample which includes observations with missing energy labels to visualize the results since I found in Section 5.1 that the results for the valuation methods are similar and robust for both samples. Therefore, I favored the visualization of the results of the sample with the highest number of observations. A higher number of observations provides a better representation of the housing market. I included a screenshot of Demonstrative Maps 3 and 4 in respectively Figures 7 and Figure 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 (SVM) makes the clusters more similar in size to each other by using the latitudes and longitudes as input. However, we can see that the screenshots of the visualization of the results by Demonstrative Maps 3 and 4 are almost identical to those of 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. We can only observe small differences in the exact subdistricts in which they are located. Furthermore, interaction with the demonstrative maps shows that the low/high total price impact of sustainability is caused by the same sustainable variables as for Demonstrative Map 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 that the areas with a high and low total price impact of sustainability in the housing prices are almost identical compared to Demonstrative Map 3. In addition to Demonstrative Map 3, Demonstrative Map 4 shows that the clusters with a high/low price impact of sustainability often have relatively high/low values for accessibility PCA components that are positively correlated with housing prices.

In conclusion, I found in the Demonstrative Maps 3 and 4 that visualized clustered observations by the support vector machine (SVM) the same areas with a high/low price impact of sustainability on housing prices compared to the demonstrative maps that clustered observations by the subdistricts. Furthermore, the reasons for this high price impact of sustainability on the housing price, which is calculated as the sum of the values of the selected sustainable variables times their coefficients in pricing Model 7, are similar. Thereby, I found further evidence to support for the conclusion to not reject the hypothesis 6 (H6).

Section 6: Conclusion and Discussion

In this research, I investigated the price impact of sustainability on housing prices in Barcelona. The price impact is assumed to be equal to the value of sustainable factors times their coefficients. I applied a data-driven approach to study the price impact of sustainability from five different dimensions: ecological, environmental, social, cultural, and economic-financial as proposed by Kauko (2019). These sustainable dimensions contain information about the quality and healthiness of the living environment around the properties and property-specific characteristics. The only difference I made compared to the proposed dimensions is the exclusion of the energy label in the sustainable variables, which is included in the housing-specific variables. The energy label was the only sustainable variable that was not location-bounded. My work elaborates earlier state-of-the-art work that only studied the effect of an increase in sustainability from one perspective or dimensions with sustainable factors such as the distance to the coast, perceived neighborhood security, and access to public services/amenities on housing (Buonanno et al., 2013; Chen & Marmolejo Duarte, 2018; Dell'Anna et al., 2019; Graells-Garrido et al., 2021; Marmolejo-Duarte & Chen, 2022). Furthermore, in addition, I visualize in my work the results for the total price impact of sustainability, defined as the value of selected sustainable factors times their coefficients, on housing prices for neighborhoods/subdistricts in clear and insightful demonstrative maps.

I estimated the price impact of the five sustainable dimensions on housing prices by semi-log hedonic pricing models. The housing prices are assumed to be equal to the asking prices shown in housing advertisements on Idealista. The asking price tends to be close to the transaction prices in stable real estate markets (McGreal et al., 2010). This requirement is fulfilled for the residential real estate market of Barcelona with a consistent increase in housing prices over the last years (Idealista, 2023b). I found in pricing models in which the five sustainable dimensions were included on an individual basis that higher values for each sustainable dimension increase the willingness to pay for housing. I stated the same conclusion for a pricing model that included every dimension of sustainability. Hence the first five hypotheses (H1 – H5), 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 my report of research. Additionally, I observed in the results of the model that included every dimension of sustainability often a lower strength for the relationship between the housing price and sustainable variables compared to the one-dimensional models. 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 results are likely to suffer from an omitted variable bias. Moreover, I found that sustainable variables with a low (in)significance can become (in)significant when controlling for other dimensions of sustainability.

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

I visualized the results in demonstrative maps to provide deeper insight into the distribution of the price impact of sustainability across Barcelona. In demonstrative maps, I defined a high/positive price impact of sustainability when the sum of the values of selected sustainable variables times their coefficients were high. I defined a low/negative price impact of sustainability as the opposite. The demonstrative maps cluster the observations in the sample based on the subdistrict in which the properties are located. I found during interaction with the demonstrative maps that the properties with a high/low total price impact of sustainability are clustered in specific areas close to each other. This provides no evidence to reject hypothesis 6 (H6), which states that the pricing of sustainability in Barcelona shows local clustering tendencies.

In more detail, Demonstrative Map 1, which included every sustainable dimension, shows that a high/low total price impact of sustainability for a cluster is caused by a wide variety of sustainable variables. The variables include 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. In Demonstrative Map 2, I excluded the sustainable variables that are strictly location-bounded or are observed to have a negative relationship between social fairness and housing prices when calculating the total price impact. These sustainable factors, which are the distance to the beach, neighborhood size, and income distribution, are not likely to be changed by policy intervention. In But also after the exclusion of these sustainable factors, I observed similar areas in which a high/low total price impact of sustainability was observed in Demonstrative Map 2 in comparison to Demonstrative Map 1. However, Demonstrative Map 2 better highlighted that the areas with a high/low price impact of sustainability have in general also low values for the access to public services/amenities sustainability factors that are positively correlated with housing prices. The clustering of the price impact of sustainability across Barcelona implies that policy intervention to address unfair pricing by distributing the price impact of sustainability more equally across the city 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. I verified the robustness of the conclusion for Demonstrative Maps 1 and 2 by clustering the properties based on the location with the support vector machine (SVM) in Demonstrative Maps 3 and 4. I shared the code to construct the maps on GitHub. In the shared code, I offer the opportunity for future work to visualize the results of the pricing models in different ways and retrieve additional insights.

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 includes in-depth information. I became more familiar with linking and combining data while maintaining interpretability. For instance by linking the geographical 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. For both, I had to learn HTML coding. Related to the results, I was expecting to find that the impact of sustainability on housing prices is locally clustered. But, I was not expecting to find that this is driven by overall bad scores for almost all sustainable factors. Concerning the research process, I learned the importance of sharing findings with people who are less familiar with the topic. This helped me to get insights from new viewpoints and better describe and interpret results.

My research has limitations and provides opportunities for future work. Firstly, I made two important assumptions in this research that could impact the results. I assumed that the asking price of the properties on Idealista is equal to the selling price since there is no actual selling price data available for my research. Besides, I removed the duplicates only by the house-id of the housing advertisements. However, it could be that the same property is listed multiple times by different agencies. I did not remove these duplicates from the sample. Secondly, I measure sustainability in the pricing models from a wide view. However, additional sustainable factors can be introduced that do not have a high correlation with the included variables. Thirdly, I visualized the results by the demonstrative maps which are only limited in terms of interpretation. The findings are not connected to the costs to improve the score of the sustainable factors. Fourthly, similar visualization maps are not reported for other cities in earlier research. Therefore I cannot verify 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. Fifthly, I apply 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. These machine learning models can offer advantages above linear models such as dealing better with big amounts of data, dealing better with bad hierarchization of data, dealing with measurement errors, and allowing for non-linear variables (Mora-Garcia et al., 2022). This resulted in various earlier research to find that machine learning models achieve a higher accuracy in predicting housing prices (Foryś, 2022; Kok et al., 2017; Mora-Garcia et al., 2022; Rico-Juan & Taltavull de La Paz, 2021). This potential of lower pricing error by machine learning models increases the fairness of real estate valuation models. (Vargas-Calderón & Camargo, 2022). The reason for not using a machine learning model as a pricing model in this research is the limited interpretation behind the price rationale of the results for most methods (Taltavull de La Paz et al., 2019). Most machine learning models that are used have not been developed to study price relations in-depth and suggest policy recommendations (Ho et al., 2021). Besides, research in this field, still mostly uses hedonic pricing models. Therefore comparability with earlier work would be harder if machine learning models were used.

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*. https://www.arcadis.com/en/knowledge-hub/perspectives/global/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.

Foryś, I. (2022). Machine learning in house price analysis: regression models versus neural networks. *Procedia Computer Science*, *207*, 435–445. https://doi.org/10.1016/j.procs.2022.09.078

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

Ho, W. K. O., Tang, B.-S., & Wong, S. W. (2021). Predicting property prices with machine learning algorithms. *Journal of Property Research*, *38*(1), 48–70. https://doi.org/10.1080/09599916.2020.1832558

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

Kok, N., Koponen, E.-L., & Martínez-Barbosa, C. A. (2017). Big Data in Real Estate? *From Manual Appraisal to Automated Valuation*. *The Journal of Portfolio Management*, *43*(6), 202–211. https://doi.org/10.3905/jpm.2017.43.6.202

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

Mora-Garcia, R.-T., Cespedes-Lopez, M.-F., & Perez-Sanchez, V. R. (2022). Housing Price Prediction Using Machine Learning Algorithms in COVID-19 Times. *Land*, *11*(11), 2100. https://doi.org/10.3390/land11112100

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

Phibbs, C. S., & Luft, H. S. (1995). Correlation of Travel Time on Roads versus Straight Line Distance. *Medical Care Research and Review*, *52*(4), 532–542. https://doi.org/10.1177/107755879505200406

pypi.org. (2023a). *undected-chromedriver 2.1.1*. Https://Pypi.Org/Project/Undetected-Chromedriver/2.1.1/.

pypi.org. (2023b, July 4). *beautifulsoup4 4.12.2*. Https://Pypi.Org/Project/Beautifulsoup4/.

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

Rico-Juan, J. R., & Taltavull de La Paz, P. (2021). Machine learning with explainability or spatial hedonics tools? An analysis of the asking prices in the housing market in Alicante, Spain. *Expert Systems with Applications*, *171*, 114590. https://doi.org/10.1016/j.eswa.2021.114590

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.

similarweb. (2023). *overview idealista.com*. Https://Www.Similarweb.Com/Website/Idealista.Com/#overview.

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.

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

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

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

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	mean	Variable	mean
House	0.02	Energy label consumption C	0.05
Mezzanine	0.05	Energy label consumption D	0.11
Ground Floor	0.10	Energy label consumption E	0.51
1st Floor	0.23	Energy label consumption F	0.09
2nd – 5th Floor	0.49	Energy label consumption G	0.17
6th Floor or higher	0.11	District Eixample	0.22
New housing development	0.02	District Ciutat Vella	0.15
Good condition	0.84	District Sant Martí	0.10
Needs renovation	0.14	District Sants-Montjuïc	0.11
Elevator	0.77	District Horta Guinardó	0.08
Terrace	0.32	District Gràcia	0.07
Heating	0.54	District Les Corts	0.04
Outdoor facilities	0.56	District Nou Barris	0.06
Parking space included	0.10	District Sarrià-Sant Gervasi	0.10
Energy label consumption A	0.03	District Sant Andreu	0.06
Energy label consumption B	0.03		

Table 5 includes the summary statistics for the housing-specific dummy variables for the sample excluding observations with missing energy labels. The sample consists of 10.104 observations.

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	Neighborhood 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	0.00	0.24	0.36	0.48	1.00	0.60	0.47
Distance to Highway/Train (km)	1.14	0.72	0.02	0.61	0.98	1.50	4.91	1.14	1.37
Distance to Beach (km)	3.28	1.40	0.01	2.13	3.43	4.61	5.00	-0.37	-1.03
Park & Garden PCA	0.40	0.39	0.00	0.01	0.50	0.51	1.00	0.36	-1.33
Viewpoint PCA	0.41	0.26	0.00	0.21	0.32	0.64	1.00	0.67	-0.79
Neighborhood size (10 ha)	13.66	12.91	2.30	8.08	11.1	14.1	142.37	6.02	50.01
Vulnerable to heat impact (1-5)	2.85	0.79	1.00	2.00	3.00	3.00	5.00	0.10	-1.02
Police PCA	0.44	0.28	0.00	0.29	0.32	0.55	1.00	0.32	-0.77
Bar & Restaurant PCA	0.50	0.35	0.00	0.19	0.34	0.99	1.00	0.26	-1.40
Secondary & Lower School PCA	0.46	0.20	0.00	0.32	0.46	0.62	1.00	0.00	-0.61
University PCA	0.35	0.38	0.00	0.05	0.24	0.63	1.00	0.76	-1.08
Pharmacy PCA	0.50	0.22	0.00	0.40	0.50	0.60	1.00	-0.02	-0.19
Hospital & Clinique PCA	0.37	0.28	0.00	0.21	0.22	0.60	1.00	0.66	-0.36
Big Shopping Place PCA	0.34	0.30	0.00	0.08	0.20	0.54	1.00	0.87	-0.56
Social Cohesion Score	0.21	0.18	0.00	0.09	0.15	0.30	1.73	1.80	4.81
Natural population growth ‰	-2.09	1.96	-8.92	-3.30	-2.50	-0.60	3.60	0.12	0.55
Net immigration rate ‰	27.27	22.42	-6.60	9.90	19.50	46.9	91.10	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	0.00	0.06	0.17	0.42	1.00	1.15	0.02
Religious Institution PCA	0.44	0.25	0.00	0.27	0.40	0.60	1.00	0.54	-0.45
Museum, Library & POI Cult. PCA	0.26	0.23	0.00	0.09	0.16	0.35	0.99	1.32	0.88
Income Distribution PCA	0.46	0.21	0.00	0.30	0.47	0.62	1.00	-0.32	-0.61
Income & Unemployment PCA	0.20	0.20	0.00	0.07	0.15	0.27	1.00	1.90	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
Save (True/False)	True	True

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
2^{nd} -5^{th} 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

Variable	Probit Model	VIF Model 1	VIF Model 2	VIF Model 3	VIF Model 4	VIF Model 5	VIF Model 6	VIF Model 7
Air Conditioning	12.90							
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.

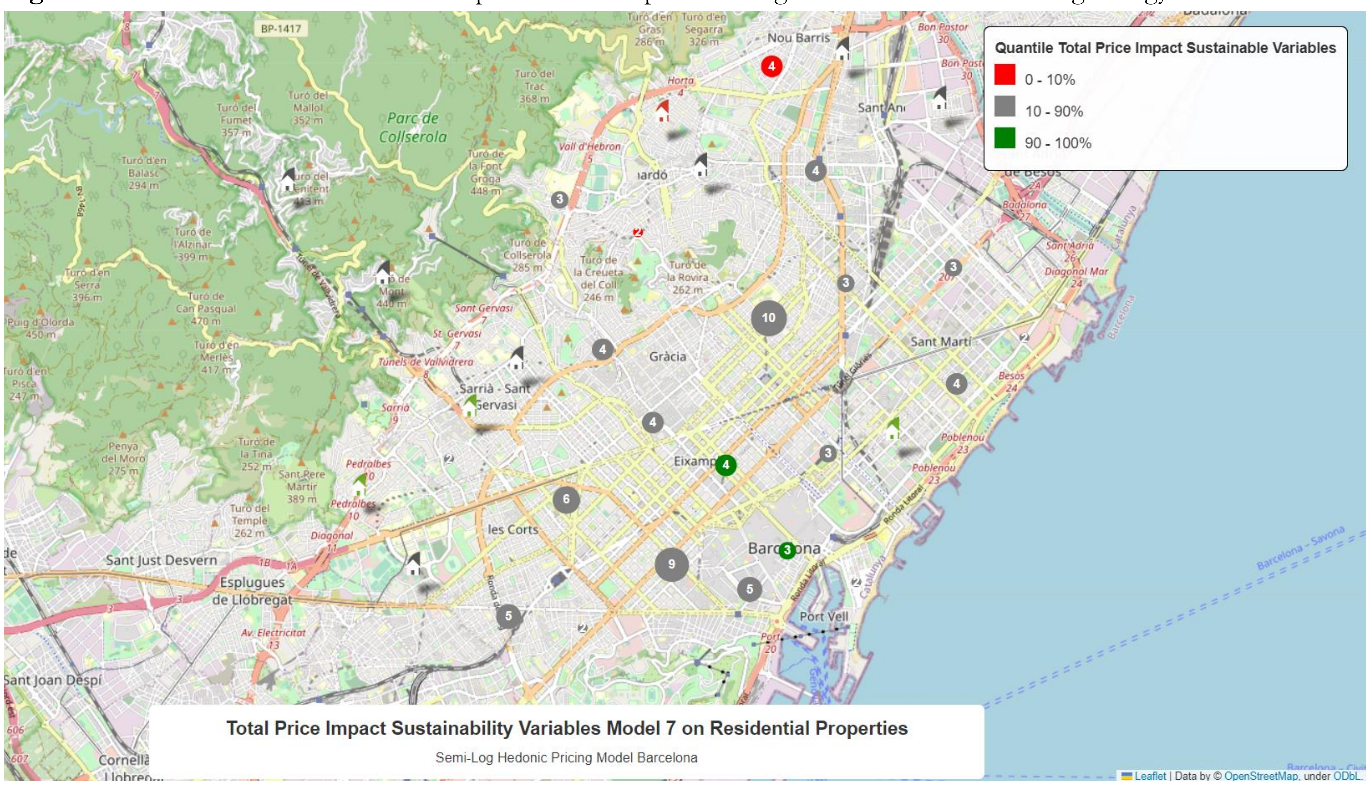
Table 18: VIF Test-Statistics of the Sample Excluding Observations with Missing Energy Label Data

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

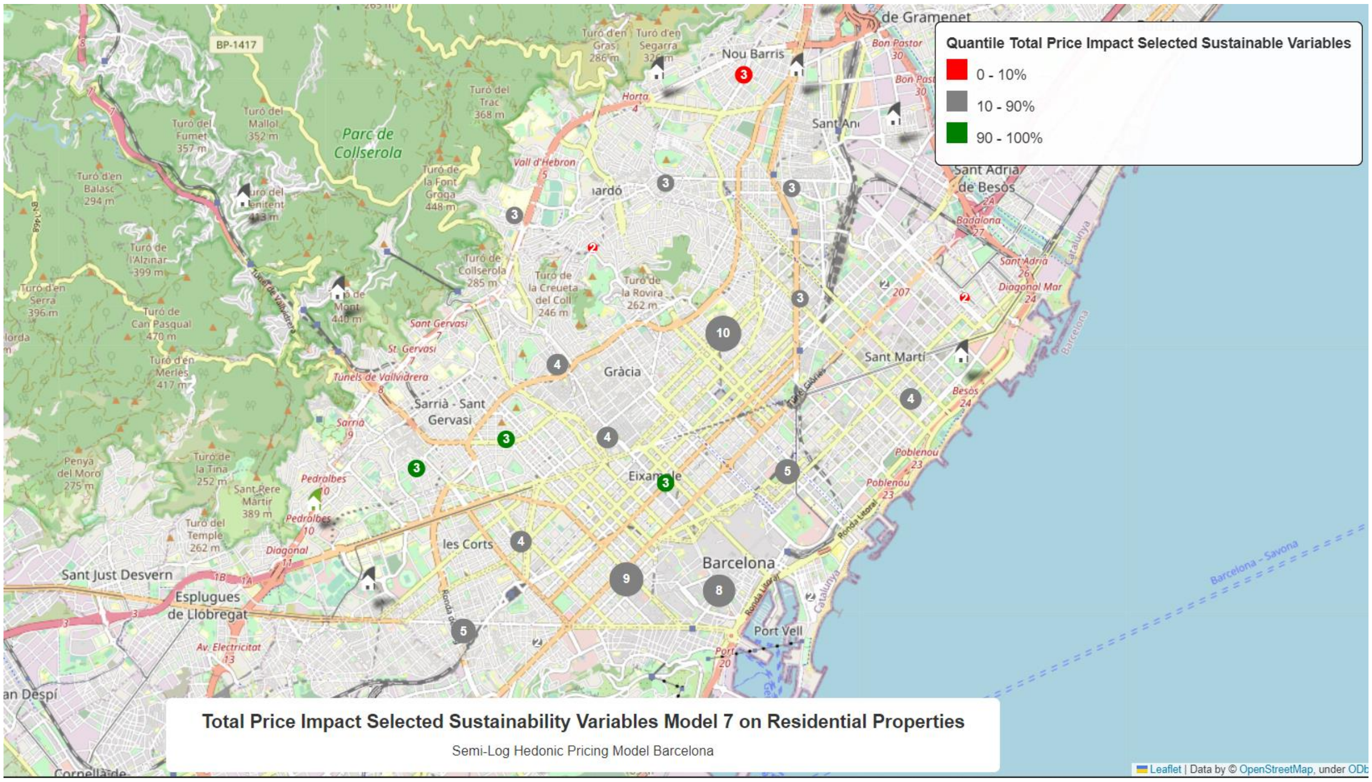
Table 18 includes the VIF test statistics for the variables included in the semi-log hedonic pricing models of the sample excluding observations with missing energy labels

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