



UNIVERSIDADE CATÓLICA PORTUGUESA

Análise dinâmica da segmentação do mercado de habitação no município do Porto

Francisco Sousa Matos

Católica Porto Business School

2024



UNIVERSIDADE CATÓLICA PORTUGUESA

Análise dinâmica da segmentação do mercado de habitação no município do Porto

Trabalho Final na modalidade de Dissertação
apresentado à Universidade Católica Portuguesa
para obtenção do grau de mestre em Business Economics

por

Francisco Sousa Matos

sob orientação de
Prof. Dr. António Pedro de Pinho de Brito Duarte Silva

Católica Porto Business School
Abril 2024

Agradecimentos

A realização de um trabalho de final de mestrado é desafiante e exigente e, por isso, as pessoas que me rodearam durante a sua realização foram cruciais para concluir esta etapa.

Começo por agradecer aos meus pais e à minha prima que foram elementos essenciais de apoio e ajuda durante a realização desta dissertação.

Agradeço ao meu orientador, o Prof. Dr. Pedro Duarte, pelo seu esforço em transmitir-me conteúdos relevantes para a realização deste trabalho e que me eram desconhecidos e pela sua total disponibilidade em direcionar o meu trabalho.

Aos meus amigos de longa data que foram cruciais para quebrar a rotina, trazendo-me mais inspiração e alegria durante esta jornada.

À Alexandra, à Inês e a outros amigos que conheci durante o mestrado agradeço as partilhas, risos, receios e o apoio que fomos uns para os outros.

Por fim, agradecer a disponibilidade do INE em fornecer os dados relevantes para esta dissertação que sem estes, este trabalho não teria sido possível.

Resumo

A configuração das habitações numa área específica geralmente reflete similaridades respeitantes a características estruturais, de localização e de vizinhança, evidenciando a formação de submercados de habitação distintos.

Desta forma, esta dissertação tem como objetivo identificar a existência de submercados de habitação no município do Porto através da aplicação da análise de *clusters* com restrições de contiguidade. Para esta aplicação as variáveis consideradas incluem características físicas e estruturais dos imóveis, bem como de localização. Tais variáveis foram extraídas da base de microdados anonimizados sobre o Imposto Municipal de Imóveis, detida pelo Instituto Nacional de Estatística.

Os resultados identificaram para o ano de 2019 e 2022, três submercados bem definidos na área do Porto. A sua delimitação complementada com a análise das características que definem estes submercados fornecem informações valiosas para os proprietários das casas, autarquias municipais, credores e investidores e promotores imobiliários.

Palavras-chave: Segmentação do Mercado de Habitação; Submercados de Habitação; Análise de *Clusters*; Município do Porto.

Abstract

The configuration of housing in a given area generally reflects similarities in terms of structural, location and neighbourhood characteristics, indicating the formation of distinct housing submarkets.

This dissertation therefore aims to identify the existence of housing submarkets in the municipality of Porto by applying cluster analysis with contiguity restrictions. For this application, the variables considered include the physical and structural characteristics of the properties, as well as their location. These variables were extracted from the anonymised microdata base on municipal property tax held by the National Statistics Institute.

The results identified three well-defined submarkets in the Porto area for 2019 and 2022. Their delimitation, complemented by an analysis of the characteristics that define these submarkets, provides valuable information for homeowners, municipalities, lenders, and real estate investors and developers.

Keywords: Housing Market Segmentation; Housing Submarkets; Cluster Analysis; Municipality of Porto.

Índice

Agradecimentos	iv
Resumo.....	vi
Abstract	vii
Índice	ix
Índice de Figuras.....	xi
Índice de Tabelas.....	xiii
Lista de Abreviaturas	xv
Introdução.....	17
Revisão da literatura.....	19
2.1 Definição de submercados de habitação.....	19
2.2 Delimitação de submercados.....	20
2.2.1 Delimitação baseada em critérios <i>a priori</i>	20
2.2.1 Delimitação baseada em procedimentos estatísticos.....	22
2.3 O Modelo de Preços Hedônicos	25
2.4 O Caso Português.....	28
Dados e Metodologia	31
3.1 Base de Dados.....	31
3.1.1 Descrição das variáveis.....	31
3.2 Metodologia	35
3.2.1 Análise de <i>clusters</i>	36
3.2.2 <i>Clustering</i> hierárquico com restrições de contiguidade.....	39
Resultados e discussão.....	41
Conclusão.....	55
Bibliografia.....	57
Apêndices.....	62
Apêndice A – Características para cálculo do coeficiente de localização	62
Apêndice B – Majorativos e Minorativos para cálculo do coeficiente de qualidade e conforto	62
Apêndice C – Script em R.....	63

Índice de Figuras

Figura 1 - Dendrograma que representa o <i>clustering</i> hierárquico com restrições de contiguidade para o ano de 2019.....	41
Figura 2 - Dendrograma que representa o <i>clustering</i> hierárquico com restrições de contiguidade para o ano de 2022.....	42
Figura 3 - Representação geográfica dos <i>clusters</i> para o ano de 2019	46
Figura 4 - Representação geográfica dos <i>clusters</i> para o ano de 2022	51

Índice de Tabelas

Tabela 1 - Descrição das variáveis utilizadas	32
Tabela 2 - Estatísticas descritivas das variáveis numéricas.....	33
Tabela 3 - Frequências absolutas das variáveis categóricas	34
Tabela 4 - Estatísticas descritivas das variáveis numéricas de cada <i>cluster</i> para 2019	43
Tabela 5 - Frequências absolutas das variáveis categóricas de cada <i>cluster</i> para 2019	44
Tabela 6 - Estatísticas descritivas das variáveis numéricas de cada <i>cluster</i> para 2022	49
Tabela 7 - Frequências absolutas das variáveis categóricas de cada <i>cluster</i> para 2019	50
Tabela 8 - Descrição sucinta dos submercados elaborados para ambos os anos	53

Lista de Abreviaturas

ACP – Análise em Componentes Principais

CBD – *Central Business District*

EFA – Análise Fatorial Exploratória

IMI – Imposto Municipal sobre Imóveis

INE – Instituto Nacional de Estatística

Capítulo 1

Introdução

O mercado imobiliário para habitação é caracterizado por ser segmentado e estruturado de acordo com um padrão complexo que tem em conta vários elementos e não apenas seguindo um processo homogéneo de organização espacial (J. L. Marques et al., 2012).

De facto, a habitação residencial apresenta várias e diversas características. Tais características, incluem características físicas, atributos do bairro e fatores de localização (Usman et al., 2020).

A apresentação de múltiplos atributos, leva a que os diferentes grupos sociais que participam no mercado e que possuem diferentes preferências e capacidades económicas se organizem em *clusters* (quer em termos territoriais quer em termos sociais e económicos) (J. L. Marques et al., 2012).

Desta forma, há uma heterogeneidade espacial no mercado da habitação e, como consequência são várias os motivos que revelam que a análise da segmentação deste mercado, nomeadamente no município do Porto, é relevante.

A elaboração e análise de submercados permite entender as particularidades de cada submercado, melhorando a capacidade dos investidores e financiadores para avaliarem o risco associado aos investimentos na habitação. Ao mesmo tempo, os próprios consumidores de habitação adquirem informação sobre como os limites dos submercados estão definidos (Goodman & Thibodeau, 2007).

Numa ótica estratégica, a definição de submercados contribui para uma melhor compreensão de possíveis problemas nas áreas específicas dos submercados. Este reconhecimento é benéfico para a avaliação fiscal e o para o desenvolvimento comunitário, fornecendo uma fonte eficaz para os planeadores

e decisores políticos investigarem a mudança dinâmica no sistema habitacional (Keskin & Watkins, 2017; Sairi et al., 2022).

O tema da habitação é uma preocupação atual e emergente em Portugal, constituindo uma motivação para estudo da temática da definição de submercados no município do Porto.

Este panorama, aliada à falta de estudos similares nesta zona do país, tornam claro que a delimitação de submercados habitacionais no município do Porto, proposta nesta dissertação é, também, emergente e relevante.

Em particular, este estudo pretende responder a duas questões de investigação: i) Como é que o mercado de habitação se segmenta no município do Porto?; ii) A segmentação do mercado de habitação no município do Porto evoluiu ao longo do tempo?

De forma a dar resposta a estas questões esta dissertação está dividida em quatro partes. A primeira diz respeito à revisão da literatura, dando destaque à explicação de como os submercados têm sido delimitados. Este capítulo é seguido pelo capítulo dos dados e metodologia, onde é descrita a estrutura dos dados, o seu tratamento, a justificação da adoção do método de *clustering* hierárquico aglomerativo com restrições de contiguidade e a sua aplicação, os resultados obtidos e a discussão. Finalmente, o último capítulo apresenta a conclusão, que sumariza fundamentalmente os principais resultados desta dissertação.

Capítulo 2

Revisão da literatura

2.1 Definição de submercados de habitação

Ao longo dos últimos anos, têm sido vários os autores que se debruçam sobre o estudo do mercado imobiliário. Apesar de ser um tema amplamente estudado, a complexidade surge desde o início, sendo a própria definição de submercado um problema (Bourassa et al., 1999; Watkins, 2001).

Uma das definições mais antigas e amplamente utilizadas, define submercado de habitação como "um conjunto de habitações que são substitutos razoavelmente próximos uns dos outros, mas que são substitutos relativamente fracos de habitações noutros submercados" (Bourassa et al., 1999; C. Wu & Sharma, 2012). Esta abordagem foi anteriormente utilizada por Rapkin e Winnick (1953) no seu livro *"Housing Market Analysis: A Study of Theory and Methods"* onde definem submercado habitacional como "(...) a área física dentro da qual todas as unidades habitacionais estão ligadas entre si numa cadeia de substituição", considerando que "(...) cada unidade habitacional num mercado local de habitação pode ser considerada um substituto de qualquer outra unidade" (J. L. Marques et al., 2012).

O conceito económico de substituição encontra lugar de destaque nas definições mencionadas por vários autores. Este conceito levanta difíceis questões sobre como identificar substitutos próximos e sobre os níveis de agregação (ou desagregação) (Bourassa et al., 1999).

Mais recentemente, Goodman & Thibodeau (2007) referem que os submercados de habitação são áreas onde o preço por unidade de quantidade de habitação (definido através de algum índice de características da habitação) é constante.

Os autores, mencionam que a própria identificação de áreas geográficas com preços constantes por unidade de habitação é um desafio. Facto justificado porque a habitação é um bem heterogéneo e o valor de mercado de uma casa (estimado pelo seu preço de transação) é uma função das características do local, da estrutura, da vizinhança e da localização da propriedade (Goodman & Thibodeau, 2007).

2.2 Delimitação de submercados

Apesar do vasto leque de definições de submercados de habitação, é consensual que estes são, normalmente, definidos em termos de áreas geográficas ou de características físicas das habitações (Bourassa et al., 2002).

Contudo, a concordância com o facto da delimitação dos submercados estar dependente de características espaciais ou físicas, não afasta a falta de consenso que existe na identificação, na prática, dos submercados (Leishman et al., 2013; Watkins, 2001; Xiao et al., 2016).

Desta feita, os desenvolvimentos nos procedimentos e metodologias para identificar e delimitar submercados são significantes. De acordo com Wu & Sharma (2012), as metodologias de classificação podem ser divididas em duas grandes categorias: classificações *a priori* e metodologias baseadas em dados.

2.2.1 Delimitação baseada em critérios *a priori*

As definições *a priori* baseiam-se em divisões espaciais convenientemente disponíveis ou em critérios pré-definidos (atributos estruturais da habitação ou características do grupo de utilizadores). É dentro deste grupo que, com base na sua metodologia, a dificuldade em identificar substitutos próximos e quais os níveis de agregação (ou desagregação), mencionada anteriormente, pode ser

ultrapassada (Bourassa et al., 1999; Kopczewska & Cwiakowski, 2021; Xiao et al., 2016).

Uma das divisões espaciais mais comuns são os limites imobiliários (Wu & Sharma, 2012). Esta abordagem foi utilizada por Bourassa et al. (2002) quando comparam um conjunto de submercados baseado em pequenas áreas geográficas definidas por avaliadores imobiliários com um conjunto de submercados executados estatisticamente. Do mesmo modo, Keskin & Watkins (2017) exploram os benefícios relativos das fronteiras dos submercados definidas por "peritos" – agentes, avaliadores e analistas de mercado. É, também, exemplo a delimitação do mercado de São Francisco feita por Palm (1978) com base em distritos dentro dos quais os agentes imobiliários trocam informações sobre as ofertas de emprego (Watkins, 2001).

Outros estudos segmentaram o mercado imobiliário em submercados tendo por base blocos de recenseamento agregados (Goodman & Thibodeau, 2003, 2007), códigos postais (Goodman & Thibodeau, 2003), rede rodoviária urbana (Xiao et al., 2016), limites da administração local (Bourassa et al., 1999) e características físicas (Watkins, 2001).

Nas divisões espaciais mencionadas anteriormente, que derivam de subdivisões do espaço já existentes ou de critérios pré-definidos, não é refletida a evolução dos padrões socioeconómicos e as preferências dos consumidores em matéria de habitação. Como resultado, as classificações espaciais *a priori*, que frequentemente possuem uma natureza estática, falham em apreender a dinâmica dos submercados de habitação (C. Wu & Sharma, 2012).

Em contraste, os submercados *a priori* baseados nas características habitacionais e demográficas podem captar a natureza evolutiva do mercado da habitação. Face ao exposto, são vários os autores que consideram este tipo de variáveis nos seus estudos.

No seu trabalho, Palm (1978) para além de ter delimitado o mercado da habitação com base na jurisdição dos conselhos imobiliários, também teve em conta a composição étnico-racial do bairro. Outros estudos segmentaram os mercados da habitação em submercados com base em características socioeconómicas (Farber, 1986), rendimento do grupo de consumidores (Schnare & Struyk, 1976), tipo de imóvel (Adair et al., 1996) e dimensão da habitação (Goodman & Thibodeau, 2007).

Alguns estudos de submercados reconheceram explicitamente a importância conjunta das características espaciais e estruturais na definição dos submercados (Watkins, 2001).

Leishman (2001) salienta que os mercados de habitação podem ser subdivididos tanto a nível espacial como estrutural, formando, deste modo, um conjunto de submercados interligados. Neste campo, vários autores propuseram classificações em camadas com base numa combinação de limites espaciais, características da habitação e dimensões socioeconómicas (Schnare & Struyk, 1976; Tu, 1997; Watkins, 2001).

2.2.1 Delimitação baseada em procedimentos estatísticos

Em contraste com as delimitações baseadas em conhecimentos *a priori*, uma abordagem alternativa consiste em deixar que os dados determinem a estrutura dos submercados. Por conseguinte, foram desenvolvidas abordagens alternativas que utilizam procedimentos estatísticos em vez de juízos *a priori*.

Goodman & Thibodeau (1998) basearam o seu trabalho em análise estatísticas, embora que baseadas parcialmente em conhecimento *a priori*. Os autores utilizaram um modelo hierárquico para delimitar as áreas onde a variação da qualidade das escolas públicas explica a variação do coeficiente hedónico do tamanho da propriedade nas 18 zonas de escolas primárias de um distrito escolar suburbano de Dallas.

Esta abordagem tem como ideia principal de que todas as habitações de uma área espacialmente concentrada partilham as comodidades associadas à localização da propriedade. Consequentemente, as características da habitação que determinam o valor de mercado de uma propriedade estão inseridas numa hierarquia - propriedades dentro de bairros, bairros dentro de zonas escolares, zonas escolares dentro de municípios, e assim por diante. Esta abordagem foi, mais tarde, aplicada num estudo que utilizou 28.000 habitações unifamiliares na área mais alargada de Dallas (Goodman & Thibodeau, 2003).

A principal crítica a este método hierárquico está relacionada com o facto de que, embora seja mais técnico e fundamentado em dados empíricos do que as abordagens iniciais, ainda depende de pressupostos prévios para determinar as fronteiras administrativas mais apropriada, como é o caso da utilização das fronteiras escolares (Keskin & Watkins, 2017).

Os estudos que utilizam uma abordagem totalmente baseada em procedimentos estatísticos, baseiam a sua metodologia em duas etapas essenciais.

A primeira diz respeito à análise da base de dados que poderá passar pela aplicação de uma metodologia de redução de dimensionalidade, como a Análise Fatorial Exploratória (EFA), ou Análise em Componentes Principais (ACP). Estas metodologias permitem substituir as variáveis originais por um conjunto mais reduzido de variáveis sintéticas, conhecidas como Fatores ou Componentes, que sumariam o essencial da informação original (Bryant & Yarnold, 1995).

A segunda, está relacionada com as técnicas que possibilitem a definição adequada de submercados, como é o caso da análise de *clusters*.

Nesta ótica, iniciando o estudo com um grande conjunto de dados, é possível determinar quais as características que mais se distinguem entre as habitações e depois agrupar as habitações de acordo com essas características. Ou, se os dados se limitarem a pequenas áreas geográficas, então pode permitir-

se que esses dados determinem a forma como as áreas devem ser agrupadas (Bourassa et al., 1999).

Por esta ser uma metodologia que integra uma vasta gama de atributos da habitação tende a ser, em oposição às classificações *a priori*, mais objetiva. Adicionalmente, tem em consideração a evolução temporal dos submercados habitacionais, uma vez que os dados imobiliários são regularmente atualizados de forma oportuna (C. Wu & Sharma, 2012).

Apesar de ser uma metodologia muito útil, o facto de poder ser utilizado um vasto número de variáveis conduz ao problema de reduzir essas variáveis a um número mais reduzido de fatores-chave. Hubbard & Allen (1987) mencionam que a utilização da ACP é útil para a extração de um pequeno número de fatores de um conjunto de dados mais vasto.

O trabalho de Dale-Johnson (1982) tem particular relevância neste campo. O autor, utiliza a análise fatorial em 13 variáveis e extrai cinco fatores que são utilizados para definir 10 submercados. A definição de submercados é feita a partir dos cinco fatores extraídos, onde para cada um deles são definidos dois submercados. Um deles contém as transações que melhor se caracterizam pelo fator (mais parecidas com o fator), enquanto o outro segmento contém as transações que melhor não se caracterizam pelo fator (mais diferentes do fator).

Bourassa et al. (1999) argumenta que esta forma da extração dos fatores para a definição dos submercados é um problema, já que esta abordagem deixa de fora a informação contida nos outros fatores. E, sugere que uma abordagem preferível é utilizar a análise de *clusters* em todos os fatores significativos.

Vários estudos utilizaram a análise de *clusters* ou a combinação da ACP e de *clusters* para obter submercados imobiliários. Bourassa et al. (1999) utilizam a ACP e a análise de *clusters* para formar submercados de habitação em Sydney e Melbourne, na Austrália. Através da ACP identificaram as características importantes dos distritos da administração local. Posteriormente, a análise de

clusters foi utilizada nos componentes principais para determinar os agrupamentos mais adequados de distritos.

No trabalho realizado por Maclennan e Tu (1996), foi elaborada uma matriz de submercados que abrangia quatro setores urbanos, bem como cinco categorias de grupos de produtos. Tais agrupamentos foram definidos com base na ACP das características das habitações e das zonas circundantes, seguida de uma análise de *clusters*.

2.3 O Modelo de Preços Hedónicos

O Modelo de Preços Hedónicos é um modelo que explica os preços dos imóveis com base nos seus atributos. Estes modelos têm sido amplamente utilizados para uma série de fins na economia da habitação, incluindo a análise de submercados (Bourassa et al., 1999).

Uma das finalidades típicas na análise dos mercados de habitação envolve a estimação de preços hedónicos para vários submercados que foram assumidos ou definidos e, em seguida, testar a estabilidade estrutural dessas equações (Bourassa et al., 2003).

Existe um vasto corpo de literatura que utiliza a metodologia dos preços hedónicos, contudo, é o trabalho seminal de Rosen (1974), que estabelece a base teórica formal para a aplicação do modelo de preço hedónico na modelação do mercado imobiliário (Usman et al., 2020). Esta abordagem parte do princípio de que uma propriedade residencial pode ser desmembrada numa série de componentes individuais, cada um com um valor implícito.

Segundo Rosen (1974), a teoria dos preços hedónicos é uma ferramenta metodológica que descreve como o preço de uma unidade habitacional é determinado pelos atributos físicos do imóvel, os atributos da vizinhança e as características da localização. Esta técnica baseia-se numa análise estatística que considera o preço da habitação como a variável dependente, enquanto as

características – do imóvel, da vizinhança e de localização – atuam como variáveis independentes na explicação desse preço (Keskin, 2008; Usman et al., 2020). Desta forma, tradicionalmente o modelo de regressão de preços hedónicos pode ser formalmente representado da seguinte forma:

$$P = f(F, V, L)$$

onde P é o preço de transação de um imóvel; F é o vetor das características físicas do imóvel; V é o vetor das características da vizinhança e L é p vetor de características da localização do imóvel.

A caraterização de um imóvel com base nos seus atributos físicos é uma prática comum no domínio da avaliação imobiliária, uma vez que o preço está fortemente dependente de variáveis quantitativas como a área, o número de divisões ou a idade. Relativamente a esta última, é expectável que o preço dos imóveis diminua à medida que envelhecem. E, são alguns os estudos que concluem que a relação entre a idade do imóvel e o preço é maioritariamente negativa (Chin & Chau, 2003; Clark et al., 2000; Rodriguez & Sirmans, 1994). Tal acontece porque as casas mais antigas acarretam mais custos de manutenção e reparação, e também têm uma utilidade reduzida devido a alterações na conceção e nos sistemas elétricos e mecânicos (Chin & Chau, 2003).

A qualidade estrutural dos imóveis é uma característica valorizada pelos compradores e, como tal, tem um impacto no preço da unidade habitacional. Porém, a medição objetiva e precisa desta característica tem sido um entrave no estudo (Chin & Chau, 2003).

Para além dos atributos físicos, o ambiente circundante ao imóvel tem um peso considerável na determinação do seu preço. É na vizinhança do imóvel onde ocorrem interações sociais e económicas, variando em intensidade de acordo com

a sua localização, o que é crucial para a vida da comunidade. Consequentemente, benefícios e desvantagens das comodidades e interações locais são refletidos nos preços dos imóveis (Usman et al., 2020).

Atributos como a existência de estacionamento, segurança, parques e instalações de lazer circundantes, composição dos vizinhos em termos étnicos, raciais, etários e de habilitações literárias são frequentemente utilizados para refletir as características da vizinhança.

Em termos práticos, os serviços públicos, como hospitais e escolas, podem valorizar os imóveis próximos, mas também podem trazer inconvenientes, como poluição e ruído, pelo que podem depreciar o valor das propriedades (Yang et al., 2018).

A localização é considerada como o fator determinante mais importante do preço do imóvel (Heyman & Manum, 2016; Keskin, 2008; Özyurt, 2014; Usman et al., 2020; Watkins, 2001). De facto, a incorporação das relações espaciais nas equações hedónicas pode melhorar significativamente a precisão do modelo de valores ao mesmo tempo que pode reduzir os erros de estimativa para os submercados (Basu & Thibodeau, 1998; Bourassa et al., 2007; Keskin, 2008).

Na visão tradicional da localização, a acessibilidade é medida em termos de acesso ao *Central Business District (CBD)* - uma área em que se assume que a atividade comercial e o emprego estão concentrados. Efetivamente, o modelo económico urbano clássico adota uma suposição monocêntrica, na qual o *CBD* é considerado o coração da cidade, recorrendo à distância linear até este ponto como o principal indicador de valor para uma localização específica (Ibeas et al., 2012; Usman et al., 2020).

Atualmente, a principal questão de preocupação não é se a localização é importante, mas sim como incorporá-la na análise do mercado imobiliário. Embora a incorporação de atributos físicos e de vizinhança no modelo hedónico seja relativamente simples, a incorporação do fator localização no modelo

hedónico é bastante difícil, subjetiva e implícita (Usman et al., 2018). Um dos métodos que permite contornar esta questão é criar uma equação distinta para cada submercado (Basu & Thibodeau, 1998; Keskin, 2008).

A substituíbilidade, referida no início deste capítulo, integra uma das componentes da definição de submercados (Bourassa et al., 1999; Keskin & Watkins, 2017; Rothenberg et al., 1993; C. Wu & Sharma, 2012). Para se averiguar a existência de substituíbilidade, isto é, para determinar se os mercados são distintos pode-se, também, recorrer à criação de uma equação hedónica distinta para cada submercado (Bourassa et al., 1999, 2002).

A HPM apresenta-se como um dos métodos adequados para definir os mercados da habitação ao encontrar áreas em que os coeficientes das equações de preços hedónicos são semelhantes (J. J. L. Marques, 2012). Para além disso, as equações hedónicas podem ser utilizadas para comparar classificações alternativas de submercados (Bourassa et al., 1999).

2.4 O Caso Português

A literatura sobre a delimitação de submercados em Portugal é limitada, e, como consequência não é refletida a importância desta temática para a compreensão da dinâmica do mercado imobiliário e das suas implicações para eventuais políticas de planeamento urbano e estratégias de investimento no setor habitacional.

Um dos principais trabalhos nesta área pertence ao José Lourenço Marques, cujo estudo *“Spatial heterogeneity across housing sub-markets in an urban area of Portugal”*, derivado da sua tese de doutoramento *“The notion of space in urban housing markets”* (J. J. L. Marques, 2012; J. L. Marques et al., 2012).

Este estudo teve como objetivo apresentar várias abordagens para definir os submercados da habitação, mas cingindo-se à área urbana de Aveiro-Ílhavo, em Portugal.

O trabalho teve como base os dados de propriedades entre 2000 e 2010 e, para além de considerar o preço, utilizou atributos físicos das habitações e características de localização e vizinhança para caracterizar submercados. A metodologia deste trabalho dividiu-se em duas abordagens principais: uma abordagem indutiva e uma abordagem analítica.

Na primeira abordagem, abordagem indutiva, o autor definiu e delimitou os submercados considerando as fronteiras administrativas, estrutura urbana, características demográficas e históricas e evolução urbana. Isto é, utilizou um método de classificação *a priori*, já explorado anteriormente.

A abordagem analítica baseou-se na análise de *clustering* espacial aplicada a diferentes critérios. Usualmente, existem três alternativas para a caracterização e delimitação dos submercados habitacionais: a primeira tem em conta apenas os atributos geográficos e de localização; a segunda delimita os submercados considerando apenas as características estruturais da habitação; a terceira, combinando as características estruturais das habitações com sua localização geográfica para formar subcategorias dentro de um mercado mais amplo (Bourassa et al., 2007; Watkins, 2001).

No seu estudo, Marques foge à abordagem tradicional e, adota cinco critérios para a delimitação dos submercados: o primeiro usa o preço da habitação por metro quadrado (em logaritmo); o segundo teve como base as características físicas e de localização da habitação (aplicação de ACP); o terceiro e quarto consideram os preços implícitos da habitação resultantes da aplicação de um modelo hedónico para explicar o valor de um imóvel (num o valor de cada zona é baseado na estrutura espacial do mercado imobiliário; noutro caso, são considerados os preços implícitos das características físicas e de localização do

imóvel); o quinto agrega e combina algumas das abordagens anteriormente descritas, ou seja, a segmentação do mercado da habitação resultante de três dimensões principais: as características da habitação; a sua importância na avaliação do imóvel (coeficientes hedônicos do modelo de regressão) e o preço por metro quadrado de uma habitação.

Desta forma, o autor apresentou uma análise explicativa onde foram analisadas múltiplas perspectivas na definição e análise de submercados, refletindo a complexidade e interdependência dos fatores que definem o valor imobiliário em áreas urbanas.

Capítulo 3

Dados e Metodologia

3.1 Base de Dados

Para a construção de uma base de dados fidedigna e útil para a aplicação de análises de *clusters* foi utilizada uma base de dados elaborada pelo Instituto Nacional de Estatística (INE) a partir do Imposto Municipal sobre Imóveis (IMI). Como tal, foram analisados os anos 2019 e 2022 para o município do Porto.

Apesar de ser a principal base de dados pública disponível para ser analisada, esta apresenta, naturalmente, algumas limitações. Primeiro, não contém variáveis passíveis de analisar características da vizinhança quer em termos sociais, económicos ou demográficos. Segundo, depois de feita a limpeza, o número de variáveis remanescentes não é exaustivo, condicionando em certa parte o objetivo desta dissertação. Por último, não contempla dados relativos a preços de imóveis transacionados, limitando-se apenas à variável que descreve o preço da avaliação para efeitos fiscais.

Tendo em conta as limitações referidas, nesta dissertação não vão ser estimados modelos de preços hedónicos. Embora a identificação aqui feita seja útil para trabalhos posteriores que façam esta estimação, e nesse caso deverá ser estimada uma equação separadamente para cada segmento.

3.1.1 Descrição das variáveis

Utilizando a base de microdados anonimizados sobre o Imposto Municipal sobre Imóveis (IMI) detida pelo INE e obtida por via administrativa da Autoridade Tributária, restringiu-se a pesquisa aos dados relativos ao município do Porto.

Após uma adequada análise das variáveis, foram consideradas pertinentes as variáveis que constam na Tabela 1, sendo que todas descrevem características respeitantes a características físicas/estruturais do imóvel com exceção da variável “coeficiente de localização” que descreve características de localização e da variável “coeficiente qualidade-conforto”, que contempla elementos que tanto descrevem características físicas como de localização – e, por isso, considerada como característica mista. Para serem feitas inferências acerca dos resultados é imprescindível ter a noção do significado de cada variável e, por essa razão, na tabela é também apresentada uma descrição de cada uma.

Tabela 1 - Descrição das variáveis utilizadas

	Variável	Descrição/Definição
Características físicas	Tipo de Prédio	Esta variável apresenta código numéricos associados a cada tipo de prédio, sendo eles: 4 – Prédio em propriedade total sem andares; 5 – Prédio em Regime de Propriedade horizontal; 6 – Prédio em propriedade total com andares.
	Tipo de Proprietário	Esta variável apresenta código numéricos associados a cada tipo de proprietário, sendo eles: 1 – Único Proprietário; 2 – Comproprietário; Outros – Usufrutuário ou Superficiário.
	Número pisos da fração	Número de pisos respeitantes à fração imobiliária
	Número de pisos	Número de pisos do imóvel, excluindo frações imobiliárias
	Área total do terreno	Área total do terreno em m ²
	Área de implantação do edifício	Área implantação edifício em m ²
	Tipologia	Descreve a tipologia (T1, T2, T3, T4) ou o número de divisões (1, 2, 3). O número de divisões contempla o número de quartos e salas; não inclui casas-de-banho, cozinhas e arrecadações.
	Idade	Idade do imóvel em anos
	Valor da avaliação	Trata-se do Valor Patrimonial Tributário (VPT), ou seja, o valor estipulado de um determinado imóvel para questões fiscais (em euros)

Característica de localização	Coeficiente de Localização	Conforme o artigo 42.º do CIMI, e a redação da Lei n. 64-B/2011, de 30 de dezembro, o coeficiente de localização (Cl) varia entre 0.4 e 3.5, podendo, em situações de habitação dispersa em meio rural, ser reduzido para 0,35. Este visa quantificar a localização do prédio para efeitos de avaliação no âmbito do IMI, de acordo com a verificação e um conjunto de características ¹ .
Característica mista	Coeficiente de qualidade e conforto	Conforme previsto no artigo 43.º do CIMI, o coeficiente de qualidade e conforto (Cq) é aplicado ao valor base do prédio edificado, podendo ser majorado até 1,7 e minorado até 0,5, e obtém-se adicionado à unidade os coeficientes majorativos e subtraindo à unidade os coeficientes minorativos ² .

Fonte: elaboração própria

O conjunto de variáveis é constituído por variáveis numéricas – número de pisos da fração, o número de pisos do imóvel, área total do terreno, área de implantação do edifício, idade, valor da avaliação, coeficiente de localização e coeficiente de qualidade e conforto e por variáveis categóricas – tipo de prédio, tipo de proprietário e tipologia.

A Tabela 2 resume as principais estatísticas descritivas das variáveis numéricas para cada um dos anos em análise.

Tabela 2 - Estatísticas descritivas das variáveis numéricas

N = 4819	2019		
Variável	Mínimo	Máximo	Média
Número pisos da fração	0,0000	5,0000	0,9647
Número de pisos	1,0000	23,000	4,6210
Área total do terreno	13,000	19639,9	961,30
Área de implantação do edifício	5,5000	19639,9	613,80
Idade	0,0000	485,000	46,050
Valor da avaliação	1660,0	1225990	70678
Coeficiente de Localização	0,0000	3,0000	1,5620
Coeficiente de qualidade e conforto	0,0000	1,6500	1,0440

¹ As características encontram-se no apêndice A.

² A lista de majorativos e minorativos encontra-se no apêndice B.

N = 6012		2022	
Variável	Mínimo	Máximo	Média
Número pisos da fração	0,0000	4,0000	1,048
Número de pisos	1,0000	22,000	5,275
Área total do terreno	8,5000	19639,9	1846,1
Área de implantação do edifício	2,2750	19138,75	950,78
Idade	0,0000	142,000	19,16
Valor da avaliação	3590,0	1933690	88573,0
Coeficiente de Localização	0,0000	3,0000	1,6210
Coeficiente de qualidade e conforto	0,0000	1,6000	1,0850

Fonte: elaboração própria

Uma análise preliminar dos dados em bruto, permite concluir que em comparação com 2019, os dados de 2022 mostram um aumento geral nas médias de várias variáveis, com exceção da idade dos imóveis. De facto, a média das idades das propriedades em 2022 é consideravelmente menor do que em 2019, o que pode indicar um aumento na construção de novas propriedades ou uma tendência para propriedades mais recentes no mercado.

A Tabela 3 apresenta as frequências absolutas das variáveis categóricas para ambos os anos.

Tabela 3 - Frequências absolutas das variáveis categóricas

	2019 N = 4812	2022 N = 6012
Variável		
Tipo de Prédio		
4 – Prédio em propriedade total sem andares	670	515
5 – Prédio em Regime de Propriedade horizontal	2849	4584
6 – Prédio em propriedade total com andares	1300	913
Tipo de Proprietário		
1 – Único Proprietário	4220	5567
2 – Comproprietário	555	437
Outros – Usufrutuário ou Superficiário	44	8
Tipologia		
T0 – Imóvel com zero quartos	1127	1514
T1 – Imóvel com um quarto	625	1823
T2 – Imóvel com dois quartos	604	905
T3 – Imóvel com três quartos	516	416

T4 – Imóvel com quatro quartos	293	-
1 – Imóvel com uma divisão	582	-
Imóveis com outro número de quartos ou divisões	1072	1354

Fonte: elaboração própria

Em ambos os anos, a maioria das propriedades são prédios em regime de propriedade horizontal, isto é, edifícios ou um conjunto de edifícios contíguos pertencentes a uma pluralidade de pessoas, tendo cada uma delas poderes exclusivos sobre uma parte específica (designada fração autónoma, ou o comumente conhecido como apartamento).

Da mesma maneira, o tipo de proprietário mais comum é o único proprietário para os dois anos.

Relativamente à tipologia, T0 e T1 são os tipos mais comuns para os anos em análise, mas é de destacar o aumento de propriedades do tipo T1 em 2022. O número de imóveis com outro número de quartos ou divisões é elevado porque esta categoria engloba várias categorias que apresentam um número reduzido de observações, mas que em conjunto contribuem para o elevado valor apresentado.

3.2 Metodologia

Foi utilizada a análise de *clusters* com restrição de contiguidade para a analisar e delimitar os submercados imobiliários no município do Porto. Para o efeito, recorreu-se ao software R.

Tal como mencionado na revisão da literatura (ver secção 2.2.1), é comum a aplicação da ACP antes da aplicação da análise de *clusters*. A ACP extrai fatores que são depois utilizados na análise de *clusters*, concentrando a análise em fatores ortogonais nos dados, em vez de múltiplas variáveis que podem estar a explicar o mesmo fator. A ACP também é interessante por si só, porque identifica as dimensões básicas que distinguem os submercados de habitação.

Contudo, tal como mencionado anteriormente, a dimensão da base de dados utilizada para o estudo e, mais especificamente, o número de variáveis possíveis a utilizar não justifica a adoção da ACP. Desta feita, o estudo inicia imediatamente a sua análise na elaboração de *clusters*.

3.2.1 Análise de *clusters*

As técnicas de análise de *clusters* surgiram nos campos das ciências biológicas e ecológicas, e têm sido extensivamente utilizadas em diversas disciplinas (Bourassa et al., 1999).

O *clustering* representa uma etapa fundamental na extração de dados, permitindo a identificação de grupos e padrões relevantes nos dados subjacentes. Os algoritmos de *clustering* categorizam os objetos de dados em conjuntos distintos, conhecidos como *clusters*, com base nas suas similaridades ou diferenças. Dentro de um *cluster* válido, os padrões tendem a ser mais similares entre si do que em relação aos padrões pertencentes a *clusters* diferentes (Frades & Matthiesen, 2009). Portanto, o principal objetivo da análise de *clusters* é aumentar a semelhança intra-grupo e a dissemelhança inter-grupo.

Sendo uma técnica que tem como base os dados e, por sua vez, as variáveis incluídas na análise, antes de aplicar esta técnica é imprescindível definir quais são as variáveis importantes para caracterizar e delimitar os submercados (J. J. L. Marques, 2012).

Como mencionado anteriormente, alguns estudos de submercados reconheceram explicitamente a importância conjunta das características espaciais e estruturais na definição de submercados (ver secção 2.2.1). O estudo de C. A. Watkins (2001) vem reforçar esta ideia ao afirmar que um modelo híbrido proporciona uma melhor abordagem empírica para delimitar os submercados.

Desta forma e uma vez que a base de dados utilizada neste estudo possui variáveis que descrevem características estruturais e espaciais, foi utilizada uma

abordagem que incorpora os dois grupos de características. Esta abordagem foi desenvolvida aplicando a análise de *clusters* com restrições de contiguidade.

Na análise de *clusters* é fundamental seguir alguns passos, dos quais dois deles têm especial importância (Hennig et al., 2015):

1. Escolha de uma medida de semelhança/dissemelhança;
2. Escolha do método de *clustering*.

A medida de distância para avaliar a semelhança ou dissemelhança entre as propriedades afeta de forma crucial a solução dos *clusters* e depende da natureza das variáveis consideradas. A Distância Euclidiana é a métrica mais utilizada quando as características/variáveis são contínuas (Frades & Matthiesen, 2009; J. L. Marques, 2012; van de Velden et al., 2018).

Nesta dissertação, tanto foram utilizadas variáveis categóricas como variáveis numéricas e, ao invés de se proceder à recodificação das variáveis foi adotada uma medida de dissemelhança que foi aplicada diretamente nos dados mistos. O coeficiente de similaridade/semelhança de *Gower* é uma das medidas de proximidade mais populares para tipos de dados mistos (Hennig et al., 2015; van de Velden et al., 2018). Desta forma, foi utilizada esta medida de dissimilaridade, seguindo o proposto por van de Velden et al. (2019).

Relativamente ao segundo passo, a análise de *clusters* pode ser feita através de vários métodos, sendo que os métodos de *clustering* hierárquico (aglomerativo) e não-hierárquico (*k-means*) são os mais utilizados (Sairi et al., 2022).

O método não hierárquico decompõe o conjunto de dados num conjunto de agrupamentos disjuntos. O algoritmo de partição mais simples e mais comumente utilizado é o *k-means* e, neste, o parâmetro *k*, que representa o número de *clusters*, deve ser especificado (Frades & Matthiesen, 2009).

Em contraste, o método hierárquico procede de forma iterativa, quer fundindo *clusters* mais pequenos em *clusters* maiores, quer dividindo *clusters* maiores, resultando numa árvore de *clusters*, chamada dendrograma, que mostra como os *clusters* estão relacionados. Através dele, é possível escolher o número de *clusters* (Frades & Matthiesen, 2009).

Uma das principais diferenças entre ambos os modelos é que o primeiro necessita de um conhecimento prévio do número de *clusters* ao passo que o segundo não requer qualquer tipo de conhecimento prévio.

Neste campo, Goodman & Thibodeau (1998) sugerem que os modelos hierárquicos fornecem uma estrutura útil para delimitar os limites do submercado habitacional.

Os passos do *clustering* hierárquico aglomerativo podem ser sumarizados da seguinte forma:

1. Computar a matriz proximidade usando uma métrica de distância particular (no caso, o coeficiente de *Gower*).
2. Cada elemento dos dados é tratado como um *cluster*.
3. Agregação dos *clusters* com baseada na métrica de distância escolhida.
4. Atualizar a matriz distância.
5. Repetir os passos 3 e 4 até restar apenas um *cluster*.

Após o cálculo da matriz distância (passo 1), este método inicia tratando cada ponto de dados como um *cluster* individual, existindo, inicialmente, exatamente o mesmo número de *clusters* e de dados.

Com a matriz de dissimilaridade que contém a “distância” entre cada par de *clusters* calculada, em cada etapa os dois *clusters* menos dissimilares são agregados, e estes mesmo *clusters* são os que vão ser agregados no passo seguinte.

Este passo é feito, sucessivamente até todos os elementos se encontrarem num único e grande *cluster*.

O critério para a escolha dos elementos que serão agregados a seguir é simples, uma vez que se baseia na escolha dos pares com menos dissemelhança. Contudo, a principal questão neste tipo de método é como se calcula a distância entre os *clusters* que se vão formando e como se atualiza a matriz de proximidade.

São várias as abordagens existentes e que respondem a esta questão. Os métodos mais comuns são: *Single Linkage*, *Complete Linkage*, *Average Linkage* e *Ward's method* (Hennig et al., 2015).

Para o objetivo desta dissertação, foi adotado o método de *Ward's* para o cálculo da distância entre *clusters*. Este método distingue-se de todos os outros porque minimiza a soma dos quadrados das distâncias de quaisquer dois *clusters* (hipotéticos) que possam ser formados em cada passo. Como resultado, é um método que tende a produzir *clusters* hipersféricos e a conter aproximadamente o mesmo número de objetos se as observações estiverem distribuídas uniformemente pelo espaço (Legendre & Legendre, 1998). A par disto, é o método amplamente adotado nos vários estudos que delimitam o submercado de habitação.

3.2.2 *Clustering* hierárquico com restrições de contiguidade

O método de *clustering* hierárquico com restrições de contiguidade descrito por Guénard & Legendre (2022) foi implementado no contexto desta dissertação.

Tradicionalmente, os estudos na área da delimitação dos submercados tendem a negligenciar o papel da contiguidade espacial uma vez que recorrem apenas a variáveis de distância, por exemplo, a distância ao *CBD*, para captar a organização espacial dos submercados habitacionais. Uma das abordagens que tem em consideração a contiguidade espacial, é o tratamento dos dados espaciais

como variáveis no processo de *clustering*, atribuindo um peso adequado a cada variável (Y. Wu et al., 2020).

O método de *clustering* hierárquico com restrições de contiguidade adotado neste estudo, vem inovar a forma como a contiguidade espacial é tratada no processo de delimitação dos submercados, dado ser um método relativamente recente e com poucas aplicações práticas, até ao momento, neste ramo.

Na sua aplicação, o método de *clustering* com restrições de contiguidade espaciais difere dos métodos sem restrições visto que não percorre todos os pares de pontos ou *clusters* para encontrar o par com a menor dissimilaridade na matriz de dissimilaridade original ou atualizada. Este método precisa apenas de considerar as dissimilaridades correspondentes aos pares contíguos. Como resultado, é geralmente um método mais rápido e que produz soluções que são menos variáveis do que as dos seus homólogos sem restrições (Guénard & Legendre, 2022).

Na prática, este modelo segue as etapas descritas anteriormente para métodos de *clustering* hierárquico aglomerativo (ver secção 3.2.1), com o acréscimo da definição da restrição de contiguidade. Para aplicar a restrição é necessário primeiro definir o significado de contíguo no contexto dos dados. Neste caso, considera-se que dois elementos são contíguos se estão suficientemente próximos um do outro.

Para este efeito, foi definida uma lista de vizinhos que ligam pares de pontos que se encontram a uma distância igual ou inferior a 500 metros. Naturalmente que durante o processo de agregação (passo 3), apenas os *clusters* que são contíguos e definidos pela lista de vizinhos, podem ser agregados.

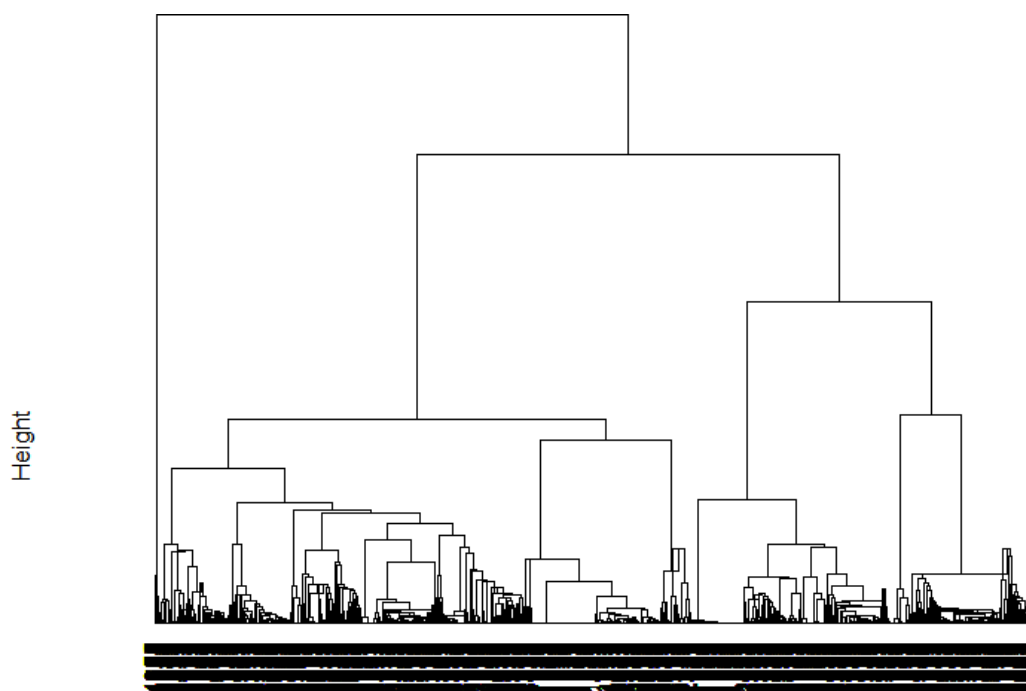
O resultado são *clusters* bem delimitados e tendo em consideração a restrição de contiguidade espacial.

Capítulo 4

Resultados e discussão

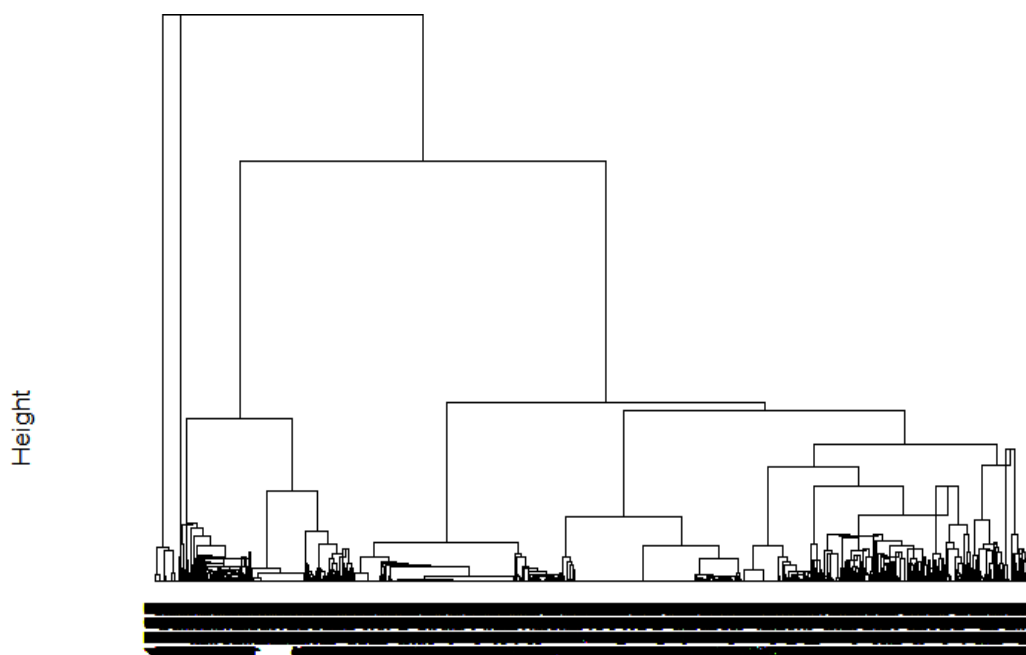
Como resultado da implementação do modelo de *clustering* hierárquico com restrições de contiguidade, as figuras 1 e 2 apresentam o dendrograma para o ano de 2019 e 2022, respectivamente.

Figura 1 - Dendrograma que representa o *clustering* hierárquico com restrições de contiguidade para o ano de 2019



Fonte: elaboração própria

Figura 2 - Dendrograma que representa o *clustering* hierárquico com restrições de contiguidade para o ano de 2022



Fonte: elaboração própria

Conforme ascendemos no dendrograma, observações similares são agrupadas em ramos, que progressivamente se juntam em pontos mais elevados da árvore. A distância vertical na qual as agregações ocorrem reflete o grau de diferença entre as observações: alturas de agregação maiores denotam menor semelhança entre as observações unidas. Assim, um nível de corte horizontal relevante do dendrograma é caracterizado por uma diferença relativamente significativa entre as alturas de dois nós sucessivos.

Tendo esta metodologia em conta foi considerado, para ambos os anos a análise de 4 *clusters* (ver figuras 1 e 2). É de notar para o ano 2019 era, também, justificável a escolha de 3 *clusters*, contudo foi explorada apenas a de 4 *clusters* por forma a tornar a comparação entre os anos mais direta e homogénea.

As tabelas 4 e 5 fornecem, para cada um dos 4 *clusters*, as estatísticas descritivas das variáveis numéricas e as frequências absolutas das variáveis categóricas para o ano de 2019 e 2022, respetivamente.

Tabela 4 - Estatísticas descritivas das variáveis numéricas de cada *cluster* para 2019

Cluster	Variável	2019		
		Mínimo	Máximo	Média
1 (N = 27)	Número pisos da fração	0,0000	3,0000	0,4074
	Número de pisos	1,0000	4,000	2,0370
	Área total do terreno	25,000	907,00	198,40
	Área de implantação do edifício	23,70	183,00	75,190
	Idade	19,00	100,00	72,810
	Valor da avaliação	6790	114160	27476
	Coeficiente de Localização	1,1000	1,1000	1,1000
	Coeficiente de qualidade e conforto	1,0000	1,1000	1,0370
	Longitude	-8,574	-8,561	-8,568
	Latitude	41,15	41,15	41,15
2 (N = 2917)	Número pisos da fração	0,0000	4,0000	1,1270
	Número de pisos	1,0000	23,000	5,6730
	Área total do terreno	19,700	19639,9	1227,0
	Área de implantação do edifício	17,800	19639,9	779,5
	Idade	0,0000	485,000	32,96
	Valor da avaliação	5530	880160	74363
	Coeficiente de Localização	0,0000	3,0000	1,5600
	Coeficiente de qualidade e conforto	0,0000	1,6500	1,0800
	Longitude	-8,685	-8,565	-8,617
	Latitude	41,14	41,18	41,15
3 (N = 1099)	Número pisos da fração	0,0000	5,0000	1,0920
	Número de pisos	1,0000	8,000	2,9950
	Área total do terreno	35,170	2068,7	468,7
	Área de implantação do edifício	32,000	1522,7	338,60
	Idade	0,0000	130,000	71,87
	Valor da avaliação	3690,0	541140	33237
	Coeficiente de Localização	1,300	2,0000	1,5740
	Coeficiente de qualidade e conforto	0,7300	1,2000	0,9369
	Longitude	-8,638	-8,588	-8,612
	Latitude	41,14	41,17	41,16
4 (N = 776)	Número pisos da fração	0,0000	4,0000	0,1946
	Número de pisos	1,0000	11,000	3,0580
	Área total do terreno	13,000	8357,0	687,00
	Área de implantação do edifício	5,5000	2532,00	399,49
	Idade	0,0000	213,000	57,720
	Valor da avaliação	1660,0	1225990	111357
	Coeficiente de Localização	0,0000	2,8000	1,5690

Coeficiente de qualidade e conforto	0,0000	1,4600	1,0630
Longitude	-8,685	-8,571	-8,619
Latitude	41,14	41,19	41,16

Fonte: elaboração própria

O *cluster* 1 caracteriza um submercado que se destaca dos demais por contemplar apenas 27 propriedades. Este pequeno submercado é formado maioritariamente por prédios em propriedade total sem andares e é nele onde se verifica uma média de idade dos imóveis mais elevada.

Os valores de avaliação dos imóveis pertencentes a este submercado são relativamente baixos, indicando a existência de propriedades menos valorizadas, provavelmente, localizadas em áreas menos centrais.

Quer o coeficiente de localização quer o coeficiente qualidade e conforto apresentam valores estáveis para o submercado em análise.

Tabela 5 - Frequências absolutas das variáveis categóricas de cada *cluster* para 2019

Variável	2019			
	C1	C2	C3	C4
Tipo de Prédio				
4 – Prédio em propriedade total sem andares	18	17	1	634
5 – Prédio em Regime de Propriedade horizontal	7	2734	1	107
6 – Prédio em propriedade total com andares	2	166	1097	35
Tipo de Proprietário				
1 – Único Proprietário	17	2490	1054	659
2 – Comproprietário	10	399	41	105
Outros – Usufrutuário ou Superficiário	0	28	4	12
Tipologia				
T0 – Imóvel com zero quartos	-	1022	90	15
T1 – Imóvel com um quarto	2	480	83	60
T2 – Imóvel com dois quartos	6	417	108	73
T3 – Imóvel com três quartos	1	385	71	59
T4 – Imóvel com quatro quartos	1	195	63	34
1 – Imóvel com uma divisão	4	153	384	52
Outros imóveis com mais quartos ou divisões	13	265	300	483

Fonte: elaboração própria

O segundo submercado é o que contém um maior número de imóveis, e, o único onde o prédio em regime de propriedade horizontal se destaca dos demais tipos de prédio, indicando um ambiente urbano com apartamentos e condomínios. Com uma análise prévia deste facto, poderá ser inferido que este submercado se encontra nas regiões fora do centro histórico e, pode representar áreas de alta densidade populacional e desenvolvimento recente. Esta inferência pode também ser corroborada pela tipologia, já que este *cluster* apresenta, de forma destacada, a existência de imóveis do tipo T0. É natural, que esta tipologia apresente maior expressão em áreas urbanas onde a procura por espaços compactos é elevada, nomeadamente por estudantes ou profissionais solteiros.

Este submercado é também o que apresenta a média da idade dos edifícios mais baixa, e que aliada com a alta variação nos valores de avaliação podem refletir uma mistura de propriedades novas e antigas, com uma possibilidade de recente desenvolvimento ou renovação na área.

Ainda, é de destacar que este submercado apresenta, em média, propriedades com maior qualidade e conforto, ainda que a diferença seja quando comparados com os restantes submercados seja mínima.

O *cluster* 3 destaca-se por dois motivos: o primeiro é que das 1 099 propriedades que o compõe, 1 097 são prédios em propriedade total com andares; o segundo é a idade média das habitações que fazem parte deste submercado ser elevada.

Ambos os factos sugerem que este submercado possa estar inserido numa zona caracterizada essencialmente por edifícios mais antigos e pertencentes a um só proprietário.

No submercado 4, aproximadamente 82% das propriedades são prédios em propriedade total sem andares e 62% são imóveis que possuem mais do que quatro quartos ou quatro divisões. Este facto pode ser suportado pela área total

Através da figura 3, fica claro que existem dois submercados que se destacam pela sua definição mais afincada, sendo eles o submercado 1 e o submercado 3.

O submercado 1 é o que se encontra mais isolado e onde todas as propriedades que dele fazem parte se encontram na freguesia de Campanhã, fora do centro urbano do Porto. As coordenadas médias deste submercado indicam que a localização ronda a zona dos bairros do Largateiro, Granja e Azevedo. Nestas zonas há prevalência de casas térreas que vêm confirmar a descrição anterior deste *cluster*, onde se concluiu a predominância de prédios em propriedade total sem andares com idades mais elevadas.

No que respeita ao submercado 3, a variação nas coordenadas é relativamente pequena, o que sugere uma concentração geográfica dos imóveis. Tendo em conta o mapa, é possível identificar que este submercado se concentra numa área central tendo imóveis presentes na união das Freguesias de Cedofeita, Santo Ildefonso, Sé, Miragaia, São Nicolau e Vitória e na freguesia de Bonfim, validando a descrição anterior sobre este submercado. Desta forma, é possível concluir que este submercado encontra-se bem delimitado e está contemplado maioritariamente na zona central do município do Porto, onde estão localizadas as principais atrações turísticas e grande parte do comércio portuense.

Para além disso, é nesta zona do porto onde se encontra o centro histórico da cidade e onde existe um Plano Diretor Municipal que prevê a preservação da identidade dos lugares urbanos para além dos imóveis individuais, tomando em consideração o cadastro, as frentes urbanas, as formas de cobertura, a geografia física, os elementos de identificação urbana, o coberto vegetal, entre outros.

Desta forma, é natural que nesta área não existam edifícios muito altos e, conseqüentemente, que não existam edifícios em regime de propriedade horizontal. Estes factos são corroborados pela descrição deste *cluster* feita acima, já que nele predominam os prédios em regime total com andares.

Outro facto a ter em conta é o tipo de proprietário. A maior parte dos prédios têm um único proprietário (ver tabela 5) e, este dado vai de encontro, por exemplo, à presença maioritária de prédios para arrendamento nomeadamente para alojamento local nesta zona do município.

Tanto o submercado 2 como o submercado 4 apresentam uma grande dispersão geográfica, tendo observações em todas as freguesias do Porto, mas ainda assim apresentam algumas concentrações.

É visível que o submercado 2 tem mais expressão na freguesia de Paranhos quando comparado com o submercado 4. Na união das Freguesias de Aldoar, Foz do Douro e Nevogilde ambos os submercados detêm observações, mas na zona costeira predomina o submercado 4, ao passo que o submercado 2 tem maior expressão na zona nordeste desta freguesia.

A freguesia de Paranhos detém o maior polo universitário da universidade do Porto, o polo Asprela. Desta forma, a habitação para estudantes tem especial impacto na conjectura do mercado habitacional desta zona. É, portanto, de esperar que esta seja uma zona onde haja predominância de apartamentos. Efetivamente, o submercado 2 é constituído na sua maioria, por este tipo de edifícios.

Apesar da tipologia predominante no *cluster 2* ser edifícios T0 e T1, este facto parece, *à priori*, não ir de encontro ao panorama da habitação para estudantes. O que é facto é que, normalmente, os edifícios para estudantes são constituídos por vários quartos por forma a rentabilizar o espaço, e assim ser mais vantajoso para os arrendatários. Contudo, é necessário ter em conta que muitos quartos podem ser construídos com base em divisões pré-existentes como é o caso de salas de estar. E, desta forma, o número de quartos para cálculo do IMI não ser o real.

O *cluster 4* é constituído por casas maiores e mais luxuosas, podendo desta forma explicar a predominância deste submercado na zona costeira da Foz do Porto, reconhecida por um mercado de alto valor imobiliário e luxuoso.

A análise dos *clusters* para o ano 2022 tem em conta as mesmas características analisadas para o ano de 2019. Desta forma, as tabelas 6 e 7 fornecem, para cada um dos 4 *clusters*, as estatísticas descritivas das variáveis numéricas e as frequências absolutas das variáveis categóricas para o ano de 2022, respetivamente.

Tabela 6 - Estatísticas descritivas das variáveis numéricas de cada *cluster* para 2022

				2022
Cluster	Variável	Mínimo	Máximo	Média
1 (N = 170)	Número pisos da fração	1,0000	1,0000	1,000
	Número de pisos	6,0000	8,000	7,8590
	Área total do terreno	689,00	4682,0	4401,0
	Área de implantação do edifício	223,40	2170,40	2033,0
	Idade	0,0000	0,0000	0,0000
	Valor da avaliação	41350,0	131000	68001
	Coeficiente de Localização	1,5000	1,5000	1,5000
	Coeficiente de qualidade e conforto	1,0300	1,0300	1,0300
	Longitude	-8,607	-8,604	-8,606
	Latitude	41,18	41,18	41,18
2 (N = 7)	Número pisos da fração	0,0000	3,0000	2,5710
	Número de pisos	2,0000	3,000	2,8570
	Área total do terreno	314,000	746,200	375,70
	Área de implantação do edifício	102,800	188,00	175,80
	Idade	1,0000	85,000	13,000
	Valor da avaliação	83870	118420	112433
	Coeficiente de Localização	1,4000	1,4000	1,4000
	Coeficiente de qualidade e conforto	1,1100	1,1700	1,1190
	Longitude	-8,577	-8,575	-8,576
	Latitude	41,18	41,18	41,18
3 (N = 1189)	Número pisos da fração	0,0000	3,0000	0,7241
	Número de pisos	1,0000	10,000	2,942
	Área total do terreno	8,5000	3440,0	539,40
	Área de implantação do edifício	8,5000	1806,00	281,71
	Idade	0,0000	142,000	64,120
	Valor da avaliação	3590	1933690	55611
	Coeficiente de Localização	0,0000	2,1000	1,4930

	Coeficiente de qualidade e conforto	0,0000	1,3400	1,0110
	Longitude	-8,654	-8,561	-8,607
	Latitude	41,14	41,18	41,16
	Número pisos da fração	0,0000	4,0000	1,1310
	Número de pisos	1,0000	22,000	5,7810
	Área total do terreno	29,7000	19639,9	2089,3
4	Área de implantação do edifício	2,2750	19138,75	1083,57
(N = 4646)	Idade	0,0000	120,000	8,361
	Valor da avaliação	5980	1115650	97725
	Coeficiente de Localização	0,0000	3,0000	1,6580
	Coeficiente de qualidade e conforto	0,0000	1,6000	1,1060
	Longitude	-8,684	-8,567	-8,621
	Latitude	41,14	41,18	41,16

Fonte: elaboração própria

O *cluster* 1 define um submercado composto inteiramente por prédios em regime de propriedade horizontal e com as tipologias T0, T1 e T2 a prevalecer quase de forma igualitária.

O *cluster* 2 destaca-se dos restantes por ser apenas composto por 7 imóveis, dos quais 6 são prédios em regime de propriedade horizontal. Este submercado exhibe a maior variação no valor de avaliação, indicando uma heterogeneidade acentuada que pode corresponder a uma área de desenvolvimento recente ou renovação.

Tabela 7 - Frequências absolutas das variáveis categóricas de cada *cluster* para 2019

	2022			
Variável	C1	C2	C3	C4
Tipo de Prédio				
4 – Prédio em propriedade total sem andares	0	1	365	149
5 – Prédio em Regime de Propriedade horizontal	170	6	17	4391
6 – Prédio em propriedade total com andares	0	0	807	106
Tipo de Proprietário				
1 – Único Proprietário	170	6	1119	4272
2 – Comproprietário	0	1	69	367
Outros – Usufrutuário ou Superficiário	0	0	1	7
Tipologia				
T0 – Imóvel com zero quartos	58	-	139	1317
T1 – Imóvel com um quarto	57	6	87	1673

T2 – Imóvel com dois quartos	53	-	56	796
T3 – Imóvel com três quartos	2	-	50	364
T4 – Imóvel com quatro quartos	-	1	11	161
1 – Imóvel com uma divisão	-	-	80	76
Outros imóveis com mais quartos ou divisões	-	-	766	259

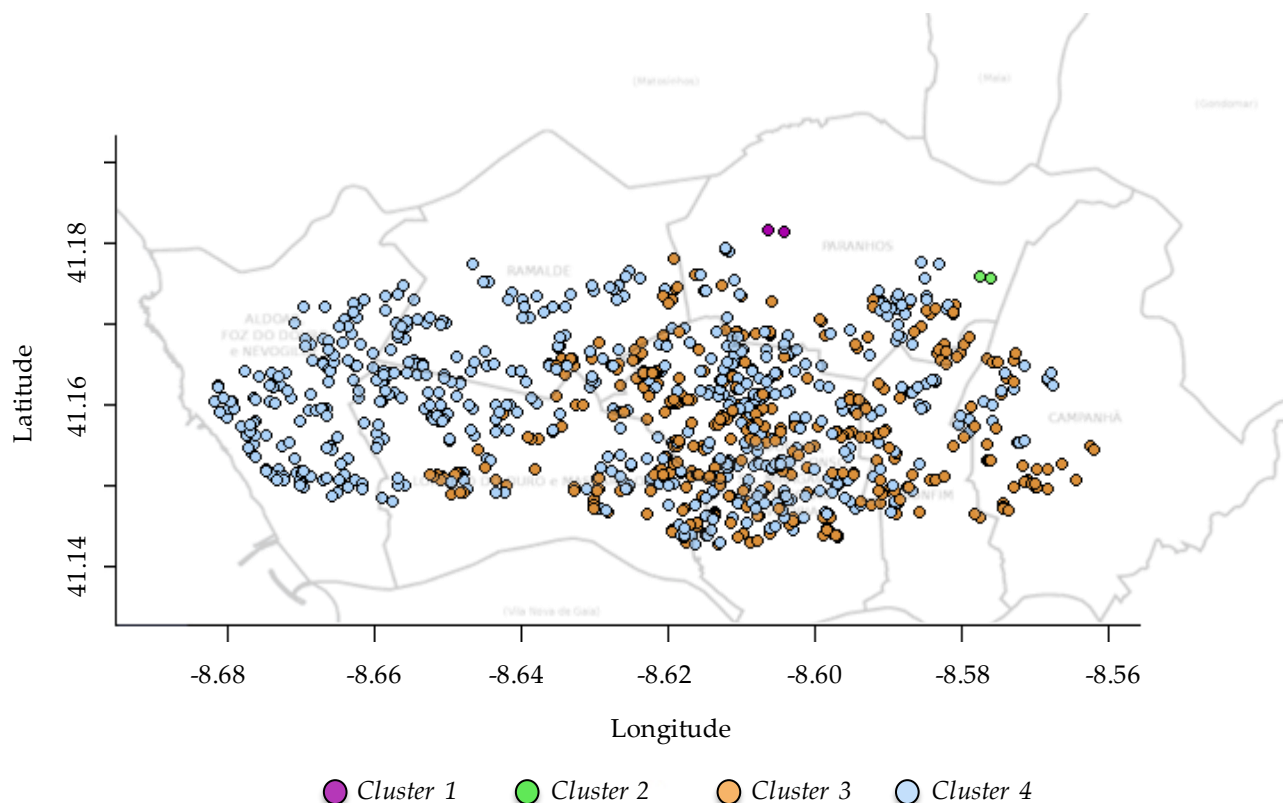
Fonte: elaboração própria

Dominado por prédios em propriedade total com andares, o *cluster* 3 é o que contempla os imóveis mais antigos de entre os restantes submercados. A sua caracterização sugere a prevalência deste submercado em zonas centrais.

O submercado 4 destaca-se por apresentar o valor médio do coeficiente de localização mais elevado e o coeficiente de qualidade conforto elevado, sugerindo que nele se incluem propriedades de luxo ou altamente valorizadas.

Como se pode verificar na figura 4, a dispersão geográfica dos *clusters* construídos para o ano 2022 apresenta, à semelhança de 2019, dois *clusters* bem definidos, contudo em locais diferentes do município do Porto.

Figura 4 - Representação geográfica dos *clusters* para o ano de 2022



Fonte: Elaboração própria

O *cluster* 1 está localizado na freguesia de Paranhos e as coordenadas geográficas médias apontam para as imediações das principais faculdades do polo Asprela. Em semelhança com o ano 2019, esta região procurada sobretudo por jovens estudantes encontra-se bem definida. Em 2022, a delimitação deste submercado torna-se ainda mais vincada dado que este *cluster* é definido apenas por propriedades pertencentes a esta zona.

O *cluster* 2, encontra-se também bem delimitado com localização na freguesia de paranhos, tal como o *cluster* 1. Contudo, este submercado detém imóveis que se encontram na fronteira entre Paranhos e Campanhã numa zona caracteristicamente diferente da definida no *cluster* 1. No que respeita ao tipo de prédio segue a tendência do submercado 1, ao apresentar quase a totalidade de prédios em regime de propriedade horizontal.

O submercado 3 concentra-se expressivamente na zona central, evidenciando maior concentração na união das Freguesias de Cedofeita, Santo Ildefonso, Sé, Miragaia, São Nicolau e Vitória. À imagem do que se verifica em 2019, este *cluster* apresenta um número muito reduzido de prédios em regime de propriedade horizontal, indo de encontro à premissa de que no centro do porto a presença de apartamentos com elevado número de pisos é diminuta. Ainda, é neste submercado onde se verifica a média de idades mais elevadas, sugerindo uma área com um parque habitacional histórico e possivelmente preservado pelo Plano Diretor Municipal.

Por fim, o *cluster* 4 é o que apresenta uma maior dispersão geográfica, no entanto verifica-se que este se encontra presente quase exclusivamente na união das Freguesias de Aldoar, Foz do Douro e Nevogilde. Tal como mencionado acima, o coeficiente médio de localização deste *cluster* é o mais elevado e, para além disso, apresenta um máximo de 3 nesta variável, sugerindo tratar-se de uma zona com elevado valor de mercado imobiliário. A par disto, este submercado agrega propriedades com valores de avaliação extremamente altos ao mesmo

tempo que apresenta um valor de avaliação médio elevado. Estes factos vêm sustentar a presença deste *cluster* numa das zonas mais nobre do município do porto – a zona da Foz.

A tabela 8 descreve sucintamente cada submercado para ambos os anos, permitindo tirar conclusões de uma forma mais concreta e objetiva.

Tabela 8 - Descrição sucinta dos submercados elaborados para ambos os anos

Descrição	
2019	
Submercado 1	Caracterizado por casas térreas antigas em bairros periféricos do Porto.
Submercado 2	Concentração de imóveis no centro histórico do Porto, com predomínio de edifícios baixos de propriedade total.
Submercado 3	Predomina na zona do polo universitário Asprela, com prevalência de apartamentos pequenos.
Submercado 4	Compreende casas grandes e luxuosas, concentradas na zona costeira da Foz do Porto.
2022	
Submercado 1	Totalmente localizado em paranhos foca em apartamentos T0 a T2, circundando o polo Asprela, visado por estudantes.
Submercado 2	Pequeno e com alta variação de valor, indica renovação na área situada da fronteira entre Paranhos e Campanhã.
Submercado 3	Imóveis históricos e prédios de propriedade total no centro do Porto, com baixa presença de apartamentos altos
Submercado 4	Caracteriza-se por imóveis de alto valor na nobre zona da Foz

Fonte: Elaboração própria

Tendo em conta a descrição condensada na tabela 8, não existiram mudança significativas na estrutura do mercado habitacional no Porto de 2019 para 2022, mantendo-se estável a distribuição e características dos submercados.

Esta conclusão revela-se expectável uma vez que a composição das propriedades de habitação numa zona geográfica não sofre alterações drásticas no espaço temporal de três anos.

A grande diferença quando comparados os dois anos, encontra-se no submercado 1 e submercado 2 dos anos 2019 e 2022, respetivamente. De facto,

estes submercados descrevem áreas bem delimitadas, mas diferentes de um ano para o outro. Isto é, cada *cluster* aparece exclusivamente num só ano, evidenciando falta de continuidade ou consistência na delimitação destas áreas ao longo do tempo.

Contrariamente, os restantes submercados delimitam as mesmas áreas em ambos os anos, reforçando a existência de três zonas bem delimitadas no município do Porto.

A primeira referente à zona central do município, onde se verifica a predominância de propriedades em regime de propriedade total, com mais antiguidade e possivelmente preservados.

A segunda situada fora do coração da cidade, mas caracterizada pelo dinamismo do crescimento urbano. Na região que contempla o principal polo da Universidade do Porto, prevalecem prédios em regime de propriedade horizontal e, provavelmente, com mais procura por parte da população estudantil ou por profissionais solteiros.

A terceira zona, situada em grande parte na foz do Porto, segmenta um setor do mercado imobiliário mais luxuoso e exclusivo, caracterizado por propriedades espaçosas, muitas com mais de quatro quartos ou divisões, e um elevado coeficiente de localização.

Capítulo 5

Conclusão

Esta dissertação delimita o mercado de habitação do município do Porto em diferentes submercados. Como base foram utilizados dados fornecidos pelo INE, mais especificamente, dados das habitações que são utilizados para o cálculo do IMI.

A abordagem utilizada passou pela adoção de um método de *clustering* hierárquico com restrições de contiguidade que, para além de ser inovador permite a criação de *clusters* que contêm propriedades contíguas – considerando as questões espaciais. A definição de cada *cluster* revela um submercado com características específicas, passíveis de serem analisadas e, assim, descrever as suas principais características.

Os principais resultados permitem responder às questões de investigação propostas.

Efetivamente, a discussão dos resultados sugere a existência de quatro submercados quer no ano de 2019 quer no ano de 2022. Dos quatro submercados, três deles são comuns aos dois anos: o primeiro, situado na zona central do município; o segundo, detentor das propriedades localizadas na Foz do Porto; e, o terceiro que contempla a zona periférica de Paranhos. Os restantes dois, caracterizam locais diferentes nos dois anos, não dando robustez suficiente para a afirmação destes submercados como pertinentes.

A segunda questão de investigação fica respondida ao analisar a dinâmica ao longo dos anos 2019 e 2022 ilustrando como os submercados evoluem no tempo. De facto, esta evolução é insignificante uma vez que as conclusões ditam que os submercados mais sólidos são definidos em ambos os anos.

A presente delimitação dos submercados serve um vasto grupo de partes de interesse, outrora mencionados na conclusão. Destaca-se o público em geral, os decisores políticos, os credores e investidores, os investigadores da área e os promotores imobiliários como os principais interessados nas conclusões desta dissertação.

Limitações e sugestões para pesquisa futura

A base de dados apresenta algumas limitações, tais como: i) inexistência de características passíveis de analisar vizinhança quer em termos sociais, económicos ou demográficos; ii) número de variáveis pertinentes para o estudo reduzido; iii) não contem dados relativos a preços dos imóveis transacionados.

Mesmo com as limitações descritas acima, a construção de submercados mostrou-se concretizável com resultados importantes para a área.

Não obstante, a delimitação de submercados no município do Porto pode ser mais robusta. Para isso, uma sugestão seria fazer a construção de variáveis que descrevessem mais detalhadamente o espaço. Isto é, calcular a distância dos imóveis a, por exemplo, hospitais, farmácias, escolas, parque da cidade, *CBD*, metro, entre outras.

Finalmente, outra sugestão seria utilizar uma base de dados que refletisse o preço de transação dos imóveis para a construção de uma abordagem de preços hedónicos que viria a complementar, tornando a análise ainda mais sólida e completa.

Bibliografia

- 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>
- Basu, S., & Thibodeau, T. G. (1998). Analysis of Spatial Autocorrelation in House Prices. In *Journal of Real Estate Finance and Economics* (Vol. 17, Issue 1). Kluwer Academic Publishers.
- Bourassa, S. C., Cantoni, E., & Hoesli, M. (2007). Spatial dependence, housing submarkets, and house price prediction. *Journal of Real Estate Finance and Economics*, 35(2), 143–160. <https://doi.org/10.1007/s11146-007-9036-8>
- Bourassa, S. C., Hamelink, F., Hoesli, M., & Macgregor, B. D. (1999). Defining Housing Submarkets*. In *Journal of Housing Economics* (Vol. 8). <http://www.idealibrary.comon>
- Bourassa, S. C., Hoesli, M., & Peng, V. S. (2002). *Do Housing Submarkets Really Matter? * Do Housing Submarkets Really Matter?*
- 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/10.1016/S1051-1377\(03\)00003-2](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 L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding multivariate statistics* (pp. 99–136). American Psychological Association.
- Chin, T. L., & Chau, K. W. (2003). A critical review of literature on the hedonic price model. In *International Journal for Housing and Its Applications* (Vol. 27, Issue 2). <http://ssrn.com/abstract=2073594> Electronic copy available at: <https://ssrn.com>

/abstract=2073594Electroniccopyavailableat:http://ssrn.com/abstract=2073594

- Clark, D. E., Herrin, W. E., & And Herrin, W. E. (2000). *The Impact of Public School Attributes on Home Sale Prices in California*.
https://epublications.marquette.edu/econ_fac
- Dale-Johnson, D. (1982). An Alternative Approach to Housing Market Segmentation Using Hedonic Price Data. In *JOURNAL OF URBAN ECONOMICS* (Vol. 11).
- Farber, S. (1986). Market Segmentation and the Effects on Group Homes for the Handicapped on Residential Property Values. In *Urban Studies* (Vol. 23).
- Frades, I., & Matthiesen, R. (2009). *Overview on Techniques in Cluster Analysis. Bioinformatics Methods in Clinical Research* (R. Matthiesen, Ed.; Vol. 593). Humana Press. <https://doi.org/10.1007/978-1-60327-194-3>
- Goodman, A. C., & Thibodeau, T. G. (1998). Housing Market Segmentation. In *JOURNAL OF HOUSING ECONOMICS* (Vol. 7).
- 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/10.1016/S1051-1377\(03\)00031-7](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/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.
<https://doi.org/10.18637/jss.v103.i07>
- Hennig, C., Meila, M., Murtagh, F., & Rocci, R. (2015). *Handbook of Cluster Analysis* (1st Edition).

- Heyman, A. V., & Manum, B. (2016). *Distances, accessibilities and attractiveness; looking at new approaches to include measures of urban form in hedonic pricing modelling?* <http://joss.bartlett.ucl.ac.uk>
- Hubbard, R., & Allen, S. J. (1987). *An Empirical Comparison of Alternative Methods for Principal Component Extraction* (Vol. 15).
- Ibeas, ángel, Cordera, R., Dell'Olio, L., Coppola, P., & Dominguez, A. (2012). Modelling transport and real-estate values interactions in urban systems. *Journal of Transport Geography*, 24, 370–382. <https://doi.org/10.1016/j.jtrangeo.2012.04.012>
- Keskin, B. (2008). Hedonic analysis of price in the Istanbul housing market. *International Journal of Strategic Property Management*, 12(2), 125–138. <https://doi.org/10.3846/1648-715X.2008.12.125-138>
- 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>
- Kopczewska, K., & Ćwiakowski, P. (2021). Spatio-temporal stability of housing submarkets. Tracking spatial location of clusters of geographically weighted regression estimates of price determinants. *Land Use Policy*, 103. <https://doi.org/10.1016/j.landusepol.2021.105292>
- Legendre, P., & Legendre, L. (1998). Numerical Ecology. In *Elsevier* (2nd English edition).
- Leishman, C. (2001). House building and product differentiation: An hedonic price approach. In *Source: Journal of Housing and the Built Environment* (Vol. 16, Issue 2). <https://about.jstor.org/terms>
- Leishman, C., Costello, G., Rowley, S., & Watkins, C. (2013). The Predictive Performance of Multilevel Models of Housing Sub-markets: A Comparative Analysis. *Urban Studies*, 50(6), 1201–1220. <https://doi.org/10.1177/0042098012466603>

- Marques, J. J. L. (2012). *The notion of space in urban housing markets* [Doutoramento em Ciências Sociais]. Universidade de Aveiro.
- Marques, J. L., Castro, E., Bhattacharjee, A., Batista, P., & Pt, J. (2012). *Spatial heterogeneity across housing sub-markets in an urban area of Portugal*.
- Özyurt, S. (2014). *Spatial dependence in commercial property prices: micro evidence from the Netherlands*. <http://www.ecb>.
- Palm, R. (1978). Spatial Segmentation of the Urban Housing Market. In *Geography* (Vol. 54, Issue 3).
- Rodriguez, M., & Sirmans, C. F. (1994). Quantifying the value of a view in single-family housing markets,. *Appraisal Journal*, 62, 600–603.
- Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. In *Source: Journal of Political Economy* (Vol. 82, Issue 1).
- Rothenberg, J., Galster, G. C., & Butler, R. V. (1993). Book reviews. In *Regional Science and Urban Economics* (Vol. 23). Chicago and London.
- 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). <https://doi.org/10.2478/remav-2022-0009>
- Schnare, A. B., & Struyk, R. J. (1976). Segmentation in Urban Housing Markets. In *JOURNAL OF URBAN ECONOMICS* (Vol. 3).
- 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., Abdullahi, A., Ibrahim, I., Tatari, A., & Polytechnic, A. (2018). Determining House Price for Mass Appraisal Using Multiple Regression Analysis Modelling in Kaduna North, Nigeria Determining House Price for Mass Appraisal Using Multiple Regression Analysis Modeling in Kaduna North, Nigeria 1 2 2. In *ATBU Journal of Environmental Technology* (Vol. 11). <https://www.researchgate.net/publication/327745987>

- Usman, H., Lizam, M., & Burhan, B. (2020). *A Review of Property Attributes Influence in Hedonic Pricing Model*.
- van de Velden, M., Iodice D'Enza, A., & Markos, A. (2018). Distance-based clustering of mixed data. In *Wiley Interdisciplinary Reviews: Computational Statistics* (Vol. 11, Issue 3). Wiley-Blackwell. <https://doi.org/10.1002/wics.1456>
- Watkins, C. A. (2001). The definition and identification of housing submarkets. *Environment and Planning A*, 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/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>
- Yang, L., Wang, B., Zhou, J., & Wang, X. (2018). Walking accessibility and property prices. *Transportation Research Part D: Transport and Environment*, 62, 551–562. <https://doi.org/10.1016/j.trd.2018.04.001>

Apêndices

Apêndice A – Características para cálculo do coeficiente de localização

- Acessibilidades, considerando-se como tais a qualidade e variedade das vias rodoviárias, ferroviárias, fluviais e marítimas;
- Proximidade de equipamentos sociais, designadamente escolas, serviços públicos e comércio;
- Serviços de transportes públicos;
- Localização em zonas de elevado valor de mercado imobiliário.

Apêndice B – Majorativos e Minorativos para cálculo do coeficiente de qualidade e conforto

Majorativos:

- Moradias unifamiliares;
- Localização em condomínio fechado;
- Garagem individual;
- Garagem coletiva;
- Piscina individual;
- Piscina coletiva;
- Campo de ténis;
- Outros equipamentos de lazer;
- Qualidade construtiva;
- Localização excecional;

- Sistema central de climatização;
- Elevadores em edifícios de menos de 4 pisos;
- Localização e operacionalidade relativas;

Minorativos:

- Inexistência de cozinha;
- Inexistência de instalações sanitárias;
- Inexistência de rede pública ou privada de água;
- Inexistência de rede pública ou privada de eletricidade;
- Inexistência de rede pública ou privada de gás;
- Inexistência de rede pública ou privada de esgotos;
- Inexistência de ruas pavimentadas;
- Inexistência de elevador em edifícios com mais de 3 pisos;
- Existência de áreas inferiores às regulamentares;
- Estado deficiente de conservação.

Apêndice C – Script em R

```
#Read data from Text
dados2019 <- read.table ("IMI_PA_2019.txt",header = TRUE,sep="|")
summary(as.numeric(dados2019$FREGUESIA))
PortoInd <- which(substring(dados2019$FREGUESIA,1,4)=="1312")
Porto2019 <- dados2019[PortoInd,]
PHabInd <- which(Porto2019$AFECTACAO==3)
Porto2019H <- Porto2019[PHabInd,]
names(meanPLcoodCP)[1] <- "CODIGOPOSTAL"
Porto2019H <- merge(Porto2019H,meanPLcoodCP)
IndV <- c(11,13,17,18,19,20,26,31,39,40,41,44,46,77,78)
```

```

SmallPorto2019 <- Porto2019H[,IndV]
SmallPorto2019$TIPOPROPRIETARIO <-
as.factor(SmallPorto2019$TIPOPROPRIETARIO)
SmallPorto2019$TIPOPREDIO <- as.factor(SmallPorto2019$TIPOPREDIO)
SmallPorto2019$TIPOLOGIA <- as.factor(SmallPorto2019$TIPOLOGIA)

SmallPorto2019$TIPOPROPRIETARIO <-
as.factor(ifelse(SmallPorto2019$TIPOPROPRIETARIO==1,1,
ifelse(SmallPorto2019$TIPOPROPRIETARIO==2,2,"Outros")))
IndOBS_ValoresFalta <- which(apply(SmallPorto2019[,-3],1,function(x)
any(is.na(x))))
if (length(IndOBS_ValoresFalta) > 0) SmallPorto2019 <- SmallPorto2019[,-
IndOBS_ValoresFalta,]
SmallPorto2019$N_PISOS_FRACCAO <-
ifelse(is.na(SmallPorto2019$N_PISOS_FRACCAO),0,
SmallPorto2019$N_PISOS_FRACCAO)
#Agrupa as observações
SmallPorto2019$NewTip <-
as.factor(ifelse(SmallPorto2019$TIPOLOGIA==1,1,
ifelse(SmallPorto2019$TIPOLOGIA=="01",1,
ifelse(SmallPorto2019$TIPOLOGIA=="T0","T0",
ifelse(SmallPorto2019$TIPOLOGIA=="T1","T1",
ifelse(SmallPorto2019$TIPOLOGIA=="T2","T2",
ifelse(SmallPorto2019$TIPOLOGIA=="T3","T3",
ifelse(SmallPorto2019$TIPOLOGIA=="T4","T4","Outros"))))))))
#conversão coordenadas library(sf)
XY_sfP <- st_as_sf(SmallPorto2019, coords = c("COORDX", "COORDY"), agr =
"constant", crs="EPSG:20790")

```

```
#Note: EPSG:20790 is the Portuguese National Grid reference coordinate system (Lisbon)
```

```
XY_sfP1 <- st_transform(XY_sfP,"EPSG:4326")
```

```
#Note: EPSG:4326 seems to be the standard latitude and longitude coordinate system
```

```
XYP_DF <- as.data.frame(st_coordinates(XY_sfP1)) colnames(XYP_DF) <- c("Longitude","Latitude")
```

```
SmallPorto2019$LongitudeT <- XYP_DF$Longitude  
SmallPorto2019$LatitudeT <- XYP_DF$Latitude
```

```
# Hierarchical clustering mix variables
```

```
library(mclust) library(cluster) library(FD) library(adespatial)
```

```
DPortoMix <- daisy(SmallPorto2019[,c(7,9,10,14,15,17,18)]) tirar a Tipologia
```

```
#Getting a list of neighbours, assuming any two points within 500m distance are neighbors
```

```
library(spdep)
```

```
nghblistP <- dnearneigh (SmallPorto2019[,c("LongitudeT","LatitudeT")], d1=0, d2=0.5, longlat = TRUE)
```

```
nghblistP <- listw2sn(nb2listw(nghblistP))[1:2]
```

```
# Space constrained hierarchical clustering
```

```
#Add coordx and coordy information to individual records and the use constr.hclust #instead of hclust and kmeans
```

```
PortoClustersR <- constr.hclust(DPortoMix, links=nghblistP, method="ward.D2", coords=SmallPorto2019[,c("LongitudeT","LatitudeT")])
```

```
PortoClustersR0 <- PortoClustersR # Creating a copy of the cluster results
class(PortoClustersR0) <- "hclust" # as an "hclust" object (so that method plot
displays a dendogram)
```

```
plot(PortoClustersR0) # Displaying the dendogram
```

```
heighdif <- diff(PortoClustersR0$height) # Finding the vertical (height)
differences in the dendogram
```

```
sortedheighdif <- sort(heighdif,decreasing=TRUE,index.return=TRUE)
head(sortedheighdif$x) # largest vertical differences
```

```
head(sortedheighdif$ix) # at steps (clusters = n - steps)
```

```
k <- 3 # Assuming 4 clusters
```

```
clesteleP <- rect.hclust(PortoClustersR,k=k)
```

```
my_colors<-c("#1985a1","#191970","#a7c7e7", "#00B3ff","#aec3b0",
"#aea391","#a7c7e7","#87ceeb")
```

```
plot(PortoClustersR,k=k,col=my_colors) # now use plot for an constr.hclust
object, that requires a k and displays a map
```

```
#Displaying summary statistics for each cluster
```

```
for (c in 1:k) {cat("Summary statistics for cluster",
c,"with",length(clesteleP[[c]]),"elements\n")
```

```
print(summary(SmallPorto2019[clesteleP[[c]],)) }
```

```
save(clesteleP,file="ResPorto2019F.RData")
```

```
load("ResPorto2019F.RData")
```