REVISTA CIENTIFICA INNOVACIÓN INTEGRAL
ISSN: 3103-1420
DOI: https://doi.org/10.70577/eag5rs05

Volúmen 1, Número 1 (2025) Pág. 16-33

Artificial Intelligence for Urban Planning in Latin America

Inteligencia Artificial para la planificación urbana en Latinoamérica

Jorge Javier Analuisa-Sánchez
jorge.analuisa@unach.edu.ec
https://orcid.org/0009-0005-4429-7752
Universidad Nacional de Chimborazo
Ecuador

Abstract.- Urban planning in Latin America faces significant challenges due to rapid urban growth, socioeconomic inequality, and environmental vulnerability. With more than 80% of the population living in urban areas and a projected 90% by 2050, it is essential to optimize resource distribution and improve public services through data-driven approaches. This article proposes the use of clustering algorithms as key tools to identify homogeneous areas within cities, facilitating more equitable and sustainable planning. Using data science techniques such as K-means and DBSCAN, urban indicators are analyzed grouped into three dimensions: infrastructure (access to water, electricity, transportation), socioeconomic (income, education, health), and territorial (land use, green spaces). These methods allow the segmentation of critical areas, such as informal settlements or areas with infrastructure deficits, improving decision-making in public policies. The analysis is based on a synthetic dataset of 5,000 records, generated with realistic statistical distributions based on recent studies. Advanced techniques such as PCA are applied to reduce dimensionality, variable normalization, and validation metrics such as the Calinski-Harabasz index. The results show a bipolar urban structure with two well-defined clusters using K-means, while DBSCAN identifies multiple transition zones and spatial noise, typical of dynamic and informal urban contexts. It is therefore concluded that the combination of clustering, geospatial analysis, and data-driven strategies offers a robust methodology for guiding urban policies in Latin America, promoting equity and resilience to climate change.

Keywords: *cluster, data, geospatial, planning, urbanism.*

Resumen.- La planificación urbana en América Latina enfrenta desafíos significativos debido al rápido crecimiento urbano, la desigualdad socioeconómica y la vulnerabilidad ambiental. Con más del 80% de su población viviendo en zonas urbanas y una proyección del 90% para 2050, es fundamental optimizar la distribución de recursos y mejorar los servicios públicos mediante enfoques basados en datos. Este artículo propone el uso de algoritmos de clustering como herramientas clave para identificar áreas homogéneas dentro de las ciudades, facilitando una planificación más equitativa y sostenible. Mediante técnicas de ciencia de datos, como K-means y DBSCAN, se analizan indicadores urbanos agrupados en tres dimensiones: infraestructura (acceso a agua, electricidad, transporte), socioeconómica (ingresos, educación, salud) y territorial (uso del suelo, espacios verdes). Estos métodos permiten segmentar áreas críticas, como asentamientos informales o zonas con déficit de infraestructura, mejorando la toma de decisiones en políticas públicas. El análisis se apoya en un conjunto de datos sintético de 5000 registros, generado con distribuciones estadísticas realistas basadas en estudios recientes. Se aplican técnicas avanzadas como PCA para reducir dimensionalidad, normalización de variables y métricas de validación como el índice de Calinski-Harabasz. Los resultados muestran una estructura urbana bipolar con dos clusters bien definidos por K-means, mientras que DBSCAN identifica múltiples zonas de transición y ruido espacial, típico de contextos urbanos dinámicos e informales, por lo que se conluye que la combinación de clustering, análisis geoespacial y estrategias basadas en datos ofrece una metodología robusta para guiar políticas urbanas en América Latina, promoviendo la equidad y la resiliencia frente al cambio climático.

REVISTA CIENTIFICA INNOVACIÓN INTEGRAL
ISSN: 3103-1420
DOI: https://doi.org/10.70577/eag5rs05

Palabras clave: *cluster, datos, geoespacial, planificación, urbanismos.*

1. Introduction

The optimization of urban planning in Latin America through clustering algorithms represents a critical intersection of data science, sustainability, and equity in response to the region's growing urban challenges. With more than 80% of the population currently residing in urban areas—projected to increase to 90% by 2050—cities such as São Paulo and Mexico City are grappling with significant socioeconomic disparities, inadequate infrastructure, and environmental vulnerabilities that threaten urban livability and inclusiveness [1].

Los algoritmos de agrupamiento, que analizan Clustering algorithms, which analyze spatial and demographic data, are increasingly used to inform urban planning strategies, allowing for a more nuanced understanding of urban dynamics and equitable resource distribution [2].

Notable case studies in cities like Bogotá, São Paulo, and Buenos Aires illustrate practical applications of clustering techniques, revealing insights into social segregation, resource allocation, and environmental conditions [3], [4], [5].

These data-driven methodologies have been fundamental in identifying underserved neighborhoods, optimizing public transport routes, and enhancing climate resilience, ultimately fostering more sustainable and equitable urban environments. However, implementing these approaches is not without challenges, including ethical considerations around data collection, governance issues, and technological disparities that may hinder effective urban management [6].

Controversies also arise regarding the effectiveness and applicability of different clustering algorithms in dynamic urban contexts, as current studies often focus on relative performance without establishing comprehensive standards for urban mapping [7].

Moreover, the reliance on data science in urban planning raises questions about inclusivity, particularly concerning how well these approaches capture the needs and voices of marginalized communities [8].

As Latin American cities continue to evolve, integrating clustering algorithms with participatory planning processes will be essential to achieving equitable and sustainable urban futures [9].

Urban Challenges in Latin America

Latin America faces a multitude of urban challenges exacerbated by rapid urbanization and socioeconomic disparities. To date, over 80% of the region's population lives in urban areas, with projections indicating this figure could reach 90% by 2050 [10]. Despite significant urban growth, the benefits of urbanization have not been distributed equitably, resulting in persistent inequalities.

Socioeconomic Inequality

The urban landscape in Latin America is characterized by marked social and spatial segregation. Approximately 26% of the urban population lives in poverty or extreme poverty. Housing policies have struggled to address the needs of the poorest populations, leaving many without access to adequate housing and essential services [11].

ISSN: 3103-1420
DOI: https://doi.org/10.70577/eag5rs05

Urban Transport and Infrastructure

Urban transport significantly contributes to energy consumption and greenhouse gas emissions in Latin American cities, resulting from urban planning that has favored private car use over sustainable public transport systems [12].
The rapid expansion of urban areas—often at rates two to three times higher than population growth—has led to increased infrastructure costs and congestion problems [13].

Cities continue to face challenges in developing efficient and sustainable transport systems, which are critical to improving urban livability and reducing environmental impact [14].

Resilience and Climate Security

As cities grow, they must also adapt to increasing threats posed by climate change. Urban areas are particularly vulnerable to climate-related events, requiring comprehensive vulnerability assessments and mitigation strategies [15].

Enhancing urban security is crucial to protect residents from crime and accidents, which are amplified in densely populated areas lacking adequate resources and planning [16].

Governance and Political Inclusion

Effective urban governance is essential to address these multifaceted challenges. Inclusive political processes are needed to meet the diverse needs of urban populations and ensure that development benefits all residents equitably [17].

Urban management practices must be restructured to promote sustainability, equity, and resilience, aligning with global initiatives such as the United Nations Conference on Sustainable Development (Rio+20), as highlighted by Al-Sehrawy [18].

Clustering Algorithms

Clustering algorithms play a fundamental role in urban planning by facilitating the identification of distinct geographic units within a city that exhibit internal homogeneity while remaining distinct from surrounding areas. These algorithms help analyze urban structure based on various factors including ethnicity, socioeconomic status, and the built environment, thereby facilitating the measurement of residential segregation and evaluation of public spaces' contributions to territorial agglomerations based on race and ethnicity [19].

Types of Clustering Algorithms

Urban clustering studies encompass several algorithm types, generally categorized as hierarchical and non-hierarchical methods. Hierarchical clustering, including both agglomerative and divisive approaches, is particularly prevalent in urban studies. In the agglomerative method, each data point starts as its own cluster and progressively merges with nearest clusters based on pairwise similarities, while the divisive method starts with a large cluster and divides it until each data point is isolated [20].

Hierarchical Clustering

Hierarchical clustering methods have notable strengths and weaknesses. They provide a comprehensive view of data relationships through dendrograms; however, they can be sensitive to outliers, computationally expensive, and difficult to interpret due to their complexity. These methods also require the number of clusters to be predetermined, which may limit their applicability in dynamic urban contexts [21].

Gaussian Mixture Models and Deep Modularity Networks

In recent years, more advanced techniques such as Gaussian Mixture Models (GMMs) and Deep Modularity Networks (DMONs) have emerged in

REVISTA CIENTIFICA INNOVACIÓN INTEGRAL
ISSN: 3103-1420
DOI: https://doi.org/10.70577/eag5rs05

the field of urban clustering. GMMs are effective for spatially distributed data, while DMONs operate on attributed graphs, allowing integration of node attributes and structural information of the urban environment. These methods enrich node representations, providing valuable insights for urban planning and analysis [22].

Empirical Validation of Clustering Techniques

To evaluate the effectiveness of different clustering methods, empirical studies have been conducted, such as one involving residents of Santiago, Chile. This study compared clustering solutions generated via GMMs and DMONs, focusing on factors like spatial resolution and data representation granularity [23].

The results of these comparisons inform the selection of appropriate algorithms based on specific urban characteristics and planning goals, emphasizing the importance of method choice for achieving sustainable and equitable urban development [24].

Data Science Approaches

Data science has emerged as a transformative tool for urban planning in Latin America, leveraging machine learning (ML) and big data to enhance decision-making processes and promote sustainable development. These approaches facilitate the analysis of complex datasets, enabling planners to identify patterns and trends that inform equitable housing policies and resource allocation [25].

Machine Learning in Urban Planning

Machine learning algorithms can process large amounts of data, including land-use patterns, demographic information, and housing market trends. By doing so, they help policymakers and developers identify suitable areas for affordable housing initiatives, fostering social inclusion and improving overall living conditions [26].

The integration of AI technologies, such as Building Information Modeling (BIM) and robotic automation, further streamlines construction processes, reducing costs and improving project timelines [27].

Moreover, machine learning can assist in early detection of construction defects, improving quality control and ensuring long-term resilience in urban development projects [28].

Challenges in Data Utilization

Despite potential benefits, several challenges hinder effective use of data science in urban planning. Limited data collection systems, fragmented datasets, and privacy concerns pose significant barriers [29].

To overcome these issues, it is crucial to improve data management practices, foster data sharing, and encourage stakeholder collaboration. Initiatives such as special data centers for sensitive public data can provide secure environments for researchers, thereby promoting better policy interventions and urban research [30].

Big Data Strategies

Cities like Cape Town, South Africa, exemplify successful implementation of big data strategies to improve local governance. The city has invested in IT infrastructure and skilled personnel to develop a comprehensive data strategy, introducing innovations such as data sandboxes. These allow data scientists to access rich datasets to address urgent policy challenges, from labor market issues to climate change adaptation [31].

Such frameworks underscore the growing recognition of data's power to optimize service delivery and enhance planning processes in Latin American and Caribbean urban environments [32].

REVISTA CIENTIFICA INNOVACIÓN INTEGRAL
ISSN: 3103-1420
DOI: https://doi.org/10.70577/eag5rs05

Future Directions

Urban planning optimization in Latin America is poised for significant advances through the integration of innovative strategies and technologies. To achieve more sustainable and equitable cities, several key areas merit attention [33].

Adopting Participatory Planning

A critical aspect of urban sustainability lies in participatory planning, ensuring that urban development reflects the needs and aspirations of local communities. This approach aligns with the right to the city, emphasized in various Latin American contexts, particularly Brazil, Ecuador, and Mexico [34], [35], [36]. By involving residents in the planning process, cities can foster social integration and counter prevailing trends of segregation and exclusion.

Leveraging Smart Technologies

The adoption of smart technologies in urban planning offers the potential to improve quality of life in cities. Automated transport systems, resource consumption sensors, and applications to optimize urban services can contribute to more efficient and responsive urban environments [37]. These technologies not only enhance service delivery but also facilitate better resource management and environmental sustainability.

Addressing Climate Change Resilience

With climate change presenting an urgent challenge, urban planning must incorporate resilience strategies. This includes designing infrastructure capable of withstanding extreme weather events, reducing carbon emissions, and promoting sustainable mobility [38]. Cities must prioritize adaptive measures that improve their capacity to cope with climate impacts while ensuring the well-being of residents.

Improving Public Spaces and Green Infrastructure

Investment in public spaces and green infrastructure is essential to promote socialization, identity, and overall quality of life [39]. By creating parks and green areas, cities can enhance urban health and well-being, as highlighted by the World Health Organization. The development of such spaces should be carried out through participatory processes to foster community ownership and improve local identity.

Using Data-Driven Approaches

Implementing clustering algorithms and data science techniques can significantly improve urban planning outcomes. By effectively analyzing urban data, planners can identify patterns and optimize land use, ensuring that development is compact and socially integrated [40]. These data-driven approaches can facilitate better decision-making and enhance the responsiveness of urban policies to the needs of diverse populations.

2. Materials and Methods

2.1 Statistical Models

Methodological Analysis of the Clustering Process

The analysis was structured in several stages to ensure data quality, reduce dimensionality, and apply spatial clustering techniques. Each of the implemented phases is detailed below:

Data Preprocessing

As an initial step, variables were normalized using *StandardScaler*. This process ensures that all features have a comparable scale, avoiding biases during distance calculations and contributing equally to subsequent analysis.

Dimensionality Reduction

ISSN: 3103-1420
DOI: https://doi.org/10.70577/eag5rs05

To simplify the feature space without losing relevant information, *Principal Component Analysis (PCA)* was applied. The optimal number of components required to explain at least 95% of the total variance of the original dataset was selected. As part of this process, a plot was generated showing the cumulative explained variance per component, allowing visualization of each component's relative contribution to the model.

Application of Clustering Techniques

Two complementary clustering methods were employed: *K-means* and *DBSCAN*, with the aim of exploring different spatial patterns in urban data.

K-means: To determine the optimal number of clusters, the *Elbow Method* and *Silhouette Analysis* were used. These tools help identify the value of *k* that maximizes internal cohesion and inter-group separation. Additionally, cluster quality was assessed using the *Silhouette coefficient*, and plots were generated to visualize both criteria.

DBSCAN: This algorithm was used to detect irregular clusters and spatial outliers or noise. The ε (epsilon) parameter, which defines the maximum neighborhood distance between points, was estimated using the *k-distance graph*. The 2% percentile of the ordered distances was adopted as a threshold, allowing the identification of high-density regions while controlling sensitivity to noise.

Evaluation of Clustering Results

The quality of the resulting clusters was assessed using the *Calinski-Harabasz Index*, a metric that measures the ratio of between-cluster dispersion to within-cluster dispersion. This measure allowed for an objective comparison of the performance of both algorithms—K-means and DBSCAN—in terms of the clarity and definition of the identified groups.

Spatial Visualization of Clusters

Finally, geographic representations of the clustering results were produced using *scatterplots* based on simulated latitude and longitude coordinates. Distinct color palettes were used to differentiate the resulting groups: a continuous scale for K-means and a contrasting combination for DBSCAN, including the points classified as noise.

This methodological approach combined advanced techniques of preprocessing, dimensionality reduction, and clustering to identify meaningful spatial patterns in urban contexts. The integration of multiple analytical strategies not only enables efficient segmentation of urban areas but also facilitates robustness evaluation through statistical metrics and geospatial visualizations.

2.2 Data Used

Description of the Synthetic Dataset Variables

The generated dataset included a total of 5,000 observations representative of urban indicators, distributed across three broad dimensions: infrastructure, socioeconomic, and territorial, along with simulated geographic coordinates. Each variable was generated using a statistical distribution selected to reflect its typical behavior in real-world contexts.

Infrastructure Variables

These variables represented the level of development and coverage of essential services in the urban environment, and were modeled using the beta distribution, which allows generating bounded values between 0 and 100, ideal for proportions or percentages: drinking water coverage, sewage coverage, access to electricity,

REVISTA CIENTIFICA INNOVACIÓN INTEGRAL
ISSN: 3103-1420
DOI: https://doi.org/10.70577/eag5rs05

access to gas and access to public transport.This choice is based on the fact that these indicators tend to show high concentration levels in consolidated urban areas, which is consistent with patterns observed in empirical studies.

Socioeconomic Variables

Socioeconomic variables described aspects related to the population and their economic and social well-being. Different distributions were used based on their nature:

Population density: Modeled using a gamma distribution, which avoids negative values and allows for positive skewness, common in this type of data.

Income level: A log-normal distribution was used, which reproduces the long-tail characteristic associated with income inequality.

Access to healthcare services and access to education: Both variables were generated with beta distributions, as were the infrastructure indicators, due to their proportional nature (range between 0 and 100).

Territorial Variables

The territorial variables describe the land use and spatial dynamics of the fictional city. Since these proportions are naturally bounded between 0 and 1, a uniform distribution was used for each: proportion of residential areas, proportion of commercial areas, proportion of industrial areas, and proportion of green spaces.

Additionally, the urban sprawl index was included, whose variability was modeled with a normal distribution, assuming symmetrical fluctuations around an average value.

Geographic Coordinates

To spatially locate the observations, fictitious coordinates were generated within a geographically plausible range for Latin America: latitude: Uniformly distributed between -35° and 15°, and longitude: Uniformly distributed between -80° and -35°.

This range covers a large portion of the South American continent, allowing for the simulation of a geographically realistic city.

Statistical Rationale and Realistic Validation

The selection of statistical distributions for each variable seeks to replicate real-world characteristics observed in recent urban studies: Beta distributions are appropriate for coverage and access variables, which tend to have higher density at high values; the log-normal distribution effectively reproduces the skewed distribution of income; land-use proportions, being bounded and without clear trends toward extremes, fit well with a uniform distribution; and geographic coordinates were limited to specific ranges to ensure regional consistency.

This synthetic dataset was designed to emulate the typical conditions of a Latin American city, providing a solid foundation for analyzing spatial patterns and benchmarking clustering algorithms in urban contexts.

3. Results

3.1 Urban Clustering Results

Dimensionality Reduction through Principal Component Analysis (PCA)

The principal component analysis revealed that 14 components were required to explain at least 95% of the total variance of the original dataset (Figure 1). This result indicates a high correlation among the original variables—a total of 15—which is consistent with the interrelated nature of urban indicators, where aspects such as infrastructure,

REVISTA CIENTIFICA INNOVACIÓN INTEGRAL
ISSN: 3103-1420
DOI: https://doi.org/10.70577/eag5rs05

access to services, and land use tend to be closely linked.

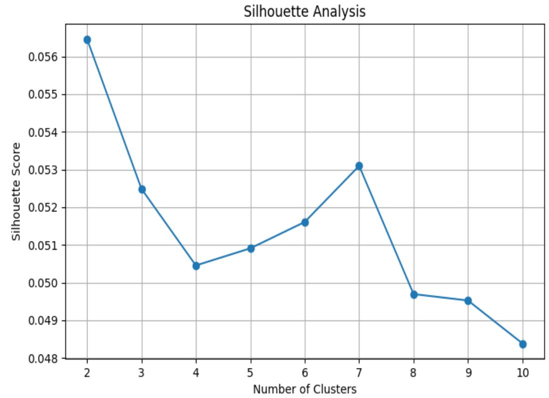

Fig 1. Dimensionality Reduction through Silhouette Analysis.

The need for a high number of components reflects a complex and multidimensional urban structure, typical of contemporary metropolitan systems. This suggests that any excessive simplification of the feature space could result in a significant loss of relevant information for spatial analysis.

Clustering with K-means

Using the Elbow Method and Silhouette Analysis, the optimal number of clusters was determined to be 2 (**Figure 2**). This dichotomy implies that, from a macro perspective, the city can be classified into two broad categories of urban areas, which could represent a division between highly developed central zones and peripheral areas with distinct characteristics.

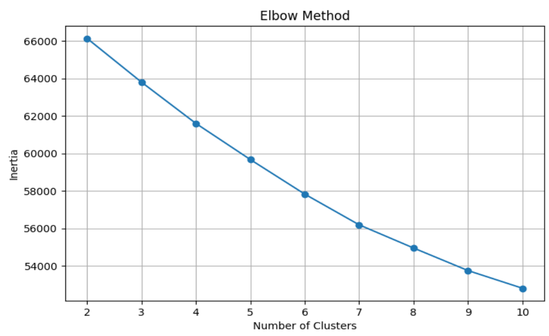

Fig 2. Dimensionality Reduction via Elbow Analysis.

Calinski-Harabasz Index for K-means
The Calinski-Harabasz index achieved a value of 291.30**,** indicating a strong separation between the identified clusters. This high score confirms the robustness of the clustering and a clear differentiation between the two groups, validating the effectiveness of the K-means algorithm in this context.

DBSCAN Clustering

The DBSCAN algorithm identified 19 dense clusters, suggesting a certain level of spatial heterogeneity within the city. However, one of the most notable findings was that 93.48% of the observations (4674 out of 5000) were labeled as noise, indicating a highly dispersed or weakly defined spatial distribution in terms of local density.

Regarding clustering quality, the Calinski-Harabasz index reached a value of 6.60**,** which is considerably lower than that of K-means. This result implies that the clusters detected by DBSCAN are less compact, less well-defined, or influenced by a high number of outlier or transitional points.

Identified Urban Structure

The segmentation into two main clusters provided by K-means (Figure 3) suggests a bipolar urban structure**,** possibly reflecting a distinction between central areas characterized by high population density, greater infrastructure, and better service coverage and peripheral zones with lower development levels and reduced access to basic

REVISTA CIENTIFICA INNOVACIÓN INTEGRAL
ISSN: 3103-1420
DOI: https://doi.org/10.70577/eag5rs05

services.

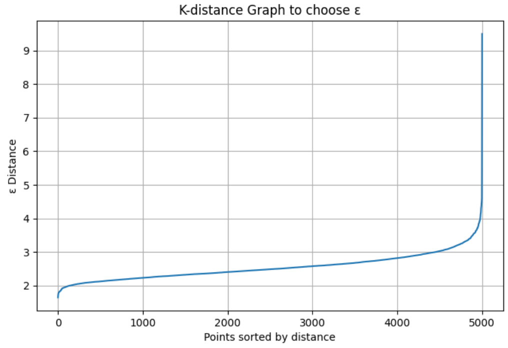

Fig 3. Segmentation into Two Main Clusters Provided by K-means.

On the other hand, the large number of points considered noise by DBSCAN reinforces the idea of a city with multiple transitional zones and a spatial organization that is not strictly homogeneous common in urban contexts undergoing expansion or informality.

Relevant Urban Insights

The results shown in Figure 4 highlight the presence of a clear division between well-defined urban areas likely corresponding to consolidated zones with broad access to infrastructure and services and others with lower levels of development. Similarly, the large amount of noise points suggests the existence of numerous transitional or recently developed zones, which deserve special attention in terms of urban planning.

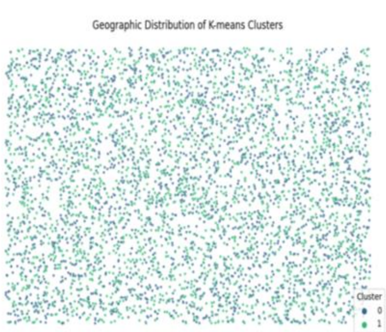

Fig 4. Geographic Distribution of K-means and Clusters**.**

These transitional areas could correspond to informal settlements, neighborhoods undergoing urbanization, or sectors experiencing dynamic changes in land use. This has direct implications for public policies aimed at reducing inequalities and improving equity in access to services.

4. Discussion

The principal component analysis revealed that 14 components were required to explain at least 95% of the total variance of the original dataset. This result indicates a high correlation among the original variables — totaling 15 — which is consistent with the interrelated nature of urban indicators, where aspects such as infrastructure, access to services, and land use tend to be closely linked [41].

The need for a large number of components reflects a complex and multidimensional urban structure, typical of contemporary metropolitan systems. This suggests that any excessive simplification of the feature space could lead to a significant loss of relevant information for spatial analysis [42]. In this regard, the use of dimensionality reduction techniques must carefully balance computational efficiency and the preservation of the rich information inherent in multivariate urban data [43].

Clustering with K-means

Using the Elbow Method and Silhouette Analysis, the optimal number of clusters was determined to be 2. This dichotomy implies that, from a macro perspective, the city can be classified into two broad categories of urban areas, which could represent a division between highly developed central zones and peripheries with distinct characteristics [44].

The obtained Calinski-Harabasz index was 291.30, indicating good separation between the identified clusters. This high value confirms the robustness

REVISTA CIENTIFICA INNOVACIÓN INTEGRAL
ISSN: 3103-1420
DOI: https://doi.org/10.70577/eag5rs05

of the clustering and the clear differentiation between both groups, validating the effectiveness of the K-means algorithm in this case. These findings align with previous studies highlighting the utility of K-means in urban contexts characterized by clear functional segmentation patterns [45].

Clustering with DBSCAN

The DBSCAN algorithm identified 19 dense clusters, suggesting some spatial heterogeneity within the city. However, one notable finding was that 93.48% (4674 out of 5000) of observations were labeled as noise, indicating a highly dispersed or poorly defined spatial distribution from the perspective of local density [46].

Regarding clustering quality, the Calinski-Harabasz index reached a value of 6.60, considerably lower than that of K-means. This result suggests that clusters detected by DBSCAN are less compact, less defined, or influenced by a large number of outliers or transitional points. This phenomenon could be due to the current model's inadequate capture of the dynamics of uneven and expansive urban development typical of many Latin American cities [47].

Identified Urban Structure

The segmentation into two main clusters provided by K-means points towards a bipolar urban structure, possibly associated with a differentiation between: Central areas: characterized by high population density, greater coverage of services and infrastructure and peripheral areas: with lower levels of development and access to basic services, where the large number of points considered noise by DBSCAN reinforces the idea of a city with multiple transition zones and a spatial organization that is not strictly homogeneous, which is common in urban contexts in the process of expansion or informality [48], [49].

These transitional areas could correspond to informal settlements, neighborhoods undergoing urbanization, or sectors with dynamic land use changes, which has direct implications for public policies aimed at reducing inequalities and improving equity in access to services [50].

Relevant Urban Insights

The results highlight the presence of a clear division between well-defined urban areas, probably corresponding to consolidated zones with broad access to infrastructure and services, versus others with lower development levels. Likewise, the large number of noise points suggests the existence of numerous transitional or recently growing zones that deserve special attention in urban planning [51].

These transitional areas could correspond to informal settlements, neighborhoods in the process of urbanization, or sectors with dynamic land use changes, which has direct implications for public policies aimed at reducing inequalities and improving equity in access to services [52]. The integration of these zones into the formal urban fabric is key to advancing towards sustainable and inclusive development models in Latin America.

5. Conclusions

The principal component analysis (PCA) revealed that 14 components were required to explain 95% of the total variance, indicating a high interrelation among the urban variables. This finding aligns with scientific literature that highlights the multidimensional nature of contemporary urban systems, where aspects such as infrastructure, access to services, and land use are intrinsically linked. Therefore, caution must be exercised when reducing dimensionality, as excessive simplifications could lead to the loss of critical information for decision-making.

Clustering with K-means identified two well-differentiated groups, suggesting a bipolar urban

ISSN: 3103-1420
DOI: https://doi.org/10.70577/eag5rs05

structure: highly developed central areas and peripheries with lower development levels. The high Calinski-Harabasz index value (291.30) confirms the robustness of the clustering and reinforces the usefulness of this method in contexts seeking clear and functional city segmentation.

On the other hand, the analysis with DBSCAN showed significant limitations: although it identified 19 dense clusters, 93.48% of the points were classified as noise, evidencing a highly dispersed spatial distribution. The low Calinski-Harabasz index score (6.60) suggests that the clusters found are poorly compact or influenced by complex spatial transitions. This could be due to the heterogeneity of urban development in Latin America, characterized by irregular expansion zones and informal settlements.

References:

[1] Son, T. H., Weedon, Z., Yigitcanlar, T., Sanchez, T., Corchado, J. M., & Mehmood, R. (2023). Algorithmic urban planning for smart and sustainable development: Systematic review of the literature. *Sustainable Cities and Society*, *94*, 104562. https://doi.org/10.1016/j.scs.2023.104562

[2] Sanchez, T. W., Shumway, H., Gordner, T., & Lim, T. (2023). The prospects of artificial intelligence in urban planning. *International journal of urban sciences*, *27*(2), 179-194. http://dx.doi.org/10.1080/12265934.2022.2102538

[3] Do Nascimento, A. C. L., Galvani, E., Gobo, J. P. A., & Wollmann, C. A. (2022). Comparison between air temperature and land surface temperature for the city of São Paulo, Brazil. *Atmosphere*, *13*(3), 491. https://doi.org/10.3390/atmos13030491

[4] Junta, U., Newiduom, L., Opuiyo, A., & Browndi, I. (2022). Predictive analysis of urban planning for through the operation of artificial cloud network. *International Journal of Science and Advanced Technology*, *62*(2022), 622-627. https://isi.ac/storage/article-files/6LuYzVzjMZ0mxAUYgoDFjoH1wObqmvqsCFSuK1qz.pdf

[5] Ogas-Mendez, A. F., Pei, X., & Isoda, Y. (2022). Squatting behavior during the COVID-19 pandemic: The case of the informal settlement "Los Hornos" in Buenos Aires. *Habitat International*, *130*, 102688. https://doi.org/10.1016/j.habitatint.2022.102688

[6] Gan, W., Zhao, Z., Wang, Y., Zou, Y., Zhou, S., & Wu, Z. (2024). UDGAN: A new urban design inspiration approach driven by using generative adversarial networks. *Journal of Computational Design and Engineering*, *11*(1), 305-324. http://dx.doi.org/10.1093/jcde/qwae014

[7] Cook, M., & Karvonen, A. (2024). Urban planning and the knowledge politics of the smart city. *Urban Studies*, *61*(2), 370-382. https://journals.sagepub.com/doi/10.1177/00420980231177688

[8] Cheng, W., Chu, Y., Xia, C., Zhang, B., Chen, J., Jia, M., & Wang, W. (2023). UrbanGenoGAN: pioneering urban spatial planning using the synergistic integration of GAN, GA, and GIS. *Frontiers in Environmental Science*, *11*, 1287858. http://dx.doi.org/10.3389/fenvs.2023.1287858

[9] Silva, C., & Vergara-Perucich, F. (2021). Determinants of urban sprawl in Latin America: evidence from Santiago de Chile. *SN social sciences*, *1*(8), 202. https://link.springer.com/article/10.1007/s43545-021-00197-4

[10] Muñoz-Erickson, T. A., Meerow, S., Hobbins, R., Cook, E., Iwaniec, D. M., Berbés-Blázquez,

ISSN: 3103-1420
DOI: https://doi.org/10.70577/eag5rs05

M., ... & Robles-Morua, A. (2021). Beyond bouncing back? Comparing and contesting urban resilience frames in US and Latin American contexts. *Landscape and Urban Planning*, *214*, 104173. https://doi.org/10.1016/j.landurbplan.2021.104173

[11] De Rosa, M., Flores, I., & Morgan, M. (2024). More unequal or not as rich? Revisiting the Latin American exception. *World Development*, *184*, 106737. https://doi.org/10.1016/j.worlddev.2024.106737

[12] Velasco Arevalo, A., & Gerike, R. (2023). Sustainability evaluation methods for public transport with a focus on Latin American cities: A literature review. *International Journal of Sustainable Transportation*, *17*(11), 1236-1253. https://doi.org/10.1080/15568318.2022.2163208

[13] Tiznado-Aitken, I., Vecchio, G., Guzman, L. A., Arellana, J., Humberto, M., Vasconcellos, E., & Muñoz, J. C. (2023). Unequal periurban mobility: Travel patterns, modal choices and urban core dependence in Latin America. *Habitat International*, *133*, 102752. https://doi.org/10.1016/j.habitatint.2023.102752

[14] Ferrari, G., Guzmán-Habinger, J., Chávez, J. L., Werneck, A. O., Silva, D. R., Kovalskys, I., ... & Fisberg, M. (2021). Sociodemographic inequities and active transportation in adults from Latin America: an eight-country observational study. *International journal for equity in health*, *20*(1), 190. https://equityhealthj.biomedcentral.com/articles/10.1186/s12939-021-01524-0

[15] Kundu, B., & Kumar, R. (2024). Enhancing crop resilience to climate change through biochar: a review. *International Journal of Environment and Climate Change*, *14*(6), 170-184 https://doi.org/10.9734/ijecc/2024/v14i64219

[16] Khan, M. I., & Al-Ghamdi, S. G. (2023). Enhancing Energy System Resilience: Navigating Climate Change and Security Challenges. *Sustainable Cities in a Changing Climate: Enhancing Urban Resilience*, 227-250. http://dx.doi.org/10.1002/9781394201532.ch14

[17] Asongu, S. A., Diop, S., & Addis, A. K. (2023). Governance, inequality and inclusive education in sub-saharan Africa. In *Forum for Social Economics* (Vol. 52, No. 1, pp. 43-68). Routledge. http://dx.doi.org/10.2139/ssrn.3734897

[18] Al-Sehrawy, R., Kumar, B., & Watson, R. (2023). The pluralism of digital twins for urban management: Bridging theory and practice. *Journal of Urban Management*, *12*(1), 16-32. https://doi.org/10.1016/j.jum.2023.01.002

[19] Ibrahim, I. M., Radie, A. H., Jacksi, K., Zeebaree, S. R., Shukur, H. M., Rashid, Z. N., ... & Yasin, H. M. (2021). Task scheduling algorithms in cloud computing: A review. *Turkish Journal of Computer and Mathematics Education*, *12*(4), 1041-1053. http://dx.doi.org/10.17762/turcomat.v12i4.612

[20] Ran, X., Xi, Y., Lu, Y., Wang, X., & Lu, Z. (2023). Comprehensive survey on hierarchical clustering algorithms and the recent developments. *Artificial Intelligence Review*, *56*(8), 8219-8264. http://dx.doi.org/10.1007/s10462-022-10366-3

[21] Zhou, Y., Zhang, X., & Ding, F. (2021). Hierarchical estimation approach for RBF-AR models with regression weights based on the increasing data length. *IEEE Transactions on Circuits and Systems II: Express Briefs*, *68*(12),

ISSN: 3103-1420
DOI: https://doi.org/10.70577/eag5rs05

3597-3601. http://dx.doi.org/10.1109/TCSII.2021.3076112

[22] Shalileh, S., & Mirkin, B. (2022). Community partitioning over feature-rich networks using an extended k-means method. *Entropy*, *24*(5), 626. https://doi.org/10.3390/e24050626

[23] Vera, C., Lucchini, F., Bro, N., Mendoza, M., Löbel, H., Gutiérrez, F., ... & Toro, S. (2022). Learning to cluster urban areas: two competitive approaches and an empirical validation. *EPJ Data Science*, *11*(1), 62. http://dx.doi.org/10.1140/epjds/s13688-022-00374-2

[24] Reyes, A., Mendoza, M., Vera, C., Lucchini, F., Dimter, J., Gutiérrez, F., ... & Reyes, A. (2024). SpatialCluster: A Python library for urban clustering. *SoftwareX*, *26*, 101739. https://doi.org/10.1016/j.softx.2024.101739

[25] Casali, Y., Aydin, N. Y., & Comes, T. (2022). Machine learning for spatial analyses in urban areas: a scoping review. *Sustainable cities and society*, *85*, 104050. https://doi.org/10.1016/j.scs.2022.104050

[26] Nagappan, S. D., & Daud, S. M. (2021). Machine learning predictors for sustainable urban planning. *International Journal of Advanced Computer Science and Applications*, *12*(7). https://thesai.org/Publications/ViewPaper?Volume=12&Issue=7&Code=IJACSA&SerialNo=87

[27] Faraji, A., Homayoon Arya, S., Ghasemi, E., Soleimani, H., & Rahnamayiezekavat, P. (2023). A constructability assessment model based on BIM in urban renewal projects in limited lands. *Buildings*, *13*(10), 2599. https://doi.org/10.3390/buildings13102599

[28] Chaturvedi, V., & de Vries, W. T. (2021). Machine learning algorithms for urban land use planning: A review. *Urban Science*, *5*(3), 68. https://doi.org/10.3390/urbansci5030068

[29] Li, Y., Zhao, Q., & Zhong, C. (2022). GIS and urban data science. *Annals of GIS*, *28*(2), 89-92. https://www.tandfonline.com/doi/pdf/10.1080/19475683.2022.2070969

[30] Sarker, I. H. (2022). Smart City Data Science: Towards data-driven smart cities with open research issues. *Internet of Things*, *19*, 100528. https://doi.org/10.1016/j.iot.2022.100528

[31] Nel, E., MacLachlan, A., Ballinger, O., Cole, H., & Cole, M. (2023). Data-Driven Decision Making in Response to the COVID-19 Pandemic: A City of Cape Town Case Study. *Sustainability*, *15*(3), 1853. http://dx.doi.org/10.3390/su15031853

[32] Schindler, S., & Kanai, J. M. (2021). Getting the territory right: Infrastructure-led development and the re-emergence of spatial planning strategies. *Regional Studies*, *55*(1), 40-51. http://dx.doi.org/10.1080/00343404.2019.1661984

[33] Liu, K., Xu, X., Huang, W., Zhang, R., Kong, L., & Wang, X. (2023). A multi-objective optimization framework for designing urban block forms considering daylight, energy consumption, and photovoltaic energy potential. *Building and Environment*, *242*, 110585. https://doi.org/10.1016/j.buildenv.2023.110585

[34] Martins, M. S., Kalil, R. M. L., & Dalla Rosa, F. (2021). Community participation in the identification of neighbourhood sustainability indicators in Brazil. *Habitat International*, *113*, 102370. https://doi.org/10.1016/j.habitatint.2021.102370

ISSN: 3103-1420
DOI: https://doi.org/10.70577/eag5rs05

[35] Vazquez, S. A., & Flores, C. C. (2022). The perception of public spaces in Mexico city, a governance approach. *Journal of Urban Management*, *11*(1), 72-81. https://doi.org/10.1016/j.jum.2021.10.002

[36] Zamora-Moncayo, E. C., Herrera, B., Larrieta, J., DuBois, A., & Miguel Esponda, G. (2024). A participatory evaluation of an urban garden project in Ecuador: Exploring factors that impact the recovery of people with severe mental health problems. *Qualitative Health Research*, *34*(14), 1472-1485. https://doi.org/10.1177/10497323241245867

[37] Peng, Z. R., Lu, K. F., Liu, Y., & Zhai, W. (2024). The pathway of urban planning AI: From planning support to plan-making. *Journal of Planning Education and Research*, *44*(4), 2263-2279. http://dx.doi.org/10.1177/0739456X231180568

[38] Maebe, K., Hart, A. F., Marshall, L., Vandamme, P., Vereecken, N. J., Michez, D., & Smagghe, G. (2021). Bumblebee resilience to climate change, through plastic and adaptive responses. *Global change biology*, *27*(18), 4223-4237. http://dx.doi.org/10.1111/gcb.15751

[39] Navarrete-Hernandez, P., Vetro, A., & Concha, P. (2021). Building safer public spaces: Exploring gender difference in the perception of safety in public space through urban design interventions. *Landscape and Urban Planning*, *214*, 104180. https://doi.org/10.1016/j.landurbplan.2021.104180

[40] Kontokosta, C. E. (2021). Urban informatics in the science and practice of planning. *Journal of Planning Education and Research*, *41*(4), 382-395. http://dx.doi.org/10.1177/0739456X18793716

[41] Michalina, D., Mederly, P., Diefenbacher, H., & Held, B. (2021). Sustainable urban development: A review of urban sustainability indicator frameworks. *Sustainability*, *13*(16), 9348. https://doi.org/10.3390/su13169348

[42] Zhu, W., & Chen, J. (2022). The spatial analysis of digital economy and urban development: A case study in Hangzhou, China. *Cities*, *123*, 103563. https://doi.org/10.1016/j.cities.2022.103563

[43] Peng, F. L., Dong, Y. H., Wang, W. X., & Ma, C. X. (2023). The next frontier: data-driven urban underground space planning orienting multiple development concepts. *Smart Construction and Sustainable Cities*, *1*(1), 3. http://dx.doi.org/10.1007/s44268-023-00003-5

[44] Oti, E. U., Olusola, M. O., Eze, F. C., & Enogwe, S. U. (2021). Comprehensive review of K-Means clustering algorithms. *criterion*, *12*, 22-23. http://dx.doi.org/10.31695/IJASRE.2021.34050

[45] Ran, X., Zhou, X., Lei, M., Tepsan, W., & Deng, W. (2021). A novel k-means clustering algorithm with a noise algorithm for capturing urban hotspots. *Applied Sciences*, *11*(23), 11202. https://doi.org/10.3390/app112311202

[46] Mohsin, S. A., & Alfoudi, A. S. (2024). Internet Traffic Classification Model Based on A-DBSCAN Algorithm. *International Journal of Intelligent Engineering & Systems*, *17*(5). https://oaji.net/articles/2023/3603-1723963830.pdf

[47] Tu, X., Fu, C., Huang, A., Chen, H., & Ding, X. (2022). DBSCAN spatial clustering analysis of urban "Production–Living–Ecological" space based on POI data: a case study of central urban Wuhan, China. *International Journal of Environmental Research and Public*

ISSN: 3103-1420
DOI: https://doi.org/10.70577/eag5rs05

Health, *19*(9), 5153. https://doi.org/10.3390/ijerph19095153

[48] Jiang, Y., Liu, Q., Zhao, S., Zhang, T., Fan, X., Zhong, R. Y., & Huang, G. Q. (2024). Heterogeneous intensity-based DBSCAN (iDBSCAN) model for urban attention distribution in digital twin cities. *Digital Engineering*, *2*, 100014. https://doi.org/10.1016/j.dte.2024.100014

[49] Cetin, Z., & Yastikli, N. (2025). Automatic Detection of Urban Trees from LiDAR Data Using DBSCAN and Mean Shift Clustering Methods in Fatih, Istanbul. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *48*, 95-102. http://dx.doi.org/10.5194/isprs-archives-XLVIII-M-6-2025-95-2025

[50] Caudillo-Cos, C. A., Montejano-Escamilla, J. A., Tapia-McClung, R., Ávila-Jiménez, F. G., & Barrera-Alarcón, I. G. (2024). Defining urban boundaries through DBSCAN and Shannon's entropy: The case of the Mexican National Urban System. *Cities*, *149*, 104969. https://doi.org/10.1016/j.cities.2024.104969

[51] Masum, M. H., Pal, S. K., Akhie, A. A., Ruva, I. J., Akter, N., & Nath, S. (2021). Spatiotemporal monitoring and assessment of noise pollution in an urban setting. *Environmental Challenges*, *5*, 100218. https://doi.org/10.1016/j.envc.2021.100218

[52] Guida, C., Carpentieri, G., & Masoumi, H. (2022). Measuring spatial accessibility to urban services for older adults: an application to healthcare facilities in Milan. *European transport research review*, *14*(1), 23. https://etrr.springeropen.com/articles/10.1186/s12544-022-00544-3#citeas:~:text=DOI-,https%3A//doi.org/10.1186/s12544%2D022%2D00544%2D3,-Share%20this%20article

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)
All authors contributed equally to the development of the article.

Sources of Funding for the Research Presented in the Scientific Article or for the Scientific Article Itself
No funding was received for conducting this study.

Conflict of Interest
The authors declare that they have no conflicts of interest relevant to the content of this article.

REVISTA CIENTIFICA INNOVACIÓN INTEGRAL
ISSN: 3103-1420
DOI: https://doi.org/10.70577/eag5rs05