

Article

Geographic Information System-Based Approach and Analysis of Housing Mismatches in Searches during Housing Selection

To Quyên Lê ^{1,*} and Le-Vinh-Lam Doan ²¹ Faculty of Civil Engineering, Ton Duc Thang University, Ho Chi Minh City 700000, Vietnam² Independent Researcher, Canberra 2614, Australia; lamlvdoan@gmail.com

* Correspondence: letoquyen@tdtu.edu.vn

Abstract: This study aimed to explore the spatial interaction between search flow and house price levels. It utilized migration flow data from 2011 and house price data from the same year in Greater Manchester, United Kingdom. The methodology included Chi-square tests and a series of visual maps to explore and visualize the relationship between current areas and search areas. The study examined five house price ranges with matches and mismatches between search patterns and current patterns and tried to illuminate the underlying causes. The research results could prove beneficial for policymakers and planners in discerning distinct migratory patterns and geographical regions, thereby facilitating the formulation of well-informed decisions.

Keywords: applied GIS; housing market; house price; housing search; housing mismatch

1. Introduction

In recent years, the rise of online platforms to find accommodation has revolutionized the way people search for housing, making it easier and more convenient than ever before. This shift towards housing property portals has led to new chances to analyze search activity. The process of searching for housing is intricate and encompasses a range of criteria, including preferences in location, budget constraints, and desired amenities [1].

The literature is increasingly examining the methods by which households seek housing, and a significant portion of this literature is centered on a search and matching model that is strongly expressed in mathematics to elucidate the behavior of housing markets with trading frictions [2–9].

This study differs from the past literature by using a geographical approach and including user-generated data, rather than analyzing housing search behavior just through a search model. To obtain a comprehensive overview of the significant literature in this field, it is recommended to consult the works of Maclennan in 1982 [10] as a starting point. More recent publications offer a geographical approach to give valuable insights into user-generated data for housing search behavior by analyzing spatial patterns and search pressure locations, aiding in understanding housing market dynamics and user preferences [11–13]. However, it should be noted that these existing sources are inadequate. Detecting the geographical dynamics of house prices utilizing advanced big data technologies from the property site could also aid in the exploration of trends, variations, and patterns at the national and subnational levels [14].

Therefore, we attempt to fill this important gap in the literature. This study aims to investigate the spatial interaction between search and migration flows and house price levels in Greater Manchester, UK, using data from 2011. Specific research questions include the following:

- How do migration flows correspond with house price levels across different geographical areas?
- What are the patterns of match and mismatch between areas of property search interest and current residence based on house price ranges?



Citation: Lê, T.Q.; Doan, L.-V.-L. Geographic Information System-Based Approach and Analysis of Housing Mismatches in Searches during Housing Selection. *Buildings* **2024**, *14*, 2232. <https://doi.org/10.3390/buildings14072232>

Academic Editors: Jie Chen, Zisheng Song and Wei Wang

Received: 16 June 2024

Revised: 16 July 2024

Accepted: 17 July 2024

Published: 19 July 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

The objectives of the paper include the following: (1) to explore the relationship between a household's current location and their search location based on house price levels; (2) to visualize the search flows between current and search locations based on house price ranges; and (3) to identify the match and mismatch between search patterns and supply patterns for particular price levels and then understand the cause.

This paper presents a case study of the housing market search extent in the United Kingdom (UK), based on posts from the housing property portal (<https://www.rightmove.co.uk/>, accessed on 2 June 2023) and house price data from the Land Registry for transactions in 2014.

The study used some simple statistical techniques, such as the K-means method, to group house prices, calculating average property prices for each merged ward based on property numbers and transactions. Furthermore, the Tukey post hoc test and the chi-square test were used to investigate cluster variation and the relationship between price levels in the current location and search locations. Parallel to quantitative analysis, the paper provides GIS maps to visually display search flows between current locations and search locations based on house price levels. The maps depict search patterns for people staying in areas with different price ranges. Quantitative statistics in SPSS and a series of maps created with QGIS software v.3.12.3 helped achieve this.

The rest of this paper is organized as follows: Section 2 presents the literature review, which focuses on real estate portals, GIS-based statistical methods, and housing price research in relation to housing pattern research. Section 3 explains the data and methods used in this study. Section 4 presents the results that align with the three research objectives. The results show the five clusters of house price ranges and a series of maps to visualize matches and mismatches between search patterns and current patterns. Finally, in Section 5, conclusions and discussions are drawn by reflecting on the benefits to planners and designers of understanding housing search activity in the rapid development of the digital age.

2. Related Literature Review

In this context of "real estate portals," users can buy, rent, or lease real estate properties via an online platform. All the options for them include the option to log in, register, browse for and explore properties, select a location, reserve a home, cancel reservations, access web services, and send emails. Now that real estate portals have spread to millions of buyers, what user-generated data from real estate portals have been left behind? Because many real estate companies provide their services online, the knowledge gained from such user-generated data may benefit these businesses. This has resulted in new user-generated data that can be utilized to analyze online search tactics and conduct research. The utilization of big data in the real estate sector has been greatly advanced by the empirical investigations carried out by property portals.

Digital real estate platforms gather standard data and use it to rate, sort, and measure people, places, and markets, which leads to classification situations [11]. The purpose of a spatial database is to aid in subject-oriented decision-making, and spatial data serve as the foundation for real estate appraisal and play an important role in determining housing prices [15]. Knowles stated that "we need to develop new graphic models to represent the complexity of social space" [16]. Furthermore, an investigation was conducted using a Geographic Information System (GIS) to study the relationship between housing prices and road density [17]. In short, though comparatively limited to the use of GIS-based methods in housing search patterns, these studies were significantly important in showing the potential of spatial analysis to visually examine movement patterns and trends.

Over the past decade, the applicable housing search data have utilized Google Trend, and now, the scope of study and novel data have extended. The area of studies using Google Trend has significant implications in these areas of business management and economics in practical, academic, and political fields. The most significant applications of big data have been mentioned in studies on the social change in housing market indicators

and prediction [18–20]. Another aspect of housing research has included a two-stage model of sectoral housing and an individual dwelling from the submarket [21], as well as behavioral factors influencing household housing choices [22,23]. The most cited studies on the analysis of housing preference patterns through search activities and the search process are based on search behavior [11]. The studies above embraced a novel approach to understanding search behavior and the search model in the decision-making process of households aligned with the submarket.

Regarding big data from the housing search portals, it can be seen as driving demand for search outcome market data. For example, analyzing house price and migration flow data can reveal market segments generated by racial segregation [24] or wage dispersion [25]. In an equilibrium market, price should be a crude measure of market segmentation. Due to their widespread use and ability to access information at a low cost, almost instantly, and at a fine-grained level of disaggregation, search engines and related information technologies are moving further away from traditional economy and business forecasting. Such real estate online data benefit participants in the industry as well as researchers for migration studies. The research on housing prices at the search portal focuses on the trends in price fluctuations within the context of big data, offering insights and methods for analyzing these data effectively [17]. Big data also contributed to research on how to avoid the negative effects of political and economic biases by focusing on the differences between urban and regional economies. The extent to which housing price variation could be explained by location has not been empirically examined using big data. In general, the movement of big data housing search portals has been helpful with predictions from social science research and shows the meaning of data in spatial linkages between migration flows and other social and economic factors. The existing literature has tended to focus on outcome indicators such as price changes and migration patterns; however, the understanding of the housing search process is inadequate. There is a mismatch between what households want and what sellers provide.

In a nutshell, this literature review is generally focused on the following three points: (1) real estate portals in practice; (2) GIS-based statistical methods that examine movement patterns and trends; and (3) housing price in relation to housing pattern research. It reveals both constraints and possibilities, while there is a possibility for us to combine the three. In line with this, this paper intends to look closely at the spatial patterns at the local level and contribute a visual approach to harness the power of big datasets including search patterns and house prices.

3. Data and Methods

3.1. Data

This study utilized the house price data from the Land Registry for the year 2014. Data collection involved gathering migration flow data and house price data from 2011 for Greater Manchester. Migration flows were likely derived from census or administrative records, while house price data may have been sourced from government records or property listings. It encompassed a total of 27,601 properties that had been sold, each associated with a unique postcode. The GIS map displayed the property postcodes in latitude and longitude coordinates within the boundaries of 215 amalgamated wards located in the Greater Manchester region.

To collect data from Rightmove Plc ([Rightmove.com](https://www.rightmove.com), accessed on 15 June 2024), we typically secure permissions for data usage and access. We then extract relevant property data, including price, location, views, and other specifics, adhering to the platform's policies. After extraction, data undergo cleaning to ensure accuracy and consistency, followed by integration with additional datasets if needed. The processed data are stored securely, complying with data protection regulations. We analyze the data using statistical and spatial methods to uncover insights into property market trends and search patterns, aiding urban planning and policy decisions. Then, spatial data analysis techniques are applied to map and analyze the relationship between migration flows and house prices.

Greater Manchester, a metropolitan county in North West England, encompasses diverse boroughs and districts each comprising multiple wards. Mapping these districts is crucial to visualizing where wards with higher or lower online search metrics are situated, revealing potential patterns influenced by socioeconomic factors, urbanization levels, and local market conditions. Districts like the City of Manchester, Bolton, Bury, Oldham, Rochdale, Salford, Stockport, Tameside, Trafford, and Wigan each exhibit distinct characteristics that affect property market dynamics. Analyzing regional variances helps identify clusters of wards with significant real estate activity or subdued market interest within each district. Understanding these socioeconomic factors and contrasting urban, suburban, and rural dynamics enhances the interpretation of online search behavior and property market trends across Greater Manchester.

The data included property price, postcode, latitude and longitude, and, importantly, detailed views, including 35,101 properties listed on the Rightmove website collected in March 2013. Each property had a number of detailed views showing how many searchers were interested in the property. The data included key metrics such as listed property count, total detailed views, average detailed views per property, sold property count, total property value, average property value, and maximum and minimum property values. We calculated the attributes regarding the online views and the property sales for each merged ward, as shown in Tables 1 and 2 below.

Table 1. The merged wards with the highest total detailed views.

Name	Listed Property Count	Total Detailed View	Average Detailed View	Sold Property Count	Total Property Value	Average Property Value	Maximum Property Value	Minimum Property Value
City Center	509	3705	7.3	322	58,055,158	180,296	1,050,000	61,500
Bowdon	296	3366	11.4	131	78,633,435	600,255	8,190,000	105,000
Didsbury West	235	2821	12.0	198	53,109,098	268,228	1,050,000	17,000
Hale Barns	211	2750	13.0	98	47,558,989	485,296	3,650,000	108,000
Hale Central	231	2689	11.6	161	66,008,528	409,991	2,650,000	48,000
Bramhall North	264	2483	9.4	144	42,926,695	298,102	830,000	85,000
Brooklands	323	2387	7.4	227	44,871,862	197,673	775,000	28,000
Brooklands	323	2387	7.4	227	44,871,862	197,673	775,000	28,000
Marple North	254	2176	8.6	163	42,418,611	260,237	725,000	90,000
St Mary's	398	2158	5.4	268	46,701,665	174,260	672,000	20,000
St Mary's	398	2158	5.4	268	46,701,665	174,260	672,000	20,000
St Mary's	398	2158	5.4	268	46,701,665	174,260	672,000	20,000
Didsbury East	177	2053	11.6	191	48,758,798	255,282	875,000	87,500
Altrincham	199	2008	10.1	182	41,798,500	229,662	850,000	38,500
Bramhall South	207	1950	9.4	164	53,114,595	323,869	1,100,000	100,000

3.2. Method

3.2.1. The Levels of House Price

To group house prices, we embraced the K-means method, which was previously used in studies of housing and urban planning [26–29]. To conduct the K-means clustering, the average property price for each merged ward was calculated in reference to property numbers and transactions in each ward. With the use of SPSS software, we selected the value of K from 2 to 15 and then chose the optimum option. The optimum option is the one

where the overall within-cluster variation (W) is the lowest and where the Tukey post hoc test is significant. These formulas can be referenced in the works of some authors [29,30].

$$W = (V_{k+1} - V_k) - (V_k - V_{k-1})$$

The values of V_{k-1} , V_k , and V_{k+1} are known as the variance ratios and are taken from the F-values of the one-way ANOVA method in SPSS.

Table 2. The merged wards with the lowest total detailed views.

Name	Listed Property Count	Total Detailed View	Average Detailed View	Sold Property Count	Total Property Value	Average Property Value	Maximum Property Value	Minimum Property Value
Milkstone and Deeplish	71	78	1.1	46	3,336,600	72,535	150,000	15,000
Central Rochdale	69	94	1.4	45	3,434,620	76,325	199,995	20,000
Coldhurst	64	102	1.6	37	2,851,950	77,080	200,000	32,000
Rumworth	62	115	1.9	71	5,506,535	77,557	249,950	28,560
Alexandra	91	144	1.6	64	4,734,400	73,975	250,000	25,000
Kingsway	145	174	1.2	75	8,382,780	111,770	250,000	30,000
Werneth	81	187	2.3	59	5,450,550	92,382	305,650	25,000
Ardwick	59	235	4.0	66	8,431,720	127,753	189,995	40,950
Broughton	86	243	2.8	94	11,359,199	120,843	360,000	43,000
Great Lever	88	249	2.8	130	9,105,994	70,046	245,000	21,000
West Middleton	113	251	2.2	53	5,222,325	98,534	200,000	34,000
Waterhead	175	287	1.6	68	5,800,595	85,303	200,000	29,000
Langworthy	85	288	3.4	60	5,193,724	86,562	240,000	40,000
Medlock Vale	123	290	2.4	71	6,767,650	95,319	400,000	30,000
St Peter's	98	290	3.0	68	5,906,850	86,865	185,000	32,000

After the calculations based on SPSS software, in this paper, we chose the value $K = 5$ for the case of house price, as this resulted in the minimum within-cluster variation value (W) and the satisfaction of the Tukey post hoc test. The analysis results are shown in Tables 3 and 4 below, as well as the five groups of the 2014 house prices for 215 merged wards in Greater Manchester.

3.2.2. Chi2 Analysis

The research performed the Chi-square test of the relationship between price levels in the current location and price levels in search locations on SPSS 20. This Chi-square analysis is widely used in housing studies [31–33]. Chi-square tests are appropriate here as they assess the association between categorical variables (e.g., migration flows categorized by origin and destination areas and house price ranges) and provide insights into patterns of match and mismatch.

In the Chi-square test, it is widely accepted that if the Asymptotic Significance (2-sided) value of the Pearson Chi-square is less than 0.05, the research rejects the hypothesis H_0 . This means that the two variables are related. If the Asymptotic Significance (2-sided) value is greater than 0.05, the null hypothesis H_0 is accepted. This means that the two variables have no relationship with each other.

Table 3. Clusters, Wk, and Tukey HSD.

Cluster Number (k)	VRC	Wk	Turkey HSD *	Optimum Solution **
2	185.52	NA	x	x
3	440.325	−180.341	x	x
4	514.789	−96.2929	x	x
5	492.96	−351.938	ok	ok
6	519.113	297.6617	ok	x
7	842.928	47.9817	x	x
8	814.805	92.88874	ok	x
9	879.571	165.4012	ok	x
10	1109.737	−25.241	x	x
11	1314.663	−259.518	x	x
12	1260.071	562.4933	ok	x
13	1767.972	−28.482	x	x
14	2247.391	−425.589	x	x
15	2301.221	NA	x	x

* Significant at 0.05 level. ** Combined based on the Turkey post hoc test result and the minimization of Wk.

Table 4. Five clusters of house prices.

Cluster Name	Ward Units	Price Range (GBP)
Significantly low price	73 ward units	63,000–110,000
Low price	79 ward units	110,000–151,000
Medium price	41 ward units	151,000–210,000
High price	18 ward units	210,000–290,000
Significantly high price	4 ward units	290,000–630,000
Total	215 ward units	

3.2.3. GIS Visualization

The overlay feature in GIS was used in this article to visually analyze the spatial search flows. The study displayed a series of maps for combined wards that included data on property price levels and search flows. Visual mapping techniques complemented the statistical analysis by spatially illustrating these patterns, enhancing understanding and interpretation.

4. Results

In this section, we first provide the analysis related to the relationship between where people were currently staying and where they searched for housing in terms of house price level. We embrace the argument that households who are currently staying in a neighborhood associated with high house prices are likely to search for houses in areas with similar or higher price levels.

Then, we examine the relationship between the areas of search polygons (SQKMs) and price levels in search locations. We support the idea that people who are looking for houses in lower price ranges may conduct wider search areas.

4.1. Relationship between Current Areas and Search Areas

The paper used SPSS software to provide the cross-tabulation analysis below to explore the association between where people were currently staying and where people searched for houses. Table 5 and Figure 1 show the price level in the search location. Table 6 shows the Chi-square tests regarding price levels in current areas and price levels in search areas.

The value Asymp. Sig of 0.000 (p -value < 0.01) indicated that there was a significantly positive association between the price levels in the areas where people were currently staying and in the destination areas where people were looking for houses.

Table 5. Cross-tabulation between price levels in current areas and price levels in search areas.

		The Price Level in the Search Location					Total
		Significantly High	High	Average	Low	Significantly Low	
The price level in the current location	Significantly high	57	0	0	0	0	57
	High	57	190	63	18	0	328
	Average	35	237	134	200	35	641
	Low	50	244	204	379	14	891
	Significantly low	4	104	272	230	406	1016
Total		203	775	673	827	455	2933

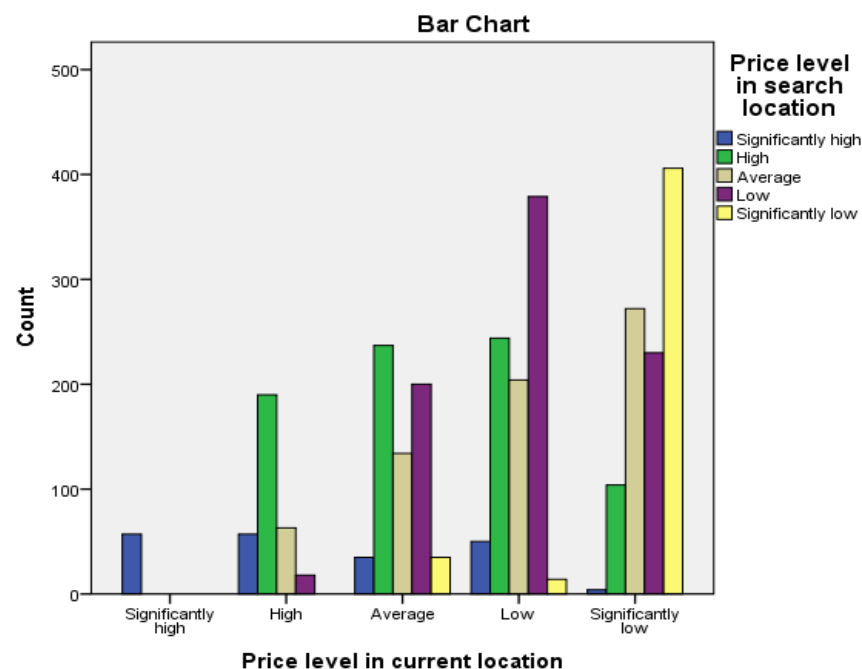


Figure 1. Relationship between price level in current location and price level in search location.

Table 6. Chi-square tests regarding price level in current location and price level in search location.

	Value	df	Asymp. Sig. (2-Sided)
Pearson Chi-Square	1879.881 ^a	16	0.000
Likelihood Ratio	1485.577	16	0.000
Linear-by-Linear Association	804.966	1	0.000
N of Valid Cases	2933		

^a One cell (4.0%) has an expected count less than 5. The minimum expected count is 3.95.

The current literature on housing studies shows that people move to areas where house prices reflect their social and economic background, and there is a connection between advantaged and disadvantaged areas [34]. However, there is a lack of research exploring the connection between advantaged and disadvantaged areas based on search flows, and this paper is expected to shed light on this matter given the importance of the housing

search process. During the search process, households may change their preferences, and this piece of research could provide more insight into their initial search preferences seen as their desired preferences. This could be compared with the outcome preference to figure out the search constraints or restrictions due to the supply side.

4.2. Search Flows between Current and Search Locations Based on House Price Levels

The previous section established the positive relationship based on house price levels between where people were currently staying and where they were looking for houses. In this section, the paper provides a series of GIS maps to visually display search flows between current locations and search locations based on house price levels. The following patterns depicted the search patterns of people who stayed in areas with five house price ranges: significantly low, low, medium, high, and significantly high.

In terms of the significantly high price level, Figure 2 reveals that Bowdon appeared to be the most attractive area among searchers who stayed in this price range. Most searchers who were looking for highly expensive properties were from inside Bowdon.

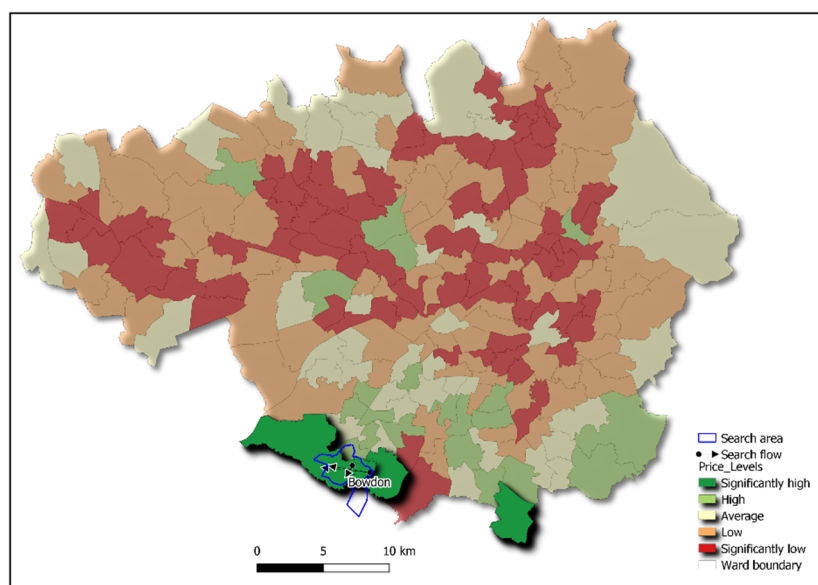


Figure 2. Where people in significantly high-price areas searched for homes.

Regarding the high price range, Figure 3 indicates that households who stayed in the areas associated with this price level dominantly searched for homes in the areas with high prices such as Priory, Brooklands, Didsbury West, Didsbury East, Chorlton, Cheadle and Gatley, and Cheadle Hulme South. Furthermore, this figure serves to broaden the mismatch of search flow; for example, searchers from high-price areas (for example, Heaton North, Priory) looked for properties in significantly high-price (Bowdon) or medium- and low-price areas (for example, Urmston).

Figure 4 illustrates the search flows from searchers in medium-price areas, with the majority of searches in the City Center and the southern areas. There was a starting point to express the mismatch of search areas from people staying in areas with average prices to areas with high and significantly high prices (St Mary's, Altrincham, Bramhall North).

Figure 5 presents where people in the low price range searched for houses with the greatest search intensity. A high proportion of searches was found in most of the areas within this submarket. There were a number of searchers who looked for houses in high- and significantly high-price areas such as Timperley and Bramhall North or properties in low-price areas such as Dukinfield.

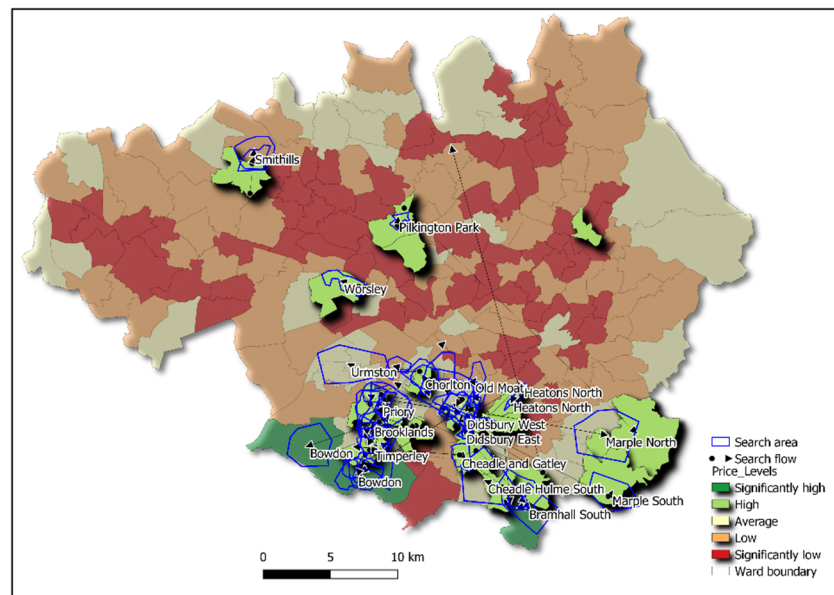


Figure 3. Where people in high-price areas searched for homes.

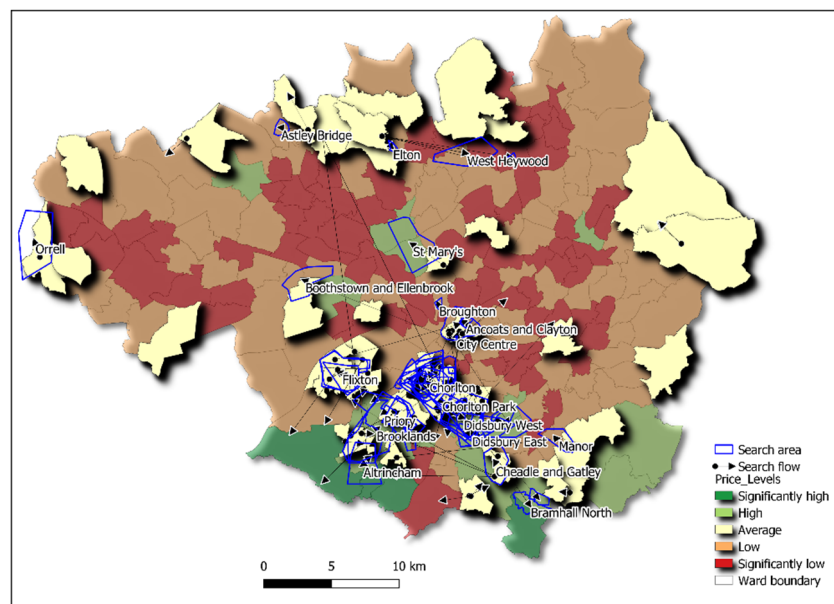


Figure 4. Where people in medium-price areas searched for homes.

For searchers who stayed in the significantly low price range in Figure 6, there was a significant mismatch between search patterns and current stay patterns. There were many searches outside of this submarket, primarily in the south and City Center, such as Broadheath, Stepping Hill, Bredbury Green, Romiley, and Lowton East.

It is worth noting that several search flows initiating from one location mean that individuals staying in one location searched for houses in multiple locations, for example, Royton North (significantly low price). This may show how actively these individuals searched for housing.

In short, the analysis of the maps above indicated that there was an association between where people were currently staying and where people searched for new housing in terms of house price levels. The levels of house prices in current locations seemed to have a relationship with the levels of house prices in the search areas.

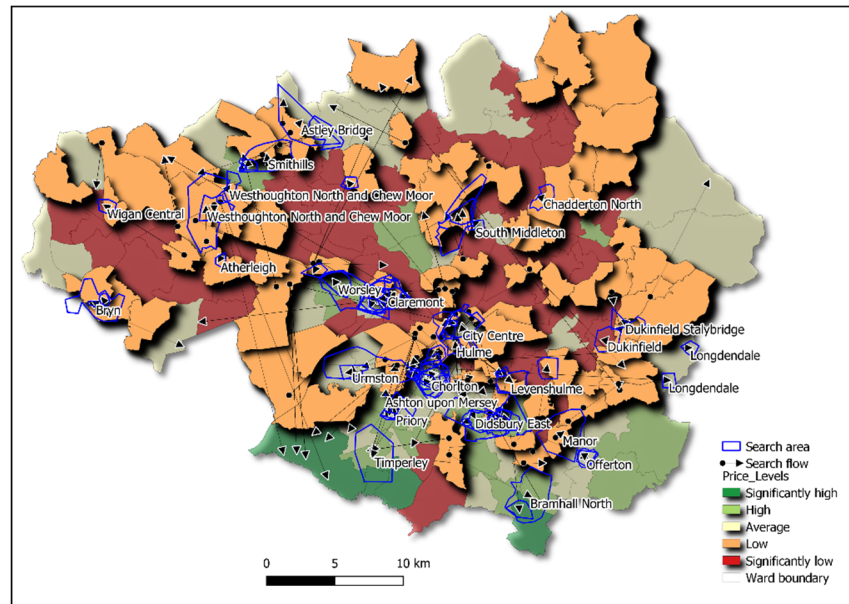


Figure 5. Where people in low-price areas searched for homes.

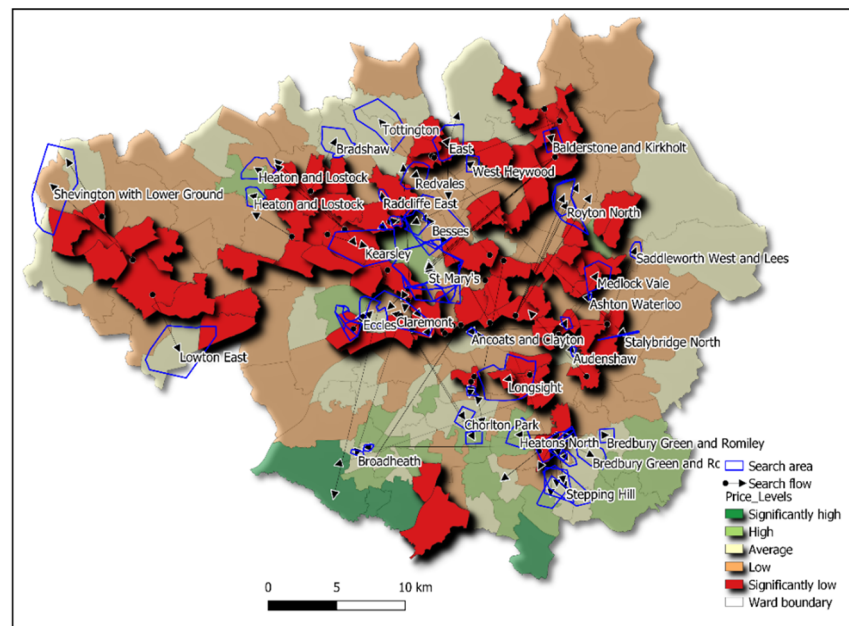


Figure 6. Where people in significantly low-price areas searched for homes.

4.3. Search Mismatch

This paper created a series of maps showing the search flows and search polygons based on the maximum search price representing the latent demand side and the current five house price levels representing the supply side. From the maps, we examined the mismatch between the search patterns and the supply patterns for particular price levels.

Figure 7 shows the search patterns for house prices below 100 K compared with the current supply patterns for this price level (the areas in red). Mismatch occurred in the south of the city, where most search patterns associated with a maximum price of 100 K focused on the significantly high- and high-price areas such as Bowdon, Priors, and Stepping Hill. Mismatch was also observed in Lowton East, Bryn, and Longendale, where people searched for low-price houses with a maximum of 100 K in the low- or average-price areas (100 K–150 K or 150–200 K).

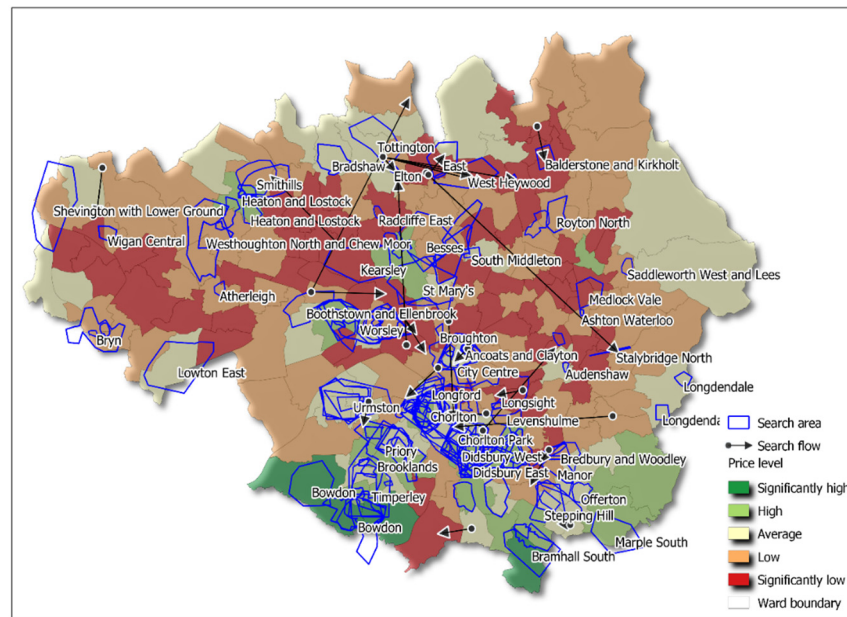


Figure 7. Mismatch of searching price below 100 K.

In the low-price range of 100 K–150 K (Figure 8), many searchers looked for houses in high-price areas where house prices started from 200 K, such as Brooklands, Cheadle Hulme South, and Aston upon Mersey. They also searched for houses in average-price areas (Longdendale, Orell) or significantly low-price areas (Claremont or West Heywood).

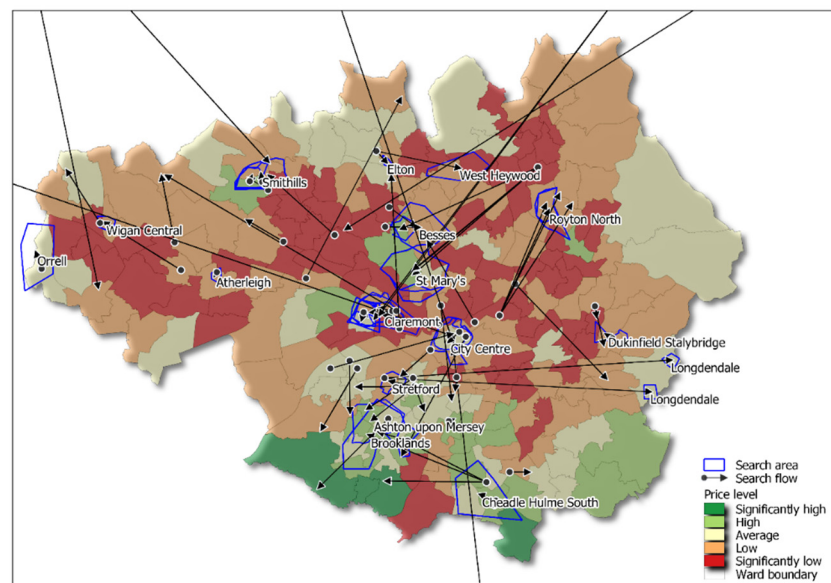


Figure 8. Mismatch of searching price of 100 K–150 K.

Figure 9 illustrates the search patterns for the medium price range of 150 K–200 K with dominant search polygons and search flows in high-end price areas. Although there was not a mismatch in Flixton and Bradshaw, a mismatch occurred in the high-price areas (Priority, Didsbury East). The figure also shows several searchers who came from other areas outside the city.

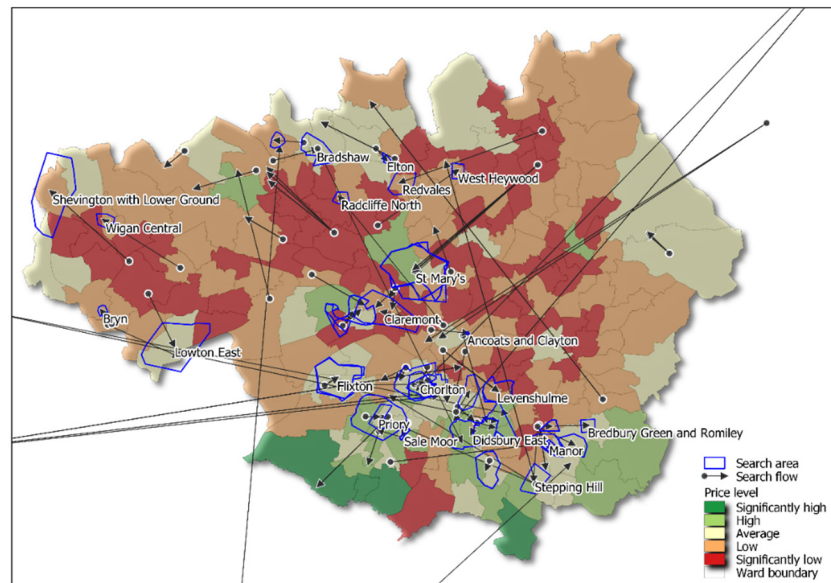


Figure 9. Mismatch of searching price of 150 K–200 K.

In terms of searching for houses in the price range of 200 K–280 K, Figure 10 shows that most searchers looked at the right locations, for example, Longford, Chorlton, Didsbury East, Priory, and Brooklands. However, people searched for houses in the wrong areas where there was no suitable supply in their budget (Holyrood, Dukinfield, Radcliffe North, Tottington, Astley Bridge). A number of searchers from other areas outside the city also looked for houses in Greater Manchester, particularly in the south, such as Bramhall North or Priory.

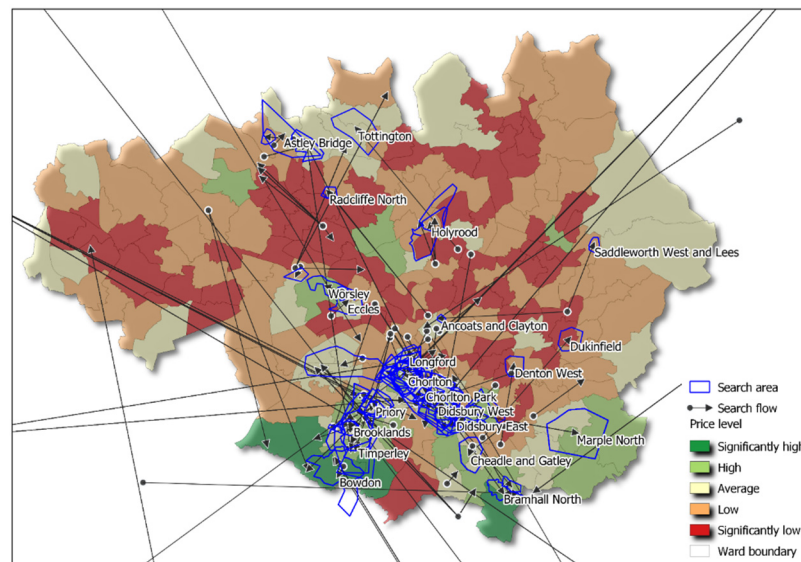


Figure 10. Mismatch of high searching price of 200 K–280 K.

Figure 11 shows the search patterns for price ranges above 280 K, with the dominant searches conducted in the wrong places in the south (Chorlton, Chorlton Park, Heaton South, Urmston, Bramhall North). Although, Bowdon saw a number of searchers who looked for the right place, several areas with the significantly low- and low-price range witnessed mismatches, such as Castleton or Smithills. The figure also indicates that several searchers outside the city also looked for properties in Bramhall North, Chorlton, and Chorlton Park.

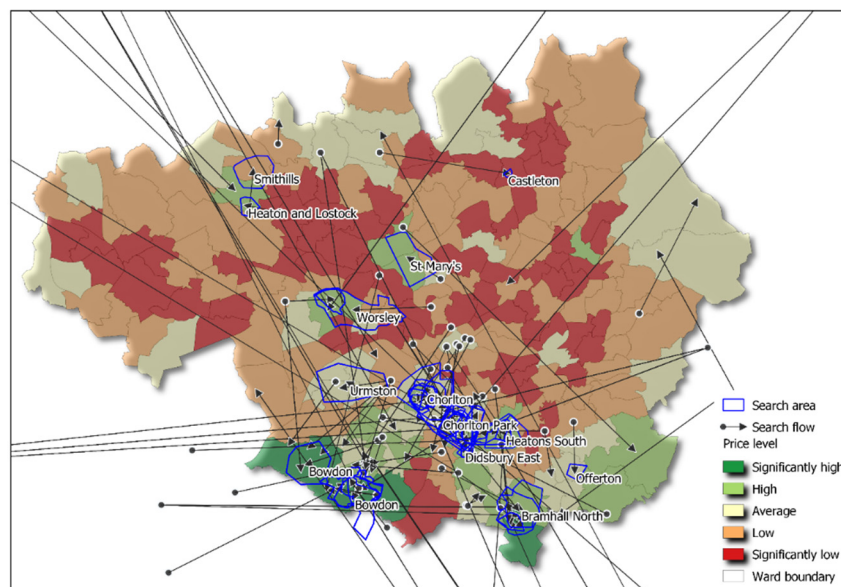


Figure 11. Mismatch of significantly high searching price above 280 K.

In short, there existed a spatial mismatch between people's searches for homes based on their current budget and the price of destination areas. This implied that these searchers had obtained a lack of market information in their search areas and demonstrated market restrictions in terms of existing housing supply.

5. Conclusions and Discussions

5.1. Key Findings

This paper investigated the relationship between five house price levels in current areas and price levels in search areas using a series of GIS maps. In our research, we focus on the origin–destination of search flow data, one of the geographic mobility metrics that concerns current locations and search locations based on house price levels. Search flow maps display in-flow and out-flow, and, together with SPSS techniques, can directly link locations and show spatial search patterns and their distributions, i.e., flow clusters of different locations and major flow patterns.

In regard to comparable house prices and locations, the paper also found evidence of a match and mismatch between search patterns and current patterns. For example, people are looking for a house with a fixed price range of 150 K, but the areas they are looking for have house prices higher than that, for example, 200 K, 250 K, or even 280 K.

Additionally, the results for exploring the match and mismatch of housing distribution in terms of housing price from users' searches on property portals were presented and applied to visualize search flows in the UK. We discovered that the dynamics of property price ranges were extremely relevant to their locations. For example, people who live in the high-price range will base their housing search in the future in the same place.

Beyond the defined findings, this study significantly contributes to the urban planning literature by integrating migration flow data, housing search patterns, and spatial analysis techniques like GIS mapping and statistical methods like Chi-square and SPSS. It elucidates spatial interactions between house price levels and housing search behaviors in Greater Manchester, UK, providing nuanced insights into how individuals' housing preferences manifest geographically. This research aids in identifying spatial clusters and submarkets within urban regions, predicting future housing demand, optimizing resource allocation, and formulating evidence-based housing policies.

Unlike many studies focusing on a singular method in studying housing market dynamics, for instance, statistical equilibrium models [35], spatial structure models [36], an index-based speculative frame [37], etc., this study employs a holistic approach to provide a comprehensive understanding of how housing preferences influence spatial movements

within urban areas, offering robust empirical evidence. Moreover, it emphasizes practical implications for policymaking by identifying match and mismatch patterns between housing search behaviors and house price levels, enhancing its relevance for urban policy and planning strategies.

5.2. Discussions

This study holds significant publication value due to its innovative integration of migration flow data, housing search patterns, and advanced spatial analysis techniques. It offers practical insights for policymakers and urban planners, identifying match and mismatch patterns in housing preferences, addressing housing affordability challenges and optimizing urban development strategies. This timely contribution enhances scholarly discourse by offering a nuanced perspective on how spatial dynamics influence residential mobility and inform evidence-based decision-making in urban planning and policy formulation. Practical implications of the findings for policymakers and urban planners can be highlighted. Insights into search and migratory patterns and their correlation with house price levels can inform housing policies, urban development strategies, and the allocation of resources.

Firstly, they help planners and designers figure out and predict varying levels of housing search demand and their corresponding locations. This aids in addressing inquiries pertaining to the optimal locations, optimal quantity, and appropriate types of structures for constructing new properties. The analysis of new housing unit supply should encompass both housing searches and paid house transactions. By integrating search patterns with established transactions or the properties listed that are presently available in the market, significant insights can be gained regarding latent demand in specific domains. This approach enables the identification of market limitations and restrictions.

Secondly, the housing search data may assist planners and designers in their efforts to identify spatial and structural submarkets in advance of the visual approach. The housing submarkets exhibit segmentation in terms of prices and locations. Consequently, the analysis of housing search data can be highly beneficial in identifying these submarkets. This can be achieved by examining the alignment or discrepancy between supply and demand within each submarket and determining how robust the current submarkets are.

Thirdly, they allow planners and designers to discuss changes in housing search demands and locations as a means of identifying good or bad policy outcomes. When housing searches suddenly increased in a neighborhood with low house prices, it meant that the neighborhood's appeal had increased due to beneficial effects such as an enhanced transportation system, better-quality housing, or favorable governmental policies. In contrast, a drop in searches in an area with a high number of home sales may indicate a reduction in the demand for homes. Such matches and mismatches in housing search demands and locations should be taken into account by planners and designers.

5.3. Limitations and Further Research

The case of the UK shows that the results and research method are practical and efficient, and they can be applied in other countries. Data accuracy and completeness, particularly in migration flows and house prices, could also impact results. Methodological limitations such as assumptions made in spatial analysis and potential biases in data collection methods should be acknowledged.

First, the limitations include the reliance on data, which may not reflect current market conditions. However, using older data from 2013 or 2014 in property market analyses serves several purposes. It provides a long-term perspective on market trends, allowing researchers to track changes over time influenced by economic cycles and policy shifts. Continuity in research projects also requires consistent datasets, ensuring reliability in findings. Moreover, processing large datasets takes time, which can delay data availability. Historical data also offer a stable baseline for comparison, helping analysts understand market responses to economic, social, and environmental factors. Despite limitations

like potential changes since data collection, researchers contextualize findings within the dataset's timeframe and relevance to current market conditions.

Second, the limitations pertain to the methodologies, which are statistical methods and GIS techniques. It is important to acknowledge that the use of the GIS technique does not provide a conclusive evaluation of mismatches. Additionally incorporating K-means clustering in urban planning provides an understanding of spatial dynamics beyond the approach regarding price per square meter. K-means clustering identifies distinct patterns and groups within urban areas based on multiple variables, offering insights into neighborhood composition, demographic trends, and socioeconomic factors that influence urban development decisions. This K-means method has been used by a number of authors in the real estate data analysis literature [38–40]. However, there are also limitations. It heavily relies on initial centroid placement, making it sensitive to starting conditions and potentially leading to suboptimal clustering outcomes. The method also assumes clusters are spherical and of similar size, which may not accurately reflect real-world data, where clusters can vary in shape and density. Moreover, K-means requires the number of clusters to be specified in advance, which can be challenging without prior knowledge of the dataset's structure. These constraints highlight the need for the careful consideration of alternatives when analyzing complex datasets with diverse cluster characteristics.

A certain matter relating to the representativity of online search data, and especially in 2014, is also discussed in this section. Online search data, particularly in domains like home searching, carry inherent biases that can skew the representativeness. Users engaging in such searches typically possess certain socioeconomic advantages, including access to technology and digital literacy skills. This demographic tends to be more affluent, educated, and technologically adept, potentially excluding less privileged groups who lack these resources or skills. Moreover, online behavior itself may favor certain demographics, such as younger or urban populations, further distorting the data's representation of broader societal trends. Consequently, while online search data offer valuable insights, it should be interpreted cautiously, complemented with other data sources to provide a more balanced and inclusive understanding of consumer behaviors and market dynamics.

Areas for further study could include longitudinal analyses to track changes over time and qualitative research to explore underlying reasons for search and migration patterns. Otherwise, future research could supplement the limitations of the GIS method by explaining the differences in choosing the search patterns based on household characteristics such as gender, ethnicity, religion, age groups, and income. The integration of user-generated search data with census data in the UK provides valuable insights into migration patterns. It is feasible to examine the potential correlation between search activity and migration patterns, as well as the geographical influence of property values on these variables.

Author Contributions: Conceptualization, T.Q.L.; methodology, L.-V.-L.D.; formal analysis, L.-V.-L.D.; resources, L.-V.-L.D.; writing—original draft preparation, T.Q.L.; writing—review and editing, T.Q.L.; visualization, L.-V.-L.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The migration flow data were freely downloaded from the Web-based Interface to Census Interaction Data (WICID) platform managed by the UK Data Service Census Support (<https://wicid.ukdataservice.ac.uk/>, accessed on 7 June 2023). The house price data were published by the HM Land Registry and were freely downloaded from the UK government website (<https://www.data.gov.uk/dataset/314f77b3-e702-4545-8bcb-9ef8262ea0fd/archived-price-paid-data-1995-to-2017>, accessed on 15 June 2024).

Conflicts of Interest: There are no financial conflicts of interest to disclose. No competing interests or funding were reported by the author(s).

References

1. Auckland Council. *Housing Choice and Preference: A Review of the Literature*; Auckland Council: Auckland, New Zealand, 2015.
2. Albrecht, J.; Anderson, A.; Smith, E.; Vroman, S. Opportunistic matching in the housing market. *Int. Econ. Rev.* **2007**, *48*, 641–664. [[CrossRef](#)]
3. Novy-Marx, R. Hot and Cold Markets. *Real Estate Econ.* **2009**, *37*, 1–22. [[CrossRef](#)]
4. Diaz, A.; Jerez, B. *House Prices, Sales and Time on the Market: A Search-Theoretic Framework*; Working Paper 09-25 Economic Series; Departamento de Economía, Universidad Carlos III de Madrid: Madrid, Spain, 2009.
5. Leung, C.K.Y.; Zhang, J. “Fire Sales” in Housing Market: Is the House-Searching Process Similar to a Theme Park Visit? MPRA: West Kelowna, BC, Canada, 2011.
6. Rocheteau, G.; Weill, P. Liquidity in Frictional Asset Markets. *J. Money Credit Bank.* **2011**, *43*, 261–282. [[CrossRef](#)]
7. Peterson, B.M. *Foiled by Search: Housing Prices, Turnover and Bubbles*; Bank of Canada: Ottawa, ON, Canada, 2009.
8. Ngai, L.R.; Tenreyro, S. Hot and Cold Seasons in the Housing Market. *Am. Econ. Rev.* **2014**, *104*, 3991–4026. [[CrossRef](#)]
9. Cheng, P.; Lin, Z.; Liu, Y.; Seiler, M.J. The Benefit of Search in Housing Markets. *J. Real Estate Res.* **2015**, *37*, 597–622. [[CrossRef](#)]
10. MacLennan, D. *Housing Economics: An Applied Approach*; Addison-Wesley Longman Ltd.: Toronto, ON, Canada, 1982.
11. Steegmans, J.; Bruin, J.D. Online Housing Search: A Gravity Model Approach. *PLoS ONE* **2021**, *16*, e0247712. [[CrossRef](#)] [[PubMed](#)]
12. Rae, A. Online Housing Search and the Geography of Submarkets. *Hous. Stud.* **2015**, *30*, 453–472. [[CrossRef](#)]
13. Doan, L.-V.-L.; Yudono, A. A Systematic Analysis of the Spatial Migration Patterns and the Linkages between Market Components in Greater Manchester, United Kingdom. *Int. J. Hous. Mark. Anal.* **2021**, *15*, 995–1018. [[CrossRef](#)]
14. Lomax, N. Internal Migration: Census Data as a Gold Standard for Insight and Visualisation. *J. Maps* **2022**, *18*, 543–550. [[CrossRef](#)]
15. Sun, L.; Zhu, H. GIS-Based Spatial Decision Support System for Real Estate Appraisal. In Proceedings of the 2009 Fourth International Conference on Computer Sciences and Convergence Information Technology, Seoul, Republic of Korea, 24–26 November 2009; IEEE: New York, NY, USA, 2009; pp. 1339–1344.
16. Knowles, A.K. Emerging Trends in Historical GIS. *Hist. Geogr.* **2005**, *33*, 7–13.
17. Li, S.; Ye, X.; Lee, J.; Gong, J.; Qin, C. Spatiotemporal Analysis of Housing Prices in China: A Big Data Perspective. *Appl. Spat. Anal. Policy* **2017**, *10*, 421–433. [[CrossRef](#)]
18. Singh, A.; Sharma, A.; Dubey, G. Big Data Analytics Predicting Real Estate Prices. *Int. J. Syst. Assur. Eng. Manag.* **2020**, *11*, 208–219. [[CrossRef](#)]
19. Jacob, C.B.; Asha, K. Real-Estate Housing Market Analytics and Prediction Using Big Data for Post Pandemic Era. In Proceedings of the 14th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2022), Seattle, WA, USA, 14–16 December 2022; Abraham, A., Hanne, T., Gandhi, N., Manghirmalani Mishra, P., Bajaj, A., Siarry, P., Eds.; Springer Nature: Cham, Switzerland, 2023; pp. 80–89.
20. Mittal, V.; Schaposnik, L.P. Housing Forecasts via Stock Market Indicators. *arXiv* **2022**, arXiv:2210.10146. [[CrossRef](#)] [[PubMed](#)]
21. Tu, Y.; Goldfinch, J. A Two-Stage Housing Choice Forecasting Model. *Urban Stud.* **1996**, *33*, 517–537. [[CrossRef](#)]
22. Levy, D.; Lee, C. The Influence of Family Members on Housing Purchase Decisions. *J. Prop. Investig. Financ.* **2004**, *22*, 320–338. [[CrossRef](#)]
23. Wong, G.K. A Conceptual Model of the Household’s Housing Decision-Making Process: The Economic Perspective. *Rev. Urban Reg. Dev. Stud.* **2002**, *14*, 217–234. [[CrossRef](#)]
24. Kushner, J.A. Apartheid in America: An Historical and Legal Analysis of Contemporary Racial Residential Segregation in the United States. *Howard Law J.* **1979**, *22*, 547–686. [[CrossRef](#)]
25. Mortensen, D.T. Wage Dispersion in the Search and Matching Model. *Am. Econ. Rev.* **2010**, *100*, 338–342. [[CrossRef](#)]
26. Bourassa: Defining Housing Submarkets—Google Scholar. Available online: https://scholar-google-com.bib-proxy.uhasselt.be/scholar_lookup?title=Defining%20housing%20submarkets&journal=Journal%20of%20Housing%20Economics&volume=8&issue=2&pages=160-183&publication_year=1999&author=Bourassa,SC&author=Hamelink,F&author=Hoesli,M&author=MacGregor,BD (accessed on 15 June 2024).
27. Zhang, D.; Zhang, X.; Zheng, Y.; Ye, X.; Li, S.; Dai, Q. Detecting Intra-Urban Housing Market Spillover through a Spatial Markov Chain Model. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 56. [[CrossRef](#)]
28. Chen, Z.; Cho, S.-H.; Poudyal, N.; Roberts, R.K. Forecasting Housing Prices under Different Market Segmentation Assumptions. *Urban Stud.* **2009**, *46*, 167–187. [[CrossRef](#)]
29. Calinski, T.; Harabasz, J. A Dendrite Method for Cluster Analysis. *Commun. Stat.—Theory Methods* **1974**, *3*, 1–27. [[CrossRef](#)]
30. Hincks, S.; Kingston, R.; Webb, B.; Wong, C. A New Geodemographic Classification of Commuting Flows for England and Wales. *Int. J. Geogr. Inf. Sci.* **2018**, *32*, 663–684. [[CrossRef](#)]
31. Stromberg Snyder, P.; Weber, M.J. A Research Note on: An Exploration of Journey to Work and Choice of Housing Location: 1974 to 1977. *Hous. Soc.* **1983**, *10*, 83–88. [[CrossRef](#)]
32. Gordon, I. Interdistrict Migration in Great Britain 1980–1981: A Multistream Model with a Commuting Option. *Environ. Plan. Econ. Space* **1988**, *20*, 907–924. [[CrossRef](#)] [[PubMed](#)]
33. Hoekstra, J. Is There a Connection between Welfare State Regime and Dwelling Type? An Exploratory Statistical Analysis. *Hous. Stud.* **2005**, *20*, 475–495. [[CrossRef](#)]
34. Doan, L.-V.-L.; Rae, A. Are We Looking in the Right Place? Housing Search Mismatches: Evidence from Greater Manchester in the UK. *Open House Int.* **2023**. ahead-of-print. [[CrossRef](#)]

35. Ömer, Ö. Dynamics of the US Housing Market: A Quantal Response Statistical Equilibrium Approach. *Entropy* **2018**, *20*, 831. [[CrossRef](#)]
36. Li, Y.; Liu, H. Research Progresses on Migration Spatial Structure Modeling. *Prog. Geogr.* **2010**, *10*, 1162–1170.
37. Brzezicka, J. The Application of the Simplified Speculative Frame Method for Monitoring the Development of the Housing Market. *Real Estate Manag. Valuat.* **2022**, *30*, 84–98. [[CrossRef](#)]
38. Sandeep Kumar, E.; Talasila, V.; Rishu, N.; Suresh Kumar, T.V.; Iyengar, S.S. Location Identification for Real Estate Investment Using Data Analytics. *Int. J. Data Sci. Anal.* **2019**, *8*, 299–323. [[CrossRef](#)]
39. Guo, K.; Wang, J.; Shi, G.; Cao, X. Cluster Analysis on City Real Estate Market of China: Based on a New Integrated Method for Time Series Clustering. *Procedia Comput. Sci.* **2012**, *9*, 1299–1305. [[CrossRef](#)]
40. Huang, Z. Extensions to the K-Means Algorithm for Clustering Large Data Sets with Categorical Values. *Data Min. Knowl. Discov.* **1998**, *2*, 283–304. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.