Real Estate Market Dynamics in the Municipality of Oporto

FRANCISCO SOUSA MATOS [a], ANTÓNIO PEDRO DUARTE SILVA [a]

[a] Católica Porto Business School, Universidade Católica Portuguesa, Porto, Portugal.

To cite this article: Silva, António Pedro Duarte, Francisco Sousa Matos. 2024. Real Estate Market Dynamics in the Municipality of Oporto, *European Review of Business Economics* IV(1): 95-114.
DOI: https://doi.org/10.26619/ERBE-2024.4.1.5.

ABSTRACT

The housing configuration in a given area generally reflects similarities in terms of structural, location and neighborhood characteristics, indicating the formation of distinct housing sub-markets. This paper aims to identify the existence and evolution of housing sub-markets in the municipality of Oporto in 2019 and 2022. These sub-markets were identified with the help of the recent methodology of hierarchical cluster analysis with contiguity restrictions. Whereas traditional clustering techniques have long been used in market segmentation studies, those studies tend to incorporate location/neighborhood restrictions in an ad hoc form. Contiguity restricted cluster analysis addresses this issue directly. Results identified three well-defined and relatively stable sub-markets. Their delimitation, complemented by an analysis of the characteristics that define them, provides valuable information for homeowners, municipalities, lenders, and real estate investors and developers.

Keywords: Real Estate Markets; Market Segmentation; Spatially Restricted Cluster Analysis.
JEL Codes: R3

I. Introduction

IN RECENT DECADES, the availability of affordable housing in Portugal has been a primary concern for the Portuguese government and society. However, this market is far from homogeneous, showing substantial differences over time and between different areas. This work aims to identify segments in the real estate market in the municipality of Oporto and the evolution of the housing market in the years 2019 and 2022.

The housing market is characterized by being segmented and structured according to a complex pattern that takes into account various elements and does not merely follow a homogeneous process of spatial organization (Marques et al., 2012). In fact, residential housing has various and diverse characteristics. These include physical characteristics, neighborhood attributes and location factors (Usman et al., 2020). The presentation of multiple attributes means that the different social groups that participate in the market and have different preferences and economic capacities organize themselves into clusters both in territorial as well as in social and economic terms (Marques, 2012). Thus, there is spatial heterogeneity in the housing market and, as a result, there are several reasons why the analysis of the segmentation of this market is needed. The development and analysis of sub-markets make it possible to understand the particularities of each sub-market, improving the ability of investors and lenders to assess the risk associated with housing investments. At the same time, housing consumers themselves acquire information on how sub-market boundaries are defined (Goodman & Thibodeau, 2007). From a strategic perspective, the definition of sub-markets contributes to a better understanding of possible problems in specific sub-market areas. This recognition benefits fiscal assessment and community development, providing an effective source for planners and policymakers to examine dynamic changes in the housing system (Keskin & Watkins, 2017; Sairi et al., 2022).

The issue of housing is a current and emerging concern in Portugal, providing a sound reason to study the issue of defining sub-markets in the municipality of Oporto. The lack of similar studies in this part of the country makes this study particularly useful for many government officials and economic agents. In this study, we identified sub-markets by hierarchical cluster analysis with contiguity restrictions (Guénard & Legendre, 2022). Although this methodology is arguably the most appropriate for any clustering problems with spatial constraints, it is also relatively new, and to the best of our knowledge, it has not been applied to any other housing segmentation study.

Our results identified four clusters in 2019 and 2022, with three of them sharing most members and characteristics for both years. The three common sub-markets may be characterized in the following way: Sub-market 1 - Mostly small apartments located predominantly around the Asprela university area. Sub-market 2 - Historic properties in Oporto center with a low presence of high-rise apartments. Sub-market 3 - High-value properties in the prestigious area of Douro's Foz. Furthermore, in 2019, our analysis suggested a fourth cluster mainly comprised of old single-story houses in the suburbs of Oporto, while in 2022, there seemed to be a small independent sub-market in the area between Paranhos and Campanhã.

The remainder of this paper is organized as follows. The next section reviews the economic literature on housing market segmentation. Section 3 describes the data and methodology used in this study. Section 4 discusses the results of our analysis, and Section 5 presents its main conclusions, limitations, and avenues for further research.

II. Segmentation of Real Estate Markets

Over the last few years, several authors have studied the real estate market. Despite being a widely studied topic, complexity arises from the outset, as the very definition of a sub-market is not straightforward (Bourassa et al., 1999; Watkins, 2001). One of the oldest and most widely used definitions defines a housing sub-market as "a set of dwellings that are reasonably close substitutes for each other but are relatively weak substitutes for dwellings in other sub-markets" (Bourassa et al., 1999; Wu & Sharma, 2012). Indeed, in their seminal book "Housing Market Analysis: A Study of Theory and Methods", Rapkin and Winnick (1953) define a housing sub-market as "(...) the physical area within which all housing units are linked together in a chain of substitution", considering that "(...) each housing unit in a local housing market can be considered a substitute for any other unit" (Marques, 2012). The economic concept of substitution features prominently in the definitions mentioned by various authors. This concept raises difficult questions about how to identify close substitutes and about levels of aggregation (or disaggregation; Bourassa et al., 1999). More recently, Goodman & Thibodeau (2007) state that housing sub-markets are areas where the price per unit of housing quantity (defined by some index of housing characteristics) is constant. However, identifying geographical areas with constant prices per unit of housing remains a challenge. This is because housing is a heterogeneous good and the market value of a house is a function of the characteristics of the site, the structure, the neighborhood and the property location (Goodman & Thibodeau, 2007). Despite the wide range of definitions of housing sub-markets, there is a consensus that they are usually defined in terms of geographical areas or the physical characteristics of dwellings (Bourassa et al., 2003). However, agreement that the delimitation of sub-markets is dependent on spatial or physical characteristics does not dispel the lack of consensus that exists in the identification of sub-markets in practice (Leishman, 2001; Watkins, 2001; Xiao et al., 2016). Developments in the procedures and methodologies for identifying and delimiting sub-markets are therefore significant. According to Wu and Sharma (2012), classification methodologies can be divided into two broad categories: a priori classifications and data-based methodologies.

A. Real Estate Segmentation Based on A Priori Criteria

A priori definitions are based on conveniently available spatial divisions or on predefined criteria such as structural attributes of the dwelling or characteristics of the user group. One of the most common spatial divisions are real estate boundaries (Wu & Sharma, 2012). This approach was used by Bourassa et al. (2003) when they compared a set of sub-markets based on small geographical areas defined by real estate appraisers with a set of statistically executed sub-markets. Similarly, Keskin & Watkins (2017) explore the relative benefits of sub-market boundaries defined by agents, valuers and market analysts. An example is Palm's (1978) delineation of the San Francisco market based on districts where real estate agents exchange information about job offers (Watkins, 2001). Other studies have segmented the real estate market into sub-markets based on aggregated census blocks, postal codes (Goodman & Thibodeau, 2003), urban road network (Xiao et al., 2016), local government boundaries (Bourassa et al., 1999) and physical characteristics (Watkins, 2001). In the aforementioned spatial divisions, which are derived from existing subdivisions of space or pre-defined criteria, the evolution of socio-economic patterns and consumer preferences in housing is not reflected. As a result, a priori spatial classifications, which are often static in nature, fail to grasp the dynamics of housing sub-markets (Wu & Sharma, 2012). In contrast, a priori sub-markets based on housing and demographic characteristics can capture the evolving nature of the housing market. In light of the above, several authors consider these types of variables in their studies. For instance, Palm (1978) considers housing market delimitations based not only on the jurisdiction of housing boards, but also on the racial/ethnic composition of the neighborhood. Other studies have segmented housing markets into sub-markets based on socio-economic characteristics (Farber, 1986), consumer group income (Schnare & Struyk, 1976), property type (Adair et al., 1996) and housing size (Goodman & Thibodeau, 2007). Some studies of sub-markets have explicitly recognized the joint importance of spatial and structural characteristics in defining sub-markets (Watkins, 2001). Leishman (2001) points out that housing markets can be subdivided both spatially and structurally, thus forming a set of interconnected sub-markets. Furthermore, several authors have proposed tiered classifications based on a combination of spatial boundaries, housing characteristics and socio-economic dimensions (Schnare & Struyk, 1976; Tu, 1997; Watkins, 2001).

B. Real Estate Segmentation Based on Statistical Criteria

In contrast to delineations based on a priori knowledge, an alternative approach is to let the data determine the structure of the sub-markets. Therefore, several alternative approaches have been developed that use statistical procedures instead of a priori judgments. Goodman and Thibodeau (1998) combined a statistical analysis, with a priori knowledge. These authors used a hierarchical model to delimit the areas where the variation in the quality of public schools explains the variation in the hedonic coefficient of the property size in the 18 elementary school zones of a suburban Dallas school district. The main idea behind this approach is that all dwellings in a spatially concentrated area share the amenities associated with the property's location. Consequently, the housing characteristics that determine a property's market value are embedded in a hierarchy - properties within neighbourhoods, neighbourhoods within school zones, school zones within municipalities, and so on. This approach was later applied in a study using 28,000 single-family homes in the wider Dallas area (Goodman & Thibodeau, 2003). The main criticism of this hierarchical method relates to the fact that, although it is more technical and grounded in empirical data than earlier approaches, it still relies on prior assumptions to determine the most appropriate administrative boundaries, such as the use of school boundaries (Keskin & Watkins, 2017).

Studies that use an approach based entirely on statistical procedures usually rely on two different steps. The first one summarises the original data using a dimensionality reduction methodology, such as Exploratory Factor Analysis (EFA) or Principal Component Analysis (PCA). These methodologies make it possible to replace the original variables with a smaller set of synthetic variables, known as Factors or Components, which represent the essential information contained in the original variables (Bryant & Yarnold, 1995). The second step builds the relevant sub-markets, using a cluster analysis technique. Thus, by starting the study with a large data set, it is possible to determine which characteristics are most distinctive among the dwellings and then group them according to these characteristics. Or, if the data is limited to small geographical areas, then this data can be used to determine how the areas should be grouped (Bourassa et al., 1999).

The work of Dale-Johnson (1982) is particularly relevant in this area. This author uses factor analysis on 13 variables and extracts five factors that are used to define ten sub-markets. The definition of sub-markets is based on the five factors extracted, where two sub-markets are defined for each one. One of them contains the transactions that are most similar to the factor, while the other segment contains the transactions that are most different from the factor.

Bourassa et al. (1999) argue that this method of extracting the factors to define the sub-markets is problematic since it leaves out the information contained in the other factors. He suggests that a preferable approach is to use cluster analysis for all significant factors. This author applied PCA and cluster analysis to form housing sub-markets in Sydney and Melbourne. Using PCA, the author identified the important characteristics of local government districts. Subsequently, cluster analysis was used on the principal components to determine the most appropriate district groupings.

Maclennan and Tu (1996) constructed a matrix of sub-markets covering four urban sectors, as well as five categories of product groups. These groupings were defined on the basis of the PCA of the characteristics of the dwellings and the surrounding areas, followed by a cluster analysis.

The literature on the delimitation of sub-markets in Portugal is limited and, as a result, does not reflect the importance of this topic for understanding the dynamics of the housing market and its implications for possible urban planning policies and investment strategies in the housing sector. One notable exception is the work of José Lourenço Marques, regarding the definition of housing sub-markets in the Aveiro-Ílhavo urban area (Marques, 2012; Marques et al., 2012). This study is based on property data concerning the 2000-2010 period and, in addition to considering price, physical attributes of housing and location and neighbourhood characteristics to characterise sub-markets were used. The methodology followed two main approaches: an inductive approach and an analytical approach. In the inductive approach, Marques defined and delimited sub-markets using a priori classification method that relies on administrative boundaries, urban structure, demographic and historical characteristics and urban evolution. The analytical approach was based on spatial clustering analysis applied to five different criteria: the first criterion used the price of housing per square metre (in logarithm); the second one is based on the physical and locational characteristics of housing (application of PCA); the third and fourth consider the implicit prices of housing resulting from the application of a hedonic model to explain the value of a property. Finally, the fifth criterion aggregates and combines some of the approaches described above, i.e. the segmentation of the housing market resulting from three main dimensions: the characteristics of the dwelling; its importance in the valuation of the property (hedonic coefficients of the regression model) and the price per square metre of a dwelling.

III. Data and Methods

A. Database

In order to build a reliable and useful database for the application of cluster analysis, a database drawn up by the National Statistics Institute of Portugal from the Municipal Property Tax (IMI) was used. The data covers 2019 and 2022 and was limited to the municipality of Oporto. Despite being the main public database available for research, it naturally has some limitations. In particular, it does not contain all relevant variables to fully characterize neighborhood characteristics in social, economic or demographic terms. Furthermore, it does not include data on the price of properties traded, limiting itself to the variable describing the price of the valuation for tax purposes. Nevertheless, this study will show that it still contains sufficient information to describe the segment and recent evolution of the Oporto real estate market.

After a careful analysis of the database, the variables in Table 1 were selected. Most of these variables describe physical/structural characteristics of the property, with the exception of the “location coefficient” variable, which describes location characteristics, and the “quality- comfort coefficient” variable, which includes elements that describe both physical and location characteristics. The set of variables is made up of eight numerical variables - number of floors in the fraction, number of floors in the property, total land area, building implantation area, age, valuation value, location coefficient and quality and comfort coefficient - and three categorical variables - type of building, type of owner and typology (Table 1).

Table 1: Description of the variables used.

	Variable	Description/Definition
Physical characteristics	Building type	This variable has a numerical code associated with each type of building: 4 - Building in total ownership without floors; 5 - Horizontally-owned building; 6 - Fully-owned building with floors.
	Type of Owner	This variable has a numerical code associated with each type of owner: 1 - Sole Owner; 2 - Co-owner. Other - Usufructuary or Superficiary.
	Number of fraction floors	Number of floors relating to the property fraction
	Number of floors	Number of floors of the property, excluding real estate fractions
	Total land area	Total land area in squared meters
	Building footprint	Building footprint in squared meters
	Type	Typology (T1, T2, T3, T4) or the number of rooms (1, 2, 3). The number of rooms includes the number of bedrooms and living rooms; it does not include bathrooms, kitchens or storage rooms.
	Age	Age of the property in years.
	Valuation value	This is the Tax Asset Value (VPT), i.e. the stipulated value of a particular property for tax purposes (in euros).
Location characteristics	Location coefficient	According to article 42 of the Municipal Property Tax (CIMI - *Códigos do Imposto Municipal sobre Imóveis*), and the wording of Law no. 64- B/2011, of December 30, the location coefficient (Cl) varies between 0.4 and 3.5, and may, in situations of dispersed housing in rural areas, be reduced to 0.35. This aims to quantify the location of the building for the purposes of IMI assessment, according to the verification of a set of characteristics
Mixed Characteristics	Quality and comfort coefficient	As set out in article 43 of the CIMI, the quality and comfort coefficient (Cq) is applied to the base value of the building and can be increased by up to 1.7 and reduced by up to 0.5, and is obtained by adding the major coefficients to the unit and subtracting the minor coefficients from the unit.

Table 2 summarizes the main descriptive statistics of the numerical variables for each of the years under analysis. A preliminary analysis of the raw data shows that compared to 2019, the 2022 data show a general increase in the averages of several variables, with the exception of the age of the properties. In fact, the average age of the properties in

2022 is considerably lower than in 2019, which may indicate an increase in the construction of new properties or a trend towards newer properties on the market.

Table 2: Descriptive statistics of the numerical variables.

***N* = 4,819**			**2019**
Variable	**Minimum**	**Maximum**	**Mean**
Number of fraction floors	0,0000	5,0000	0,9647
Number of floors	1,0000	23,000	4,6210
Total land area m²	13,000	19639,9	961,30
Building footprint	5,5000	19639,9	613,80
Age	0,0000	485,000	46,050
Evaluation value	1660,0	1225990	70678
Location coefficient	0,0000	3,0000	1,5620
Quality and comfort coefficient	0,0000	1,6500	1,0440
***N* = 6,012**			**2022**
Variable	**Minimum**	**Maximum**	**Mean**
Number of fraction floors	0,0000	4,0000	1,048
Number of floors	1,0000	22,000	5,275
Total land area m²	8,5000	19639,9	1846,1
Building footprint	2,2750	19138,75	950,78
Age	0,0000	142,000	19,16
Evaluation value	3590,0	1933690	88573,0
Location coefficient	0,0000	3,0000	1,6210
Quality and comfort coefficient	0,0000	1,6000	1,0850

Table 3 shows the absolute frequencies of the categorical variables for both years. In both years, most of the properties were horizontal properties, i.e. buildings or a group of contiguous buildings owned by several people, each of whom has exclusive powers over a specific part (called an autonomous fraction, or what is commonly known as an apartment).

Table 3: Absolute frequencies of the categorical variables.

Variable	**2019** *N* = 4,812	**2022** *N* = 6,012
Building type		
4 - Building in total ownership without floors;	670	515
5 - Horizontally-owned building	2,849	4,584
6 - Wholly-owned building with floor	1300	913
Type of Owner		
1 - Sole Owner;	4,220	5,567
2 - Co-owner;	555	437
Other - Usufructuary or Superficiary	44	8
Type		
T0 - Property with zero bedrooms	1,127	1,514
T1 - One bedroom property	625	1,823
T2 – Two-bedroom property	604	905
T3 – Three-bedroom property	516	416
T4 - Property with four bedrooms	293	-
1 - Property with one room	582	-
Properties with a different number of rooms or divisions	1,072	1354

Similarly, the most common type of owner is the sole owner for both years. With regard to type, T0 and T1 are the most common types for the years under analysis, but the increase in T1 properties in 2022 is noteworthy. The number of properties with a different number of rooms or divisions is high because this category encompasses several categories that have a small number of observations, but which together contribute to the high value presented.

B. Methods

Cluster analysis with contiguity restrictions (Guénard & Legendre, 2022) was used to analyze and delimit the real estate sub-markets in the municipality of Oporto. As mentioned in Section 1.2.1, it is common to apply PCA before applying cluster analysis. PCA extracts factors that are then used in cluster analysis, focusing the analysis on orthogonal factors in the data, rather than multiple variables that may be explaining the same factor. PCA is also interesting in its own right because it identifies the basic dimensions that distinguish housing sub-markets.

However, PCA is not applicable in this study, since it is designed for continuous data, and our data set includes some categorical variables. Furthermore, as we only have a moderate number of relevant variables, the argument for using dimensionality reduction techniques is particularly compelling, and we choose to perform a cluster analysis directly on the whole set of original variables.

Cluster analysis techniques emerged in the fields of biological and ecological sciences and have been used extensively in various disciplines (Bourassa et al., 1999). Clustering represents a fundamental step in data extraction, allowing the identification of relevant groups and patterns in the underlying data. Clustering algorithms categorize data objects into distinct sets, known as clusters, based on their similarities or differences. Within a valid cluster, patterns tend to be more similar to each other than to patterns belonging to different clusters (Frades & Matthiesen, 2010). Therefore, the main objective of cluster analysis is to increase intra-group similarity and inter-group dissimilarity. As it is a technique based on data and, in turn, the variables included in the analysis, before applying this technique it is essential to define which variables are important for characterizing and delimiting the sub-markets (Marques, 2012). As mentioned above, some sub-market studies have explicitly recognized the joint importance of spatial and structural characteristics in defining sub-markets (see section 1.2.1). The study by Watkins (2001) reinforces this idea by stating that a hybrid model provides a better empirical approach to delimiting sub-markets. Therefore, since the database used in this study has variables that describe both structural and spatial characteristics, an approach that incorporates both groups of characteristics was used. This approach was developed by applying cluster analysis with contiguity restrictions.

The distance measure used to assess the similarity or dissimilarity between properties crucially affects the solution of clusters and depends on the nature of the variables considered.

Euclidean distance is the most commonly used metric when the variables are continuous (Frades & Matthiesen, 2010; Marques, 2012; van de Velden et al., 2019). In this study, both categorical and numerical variables were used, and instead of recoding the variables, a dissimilarity measure was adopted and applied directly to the mixed data. The Gower similarity coefficient is one of the most popular proximity measures for mixed data types (Hennig et al., 2015; van de Velden et al., 2019). Therefore, this measure of dissimilarity was used, as proposed by van de Velden et al. (2019).

Cluster analysis can be carried out using various methods, which can be divided into hierarchical and non-hierarchical clustering methods (Sairi et al., 2022). The non-hierarchical method breaks down the data set into a set of disjoint clusters. The simplest and most commonly used partitioning algorithm is k-means, in which the parameter k, which represents the number of clusters, must be specified (Frades & Matthiesen, 2010).

One of the main differences between the two methods is that the latter requires prior knowledge of the number of clusters, whereas the former does not require any prior knowledge. For housing sub-market applications, Goodman & Thibodeau (1998) suggest that hierarchical models provide the most useful framework for delimiting the market boundaries. The agglomerative hierarchical clustering steps can be summarized as follows:

1. Compute the proximity matrix using a particular distance metric.
2. Initially define n (one for each single element of the data) clusters, each one containing only one unique element.
3. Aggregate the clusters based on the chosen distance metric.
4. Update the distance matrix.
5. Repeat steps 3 and 4 until only one cluster remains.

The main methodological choice in this methodology concerns the computation of distance between clusters and the resulting update of the dissimilarity matrix. There are several existing approaches regarding these issues. The most common methods are: Single Linkage, Complete Linkage, Average Linkage and Ward's method (Hennig et al., 2015). We adopted Ward's method, which is based on the minimization of the sum of the squares of the distances of any two (hypothetical) clusters that could be formed at each step. As a result, it is a method that tends to produce hyper-spherical clusters and to contain approximately the same number of objects if the observations are evenly distributed throughout space (Legendre & Legendre, 2012). In addition, it is the method widely adopted in the various studies that delimit the housing sub-market.

In this work, we adopted the hierarchical clustering method with contiguity restrictions (Guénard & Legendre, 2022). Traditionally, studies in the area of sub-market delimitation have tended to neglect the role of spatial contiguity since they only use distance variables, such as the distance to the central business district (CBD), to capture the spatial organization of housing sub-markets. One approach that takes spatial contiguity into account is to treat spatial data as variables in the clustering process, assigning an appropriate weight to each variable (Wu et al., 2020). The hierarchical clustering method with contiguity restrictions adopted in this study innovates the way in which spatial contiguity is dealt with in the process of delimiting sub-markets. This is a relatively recent method, and we are not aware of any other application of this technique in the field of housing market segmentation.

The clustering method with spatial contiguity restrictions differs from unrestricted methods in that it does not go through all the pairs of points or clusters to find the pair with the lowest dissimilarity in the original or updated dissimilarity matrix. This method only needs to consider the dissimilarities corresponding to contiguous pairs. As a result, it is generally a faster method and produces solutions that are less variable than their unrestricted counterparts (Guénard & Legendre, 2022). In practice, this model follows the steps described above for hierarchical agglomerative clustering methods, with the addition of defining the contiguity constraint. To apply the restriction, it is first necessary to define the meaning of contiguous in the context of the data. In this case, two elements are considered to be contiguous if they are close enough to each other. For this purpose, after trying several alternative thresholds, a list of neighbors was defined that connects pairs of points that are 500 meters apart or less. Naturally, during the aggregation process (step 3), only clusters that are contiguous and defined by the list of neighbors can be aggregated. The result is well-defined clusters taking into account the spatial contiguity constraint.

IV. Results

As a result of implementing the hierarchical clustering model with contiguity restrictions, Figures 1 and 2 show the dendrogram for 2019 and 2022, respectively.

As we move up the dendrogram, similar observations are grouped into branches, which progressively come together at higher points in the tree. The vertical distance at which the aggregations occur reflects the degree of difference between the observations: greater aggregation heights denote less similarity between the joined observations. Thus, a relevant horizontal cut-off level of the dendrogram is characterized by a relatively significant difference between the heights of two successive nodes. Taking this methodology into account, the analysis of four clusters was considered for both years (see Figures 1 and 2).

Figure 1
Dendrogram representing hierarchical clustering with contiguity restrictions for 2019.

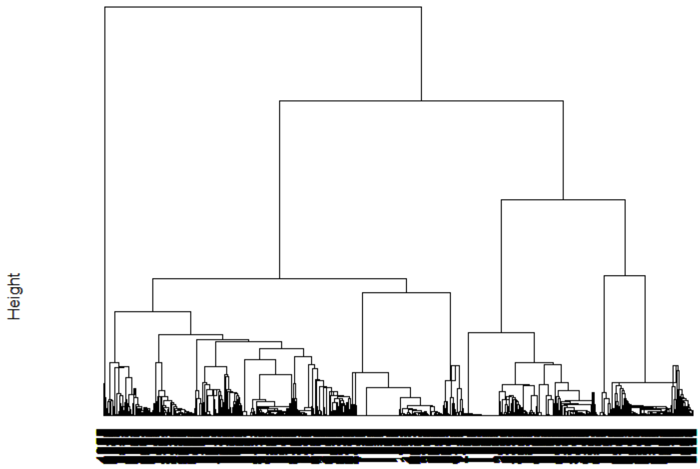
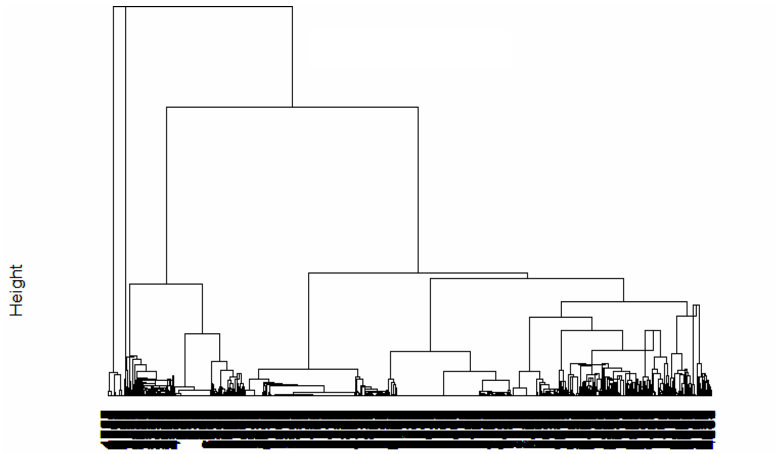


Figure 2
Dendrogram representing hierarchical clustering with contiguity restrictions for 2022



It should be noted that the choice of three clusters was also justifiable for 2019, but only four clusters were explored in order to make the comparison between the years more direct and homogeneous. Tables 4 and 5 provide, for each of the four clusters, the descriptive statistics of the numerical variables and the absolute frequencies of the categorical variables for 2019, respectively.

Table 4: Descriptive statistics of the numerical variables of each cluster for 2019.

Cluster	Variable	Minimum	Maximum	Mean
	Number of fraction floors	0,0000	3,0000	0,4074
	Number of floors	1,0000	4,000	2,0370
	Total land area m²	25,000	907,00	198,40
	Building footprint	23,70	183,00	75,190
1	Age	19,00	100,00	72,810
(*N* = 27)	Evaluation value	6790	114160	27476
	Location coefficient	1,1000	1,1000	1,1000
	Quality and comfort coefficient	1,0000	1,1000	1,0370
	Longitude	-8,574	-8,561	-8,568
	Latitude	41,15	41,15	41,15
	Number of fraction floors	0,0000	4,0000	1,1270
	Number of floors	1,0000	23,000	5,6730
	Total land area m²	19,700	19639,9	1227,0
	Building footprint	17,800	19639,9	779,5
2	Age	0,0000	485,000	32,96
(*N* =	Evaluation value	5530	880160	74363
2917)	Location coefficient	0,0000	3,0000	1.5600
	Quality and comfort coefficient	0,0000	1,6500	1,0800
	Longitude	-8,685	-8,565	-8,617
	Latitude	41,14	41,18	41,15

	Number of fraction floors	0,0000	5,0000	1,0920
	Number of floors	1,0000	8,000	2,9950
	Total land area m²	35,170	2068,7	468,7
	Building footprint	32,000	1522,7	338,60
3	Age	0,0000	130,000	71,87
(*N* =	Evaluation value	3690,0	541140	33237
1099)	Location coefficient	1,300	2,0000	1,5740
	Quality and comfort coefficient	0,7300	1,2000	0,9369
	Longitude	-8,638	-8,588	-8,612
	Latitude	41,14	41,17	41,16
	Number of fraction floors	0,0000	4,0000	0,1946
	Number of floors	1,0000	11,000	3,0580
	Total land area m²	13,000	8357,0	687,00
	Building footprint	5,5000	2532,00	399,49
4	Age	0,0000	213,000	57,720
(*N* = 776)	Evaluation value	1660,0	1225990	111357
	Location coefficient	0,0000	2,8000	1,5690
	Quality and comfort coefficient	0,0000	1,4600	1,0630
	Longitude	-8,685	-8,571	-8,619
	Latitude	41,14	41,19	41,16

Table 5: Absolute frequencies of the categorical variables in each cluster for 2019.

Variable	C1	C2	C3	C4
Building type				
4 - Building in total ownership without floors	18	17	1	634
5 - Horizontally-owned building	7	2734	1	107
6 - Wholly-owned building with floor	2	166	1097	35
Type of owner				
1 - Sole Owner;	17	2490	1054	659
2 - Co-owner;	10	399	41	105
Other - Usufructuary or Superficiary	0	28	4	12
Type				
T0 - Property with zero bedrooms	-	1022	90	15
T1 - One bedroom property	2	480	83	60
T2 – Two-bedroom property	6	417	108	73
T3 – Three-bedroom property	1	385	71	59
T4 - Property with four bedrooms	1	195	63	34
1 – Property with one room	4	153	384	52
Properties with a different number of rooms or divisions	13	265	300	483

Cluster 1 is a sub-market that stands out from the others because it only has 27 properties. This small sub-market is mostly made up of freehold buildings with no floors and is where the average age of the properties is highest. The appraisal values of the properties belonging to this sub-market are relatively low, indicating the existence of less valued properties, probably located in less central areas. Both the location coefficient and the quality and comfort coefficient show stable values for the sub-market under analysis.

The second sub-market is the one with the largest number of properties, and the only one where the horizontal property regime stands out from the other types of buildings, indicating an urban environment with apartments and condominiums. With a preliminary analysis of this fact, it could be inferred that this sub-market is located in regions outside the historic center and may represent areas of high population density and recent development. This inference can also be corroborated by the typology, as this cluster features a large number of studio properties. It is only natural that this type is more prevalent in urban areas where demand for compact spaces is high, particularly among students and single professionals. This sub-market is also the one with the lowest average age of buildings, which together with the high variation in appraisal values, may reflect a mixture of new and old properties, with the possibility of recent development or renovation in the area.

Cluster 3 stands out for two reasons: the first is that of the 1,099 properties that make it up, 1,097 are freehold buildings with floors; the second is that the average age of the homes in this sub-market is high. Both facts suggest that this sub-market may be located in an area characterized essentially by older buildings and belonging to a single owner.

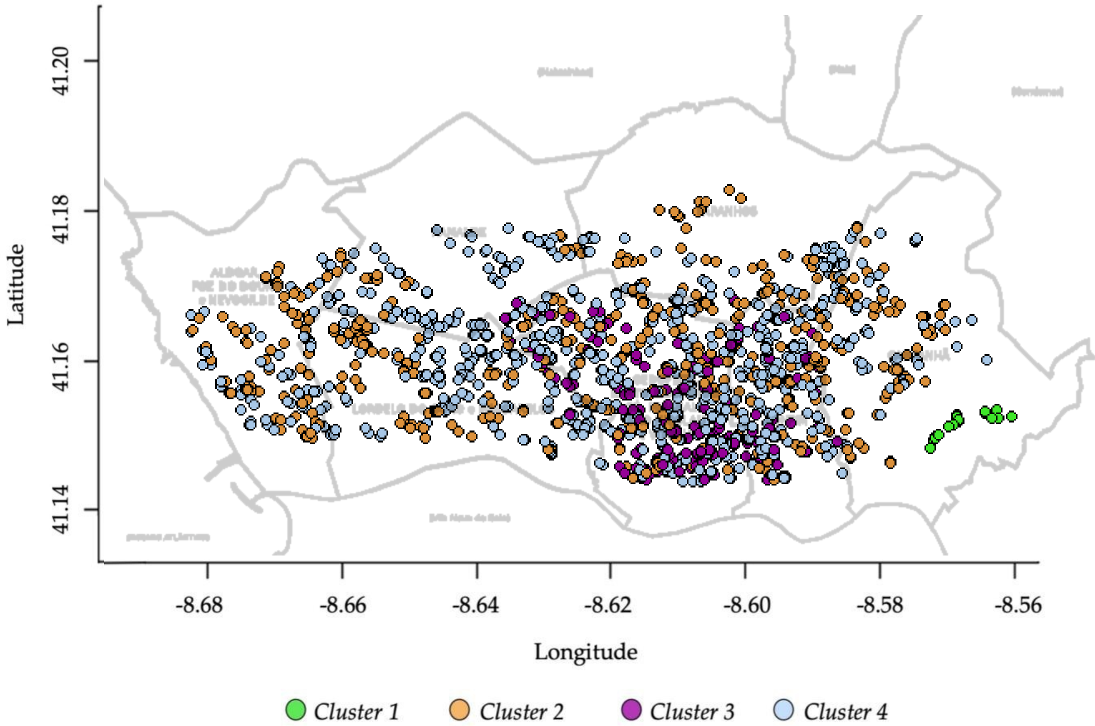
In sub-market 4, approximately 82% of the properties are freehold buildings and 62% are properties with more than four bedrooms or four rooms. This can be supported by the total area of the land and the building's footprint, as these variables have considerably high values. The average valuation of properties in this sub-market is the highest and, combined with the fact that this sub-market has a high average value for the location coefficient and, at the same time, a maximum greater than 2, this may reflect a luxury sub-market or areas of high real estate value.

The geographical representation of the clusters for 2019 is shown in Figure 3. It is clear from Figure 3 that there are two sub-markets that stand out due to their more precise definition: sub-market 1 and sub-market 3.

Sub-market 1 is the most isolated and all the properties that are part of it are located in the parish of Campanhã, outside the urban center of Porto. The average coordinates of this sub-market indicate that it is located around the neighborhoods of Lagarteiro, Granja and Azevedo. In these areas, single-storey houses are prevalent, which confirms the previous description of this cluster, where it was concluded that there is a predominance of buildings with total ownership with no older floors.

With regard to sub-market 3, the variation in coordinates is relatively small, which suggests a geographical concentration of properties. According to the map, it is possible to identify that this sub-market is concentrated in a central area with properties where the parishes of Cedofeita, Santo Ildefonso, Sé, Miragaia, São Nicolau and Vitória intersect and in the parish of Bonfim, confirming the previous description of this sub-market. We can therefore conclude that this sub-market is clearly delimited and is mainly located in the central area of the municipality of Oporto, where the main tourist attractions and most of Oporto's shops are located. In addition, it is in this area of Oporto that the historic city center is located and where there is a Municipal Master Plan that provides for the preservation of the identity of urban places beyond individual properties, taking into account the cadaster, urban fronts, forms of coverage, physical geography, elements of urban identification, vegetation cover, among others. Hence, it is natural that the buildings in this area are not very tall and, consequently, that there are no buildings under horizontal ownership. These facts are corroborated by the description of this cluster given above, since it predominately consists of buildings in full regime with floors. Another fact to bear in mind is the type of owner. Most of the buildings have a single owner (see Table 5) and this is in line, for example, with the majority of rental buildings, particularly for local accommodation, in this part of the municipality.

Figure 3
Geographical representation of the clusters for 2019



Both sub-market 2 and sub-market 4 are very geographically dispersed, with observations in all of Oporto's parishes. Cluster 4 is made up of larger and more luxurious homes.

The analysis of the clusters for 2022 takes into account the same characteristics analyzed for 2019. Tables 6 and 7, therefore, provide for each of the four clusters, the descriptive statistics of the numerical variables and the absolute frequencies of the categorical variables for 2022, respectively. It is clear that sub-market 2 is more prevalent in the parish of Paranhos when compared to sub-market 4. In the intersection of the parishes of Aldoar, Foz do Douro and Nevogilde, both sub-markets have observations, but sub-market 4 predominates in the coastal area, while sub-market 2 is more prevalent in the northeast of this parish.

Cluster 1 defines a sub-market made up entirely of buildings with horizontal properties and with the T0, T1 and T2 typologies prevailing almost equally.

Cluster 2 stands out from the rest because it only consists of seven properties, six of which are horizontal. This sub-market shows the most significant variation in appraisal value, indicating a marked heterogeneity that may correspond to an area of recent development or renovation. Dominated by freehold buildings with floors, cluster 3 is the one with the oldest properties among the other sub-markets. Its characterization suggests the prevalence of this sub-market in central areas. Sub-market 4 stands out for having the highest average location coefficient and the highest comfort quality coefficient, suggesting that it includes luxury or highly valued properties.

Table 6: Descriptive statistics of the numerical variables of each cluster for 2022.

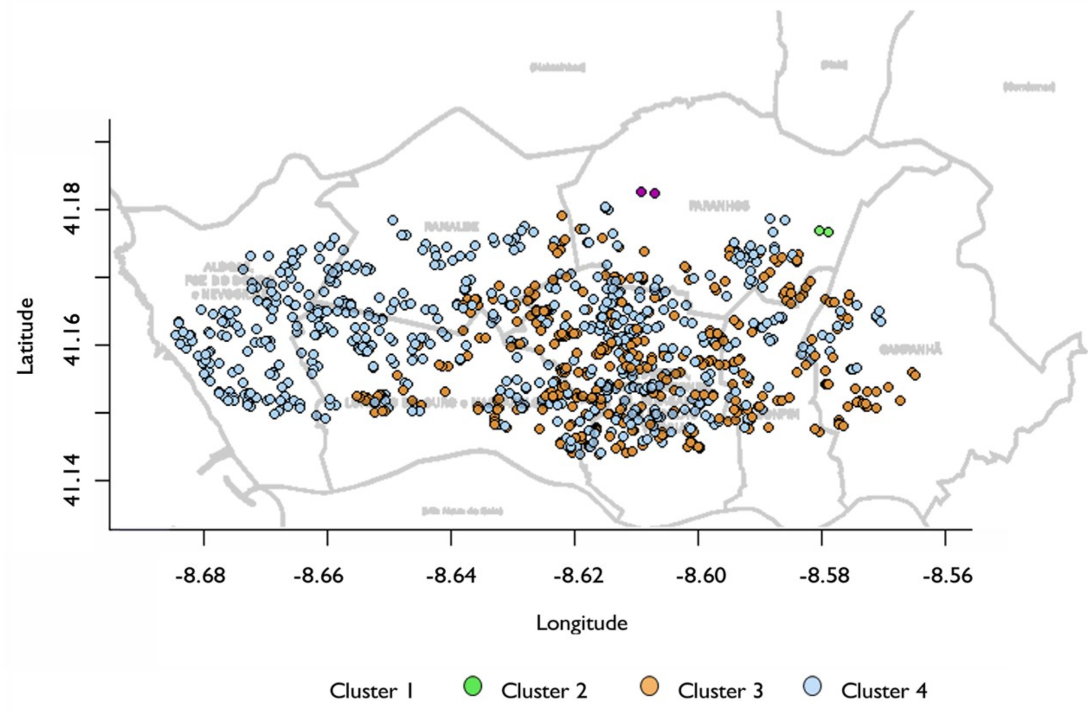
Cluster	Variable	Minimum	Maximum	Mean
	Number of fraction floors	1,0000	1,0000	1,000
	Number of floors	6,0000	8,000	7,8590
	Total land area m²	689,00	4682,0	4401,0
	Building footprint	223,40	2170,40	2033,0
1	Age	0,0000	0,0000	0,0000
(N=170)	Evaluation value	41350,0	131000	68001
	Location coefficient	1,5000	1,5000	1,5000
	Quality and comfort coefficient	1,0300	1,0300	1,0300
	Longitude	-8,607	-8,604	-8,606
	Latitude	41,18	41,18	41,18
	Number of fraction floors	0,0000	3,0000	2,5710
	Number of floors	2,0000	3,000	2,8570
	Total land area m²	314,000	746,200	375,70
	Building footprint	102,800	188,00	175,80
2	Age	1,0000	85,000	13,000
(N=7)	Evaluation value	83870	118420	112433
	Location coefficient	1,4000	1,4000	1,4000
	Quality and comfort coefficient	1,1100	1,1700	1,1190
	Longitude	-8,577	-8,575	-8,576
	Latitude	41,18	41,18	41,18
	Number of fraction floors	0,0000	3,0000	0,7241
	Number of floors	1,0000	10,000	2,942
	Total land area m²	8,5000	3440,0	539,40
3	Building footprint	8,5000	1806,00	281,71
(N=1189)	Age	0,0000	142,000	64,120
	Evaluation value	3590	1933690	55611
	Location coefficient	0,0000	2,1000	1,4930
	Quality and comfort coefficient	0,0000	1,3400	1,0110
	Longitude	-8,654	-8,561	-8,607
	Latitude	41,14	41,18	41,16
	Number of fraction floors	0,0000	4,0000	1,1310
	Number of floors	1,0000	22,000	5,7810
	Total land area m²	29,7000	19639,9	2089,3
4	Building footprint	2,2750	19138,75	1083,57
(N=4646)	Age	0,0000	120,000	8,361
	Evaluation value	5980	1115650	97725
	Location coefficient	0,0000	3,0000	1,6580
	Quality and comfort coefficient	0,0000	1,6000	1,1060
	Longitude	-8,684	-8,567	-8,621
	Latitude	41,14	41,18	41,16

Table 7: Descriptive statistics of the numerical variables of each cluster for 2022.

Variable	C1	C2	C3	C4
Building type				
4 - Building in total ownership without floors	0	1	365	149
5 - Horizontally-owned building	170	6	17	4391
6 - Wholly-owned building with floor	0	0	807	106
Type of owner				
1 - Sole Owner;	170	6	1119	4272
2 - Co-owner;	0	1	69	367
Other - Usufructuary or Superficiary	0	0	1	7
Type				
T0 - Property with zero bedrooms	58	-	139	1317
T1 - One bedroom property	57	6	87	1673
T2 – Two-bedroom property	53	-	56	796
T3 – Three-bedroom property	2	-	50	364
T4 - Property with four bedrooms	-	1	11	161
1 – Property with one room	-	-	80	76
Properties with a different number of rooms or divisions	-	-	766	259

As can be seen in Figure 4, the geographical dispersion of the clusters built for the year 2022 shows, as in 2019, two well-defined clusters, but in different locations in the municipality of Oporto. Cluster 1 is located in the parish of Paranhos and the average geographical coordinates point to the vicinity of the main faculties of the Asprela hub. Similar to 2019, this region, which is mainly sought after by young students, is well-defined. In 2022, the delimitation of this sub-market becomes even more marked, as this cluster is defined only by properties belonging to this area.

Figure 4
Geographical representation of the clusters for 2022



Cluster 2 is also well-defined and is located in the parish of Paranhos, just like cluster 1. However, this sub-market contains properties on the border between Paranhos and Campanhã in an area. As far as the type of building is concerned, it follows the trend of sub-market 1, with almost all of the buildings under horizontal ownership.

Despite being quite dispersed, submarkets 3 and 4 make up the central area of the city of Porto. As was the case in 2019, cluster 3 has a very small number of buildings under horizontal ownership, in line with the premise that in the center of Oporto there are few apartments with a high number of floors. This sub-market also has the highest average age, suggesting an area with a historic housing stock, possibly preserved by the Municipal Master Plan.

Finally, cluster four is the most geographically dispersed, although it is found almost exclusively in the union of the parishes of Aldoar, Foz do Douro and Nevogilde. As mentioned above, the average location coefficient of this cluster is the highest and it also has a maximum of three in this variable, suggesting that it is an area with a high real estate market value. In addition, this sub-market aggregates properties with extremely high appraisal values and at the same time has a high average appraisal value. Table 8 briefly describes each sub-market for both years.

Taking into account the description condensed in Table 8, there were no significant changes in the structure of the housing market in Oporto from 2019 to 2022, with the distribution and characteristics of the sub-markets remaining stable. This conclusion is to be expected since the composition of housing properties in a geographical area does not change drastically over a three-year period.

Table 8: Descriptive statistics of the numerical variables of each cluster for 2022.

2019	
Sub-market 1	Characterized by old single-storey houses on the outskirts of Oporto.
Sub-market 2	Prevalence of small apartments.
Sub-market 3	Concentration of properties in the historic center of Oporto, predominately low-rise, wholly-owned buildings.
Sub-market 4	It includes large, luxurious houses.
2022	
Sub-market 1	Entirely located in the Paranhos area, it focuses on studio and two-bedroom apartments surrounding the Asprela hub, which is a hot spot for students.
Sub-market 2	Small and with a significant variation in value, there are several renovation projects in the area on the border between Paranhos and Campanhã.
Sub-market 3	Historic properties and wholly-owned buildings, with a low presence of high-rise apartments.
Sub-market 4	Characterized by high-value properties.

The largest difference when comparing the two years is found in sub-market 1 and sub-market 2 in 2019 and 2022, respectively. In fact, these sub-markets describe well-defined areas, but they differ from one year to the next. In other words, each cluster appears exclusively in a single year, showing a lack of continuity or consistency in the delineation of these areas over time. In contrast, the other sub-markets either delimit the same areas in both years or represent properties with similar characteristics. The first refers to the central area of the municipality, where there is a predominance of wholly-owned properties, which are older and possibly preserved buildings. The second is located outside the heart of the city but is characterized by dynamic urban growth. In the region that includes the main hub of the University of Oporto, horizontal properties prevail, and these are more likely to be in high demand from the student population or single professionals. The third segments a more luxurious and exclusive sector of the real estate market, characterized by spacious properties, many with more than four bedrooms or divisions, and a high location quotient.

V. Conclusions

This study delimits the housing market in the municipality of Oporto into different sub-markets. Data supplied by the National Statistics Institute of Portugal was used as a basis, more specifically, housing data used to calculate the IMI. The approach used was to adopt a hierarchical clustering method with contiguity restrictions which, as well as being innovative, allows for the creation of clusters containing contiguous properties - taking spatial issues into account. The definition of each cluster reveals a sub-market with specific characteristics that can be analyzed and its main features described.

The discussion of the results suggests the existence of four sub-markets in both 2019 and 2022. Of the four sub-markets, three are common to both years: the first, located in the central area of the municipality; the second, characterized by small horizontal properties in the outskirts of Porto, including the outlying area of Paranhos; and the third where high-value properties prevail. The remaining two sub-markets are characterized by different locations in the two years, which is not sufficient to affirm these sub-markets as relevant. By analyzing the dynamics in 2019 and 2022, we can grasp how the sub-markets evolved over time. In fact, this evolution is relatively mild, since the conclusions dictate that the most solid sub-markets are defined in both years.

This delineation of sub-markets serves a wide range of stakeholders, as mentioned earlier in the conclusion. The general public, policymakers, lenders and investors, researchers in the field and real estate developers are the main stakeholders interested in the conclusions of this study.

The database has some limitations, such as: i) there are no characteristics that can be analyzed in the neighborhood, whether in social, economic or demographic terms; ii) the number of variables relevant to the study is small; iii) it does not contain data on the prices of properties traded. Even with the limitations described above, the construction of sub-markets proved to be feasible with important results for the area. Nevertheless, the delimitation of sub-markets in the municipality of Oporto could be more robust. One suggestion would be to construct variables that describe the area in more detail. In other words, calculate the distance of properties from, for example, hospitals, pharmacies, schools, the city park, the CBD, and the metro, among others.

Finally, another suggestion would be to use a database that reflects the transaction price of properties to build a hedonic pricing approach that would complement and make the analysis even more solid and complete.

Conflicts of Interest: The authors declare no conflict of interest.

REFERENCES

Adair, A. S., Berry, J. N., & McGreal, W. S. (1996). Hedonic modelling, housing submarkets and residential valuation. *Journal of Property Research*, *13*(1), 67-83. https://doi.org/10.1080/095999196368899

Bourassa, S., Hamelink, F., Hoesli, M., & MacGregor, B. (1999). Defining housing sub-markets. *Journal of Housing Economics*, *8*(2), 160-183. https://doi.org/10.1006/jhec.1999.0246

Bourassa, S. C., Hoesli, M., & Peng, V. S. (2003). Do housing submarkets really matter? *Journal of Housing Economics*, *12*(1), 12-28. https://doi.org/https://doi.org/10.1016/S1051-1377(03)00003-2

Bryant, F. B., & Yarnold, P. R. (1995). Principal-components analysis and exploratory and confirmatory factor analysis. In *Reading and understanding multivariate statistics.* (pp. 99-136). American Psychological Association.

Dale-Johnson, D. (1982). An alternative approach to housing market segmentation using hedonic price data. *Journal of Urban Economics*, *11*(3), 311-332. https://doi.org/https://doi.org/10.1016/0094-1190(82)90078-X

Farber, S. (1986). Market Segmentation and the Effects on Group Homes for the Handicapped on Residential Property Values. *Urban Studies*, *23*(6), 519-525. https://doi.org/10.1080/00420988620080801

Frades, I., & Matthiesen, R. (2010). Overview on Techniques in Cluster Analysis. In R. Matthiesen (Ed.), *Bioinformatics Methods in Clinical Research* (pp. 81-107). Humana Press. https://doi.org/10.1007/978-1-60327-194-3_5

Goodman, A. C., & Thibodeau, T. G. (1998). Housing Market Segmentation. *Journal of Housing Economics*, *7*(2), 121-143. https://doi.org/https://doi.org/10.1006/jhec.1998.0229

Goodman, A. C., & Thibodeau, T. G. (2003). Housing market segmentation and hedonic prediction accuracy. *Journal of Housing Economics*, *12*(3), 181-201. https://doi.org/https://doi.org/10.1016/S1051-1377(03)00031-7

Goodman, A. C., & Thibodeau, T. G. (2007). The Spatial Proximity of Metropolitan Area Housing Submarkets. *Real Estate Economics*, *35*(2), 209-232. https://doi.org/https://doi.org/10.1111/j.1540-6229.2007.00188.x

Guénard, G., & Legendre, P. (2022). Hierarchical Clustering with Contiguity Constraint in R. *Journal of Statistical Software*, *103*(7), 1 - 26. https://doi.org/10.18637/jss.v103.i07

Hennig, C., Meila, M., Murtagh, F., & Rocci, R. (2015). *Handbook of cluster analysis*. CRC Press.

Keskin, B., & Watkins, C. (2017). Defining spatial housing submarkets: Exploring the case for expert delineated boundaries. *Urban Studies*, *54*(6), 1446-1462. https://doi.org/10.1177/0042098015620351

Legendre, P., & Legendre, L. (2012). *Numerical Ecology* (3 ed.). Elsevier.

Leishman, C. (2001). House building and product differentiation: An hedonic price approach. *Journal of Housing and the Built Environment*, *16*(2), 131-152. https://doi.org/10.1023/A:1017924503356

Maclennan, D., & Tu, Y. (1996). Economic perspectives on the structure of local housing systems. *Housing Studies*, *11*(3), 387-406. https://doi.org/10.1080/02673039608720864

Marques, J. J. L. (2012). *The notion of space in urban housing markets* Universidade de Aveiro (Portugal)]. Aveiro, Portugal.

https://www.proquest.com/openview/ce328f597250e5df6f52ef7cd651af78/1?cbl=2026366&pq-origsite=gscholar

Marques, J. L., Castro, E., Bhattacharjee, A., & Batista, P. (2012). *Spatial heterogeneity across housing sub-markets in an urban area of Portugal* 52nd Congress of the European Regional Science Association: "Regions in Motion - Breaking the Path", Bratislava, Slovakia. https://hdl.handle.net/10419/120764

Palm, R. (1978). Spatial Segmentation of the Urban Housing Market. *Economic Geography*, *54*(3), 210-221. https://doi.org/10.2307/142835

Sairi, N. A. M., Burhan, B., & Safian, E. E. M. (2022). Identifying housing submarkets in Johor Bahru and Kulai, Malaysia: a data-driven method. *Real Estate Management and Valuation*, *30*(2), 1-11. https://doi.org/10.2478/remav-2022-0009

Schnare, A. B., & Struyk, R. J. (1976). Segmentation in urban housing markets. *Journal of Urban Economics*, *3*(2), 146-166. https://doi.org/https://doi.org/10.1016/0094-1190(76)90050-4

Tu, Y. (1997). The Local Housing Sub-market Structure and Its Properties. *Urban Studies*, *34*(2), 337-353. https://doi.org/10.1080/0042098976203

Usman, H., Lizam, M., & Burhan, B. (2020). A review of property attributes influence in hedonic pricing model. Proceedings of the 2nd African International Conference on Industrial Engineering and Operations Management.

van de Velden, M., Iodice D'Enza, A., & Markos, A. (2019). Distance-based clustering of mixed data. *WIREs Computational Statistics*, *11*(3), e1456. https://doi.org/https://doi.org/10.1002/wics.1456

Watkins, C. A. (2001). The Definition and Identification of Housing Submarkets. *Environment and Planning A: Economy and Space*, *33*(12), 2235-2253. https://doi.org/10.1068/a34162

Wu, C., & Sharma, R. (2012). Housing submarket classification: The role of spatial contiguity. *Applied Geography*, *32*(2), 746-756. https://doi.org/https://doi.org/10.1016/j.apgeog.2011.08.011

Wu, Y., Wei, Y. D., & Li, H. (2020). Analyzing Spatial Heterogeneity of Housing Prices Using Large Datasets. *Applied Spatial Analysis and Policy*, *13*(1), 223-256. https://doi.org/10.1007/s12061-019-09301-x

Xiao, Y., Webster, C., & Orford, S. (2016). Can street segments indexed for accessibility form the basis for housing submarket delineation? *Housing Studies*, *31*(7), 829-851. https://doi.org/10.1080/02673037.2016.1150433