social sciences

MDPI

Article

Migration and Segregated Spaces: Analysis of Qualitative Sources Such as Wikipedia Using Artificial Intelligence

Javier López-Otero [1], Ángel Obregón-Sierra [1] and Antonio Gavira-Narváez [2,*]

[1] Facultad de Letras y de la Educación, Universidad Internacional de La Rioja, Avenida de la Paz 137, 26006 Logroño, La Rioja, Spain; javier.lopezotero@unir.net (J.L.-O.); angel.obregon@unir.net (Á.O.-S.)
[2] Departament of Ciencias Sociales, Filosofía, Geografía y Traducción e Interpretación, Universidad de Córdoba, Square of Cardenal Salazar, 14003 Córdoba, Andalusia, Spain
* Correspondence: antonio.gavira@uco.es

Abstract: The scientific literature on residential segregation in large metropolitan areas highlights various explanatory factors, including economic, social, political, landscape, and cultural elements related to both migrant and local populations. This paper contrasts the impact of these factors individually, such as the immigrant rate and neighborhood segregation. To achieve this, a machine learning analysis was conducted on a sample of neighborhoods in the main Spanish metropolitan areas (Madrid and Barcelona), using a database created from a combination of official statistical sources and textual sources, such as Wikipedia. These texts were transformed into indexes using Natural Language Processing (NLP) and other artificial intelligence algorithms capable of interpreting images and converting them into indexes. The results indicate that the factors influencing immigrant concentration and segregation differ significantly, with crucial roles played by the urban landscape, population size, and geographic origin. While land prices showed a relationship with immigrant concentration, their effect on segregation was mediated by factors such as overcrowding, social support networks, and landscape degradation. The novel application of AI and big data, particularly through ChatGPT and Google Street View, has enhanced model predictability, contributing to the scientific literature on segregated spaces.

Keywords: urban segregation; urban landscape; Wikipedia; ChatGPT; machine learning; immigration

check for updates

Citation: López-Otero, Javier, Ángel Obregón-Sierra, and Antonio Gavira-Narváez. 2024. Migration and Segregated Spaces: Analysis of Qualitative Sources Such as Wikipedia Using Artificial Intelligence. *Social Sciences* 13: 664. https://doi.org/10.3390/socsci13120664

Academic Editors: Federico Benassi, Ricardo Iglesias-Pascual, Salvatore Strozza and Xiaoling Shu

Received: 25 October 2024
Revised: 24 November 2024
Accepted: 6 December 2024
Published: 11 December 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/).

1. Introduction

Residential segregation is a complex phenomenon related to the social, economic, political, and cultural dynamics of a society. The influencing factors are heterogeneous, maintaining reciprocal relationships that complicate a detailed analysis. The individualism inherent in human beings and the fear of the unknown safeguard these spaces that seek homogeneity in the characteristics of the resident population (García-Hernández 2020). This phenomenon tends to manifest itself, especially in metropolitan areas, where the concentration of the immigrant population in certain neighborhoods often ends up generating spaces segregated from the rest of the city.

Several studies have examined spatial segregation, identifying how various factors contribute to segregation. In this regard, the income of residents, land prices, the concentration of immigrants from the same geographic origin, and the characteristics of the urban landscape are significant factors (Vaalavuo et al. 2017). Thus, it has been established that neighborhoods with a low average income tend to concentrate a higher proportion of immigrants, which in turn increases the likelihood of segregated enclaves (Van Ham et al. 2021). Similarly, the key importance of the urban landscape and negative narratives about certain neighborhoods has been identified, which can certainly accentuate the social exclusion of certain neighborhoods, creating a vicious cycle that hinders the social and residential mobility of immigrants (Sýkora and Bouzarovski 2012).

Despite the abundance of literature on residential segregation in Spanish cities, there are significant gaps regarding the relationship between the different factors that contribute to this phenomenon. Thus, little research has been conducted on the combined impact of variables such as income, land prices, urban landscape, and the geographic concentration of immigrants on segregation and how these factors interact differently according to the geographic origin of migrant populations. In this regard, Rodríguez-García (2022) and Lacayo (2016) point out that segregation does not manifest itself in the same way in all immigrant groups. However, the current literature does not provide sufficient detail on the differences between populations from Asia, Africa, and Latin America in Spanish cities.

This research gap poses a fundamental question: How do average income, the urban landscape and the concentration of immigrants from different geographic origins interact in the formation of segregated spaces in Spain's main metropolitan areas? The answer to this question will allow us not only to identify the most influential factors in the formation of segregated spaces but also to identify differences in the processes of urban segregation according to the geographic origin of immigrants.

On the other hand, there are abundant studies on the segregation of spaces (Matossian 2018), although statistical sources that estimate the segregation of urban spaces with little rigor are often resorted to due to limitations imposed by statistical legislation on the characteristics of the population and because the methodologies for obtaining data evolve at a slower pace than the current social dynamics. That is why it may be convenient to resort to other textual sources that adapt more quickly such as social networks or an encyclopedia such as Wikipedia.

Therefore, this paper seeks to advance the understanding of residential segregation processes in neighborhoods with high immigrant concentrations in Madrid and Barcelona, addressing gaps in the existing literature on the interplay between economic, spatial, and demographic factors in shaping segregated spaces. By employing a comparative analysis that uses both traditional demographic data and unconventional sources such as Wikipedia, this study offers a novel methodological approach. Wikipedia's dynamic and community-driven content provides a unique perspective on urban spaces, capturing narratives and representations that evolve more rapidly than traditional statistical sources.

These analyses, combined with indices measuring the geographic concentration of immigrant populations by origin (Asia, Africa, Latin America, and Europe), aim to uncover nuanced differences in segregation patterns across diverse migrant groups. By integrating quantitative measures with insights from publicly accessible knowledge bases, this study not only contributes to the academic debate on ethnic residential segregation but also highlights the importance of alternative data sources for understanding complex urban phenomena. Ultimately, this research aspires to equip policymakers and urban planners with actionable insights to foster more equitable and inclusive cities.

1.1. Main Factors Influencing Urban Segregation

Income levels have a direct impact on where immigrant populations settle in cities, as they create barriers to housing and access to certain neighborhoods. As a result, immigrants, typically with fewer resources, concentrate in areas with more affordable housing, which increases residential segregation (Jones 2011). Neighborhoods with lower incomes tend to be more densely populated, as people are forced to share smaller, more accessible spaces. In addition, the concentration of low-income immigrants in neighborhoods fosters the creation of segregated communities, where opportunities for social and residential mobility are limited. Furthermore, segregation creates disparities in access to public services, infrastructure, and job opportunities, thereby exacerbating economic and social inequalities in cities. As Musterd et al. (2016) point out, this dynamic reinforces patterns of urban inequality, which in turn weakens social cohesion and the integration of immigrants into the urban fabric.

1.1.1. Price of Land

Housing prices and land values are not only linked to accessibility to goods and services but also to socioeconomic and political dynamics that influence residential segregation. In this sense, Skifter Andersen et al. (2015) argue that segregation is not only caused by economic factors but that urban planning policies also represent an essential factor in the distribution of population in cities. The concentration of social housing in certain areas, instead of a more equitable distribution throughout the urban space, contributes to the creation of segregated low-income neighborhoods, which reinforces spatial segregation. In this way, the price of land, influenced by real estate speculation, limits the potential for developing affordable housing in central or better-serviced areas.

Trounstine (2020) argues that urban policies have historically fostered patterns of segregation that continue to this day, with zoning decisions that exclude certain social groups from more affluent neighborhoods. These decisions have had a disproportionate impact on immigrant and minority communities, who are not only affected by economic barriers but also by discriminatory practices that restrict their access to affordable rental housing in areas with higher real estate values and better infrastructure. As Cuena (2023) points out, this spatial segregation is reinforced by the lack of inclusive policies that promote genuine social and economic integration in cities. Instead, the concentration of investments in areas of more profitable areas increases the value of land in these spaces, displacing the population with lower incomes to the urban periphery.

Thus, as Sampson (2012) indicates, immigrants' choice of residence is not so much a matter of preference but of constraints set by the real estate market and the discriminatory practices of the different groups involved in the supply of housing. Social networks, such as family and friends, play an important role in the concentration of this population in specific neighborhoods, reinforcing segregation. In addition, failed urban policies contribute to the fact that the neighborhoods where the immigrant population is concentrated are affected by the lack of investment in public services, which in turn contributes to prolonging socioeconomic and spatial inequalities (Ministerio de Vivienda y Agenda Urbana 2024).

1.1.2. Compatriot Networks and Immigrant Placement

Cultural, ethnic, and social identifications are primary factors in the residential segregation of the immigrant population, as they influence the preferences they may have in choosing their residence. Often, the immigrant population is confronted with the practices of some landlords who limit rental options in certain neighborhoods, resulting in concentration in neighborhoods with less strict requirements. Likewise, the immigrant population tends to consult relatives, friends, or compatriots about the most appropriate places to look for housing, resulting in concentration in certain city neighborhoods (Sampson 2012).

Likewise, urban planning often has a direct impact on the configuration of cities, as well as on segregation patterns. One example is the development of social or public housing in specific areas, as opposed to atomization in large cities, which can lead to greater inclusion of the population (Musterd and Ostendorf 1998).

1.1.3. The Urban Landscape

The urban landscape is an essential factor in shaping social dynamics within the city, as it represents existing inequalities and helps to perpetuate spatial segregation. In the most degraded neighborhoods, deterioration not only discourages investment and improvement of urban space but also contributes to reinforcing barriers between different social groups. According to Knorr (2017), the presence of visual elements such as dilapidated buildings, poorly maintained public spaces, dirty areas, and visible signs of neglect such as graffiti is a visible reminder of socioeconomic divisions, which in turn discourages non-residents from accessing these areas, thereby reinforcing segregation. Such devaluation of the urban landscape, which is often related to a lack of resources and public attention, contributes to these spaces being perceived as marginal areas, further limiting integration and improvement opportunities for their inhabitants (Knorr 2017).

Likewise, LaFleur (2021) argues that cultural appropriation of the urban landscape is a process by which immigrant or marginalized groups adapt and transform the environment to meet their needs and values. Through this resignification of space, inhabitants create forms of resistance against imposed marginalization, generating new uses for the spaces they inhabit. In this sense, Knorr (2017) highlights that, although both factors based on cultural appropriation and landscape degradation are different, both coincide in that the urban environment has significant power to reinforce or transform social divisions.

Similarly, Zukin (2010) and Harvey (2012) have pointed out that the relationship between the urban landscape and segregation is not only embodied in its physical deterioration but also in the distribution of a symbology and control of resources in the city. Inequality in access to quality spaces, both in terms of housing and public infrastructure, perpetuates the marginalization of certain groups. In this sense, urban exclusion is not only a physical issue, but also a cultural one, as spaces become representative of who has the right to the city and who does not (Zukin 2010).

Finally, as Roberto and Hwang (2015) point out, in cities there are also elements that influence residential segregation such as the existence or construction of natural barriers that can be, among others, rivers or mountains, as well as those related to transport road infrastructures. Barriers to accessibility significantly impact population distribution, often limiting access to essential resources and services.

1.1.4. Local Demographics and Immigration

Population size and density are important contributors to urban segregation, as more densely populated neighborhoods are associated with a higher proportion of immigrants, due to the need to access more affordable housing in limited areas. This higher density increases competition for space and contributes to reinforce residential segregation, as immigrants with lower incomes are forced to live in overcrowded neighborhoods. In this sense, Crowder et al. (2011) argue that an increase in the density of immigrants in a neighborhood favors the probability that residents move out, thus favoring segregation. In turn, Owens (2019) notes that income segregation is higher in areas where housing segregation is already significant, which favors urban disparities. However, although there has been considerable study on the impact of population density, there is a significant gap in studies analyzing the impact of population volume on the formation of segregated spaces.

1.1.5. Segregated Spaces

Methods for quantifying segregated spaces are varied. In Spain, the latent problem of inequality and vulnerability of the population involves 22 observatories whose analysis is carried out at different scales, from national to local (Giménez-Bertomeu 2020). However, the main data sources usually used to analyze these problems at the local level are the Population and Housing Census and the Atlas of Household Income Distribution of the National Institute of Statistics. The latter source has a level of detail that reaches census districts and sections. It also provides information on the average income per person and household and the Gini index, among other variables.

Another primary source in Spain is the Urban Vulnerability Observatory of the Ministry of Housing and Urban Agenda, which has different applications, including the Atlas of Urban Vulnerability and the Urban Analysis of Vulnerable Neighborhoods in Spain, which includes the Catalog of Vulnerable Neighborhoods and the Vulnerable Neighborhoods Catalog Viewer. It also includes several reports on urban vulnerability in some large Spanish cities, such as Madrid and Barcelona. In this context of data search and analysis, Wikipedia is presented as a valuable tool, since it receives almost 30 million daily visits in its Spanish version and more than 250 million in its English version (Pageviews 2024).

1.1.6. Immigrant Resilience and Adaptation

Some threads of the scholarly literature have also focused on how immigrants often demonstrate remarkable resilience and adaptability when facing the structural challenges

of residential segregation. Despite the isolating effects of these barriers, immigrants very often show resilience by reshaping the urban spaces they inhabit to recreate an environment consistent with their cultural, economic, and social needs. This process, often referred to as 'place-making', involves transforming neighborhoods into hubs of cultural and economic activity through the establishment of ethnic businesses, cultural centers, and informal support networks (Simone 2018).

Other scholars, such as Cordero-Guzmán (2005), have highlighted how immigrant communities resist marginalization by creating grassroots efforts, such as collective bargaining for housing rights or the formation of local advocacy groups, exemplifying their ability to resist exclusion and create pathways for inclusion. This resilience is particularly evident in neighborhoods like Raval in Barcelona or Lavapiés in Madrid, where immigrant populations have revitalized urban spaces despite systemic neglect (Zapata-Barrero 2024).

This resilience not only demonstrates their capacity to adapt but also highlights their role in challenging urban exclusion and fostering social cohesion. As Lefebvre (1991) and Zukin (2010) emphasize, the 'right to the city' is actively reclaimed by marginalized groups through their everyday practices and transformations of urban environments, countering the divisive impacts of segregation.

1.2. Wikipedia

Wikipedia is a free online encyclopedia that can be consulted and edited by anyone. Since its creation in 2001, there have been numerous scientific articles analyzing it, but most importantly, it has established itself as a very important tool in the access and dissemination of information across various fields of knowledge (Obregón and González 2018). It is widely used by users around the world, as well as by third-party services, such as search engines (Google, Bing ...) and knowledge bases. It is perceived as a reliable source of information, although the reliability of its content is constantly debated, as anyone can edit it (Obregón-Sierra et al. 2023). However, the trust that users and third-party services place in Wikipedia seems to be justified by its quality and reliability (Singh et al. 2021).

With more than six million articles in its English version and millions of articles in other languages, Wikipedia covers a wide range of topics, including demographic, geographic, and urban aspects (List of Wikipedias 2024). For studies of residential segregation, Wikipedia provides an invaluable resource by offering data on the history, development, and current characteristics of neighborhoods and cities around the world. These articles are often continually updated, reflecting recent changes in the demographic and urban composition of these places (Deltell and Claes 2021).

Wikipedia's collaborative structure allows users worldwide to edit geographic articles, incorporating local perspectives that enrich the analysis of factors influencing residential segregation. Although it is not a primary source per se, Wikipedia acts as a repository of references and data from official sources, such as censuses or government reports, as well as information from the media, making it a complementary platform for academic research (Obregón and González 2021).

One of the key points that favors its use in studies of this type is the availability of articles in multiple languages. Wikipedia exists in 332 languages (List of Wikipedias 2024). Although with large differences between versions, this high number allows comparison of the representation of the same area in different languages, which may reveal differences in perception and emphasis on factors such as immigration or urban vulnerability.

Factors such as the number of editors who contributed to an article or discussion pages are important throughout the life cycle of an article (Zhang et al. 2018), but the existence of articles for each neighborhood allows us to analyze the content of these to identify demographic and geographic patterns. It is important to review the categorization of neighborhoods or districts in each Wikipedia article, as it allows us to track aspects related to classification with respect to immigration, urban development, and infrastructure (Aghaebrahimian et al. 2022). For their part, the links between articles provide insight into

the interconnectedness between different areas, helping to better contextualize the drivers of segregation.

However, we consider that what is most important is the content of each article. Although users use references to verify information, Wikipedia reflects not only factual data but also public perceptions and debates on various issues. The selection of certain biased sources can lead to biased content (Zubimendi et al. 2024).

If these perceptions are embodied in Wikipedia articles, which will be read by millions of people, they can generate opinions and biases about specific individuals or places. Performing a content analysis of Wikipedia articles requires careful review, as reading them may not always be neutral and often involves processing a large amount of information about those places.

1.3. Artificial Intelligence

Artificial intelligence (AI) has revolutionized the traditional way of working, as well as the analysis and understanding of complex phenomena. Since the launch of ChatGPT in November 2022, several areas have focused their research on this artificial intelligence chatbot application developed by OpenAI. These include customer care, health care, education, finance, and communication technology (Singh and Singh 2023).

In the short time since its appearance, there are already several articles that offer a literature review on ChatGPT, mainly referring to health and education (Aydın and Karaarslan 2022; Li et al. 2024; Montenegro-Rueda et al. 2023), but also about its applications and its potential impact in different fields (Gabashvili 2023).

In the latter, for example, generative AI is used to obtain cognitive collaboration, leveraging its potential to learn "with" the technology rather than "from" it (Fuertes 2024). In this field, a great influence of the use of ChatGPT on the quality of the submitted papers has been observed (Baldrich and Domínguez-Oller 2024). In health care, benefits are also noted, but the need for accurate and current data is revealed as one of the main obstacles to adopting ChatGPT (Javaid et al. 2023).

In other areas such as geography, ChatGPT is also used for various studies. For example, it provides invaluable assistance in creating large-scale maps and understanding geographic space (Tao and Xu 2023), serves as a cartographic assistant, considerably enriching maps (Juhász et al. 2023) or to study the dynamics of coastal water quality (Aziz et al. 2024), though overall, research in this area remains limited.

In a context where population distribution patterns and urban dynamics are constantly changing, the ability of AI to process large volumes of data quickly and accurately provides a significant advantage for researchers and urban planners (Bermúdez-Quimis 2024). AI not only helps identify correlations between socioeconomic, demographic, and geographic factors but also detects hidden trends, makes predictions, and analyzes data with a level of accuracy that surpasses traditional methods (Tejedor 2024).

In addition, AI facilitates information analysis by processing multiple sources, such as Excel files containing national censuses, PDF documents, and web pages (OpenAI 2024). This ability to understand datasets and combine structured and unstructured data, along with its ability to analyze information in real time, allows researchers to obtain a more holistic and up-to-date view.

Another key aspect is that AI has significantly improved digital content analysis techniques, idea generation, and innovation (Livberber 2023). Its advanced algorithms can efficiently analyze large volumes of texts and identify discourse patterns, highlighting how narratives in those texts are structured. For this reason, the aim of this paper is to analyze the factors that favor the formation of segregated spaces due to the concentration of immigrants in certain territories of a city, analyzing Wikipedia content using ChatGPT.

2. Materials and Methods

The bibliographic analysis presented earlier identified the factors that favor the concentration of the migrant population and contribute to segregation processes. This identifi-

cation allowed us to propose the following causal relationship in the form of an equation: Segregated spaces = Income + Price of land + Concentration of residents of the same geographical origin + Urban landscape + Neighborhood demographics.

2.1. Dependent Variables

Segregated spaces have been approximated using two groups of variables: the percentage of the immigrant population in the neighborhood and two indexes derived from processing information contained in a large knowledge base. The first factor assesses only the concentration of immigrants, not the potential marginality associated with it. In contrast, the indexes have been developed to evaluate neighborhood segregation by immigration from two different perspectives, which will be explained below.

Both indexes were obtained from a set of parameters sourced from Wikipedia. Specifically, the degree of segregation in the sample neighborhoods has been analyzed through the interpretation of the information by a Large Language Model, specifically ChatGPT. Executing the following prompt on the information from each Wikipedia article, for each neighborhood and language, allowed us to obtain a series of parameters regarding the degree of segregation in each neighborhood.

The prompt-questionnaire includes 12 items, most of which are parameters that help infer the degree of marginality in the neighborhood. These parameters were designed based on the analyses of Koopmans and Schaeffer (2016), McCann and Boateng (2020), Laurence and Bentley (2016), and Dinesen and Sønderskov (2015). After designing the prompt, it was applied to the texts collected from Wikipedia, and the indicators corresponding to each question were obtained. Additionally, an overall assessment of neighborhood segregation was requested from the text. Finally, it should be specified that all the information requested from Wikipedia was obtained in 2024.

Specifically, the prompt applied to each neighborhood would be as follows:

Variables and evaluation items

1. Economic inequality: Does the article disproportionately emphasize economic difficulties? Likert scale: 1 (Very negative or absent)–5 (Very positive or present).
2. Criminality: Does the article excessively highlight crime issues? Does the article link crime to the foreign population and immigration? Likert scale: 1 (Very negative or absent)–5 (Very positive or present).
3. Cultural representation: Does the article omit cultural diversity and the positive contributions of immigrant communities? Likert scale: 1 (Very negative or absent)–5 (Very positive or present).
4. Negative language: Does the article use language that perpetuates negative stereotypes about immigrants and minorities? Does the article highlight the presence of certain immigrant nationalities? Likert scale: 1 (Very negative or absent)–5 (Very positive or present).
5. Images and examples: Does the article use images that only show negative aspects, such as deteriorated areas or people in precarious situations? Likert scale: 1 (Very negative or absent)–5 (Very positive or present).
6. Incomplete history: Does the article fail to provide adequate historical context on how certain neighborhoods were formed and the policies that led to their current situation? Likert scale: 1 (Very negative or absent)–5 (Very positive or present).
7. Public policies: Does the article ignore public policies that have affected the community? Likert scale: 1 (Very negative or absent)–5 (Very positive or present).
8. Local successes: Does the article ignore success stories and improvements in the neighborhood, such as new businesses, infrastructure improvements, and reduced crime? Does the article highlight social problems such as isolation, persecution, and discrimination? Likert scale: 1 (Very negative or absent)–5 (Very positive or present).
9. Presenting a balanced view: Does the article present both the positive and negative aspects of the neighborhood, providing a balanced perspective? Likert scale: 1 (Very negative or absent)–5 (Very positive or present).

10. Use of diverse sources: Does the article use diverse sources and compare information to avoid bias and provide a more complete and objective view? Does the article make unverified negative claims? Likert scale: 1 (Very negative or absent)–5 (Very positive or present).
11. Updated statistics: Does the article use updated statistics and provide links to reliable sources so that readers can verify and explore the information further? Likert scale: 1 (Very negative or absent)–5 (Very positive or present).
12. Highlighting positive contributions: Does the article highlight the positive contributions of immigrant communities and the initiatives that are improving the neighborhood? Likert scale: 1 (Very negative or absent)–5 (Very positive or present).
13. Balanced images and examples: Does the article use images and examples that reflect both the challenges and the achievements and positive aspects of the neighborhood? Likert scale: 1 (Very negative or absent)–5 (Very positive or present).
14. Community involvement: Does the article encourage the participation of local residents, including immigrant communities, in editing and updating the content? Likert scale: 1 (Very negative or absent)–5 (Very positive or present).
15. Overall assessment: Evaluate the overall contribution of the Wikipedia article to spatial segregation, where 1 is minimal contribution and 5 is great contribution. Likert scale: 1 (Minimal contribution)–5 (Great contribution).

Only include numerical ratings.

This information was obtained for a sample of 74 neighborhoods in two metropolitan areas, Barcelona and Madrid. It should be noted that this database represents 80% of the neighborhoods. The data were gathered from neighborhoods where the percentage of the foreign population exceeds 25% of the total, focusing only on Spain's main metropolitan areas, specifically Madrid and Barcelona.

Moreover, Wikipedia information on each neighborhood was consulted in four languages: Spanish, English, French, and Catalan. Finally, all the information from each language was aggregated at the neighborhood level using an average.

A factor analysis of the obtained values was performed, and the interpretation of each factor was used as a dependent variable.

Thus, to conclude, three dependent variables were created in this research:

1. Factor Analysis Variable: This variable was obtained from a specific factor analysis, capturing spatial segregation related to immigration, based on data obtained by ChatGPT from Wikipedia. (Referred to later as "Factor Analysis Variable").
2. Foreigners Proportion: The percentage of foreigners relative to the total number of neighborhoods. (Referred to later as "Foreigners Proportion").
3. Index for Spatial Segregation: An index related to spatial segregation due to immigration, obtained by ChatGPT from Wikipedia and available in question 15 of the prompt. (Referred to later as "Index for Spatial Segregation").

2.2. Independent Variables

The independent variables of the model have been grouped into four categories: those that value the demographic size of the neighborhood, those that value the income of the neighborhood, those that value the formation of clusters of immigrants by geographic origin, those that analyze the price of land and those that value the urban landscape. Some variables have been obtained directly from different sources, while another group of variables are indexes, which will be explained below.

First, the index of geographic concentration of immigrants by geographic origin should be highlighted. The calculation of this index was based on the application of the Local Index of Spatial Autocorrelation (LISA) by Anselin (2013), as it is one of the most widely used methods for identifying the geographic concentration of variables. This algorithm detects local spatial patterns, providing an accurate measure of spatial autocorrelation in specific areas. Specifically, four analyses were performed—one for each of the four geographic origins: Asia, Africa, America, and Europe, excluding Spain. Since the National Institute

of Statistics does not provide neighborhood-level information on this, it was necessary to obtain such data at the census section level, which offers a notably smaller geographic unit, with each neighborhood containing between 6 and 7 census sections on average. After applying the analysis to the set of census sections, immigrant hotspots were identified for each geographic origin, as shown in Figures 1 and 2.

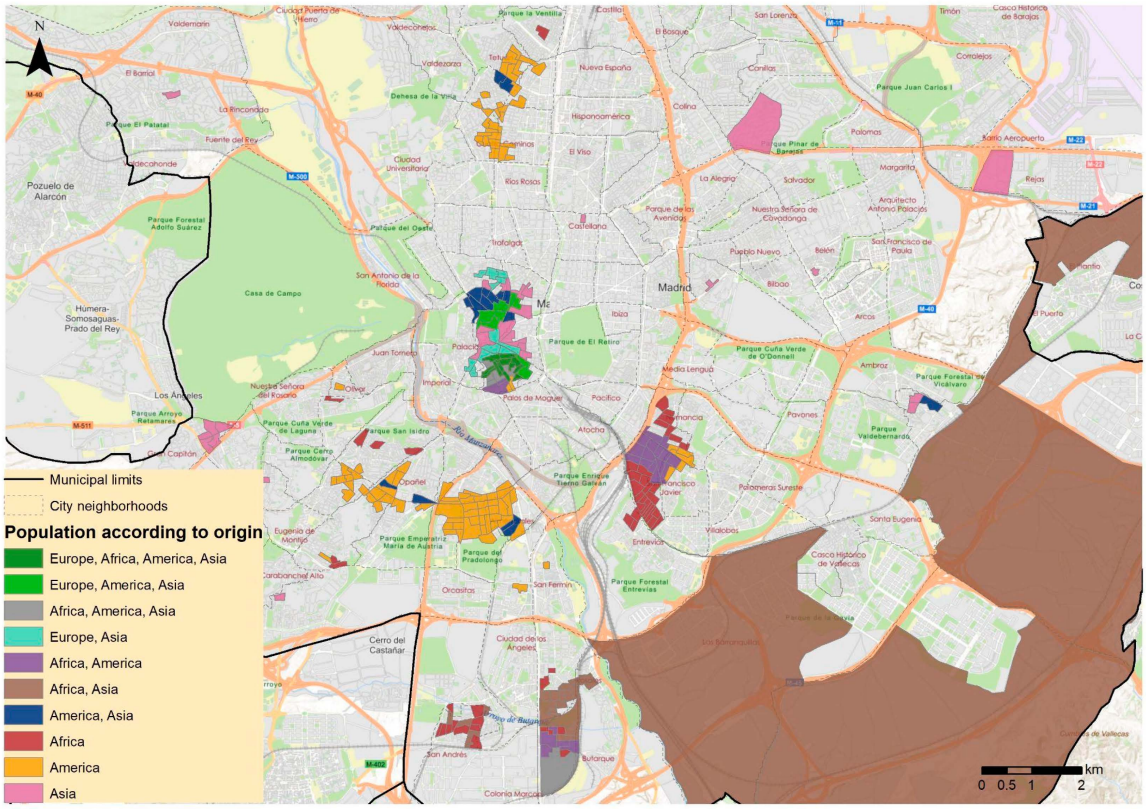


Figure 1. Concentration of immigrants in Madrid. Source: author.

Next, only those census sections that met the following criteria were selected: significant at the 95% confidence level, a positive High–High correlation, and at least 25% of the population being immigrants. Finally, the census sections identified for each geographic origin are presented below.

Once the LISA value for each census section was identified, the values for all the census sections in each neighborhood were averaged. Then, the percentage of these averaged values relative to the total sum was calculated. This proportion constituted the clustering indicator for each neighborhood and each geographic origin: Europe (Europe_Percentage), Latin America (Latino_Percentage), Africa (Africa_Percentage), and Asia (Asia_Percentage).

Another index, called Max_Clustering_Index, identifies the highest percentage of concentration for any continental ethnic group within a given neighborhood.

The final index (Urban_Landscape_Index) involves an assessment of the urban landscape using AI. Specifically, for each neighborhood, an image of a random street was selected from Google Street View. In neighborhoods with census sections that indicated a high concentration of a particular ethnic group, the street image was selected from those areas. The image was then downloaded and attached to ChatGPT-4.0, along with the following prompt: "Rate the urban landscape of the attached image using a Likert scale."

These variables were then entered along with the rest into a database with the following characteristics (see Table 1):

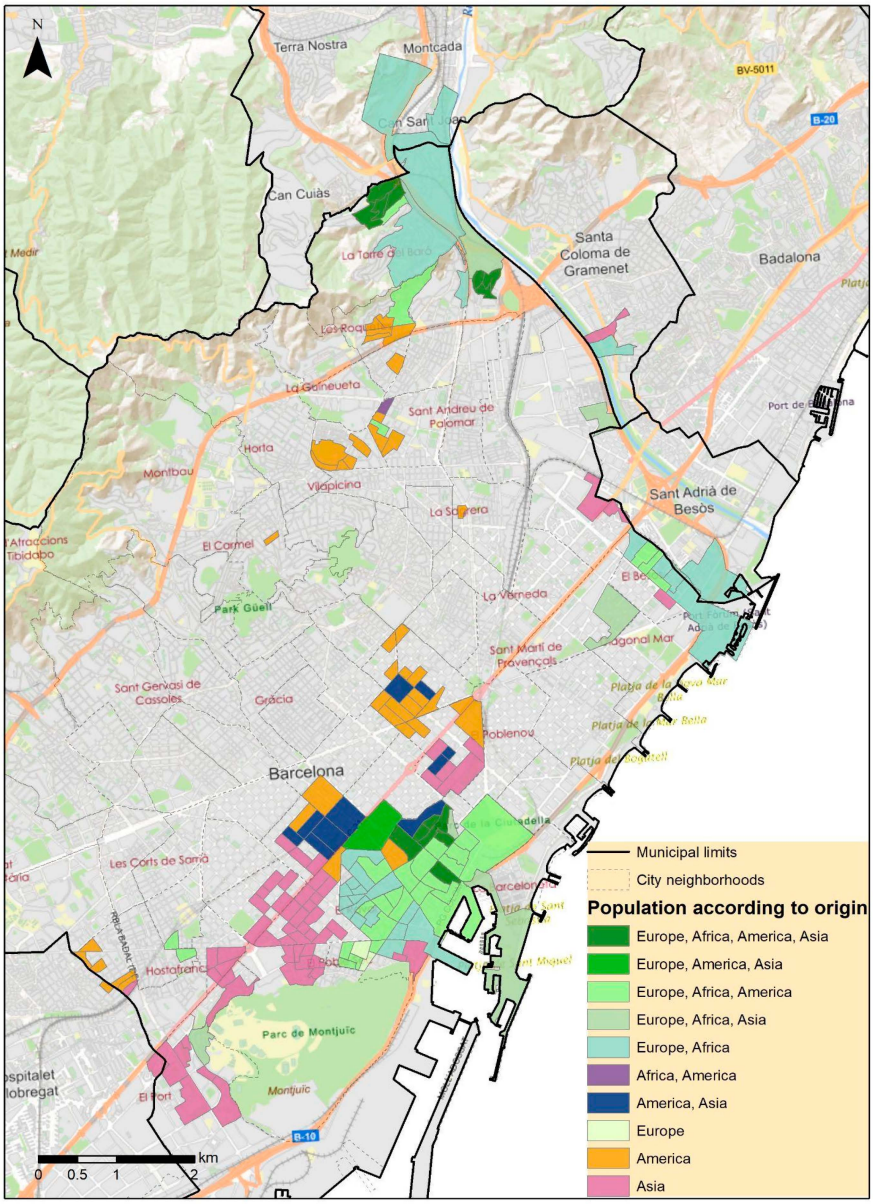


Figure 2. Concentration of immigrants in Barcelona. Source: author.

Finally, a regression analysis was carried out using four machine learning algorithms: Lasso, ridge regression, gradient boosting, and random forest. Algorithm selection was performed by comparing the mean square error (MSE) and the coefficient of determination (r^2) of the resulting models for each algorithm. The algorithm that showed the best performance in both parameters was selected for use in the research.

These four models are relevant because they are particularly suitable for this type of analysis. Lasso and ridge regression are regularized linear regression models, useful when the goal is to simplify the model by eliminating or reducing the weight of less relevant variables (Hastie et al. 2015; Zou and Hastie 2005). On the other hand, gradient boosting analyzes complex nonlinear interactions between variables and improves predictive performance by sequentially combining several weak models, such as decision trees (Friedman 2001; Chen and Guestrin 2016). Random forest, in contrast, is robust in avoiding overfitting, as it generates multiple decision trees randomly and averages their predictions, giving it high generalization capability (Breiman 2001). These characteristics make these algorithms particularly effective for analyzing data with multiple complex relationships and interaction structures between variables, as in the present study.

Table 1. Variables collected in the database. Source: own elaboration.

Variable	Description	Measurement	Source	Category
Name	Name of the neighborhood or district.	Name of the geographical entity	National Statistics Office	-
Africa_Clustering_Index	Clustering index for African geographic origin in the district.	Anselin LISA algorithm (value)	(Instituto Nacional de Estadística 2024b)	Index
Asia_Clustering_Index	Clustering index for Asian geographic origin in the district.	Anselin LISA algorithm (value)	(Instituto Nacional de Estadística 2024b)	Index
Europe_Clustering_Index	Clustering index for European geographic origin in the district.	Anselin LISA algorithm (value)	(Instituto Nacional de Estadística 2024b)	Index
Latino_Clustering_Index	Clustering index for Latin American geographic origin in the district.	Anselin LISA algorithm (value)	(Instituto Nacional de Estadística 2024b)	Index
Africa_Percentage	Proportion of African population in the district out of the total population.	Percentage	(Instituto Nacional de Estadística 2024b)	Index
Asia_Percentage	Proportion of Asian population in the district out of the total population.	Percentage	(Instituto Nacional de Estadística 2024b)	Index
Europe_Percentage	Proportion of European population in the district out of the total population.	Percentage	(Instituto Nacional de Estadística 2024b)	Index
Latino_Percentage	Proportion of American population in the district out of the total population.	Percentage	(Instituto Nacional de Estadística 2024b)	Index
Max_Clustering_Index	Maximum clustering index.	Maximum value of the clustering index	Index	Index
Resident_Population	Total number of people residing in the district.	Number of persons	Instituto Nacional de Estadística (2024a)	Demographic size
Net_Annual_ Income_Per_Capita	Average annual net income per inhabitant of the district.	Euros	Instituto Nacional de Estadística (2024a)	Renta
House_Price_Index	Average price of houses in the district.	Euros per square meter	Idealista (2024)	Land price
City_Average_Index	Average of city averages for different variables.	Numerical value	Idealista (2024)	Land price
Relative_House_Price_Index	Relative price of homes in the district compared to other areas of the city.	Numerical value	Idealista (2024)	Land price
Urban_Landscape_Index	Index of the district's urban landscape.	Numerical value	Google Street View (2024)	Index

To interpret the results of this model, we used SHAP (Shapley Additive Explanations), a technique based on game theory that decomposes the prediction of each observation and reveals the contribution of each variable. SHAP not only provides interpretability in terms of which variables are most important but also shows how each explanatory variable influences the dependent variable positively or negatively (Lundberg and Lee 2017). This is particularly useful for our analysis, as it allows us to visualize clearly whether factors such as income or streetscape are increasing or decreasing the likelihood of segregated spaces forming. SHAP offers an accurate and understandable representation of the causal

relationships between the independent variables and the observed outcomes, helping to generate more robust conclusions (Molnar 2020).

3. Results

Once the database has been completed and the analyses explained in the Section 2 have been applied, the results of the analysis are explained.

3.1. Correlation Analysis

Thus, initially, an analysis of correlations between the dependent and independent variables considered was conducted (see Figure 3). Thus, it was found that in general, the correlation of the variables in the model is low, with some groups of variables standing out in particular. Specifically, there is a significant expected correlation between the average sales price of the neighborhoods and that of the city. Likewise, significant expected correlations are identified between the urban landscape and the average rent of buildings or the average price of housing. Finally, the concentration of immigrants in a neighborhood is particularly correlated with the concentration of populations from Asia and Africa.

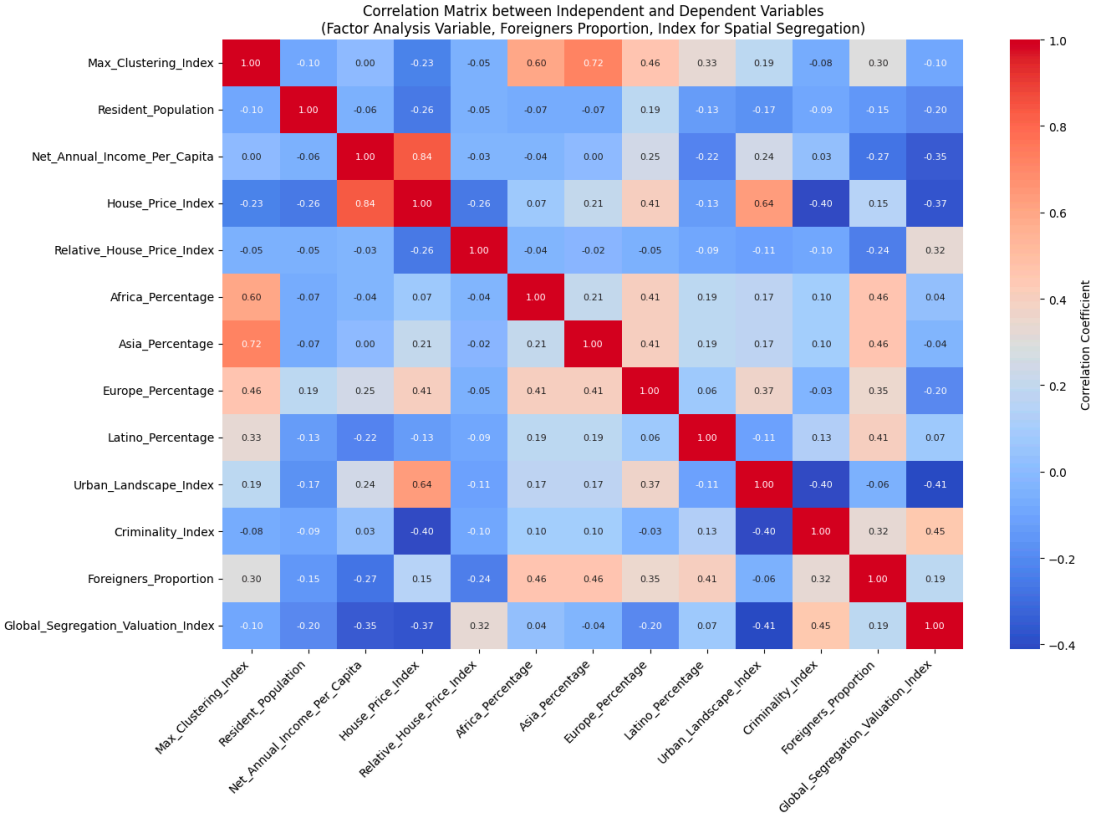


Figure 3. Table of correlations between the variables of the model. Source: author.

3.2. Factor Analysis Results

The results of the factor analysis (FA) showed that Factor 3 is strongly influenced by variables related to negative social and economic narrative (0.828), economic inequality (0.784), criminality (0.628), and community participation (0.409). This indicates that this factor is influenced by negative perceptions or discourses about the neighborhood. It is also noteworthy that most of the articles mention the structural inequality of the neighborhood. A large number of these neighborhoods cite insecurity due to high crime. Finally, the low involvement of residents suggests that the neighborhood is neglected, leading to poorer preservation. "Factor Analysis Variable," will be obtained from factor 3 and can be defined as a "Factor of negative social perception and structural inequality of a neighborhood" (see Table 2).

Table 2. Values returned by ChatGPT for Factor Analysis Variable (Factor 3).

Factor 3	Valor
Negative_Narrative_Index	0.828017
Economic_Inequality_Index	0.783713
Criminality_Index	0.628341
Community_Involvement_Index	0.409855
Images_Examples_Index	0.315339
Cultural_Representation_Index	0.256997
Incomplete_History_Index	0.204445
Balanced_Images_Index	0.190039
Diverse_Sources_Index	0.143111
Balanced_View_Index	0.104003
Public_Policies_Index	0.068489
Positive_Contributions_Index	−0.083090
Local_Success_Index	−0.118917
Updated_Statistics_Index	−0.195487

3.3. Regression Results Using the Factor Analysis Variable as Target Variable

On the other hand, the model run showed that ridge regression was the most efficient for the variable from the FA (factor of negative social perception and structural inequality of a neighborhood) and for the variable Foreigners Proportion, while random forest was the most outstanding for the index generated by ChatGPT from the information of the neighborhoods in Wikipedia.

Specifically, of the total tests performed, the most outstanding models for each of the three dependent variables were the following:

Results with ridge regression for the dependent variable (FA) Factor 3 (negative social perception factor and structural inequality of a neighborhood) are presented in Figure 4 and include the following parameters.

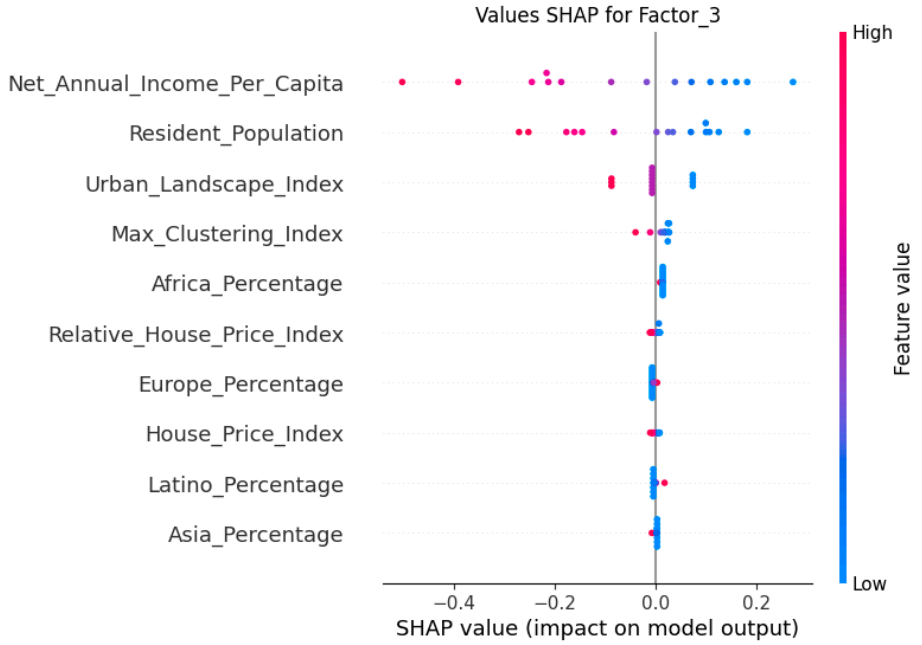


Figure 4. SHAP values for the dependent variable "Factor Analysis Variable (Factor 3)". Source: author.

R2: 0.2497
MSE: 1.2135
RMSE: 1.1016
MAE: 0.8894

Interpretation of the model by SHAP representation on the most prominent variables, i.e., those located from the top down.

- Net_Average_Income: High values in red have a negative impact on Factor 3, while low values in blue amplify it. From this, it can be inferred that the higher the income of the residents, the lower the contribution to Factor 3, indicating a negative relationship.
- Resident_Population shows a negative trend similar to that of average income, although its importance is less, as fewer observations are located at the extremes of the graph.
- The next variable in importance would be Urban_Landscape_Index. This parameter presents a lower importance and presents a negative relationship in relation to Factor Analysis Variable. It is found that the factor's negative social perception and structural inequality of a neighborhood increase as the landscape deteriorates.
- Max_Clustering_Index, which represents the concentration of immigrant populations in a neighborhood, also has a negative relationship, i.e., the higher the concentration, the higher the values of Factor 3.

The remaining variables have limited importance in the model.

Results with ridge regression for the dependent variable proportion of Foreigners Proportion are presented in Figure 5 and include the following parameters.

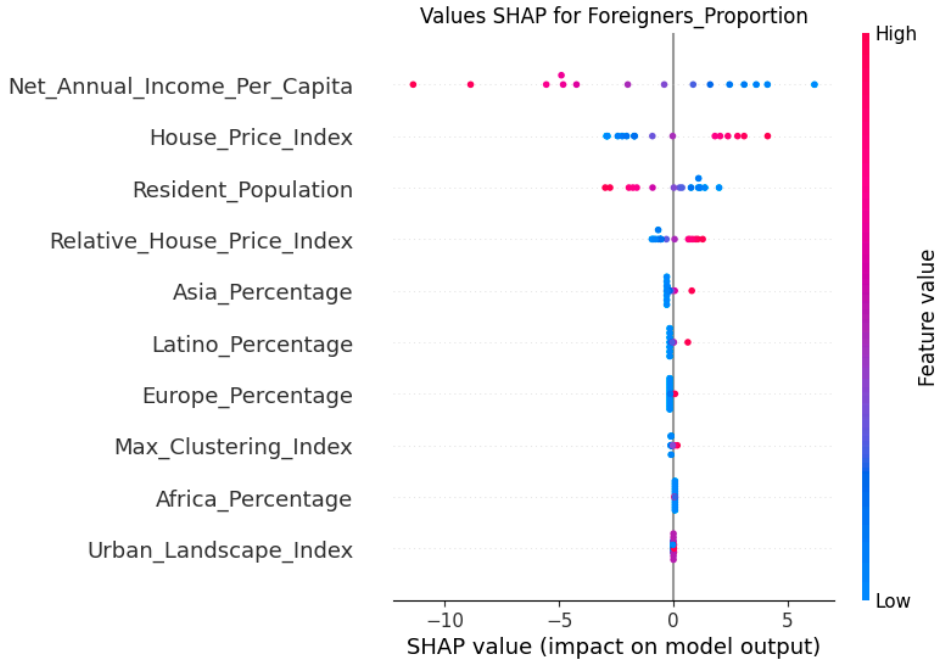


Figure 5. SHAP values for the dependent variable Foreigners Proportion. Source: author.

R2: 0.3282
MSE: 18.4581
RMSE: 4.2962
MAE: 3.7219

The following model exclusively values the accumulation of immigrants, regardless of whether they have become segregated spaces or not. The dependent variable is Foreigners Proportion. Thus, the interpretation of the most relevant variables according to the SHAP values, would be the following:

- Average_Neta_Income: As in the previous variable, the average income tends to be negative with respect to the immigration rate, so the lower the income, the higher the number of foreigners in the neighborhood, as expected, and conversely the higher the values the lower the immigration rate.
- House_Price_Index: This parameter has a positive relationship indicating that the higher the price of land the higher the proportion of foreigners residing in an area.
- Resident_Population: The volume of population has a negative and expected impact with respect to the volume of population. Thus, the immigrant population tends to be concentrated in sparsely populated neighborhoods.
- Percentage_Asian and Percentage_Latino: These variables have a moderate influence, with the percentage of Asians positively impacting the proportion of foreigners.

Random forest results for the dependent variable of ChatGPT with Wikipedia "Index for Spatial Segregation" are presented in Figure 6 and include the following parameters:

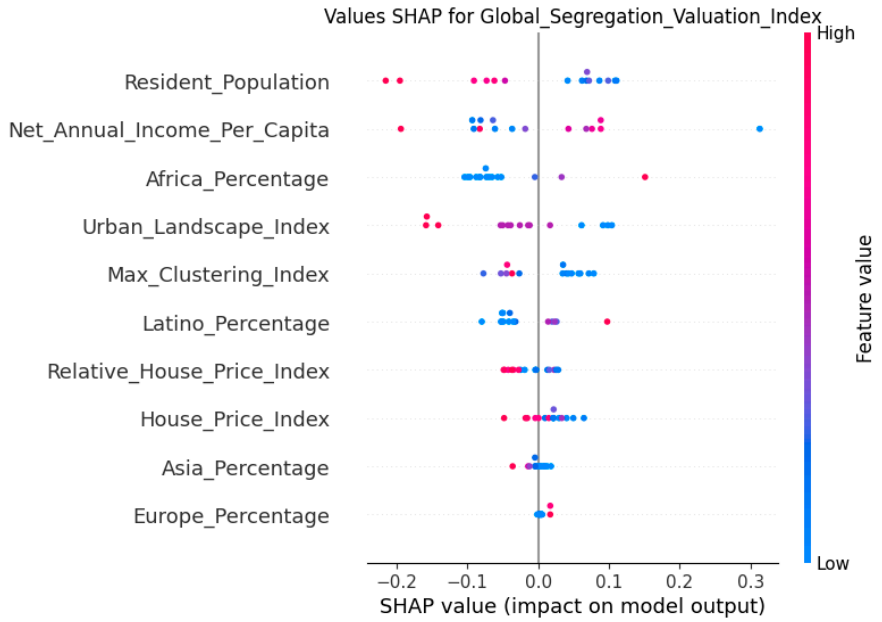


Figure 6. SHAP values for Index for Spatial Segregation. Source: author.

R2: 0.4096v
MSE: 0.2404
RMSE: 0.4903
MAE: 0.3643

Finally, we will proceed to analyze the spatial segregation parameter, which does consider the impact of the immigrant population on spatial segregation. Next, the most relevant variables will be interpreted according to the SHAP values, as with the previous models:

- Resident_Population: As in the previous model, a clear negative relationship between population volume and spatial segregation is seen. So, areas with less population tend to be more segregated compared to the rest of the city. It is also worth noting that since the variable is located at the top of the model, it is the most relevant parameter in the model.
- Average_Neta_Income: The average income variable, unlike the first model, presents a nonlinear relationship, where extremely high-income values tend to discourage neighborhood segregation and extremely low values tend to encourage it, but there is also an intermediate situation in which moderately high-income values can favor segregation and moderately low values discourage it.
- Urban_Landscape_Index: This parameter, like the previous ones, presents a negative relationship with respect to social segregation, where poorer landscape quality is associated with greater neighborhood segregation.

- Max_Clustering_Index also shows a nonlinear relationship, where a strong concentration of foreigners from the same geographical origin does not necessarily lead to spatial segregation in the neighborhood.
- Africa_Percentage and Latino_Percentage: Both parameters present a clear positive relationship with respect to the segregation of spaces, being especially prominent in the case of African populations.

Finally, the average price of housing and its relative price with respect to the city's total is also negative, although it is one of the parameters with the least influence on the model.

In comparing the models, it is worth noting that they exhibit acceptable evaluation parameters (MSE, RMSE, and MAE), though the explanation of variability is somewhat limited. In this regard, it is worth mentioning that the model with the highest predictive capacity is the one that includes the general assessment of the neighborhoods based on ChatGPT as the dependent variable.

4. Discussion and Conclusions

The analysis of the model results allows us to draw several conclusions about the comparison between these approaches. The comparative analysis of the two types of variables shows that the variable generated by ChatGPT's textual interpretation has contributed to building a more predictive model than those using variables processed through factor analysis or those that explain neighborhood segregation by simply grouping immigrants.

Average income is a relevant variable in all the models. In general, it has a negative relationship with both the proportion of immigrants and segregation, although this relationship is not linear in the ChatGPT general valuation model. This finding confirms conclusions from previous studies that the proportion of immigrants and segregation are higher in lower-income neighborhoods (Jiménez et al. 2020).

Resident population is also a significant variable in all three models, although it is more influential in the models that assess segregation than in those analyzing the percentage of immigrants in neighborhoods. This difference can be explained by the fact that neighborhoods with smaller populations are more easily segregated and transformed than those with larger populations. In larger neighborhoods, a significant proportion of the population must be segregated to effect a noticeable change, while in smaller neighborhoods, transformation is simpler. This is a novel finding compared to the existing literature. Although it aligns with the results of Bellman et al. (2018) and Logan et al. (2023), which do not establish a direct relationship between total neighborhood population and segregation.

The urban landscape has a significant impact on the Index of Segregated Spaces variable, which is consistent with the findings of Vargas (2020). However, this situation is nuanced, as the concentration of foreigners has a low correlation with the landscape valuation parameter, as shown in Figure 3. This impact likely results from a long-term interaction between immigrant concentration and landscape degradation.

Neighborhoods with more degraded or poorly maintained urban landscapes tend to become less attractive to the broader population, ultimately leading to a reduction in property prices and, consequently, greater availability of affordable housing for low-income residents, including immigrants. As a result, degradation fosters long-term segregation.

On the other hand, the lack of impact of the landscape valuation variable on the dependent variable proportion of foreigners may seem contradictory, but it can be explained. In many cases, neighborhoods with a high percentage of foreign residents host long-term international students, workers, and tourists alongside other immigrants. These populations often reside in areas with high landscape quality. Therefore, it is logical that the model does not find this parameter significant for the proportion of foreigners, although it is significant for the Spatial Segregation Index.

Furthermore, it should be noted that two tools related to big data and AI have been essential in reaching this conclusion. The use of Google Street View alongside AI analysis has allowed us to assess the importance of this variable. Obtaining these data through conventional methods would have been time-consuming and costly, making it impractical.

Additionally, this research highlights the impact of landscape degradation on the segregation of spaces, which contrasts with the cultural appropriation of the landscape discussed by LaFleur (2021) and Knorr (2017). It would have been interesting to introduce an assessment of the impact of other elements of the urban landscape, such as accessibility to products and services specific to immigrant cultures. This aligns with the findings of den den Butter et al. (2004) regarding the closed economies of ethnic groups within neighborhoods. In relation to this, while Roberto and Hwang's (2015) contributions on the barrier effect was not considered in the model, many neighborhoods with a high concentration of immigrants are not physically separated, as seen in the maps in Figures 1 and 2, and therefore, this variable was not included in the model.

This research has also confirmed the relative importance of local networks of compatriot immigrants, although the result is not homogeneous across all the models. The conclusion drawn is that the concentration of the population of African origin tends to create more spatial segregation than immigrants from other areas such as America or Asia. In fact, as can be seen in Figures 1 and 2, the degree of spatial concentration is lower for immigrants from the Americas and Asia, while it is higher for immigrants from the African continent. Language difficulties may be a factor that contributes to African immigrants being more affected by social and economic barriers, which limits their opportunities for residential mobility and access to more diverse areas (Espaliú 2022). All of this can lead to a concentration of the population of African origin in neighborhoods where compatriots predominate, forming segregated enclaves. These results are consistent with those obtained by Pan Ké Shon (2010) and Tesfai and Thomas (2020); however, they also represent a contribution to the literature in that, to our knowledge, the impact of the concentration of immigrants of various ethnic groups had not been compared on a large scale as in this study.

Finally, it should be noted that the price of land has not shown a significant relationship with the indexes that focus specifically on segregation, but it has shown a significant and positive relationship with the dependent variable Foreigners Proportion, which measures the raw percentage of immigrants in the city. This conclusion contradicts the findings of authors such as Skifter Andersen et al. (2015) and Trounstine (2020).

A plausible explanation may lie in the fact that the Foreigners Proportion variable includes long-term tourists, foreign long-term high-skilled workers, and international students who are less sensitive to housing prices, as they tend to live in expensive central areas with high-quality landscapes due to their higher incomes. On the other hand, the presence of immigrants, presumably from lower-income backgrounds, in high-price housing areas may be explained by other factors that make them less sensitive to land prices, despite having significantly lower incomes. Specifically, factors such as social support networks of compatriots in certain neighborhoods, illegal occupation of vacant housing, or overcrowding may mitigate the impact of land prices on residential choices. Additionally, the degradation of the urban landscape can significantly reduce the price of specific buildings. It should also be noted that the sale price of housing is an average applied to the entire neighborhood, not to individual buildings, which may have significantly lower prices due to their condition. In fact, the correlation between the urban landscape and land price is the strongest in the correlation table shown in Figure 3, so severe landscape degradation should have a very negative impact on land prices.

Therefore, land price does not always act as a direct predictor of segregation, as it may be mediated by other factors such as the urban landscape, spatial concentration of compatriots, overcrowding, and squatting. Thus, this analysis contributes to the literature by qualifying the direct relationship between land price and segregation, showing that other social factors—such as compatriot networks, squatting, and the degradation of the urban landscape—may mediate this relationship.

Finally, the use of Wikipedia as a data source for identifying segregated spaces is a valuable tool for practitioners and academics in the field of spatial segregation management and analysis. It not only helps to identify segregated areas but also the causal factors

behind them, which can aid in mitigating segregation. In this regard, it is important to note that both the areas and the factors contributing to segregation can evolve over time. Knowledge sources such as Wikipedia, which evolve relatively quickly, can thus be valuable for managing these spaces. Moreover, Wikipedia appears more reliable than other sources, such as the INE, for identifying segregated neighborhoods, since neighborhood segregation is not always accurately captured by generic data on immigrants, as observed during model testing. In any case, the analysis of Wikipedia information would not have been possible without the help of AI Natural Language Processing (NLP) technology, which facilitates text comprehension. In this regard, a generative AI like ChatGPT can interpret and execute NLP-processed commands. Without these technologies, this work would have been enormous and unfeasible. Therefore, the technological capabilities of AI can be of great use in analyzing segregated spaces using various textual sources.

Among the main limitations of this research, two should be highlighted. First, the sample size is not particularly large due to the lack of data from all neighborhoods in Madrid and Barcelona from Wikipedia. Additionally, the inclusion of some neighborhoods in Spain that slightly exceed the 25% threshold would have provided observations from other Spanish cities, which might have altered the models. Nevertheless, the current database accounts for 80% of all neighborhoods in Spain that meet this condition, as previously mentioned.

Although the models' predictability is reasonable, there is significant room for improvement in explaining data variability. Introducing additional factors could enhance predictability.

Another limitation is that while the information consulted from Wikipedia was obtained in 2024, the information requested and used for analysis is constrained by the last update, which may not reflect the current date or have been updated at all.

As future lines of research, an analysis of the references used by Wikipedia could be undertaken to check their consistency with the statements made in the articles, identifying possible positive or negative biases. This could be a limitation in the current study's database and could serve as a case study on the reliability of Wikipedia as a knowledge base.

Author Contributions: All the authors participated in the conception and design of the work. J.L.-O. analyzed the data, Á.O.-S. analyzed everything related to Wikipedia, A.G.-N. elaborated cartography. All the authors wrote the first draft. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The authors declare that the Institutional Review Board Statement is not necessary.

Informed Consent Statement: The authors declare that no Declaration of Informed Consent is necessary.

Data Availability Statement: The authors declare that the data are not available.

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

References

Aghaebrahimian, Ahmad, Andy Stauder, and Michael Ustaszewski. 2022. Testing the validity of Wikipedia categories for subject matter labelling of open-domain corpus data. *Journal of Information Science* 48: 686–700. [CrossRef]

Anselin, Luc. 2013. *Spatial Econometrics: Methods and Models*. Berlin/Heidelberg: Springer Science & Business Media, vol. 4.

Aydın, Ömer, and Enis Karaarslan. 2022. OpenAI ChatGPT Generated Literature Review: Digital Twin in Healthcare. In *Emerging Computer Technologies 2*. Edited by Ömer Aydın, İzmir Akademi Dernegi and Emerging Computer Technologies. İzmir: İzmir Akademi Dernegi, pp. 22–31. [CrossRef]

Aziz, Mohamed Alkhuzamy, Ahmed El-Zeiny, Fayrouz. M. Hassan, and Doaa M. Naguib. 2024. Coastal water quality dynamics of the Red Sea, southeast coast of Egypt using GeoAI and ChatGPT. *Journal of African Earth Sciences* 219: 105409. [CrossRef]

Baldrich, Kevin, and Juana Celia Domínguez-Oller. 2024. El uso de ChatGPT en la escritura académica: Un estudio de caso en educación [The use of ChatGPT in academic writing: A case study in Education]. *Pixel-Bit, Revista de Medios y Educación* 71: 141–57. [CrossRef]

Bellman, Benjamin, Seth E. Spielman, and Rachel S. Franklin. 2018. Local Population Change and Variations in Racial Integration in the United States, 2000–2010. *International Regional Science Review* 41: 233–55. [CrossRef] [PubMed]
Bermúdez-Quimis, Neptalí Cristóbal. 2024. Inteligencia artificial y gobiernos de la información. *Revista Científica INGENIAR: Ingeniería, Tecnología e Investigación* 7: 282–88. Available online: https://journalingeniar.org/index.php/ingeniar/article/view/219 (accessed on 28 November 2024).
Breiman, Leo. 2001. Random Forests. *Machine Learning* 45: 5–32. [CrossRef]
Chen, Tianqi, and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. Paper presented at 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13–17; pp. 785–94. [CrossRef]
Cordero-Guzmán, Héctor R. 2005. Community-based organizations and migration in New York City. *Journal of Ethnic and Migration Studies* 31: 889–909. [CrossRef]
Crowder, Kyle, Matthew Hall, and Stewart E. Tolnay. 2011. Neighborhood Immigration and Native Out-Migration. *American Sociological Review* 76: 25–47. [CrossRef]
Cuena, Matilde. 2023. La ocupación ilegal de inmuebles: Un necesario enfoque global. *Cuadernos de Derecho Transnacional* 15: 293–336. [CrossRef]
Deltell, Luis, and Florencia Claes. 2021. Conocimiento libre en tiempos de pandemia. Estudio de los artículos "COVID-19" y "Pandemia por COVID-19" en Wikipedia*. *Interface Comunicação, Saúde, Educação* 25: e200329. [CrossRef]
den Butter, Frank. A. G., Enno Masurel, and Robert H. J. Mosch. 2004. *The Economics of Co-Ethnic Employment: Incentives, Welfare Effects and Policy Options*. Tinbergen Institute Discussion Paper No. TI 2004-027/3. Amsterdam: Tinbergen Institute, Amsterdam and Rotterdam. [CrossRef]
Dinesen, Peter Thisted, and Kim Mannemar Sønderskov. 2015. Ethnic diversity and social trust: Evidence from the micro-context. *American Sociological Review* 80: 550–73. [CrossRef]
Espaliú, Carlos. 2022. La labor de los centros de participación e integración de inmigrantes (CEPI) en la prevención, detección y canalización de problemas y conflictos derivados del empleo del idioma español en el contexto de la inmigración en la Comunidad de Madrid. *Lengua y Migración* 14: 39–60. Available online: http://hdl.handle.net/10017/53097 (accessed on 3 October 2024). [CrossRef]
Friedman, Jerome H. 2001. Greedy function approximation: A gradient boosting machine. *Annals of Statistics* 29: 1189–232. [CrossRef]
Fuertes, Marc. 2024. Enmarcando las aplicaciones de IA generativa como herramientas para la cognición en educación [Framing Generative AI applications as tools for cognition in education]. *Pixel-Bit, Revista de Medios y Educación* 71: 42–57. [CrossRef]
Gabashvili, Irene S. 2023. The impact and applications of ChatGPT: A systematic review of literature reviews. *arXiv* arXiv:2305.18086. [CrossRef]
García-Hernández, Juan Samuel. 2020. Territorial stigmatization in the neoliberal city: Citizen perceptions and reactions in the social peripheries of Santa Cruz de Tenerife (Canary Islands, Spain). *Boletín de la Asociación de Geógrafos Españoles* 86. [CrossRef]
Giménez-Bertomeu, Víctor M. dir. 2020. *Vulnerabilidad Territorial: Indicadores Para su Medición Desde los Servicios Sociales*. Alicante: Limencop. Available online: http://hdl.handle.net/10045/103667 (accessed on 2 October 2024).
Google Street View. 2024. Street View. Available online: https://www.google.com/intl/es/streetview/ (accessed on 3 October 2024).
Harvey, David. 2012. *Rebel Cities: From the Right to the City to the Urban Revolution*. London: Verso.
Hastie, Trevor, Robert Tibshirani, and Martin Wainwright. 2015. Statistical learning with sparsity. In *Monographs on Statistics and Applied Probability*. Boca Raton: CRC Press, vol. 143. [CrossRef]
Idealista. 2024. Buscar Valoraciones e Información de una Vivienda. Available online: https://www.idealista.com/maps/ (accessed on 3 October 2024).
Instituto Nacional de Estadística. 2024a. Atlas de Distribución de Renta de los Hogares. Available online: https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736177088&menu=ultiDatos&idp=1254735976608 (accessed on 3 October 2024).
Instituto Nacional de Estadística. 2024b. Censos de Población y Viviendas. Últimos datos. Available online: https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176992&idp=1254735572981 (accessed on 3 October 2024).
Javaid, Mohd, Abid Haleem, and Ravi Pratap Singh. 2023. ChatGPT for healthcare services: An emerging stage for an innovative perspective. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations* 3: 100105. [CrossRef]
Jiménez, Beatriz Cristina, Rosa Resino, Milagros Mayoral, and Silvana Sassano. 2020. Inmigración y segregación residencial en la ciudad de Madrid. *Anales de Geografía de la Universidad Complutense* 40: 393–418. [CrossRef]
Jones, Antwan. 2011. Income, Homeownership and Residential Assorting among Latinos in the US. *Advances in Applied Sociology* 1: 1–12. [CrossRef]
Juhász, Levente, Peter Mooney, Hartwig H. Hochmair, and Boyuan Guan. 2023. ChatGPT as a mapping assistant: A novel method to enrich maps with generative AI and content derived from street-level photographs. In *Spatial Data Science Symposium 2023 Paper Proceedings*. Santa Barbara: UC Santa Barbara Center for Spatial Studies. [CrossRef]
Knorr, Lilian. 2017. Divided landscape: The visual culture of urban segregation. *Landscape Journal* 35: 109–26. Available online: https://faculty.uml.edu/sgallagher/Divided_Landscape_Visual_Culture_of_Urban_Segregation.pdf (accessed on 3 October 2024). [CrossRef]
Koopmans, Ruud, and Merlin Schaeffer. 2016. Statistical and Perceived Diversity and Their Impacts on Neighborhood Social Cohesion in Germany, France and the Netherlands. *Social Indicators Research* 125: 853–83. [CrossRef]
Lacayo, Celia. 2016. Latinos Need to Stay in Their Place: Differential Segregation in a Multi-Ethnic Suburb. *Societies* 6: 25. [CrossRef]

LaFleur, Jennifer. 2021. The Race That Space Makes: The Power of Place in the Colonial Formation of Social Categorizations. *Sociology of Race and Ethnicity* 7: 512–26. [CrossRef]
Laurence, James, and Lee Bentley. 2016. Does ethnic diversity have a negative effect on attitudes towards the community? A longitudinal analysis of the causal claims within the ethnic diversity and social cohesion debate. *European Sociological Review* 32: 54–67. [CrossRef]
Lefebvre, Henri. 1991. *The Production of Space*. Hoboken: Wiley-Blackwell.
Li, Jianning, Amin Dada, Behrus Puladi, Jens Kleesiek, and Jan Egger. 2024. ChatGPT in healthcare: A taxonomy and systematic review. *Computer Methods and Programs in Biomedicine* 245: 108013. [CrossRef] [PubMed]
List of Wikipedias. 2024. En Wikipedia. Available online: https://meta.wikimedia.org/w/index.php?title=List_of_Wikipedias&oldid=27488289 (accessed on 20 September 2024).
Livberber, Tuba. 2023. Toward non-human-centered design: Designing an academic article with ChatGPT. *Profesional de la Información* 32: e320512. [CrossRef]
Logan, John R., Samuel Kye, H. Jacob Carlson, Elisabeta Minca, and Daniel Schleith. 2023. The Role of Suburbanization in Metropolitan Segregation After 1940. *Demography* 60: 281–301. [CrossRef]
Lundberg, Scott, and Su-In Lee. 2017. A unified approach to interpreting model predictions. *arXiv* arXiv:1705.07874. [CrossRef]
Matossian, Brenda. 2018. Debates sobre la segregación urbana: Una revisión teórico-metodológica. *Estudios Sociales Contemporáneos* 19: 117–38.
McCann, Wesley S., and Francis D. Boateng. 2020. An Examination of American Perceptions of the Immigrant-Crime Relationship. *American Journal of Criminal Justice* 45: 973–1002. [CrossRef]
Ministerio de Vivienda y Agenda Urbana 2024. Observatorio de la Vulnerabilidad Urbana. Available online: https://www.mivau.gob.es/urbanismo-y-suelo/urbanismo/observatorio-de-la-vulnerabilidad-urbana (accessed on 2 October 2024).
Molnar, Christoph. 2020. *Interpretable Machine Learning*. Morrisville: Lulu Press.
Montenegro-Rueda, Marta, José Fernández-Cerero, José María Fernández-Batanero, and Eloy López-Meneses. 2023. Impact of the Implementation of ChatGPT in Education: A Systematic Review. *Computers* 12: 153. [CrossRef]
Musterd, Sako, and Wim Ostendorf. 1998. *Urban Segregation and the Welfare State: Inequality and Exclusion in Western Cities*. Milton Park: Routledge.
Musterd, Sako, Szymon Marcińczak, Maarten van Ham, and Tiit Tammaru. 2016. Socioeconomic segregation in European capital cities. *Increasing Separation Between Poor and Rich. Urban Geography* 38: 1062–83. [CrossRef]
Obregón, Ángel, and Natalia González. 2018. La Wikipedia en las facultades de educación españolas. Diseño y validación de herramientas diagnósticas cuantitativas y cualitativas. *Revista Iberoamericana de Educación* 77: 55–76. [CrossRef]
Obregón, Ángel, and Natalia González. 2021. ¿Utilizan correctamente Wikipedia los estudiantes universitarios españoles? *Revista Lusófona de Educação* 53: 159–74. [CrossRef]
Obregón-Sierra, Ángel, Javier López-Otero, Antonio Gavira-Narváez, and Rafael Fernando Vega-Pozuelo. 2023. Methodology for the incorporation of geographic information in Wikidata. *Revista de Estudios Andaluces* 45: 260–64. [CrossRef]
OpenAI. 2024. Improvements to Data Analysis in ChatGPT. Available online: https://openai.com/index/improvements-to-data-analysis-in-chatgpt/ (accessed on 16 May 2024).
Owens. 2019. Building inequality: Housing segregation and income segregation. *Sociological Science* 6: 497. Available online: https://www.sociologicalscience.com/download/vol-6/august/SocSci_v6_497to525.pdf (accessed on 3 October 2024).
Pageviews. 2024. Siteviews Analysis. Available online: https://pageviews.wmcloud.org/siteviews/ (accessed on 3 October 2024).
Pan Ké Shon, Jean-Louis. 2010. The Ambivalent Nature of Ethnic Segregation in France's Disadvantaged Neighbourhoods. *Urban Studies* 47: 1603–23. [CrossRef]
Roberto, Elisabeth, and Jackelyn Hwang. 2015. Barriers to integration: Physical boundaries and the spatial structure of residential segregation. *arXiv* arXiv:1509.02574. [CrossRef]
Rodríguez-García, Dan. 2022. The Persistence of Racial Constructs in Spain: Bringing Race and Colorblindness into the Debate on Interculturalism. *Social Sciences* 11: 13. [CrossRef]
Sampson, Robert J. 2012. *Great American City: Chicago and the Enduring Neighborhood Effect*. Chicago: University of Chicago Press.
Simone, AbdouMaliq. 2018. *Improvised Lives: Rhythms of Endurance in an Urban South*. Cambridge: Polity Press.
Singh, Harjit, and Avneet Singh. 2023. ChatGPT: Systematic Review, Applications, and Agenda for Multidisciplinary Research. *Journal of Chinese Economic and Business Studies* 21: 193–212. [CrossRef]
Singh, Harshdeep, Robert West, and Giovanni Colavizza. 2021. Wikipedia citations: A comprehensive data set of citations with identifiers extracted from English Wikipedia. *Quantitative Science Studies* 2: 1–19. [CrossRef]
Skifter Andersen, Hans, Roger Andersson, Terje Wessel, and Katja Vilkama. 2015. The impact of housing policies and housing markets on ethnic spatial segregation: Comparing the capital cities of four Nordic welfare states. *International Journal of Housing Policy* 16: 1–30. [CrossRef]
Sýkora, Luděk, and Stefan Bouzarovski. 2012. Multiple Transformations: Conceptualising the Post-communist Urban Transition. *Urban Studies* 49: 43–60. [CrossRef]
Tao, Ran, and Jinwen Xu. 2023. Mapping with ChatGPT. *ISPRS International Journal of Geo-Information* 12: 284. [CrossRef]
Tejedor, Joan Miguel. 2024. Inteligencia artificial para la investigación económica. *Revista Finanzas y Política Económica* 16: 9–16. [CrossRef]

Tesfai, Rebbeca, and Kevin J. A. Thomas. 2020. Dimensions of Inequality: Black Immigrants' Occupational Segregation in the United States. *Sociology of Race and Ethnicity* 6: 1–21. [CrossRef]

Trounstine, Jessica. 2020. The Geography of Inequality: How Land Use Regulation Produces Segregation. *American Political Science Review* 114: 443–55. [CrossRef]

Vaalavuo, Maria, Maarten van Ham, and Timo M. Kauppinen. 2017. *Income Increase and Moving to a Better Neighbourhood: An Enquiry into Ethnic Differences in Finland*. IZA Discussion Paper; No. 11076. Bonn: Forschungsinstitut zur Zukunft der Arbeit/Institute for the Study of Labor (IZA). Available online: https://pure.tudelft.nl/ws/portalfiles/portal/51457256/dp11076.pdf (accessed on 3 October 2024).

Van Ham, Maarten, Tiit Tammaru, Rūta Ubarevičienė, and Heleen Janssen. 2021. *Urban Socio-Economic Segregation and Income Inequality: A Global Perspective*. Cham: Springer. [CrossRef]

Vargas, María Dolores. 2020. ¿Somos racistas? Caminando a través de los conceptos de racismo, exclusión y segregación urbana y social. In *Racismo, Etnicidad e Identidad en el Siglo XXI*. Coordinated by Rachid El Quaroui. Badajoz: AntropiQa 2.0, pp. 79–99. Available online: https://dialnet.unirioja.es/descarga/libro/778912.pdf (accessed on 3 October 2024).

Zapata-Barrero, Ricard. 2024. De-bordering policies at the city scale: Strategies for building resilience in Barcelona's migration governance. *Comparative Migration Studies* 12: 2. [CrossRef]

Zhang, Haifeng, Yuqing Ren, and Robert E. Kraut. 2018. *Mining and Predicting Temporal Patterns in the Quality Evolution of Wikipedia Articles*. Valhalla: Academy of Management. [CrossRef]

Zou, Hui, and Trevor Hastie. 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B Statistical Methodology* 67: 301–20. [CrossRef]

Zubimendi, Mikel, Jorge Julián Cueto, Marien Béguelin, and Fernando M. Archuby. 2024. Excavando Wikipedia: Apuntes sobre la comunicación pública de la arqueología en una enciclopedia libre y colaborativa. *Revista del Museo de Antropología* 17: 189–206. [CrossRef]

Zukin, Sharon. 2010. *Naked City: The Death and Life of Authentic Urban Places*. Oxford: Oxford University Press.

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.