

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 concluye 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.

Palabras clave: Cluster, Datos, Geoespacial, Planificación, Urbanismos.

1. Introducción

La optimización de la planificación urbana en América Latina a través de algoritmos de agrupamiento representa una intersección crítica de la ciencia de datos, la sostenibilidad y la equidad en respuesta a los crecientes desafíos urbanos de la región. Con más del 80% de la población residiendo actualmente en áreas urbanas, proyectado a aumentar al 90% para 2050, ciudades como São Paulo y Ciudad de México están lidiando con importantes disparidades socioeconómicas, infraestructura inadecuada y vulnerabilidades ambientales que amenazan la habitabilidad y la inclusividad urbana [1].

Los algoritmos de agrupamiento, que analizan datos espaciales y demográficos, se utilizan cada vez más para informar las estrategias de planificación urbana, permitiendo una comprensión más matizada de la dinámica urbana y la distribución equitativa de los recursos [2].

Estudios de caso notables en ciudades como Bogotá, São Paulo y Buenos Aires ilustran las aplicaciones prácticas de las técnicas de agrupamiento, revelando información sobre la segregación social, la asignación de recursos y las condiciones ambientales [3], [4], [5].

Estas metodologías basadas en datos han sido fundamentales para identificar barrios desatendidos, optimizar las rutas del transporte público y mejorar la resiliencia climática, fomentando en última instancia entornos urbanos más sostenibles y equitativos. Sin embargo, la implementación de estos enfoques no está exenta de desafíos, incluyendo

consideraciones éticas en torno a la recolección de datos, problemas de gobernanza y disparidades tecnológicas que pueden obstaculizar la gestión urbana efectiva [6].

También surgen controversias respecto a la efectividad y aplicabilidad de los diferentes algoritmos de agrupamiento en contextos urbanos dinámicos, ya que los estudios actuales a menudo se centran en el rendimiento relativo sin establecer estándares integrales para la cartografía urbana [7].

Además, la dependencia de la ciencia de datos en la planificación urbana plantea preguntas sobre la inclusividad, particularmente en lo que respecta a qué tan bien estos enfoques capturan las necesidades y voces de las comunidades marginadas [8].

A medida que las ciudades de América Latina continúan evolucionando, la integración de algoritmos de agrupamiento y procesos de planificación participativa será esencial para lograr futuros urbanos equitativos y sostenibles [9].

Desafíos Urbanos en América Latina

América Latina enfrenta una multitud de desafíos urbanos exacerbados por la rápida urbanización y las disparidades socioeconómicas. Hasta ahora, más del 80% de la población de la región reside en áreas urbanas, con proyecciones que indican que esta cifra podría aumentar al 90% para 2050 [10], por lo que, a pesar del crecimiento urbano significativo, los beneficios de la urbanización no se han distribuido equitativamente, lo que ha resultado en desigualdades persistentes.

Desigualdad Socioeconómica

El paisaje urbano en América Latina se caracteriza por una marcada segregación social y espacial. Aproximadamente el 26% de la población urbana vive en pobreza o pobreza extrema, por lo que las políticas de vivienda han tenido dificultades para abordar las necesidades de las poblaciones más pobres, dejando a muchos sin acceso a viviendas adecuadas y servicios esenciales [11].

Transporte Urbano e Infraestructura

El transporte urbano es un contribuyente significativo al consumo de energía y las emisiones de gases de efecto invernadero en las ciudades latinoamericanas, consecuencia de una planificación urbana que ha favorecido el uso del automóvil privado sobre los sistemas de transporte público sostenibles [12].

La rápida expansión de las áreas urbanas, a menudo a tasas de dos a tres veces más altas que el crecimiento de la población, ha llevado a un aumento en los costos de infraestructura y problemas de congestión [13].

Las ciudades continúan enfrentando desafíos para desarrollar sistemas de transporte eficientes y sostenibles, que son críticos para mejorar la habitabilidad urbana y reducir el impacto ambiental [14].

Resiliencia y Seguridad Climática

A medida que las ciudades crecen, también deben adaptarse a las crecientes amenazas que plantea el cambio climático. Las áreas urbanas son particularmente vulnerables a los eventos relacionados con el clima, lo que requiere evaluaciones exhaustivas de vulnerabilidad climática y estrategias de mitigación [15].

Aumentar la seguridad urbana es crucial para proteger a los residentes del crimen y los accidentes, que se amplifican en áreas densamente pobladas que carecen de recursos y planificación adecuados [16].

Gobernanza e Inclusión Política

La gobernanza urbana efectiva es esencial para abordar estos desafíos multifacéticos. Se requieren procesos políticos inclusivos para satisfacer las diversas necesidades de las poblaciones urbanas y asegurar que el desarrollo beneficie equitativamente a todos los residentes [17].

Las prácticas de gestión urbana deben reestructurarse para promover la sostenibilidad, la equidad y la resiliencia, alineándose con iniciativas globales como la Conferencia de las Naciones Unidas sobre el Desarrollo Sostenible (Río+20) como lo plantea Al-Sehrawy [18].

Algoritmos de Agrupamiento

Los algoritmos de agrupamiento desempeñan un papel fundamental en la planificación urbana al facilitar la identificación de unidades geográficas distintas dentro de una ciudad que exhiben homogeneidad interna mientras permanecen distintas de las áreas circundantes. Estos algoritmos ayudan a analizar la estructura urbana basándose en diversos factores, incluyendo la etnicidad, el estatus socioeconómico y el entorno construido, facilitando así la medición de la segregación residencial y la evaluación de las contribuciones de los espacios públicos a las aglomeraciones territoriales basadas en la raza y la etnicidad [19].

Tipos de algoritmos de agrupamiento

El estudio del agrupamiento urbano abarca varios tipos de algoritmos, que generalmente se pueden categorizar en métodos de agrupamiento

jerárquico y no jerárquico. El agrupamiento jerárquico, que incluye tanto el enfoque de agregación ascendente como el de división descendente, es particularmente prevalente en los estudios urbanos. En el método basado en la agregación, cada punto de datos comienza como su propio clúster y se fusiona progresivamente con los clústeres más cercanos según las similitudes por pares, mientras que el método basado en la división comienza con un clúster grande y lo divide hasta que cada punto de datos está aislado [20].

Agrupamiento Jerárquico

Los métodos de agrupamiento jerárquico tienen fortalezas y debilidades notables. Proporcionan una visión integral de las relaciones de los datos a través de dendrogramas; sin embargo, pueden ser sensibles a los valores atípicos, costosos computacionalmente y difíciles de interpretar debido a su complejidad. Estos métodos también requieren que el número de clústeres sea predeterminado, lo que puede limitar su aplicabilidad en contextos urbanos dinámicos [21].

Modelos de Mezcla Gaussiana y Redes de Modularidad Profunda

En los últimos años, han surgido técnicas más avanzadas como los Modelos de Mezcla Gaussiana (GMMs) y las Redes de Modularidad Profunda (DMONs) en el campo de la agrupación urbana. Los GMM son efectivos para datos distribuidos espacialmente, mientras que los DMON operan en grafos atribuidos, permitiendo la integración de atributos de nodos e información estructural del entorno urbano. Estos métodos enriquecen las representaciones de los nodos, proporcionando valiosos conocimientos para la planificación y el análisis urbano [22].

Validación Empírica de Técnicas de Agrupamiento

Para evaluar la efectividad de diferentes métodos de agrupamiento, se han realizado estudios empíricos, como uno que involucra a residentes de Santiago, Chile. Este estudio comparó las soluciones de agrupamiento generadas a través de GMMs y DMONs, centrándose en varios factores como la resolución espacial y la granularidad de la representación de datos [23].

Los resultados de estas comparaciones informan la selección de algoritmos apropiados basados en características urbanas específicas y objetivos de planificación, enfatizando la importancia de la elección del método para lograr un desarrollo urbano sostenible y equitativo [24].

Enfoques de Ciencia de Datos

La ciencia de datos ha surgido como una herramienta transformadora para la planificación urbana en América Latina, aprovechando el aprendizaje automático (ML) y los grandes datos para mejorar los procesos de toma de decisiones y promover el desarrollo sostenible. Estos enfoques facilitan el análisis de conjuntos de datos complejos, permitiendo a los planificadores identificar patrones y tendencias que informan políticas de vivienda equitativas y la asignación de recursos [25].

Aprendizaje Automático en la Planificación Urbana

Los algoritmos de aprendizaje automático pueden procesar grandes cantidades de datos, incluyendo patrones de uso del suelo, información demográfica y tendencias del mercado de la vivienda. Al hacerlo, pueden ayudar a los responsables de políticas y a los desarrolladores a identificar áreas adecuadas para iniciativas de vivienda asequible, fomentando la inclusión social y mejorando las condiciones de vida en general [26].

La integración de tecnologías de IA, como el Modelado de Información de Construcción (BIM) y la automatización robótica, agiliza aún más los procesos de construcción, reduciendo costos y mejorando los plazos de los proyectos [27].

Además, el aprendizaje automático puede ayudar en la detección temprana de defectos de construcción, mejorando el control de calidad y asegurando la resiliencia a largo plazo en los proyectos de desarrollo urbano [28].

Desafíos en la Utilización de Datos

A pesar de los beneficios potenciales, varios desafíos obstaculizan la utilización efectiva de la ciencia de datos en la planificación urbana. Los sistemas de recolección de datos limitados, los conjuntos de datos fragmentados y las preocupaciones de privacidad presentan barreras significativas [29].

Para superar estos problemas, es fundamental mejorar las prácticas de gestión de datos, fomentar el intercambio de datos y la colaboración entre las partes interesadas. Iniciativas como centros de datos especiales para datos públicos sensibles pueden proporcionar entornos seguros para los investigadores, promoviendo así mejores intervenciones políticas e investigación urbana [30].

Estrategias de Big Data

Ciudades como Ciudad del Cabo, Sudáfrica, ejemplifican la implementación exitosa de estrategias de big data para mejorar la gobernanza local. La ciudad ha invertido en infraestructura informática y personal calificado para desarrollar una estrategia de datos integral, introduciendo innovaciones como los data sandboxes. Estos permiten a los científicos de datos acceder a conjuntos de datos ricos para resolver desafíos políticos urgentes, desde

problemas del mercado laboral hasta la adaptación al cambio climático [31].

Tales marcos subrayan el reconocimiento creciente del poder de los datos para optimizar la prestación de servicios y mejorar los procesos de planificación en entornos urbanos de América Latina y el Caribe [32].

Direcciones Futuras

La optimización de la planificación urbana en América Latina está lista para avances significativos mediante la integración de estrategias y tecnologías innovadoras. Para lograr ciudades más sostenibles y equitativas, hay varias áreas clave que merecen atención [33].

Adoptando la Planificación Participativa

Un aspecto crítico de la sostenibilidad urbana radica en la planificación participativa, que asegura que el desarrollo urbano refleje las necesidades y aspiraciones de las comunidades locales. Este enfoque se alinea con el derecho a la ciudad, como se enfatiza en varios contextos latinoamericanos, particularmente en Brasil, Ecuador y México [34], [35], [36]. Al involucrar a los residentes en el proceso de planificación, las ciudades pueden fomentar la integración social y contrarrestar las tendencias prevalentes de segregación y exclusión.

Aprovechando las Tecnologías Inteligentes

La adopción de tecnologías inteligentes en la planificación urbana ofrece el potencial para mejorar la calidad de vida en las ciudades. Los sistemas de transporte automatizados, los sensores de consumo de recursos y las aplicaciones para optimizar los servicios urbanos pueden contribuir a entornos urbanos más eficientes y receptivos [37]. Estas tecnologías no solo mejoran la prestación de servicios, sino que también facilitan una mejor

gestión de los recursos y la sostenibilidad ambiental.

Abordar la Resiliencia al Cambio Climático

Con el cambio climático presentando un desafío urgente, la planificación urbana debe incorporar estrategias de resiliencia. Esto incluye diseñar infraestructuras que puedan resistir eventos climáticos extremos, reducir las emisiones de carbono y promover la movilidad sostenible [38]. Las ciudades deben priorizar medidas adaptativas que mejoren su capacidad para hacer frente a los impactos climáticos mientras aseguran el bienestar de sus habitantes.

Mejorando los Espacios Públicos y la Infraestructura Verde

La inversión en espacios públicos e infraestructura verde es esencial para promover la socialización, la identidad y la calidad de vida en general [39]. Al crear parques y áreas verdes, las ciudades pueden mejorar la salud y el bienestar urbano, como lo destaca la Organización Mundial de la Salud. El desarrollo de tales espacios debe llevarse a cabo mediante procesos participativos para fomentar la apropiación comunitaria y mejorar la identidad local.

Utilizando Enfoques Basados en Datos

La implementación de algoritmos de agrupamiento y técnicas de ciencia de datos puede mejorar significativamente los resultados de la planificación urbana. Al analizar los datos urbanos de manera efectiva, los planificadores pueden identificar patrones y optimizar el uso del suelo, asegurando que el desarrollo sea compacto e integrado socialmente [40]. Estos enfoques basados en datos pueden facilitar una mejor toma de decisiones y mejorar la capacidad de respuesta de las políticas urbanas a las necesidades de diversas poblaciones, por lo que el objetivo de esta investigación fue evaluar el

potencial de la inteligencia artificial para la planificación urbana en Latinoamérica.

2. Materiales y Métodos

2.1 Modelos Estadísticos

Análisis Metodológico del Proceso de Clustering

El análisis se estructuró en varias etapas que permitieron asegurar la calidad de los datos, reducir su dimensionalidad y aplicar técnicas de clustering espacial. A continuación, se detalla cada una de las fases implementadas:

Preprocesamiento de Datos

Como paso inicial, se realizó la normalización de las variables utilizando el StandardScaler. Este proceso garantiza que todas las características tengan una escala comparable, evitando sesgos durante el cálculo de distancias y contribuyendo equitativamente al análisis posterior.

Reducción de Dimensionalidad

Para simplificar el espacio de características sin perder información relevante, se aplicó el Análisis de Componentes Principales (PCA). Se seleccionó el número óptimo de componentes necesarios para explicar al menos el 95% de la varianza total del conjunto de datos original. Como parte de este proceso, se generó un gráfico que muestra la varianza explicada acumulada por componente, lo cual permite visualizar la contribución relativa de cada uno de ellos al modelo.

Aplicación de Técnicas de Clustering

Se emplearon dos métodos complementarios de agrupamiento: K-means y DBSCAN, con el objetivo de explorar diferentes patrones espaciales en los datos urbanos.

K-means : Para determinar el número óptimo de clusters, se utilizó el método del codo (Elbow Method) y el análisis de silueta (Silhouette Analysis). Estas herramientas ayudan a identificar el valor de k que maximiza la cohesión interna y la separación entre grupos. Además, se evaluó la calidad de los clusters mediante el coeficiente de Silhouette, y se generaron gráficos para visualizar ambos criterios.

DBSCAN : Este algoritmo fue utilizado para detectar clusters irregulares y puntos atípicos o ruido espacial. El parámetro ϵ (epsilon), que define la vecindad máxima entre puntos, se estimó a partir del gráfico de k-distancia. Se adoptó como umbral el percentil 2% de las distancias ordenadas, lo cual permitió identificar regiones de alta densidad mientras se controlaba la sensibilidad al ruido.

Evaluación de los Resultados de Clustering

La calidad de los clusters obtenidos se evaluaron mediante el índice de Calinski-Harabasz , una métrica que mide la relación entre la dispersión entre clusters y la dispersión dentro de los clusters. Esta medida permitió comparar objetivamente el desempeño de ambos algoritmos, K-means y DBSCAN, en términos de la claridad y definición de los grupos identificados.

Visualización Espacial de los Clusters

Finalmente, se realizaron representaciones geográficas de los resultados de clustering, empleando diagramas de dispersión (scatterplots) basados en las coordenadas latitud y longitud simuladas. Para diferenciar los grupos resultantes, se utilizaron paletas de colores distintas para cada método: una escala continua para K-means y una combinación contrastante para DBSCAN, incluyendo también los puntos clasificados como ruido.

Este enfoque metodológico combina técnicas avanzadas de preprocesamiento, reducción de dimensionalidad y clustering, con el fin de identificar patrones espaciales significativos en contextos urbanos. La integración de múltiples estrategias analíticas permite no solo segmentar áreas urbanas de forma eficiente, sino también evaluar la robustez de los resultados a través de métricas estadísticas y visualizaciones geoespaciales.

2.2 Datos utilizados

Descripción de las Variables del Conjunto de Datos Sintético

El conjunto de datos generado comprendió un total de 5,000 observaciones representativas de indicadores urbanos, distribuidas en tres grandes dimensiones: infraestructura, socioeconómica y territorial, junto con coordenadas geográficas simuladas. Cada variable fue generada utilizando una distribución estadística seleccionada para reflejar su comportamiento típico en contextos reales.

Variables de Infraestructura

Estas variables representaron el nivel de desarrollo y cobertura de servicios esenciales en el entorno urbano, y se modelaron utilizando la distribución beta , que permite generar valores acotados entre 0 y 100, ideal para proporciones o porcentajes: cobertura de agua potable, cobertura de alcantarillado, acceso a electricidad, acceso a gas y acceso a transporte público.

Esta elección se fundamenta en que estos indicadores suelen mostrar una concentración hacia niveles altos de cobertura en áreas urbanas consolidadas, lo cual es consistente con patrones observados en estudios empíricos.

Variables Socioeconómicas

Las variables socioeconómicas describieron aspectos relacionados con la población y su bienestar económico y social. Para su generación, se utilizaron distintas distribuciones según su naturaleza:

Densidad poblacional: Se modeló mediante una distribución gamma , que evita valores negativos y permite asimetría positiva, común en este tipo de datos.

Nivel de ingresos: Se utilizó una distribución log-normal , que reproduce la característica cola larga asociada a la desigualdad en los ingresos.

Acceso a servicios de salud y acceso a educación :Ambas variables fueron generadas con distribuciones beta , al igual que los indicadores de infraestructura, debido a su naturaleza proporcional (rango entre 0 y 100).

Variables Territoriales

Las variables territoriales describen el uso del suelo y la dinámica espacial de la ciudad ficticia. Dado que estas proporciones están naturalmente acotadas entre 0 y 1, se empleó una distribución uniforme para cada una de ellas: proporción de áreas residenciales, proporción de áreas comerciales, proporción de áreas industriales y proporción de espacios verdes

Adicionalmente, se incluyó el índice de expansión urbana , cuya variabilidad fue modelada con una distribución normal , suponiendo fluctuaciones simétricas alrededor de un valor promedio.

Coordenadas Geográficas

Para ubicar espacialmente las observaciones, se generaron coordenadas ficticias dentro de un rango geográficamente plausible para América Latina: latitud : Distribución uniforme entre -35° y 15° y longitud : Distribución uniforme entre -80° y -35°.

Este rango cubre gran parte del continente sudamericano, permitiendo simular una ciudad realista en términos geográficos.

Fundamento Estadístico y Validación Realista

La selección de distribuciones estadísticas para cada variable busca replicar características reales observadas en estudios urbanos recientes: Las distribuciones beta son adecuadas para variables de cobertura y acceso, que tienden a tener mayor densidad en valores altos., la distribución log-normal reproduce de forma efectiva la distribución sesgada de los ingresos, las proporciones de uso del suelo , al estar acotadas y sin tendencias claras hacia extremos, se ajustan bien a una distribución uniforme y las coordenadas geográficas se limitaron a rangos específicos para garantizar coherencia regional.

Este conjunto de datos sintéticos fue diseñado para emular las condiciones típicas de una ciudad latinoamericana, ofreciendo una base sólida para el análisis de patrones espaciales y la evaluación comparativa de algoritmos de clustering en contextos urbanos.

3. Resultados

3.1 Resultados del Clustering Urbano

Reducción de Dimensionalidad mediante Análisis de Componentes Principales (PCA)

El análisis de componentes principales reveló que se requirieron 14 componentes para explicar al menos el 95% de la varianza total del conjunto de datos original (Figura 1). Este resultado indica una alta correlación entre las variables originales —en total 15— lo cual es coherente con la naturaleza interrelacionada de los indicadores urbanos, donde aspectos como infraestructura, acceso a servicios y uso del suelo tienden a estar estrechamente vinculados.

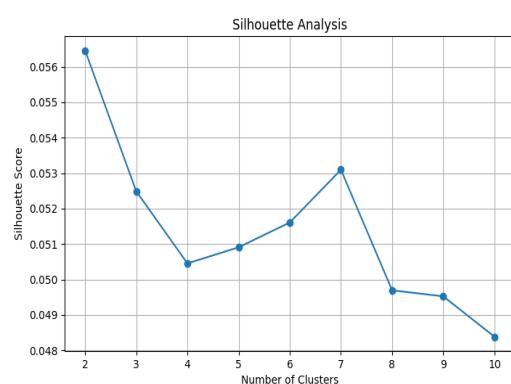


Fig 1. Reducción de Dimensionalidad mediante Análisis de Silhouette.

La necesidad de un número elevado de componentes refleja una estructura urbana compleja y multidimensional, típica de sistemas metropolitanos contemporáneos. Esto sugiere que cualquier simplificación excesiva del espacio de características podría llevar a una pérdida significativa de información relevante para el análisis espacial.

Agrupamiento con K-means

Mediante el método del codo (Elbow Method) y el análisis de silueta (Silhouette Analysis), se determinó que el número óptimo de clusters el cual fue 2 (Figura 2). Esta dicotomía implica que, desde una perspectiva macro, la ciudad puede clasificarse en dos grandes categorías de áreas urbanas, lo cual podría representar una división entre zonas centrales altamente desarrolladas y periferias con características distintas.

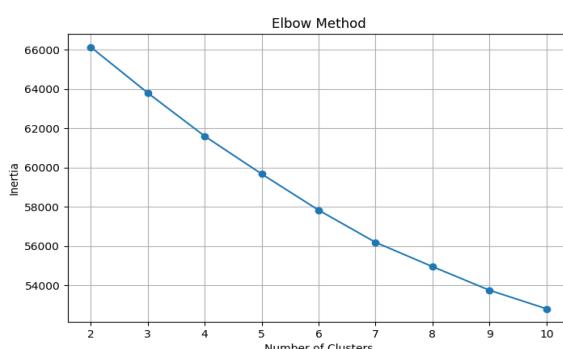


Fig 2. Reducción de Dimensionalidad mediante Análisis de Elbow.

El índice de Calinski-Harabasz obtenido fue de 291.30, lo cual denota una buena separación entre los clusters identificados. Este valor alto confirma la solidez del agrupamiento y la diferenciación clara entre ambos grupos, validando la efectividad del algoritmo K-means en este caso.

Agrupamiento con DBSCAN

El algoritmo DBSCAN logró identificar 19 clusters densos, lo cual sugiere cierta heterogeneidad espacial dentro de la ciudad. Sin embargo, uno de los hallazgos más notables fue que el 93.48% (4674 de 5000) de las observaciones fueron etiquetadas como ruido, lo que indica una distribución espacial altamente dispersa o poco definida desde el punto de vista de la densidad local.

En cuanto a la calidad del agrupamiento, el índice de Calinski-Harabasz alcanzó un valor de 6.60, considerablemente inferior al de K-means. Este resultado sugiere que los clusters detectados por DBSCAN son menos compactos, menos definidos o están influenciados por una gran cantidad de puntos atípicos o de transición.

Estructura Urbana Identificada

La segmentación en dos clusters principales proporcionada por K-means (Figura 3) apunta hacia una estructura urbana bipolar, posiblemente asociada a una diferenciación entre áreas centrales caracterizadas por alta densidad poblacional, mayor cobertura de servicios e infraestructura y áreas periféricas con menores niveles de desarrollo y acceso a servicios básicos.

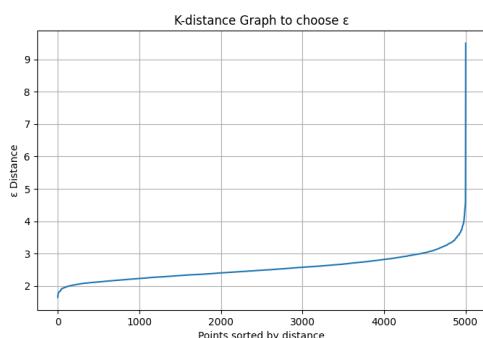


Fig 3. segmentación en dos clusters principales proporcionada por K-means.

Por otro lado, la gran cantidad de puntos considerados ruido por DBSCAN refuerza la idea de una ciudad con múltiples zonas de transición y una organización espacial no estrictamente homogénea, lo cual es común en contextos urbanos en proceso de expansión o informalidad.

Insights Urbanos Relevantes

Los resultados mostrados en la Figura 4, destacan la presencia de una división clara entre áreas urbanas bien definidas, probablemente correspondientes a zonas consolidadas con acceso amplio a infraestructura y servicios, versus otras con menor nivel de desarrollo. Asimismo, la gran cantidad de puntos de ruido sugiere la existencia de numerosas zonas de transición o de crecimiento reciente, que merecen atención especial en términos de planificación urbana.

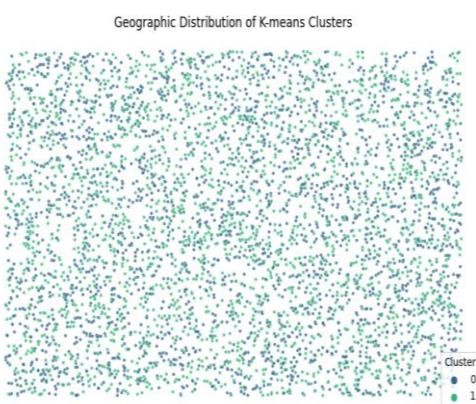


Fig 4. Distribución geográfica de K-means y cluster.

Estas áreas de transición podrían corresponder a asentamientos informales, barrios en proceso de urbanización o sectores con cambios dinámicos en el uso del suelo, lo cual tiene implicaciones directas para políticas públicas orientadas a reducir desigualdades y mejorar la equidad en el acceso a servicios.

4. Discusión

El análisis de componentes principales reveló que se requirieron 14 componentes para explicar al menos el 95% de la varianza total del conjunto de datos original. Este resultado indica una alta correlación entre las variables originales —en total 15— lo cual es coherente con la naturaleza interrelacionada de los indicadores urbanos, donde aspectos como infraestructura, acceso a servicios y uso del suelo tienden a estar estrechamente vinculados [41].

La necesidad de un número elevado de componentes refleja una estructura urbana compleja y multidimensional, típica de sistemas metropolitanos contemporáneos. Esto sugiere que cualquier simplificación excesiva del espacio de características podría llevar a una pérdida significativa de información relevante para el análisis espacial [42]. En este sentido, el uso de técnicas de reducción de dimensionalidad debe equilibrarse cuidadosamente entre la eficiencia computacional y la preservación de la riqueza informativa inherente a los datos urbanos multivariados [43].

Agrupamiento con K-means

Mediante el método del codo (Elbow Method) y el análisis de silueta (Silhouette Analysis), se determinó que el número óptimo de clusters es 2. Esta dicotomía implica que, desde una perspectiva macro, la ciudad puede clasificarse en dos grandes categorías de áreas urbanas, lo

cual podría representar una división entre zonas centrales altamente desarrolladas y periferias con características distintas [44].

El índice de Calinski-Harabasz obtenido fue de 291.30 , lo cual denota una buena separación entre los clusters identificados. Este valor alto confirma la solidez del agrupamiento y la diferenciación clara entre ambos grupos, validando la efectividad del algoritmo K-means en este caso. Estos hallazgos coinciden con estudios previos que destacan la utilidad de K-means en contextos urbanos caracterizados por patrones claros de segmentación funcional [45].

Agrupamiento con DBSCAN

El algoritmo DBSCAN logró identificar 19 clusters densos , lo cual sugiere cierta heterogeneidad espacial dentro de la ciudad. Sin embargo, uno de los hallazgos más notables fue que el 93.48% (4674 de 5000) de las observaciones fueron etiquetadas como ruido, lo que indica una distribución espacial altamente dispersa o poco definida desde el punto de vista de la densidad local [46].

En cuanto a la calidad del agrupamiento, el índice de Calinski-Harabasz alcanzó un valor de 6.60 , considerablemente inferior al de K-means. Este resultado sugiere que los clusters detectados por DBSCAN son menos compactos, menos definidos o están influenciados por una gran cantidad de puntos atípicos o de transición. Este fenómeno podría deberse a que el modelo actual no capture adecuadamente la dinámica de desarrollo urbano desigual y expansivo típico de muchas ciudades latinoamericanas [47].

Estructura Urbana Identificada

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].

Estas áreas de transición podrían corresponder a asentamientos informales, barrios en proceso de urbanización o sectores con cambios dinámicos en el uso del suelo, lo cual tiene implicaciones directas para políticas públicas orientadas a reducir desigualdades y mejorar la equidad en el acceso a servicios [50].

Insights Urbanos Relevantes

Los resultados destacan la presencia de una división clara entre áreas urbanas bien definidas , probablemente correspondientes a zonas consolidadas con acceso amplio a infraestructura y servicios, versus otras con menor nivel de desarrollo. Asimismo, la gran cantidad de puntos de ruido sugiere la existencia de numerosas zonas de transición o de crecimiento reciente, que merecen atención especial en términos de planificación urbana [51].

Estas áreas de transición podrían corresponder a asentamientos informales, barrios en proceso de urbanización o sectores con cambios dinámicos en el uso del suelo, lo cual tiene implicaciones directas para políticas públicas orientadas a reducir desigualdades y mejorar la equidad en el acceso a servicios [52]. La integración de estas zonas en el tejido urbano formal es clave para avanzar hacia modelos de desarrollo sostenible e inclusivo en América Latina.

5. Conclusiones

El análisis de componentes principales (PCA) reveló que se requirieron 14 componentes para explicar el 95% de la varianza total, lo cual indica una alta interrelación entre las variables urbanas. Este hallazgo es coherente con la literatura científica que señala la naturaleza multidimensional de los sistemas urbanos contemporáneos, donde aspectos como infraestructura, acceso a servicios y uso del suelo están intrínsecamente vinculados. Por tanto, se debe tener cuidado al reducir dimensionalidad, ya que simplificaciones excesivas podrían llevar a pérdida de información crítica para la toma de decisiones.

El clustering con K-means identificó dos grupos bien diferenciados, lo cual sugiere una estructura urbana bipolar : áreas centrales altamente desarrolladas y periferias con menor nivel de desarrollo. El alto valor del índice Calinski-Harabasz (291.30) confirma la solidez del agrupamiento y refuerza la utilidad de este método en contextos donde se busca una segmentación clara y funcional de la ciudad.

Por otro lado, el análisis con DBSCAN mostró limitaciones significativas: aunque logró identificar 19 clusters densos, el 93.48% de los puntos fueron clasificados como ruido , lo que evidencia una distribución espacial altamente dispersa. La baja puntuación del índice Calinski-Harabasz (6.60) sugiere que los clusters encontrados son poco compactos o están influenciados por transiciones espaciales complejas. Esto podría deberse a la heterogeneidad del desarrollo urbano en América Latina, caracterizado por zonas de expansión irregular y asentamientos informales.

Referencias:

- [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/10.1177/0420980231177688>

- [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, 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), 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., & Rahnamayizekavat, 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>
- [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/10.1177/1049732324124586>
- [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 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.

Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)

This article is published under the terms of the Creative Commons Attribution License 4.0.

<https://creativecommons.org/licenses/by/4.0/deed.es>