



Dignidad, un derecho inalienable: Caracterización de grupos de personas que habitan la calle en 21 municipios de Colombia con metodologías de segmentación y clasificación

Laura Estefanía Noriega Perlaza
Juan Diego Villa Franco

Monografía presentada para optar al título de Especialista en Analítica y Ciencia de Datos

Asesor
Daniel Escobar Saltarén, Magister (MsC) en Ingeniería

Universidad de Antioquia
Facultad de Ingeniería
Especialización en Analítica y Ciencia de Datos
Medellín, Antioquia, Colombia
2024

Cita	Noriega Laura y Villa Juan [1]
Referencia	[1] L. E. Noriega Perlaza y J. D. Villa Franco, “Dignidad, un derecho inalienable: Caracterización de grupos de personas que habitan la calle en 21 municipios de Colombia con metodologías de segmentación y clasificación. [Trabajo de grado especialización]. Universidad de Antioquia, Medellín, Colombia.
Estilo IEEE (2020)	



Especialización en Analítica y Ciencia de Datos, Cohorte VII.

Grupo de Investigación Intelligent Information Systems Lab – In2Lab.

Centro de Investigación Ambientales y de Ingeniería (CIA).



Centro de Documentación Ingeniería (CENDOI)

Repositorio Institucional: <http://bibliotecadigital.udea.edu.co>

Rector: John Jairo Arboleda Céspedes.

Decano: Julio Cesar Saldarriaga Molina

Jefe departamento: Danny Alexandro Múnera Ramírez

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

Dignidad, un derecho inalienable: Caracterización de grupos de personas que habitan la calle en 21 municipios de Colombia con metodologías de segmentación y clasificación.

Abstracto — Los habitantes de calle en Colombia son una población que bajo el margen constitucional, son sujetos con derechos que deben ser protegidos por políticas públicas que el gobierno extiende para garantizar la salud de estas personas. Aun así, el habitar la calle es una condición en la que una persona por diferentes motivos personales se ve obligada a asumir espacios, territorios y formas de vida particulares. El DANE realizó un censo en el 2019 a esta población en 21 municipios de 5 departamentos de Colombia, insumo en base de datos anonimizada que dispone en microdatos libres para su uso en investigación. En este artículo se encuentran 7 clústeres de personas usando metodologías estadísticas y analíticas de aprendizaje automático como K-Means, se valida con Árboles de decisión la predicción de estos grupos y con el algoritmo LIME se encuentran las principales características de cada clúster. Con los resultados se espera, ampliar el estado del arte en análisis cuantitativo de esta temática y que los resultados contribuyan a proponer soluciones basadas en las necesidades de cada grupo..

Índice de términos — Habitantes de Calle, Analítica de Datos, Clustering, K-Means, Segmentación, Clasificación, Caracterización.

I. INTRODUCCIÓN

La Organización Mundial de la Salud (OMS) al referirse a condiciones de exclusión social, define a los habitantes de calle como “un grupo especialmente vulnerable; que comprende grupos en los que se podría incluir personas que generalmente no tienen recursos económicos, sin redes de salud, sociales o familiares de apoyo”. [1] Por su parte, el Ministerio de Salud de Colombia según la Ley 1641 del 12 de julio de 2013, define el término “Habitante de la calle” como la persona sin distinción de: sexo, raza o edad; que hace de la calle su lugar de habitación, ya sea de forma permanente o transitoria. [2] El foro de economía mundial en 2021 reporta que alrededor del mundo 150 millones de personas viven en condición de calle. [3]

Para la población Colombiana, según el Censo elaborado por el DANE en Bogotá en el 2017 y para el resto de municipio en 2019, el 0,13% está en situación de calle,

lo que corresponde a un total de 22.790 personas. Los habitantes de calle se concentran principalmente en Bogotá (41,85 %), Cali (20,84 %), y Medellín (14,10 %) [4].

Según una caracterización aplicada a este grupo poblacional realizada entre 2017 y 2021 en territorio Colombiano, “la mayoría de las personas habitantes de calle se ubican en la región andina; son jóvenes y adultas entre los 25 y 44 años sin una pertenencia étnica definida y con limitado acceso a la educación, en general cuentan solo con educación básica primaria [4]. Los principales motivos para iniciar la vida en calle son: el consumo de sustancias psicoactivas (SPA) en los hombres; y los conflictos y dificultades familiares, lo son en las mujeres. [5] Aproximadamente dos (2) de cada cinco (5) habitantes de la calle llevan 10 o más años en esta situación, generalmente inician la vida en calle entre los 14 y 29 años [4]. Se considera que cerca de la tercera parte de esta población se encuentra en un alto grado de exclusión, quienes mencionan que no cuentan con: redes de apoyo de personas ni de instituciones, limitado acceso a servicios de salud y sin un enfoque diferencial [6].

Las personas sin hogar enfrentan una vulnerabilidad extrema debido a la falta de vivienda estable. Están expuestas a condiciones climáticas adversas, violencia, explotación y abuso. Esta situación puede llevar a una serie de problemas de salud física y mental [7]. A su vez la falta de vivienda conlleva a un acceso limitado a servicios básicos como: atención médica, saneamiento, agua potable, alimentación adecuada y servicios de higiene. Esto puede dar lugar a problemas de salud pública como: enfermedades infecciosas y parasitarias. La falta de acceso regular a los servicios de salud y medicina preventiva aumenta el riesgo de padecer otras patologías crónicas como: diabetes, enfermedades cardíacas, enfermedades respiratorias, enfermedades transmitidas por vectores y en el caso de las mujeres cánceres prevenibles [2], como el cáncer de cuello uterino, debido a la mayor vulnerabilidad de adquirir infecciones por Virus del Papiloma Humano (VPH) de alto riesgo. [8] Muchas personas sin hogar sufren principalmente de problemas de salud mental no tratados como: depresión,

ansiedad, trastorno de estrés postraumático y esquizofrenia.[9]

Debido al gran reto social y sanitario que representa la situación de los habitantes de calle en el país, se evidencian importantes brechas de género, particularmente en el consumo de drogas.[2] Los estudios indican que las mujeres en situación de calle presentan vulnerabilidades específicas: son más estigmatizadas socialmente, tienden a un mayor consumo recreativo de medicamentos y experimentan el síndrome de abstinencia de manera diferente, lo que afecta significativamente su estabilidad emocional y salud mental.

Además, existe una marcada desigualdad en los medios utilizados para conseguir drogas: mientras los hombres suelen recurrir a actividades como mendigar, robar o reciclar; las mujeres frecuentemente se ven forzadas a utilizar la seducción y la prostitución como mecanismos de supervivencia, exponiéndolas a riesgos adicionales para su salud física y mental. [10]

Estos riesgos se ven agravados por las barreras significativas que enfrentan los habitantes de calle tanto hombres como mujeres para acceder a servicios de salud, a pesar de que este es un derecho inalienable e indiscutible que, bajo el principio de universalidad, debería garantizar una cobertura total de la población.[11] Sin embargo, la materialización de este derecho se ve obstaculizada por múltiples factores: la falta de capacitación del personal sanitario para atender esta población, la desconfianza arraigada de los habitantes de calle hacia el sistema de salud y la inadecuación de los mecanismos de afiliación existentes.[12] Aun así el ministerio cuenta con documentación llamada: “*Lineamiento para la atención integral en salud de la población en situación de calle*”[13]

Esta situación se complica aún más por factores adicionales como la carencia de documentos de identificación, la discriminación en los centros de salud, los horarios poco flexibles de atención que no se ajustan a sus dinámicas de vida y la falta de seguimiento a los tratamientos debido a su condición de movilidad constante. En el caso particular de las mujeres habitantes de calle, estas barreras se intensifican cuando requieren atención en salud sexual y reproductiva, atención prenatal o servicios de salud mental; lo que aumenta su vulnerabilidad y riesgo de complicaciones médicas a mediano o largo plazo.[14]

Aunque es evidente la necesidad de abordar e intervenir la situación de este grupo poblacional, es fundamental tener presente el principio de autonomía dentro de este fenómeno social. Este principio cobra especial relevancia considerando que gran parte de las personas habitantes de calle optan bajo su autonomía de permanecer en las calles, siendo el consumo de sustancias psicoactivas una característica constante en esta población. (Entre la

salud y la libertad de los habitantes de calle en Colombia)

Es pertinente resaltar que hay una extensa literatura sobre la población habitante de calle a nivel mundial, sin embargo, gran parte de los estudios que se disponen al respecto se abordan desde una naturaleza cualitativa, enfocados en narrativas y testimonios de vida. Es evidente la ausencia de técnicas avanzadas de análisis cómo lo son la segmentación y clasificación, limitando el enfoque y la capacidad de identificar patrones o subgrupos dentro de esta misma población, quienes podrían beneficiarse en intervenciones específicas de ámbito social, sanitario y/o económico en un futuro.

En Colombia al investigar esta comunidad, es de notar que la muestra poblacional con la que exploran diversos comportamientos o eventos de personas habitantes de calle es ínfima con relación al tamaño real de esta población, dado que los investigadores enfocan sus análisis concluyentes generalizando a estos sujetos. [15], [16], [17], [18]

Se entiende que la dinámica de actualización en censos e investigaciones de la población de estudio es compleja, por lo que no se hace hincapié en caracterizar a estos individuos. Con esto surge ciertas preocupaciones, y desde el Concejo de Bogotá recientemente se publica el inquietante crecimiento y las dificultades que presenta esta comunidad, en el que “se hace un llamado urgente a la administración distrital para que prioricen la atención a los habitantes de calle, garantizando su acceso a servicios de salud, vivienda y programas de rehabilitación, con un enfoque diferencial que atienda las necesidades particulares de esta población, especialmente en términos de salud mental y adicciones”. [19]

Por todo lo anterior, es considerable indagar e investigar acerca de este evento y sus causas subyacentes como: la vulnerabilidad extrema, la falta de acceso a servicios básicos, los problemas de salud mental y adicciones, el riesgo de enfermedades crónicas, los problemas de higiene y saneamiento, además de los costos sociales y económicos asociados. Abordar este hecho desde un enfoque exploratorio puede proporcionar información valiosa para aproximarse a las raíces de esta problemática.

El presente trabajo tiene como objetivo general segmentar la población de habitantes de calle encuestada por el DANE en el 2019, en las principales ciudades de Colombia para identificar patrones de comportamiento específicos, de manera que se contribuya a la toma de decisiones basadas en las necesidades de cada grupo. Con este fin, se plantean objetivos específicos para identificar características comunes en las clasificaciones que surjan del proceso de segmentación, y detectar agrupaciones de individuos con comportamientos similares.

II. METODOLOGÍA

A. Origen de los datos

La información fue recolectada a través de un cuestionario con 37 preguntas elaborado por el DANE, que fue diseñado para ser aplicado mediante entrevista directa, bajo el cumplimiento dispuesto en la ley 1641 de 2013. Esta institución gubernamental censó en el 2019 a la población de 21 municipios de 5 departamentos: Atlántico, Antioquia, Santander, Caldas y Valle del Cauca. El instrumento cumple con la condición de que la información resultante se contribuya a tener insumos para la política previamente descrita, en el que se contemplen componentes de dignificar la vida. [20]

La investigación se focalizó en encontrar patrones de la población habitante de calle que puedan dar indicios de comportamientos y/o condiciones de estos individuos que ayuden a ampliar la literatura sobre la comunidad que habita la calle en Colombia. Todo esto tiene sentido si se parte del fundamento de verdad en la respuesta a la encuesta. La campaña tuvo como resultado un total de 13252 encuestados, de los cuales 152 fueron descartadas como formularios incompletos por la institución, y en un posterior análisis de la información en nuestra investigación, se descartaron como incompletas 3084 encuestas, teniendo un total final de 10044 registros.

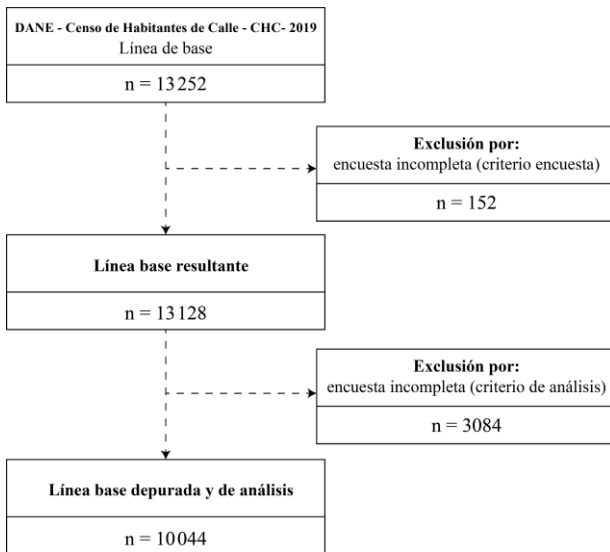


Figura 1. Flujograma de la selección de encuestas para análisis

B. Limpieza de la Base de Datos.

El proceso de limpieza de la base de datos inició con la eliminación de los registros identificados como "formulario incompleto", según lo indicado por el encuestador en el

campo correspondiente. Posteriormente, se llevó a cabo un análisis exploratorio que permitió identificar que los individuos que no proporcionaron información sobre "cuántos años llevan en la calle" presentaban valores faltantes en la mayoría de las variables; por esta razón, se decidió excluir también estos registros. Finalmente, se realizó un análisis descriptivo para evaluar la concentración de las respuestas en las variables. Aquellas variables en las que una de las categorías concentraba más del 95% de las respuestas fueron consideradas desbalanceadas y, por lo tanto, descartadas del análisis para garantizar la relevancia y la calidad de los resultados.

C. Herramientas Utilizadas

El análisis de datos es generado directamente en un notebook de Google Colab[21], utilizando el lenguaje de programación Python (versión 3.10)[22]. Se emplean las bibliotecas: NumPy[23], Pandas[24], el formato JSON[25] se utiliza para leer archivos de metadatos que contienen información de la estructura del conjunto de datos; las visualizaciones estadísticas de datos se generan con Matplotlib[26] y con Seaborn[27]; se usa GeoPandas[28] para análisis geoespacial. Para la manipulación, visualización y segmentación se utiliza el paquete Scikit-learn[29] y para interpretar los resultados del clustering, se usa el método LIME[30], que permite generar explicaciones locales y comprensibles para las predicciones de los segmentos del clasificador.

D. Imputación de Valores Faltantes

Debido a la naturaleza de las encuestas y la presencia de valores faltantes, se realizaron los siguientes procesos de imputación:

1. **Variables cualitativas:** La imputación de estas se realizó asumiendo que la ausencia de una respuesta equivale a una negación. Esto se fundamenta en el diseño de las preguntas de selección múltiple presentes en la encuesta, donde únicamente se registran respuestas afirmativas en caso de que aplique la categoría. Por lo tanto, los valores en blanco en estas variables se interpretaron como un "no", ya que la ausencia de datos implica que los encuestados no seleccionaron dichas opciones como válidas para su caso.
2. **Variables numéricas:** Para la mayoría de las variables numéricas, los valores faltantes designados por el lenguaje en inglés como "Not At Number" (NaN) fueron imputados utilizando la media de cada columna. En el caso específico de la edad, se utilizó el comodín 99 para imputar valores faltantes. Esto se justifica debido a su posible uso como un valor de referencia o *placeholder* que indica la falta de información, permitiendo conservar el registro en el análisis sin generar distorsiones significativas en los resultados. No se optó en este último caso por el valor 0 o negativo, puesto que muchas edades de comienzo de consumo de sustancias son cercanas a este valor, lo que causaría un sesgo en caso de usarse.

E. Transformación y Preparación de Variables

1. Dicotomización: Las variables con más de dos categorías se “dummifican” para convertirlas en un formato binario.
2. Concatenación de Variables: Se crean nuevas variables que combinan información de columnas relacionadas, en el caso de la edad y el sexo, cuando estas no fueron declaradas explícitamente por los participantes, fueron combinadas en una nueva columna basada en las observaciones del encuestador. También se agrupan variables que miden aspectos similares (capacidades físicas, salud mental, consumo de sustancias), con el fin de tener la correcta identificación del encuestado.[20]
3. Agrupación de Variables: Algunas preguntas del cuestionario que medían dimensiones similares se combinan para crear indicadores más robustos:
 - Capacidades físicas y mentales: Se agrupan preguntas sobre movimiento, interacción social y percepción del entorno en variables más generales.
 - Problemas de salud: Se combinan preguntas relacionadas con intoxicaciones, salud mental, y problemas respiratorios en variables globales.
 - Consumo de sustancias: Se agrupan diferentes tipos de consumo (drogas recreativas, estimulantes, opioides) en categorías amplias.

F. Clustering

Se utilizó el Método del Codo para identificar el número óptimo de clústeres en el algoritmo K-Means. Se evaluó la inercia para diferentes números de clústeres desde 2 a 10 agrupaciones.

G. Árbol de Decisión

Se usa la metodología de *Árbol de Decisión* como modelo supervisado con el fin de lograr determinar la correcta clasificación de los clústeres resultantes y proponer una caracterización de cada uno de estos. Las variables predictoras incluyeron todas las características resultantes de la depuración de la base de datos inicial. La columna de etiquetas de clúster resultante de K-Means se usa como variable desenlace. Los datos fueron divididos en conjuntos de entrenamiento (70%) y prueba (30%) para evaluar el desempeño del modelo. Se calcula la precisión en el conjunto de prueba para medir el desempeño global del modelo. Se genera un reporte de clasificación para cada clúster que incluye métricas como: *precisión*, *recall*, y *F1-Score*. Las contribuciones de las variables en el modelo fueron evaluadas y visualizadas para entender cuáles características influyeron más en la correcta clasificación de los clústeres. Se crea una matriz de confusión para evaluar la capacidad del modelo de árbol de decisión en clasificar correctamente las observaciones en sus respectivos clústeres.

H. LIME (Local Interpretable Model-Agnostic Explanations)

Se emplea el método LIME para interpretar las predicciones del árbol de decisión a nivel local. Cada una de las agrupaciones fue evaluada con esta metodología, obteniendo las 10 características más relevantes para su

clasificación en un clúster particular y luego poder identificar cuáles son las características particulares de cada clúster.

III. RESULTADOS

A. Análisis descriptivo.

La Tabla 1 y la figura 2, muestran la distribución de los encuestados en estudio según los 5 departamentos donde se recogieron los datos sumando un total de 10044. El departamento con mayor presencia en la muestra es Valle del Cauca, con 3,522 registros y un 35.1% del total. Antioquia ocupa el segundo lugar con 2,732 registros (27.2%). Atlántico y Santander tienen un peso similar, con 1,772 (17.6%) y 1,462 (14.6%) registros respectivamente. Finalmente, Caldas es el departamento con menor representación, con 556 registros que constituyen el 5.5% de la muestra total.

Departamento	Frecuencia	Porcentaje (%)
Valle del Cauca	3522	35,1
Antioquia	2732	27,2
Atlántico	1772	17,6
Santander	1462	14,6
Caldas	556	5,5
Total	10044	100

Tabla 1. Frecuencia de registros por departamento

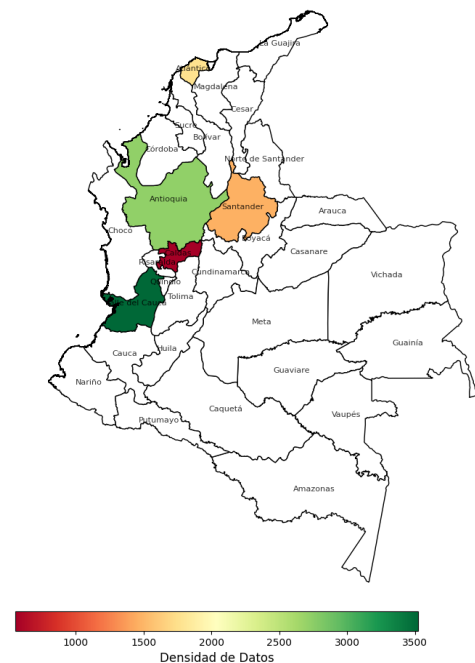


Figura 2. Densidad de encuestas por departamento en Colombia.

El 87.8% de los participantes son hombres, frente a las mujeres quienes representan el 12.2% de la población encuestada.

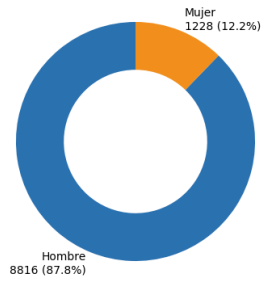


Figura 3. Diagrama circular de proporción por sexo.

La base con mayores proporciones corresponde a los grupos de edad entre los 25 y los 39 años, tanto para los hombres como para las mujeres.

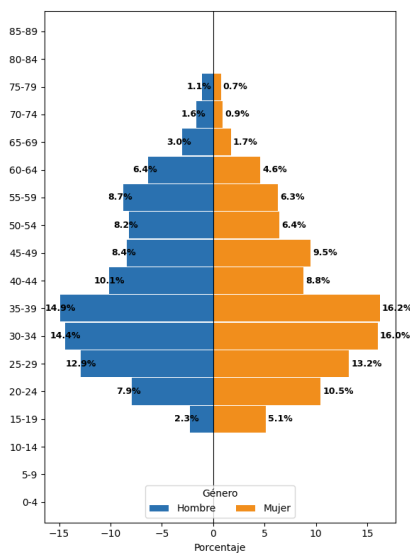
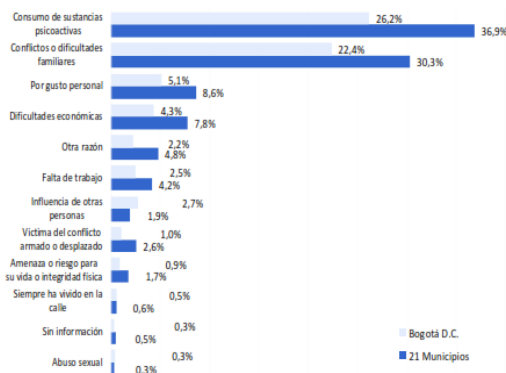


Figura 4. pirámide poblacional.

Las dos principales razones por las que una persona inicia su vida en la calle son: el consumo de sustancias psicoactivas y los conflictos familiares, como se puede apreciar en la Figura 5.



Fuente: DANE Censo habitante de calle 2017 para Bogotá D.C y 2019 para 21 municipios principales

Figura 5. Diagrama de barras de proporción según la razón por la

que inició vida en la calle.

La edad de inicio de consumo de las diferentes sustancias psicoactivas, ver figura 6, se da a lo largo de la vida, yendo desde los 8 hasta los 80 años. Esta iniciación tiene su concentración del 75% por debajo de los 20 años en cada una de las sustancias preguntadas en el cuestionario.

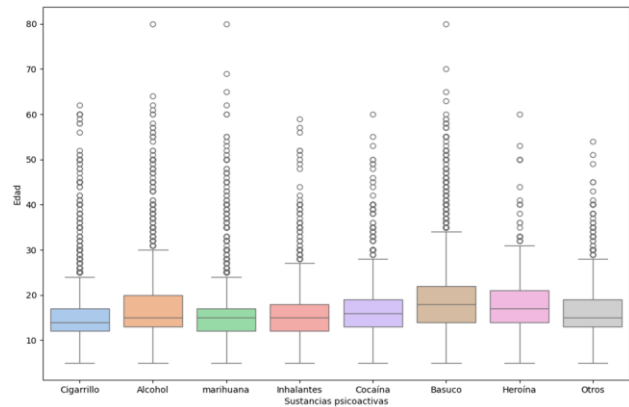


Figura 6. Diagrama de cajas y bigotes según la edad de inicio de consumo de sustancias psicoactivas.

Luego de analizar la información relevante de forma descriptiva, se procede a implementar las metodologías de Aprendizaje automático, también conocida como *machine learning* para segmentar a la población. Se usa K-Means y a través del método de codo se toma el punto en el gráfico donde la disminución de la inercia comienza a estabilizarse, quien indica el número óptimo de clústeres, se decide con este criterio 7 clústeres.

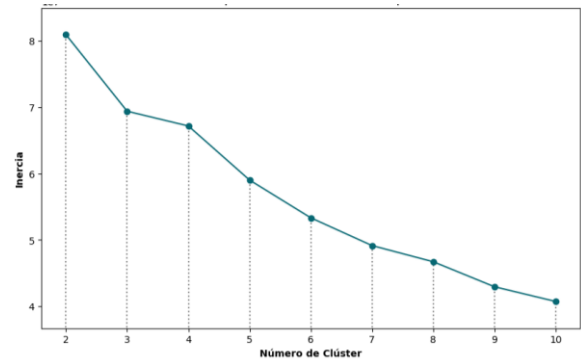


Figura 7. Gráfica de la determinación del número óptimo de clusters mediante el Método del Codo.

Se valida con modelación usando Árbol de decisión, y por medio de métricas como *Precisión*, *Recall* y *F1-Score*, se evalúa el modelo.

- **Precisión:** Los valores de precisión son altos, oscilando entre 0.95 y 0.99, indicando que el modelo clasifica adecuadamente la mayoría de las instancias en sus clústeres correspondientes.
- **Recall:** Los valores de *recall* también son elevados, entre 0.97 y 0.99, detectando eficazmente la mayoría de las instancias pertenecientes a cada clúster.
- **F1-score:** Los valores de F1-score son igualmente muy buenos,

situándose entre 0.96 y 0.99. Esto refleja un equilibrio entre la precisión y el *recall* del modelo.

Cluster	Precision	Recall	F1-Score	Support
0	0,95	0,97	0,96	248
1	0,99	0,99	0,99	553
2	0,96	0,96	0,96	165
3	0,99	0,99	0,99	320
4	0,99	0,99	0,99	633
5	0,99	0,98	0,99	716
6	0,98	0,99	0,98	379
accuracy				
macro avg	0,98	0,98	0,98	3014
weighted avg	0,99	0,99	0,99	3014
Precisión del modelo en el conjunto de prueba: 0,99				

Tabla 2. Desempeño del modelo de clasificación por cluster. incluyendo variables de edad

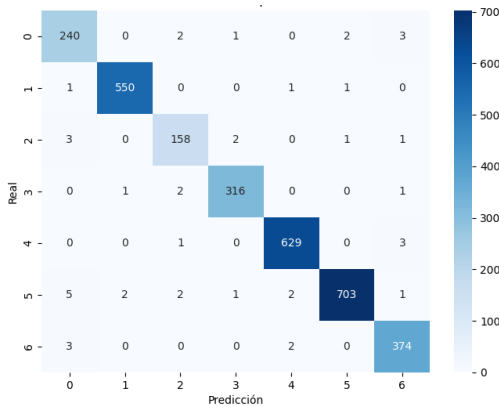


Figura 8. Matriz de confusión según cada clúster.

La matriz de confusión también da cuenta del buen desempeño al modelar los datos, puesto que los datos están concentrados en la diagonal principal.

A su vez, se quiso entrenar un modelo de Árbol de decisión removiendo las variables de edad, y con validación cruzada se evalúa la efectividad de este nuevo modelo, en el que se veía un sobreajuste y se evaluó igualmente la efectividad de este.

Clúster	Precision	Recall	F1-Score	Support
0	0.76	0.81	0.78	361
1	0.88	0.91	0.90	407
2	0.92	0.91	0.92	336
3	0.97	0.96	0.96	702
4	0.92	0.90	0.91	443
5	0.93	0.91	0.92	388
6	0.88	0.86	0.87	377
accuracy				
macro avg	0.90	0.90	0.90	3014

weighted avg	0.90	0.90	0.90	3014
Precisión promedio de los k-folds: 0,9001				
Puntaje de precisión en el conjunto de validación: 0.9027				

Tabla 3. Desempeño del modelo de clasificación por clúster sin variables de edad.

La precisión de este segundo modelo disminuye, como se observa en la tabla 3. Por lo que se parte de este para realizar la caracterización de la población. Para ello se usa LIME, un algoritmo que puede explicar las predicciones de cualquier clasificador de una manera fiel. Así los clústeres están representados por diversas características y condiciones de la siguiente manera:

- **Consumidores de inhalantes o estimulantes que requieren apoyo básico (Clúster 0):** este grupo no consume bazuco, pero si consumen drogas como inhalantes o estimulantes, necesitan ayuda básica para subsistir. Reciben apoyo de instituciones y requieren asistencia en salud y bienestar. Aunque no cuentan con ayuda de personas cercanas ni familiares, son conscientes del entorno que los rodea.
- **Consumidores autovalentes que no buscan ayuda (Clúster 1):** este grupo consume bazuco y otras drogas inhalantes o estimulantes, no requiere ayuda básica para subsistir ni aceptan apoyo de instituciones, familiares o personas cercanas. Pueden valerse por sí mismos, perciben su entorno y no buscan resolver sus problemas de salud.
- **Consumidores con apoyo externo (Clúster 2):** consumen bazuco y otras drogas inhalantes o estimulantes, necesitan ayuda básica para subsistir. Reciben apoyo de instituciones y asistencia para su salud y bienestar. Son conscientes de su entorno y tratan activamente de resolver sus problemas de salud, lo que los distingue como un grupo que busca soluciones.
- **Autovalentes en situación de calle con apoyo institucional (Clúster 3):** este grupo requiere ayuda básica para subsistir y recibe apoyo de instituciones, tanto para necesidades básicas como de salud y bienestar. Consumen inhalantes, duermen habitualmente en instituciones, no presentan problemas físicos significativos y tratan de conseguir dinero para su sustento.
- **Dependientes de familiares (Clúster 4):** este grupo recibe ayuda básica para subsistir, pero no de instituciones. Su principal fuente de apoyo proviene de familiares o personas cercanas. Pueden valerse por sí mismos, interactúan socialmente, perciben su entorno y continúan en la calle debido a razones personales específicas.
- **Consumidores que buscan soluciones activamente (Clúster 5):** este grupo consume bazuco y necesita ayuda básica para subsistir. Reciben apoyo de instituciones, familiares y para su bienestar general. Son conscientes de su entorno, pueden valerse por sí mismos, interactúan socialmente y buscan activamente resolver sus problemas.
- **No consumidores e independientes (Clúster 6):** este grupo no consume bazuco y no recibe ayuda básica, ni de instituciones ni de personas cercanas. Consumen inhalantes y pueden valerse por sí mismos. Aunque perciben su entorno y tratan de resolver sus problemas de salud, no recurren a terapias alternativas ni a profesionales de la salud.

Características	0	1	2	3	4	5	6
Ayuda básica para subsistir	1	0	1	1	1	1	0
Ayuda de instituciones oficiales, privadas o religiosas	1	0	1	1	0	1	0
Ayuda de personas cercanas o del entorno social	0	0			1	1	0
Ayuda para la salud y bienestar	1	0	1	1		1	0
Cómo consigue usted dinero: Construcción				0			
Cómo consigue usted dinero: prostitución	0						
Consumo drogas inhalantes o estimulantes	1	1	1	1	0	1	1
Consumo principalmente: Basuco	0	1	1	0	0	1	0
Consumo principalmente: Pepas				0			
Principal fuente de ayuda que usted recibe: de algún familiar	0						
Principal razón por la que continúa viviendo en la calle: Enfermedad				0			
Principal razón por la que continúa viviendo en la calle: Hermano					0		
Puede hacer las actividades sin problemas cardíacos/respiratorios				1			
Razón por la que empezó a vivir en la calle: Abuso sexual			0		0		
Tratar problema de salud acudió a: brigadas de salud de ONGs			0				
Tratar problema de salud acudió a: terapias alternativas		0	0				0
Tratar problema de salud acudió a: otro médico de un grupo étnico	0		0	0			
Tratar problema de salud acudió a: profesional de la salud/particular							0
Usted duerme habitualmente: en una institución				1			
Usted puede moverse y valerse por sí mismo en las actividades cotidianas		1			1	1	1
Usted puede oír, hablar e interactuar socialmente		1			1	1	1
Usted puede percibir su entorno	1	1	1		1	1	1

Tabla 4. Características más relevantes para la clasificación de cada clúster.

Se generan diagramas de cajas y bigotes para observar cómo se distribuyen las variables de edad de inicio de consumo de sustancias, que son importantes entre los clústeres. Se tiene que el consumo de cigarrillo (B), Alcohol (C) y Marihuana (D), son relevantes en muchos de los clústeres. El basuco (G) tiene mayor concentración de consumidores en los clústeres 1, 5 y 6. Por otro lado, la heroína (H), pepas (I) y otras drogas (J) tienen un consumo más elevado en el clúster 0.

IV. CONCLUSIONES

Analizamos que la edad de inicio de consumo de cada una de las sustancias está en su mayoría en una edad temprana por debajo de los 20 años, lo cual resalta la importancia de intervenir en edades tempranas y adolescentes, ya sea con campañas de educación, prevención y sensibilización dirigidas a estos grupos etarios. Se evidencia que la mayor cantidad de encuestados fueron hombres, por lo que se puede decir que 1 de cada 10 personas habitantes de calle es mujer. Con el análisis de esta investigación se encuentra que mayoría de esta población cuenta con una edad entre los 20 a 40 años, tanto para hombres como mujeres, lo que corrobora lo mencionado con la literatura en la introducción y el boletín del censo reportado por el DANE

Los resultados obtenidos en la encuesta revelan una clara desproporción en la participación entre hombres y mujeres, con un 88% de los encuestados siendo hombres y solo un 12% mujeres." A partir de esto, se podrían discutir varios puntos, como la necesidad de investigar las razones detrás de esta disparidad. Se plantea preguntarse si ¿Existen menos mujeres en la calle o ellas mismas participan menos en este tipo de encuesta? De ser así, podría sugerirse el uso de estrategias que incentiven la participación de las mujeres.

Para segmentar la población habitante de calle en las principales ciudades de Colombia y lograr identificar patrones específicos, el análisis del método del codo permitió identificar que el número óptimo de clústeres es 7, garantizando una segmentación equilibrada sin perder capacidad explicativa. Por lo tanto, estos 7 grupos representan grupos heterogéneos entre sí según factores: sociales, de

salud, de consumo de sustancias y de si reciben o requieren

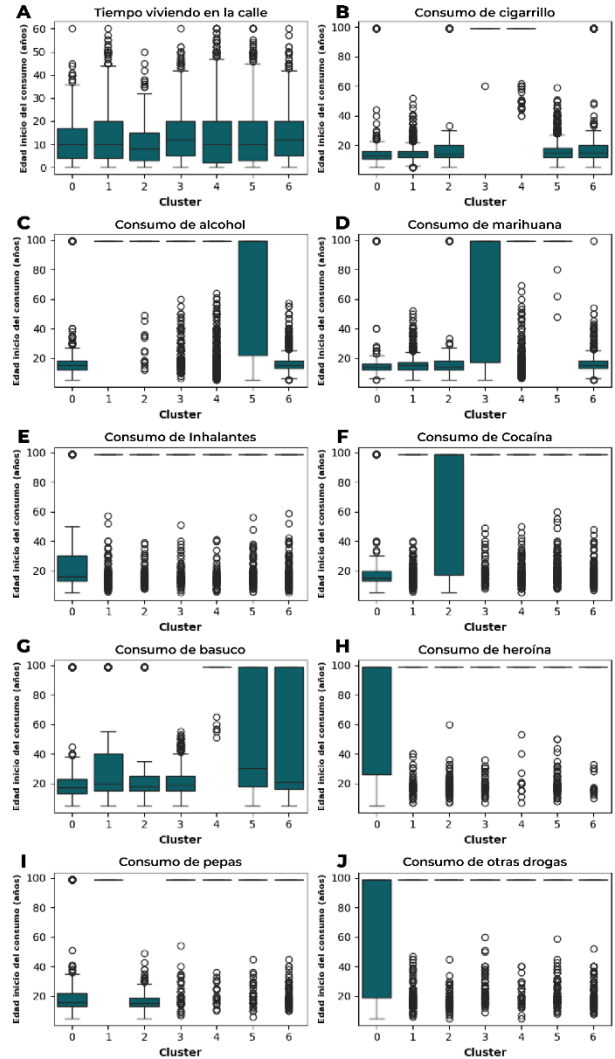


Figura 9. Matriz de Diagrama de cajas y bigotes de edades de inicio de consumo de sustancias psicoactivas.

algún tipo de ayuda.

En cada uno de los clústeres se evidencia el requerimiento de estrategias específicas óptimas y eficaces para tener en cuenta dentro de las políticas públicas orientadas en la intervención y tratamientos de estos grupos poblacionales. Para los clústeres 1, 2 y 5, que agrupan a consumidores de basuco y otras drogas como inhalantes o estimulantes, se propone desarrollar planes de acción orientados a concientizar sobre los riesgos neurológicos y sociales del consumo constante, ya que este deterioro afecta su interacción social, a pesar de que muchos perciben valerse por sí mismos al momento de la encuesta. Por otra parte para los agrupados en los clústeres 0 y 3 que hacen referencia a "consumidores de inhalantes y estimulantes que a su vez reciben ayudas", se recomienda que se siga brindando ese apoyo y acompañamiento para mantener esta población supervisada y que en un futuro se dé la reinserción social de forma gradual para ellos.

Las personas que clasificaron en el clúster 6 que indican que también son consumidores, como se describió para los clústeres 0 y 3, y que por el contrario no reciben algún

tipo de ayuda, estos individuos también deben ser considerados en la reinserción social que se propone y ofrecerles programas de apoyo que incentiven la ayuda institucional y acompañamiento familiar. Por último, con respecto al clúster 4 que agrupa a los habitantes de calle que no son consumidores, que se valen por sí mismos y reciben ayuda por parte de personas cercanas, pero no de instituciones, se resalta la importancia de no dejarlos desatendidos. Es fundamental que este grupo sea considerado en programas y campañas de prevención y promoción, dado que por su perfil presentan mayores posibilidades de ser abordados con éxito en intervenciones dirigidas a facilitar su reintegración en la sociedad. Promover iniciativas específicas para este grupo podría resultar en procesos de inclusión social más efectivos y sostenibles.[31]

En términos generales para lograr la reinserción efectiva que se propone a los grupos clasificados en esta investigación, resulta elemental reforzar programas orientados al empleo formal y la promoción de la autonomía económica, donde estos ingresos se destinen a propósitos constructivos, fomentando una economía consciente que permita a los beneficiarios contribuir de manera productiva a la sociedad. Asimismo, se destaca la necesidad de fortalecer estrategias que consoliden los lazos con sus familias y comunidades, facilitando así su integración social.

Las características de los clústeres muestran que las políticas deben considerar aspectos multidimensionales que incluyan los asuntos de: educación, empleo, vivienda, salud y una adecuada asignación de recursos en las ciudades más afectadas y sobre todo en los grupos que representan mayor vulnerabilidad. No obstante, estas políticas deben ser medidas a través de indicadores de impacto como: tasa de reinserción laboral, acceso a vivienda digna, cobertura en servicios de salud, reducción del consumo de sustancias psicoactivas, fortalecimiento de vínculos familiares, reducción en la reincidencia a la vida en calle, eficiencia en la asignación de recursos, tasa de escolarización y formación. Estos indicadores y los que puedan surgir al respecto permitirán medir de manera sistemática los avances de las políticas, identificar áreas de mejoras y garantizar la sostenibilidad de estos programas en el tiempo.[32]

En definitiva, esta investigación logró cumplir con los objetivos planteados al segmentar la población de habitantes de calle encuestada por el DANE en 2019 en 21 municipios de áreas metropolitanas de Colombia. Los modelos implementados demostraron ser robustos y consistentes, reforzando la validez de los resultados obtenidos. Esto permitió identificar patrones específicos de comportamiento en esta población, contribuyendo de manera significativa al conocimiento de sus necesidades y características.

Los resultados no solo revelaron agrupaciones de individuos con comportamientos similares, sino que también facilitaron la identificación de características comunes dentro de cada clúster, aportando información valiosa para la toma de decisiones orientadas a estrategias más sostenibles e inclusivas que promuevan el bienestar de los habitantes de calle y fomenten su reintegración en la sociedad.

A. LIMITACIONES

Una de las principales limitaciones identificadas durante el desarrollo de esta investigación es que los resultados están restringidos a la población encuestada en el año 2019, lo que implica que podrían no ser representativos de la situación actual. Esto resalta la necesidad de intentar establecer censos periódicos para esta población, teniendo en cuenta que se busca tener un cuestionario más ligero con preguntas claves de identificación e información relevante, lo que permitiría evaluar cómo los indicadores evolucionan a lo largo del tiempo y facilitar la implementación de acciones de mejora basadas en evidencia.

Otra dificultad importante radica en la limitada documentación disponible con análisis cuantitativos sobre esta población en el estado del arte colombiano. Existe una carencia de informes, investigaciones y datos históricos accesibles, lo que dificulta un análisis más robusto y longitudinal. Se propone que investigaciones como la nuestra, que contengan análisis estadísticos y metodologías de segmentación, sean cobijados por proyectos de ley de entes estatales que ayuden a encontrar patrones de esta población.

Para que estas propuestas sean viables, es esencial un esfuerzo conjunto entre las instituciones gubernamentales, organizaciones no gubernamentales y la sociedad civil, respaldado por voluntad política. Esto garantiza un acceso equitativo a servicios clave como: salud, vivienda y programas de rehabilitación; implementados bajo un enfoque diferencial que optimice los recursos disponibles.

También se propone incorporar un análisis geográfico con una perspectiva temporal. Este análisis incluiría datos espaciales para identificar regiones con mayor concentración de esta población para evaluar cómo varían los patrones de movilidad en función del tiempo, evitando una entrevista directa con el individuo de estudio. Esta dimensión espacial ofrecería una herramienta adicional para la planificación de estrategias localizadas y la asignación eficiente de recursos en zonas prioritarias.

V. REFERENCIAS

- [1] J. P. Mayorga, C. Navarrete, y A. Uquillas, «Norma técnica para el servicio de atención a personas habitantes de calle.», *Minist. Inclusión Económica Soc.*, p. 36, 2021.
- [2] EL CONGRESO DE COLOMBIA, *Política pública de Habitante en Calle*. 2023. Accedido: 24 de noviembre de 2024. [En línea]. Disponible en: <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/INEC/IGUB/ley-1641-de-2013.pdf>
- [3] «PREVIOUS HOMELESS WORLD CUPS», *Homeless World Cup*. Accedido: 24 de noviembre de 2024. [En línea]. Disponible en: <https://www.homelessworldcup.org/previous-tournaments>
- [4] Ministerio de Salud y Protección Social, «Boletines Poblacionales: Personas en Situación de Calle». diciembre de 2020. [En línea]. Disponible en: <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDi>

- gital/RIDE/DE/PS/280920-boletines-poblacionales-habitante-calle-2020.pdf
- [5] H. A. Alejo y A. S. A. Espinosa, «Caracterización del abuso de sustancias psicoactivas en población habitante de calle», *Subdir. Vigil. En Salud Pública Alcaldía Bogotá*, 2015, [En línea]. Disponible en: <https://www.minjusticia.gov.co/programas-co/ODC/Publicaciones/Publicaciones/CO031052015-caracterizacion-abuso-sustancia-psicoactivas-habitantes-calle.pdf>
- [6] L. V. Q. Pacheco, «La exclusión social en “habitantes de la calle” en Bogotá. Una mirada desde la bioética», vol. 3, 2008.
- [7] P. Corrigan, «How stigma interferes with mental health care», *Am. Psychol.*, vol. 59, n.º 7, pp. 614-625, oct. 2004, doi: 10.1037/0003-066X.59.7.614.
- [8] R. Asgary *et al.*, «Cervical cancer screening among homeless women of New York City shelters», *Matern. Child Health J.*, vol. 20, n.º 6, pp. 1143-1150, jun. 2016, doi: 10.1007/s10995-015-1900-1.
- [9] Ministerio de Salud y Protección Social, «Caracterización de personas Habitantes de la Calle en Colombia entre 2017 y 2021: Actualización de línea de base de la Política Pública Social para Habitantes de la Calle», octubre de 2022. [En línea]. Disponible en: <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/DE/PS/caracterizacion-habitantes-calle-colombia-2017-2021.pdf>
- [10] J. D. Zabala-Sandoval, M. P. López-Parra, J. D. Zabala-Sandoval, y M. P. López-Parra, «FACTORES DE VULNERABILIDAD COGNITIVA DE LA PERMANENCIA EN CALLE: DESESPERANZA Y DERELICCIÓN», *Límite Arica*, vol. 16, pp. 0-0, 2021, doi: 10.4067/s0718-50652021000100202.
- [11] F. F. Dueñas Gaitán, D. F. Peña Garnica, G. S. Acuña Daza, y A. D. P. García Donato, *Habitantes de calle y responsabilidad social*, 1.ª ed. Corporación Universitaria Minuto de Dios - UNIMINUTO, 2022. doi: 10.26620/uniminuto/978-958-763-596-6.
- [12] P. Omerov, Å. G. Craftman, E. Mattsson, y A. Klarare, «Homeless persons’ experiences of health- and social care: A systematic integrative review», *Health Soc. Care Community*, vol. 28, n.º 1, pp. 1-11, 2020, doi: 10.1111/hsc.12857.
- [13] Oficina de promoción social, «Lineamiento para la atención integral en salud de la población en situación de calle». 2021.
- [14] M. Lemus y V. Bohórquez, «Entre la salud y la libertad de los habitantes de calle en Colombia», 2023.
- [15] M. J. O. Salazar, E. J. C. Cencio, y M. G. Vargas, «Reconocimiento social en habitantes de calle en Medellín, Colombia», *Rev. Colomb. Cienc. Soc.*, vol. 12, n.º 2, Art. n.º 2, ago. 2021, doi: 10.21501/22161201.3548.
- [16] C. G. Urueta, «perspectiva social-preventiva».
- [17] O. A. Alfonso R *et al.*, «El ciclo mortal de los habitantes de calle en Bogotá. Teorías, olvidos, políticas y desenlaces fatales», *Rev. Econ. Inst.*, vol. 21, n.º 41, pp. 99-131, dic. 2019, doi: 10.18601/01245996.v21n41.05.
- [18] J. Jaramillo, T. Fernández, y S. Bedoya, «Habitantes de calle: Entre el mito y la exclusión», 2017. [En línea]. Disponible en: <https://revistas.ucatolicaluisamigo.edu.co/index.php/poiesis/article/download/2311/1748/8982>
- [19] «Concejo de Bogotá D.C. - Habitante de Calle: Un Fenómeno que Requiere Atención Urgente». Accedido: 29 de noviembre de 2024. [En línea]. Disponible en: <https://concejodebogota.gov.co/habitante-de-calle-un-fenomeno-que-requiere-atencion-urgente/cbogota/2024-09-23/150526.php>
- [20] «Colombia - Censo de Habitantes de Calle - CHC-2021». Accedido: 29 de noviembre de 2024. [En línea]. Disponible en: <https://microdatos.dane.gov.co/catalog/720>
- [21] Google, «colab.google», colab.google. Accedido: 29 de noviembre de 2024. [En línea]. Disponible en: <http://0.0.0.0:8080/>
- [22] Van Rossum, G. and D. Jr, y Fred L, «Python reference manual». Centrum voor Wiskunde en Informatica Amsterdam, 1995. Accedido: 29 de noviembre de 2024. [En línea]. Disponible en: <https://www.python.org/about/>
- [23] C. R. Harris *et al.*, «Array programming with NumPy», *Nature*, vol. 585, pp. 357-362, 2020, doi: 10.1038/s41586-020-2649-2.
- [24] W. McKinney y others, «Data structures for statistical computing in python», en *Proceedings of the 9th Python in Science Conference*, Austin, TX, 2010, pp. 51-56.
- [25] F. Pezoa, J. L. Reutter, F. Suarez, M. Ugarte, y D. Vrgoč, «Foundations of JSON schema», en *Proceedings of the 25th International Conference on World Wide Web*, International World Wide Web Conferences Steering Committee, 2016, pp. 263-273.
- [26] J. D. Hunter, «Matplotlib: A 2D graphics environment», *Comput. Sci. Eng.*, vol. 9, n.º 3, pp. 90-95, 2007.
- [27] M. Waskom *et al.*, *mwaskom/seaborn: v0.8.1 (September 2017)*. (septiembre de 2017). Zenodo. doi: 10.5281/zenodo.883859.
- [28] K. Jordahl, «GeoPandas: Python tools for geographic data», *URL <https://github.com/geopandas/geopandas>*, 2014.
- [29] F. Pedregosa *et al.*, «Scikit-learn: Machine learning in Python», *J. Mach. Learn. Res.*, vol. 12, n.º Oct, pp. 2825-2830, 2011.
- [30] M. T. Ribeiro, «lime: Local Interpretable Model-Agnostic Explanations for machine learning classifiers», *J. Mach. Learn. Res.*, 2016.
- [31] C. M. Baptista, G. E. Herrera, y L. Z. Piedrahita, «Consumo De Bazuco Y La Ruta Hacia La Situación De Calle: Crónicas Y Escenarios De La Vivencia Familiar», *Investig. Andina*, vol. 21, n.º 39, pp. 307-324, 2019.
- [32] «Colombia - Censo de Habitantes de Calle - CHC-2021». Accedido: 29 de noviembre de 2024. [En línea]. Disponible en: <https://microdatos.dane.gov.co/index.php/catalog/720/>