uc3m | Universidad Carlos III de Madrid

Master Degree in Computational Social Sciences

2022/2023

Master Thesis

"Cluster analysis approach to the relation between the neighborhoods in Madrid and gentrification"

Jorge Pascual Segovia

Supervisor: Gonzalo Génova Fuster

Madrid, 2023

ABSTRACT

This thesis investigates gentrification in Madrid, focusing on identifying at-risk neighborhoods and understanding their relationship to the gentrification process. By analyzing demographic characteristics, socioeconomic indicators, housing prices, and Airbnb data primarily obtained from the Ayto. de Madrid (Madrid City Council), a cluster analysis was performed. The findings reveal the spatial distribution of high-risk gentrification areas and provide insights into influential factors. This study enhances the understanding of gentrification dynamics and patterns in Madrid, contributing to the broader knowledge of the process and its identification.

Associated public repository:
https://github.com/JorPS/Cluster-analysis-Gentrification-in-Madrid

1. THEORETICAL FRAMEWORK: A CONCISE OVERVIEW OF THE PHENOMENON OF GENTRIFICATION

The Right to the city is defined by Henri Lefebvre (1967) as the right of the citizens to build, create and decide on their city. This is not recognized as a human right, but the concept is strictly linked with the idea of democracy and social equality: every citizen should have the same right to create, build and decide on their city. Gentrification is a problem as it represents class privileges and creates disparities between classes regarding the ability to choose where to live, when to stay, and who can benefit from residing in an area with lucrative potential. Moreover, it fundamentally affects the right to shape the direction of the city's development.

Gentrification is the historical social transformation of city centers that began in certain cities in Great Britain and the USA during the latter part of the 20th century, such as Manhattan. The emerging middle classes with highly valued professional skills, associated with the global economy, known as the 'gentry' in Great Britain, gave their name to this process as they began demanding neighborhoods in city centers. Despite these neighborhoods typically being traditional working-class areas, the city underwent institutional changes to facilitate the gentry's residence in these areas, often compelling existing residents to disperse and relocate throughout the city, away from their homes. Nowadays, this process has expanded across the Western world with the rise of this social class, affecting cities that exhibit the characteristics of a 'global city' deeply integrated into the global economy network. (Sorando & Ardura, 2018) (Sequera, 2014) (Walliser & Sorando, 2019).

There are many examples of this phenomenon, back in the late 20th Century, but also recently in Spain, where it’s supposing an actual problem nowadays, caused by touristification, real estate pressure and speculation. El Raval (Barcelona), Malasaña (Madrid) and Lavapiés (Madrid) are common examples of gentrified neighborhoods in Spain. According to Sorando & Ardura (2016), gentrification has an identifiable process they describe as the *creative destruction of the city*:

1. First, there is a working class city center area of interest inside a globalized city which can be economically exploited by companies and demanded by this new kind of middle class.

2. The government also has plans to transform a specific area, due to the potential benefits it can bring and its ability to attract companies with substantial capital. In the case of el Raval, the old “Barrio Chino”, even a company funded by the government in a half, while the other half was holded by companies as BBVA or Telefónica, is the responsible of administering and coordinating the renewal plans of the most humble neighborhoods in Barcelona[1].

3. The stigmatization of neighborhoods and the decline of their quality of life can result from various factors, including heightened police presence, or the lack of investments in neighborhood issues can contribute to its gradual deterioration. The presence of poverty, insecurity, and increased police presence further reinforces the perception of a neighborhood as "dangerous" or "problematic," thus legitimizing interventions and fostering apathy among its residents.

4. Once public authorities have gained credibility and the ability to intervene in these neighborhoods, the actions taken lead to the displacement of residents. This displacement can occur through various means, such as pressuring inhabitants to leave, negotiating with companies on unfavorable terms for the sale of their houses, and ultimately resorting to expropriation of properties at prices that may not adequately reflect their true value.

5. Finally, the estates are liberalized and sold at market price, with promising investments to renew the neighborhood, attracting the capital from big companies.

6. The consequences of gentrification are the dispersion of poverty, segregation and the occupation of inhabitants by a privileged class.

---

[1] Foment de Ciudad, SA, official web page: Foment de Ciudad, SA, es una empresa municipal especializada en la gestión de proyectos integrales, transversales y con implicación ciudadana y territorial que lidera proyectos de ciudad de gran trascendencia | Fomento de Ciudad (barcelona.cat)

2. RESEARCH SUMMARY

This thesis aims to gather information on various variables related to gentrification and utilize the data to conduct a cluster analysis that enables the grouping of different neighborhoods based on their association with gentrification. The primary objective is to identify areas that are more susceptible to the occurrence of gentrification and define the relationship between neighborhoods and the gentrification process in Madrid.

Objectives:

- ➢ To compile relevant data on variables related to gentrification, including but not limited to demographic characteristics, socioeconomic indicators, housing prices, and urban development.
- ➢ To perform a cluster analysis on the collected data in order to identify distinct patterns and group neighborhoods based on their level of association with gentrification.
- ➢ To determine the areas most susceptible to gentrification in Madrid by identifying clusters characterized by indicators commonly associated with the process.
- ➢ To establish a clear relationship between neighborhoods and the occurrence of gentrification, providing insights into the spatial distribution and dynamics of this phenomenon in Madrid.

Hypotheses:

- ➢ Neighborhoods with lower education levels, higher unemployment rates and a higher Population Aging index will exhibit a higher likelihood of experiencing gentrification.
- ➢ Neighborhoods in close proximity to city centers, transportation hubs, and areas with high cultural amenities, and thus experiencing a significant tourist influx, are likely to exhibit a stronger association with gentrification.
- ➢ Areas with significant urban redevelopment projects and rising property prices will be more prone to gentrification.

The findings of this thesis will contribute to the existing literature on gentrification by providing a comprehensive analysis of the factors and neighborhood characteristics associated with this urban process. By identifying areas at higher risk of gentrification, this research will support policymakers, urban planners, and stakeholders in implementing

targeted interventions and strategies to mitigate the negative consequences and promote inclusive development in Madrid. These contributions could have a profound and positive impact on addressing and mitigating the growing inequalities among citizens. Furthermore, it can help reduce the dispersion of poverty and provide better opportunities for underprivileged classes and marginalized groups who are at risk of social exclusion.

This unsupervised model is expected to be useful for identifying and preventing this process to occur in an area, protecting the inhabitants's interests and right to the city, as they are the essence and the main components of the neighborhoods. Also, this thesis will provide important information about the different neighborhoods to decide on which of them are more appropriate for treatment experiments or testing urban policies.

3. METHODOLOGY

The methodology of this thesis is to develop a cluster analysis for neighborhoods in Madrid that detects the vulnerability of working class neighborhoods in each of the clusters, by grouping neighborhoods by similarity according to variables that are strongly related to gentrification, i.e. their vulnerability or risk for being gentrified, selected after an exhaustive revision of the literature. Based on the revision done on the sources mentioned below, I decided to select the following list of variables, for the moment, as they respond to a causal explanation of the gentrification phenomenon, according to the bibliography:

- Education level. For getting insights about the social class structure in the different neighborhoods and clusters.
- Immigration proportion and growth of immigration from the previous 4 years.
- Real estate prices and its increment since the previous 4 years.
- Airbnb prices, presence and their increment during the previous 4 years. It will give valuable information about the touristic attractiveness of the areas.
- Unemployment and its growth the past 4 years. For getting insights about the social class structure in the different neighborhoods and clusters.
- Loss of population (Walliser & Sorando, 2019).
- Population aging (Sequera, 2014).

The causality established by theoretical perspectives of the revised literature will be tested statistically to consolidate or discuss the assumed causality depending on the patterns found. With that being said, the present thesis does not engage in an examination of the veracity or existence of gentrification as a prevalent issue, as this topic has undergone extensive scrutiny and comprehensive examination, thus affirming the existence of such a phenomenon within the Spanish context, even though there are political and economical interests behind this matter that triggers popular but not scientific controversy (Rubiales, 2014).

Clustering methodology

For the clustering analysis, I chose to use the k-means method, as it fits properly with numerical data, a wide number of variables and fixed clusters, which is adequate for my objective. Hierarchical clustering will also be tested, as the k-means method is limited to the assumption of equal sized spherical clusters, which is not the case as we will see. The hierarchical clustering model was performed using the Euclidean distance by Ward method. Both clustering models will be compared and evaluated to choose the method that better minimizes the error, as well as the differences within clusters, and better differentiate between them.

Before performing the different analysis on the data, some variables were transformed to avoid scale problems and avoid bias on interpreting the results. Variables that showed and skewed distribution underwent a logarithmic transformation. This is the case of the Real Estate prices, number of Airbnb listings and its growth. Robust scaling was applied over other variables that showed distributions prone to outliers, these variables were the growth of the population and foreign proportion during the previous 4 years. The Airbnb mean price growth showed a normal distribution and was scaled by a *z score* standardization.

Data processing strategy

Using RStudio with R language, I stored the data available of the already mentioned variables, in the years 2015 and 2019 (in order to compute the increment of certain variables) that are involved in the gentrification process of a neighborhood. R allows me to compute new variables or summarize them, as well as merge all of the data together, develop the cluster analysis and visualizations of the results. In the Appendix 1, it can be seen the source used in this thesis for each of the variables.

It should be taken into account that there are significant limitations in obtaining up-to-date open information about variables related to gentrification. This is primarily due to an apparent lack of interest by Spanish institutions in collecting and regularly publishing data, especially regarding the social class structure, such as occupational data (Rubiales ,2014).

There were several decisions that had to be taken due to the format of the data available and its characteristics when manipulating and merging the data. The diverse data sources

presented a challenge in terms of compatibility, as they varied in file types, structures, and naming conventions. In order to ensure smooth data integration, careful consideration was given to selecting the appropriate data transformation techniques. These decisions were crucial in maintaining data integrity and ultimately providing a solid foundation for further analysis and decision-making.

Some neighborhoods had different denominations in the different datasets, for example Peñagrande (Peña Grande); a name changing throughout the years like Palos de Moguer to Palos de la Frontera and some others were new in 2019, which is the case of Ensanche de Vallecas. To solve the denomination problems, I used the neighborhood code columns to identify the specific denomination of the neighborhoods with this issue and built string manipulation syntaxes that I could use with different datasets. To address the issues related to the new neighborhoods, which lacked data for 2015 and growth variables, a decision was made to exclude them from the analysis. This choice was based on several factors: the new neighborhoods exhibited a distinct socio-demographic structure compared to the rest of the neighborhoods, they were not undergoing gentrification but rather other urbanism processes, and the data available for these neighborhoods was incomplete. Additionally, when subjected to cluster analysis, these neighborhoods formed a separate and exclusive cluster, further justifying their removal from the analysis.

There were several neighborhoods like Atocha that presented missing values in the Real estate prices dataset from Ayto. de Madrid. A missing value proportion of 12.2% in 2015 data, and 9.9% in 2019 data. As it was a fairly important proportion, and the neighborhoods missing didn't show any pattern, I decided to estimate the missing values using "Mice" by Predictive Mean Matching imputation method with kN-Neighbors algorithm.

The last important limitation I faced when processing the data, was that there were many variables that I wanted to add, like the movements of the population in terms of emigrants and immigrants of each neighborhood, or the police presence. The problem was that the data was only available for districts, which was a big limitation that made me decide not to include these variables nor estimate them. Also, the income information was available in a map format, but the data wasn't available in "csv." or other formats that could be treated with R.

3. RESULTS

As Hierarchical and K-means methods require a predefined number of clusters, I decided to estimate it with Gap Statistics, a Machine Learning method that consists of quantifying the clustering quality based on the gap between the observed within-cluster dispersion and its expected value under null reference distributions. This approach takes into account both the compactness of the clusters and the separation between them. Gap Statistics offers a data-driven approach to cluster analysis, providing valuable insights into the underlying structure of the data and aiding in informed decision-making for clustering tasks. In this case, the optimal number of clusters computed is 3, which is a low number of clusters for describing the differences between neighborhoods properly. The next optimal value was selected, designing a cluster analysis for **6** groups.

To determine whether to use hierarchical clustering or k-means clustering, the within-cluster similarity was computed for both models. The following results indicate that the clusters generated by the hierarchical method exhibit more optimal values and demonstrate greater and more balanced homogeneity, as we can see in the following table of my own elaboration. Therefore, the results from the hierarchical clustering will be used for the analysis.

K-means method

- Cluster 1 homogeneity: 0.4259097
- Cluster 2 homogeneity: 0.7700548
- Cluster 3 homogeneity: 0.3917723
- Cluster 4 homogeneity: 0
- Cluster 5 homogeneity: 0.7222127
- Cluster 6 homogeneity: 0.1633048

Hierarchical method

- Cluster 1 homogeneity: 0.3153837
- Cluster 2 homogeneity: 0.4401724
- Cluster 3 homogeneity: 0.3808216
- Cluster 4 homogeneity: 0.4505045
- Cluster 5 homogeneity: 0.2591299

➢ Cluster 6 homogeneity: 0.265247

The results of the Hierarchical clustering (Appendix 2, Figure 1 and Figure 2) of the different neighborhoods show apparent gentrification patterns in some of the areas, but don't define strictly the areas that are or are not being gentrified, although it provides valuable insights about the relationship between each group of neighborhoods and the gentrification process or interest.

FIGURE 1.1

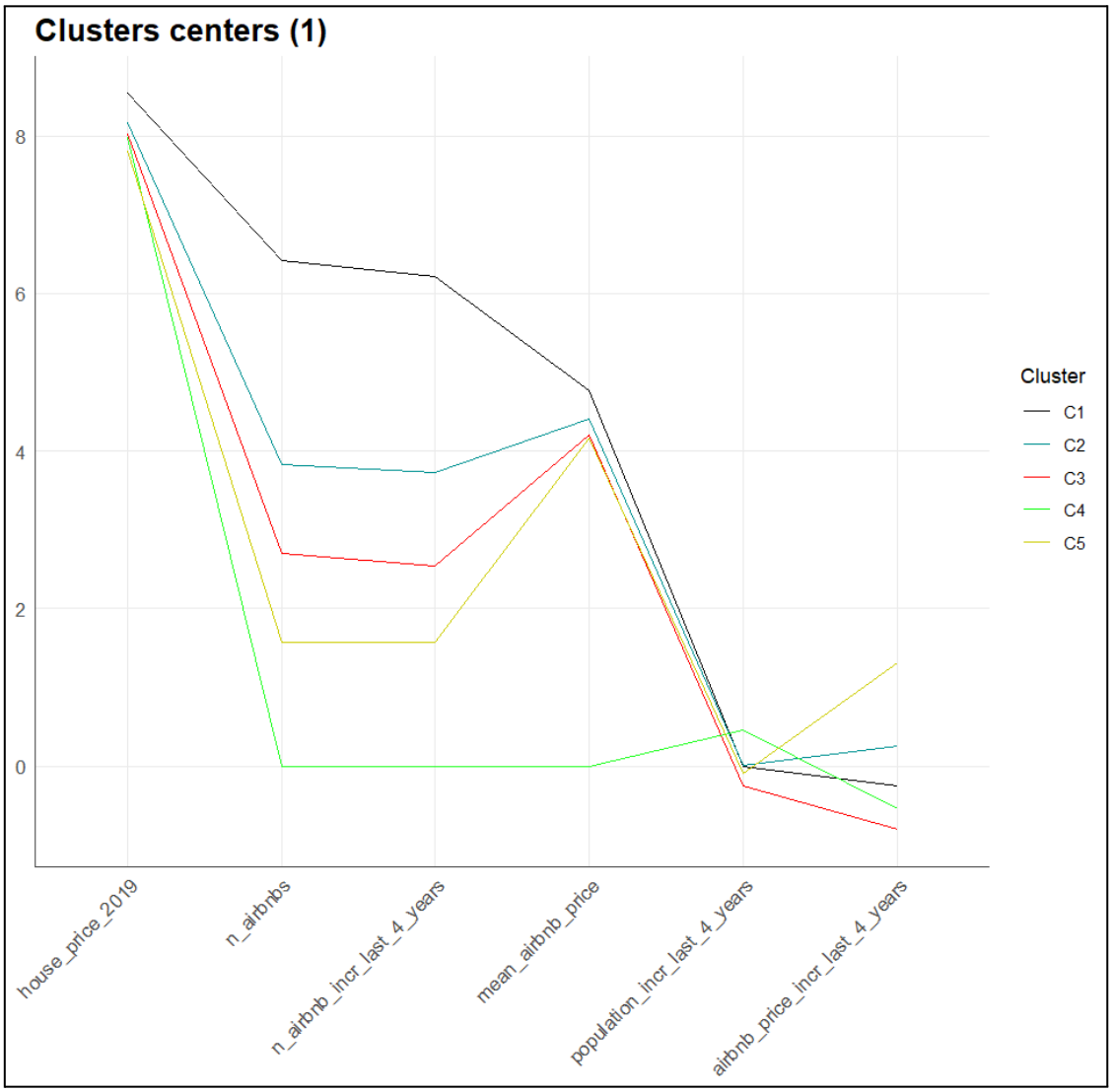
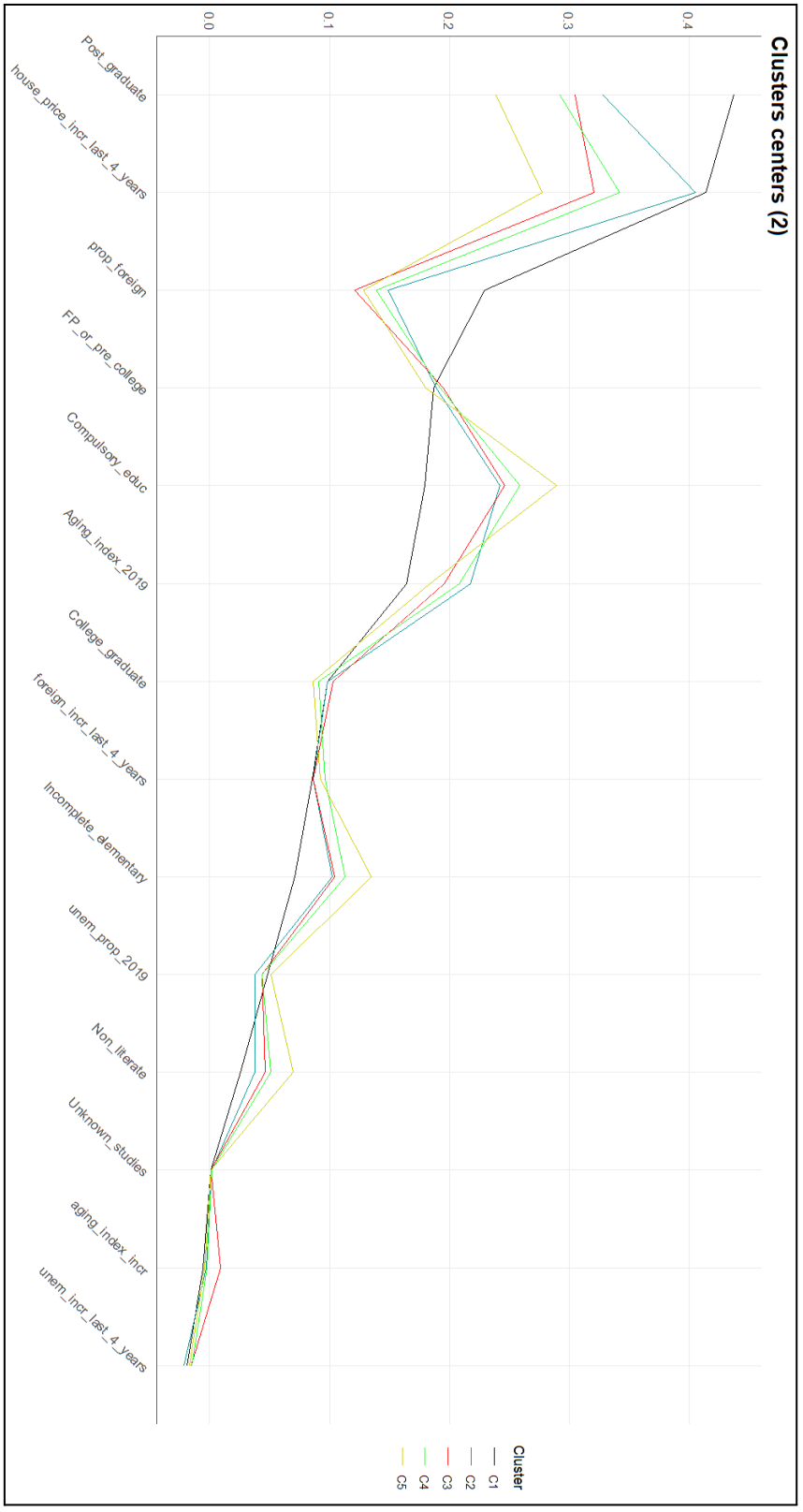


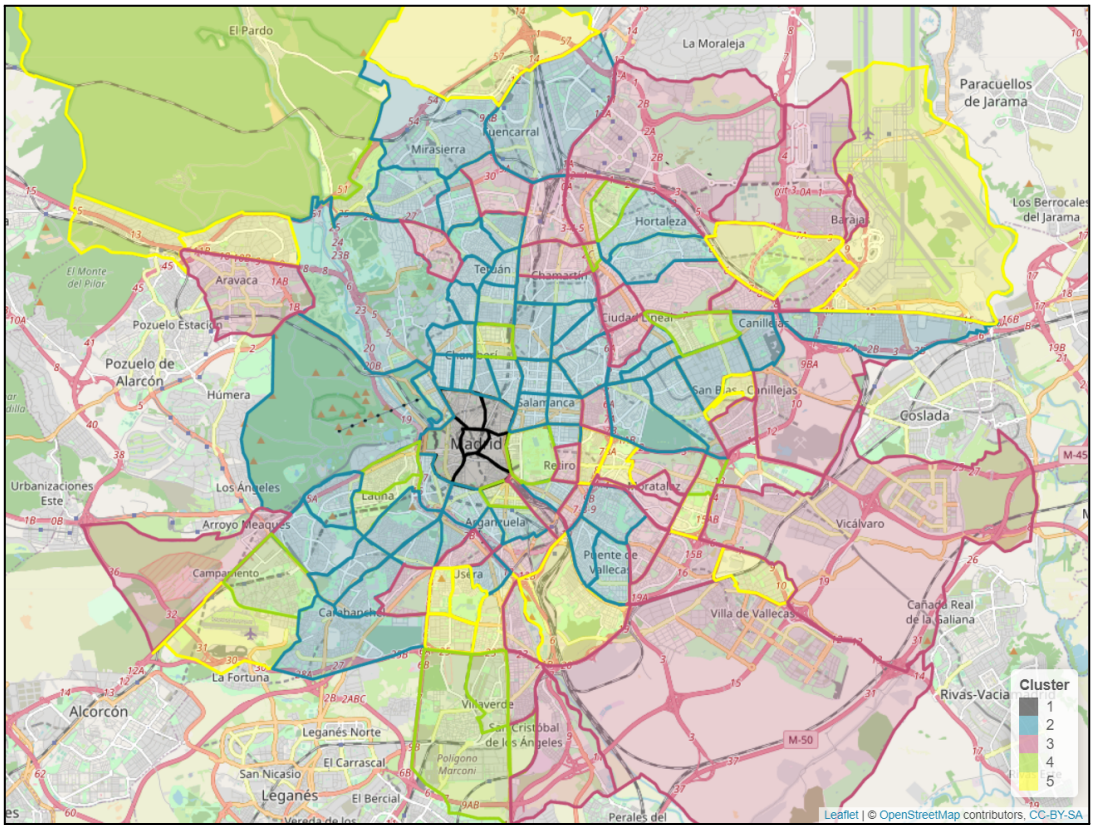
FIGURE 1.2


Clusters centers (2)
0.0
0.1
0.2
0.3
0.4
Post_graduate
house_price_incr_last_4_years
prop_foreign
FP_or_pre_college
Compulsory_educ
Aging_Index_2019
College_graduate
foreign_incr_last_4_years
Incomplete_elementary
unem_prop_2019
Non_literate
Unknown_studies
aging_Index_incr
unem_incr_last_4_years
Cluster
C1
C2
C3
C4
C5

For instance, we can find clusters defined by a working social class with relevant indicators of gentrification, and others that are not being affected because of a lack of interest in those areas due to the social class structure, the lack of amenities in the area or the distance from the city center. We can observe as well some clusters characterized by a privileged social class structure which shows signs either of previous gentrification or that are traditionally privileged class neighborhoods. The 6th cluster wasn't taken into account as it only contains one neighborhood, Valdebernardo, where the real estate prices have decreased by 59.82%.

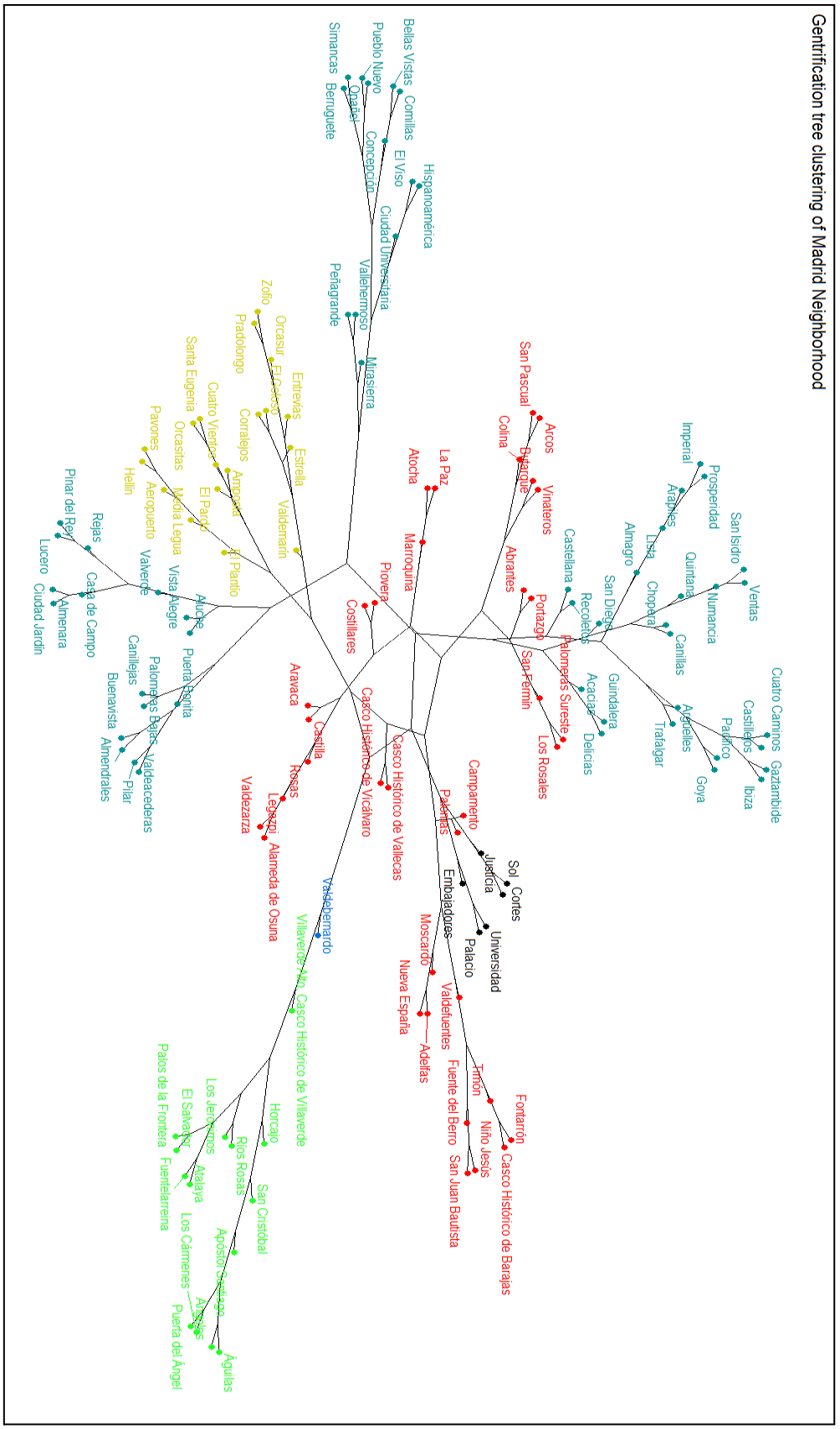
In the following dendrogram (Figure 2.2[2]), you can find the cluster to which each neighborhood belongs. Also, we can visualize the neighborhood in the following map (Figure 2.1), created with the leaflet package in R:

FIGURE 2.1



[2] You can download the interactive map from the Github repository: https://github.com/JorPS/Cluster-analysis-Gentrification-in-Madrid/blob/main/Plots/Interactive_maps/Interactive_Madrid_Gentrification_Hierarchical_Clusters_Map.html

FIGURE 2.2


Gentrification tree clustering of Madrid Neighborhood
Simancas
Berruguete
Bellas Vistas
Pueblo Nuevo
Quintana
Concepción
Comillas
El Viso
Hispanoamérica
Ciudad Universitaria
Valdemarín
Peñagrande
Valdehermoso
Mirasierra
Zofío
Pradolongo
Orcasur
Entrevías
Estrella
Corralejos
Amposta
El Pardo
El Plantío
Cuatro Vientos
Santa Eugenia
Pavones
Orcasitas
Media Legua
Aeropuerto
Helín
Pinar del Rey
Rejas
Valverde
Vista Alegre
Aluche
Lucero
Ciudad Jardín
Almenara
Casa de Campo
Canillejas
Buenavista
Almendrales
Pilar
Palomeras Bajas
Valdeacederas
Puerta Bonita
Atocha
La Paz
Mantequina
Pioveta
Costillares
Colina
San Pascual
Arcos
Valiente
Viñateros
Abrantes
Portazgo
San Fermín
Palomeras Sureste
Los Rosales
Castellana
Recoletos
San Diego
Imperial
Prosperidad
Arapiles
Almagro
Lista
Chopera
Canillas
Quintana
San Isidro
Ventas
Numancia
Trafalgar
Acacias
Guindalera
Delicias
Arguelles
Pacífico
Goya
Cuatro Caminos
Castilleros
Gaztambide
Ibiza
Ayavaca
Castilla
Rosas
Legazpi
Valdezarza
Alameda de Osuna
Casco Histórico de Vicálvaro
Casco Histórico de Vallecas
Campamento
Palomeras
Sol
Cortes
Justicia
Universidad
Palacio
Embajadores
Vicálvaro
Villaverde Alto
Casco Histórico de Villaverde
Moscardó
Nueva España
Adelfas
Valdefuentes
Fuente del Berro
San Juan Bautista
Trrión
Niño Jesús
Fontarrón
Casco Histórico de Barajas
Horcajo
Ríos Rosas
San Cristóbal
Los Jerónimos
El Salvador
Atalaya
Fuentelatreina
Patos de la Frontera
Apóstol Santiago
Los Cármenes
Aurelias
Puerta del Ángel
Águilas

Most explanatory variables of the cluster analysis

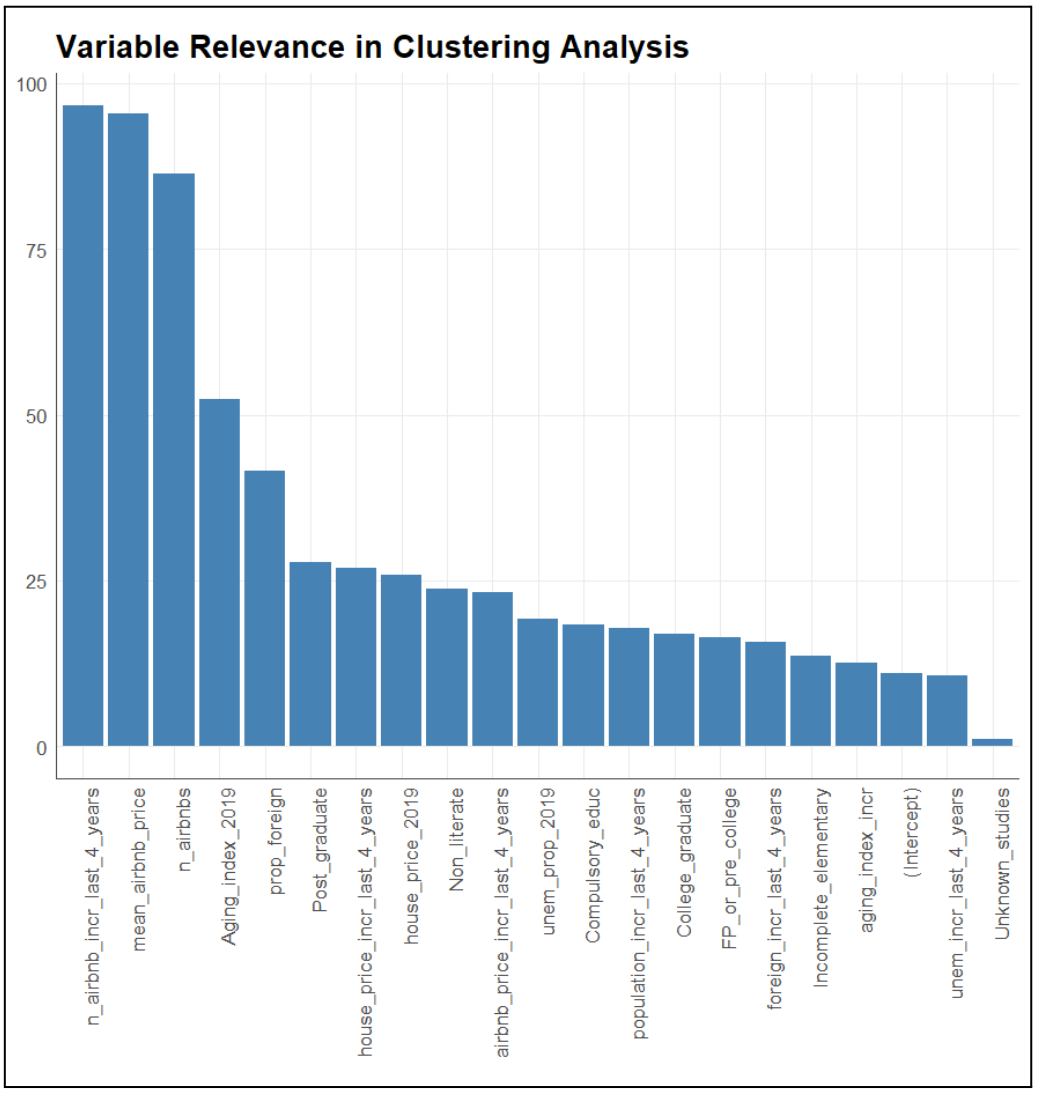
To identify the most relevant variables in understanding the clustering process, a multinomial classification model was conducted using the *nnet* package in R. The mean coefficients of the variables were extracted for each cluster, providing insights into their relative explanatory power. The results, depicted in Figure 2.3 (Appendix 5), highlight the differences in explanatory power among the variables.

Among the variables, the increment of Airbnb listings during the previous 4 years (96.71), the Mean Airbnb listings price (95.40), the number of Airbnb listings (86.40), and the Aging index (52.29) emerged as the most influential factors. These findings suggest that changes in Airbnb activity, pricing, and the presence of listings have a significant impact on the clustering patterns. Additionally, the aging index, which reflects the demographic composition of the population, plays a prominent role in shaping the clustering outcomes.

Conversely, the proportion of the population with unknown education levels and the unemployment evolution over the past 4 years were identified as the variables with the least explanatory power for the dataset's variance. These variables may have minimal impact on the clustering patterns observed.

Notably, the results align with expectations, as the influence of Airbnb activity and population aging emerges as strong contributors to the data variance in 2019. Furthermore, the variable related to the educational level that exhibits the highest explanatory power is the proportion of the population with a postgraduate degree, indicating its significance in understanding the clustering process.

FIGURE 2.3



Cluster 1: The city center

These neighborhoods in the center are characterized by an estimated high social class, as it shows a relatively low unemployment rate(4.84%), which has decreased 1.89% during the last 4 years. In education terms, it shows the biggest proportion of post-graduated people (43.74%) and high real estate prices (€5.15 M), while maintaining the lowests proportions of

people with studies below Spanish FP or precollege.It is the most affected cluster by airbnb number of housings (659, increasing by a mean of 531 per neighborhood) and it has the highest mean price of $116.49, which has slightly increased by $10.63 during the last 4 years, if we compare it with the next clusters. This would represent already gentrified neighborhoods in the center, or neighborhoods that traditionally had these characteristics. In this cluster we find Malasaña, which is one of the canonical examples of gentrified neighborhoods in Madrid (Ruiz et. al., 2021).

In the first cluster, the neighborhood of *Embajadores* stands out as the most prominent. It can be inferred that the residents in this neighborhood belong to a less privileged social class compared to the other neighborhoods in the cluster. Embajadores exhibits several distinct characteristics: a notably higher unemployment rate of 5.65%, the lowest real estate prices within the cluster at €4.48 M, and a higher proportion of residents who have completed their education before compulsory education and elementary levels. Conversely, Embajadores has the lowest proportion of post-graduates and college graduates in the cluster.

Furthermore, it is worth noting that Embajadores has experienced a significant increase in the number of Airbnb amenities, with 860 listings compared to the cluster average of 531.67. Based on this information, it can be concluded that the gentrification process in Embajadores is less consolidated compared to other neighborhoods in the city center. Additionally, Embajadores appears to be the most vulnerable neighborhood within the cluster, indicating a higher likelihood of experiencing the gentrification process.

Cluster 2: High relationship with gentrification area

The second cluster comprises a diverse range of neighborhoods, spanning from the periphery of the city such as Fuencarral, Mirasierra, Buenavista, and Rejas, to the surrounding areas of the city center including Argüelles, Gaztambide, Acacias, and Recoletos. As a result, there is a wide variation in the level of gentrification influence within this cluster. The defining characteristics of the second cluster include a higher proportion of high social class residents and a significant presence of middle-class individuals, as evidenced by the data. It is noteworthy that this cluster exhibits the second-highest house prices among the different clusters (€3.76 M), the second-highest proportion of post-graduates (32.85%), and the lowest unemployment rate (3.78%).

However, it is important to highlight that this cluster also displays patterns that do not exhibit a consolidated presence of high social class inhabitants, unlike the first cluster. The values in this cluster are closer to those observed in the rest of the clusters, particularly in terms of the proportion of individuals who have completed compulsory education or have not completed elementary education. This observation, considering the hierarchical clustering method employed, suggests that there is a greater diversity among neighborhoods in this cluster. While this cluster can be defined as privileged in terms of social class, its proximity to other clusters and the variation within the neighborhoods indicate that there is considerable diversity within the group.

Within this cluster, the influence of Airbnb is the second highest, characterized by 52 listings and an average price of $84.86, which has witnessed an increase of $29 over the past four years. The real estate prices in this cluster have risen by 40.58%, placing them closer to the neighborhoods in the city center cluster (41.45%) rather than the other clusters, where the price growth is less than 35%.

In conclusion, neighborhoods within this cluster that have lower education levels, lower real estate prices, higher unemployment rates, and a stronger presence of Airbnb listings are more susceptible to experiencing the effects of gentrification. It is also important to note that this cluster has the highest average Population Aging Index, further increasing its vulnerability to the impacts of gentrification.

Among the neighborhoods in this cluster, San Diego, Almendrales, Numancia, San Isidro, Vista Alegre, Aluche, and Opañel are particularly vulnerable (Appendix 6, Figures 3.1, 3.2, 3.3). These neighborhoods exhibit a high concentration of Airbnb listings and indicators that suggest a predominance of working-class citizens. Despite this, the real estate prices in these neighborhoods are clearly increasing, as we can see in the following table. It’s important to notice that these neighborhoods show significant statistical similarities with wealthier neighborhoods like Castellana and Recoletos. For instance, Castellana has an average house price of €7.12 million, Recoletos has an average price of €8.44 million, while the cluster's overall average is €3.76 million. Additionally, Castellana and Recoletos have a high proportion of post-graduates (58.67% and 55.94%, respectively), compared to the cluster average of 32.85%.

FIGURE 3.1

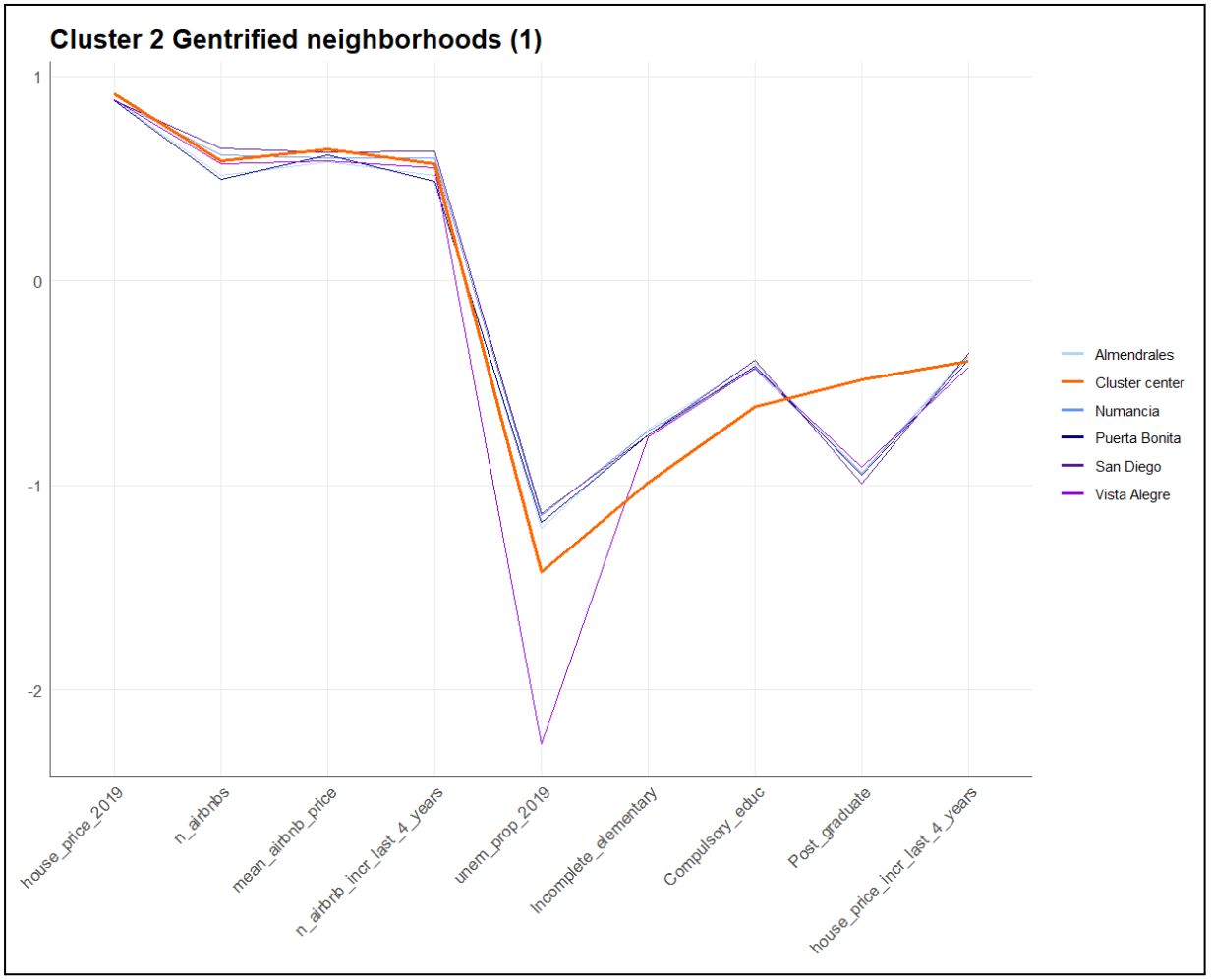

Cluster 2 Gentrified neighborhoods (1)
1
0
-1
-2
house_price_2019
n_airbnbs
mean_airbnb_price
n_airbnb_incr_last_4_years
unem_prop_2019
Incomplete_elementary
Compulsory_educ
Post_graduate
house_price_incr_last_4_years
Almendrales
Cluster center
Numancia
Puerta Bonita
San Diego
Vista Alegre

FIGURE 3.2

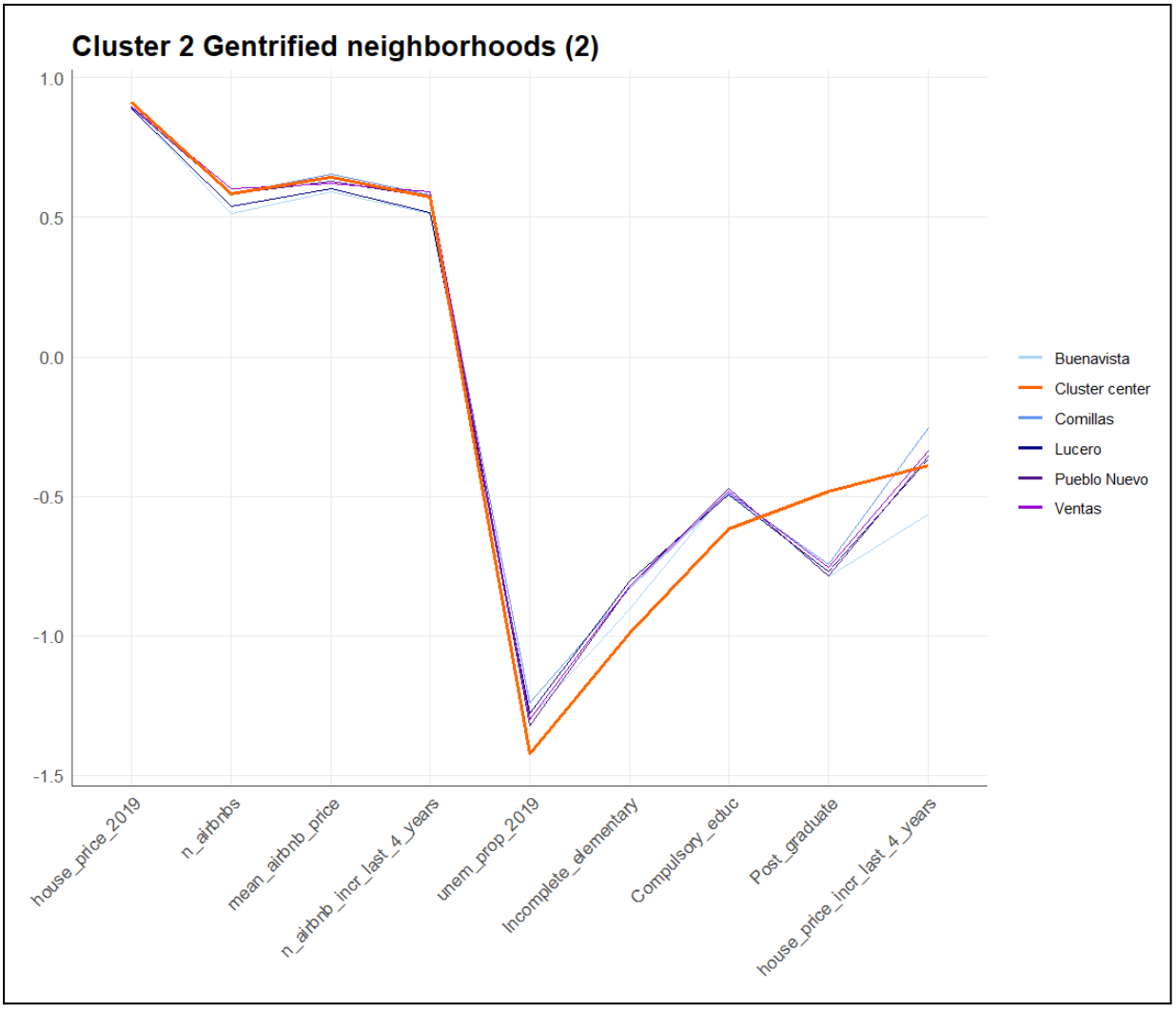
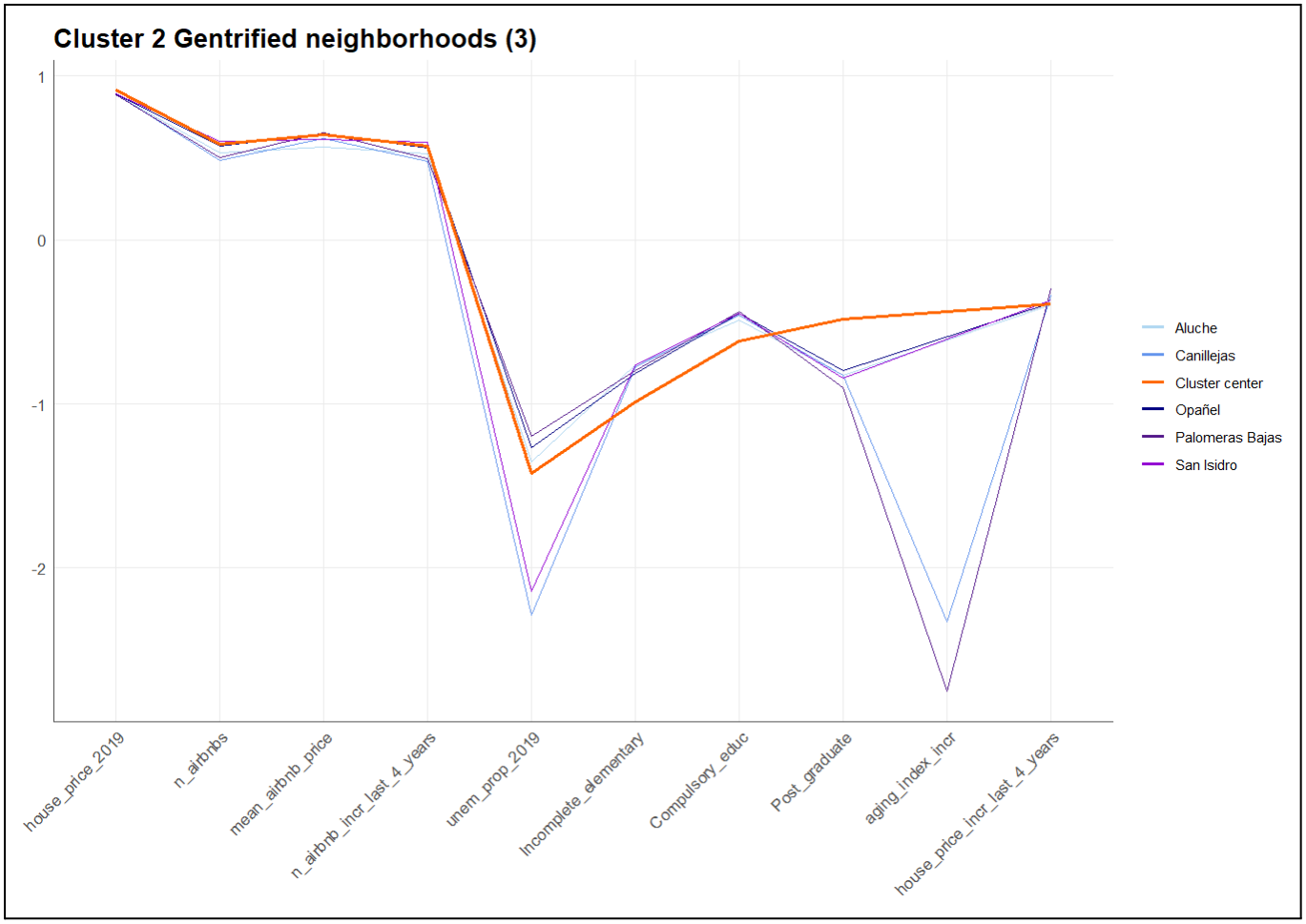

Cluster 2 Gentrified neighborhoods (2)
1.0
0.5
0.0
-0.5
-1.0
-1.5
house_price_2019
n_airbnbs
mean_airbnb_price
n_airbnb_incr_last_4_years
unem_prop_2019
Incomplete_elementary
Compulsory_educ
Post_graduate
house_price_incr_last_4_years
Buenavista
Cluster center
Comillas
Lucero
Pueblo Nuevo
Ventas

FIGURE 3.3



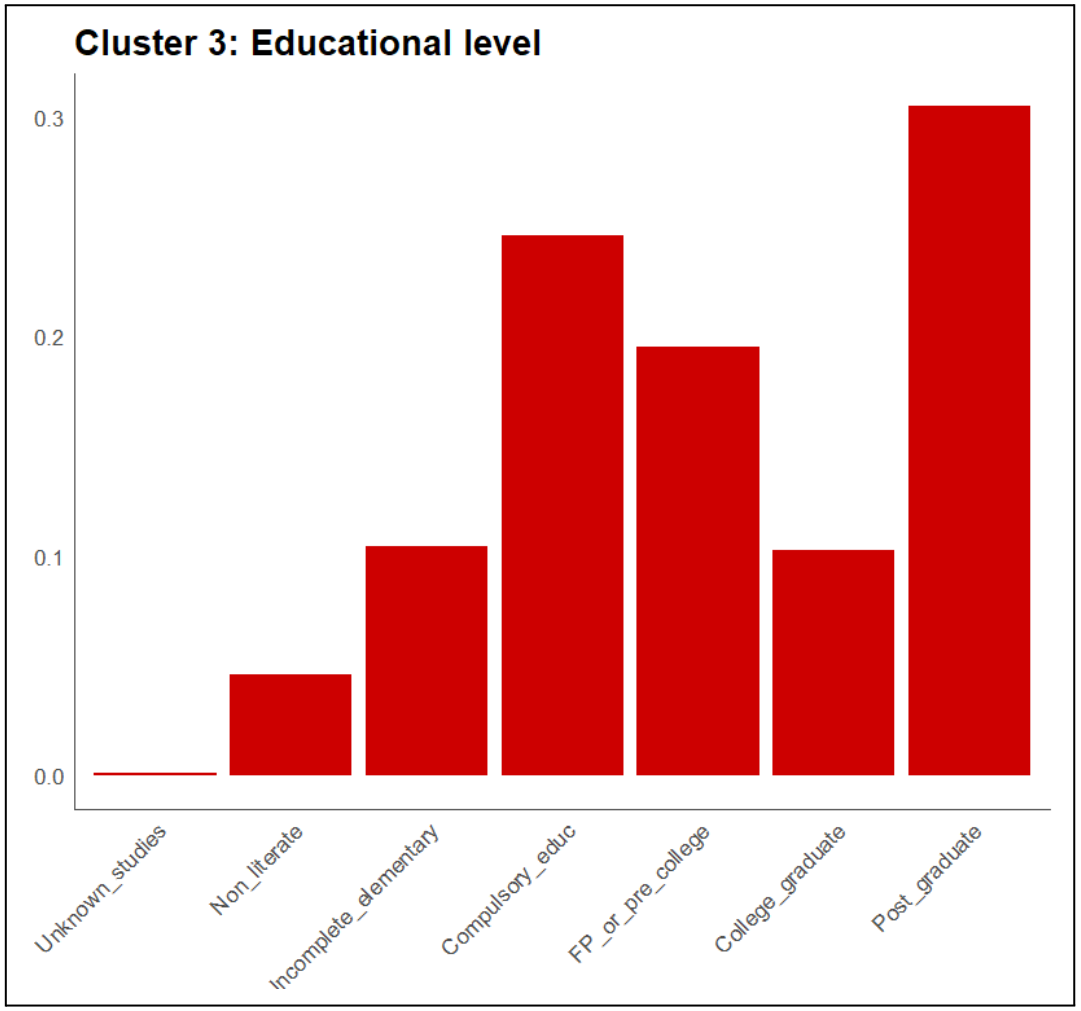
Cluster 3: Polarized area

Observing the location of the neighborhoods in this cluster we find that they are located mostly in the periphery of the city and are surrounded mainly by neighborhoods belonging to the second cluster. The neighborhoods of Retiro and Atocha are the nearest to the city center. One of the main characteristics of this cluster is that the population is decreasing dislike the rest of the clusters

If we take a glance at the centers of the cluster (Figure 4), we find that there is a lower gap between people who finished post-graduate studies (30.47%) and those with lower education level (10% of people with uncompleted elementary studies). In fact, there are more people who just finished compulsory education, 24.61% or pre-college studies, 19.51%, than college

graduates, which represents a mean of 10.28% of the neighborhood's population. On the other hand, the unemployment rate (4.35%) is similar to the unemployment mean between clusters, 4.49%.

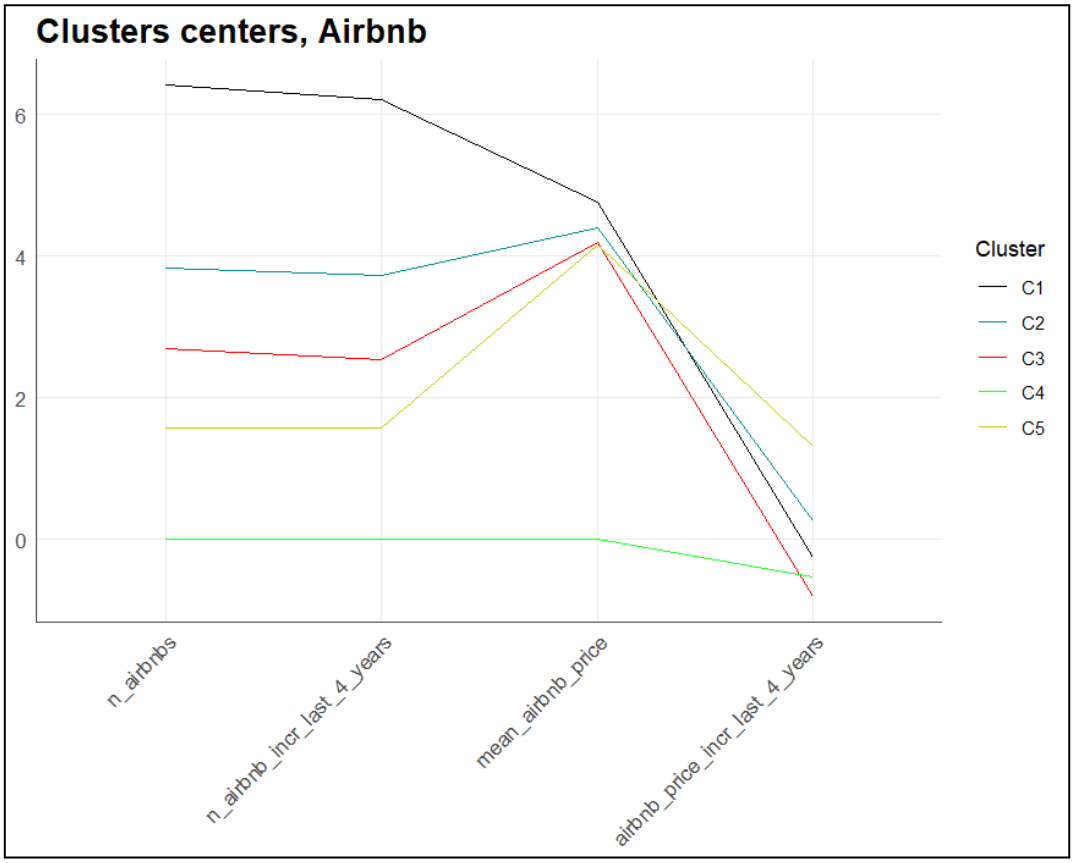
FIGURE 4



From this information, we can confirm that the social structure of these neighborhoods are mainly based on middle classes, but also by a diversity of working class and privileged class patterns. Therefore, the question here is which is the relationship that these socially diverse neighborhoods keep with gentrification? If we observe the airbnb influence, it's significantly lower than the first two clusters, but the evolution of the Airbnb influence is not clear as, for example, the prices are decreasing while the supply is increasing, suggesting that this issue is also polarized between different neighborhoods (Figure 5). This can mean that the gentrification process is less consolidated than in the first clusters, but it's also present in

some areas as we can observe that in these clusters there are working class neighborhoods with significant statistical similarities to middle classes and privileged classes, according to the considered variables.

FIGURE 5



In the following plots (Figures 6.1 and 6.2), you can find the various neighborhoods that exhibit indications of being linked to the gentrification process in this cluster for different reasons (Appendix 7). These neighborhoods display notable characteristics, such as escalating property prices, number of Airbnb listings and characteristics indicating a middle or working class social structure in the neighborhood, such as higher unemployment rates. These factors collectively contribute to the transformation of these areas, attracting investment and potentially resulting in harmful consequences for the most modest neighbors.

FIGURE 6.1

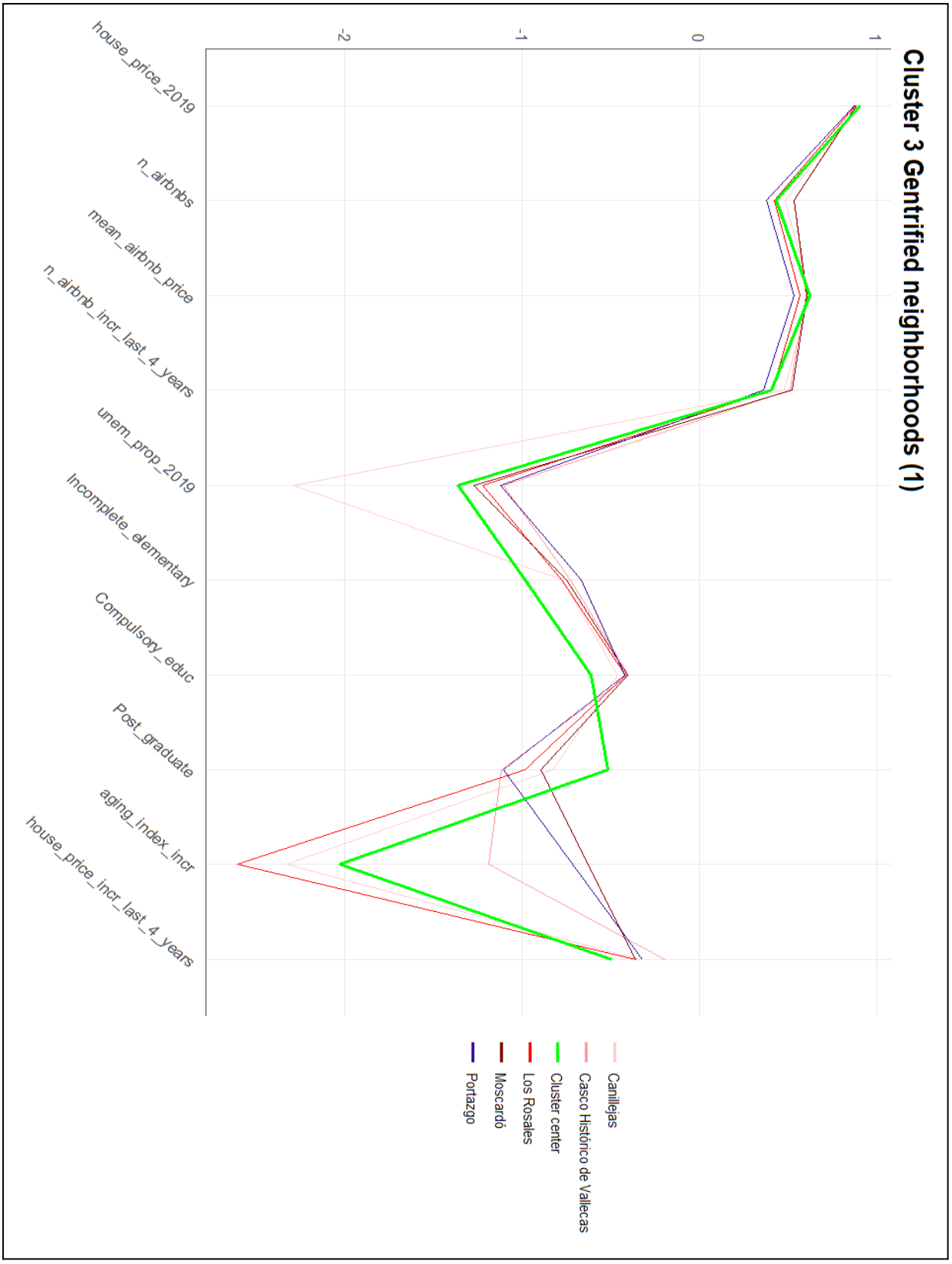
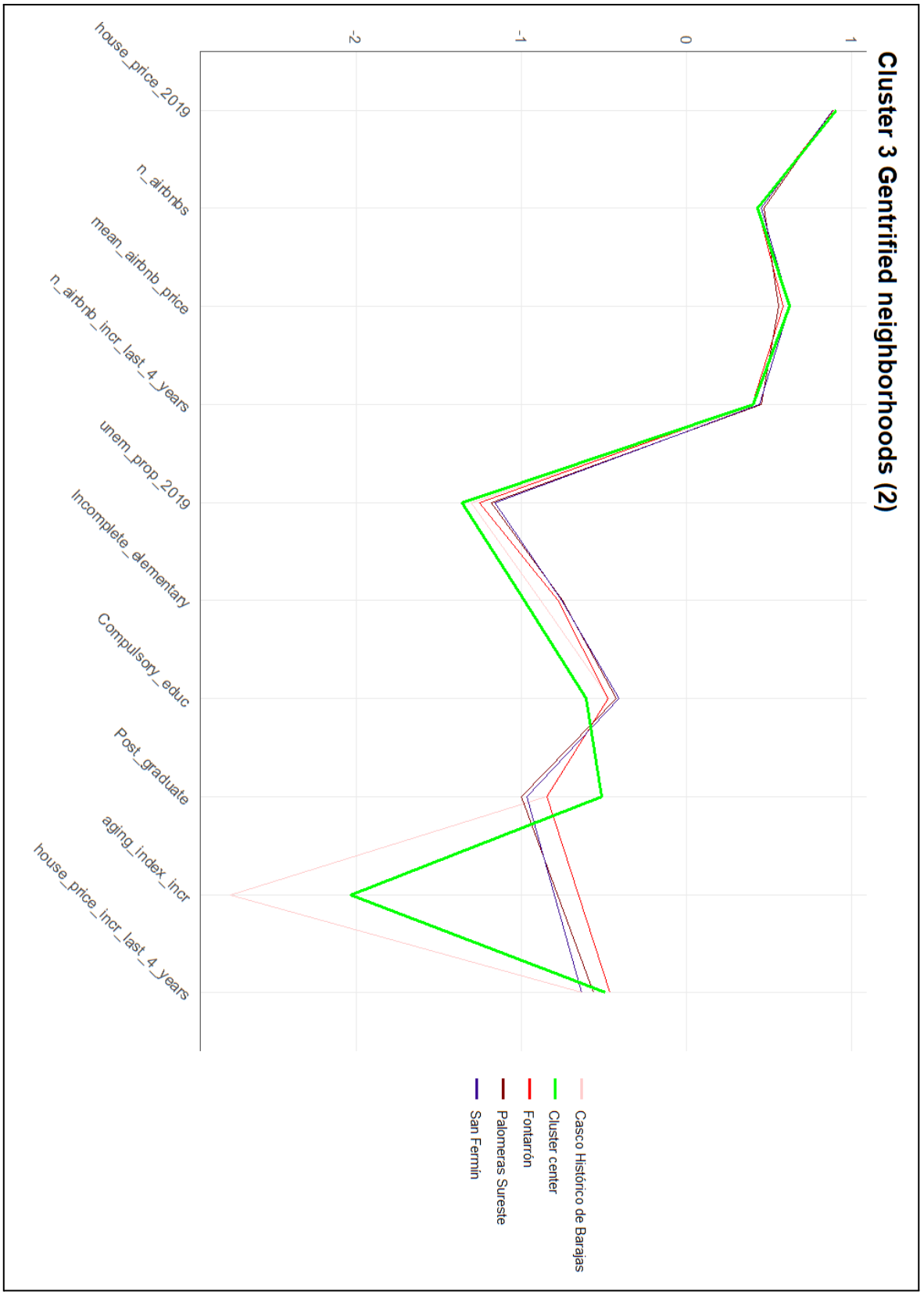
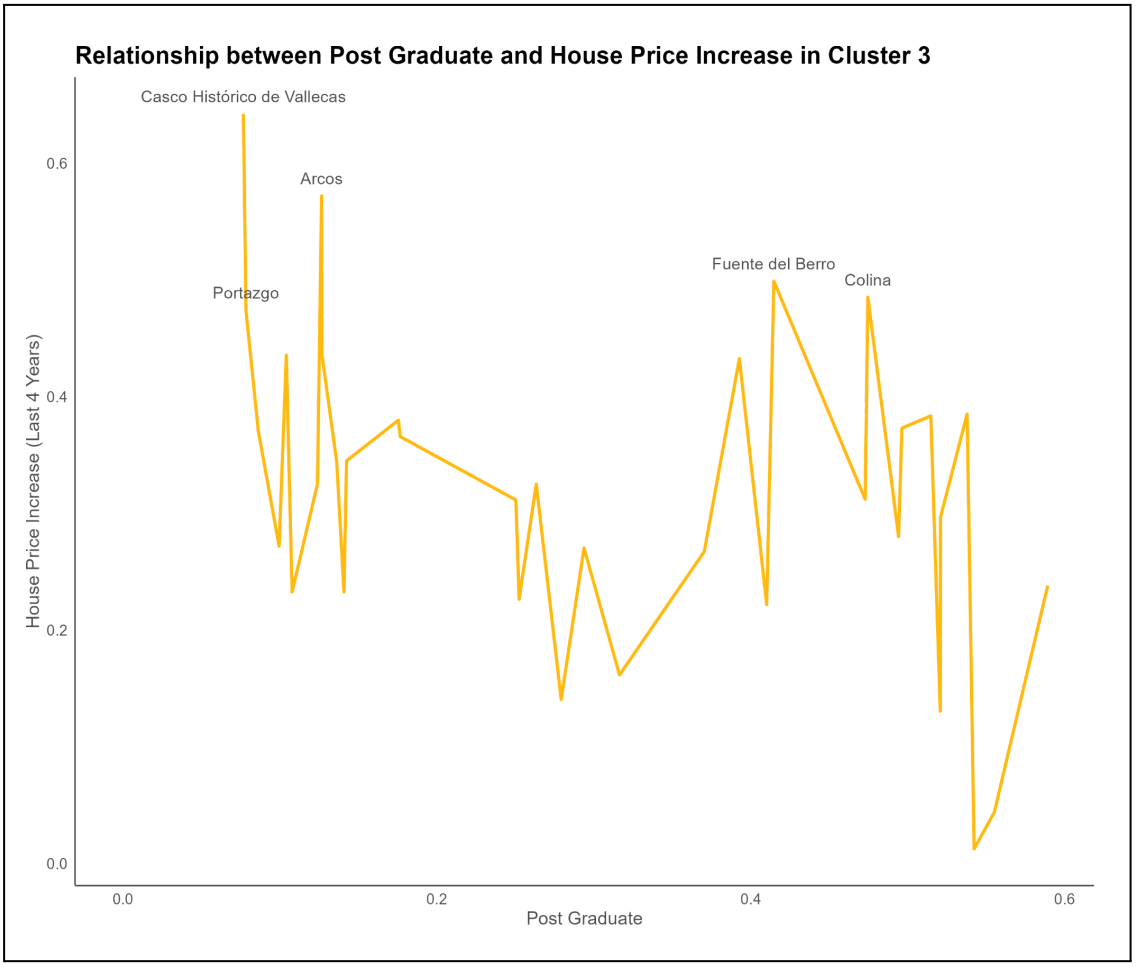

Cluster 3 Gentrified neighborhoods (1)
house_price_2019
n_airbnbs
mean_airbnb_price
n_airbnb_incr_last_4_years
unem_prop_2019
Incomplete_elementary
Compulsory_educ
Post_graduate
aging_index_incr
house_price_incr_last_4_years
1
0
-1
-2
Canillejas
Casco Histórico de Vallecas
Cluster center
Los Rosales
Moscardó
Portazgo

FIGURE 6.2


Cluster 3 Gentrified neighborhoods (2)
house_price_2019
n_airbnbs
mean_airbnb_price
n_airbnb_incr_last_4_years
unem_prop_2019
Incomplete_elementary
Compulsory_educ
Post_graduate
aging_index_incr
house_price_incr_last_4_years
1
0
-1
-2
Casco Histórico de Barajas
Cluster center
Fontarrón
Palomeras Sureste
San Fermín

When taking into account the data from the different neighborhoods in this cluster, we find interesting patterns that clearly indicate a considerable risk of gentrification for specific neighborhoods in this cluster. A very interesting pattern is that there are some neighborhoods in which the proportion of post-graduated inhabitants is very low and the house price is increasing more than the neighborhoods with the higher post-graduates proportion, as we can observe in the following graph (Figure 7).

FIGURE 7



The neighborhoods exhibiting more severe signs of gentrification within this cluster include Casco Histórico de Vallecas, Portazgo, Arcos, Los Rosales, and Moscardó. In particular, Casco Histórico de Vallecas and Portazgo have the lowest proportions of individuals with post-graduate education (7.68% and 7.84%, respectively) and a significant proportion of residents who have only completed compulsory education (40.04% and 37.85%, respectively)

or have not finished elementary school (18.59% and 21.6%, respectively). Additionally, these neighborhoods exhibit elevated unemployment rates (7.82% and 7.5%, respectively), further indicating a predominantly working-class population.

Furthermore, Casco Histórico de Vallecas and Moscardó have a higher presence of Airbnb listings, with 29 listings compared to the cluster's average of 15.63. This suggests an increased influence of short-term rentals and potentially higher demand from tourists or visitors in these neighborhoods. The remaining neighborhoods in the past table exhibit similar patterns associated with gentrification, although slightly less pronounced.

The combination of these factors points to a more advanced stage of gentrification in Casco Histórico de Vallecas, Portazgo, Arcos, Los Rosales, and Moscardó. These findings highlight the socio-economic shifts occurring in these areas and the potential challenges faced by the local communities in the face of gentrification.

Cluster 4: No Airbnb supply area.

The most notable feature of the neighborhoods in this cluster is the absence of Airbnb listings, which can be attributed to various factors. Furthermore, this cluster comprises a relatively small number of neighborhoods, specifically 14. Analyzing the central tendencies within this cluster, it becomes apparent that the social class composition closely resembles that of the third cluster, as demonstrated in the table below (FIGURE 8) and in the the plot of cluster centers showed before:

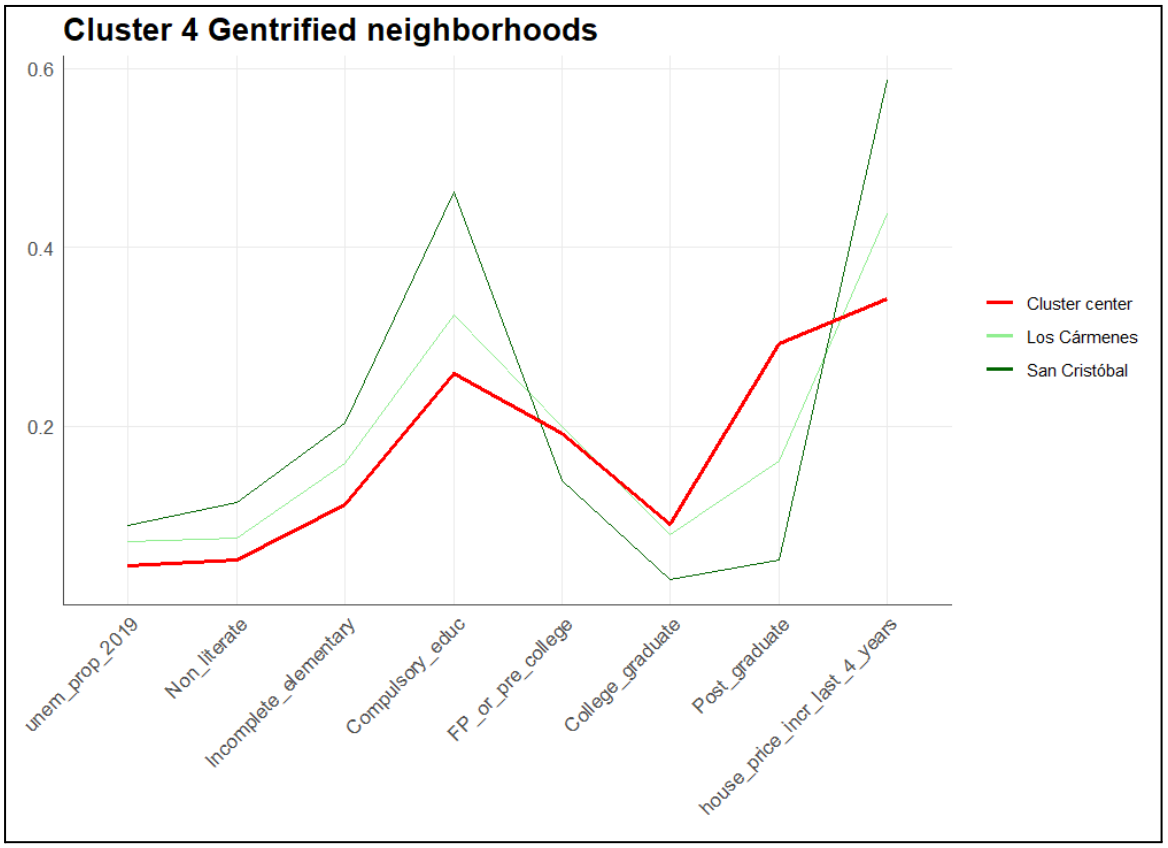
FIGURE 8

	C3	C4
house_price_2019	3209.6524	**3226.5714**
house_price_incr_last_4_years	0.3206	**0.3424**
unem_prop_2019	0.0435	**0.0440**
unem_incr_last_4_years	-0.0158	**-0.0150**
Unknown_studies	0.0012	**0.0015**
Non_literate	0.0461	**0.0507**
Incomplete_elementary	0.1041	**0.1128**

Compulsory_educ	0.2461	**0.2592**
FP_or_pre_college	0.1951	**0.1925**
College_graduate	0.1028	**0.0908**
Post_graduate	0.3047	**0.2925**
Aging_index_2019	0.1955	**0.2084**

The neighborhoods within this cluster that are at high risk of experiencing gentrification are primarily San Cristobal and Los Cármenes. These neighborhoods exhibit significantly higher housing price growth, reaching 34.24%, compared to the cluster mean. Additionally, they demonstrate distinct patterns of a working-class social structure, evident in the unemployment rates of 7.14% and 8.96%, respectively. The distribution of educational levels also aligns with a working-class profile, as depicted in the following plot (Figure 9. Table in Appendix 8):

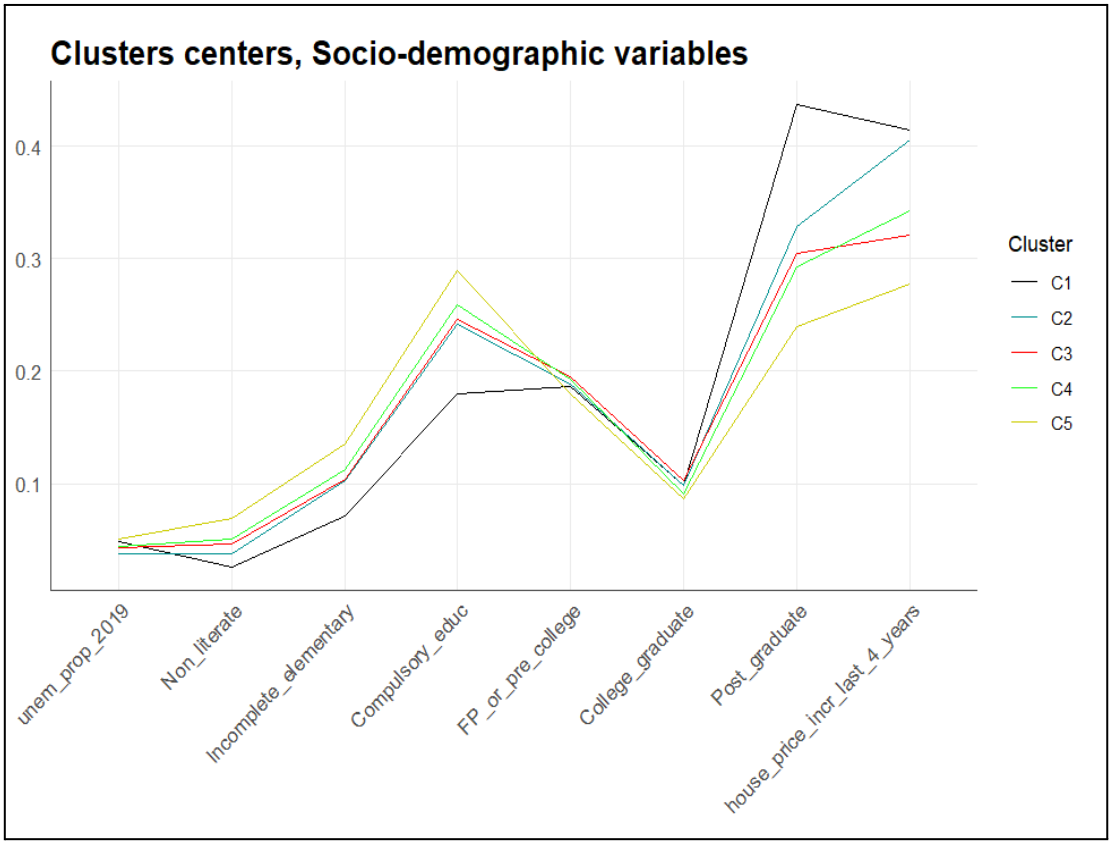
FIGURE 9



Cluster 5: Recent Airbnb presence area

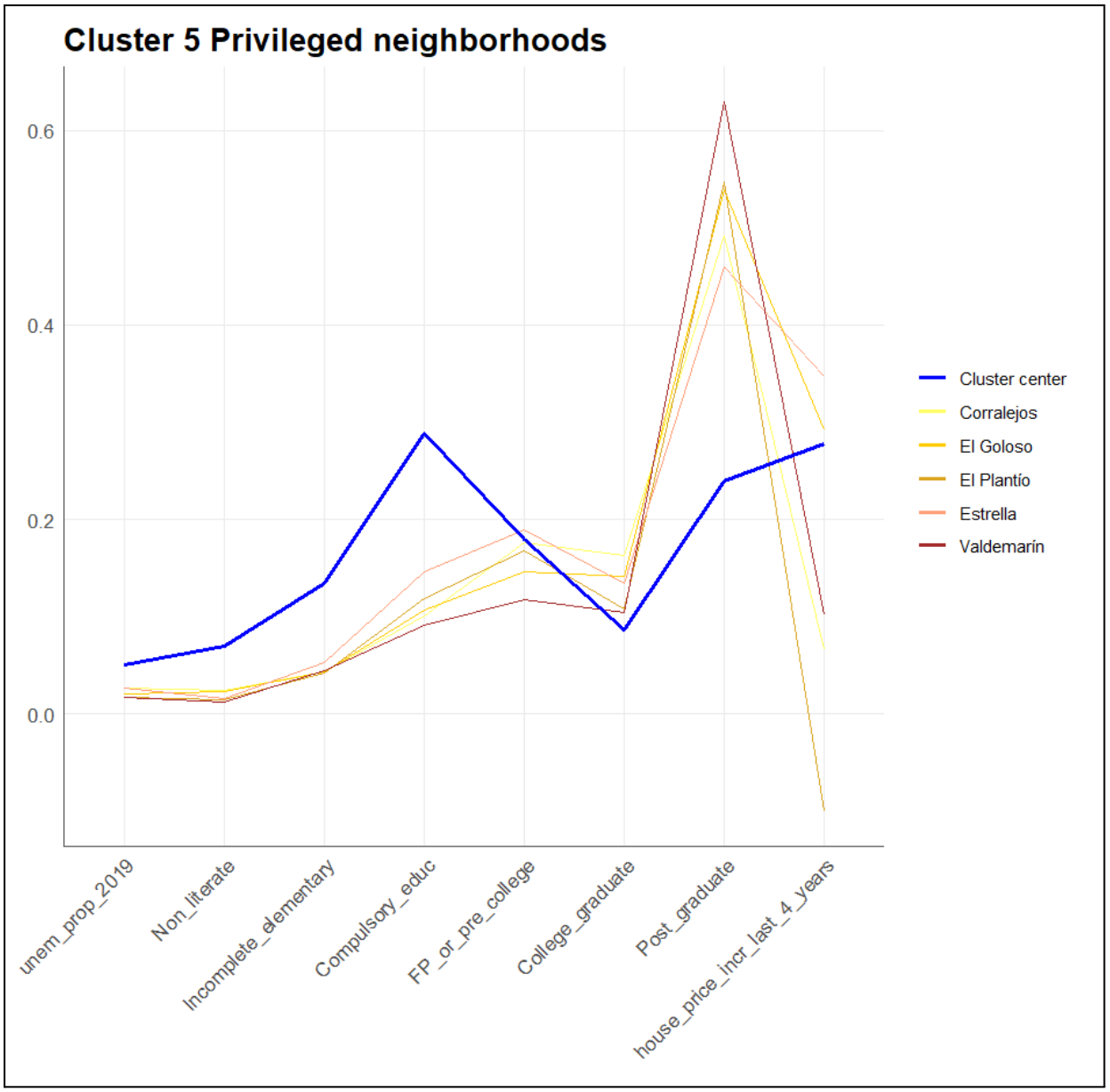
In this cluster, we specifically observe neighborhoods where the presence of Airbnb is relatively new, with no listings in 2015. The number of Airbnb listings and its increment during the previous 4 years are equal, at 4.56. These neighborhoods typically exhibit a more modest social structure, likely influenced by the relatively low Airbnb prices. Additionally, this cluster has the highest mean unemployment rate (5.08%) and the lowest center for real estate prices (€2.59 M) and its growth over the previous 4 years (27.74%). Furthermore, the proportion of people who have completed Post-graduate, College-graduate, and Pre-college graduate studies is notably lower in this cluster, as shown in the following plot (Figure 10).

FIGURE 10



Despite the observed patterns, we can still identify neighborhoods in this cluster that exhibit privileged or middle-class structures, if we observe at the distance from the cluster center. These neighborhoods coexist in the cluster with predominantly working-class areas within the same cluster, primarily because of the recent and relatively low influence of Airbnb. In the following graph, we can find these neighborhoods (Figure 11. Table in Appendix 9).

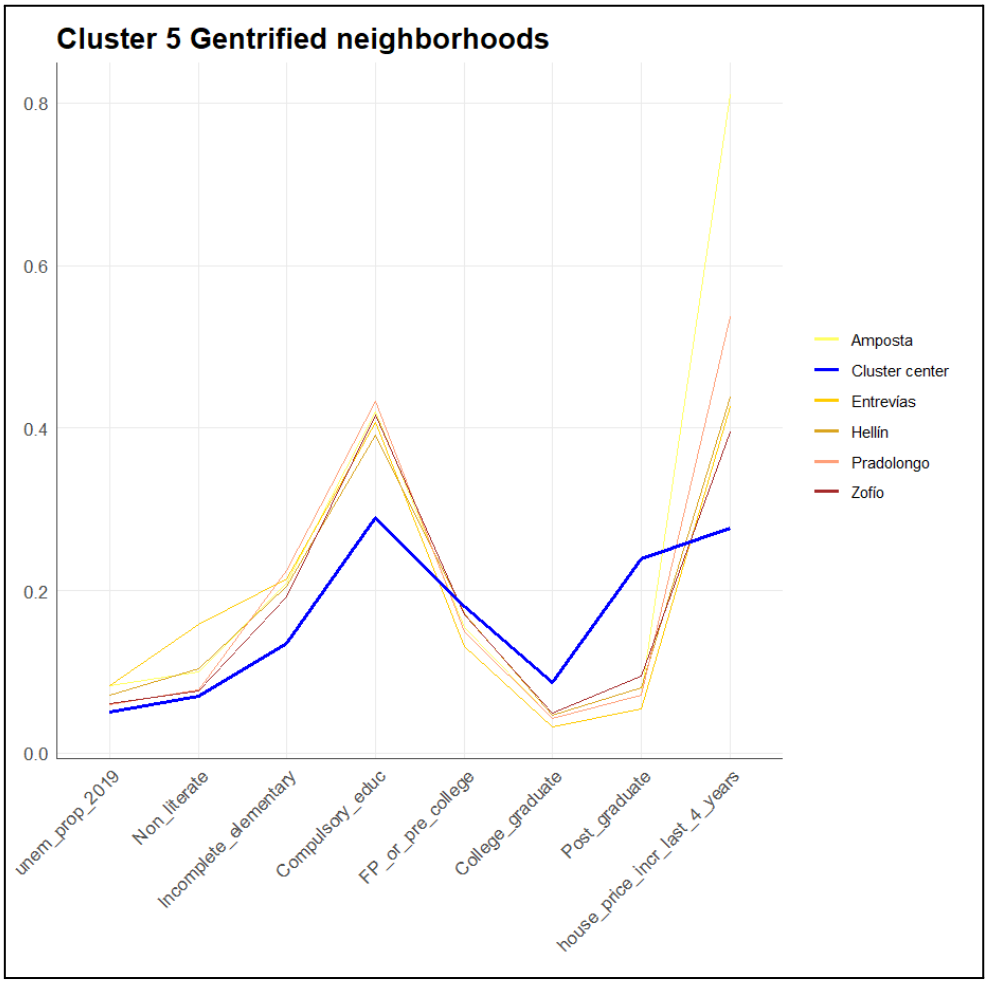
FIGURE 11



On the other hand, we find neighborhoods in which the price is increasing considerably higher than the cluster's mean and, besides, the social structure indicates a predominance of working class because their variables related to the social class structure have a significant distance with the centers, showing a tendency of high unemployment and lower education level: Entrevias is the neighborhood with the highest rate of college-graduated (4.6%) and post-graduated neighbors (8.08%), which is already significantly lower than the center (23.92%).

Upon examining the following plot of vulnerable neighborhoods at risk of gentrification (Figure 12.Table in Appendix 10), it becomes evident that the proportion of individuals who completed their education below the pre-college level is significantly higher compared to the cluster's centers. The two neighborhoods in which the gentrification risk is more clear are Amposta and Pradolongo, which are experiencing an important growth of the real estate market prices (81.05% and 53.74%, respectively), while Amposta shows a 8.23% of unemployment rate, which is considerably high in comparison with the rest of the neighborhoods. Although, the specific relationship between this cluster and the process of gentrification is that it shows early gentrification patterns, as I mentioned, such as the recent Airbnb influence on the neighborhood and a predominance of working class structure.

FIGURE 12



4. CONCLUSIONS

Summary

At the beginning of this thesis the process of gentrification was explained as a socio-cultural historical change of urban cities in the context of the globalized economy that appeared in the middle of the second half of the twentieth century in the Western countries mostly. Also, we defined the process of gentrification step by step and defining the consequent problem produced by this process that affects negatively the working classes in urban areas, as they are excluded from the benefits of the changes in the neighborhoods where they live and being forced to move out of their home and generating structural issues with negative consequences on the society and vulnerating the equal right to the city of the citizenship. (Sorando & Ardura, 2018) (Sequera, 2014) (Walliser & Sorando, 2019).

Gentrification can contribute to a decrease in affordable housing options. As property values increase, affordable housing units may be demolished or converted into higher-priced housing, reducing the availability of affordable homes in the area. This process also disrupts long-standing communities and social networks. As neighborhoods change and longtime residents are displaced, the sense of community and social cohesion can be eroded, leading to a loss of social support systems that are specially necessary for working classes and a weakened sense of belonging. As wealthier individuals and families concentrate in gentrified neighborhoods, the divide between affluent and low-income areas can widen, exacerbating socioeconomic disparities and causing segregation (Sorando & Ardura, 2016 & 2018) (Ruiz et. al., 2021) (Walliser & Sorando, 2019).

This thesis aims to tackle the issue of gentrification in Madrid, Spain, through a comprehensive analysis. The primary objective is to conduct a hierarchical cluster analysis, which groups the various neighborhoods based on their similarity in relation to several key variables. These variables have been identified as significant contributors to the gentrification process, based on extensive literature review. To gather the necessary data, the official website and API of the Ayuntamiento de Madrid, the governing body of the municipality of Madrid, have been utilized. Data from July 2015 and July 2019 have been collected, providing a comprehensive timeframe for analysis.

Limitations

As I mentioned in the beginning of this document, there are several limitations present when studying the gentrification in Madrid city, due to a lack of periodical and complete data about the different neighborhoods, specially with socio-economic indicators such as the mean income of a Neighborhood (Rubiales, 2014). It was considered to add to this study the data of the residence moves of the different districts, as it's a key variable in order to confirm if there has been a gentrification process in a Neighborhood, even though there are several neighborhoods mentioned such as Embajadores or Casco Histórico de Vallecas that show, as we've discussed, a significantly high increase of the real estate market prices, a clear working class social structure and an elevated number of Airbnb listings. It was not included because, as well as many other variables, the data is only available by districts and it's not provided at the neighborhood level. Estimating or collecting this data could definitely help to reinforce the findings of the present thesis and observe clearer patterns.

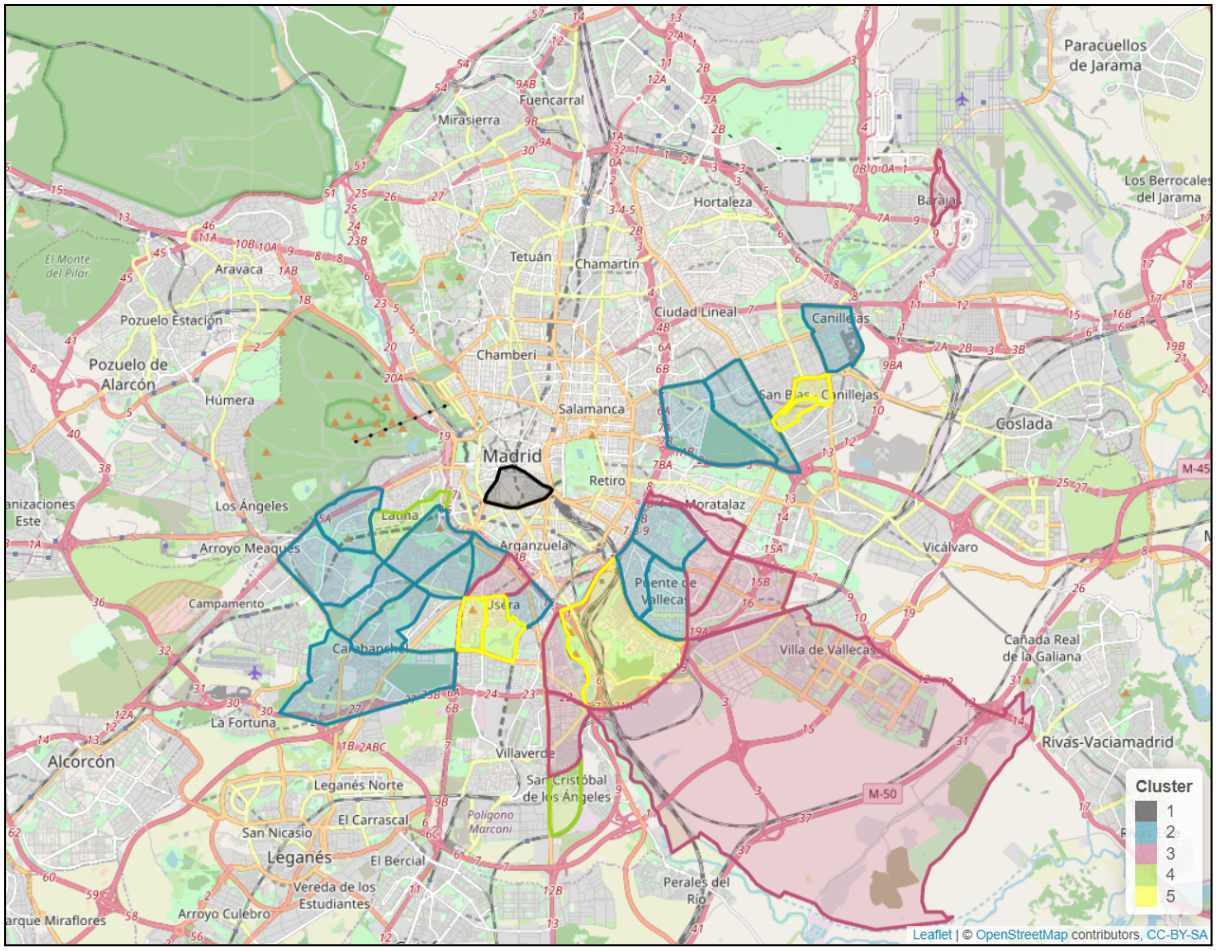
Another limitation of the used methodology is that it relies almost completely on quantitative methods and information, while studying gentrification requires getting involved in the neighborhood's history, changes, customs and actual issues of the neighbors (Rubiales, 2014). Thus, a mixed methodology with wider resources and format limitations could result in important improvements to this cluster analysis.

The unsupervised nature of this cluster analysis limits its ability to predict and prevent gentrification in specific neighborhoods using machine learning or other supervised predictive methods. However, the model can still provide valuable insights into the characteristics associated with gentrification. For instance, within the 5th cluster, neighborhoods like Amposta and Pradolongo exhibit clear signs of early gentrification. These neighborhoods demonstrate a well-established working-class social structure, significant price increases, and a relatively recent emergence of Airbnb supply. While the model itself is not predictive, it highlights important indicators and patterns related to gentrification. This prediction limitation is also caused by using 2019 as the most recent data, taking into account that the COVID 19 pandemic was suffered mainly in the years 2020 and 2021, producing unknown and not studied possible effects on the considered variables.

Findings

By doing this cluster analysis, there were identified different neighborhoods which are at greater risk of gentrification due to its relation with the gentrification and the degree to which the prices are growing in the real estate market and the increase of Airbnb listings. These are the main considerations taken, even though there are other variables like the population aging that don't show a clear gentrification pattern but help to identify it in some cases. In the following map (Figure 13[3]) you can find the different neighborhoods at greater risk of being gentrified, according to the findings of this thesis.

FIGURE 13



In conclusion, the clusters derived from the analysis reveal distinct relationships between neighborhoods and the phenomenon of gentrification. These clusters not only capture the

[3] You can download the interactive map from the Github repository: https://github.com/JorPS/Cluster-analysis-Gentrification-in-Madrid/blob/main/Plots/Interactive_maps/Interactive_Madrid_Gentrified_neighborhoods_Map.html

diversity of gentrification experiences within the study area but also highlight similarities with neighborhoods in Madrid where gentrification is not prevalent. Also, it clearly shows how the risk of gentrification is mostly present in the South and East of Madrid.

Cluster 1 signifies the presence of gentrification in the city center, while Cluster 2 identifies neighborhoods of high value attributed to factors such as social class structure, proximity to the center, and various points of interest. Cluster 3 indicates a connection between working-class neighborhoods and recent gentrification, characterized by the persistence of working-class social patterns. Cluster 4 encompasses areas lacking Airbnb supply, indicating a lesser impact of gentrification on a smaller number of neighborhoods. Finally, Cluster 5 represents neighborhoods where early signs of gentrification are observed, driven by the recent influence of Airbnb and a predominantly working-class social structure.

Overall, these findings contribute to a comprehensive understanding of the diverse dynamics and factors associated with gentrification in Madrid. This thesis has the potential to rethink the way the city is evolving and to reduce or avoid gentrification, a problem that negatively affects a lot of neighbors of the urban working class in Madrid, protecting their Right to the City.

5. REFERENCES:

- De la Calle Vaquero, M. (2019). Turistificación de centros urbanos: clarificando el debate. *Boletín de la Asociación de Geógrafos Españoles*, vol. 83, nº 2829, pp. 1–40. http://dx.doi.org/10.21138/bage.2829

- Foment de Ciudad, SA, official web page: https://ajuntament.barcelona.cat/fomentdeciutat/es/quienes-somos

- Lefebvre, H. (1975). "El derecho a la ciudad". *Península*, 3ª ed. Barcelona. [1967].

- Rubiales, M. (2014). "¿Medir la gentrificación? Epistemologías, metodologías y herramientas de investigación de carácter cuantitativo y mixto". *Contested Cities*. Barcelona.

- Ruiz, N., Expósito, V., & Mendoza, P. (2021) "Political critique in madrid's urban art scene: From the late '90s until now". *Communication & Society*, nº2, vol. 34, 387-401.

  DOI: https://doi-org.bucm.idm.oclc.org/10.15581/003.34.2.387-401

- Sequera, J. (2014) "Gentrificación en el centro histórico de Madrid: el caso de Lavapiés". In: Hidalgo, R. & Janoschka, M. (Eds.) "La ciudad neoliberal. Gentrificación y exclusión en Santiago de Chile, Buenos Aires, Ciudad de México y Madrid". *Serie Geolibros* nº 19, pp. 233-25. Santiago de Chile.

  Available at: https://dialnet.unirioja.es%2Fdescarga%2Farticulo%2F3262724.pdf&usg=AOvVaw3KPE0aLFel21o_p99trC9W

- Sorando, D. & Ardura, A. (2016) "First we take Manhattan". *Los Libros de la Catarata*, Madrid.

  Available at: https://bibliotecacomplutense.odilotk.es/info/00544427

- Sorando, D. & Ardura, A. (2018) "Procesos y dinámicas de gentrificación en las ciudades españolas". *Institut d'Estudis Regionals i Metropolitans de Barcelona (ERMB)*. Barcelona.

- Walliser, A. & Sorando, D. (2019) “Las ciudades en España y el impacto de la globalización sobre los sistemas urbanos”. *Informe España 2019*, Madrid. Available at: https://blogs.comillas.edu/informeespana/informe-espana-2019/

6. APPENDICES

Appendix 1: Sources used to extract the variables data

VARIABLE	*SOURCE*
Education level	*Ayto. de Madrid, Open Data Portal. Information panel per dsitrict and neighborhood*: https://datos.madrid.es/portal/site/egob/menuitem.c05c1f754a33a9fbe4b2e4b284f1a5a0/?vgnextoid=71359583a773a510VgnVCM2000001f4a900aRCRD&vgnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCRD&vgnextfmt=default
Immigration	*Ayto. de Madrid, Banco de datos. Population per district, neighborhood and nationality*: Source (2019): https://www-s.madrid.es/CSEBD_WBINTER/seleccionSerie.html?numSerie=0307010000012 Source (2015): https://www-s.madrid.es/CSEBD_WBINTER/seleccionSerie.html?numSerie=0307010000011
Real Estate prices	*Ayto. de Madrid, Banco de datos. Evolution of Real Estate prices of pre-owned housings*: https://www-s.madrid.es/CSEBD_WBINTER/seleccionSerie.html?numSerie=0504030000200
Airbnb data	*Aggregated by Inside Airbnb, the following webpage, web-scraping public information in airbnb.com*: http://insideairbnb.com/get-the-data
Unemployment	*Ayto. de Madrid, Banco de datos. Registered unemployment per neighborhood*:

	Source (2019): https://www-s.madrid.es/CSEBD_WBINTER/seleccionSerie.html?numSerie=0904040000013 Source (2015): https://www-s.madrid.es/CSEBD_WBINTER/seleccionSerie.html?numSerie=0904040000011
Loss of population	Immigration data also contains the population per neighborhood for each year.
Population aging	https://www-s.madrid.es/CSEBD_WBINTER/seleccionSerie.html?numSerie=0301000000001

Appendix 2: Dataset

Link:
https://github.com/JorPS/Cluster-analysis-Gentrification-in-Madrid/blob/main/Datasets%20-%20TFM/Cleaned%20datasets/Madrid_gentrification_data_2019

Appendix 3: Variables codebook

Features	Explanation
Aging_index_2019	Aging index in 2019
College_graduate	Proportion of college graduates in the neighborhood
Compulsory_educ	Proportion of people in the neighborhood who finished their studies in Compulsory Education
FP_or_pre_college	Proportion of people in the neighborhood who finished their studies in Pre-college studies
Incomplete_elementary	Proportion of people in the neighborhood who didn't finish elementary studies
Non_literate	Proportion of non-literate people in the neighborhood.
Post_graduate	Proportion of postgraduates in the neighborhood
Unknown_studies	Proportion of people in the neighborhood with unknown studies
aging_index_incr	Aging index growth during the last 4 years
airbnb_price_incr_last_4_years	Airbnb mean price in the neighborhoods growth during the last 4 years
foreign_incr_last_4_years	Foreign proportion of people in the neighborhood growth during the last 4 years
house_price_2019	Pre-owned housing mean prices in the neighborhood
house_price_incr_last_4_years	Growth of th pre-owned housing mean prices in the neighborhood
mean_airbnb_price	Mean Airbnb price in the neighborhood
n_airbnb_incr_last_4_years	Increment of the number of Airbnb listings in a Neighborhood
n_airbnbs	Number of Airbnb listings in a Neighborhood
population_incr_last_4_years	Population growth during the previous 4 years
prop_foreign	Proportion of foreign people in the neighborhood
unem_incr_last_4_years	Unemployment rate growth during the previous 4 years
unem_prop_2019	Unemployment rate in the neighborhood

Appendix 4: Clusters centers

	C1	C2	C3	C4	C5
prop_foreign	0.2296	0.1492	0.1213	0.1393	0.1284
foreign_incr_last_4_years	0.0242	0.0249	0.0249	0.0343	0.0310
house_price_2019	5148	3760.9907	3209.6524	3226.5714	2594.1667
house_price_incr_last_4_years	0.4145	0.4058	0.3206	0.3424	0.2774
n_airbnbs	659.1667	51.8704	15.6286	0.0000	4.5556
mean_airbnb_price	116.4903	84.8615	70.8708	0.0000	69.3557
n_airbnb_incr_last_4_years	531.6667	46	13.2	0.0000	4.5556
airbnb_price_incr_last_4_years	10.6286	29.6938	-10.2087	0.0000	69.3557
unem_prop_2019	0.0484	0.0378	0.0435	0.0440	0.0508
population_incr_last_4_years	1078.1667	1175.7222	-347.4	3728.5	601.5
unem_incr_last_4_years	-0.0189	-0.0214	-0.0158	-0.0150	-0.0172
Unknown_studies	0.0007	0.0016	0.0012	0.0015	0.0008
Non_literate	0.0257	0.0380	0.0461	0.0507	0.0695
Incomplete_elementary	0.0713	0.1026	0.1041	0.1128	0.1349
Compulsory_educ	0.1797	0.2421	0.2461	0.2592	0.2893
FP_or_pre_college	0.1868	0.1891	0.1951	0.1925	0.1800
College_graduate	0.0984	0.0981	0.1028	0.0908	0.0863
Post_graduate	0.4374	0.3285	0.3047	0.2925	0.2392
Aging_index_2019	0.1642	0.2175	0.1955	0.2084	0.1839
aging_index_incr	-0.0054	-0.0033	0.0093	-0.0024	-0.0044

Appendix 5: Variables importance in clustering analysis

Features	Importance
(Intercept)	11.02
Aging_index_2019	52.29
College_graduate	17.01
Compulsory_educ	18.30
FP_or_pre_college	16.53
Incomplete_elementary	13.63
Non_literate	23.82
Post_graduate	27.77
Unknown_studies	1.04
aging_index_incr	12.71
airbnb_price_incr_last_4_years	23.17
foreign_incr_last_4_years	15.83
house_price_2019	25.92
house_price_incr_last_4_years	26.91
mean_airbnb_price	95.40
n_airbnb_incr_last_4_years	96.71
n_airbnbs	86.40
population_incr_last_4_years	17.90
prop_foreign	41.54
unem_incr_last_4_years	10.68
unem_prop_2019	19.19

Appendix 6: Neighborhoods in risk of gentrification in Cluster 2

	house price	nº airbnbs	nº airbnbs incr	airbnb price	unemployment	post graduate	compulsory educ	no elementary	aging ind. incr	house price incr
San Diego	1.92	82	71	703.780	0.0729	0.1021	0.4089	0.1769	-0.0231	0.4399
Puerta Bonita	1.94	22	20	607.727	0.0658	0.1121	0.3841	0.1761	-0.0205	0.4103
Almendrales	2.18	25	25	436.400	0.0616	0.1149	0.3963	0.1880	-0.0203	0.4418
Numancia	2.08	59	53	528.305	0.0712	0.1156	0.3709	0.1843	-0.0076	0.4260
Vista Alegre	2.04	40	34	452.250	0.0055	0.1230	0.3748	0.1717	-0.0139	0.3801
Palomeras Bajas	2.10	23	22	894.348	0.0639	0.1260	0.3640	0.1599	0.0018	0.5036
San Isidro	2.30	53	49	599.811	0.0072	0.1428	0.3539	0.1727	-0.0112	0.4295
Canillejas	2.36	20	19	643.500	0.0052	0.1490	0.3479	0.1686	0.0047	0.4595
Aluche	2.21	29	27	381.724	0.0446	0.1494	0.3256	0.1730	-0.0053	0.3991
Opañel	2.33	40	36	821.500	0.0540	0.1591	0.3553	0.1533	-0.0110	0.4099
Buenavista	2.50	25	25	487.200	0.0507	0.1629	0.3364	0.1241	-0.0052	0.2727
Pueblo Nuevo	2.49	46	41	685.435	0.0475	0.1640	0.3372	0.1502	-0.0076	0.4403
Lucero	2.38	31	26	544.194	0.0529	0.1697	0.3191	0.1571	-0.0147	0.4294
Ventas	2.66	54	49	650.185	0.0499	0.1755	0.3279	0.1503	-0.0167	0.4595
Comillas	2.70	46	44	896.739	0.0575	0.1805	0.3242	0.1485	-0.0076	0.5538

Appendix 7: Neighborhoods in risk of gentrification in Cluster 3

	Real estate price	Real estate prices increase	Nº of airbnbs listings	Post graduate proportion	Unemployment rate
Casco Histórico de Vallecas	2356.5	0.6422	29	0.0768	0.0782
Arcos	2130.0	0.5720	16	0.1266	0.0622
Portazgo	1826.0	0.4738	10	0.0784	0.0750
Los Rosales	1869.0	0.4355	13	0.1041	0.0601
Moscardó	2209.0	0.4353	29	0.1268	0.0534
Palomeras Sureste	2012.0	0.2718	18	0.0995	0.0659
San Fermín	2041.0	0.2325	16	0.1078	0.0691
Casco Histórico de Barajas	3104.0	0.2327	16	0.1408	0.0494
Fontarrón	2207.0	0.3449	14	0.1424	0.0554

Appendix 8: Neighborhoods in risk of gentrification in Cluster 4.

	Los Cármenes	San Cristóbal
house_price_incr_last_4_years	0.4376	0.5859
unem_prop_2019	0.0714	0.0896
Non_literate	0.0755	0.1153
Incomplete_elementary	0.1586	0.2035
Compulsory_educ	0.3246	0.4613
FP_or_pre_college	0.2002	0.1397
College_graduate	0.0784	0.0284
Post_graduate	0.1608	0.0510

Appendix 9: Privileged and Middle class social structure neighborhoods in Cluster 5.

	Valdemarín	El Plantío	El Goloso	Corralejos	Estrella	**Cluster 5 center**
house_price_2019	4404	2638	4575	3476	4155	**2594.16**
unem_prop_2019	0.0174	0.0169	0.0203	0.0268	0.0268	**0.0508**
Non_literate	0.0122	0.0152	0.0231	0.0241	0.0163	**0.0695**
Incomplete_elementary	0.0447	0.0425	0.0437	0.0439	0.0529	**0.1349**
Compulsory_educ	0.0909	0.1189	0.1064	0.1013	0.1458	**0.2893**
FP_or_pre_college	0.1182	0.1675	0.1463	0.1760	0.1900	**0.1800**
College_graduate	0.1041	0.1078	0.1417	0.1631	0.1344	**0.0863**
Post_graduate	0.6294	0.5476	0.5387	0.4916	0.4596	**0.2392**

Appendix 10: Neighborhoods in risk of gentrification in Cluster 5.

	Amposta	Pradolongo	Hellín	Entrevías	Zofío	**Cluster 5 center**
house_price_incr_last_4_years	0.8105	0.5374	0.4376	0.4267	0.3951	**0.2774**
unem_prop_2019	0.0823	0.0593	0.0708	0.0827	0.0614	**0.0508**
Non_literate	0.1004	0.0780	0.1036	0.1590	0.0771	**0.0695**
Incomplete_elementary	0.2093	0.2247	0.2049	0.2150	0.1926	**0.1349**
Compulsory_educ	0.4197	0.4327	0.3918	0.4072	0.4161	**0.2893**
FP_or_pre_college	0.1542	0.1495	0.1708	0.1314	0.1702	**0.1800**
College_graduate	0.0433	0.0430	0.0460	0.0320	0.0491	**0.0863**
Post_graduate	0.0717	0.0714	0.0808	0.0547	0.0944	**0.2392**