1. **Introduction:**

In recent years, significant changes are occurring in urban centers and internal structures due to shifting demographics and evolving urban dynamics. However, administrative boundaries have remained largely unchanged, highlighting the necessity for a reevaluation in light of these transformations. The city of Lisbon is no exception, and in response to these challenges, the Portuguese government implemented administrative reforms in July 2012. The Parliament approved a proposal, including Chapter Two outlining the plan to reconfigure Lisbon's parish map. This chapter stipulated that the reconfiguration would be based on the principles of rationalization and adjustment of territorial organization, with the aim of creating larger and more balanced parishes. The original 53 parishes were subjected to modifications such as merging or extending, resulting in the current 24 parishes that came into effect in January 2013. The implementation of Lisbon's administrative reform was a complex and challenging process, but it resulted in a more modern and efficient administrative structure for the city. Notably, the most modern and efficient parish, Parque das Nações(Expo), was born from this reform. However, as Lisbon's population continues to diversify and urbanize, the current administrative framework may no longer be sufficient to meet the needs of the city's residents. Therefore, there is an urgent need for a comprehensive and efficient redistricting plan that can adapt to evolving circumstances, consider the diverse needs of the city, and ensure equitable representation for all communities.

In the classic paper on re-scaling cities (Friedmann & Wolff, 1982), the following important points are expressed viz: The importance of understanding the changing territorial organization of cities and regions in the context of global capitalism, as this can impact the spatial distribution of economic activities, populations, and resources. The need for a comprehensive and nuanced approach to studying world cities that takes into account multiple factors, such as political economy, culture, and social processes. The potential for world cities to be sites of both inequality and innovation, where the concentration of economic and social activities can create both opportunities and challenges for different segments of the population.

This study aims to propose a redistricting plan for the Lisbon administrative area based on a data-driven approach, utilizing the 2011 census data to inform the decision-making process. The objective of this research is to provide a comprehensive and well-informed redistricting proposal that takes into account the current and future needs of Lisbon's diverse population. By analysis key demographic variables, such as population density, age distribution, education, employment, housing, and migration patterns, we seek to identify trends, disparities, and opportunities that can inform the redistricting process. In doing so, we aim to ensure that the proposed administrative boundaries are equitable, functional, and reflective of the city's changing demographics.

First of all, the study will commence with a meticulous examination of Lisbon's current administrative structure, followed by an analysis of the 2011 census data to underscore the patterns and trends that will inform the proposed redistricting plan. Subsequently, we will utilize advanced methods such as spatial cluster analysis from data science disciplines, regionalization studies, and other relevant tools to pinpoint Lisbon's specific challenges and opportunities at the granular level of the smallest geographical units. This will allow us to outline the goals and objectives of the redistricting process more effectively. Based on these findings, a comprehensive redistricting plan will be developed, with recommendations for redefining administrative boundaries, reallocating resources, and implementing new policies and plans to address the identified needs.

Ultimately, this research aims to contribute to the creation of a more equitable, functional, and responsive administrative framework that can better serve the diverse needs of Lisbon's residents, both now and in the future.

To propose the different planning or solution for lisbon

The structure of this paper is as follows: In Section 2, a comprehensive analysis of previous and related studies is presented. Section 3 provides an in-depth discussion on the motivation and significance of this research. Section 4 outlines the methodology employed in this study, while Section 5 elaborates on the step-by-step process from cluster analysis to the final re-regional division plan. The discussion of the results and their implications are presented in Section 6. Finally, Section 7 concludes the paper and provides suggestions for future research endeavors.

**2. Literature review**

**2.1 Evolution of Urban Planning: From Ancient Origins to Modern Concepts**

**The difference between a city and other human settlements lies not only in its relatively large size but also in its functions and special symbolic status, which may be granted by central authorities. The term can also refer to the streets and physical structures of the urban area, as well as the group of people living there, and can be used in a general sense to denote urban as opposed to rural areas (Lynch, 2008). The history of urban planning can be traced back to some of the earliest known cities, particularly in the Indus Valley and Mesoamerican civilizations, who built their cities on grids and divided them into different zones apparently for different purposes. The influence of planning is ubiquitous in today's world and can be seen most clearly in the layout of planned communities, which are designed comprehensively prior to construction, often considering interrelated physical, economic and cultural systems (Smith, 2002). Hippodamus of Miletus (498-408 B.C.), an ancient Greek architect and town planner, is widely recognized as the "father of European town planning." He is best known for his influential "Hippodamian Plan" (grid plan), which revolutionized the layout of cities (Glaeser, 2011). During the Second French Empire, under the leadership of Napoleon III, Baron Georges-Eugene Haussmann redesigned the city of Paris into a more modern capital, featuring long, straight, and wide boulevards (Jordan, 1992).**

**In the second half of the 20th century, urban planners gradually shifted their focus towards individualism and diversity in the city center (Routley, 2018). Urban planning guides the development of cities, suburbs, and rural areas in an organized manner by addressing questions related to how people will live, work, and entertain in specific regions (Caves, 2005).Zoning is an urban planning method that divides land into areas with specific regulations for new development. The rules for each zone determine whether planning permission for a given development may be granted. The guidelines set for zoning can include the types of land use allowed, size and dimensions of lots, and the form and scale of buildings. This helps guide urban growth and development in a municipality or other tier of government (Urban Stormwater Management in the United States, 2009).**

**Urban regionalization analysis through spatial data analysis already exists in a number of studies, such as the delineation of policy-relevant urban area boundaries in England through the analysis of spatial economic data (Coombes, 2014). One study suggests that urban planning needs to be reconsidered, requiring a new approach to define the "scale" of spatial practices to keep up with the evolving territorial organization of global capitalism by the late 20th century (Brenner, 1999). A theory of urban reform that was developed a hundred years ago: mobile data sets will provide the basis for defining the boundaries of urban areas (Geddes, 1915). This work has had a profound influence on urban zoning, for example, and later researchers have drawn a very general principle from it that most of the time it is more appropriate to define regional sets first before zoning within a region, and some urban areas are polycentric. (Parr, 2005). Previous studies have also shown that New urban systems are emerging with increasingly polycentric geometries that challenge the traditional urban center model (Keil, 1994).The degree of urbanization is a modern indicator that helps define the composition of a city: "a population of at least 50,000 inhabitants in continuous dense grid cells (> 1,500 inhabitants per square kilometer)" (DIJKSTRA et al., 2020).**

**In many European countries, cities, particularly the larger ones, are further subdivided into administrative units for more granular governance and management. In the UK, cities may be divided into boroughs or districts, which are further broken down into wards and, in some cases, parishes. France's cities, like Paris, are categorized into 'communes', and larger cities have 'arrondissements'. Germany's cities, known as 'Städte', can have subdivisions called 'Stadtbezirke', with smaller neighborhoods termed as 'Ortsteile'. Italian cities, or 'Comuni', have city sectors referred to as 'Quartieri or Circoscrizioni'. In Spain, cities called 'Municipios' have divisions known as 'Distritos'. Dutch cities, termed 'Gemeenten', might be segmented into 'Stadsdelen' or 'Wijken', while Poland's 'Miasta' can be divided into 'Dzielnice'. These categorizations are general, and exact divisions may vary based on the specific country or region.**

2.2 Planning methods : Case studies of transformation in Regional Urban Systems

Regional Studies. In approaching the concept of the regional urban system, attention is initially drawn to the better-known types of economic region. The distinctive nature of the regional economy is next examined, and it is argued that its spatial structure represents an important dimension. Spatial structure can be characterized in a variety of ways, the most comprehensive of which employs the perspective of an urban system. This is examined firstly in terms of particular models from location theory, which provide important points of reference, and then within the setting of the present-day city-region (Parr, 2014).

Following more than 25 years of animated debate among French elected officials and civil servants about territorial reform, the government have finally decided to reduce the number of regions from 22 to 13 in a process dubbed ‘le big bang des régions’ by the French media.The previous structure was notoriously complicated, with 36,700 communes, 2,600 inter-communal groupings, 101 departments and 22 regions, leading to waste, duplication, and a lack of transparency. The aim of the reform is to make the public sector in France more dynamic, responsive, and better adapted to the geography of the modern economy, while freeing up the regions to focus on economic development by devolving power to local levels on more everyday issues (Mcnally, 2016).

In 1999, Poland underwent administrative reforms that reduced the number of top-level administrative districts from 49 to 16. The focus of the responsibilities of these districts was also restructured to prioritize overall regional development, higher education, regional infrastructure, and the management of EU funds. This resulted in the creation of 16 voivodeships, 308 counties, 66 county towns, and 2,478 municipalities. When considering changes to the territorial division, it is important to consider combining urban and rural areas in sparsely populated regions or areas with high concentrations of service infrastructure in central towns. Any proposed changes should be reviewed by experts and the public before implementation. Changes that have sufficient justification and social acceptance can be introduced gradually or through administrative pilots, which typically test more complex reforms in selected administrative units or in more administratively efficient regions (Kaczmarek, 2016).

Denmark also underwent administrative reforms in early 2007. The Danish territorial reform aimed to create municipalities and regions that were financially and professionally sustainable. Before the reform, municipalities in some areas of the country had already started merging, with referendums leading to the formation of single municipalities on islands like Bornholm, Ærø, and Langeland. The structural agreement recommended a minimum size of 30,000 inhabitants for new municipalities, with a minimum of 20,000 inhabitants required for the formation of new municipalities or binding cooperation with neighboring municipalities. As a result, after the reform, the average municipality size increased from around 20,000 inhabitants to about 55,000 inhabitants, and the number of municipalities with less than 20,000 inhabitants decreased significantly. The regions were also created to have an improved professional and financial basis to perform their tasks in providing high-quality healthcare to their populations. The regions have a population of between approximately 0.6 million to 1.6 million inhabitants, significantly larger than the counties before the reform (*Evaluation of the Local Government Reform*, 2013).

The network partitioning method of area division has been studied (Hamilton & Rae, 2020). The author used a network partitioning algorithm called Combo to analyze commuting data from the 2011 UK Census. The goal was to improve the methodology of regional partitioning and produce practical results. By using this approach, the author was able to create 17 new regions for Scotland, which is a reduction from the existing 32 regions.

2.3 The case of Lisbon

The administrative zoning reform in Lisbon can be traced back to the 1880s. The Law of July 18, 1885, consisting of 230 articles, was meticulously crafted through extensive consultations and legislative processes. This law provided a comprehensive legal framework for the administrative division of the city of Lisbon. Recognizing the unique needs of this important city, the government and legislative institutions of that era realized the necessity for a specialized management model. Consequently, they developed this law, which divided Lisbon into four administrative boroughs.

The most recent administrative zoning reform occurred in 2012. Portuguese Law No. 56/2012, enacted on November 8, introduced a new administrative reform for the City of Lisbon. This reform led to the decentralization of power from the Lisbon City Hall to each of the 24 "Junta de Freguesia" (parish councils) that make up the municipality. The motivation behind this administrative reorganization stemmed from the need to modernize and adapt the governance model of the city of Lisbon. This need was further compounded by the fact that Lisbon serves as the national capital and hosts various national government institutions. Additionally, it was influenced by the size and administrative disparities among the existing parishes within the county. Between 2009 and 2012, the Lisbon City Council, with the crucial involvement of Seixas, actively prepared and approved an administrative reform. This reform addressed both vertical changes at the parish level and horizontal restructuring within municipal departments (Marcuse, 2010). Thus, the reform plan was developed through extensive consultation and discussions to better meet the administrative and governance needs of Lisbon.

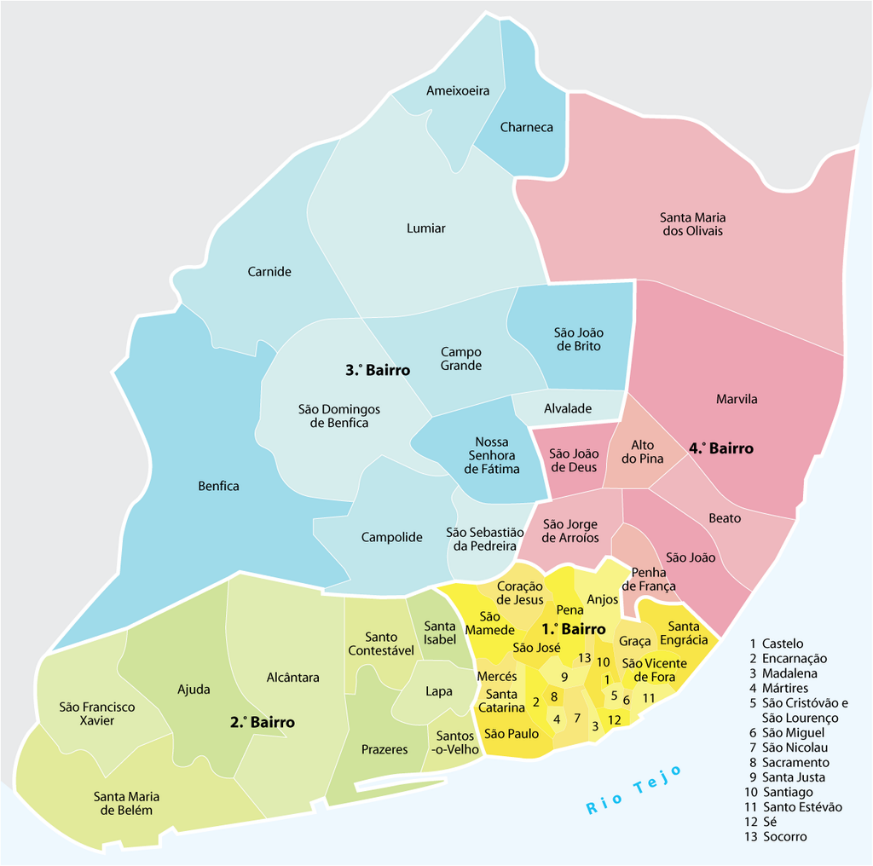
From this historical perspective, it's evident that most of Lisbon's administrative reforms have been developed based on negotiation and discussion, often in response to various objective factors. Therefore, we have decided to conduct this research with the aim of formulating an administrative reform plan for Lisbon under a data-driven approach.

1. **Methodology**

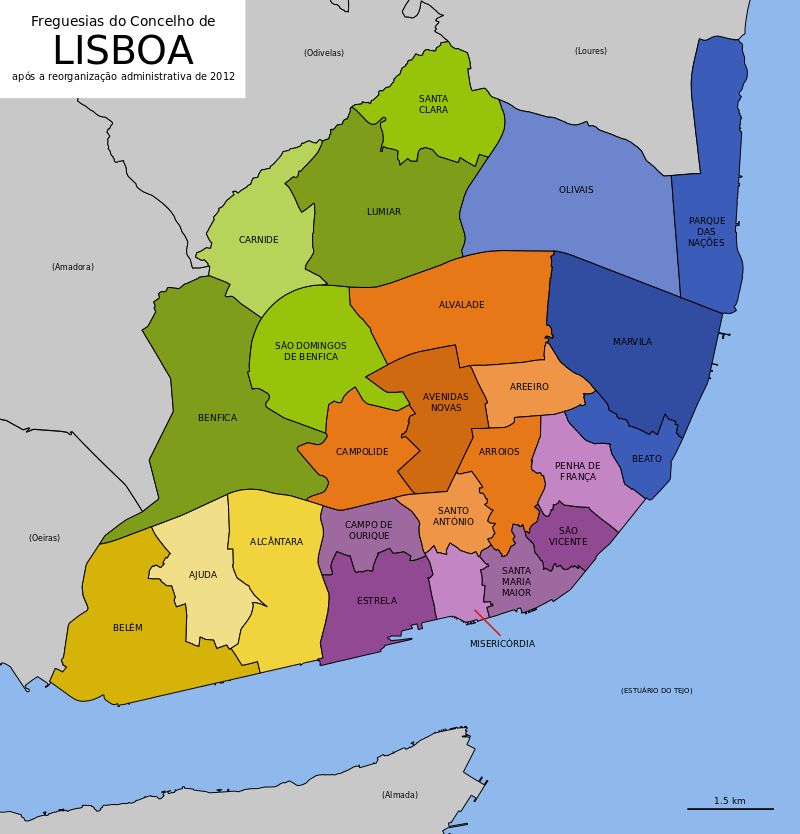
****3.1 Overview of demographic data and information in Lisbon****

**This study focuses on developing a comprehensive redistricting plan for the Lisbon administrative area. The primary goal is to ensure that the proposed administrative boundaries are equitable, functional, and reflective of Lisbon's changing demographics. This study employs a mixed-methods approach, combining quantitative and qualitative data analysis techniques.**

**In 2011, the Área Metropolitana de Lisboa had a population of 26.7% of the national total, with an annual growth rate of 0.6%. The population density was 940.0 individuals per km², and the index of aging was 117, indicating that there were more elderly people than young people. The majority of families consisted of two or more people, with a growing number of people living alone. The percentage of foreign residents was 7.2%, and the population with higher education was 19.6%,** **From Pordata (a database of certified statistics about Portugal) .**



**In 2021, the population of the Área Metropolitana de Lisboa increased to 27.8%, with a lower annual growth rate of 0.2%. The population density also increased to 951.9 individuals per km². The index of aging also increased to 151, indicating a more significant number of elderly people compared to young people. The percentage of foreign residents increased to 8.9%, and the population with higher education increased to 26.6%. The number of households living alone and using transportation colectivo also increased,** **From Pordata.**



****3.2 Data** Preparation**

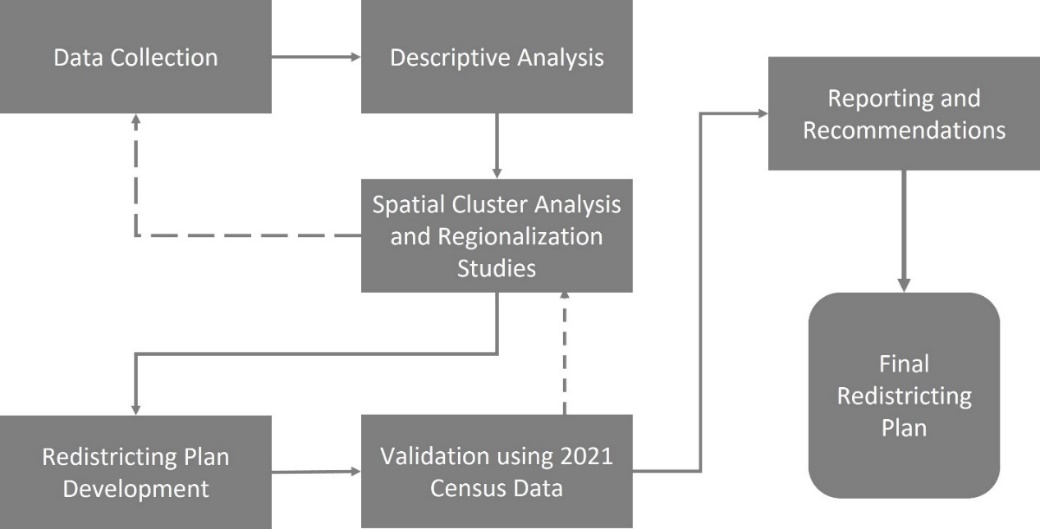
**The primary data source for this research is the 2011 Portuguese National Census conducted by the Portuguese National Institute of Statistics (INE), as mandated by law every ten years, from March 21 to May 2011. Census enumerators collected data from households and individuals across the country. The INE was responsible for organizing and implementing the census, ensuring the accuracy of data collection and analysis. The comprehensive results and data from the census were gradually released starting on July 29, 2011. According to the census, Lisbon's population in 2011 was 547,733, a decrease from 564,477 in 2001, with a population density of approximately 6,378 people per square kilometer. Lisbon's population was aging, with a higher proportion of elderly people and a median age of around 42.2 years. Compared to other regions in Portugal, Lisbon had smaller average household sizes, with 2.4 people per household. Lisbon had a relatively high number of vacant housing units, accounting for about 12.2% of the total housing stock. The educational level of Lisbon's population was higher than that of other regions in Portugal, with a higher proportion of individuals completing higher education. The unemployment rate was higher than the national average, but the city also had a higher proportion of individuals engaged in professional, scientific, and technical activities.**

**Before conducting any analysis, the raw census data will be preprocessed, cleaned, and transformed as needed to ensure its usability for the study. In this study, we utilized features derived from census data. Since all the data values are non-negative, starting from zero and upwards, we employed the Non-negative Matrix Factorization (NMF) technique for dimensionality reduction. NMF is particularly suitable for datasets with non-negative values as it factorizes the original data into two lower-dimensional matrices, ensuring that all values remain non-negative. This property of NMF makes it an ideal choice for preserving the inherent structures and patterns present in datasets like census data, where negative values wouldn't have a meaningful interpretation.**

**3.3 Redistricting Lisbon parishes using a data science approach**

3.3.1 Processing Redistricting plan

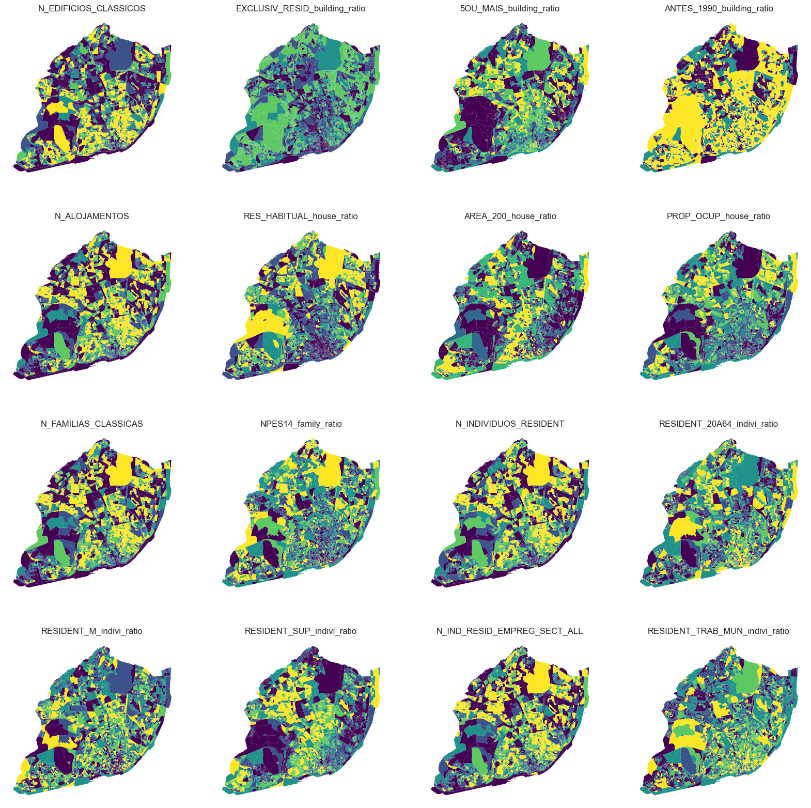
The redistricting plan development will involve integrating the findings from the descriptive analysis and spatial cluster analysis to propose new administrative boundaries, resource allocations, and policies. Here is the flow chart to process the redistricting plan.



* + 1. Descriptive Analysis

Initially, we constructed a list which encompassed all the proportion variables derived in the previous steps. Additionally, we generated a list containing all the proportion variables alongside the primary variables of distinct categories, including the total number of classic buildings, total number of accommodations, total number of classical families, and total number of resident individuals. These lists serve as a systematic approach to organizing the variables for subsequent analysis and ensuring a comprehensive examination of the relevant factors.

1. The spatial distribution maps



From the figure, it can be observed that there is a higher concentration of buildings in the center of Lisbon. However, the proportion of exclusively residential buildings is low in the center, while the proportion of tall buildings is higher in the north of the center. Additionally, the proportion of old buildings is quite high in most areas, except for some regions in the north.

The number of accommodations is also high in the city center, but the proportion of conventional dwellings for habitual residence is notably low in the center. In the western part of the city center and Expo, the proportion of dwellings larger than 200 m2 is high. Furthermore, the proportion of owner-occupied conventional dwellings is higher in some northern areas and Expo.

Excluding the Avenida Liberdade area, the number of families is high in most of the central areas of Lisbon. The proportion of families with children under 15 years old is higher around the city center. The visualization of the number of resident individuals is similar to that of families. The proportion of residents aged 20 to 64 years old is dispersed throughout the city, with noticeable concentrations in the center, Expo, and Benfica. The proportion of female individuals is evenly distributed across the city, as illustrated in the chart. The proportion of residents who have completed their university studies is high in areas ranging from the west of the center to Benfica, as well as in Belem and Expo.

The number of employees is high in most areas, with the exception of the city's fringes and certain border regions. The proportion of resident individuals working within their municipality of residence increases as one moves closer to the city center, but there are also a few highlighted fringe areas.

* + 1. Spatial Cluster Analysis and Regionalization Studies

1. Max-P Regionalization

The max-p problem focuses on the aggregation of various geographical zones into as many homogenous regions as possible, ensuring that a spatially extensive attribute exceeds a predetermined benchmark. This attribute can be designed to guarantee that every region maintains a significant population size or contains a minimum number of enumeration units. Unlike methods where a set number of regions is predetermined, the count of regions in the max-p problem emerges organically from the data, making it apt for regionalization scenarios where there's no need to establish a region count beforehand.

Initially conceptualized as a mixed-integer problem by Duque, Anselin, and Rey in 2012, the max-p is acknowledged as an NP-hard challenge, with precise solutions being practical only for smaller datasets. Consequently, several heuristic methodologies have been proposed to address it. The heuristic strategy detailed in Wei, Rey, and Knaap's 2020 study is incorporated into PySAL.

3.3.4 Comparison based on Jaccard Index.

The Jaccard Index, also known as the Jaccard similarity coefficient, is a statistical measure used to gauge the similarity between two sample sets. When applied to geographical imagery or spatial data, it aids in ascertaining the level of similarity or overlap between two distinct geographic regions or images.

Specifically, the Jaccard Index is calculated as:

where and are two sample sets.

For geographical imagery or Geographic Information System (GIS) data, these sets could represent specific regions of geographical space, such as land use types, land cover, or other geographic features. The value of the Jaccard Index ranges between 0 and 1, with 1 indicating complete overlap of the two spatial sets, and 0 indicating no overlap at all.

For instance, consider two land use images, A and B. If you wish to discern how similar the land use types in these two images are, the Jaccard Index can be employed to measure their similarity. By examining the overlap and union of land-use pixels in A and B, a value for the Jaccard Index can be obtained.

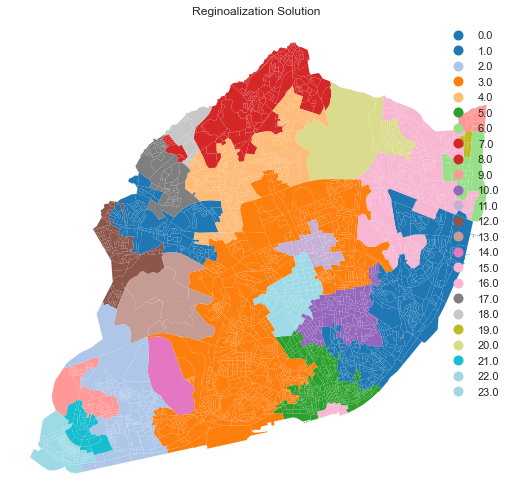
In summary, the Jaccard Index offers an effective means to quantify and compare the similarity between geographical images or spatial datasets.

1. **Results**

4.1 NMF

Firstly, we developed a program to determine the optimal number of components for the NMF (Non-negative Matrix Factorization) model. This program calculates results and reconstruction error scores for components ranging from one to twenty. Then, it uses the "elbow method" to identify the optimal number of components. The elbow method involves measuring the distance of each point to the two points adjacent to it. The point where this distance is maximized is considered the "elbow." Finally, all results, along with the optimal number of components, are visually presented in a chart.

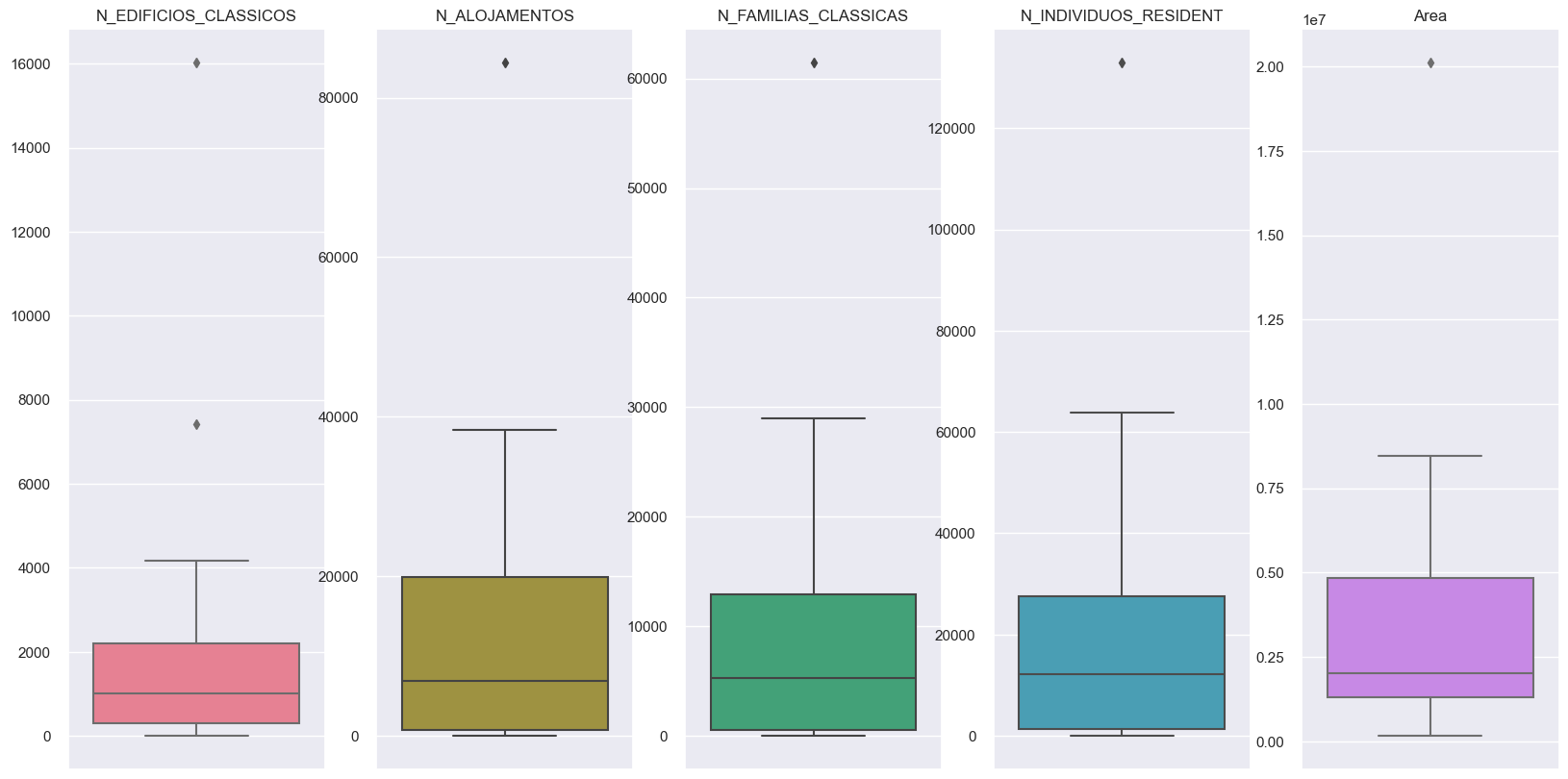
* 1. Redistricting Plan



To integrate the findings from the descriptive and spatial cluster analysis, we first create a new database that includes the result labels and important features with significant practical significance in the original data (prior to normalization processing). The key considerations during this process include ensuring the legitimacy of community areas, maintaining community integrity, and promoting effective governance.

For the selected raw data, we will include the number of buildings, number of residents, number of households, area of the region, and geographic data columns. Additionally, we will generate new features based on the original data, such as the total number of residents/total area, total number of residences/total area, total number of residences/total number of buildings, total number of families/total number of residences, total number of residents under the age of 19/total number of residents, total number of residents over the age of 65/total number of residents, total number of university graduates over 19 years old/total number of adult (20-65 years old) residents, total number of adult female residents/total number of adult residents, total number of workers/total number of adult residents, and total number of local workers/total number of workers.

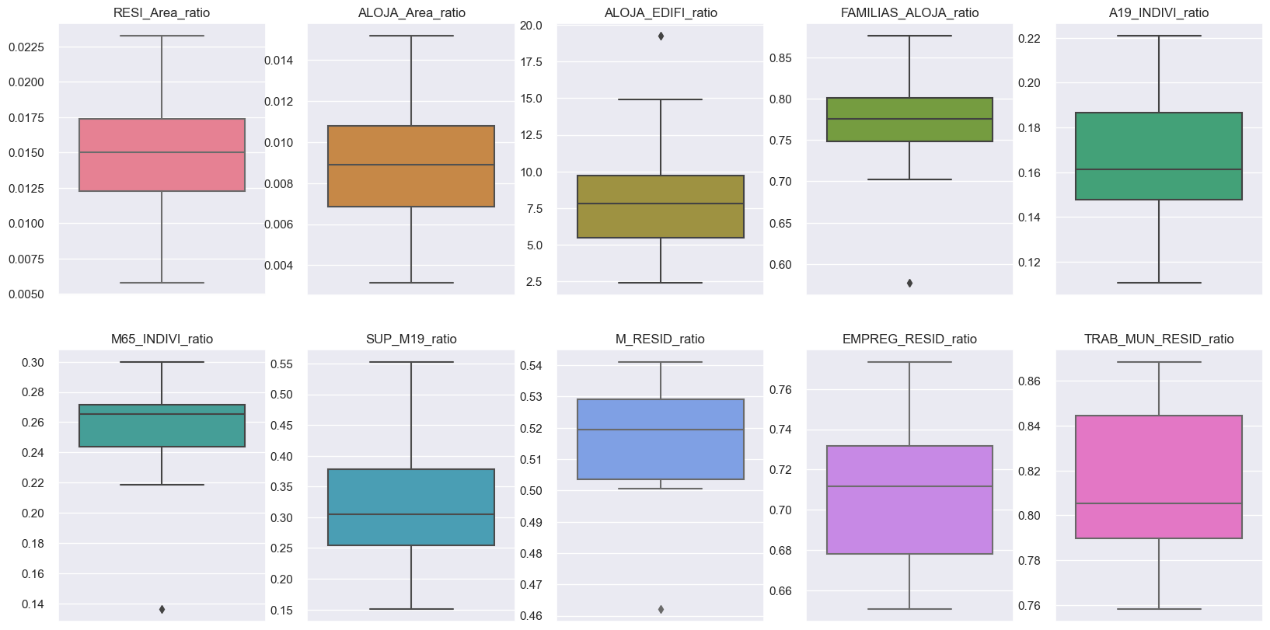
Next, we will dissolve our database to obtain the raw data for each label group. During this step, we will sum the quantities, average the proportions, and calculate the number of subsections for each label. And we will analyze the numerical data and visualize the results in the form of box plots as shown below.



Upon analyzing the quantity data, we observed several discrepancies, such as extremely large and close-to-zero values. While large values are understandable for regions with large areas and high building densities, values close to 0 are unrealistic. In urban planning, it is crucial to ensure that each district's population area falls within a reasonable range. A district with an extremely small area or population is meaningless and may result in administrative bloat and misuse of public resources. Therefore, we plan to merge such areas with the closest geographically adjacent regions, similar to how we previously handled the Lisbon Airport area. This approach will enable us to ensure the legitimacy of community areas, maintain community integrity, and promote effective governance.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Parish labels** | **Buildings** | **Houses** | **Families** | **Residents** | **Area** | **SubSections** |
| **19** | 1 | 1 | 0 | 0 | 1.62E+05 | 7 |
| **15** | 3 | 10 | 5 | 242 | 2.04E+05 | 13 |
| **18** | 36 | 113 | 93 | 212 | 7.67E+05 | 13 |
| **6** | 46 | 860 | 550 | 1382 | 1.39E+06 | 21 |
| **14** | 64 | 109 | 81 | 168 | 1.64E+06 | 1 |
| **13** | 79 | 95 | 78 | 306 | 4.02E+06 | 6 |

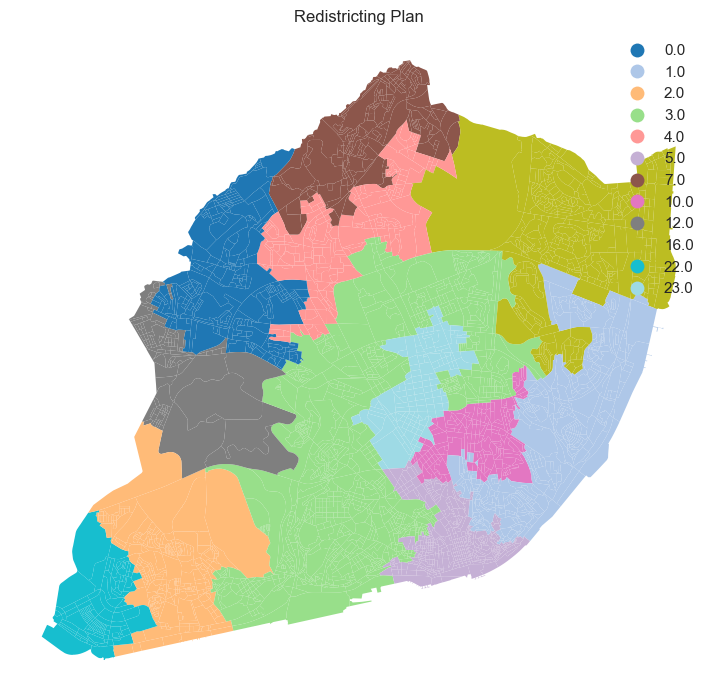
The above table presents label groups whose quantity values are below the first quartile. In order to improve the rationality of our analysis, we will reclassify these labels join into the nearest labels. Then we continue to analyze the ratio values we created after integrating the data, as shown below.



Continuing the analysis of our regional division plan, we found that all the proportion data basically conform to the normal distribution or partial normal distribution, indicating that our plan is quite reasonable so far. We have also identified several outliers in the figure, which can be used to explain the particularity of the areas to which they belong. For example, the high ratio of accommodation to buildings in the No. 4 label area indicates the presence of many high-rise or large buildings in this area. Additionally, the extremely low proportion of the number of 65-year-olds in this area also suggests that this area is a young and modern area.

Moreover, we found that the extremely low household-to-accommodation ratio in the No. 5 labeled area indicates that there are fewer families living in this area. Furthermore, the proportion of adult women in this area is also extremely low, which may indicate a lack of job opportunities or other amenities that are typically more attractive to women.

Overall, our analysis has shown that our regional division plan is based on a sound methodology and is supported by empirical evidence. We believe that our plan will ensure the legitimacy of community areas, maintain community integrity, and promote effective governance, by taking into consideration important factors such as population density, demographic characteristics, and urban planning principles. Below is the final Redistricting Plan.



1. **Discussion**

Regarding the number of clusters, it is important to choose an appropriate number of clusters that provides meaningful information and insights without overfitting the data. Overfitting occurs when the number of clusters is too high, causing the model to fit the noise in the data, rather than the underlying patterns. On the other hand, if the number of clusters is too low, important patterns in the data may be missed. The limitation on the number of clusters refers to the fact that clustering algorithms, such as K-means, require the user to specify the number of clusters beforehand. This can be difficult in situations where the optimal number of clusters is not immediately obvious, and there is a risk of choosing an inappropriate number of clusters which can lead to suboptimal results. In addition, as the number of clusters increases, the interpretability of the results can become more difficult, as it becomes harder to assign meaningful labels to each cluster.

Additionally, while data can provide valuable insights, it cannot explain everything. Other factors, such as historical and cultural contexts, political dynamics, and individual experiences and behaviors, may play a significant role in shaping the trends and patterns observed in the data. Therefore, it is important to consider data in conjunction with other sources of information to gain a more comprehensive understanding of a given phenomenon. Data can't explain everything, this means that there are always limitations to what we can learn from data. For example, data may not capture all the relevant factors that influence a particular phenomenon, or there may be inherent biases or measurement errors in the data. Additionally, while data can provide us with insights into patterns and trends, it cannot always explain why these patterns and trends exist, and further research and analysis may be required to fully understand a particular phenomenon. Therefore, while data can be a valuable tool for understanding the world around us, it is important to approach it with a critical eye and be mindful of its limitations.