Evaluación de patrones de comportamiento criminal en Bucaramanga mediante machine learning no supervisado

Juan Camilo Betancourt Monsalve

Asesor

Isaac Esteban Camargo Freile

Universidad Nacional Abierta y a Distancia UNAD

Escuela de Ciencias Básicas, Tecnología e Ingeniería ECBTI

Especialización en Ciencia de Datos y Analítica

2025

Isaac Esteban Camargo Freile

_____

Director de Trabajo de Grado

Julio Eduardo Mejía Manzano

_____

Jurado

2025

Dedicatoria

A mis padres, Gloria y Ángel, por su apoyo incondicional, y a mi hermano Felipe, por sus consejos.

Agradecimientos

Agradezco a la Universidad Nacional Abierta y a Distancia por brindarme la oportunidad de adquirir las herramientas conceptuales y metodológicas que hicieron posible el desarrollo de este proyecto. Extiendo un especial agradecimiento al profesor Isaac Esteban Camargo Freile por su valioso acompañamiento, sus observaciones rigurosas y su dedicación durante todo el proceso. Su orientación fue fundamental para enriquecer este trabajo.

Resumen

Este trabajo identifica y analiza patrones espaciales de criminalidad en zonas de Bucaramanga mediante técnicas de aprendizaje automático no supervisado, específicamente los algoritmos K-Means y clustering jerárquico. A partir de datos del periodo 2016–2023 extraídos de la plataforma Datos Abiertos Colombia, se construyó una base consolidada con variables delictivas, demográficas y espaciales, complementadas con tasas ajustadas como la Razón de Morbilidad Estandarizada (RME). La segmentación permitió identificar cinco clústeres útiles con características delictivas diferenciadas entre zonas. Entre los hallazgos clave se destacan: zonas con alta incidencia de violencia intrafamiliar y afectación a población infantil, zonas críticas de crimen organizado y delitos patrimoniales en sectores comerciales, así como sectores residenciales con baja incidencia delictiva. La coherencia de los agrupamientos fue validada mediante métricas como el índice de silueta, y los resultados se integraron en una aplicación interactiva desarrollada con Streamlit. Esta herramienta, de acceso público, incluye un módulo asistido por modelos de lenguaje (LLM) que facilita la interpretación automatizada y la generación de recomendaciones prácticas. Los hallazgos aportan insumos clave para el diseño de estrategias diferenciadas de seguridad pública, promoviendo una planificación territorial más focalizada, preventiva y basada en evidencia.

***Palabras clave*:** Clustering, K-Means, Jerárquico, criminalidad, Bucaramanga, zonas, RME, Streamlit, LLM.

Abstract

This study identifies and analyzes spatial patterns of criminality in zones of Bucaramanga using unsupervised machine learning techniques, specifically the K-Means and hierarchical clustering algorithms. Based on data from the 2016–2023 period extracted from the Datos Abiertos Colombia platform, a consolidated database was built incorporating criminal, demographic, and spatial variables, complemented by adjusted rates such as the Standardized Morbidity Ratio (RME). The segmentation process identified five useful clusters with distinct criminal characteristics across different zones. Key findings include: zones with a high incidence of domestic violence and impact on children, critical areas of organized crime and property offenses in commercial sectors, as well as residential zones with lower crime rates. The coherence of the groupings was validated using metrics such as the silhouette index, and the results were integrated into an interactive application developed with Streamlit. This publicly accessible tool includes a module powered by large language models (LLMs) that facilitates automated interpretation and the generation of practical recommendations. The findings provide key inputs for designing differentiated public security strategies, promoting more focused, preventive, and evidence-based territorial planning.

***Keywords*:** Clustering, K-Means, Hierarchical, criminality, Bucaramanga, zones, RME, Streamlit, LLM.

Tabla de Contenido

Lista de Tablas

Lista de Figuras

Lista de Apéndices

Introducción

La creciente inseguridad ciudadana en muchas ciudades latinoamericanas ha configurado un panorama de desafíos complejos en materia de seguridad pública. En el caso de Colombia, la criminalidad ha presentado dinámicas cambiantes en los últimos años, con un incremento del 4 % en los delitos registrados durante 2022, principalmente en aquellos contra el patrimonio económico, como el hurto y la extorsión, mientras que los delitos contra la vida e integridad personal, como el homicidio y las lesiones, mostraron una tendencia a la baja (Núñez et al., 2023, p.187).

En este contexto, Bucaramanga ha evidenciado un incremento en la incidencia de delitos, afectando de manera directa la calidad de vida de sus habitantes y el desarrollo económico y urbano de la región. Esta problemática se encuentra estrechamente vinculada a factores estructurales como la desigualdad social, el desempleo y la informalidad, los cuales generan un entorno propicio para la violencia intrafamiliar, la desintegración familiar y otras manifestaciones delictivas. La inseguridad ha erosionado la cohesión social, debilitando la convivencia y fomentando la justicia por mano propia ante la falta de respuesta estatal. Factores como el microtráfico, el hurto y la debilidad del sistema judicial alimentan un ciclo de impunidad que afecta la confianza institucional. Esta situación también impacta negativamente la economía local, desincentivando la inversión y limitando la capacidad pública para impulsar el desarrollo y el bienestar social. (Alcaldía de Bucaramanga, 2024, p. 258).

Frente a este escenario, se vuelve prioritario fortalecer las capacidades analíticas del Estado y de los gobiernos locales para comprender la distribución territorial del delito y sus determinantes. Una aproximación integral que incorpore variables espaciales y sociodemográficas permitiría orientar intervenciones más eficaces, focalizadas y sostenibles. En

este contexto, el uso de herramientas de análisis basadas en datos se presenta como una oportunidad estratégica para transformar la información delictiva en conocimiento útil para la toma de decisiones.

El presente estudio se justifica en esta necesidad. A través de enfoques de análisis cuantitativo y modelos de aprendizaje automático no supervisado, se propone segmentar las comunas de Bucaramanga según sus características delictivas, con el fin de promover una gestión de seguridad pública más focalizada, preventiva y basada en evidencia.

Se parte de la hipótesis de que la criminalidad en Bucaramanga presenta una distribución no aleatoria, estructurada en clústeres con características diferenciadas que pueden ser detectadas mediante técnicas de agrupamiento. En consecuencia, el objetivo general del estudio es evaluar los patrones de comportamiento criminal mediante el uso de algoritmos de clustering, específicamente K-Means y agrupamiento jerárquico, para segmentar las zonas de la ciudad en función de sus características delictivas y demográficas. Asimismo, se busca identificar y organizar las variables más relevantes en una base de datos estructurada, sintetizar los clústeres formados y facilitar su interpretación a través de herramientas interactivas.

Para cumplir estos objetivos, se emplea un enfoque cuantitativo no supervisado basado en técnicas de minería de datos. El análisis se sustenta en registros delictivos de la ciudad entre 2016 y 2023, y se inicia con la identificación y estructuración de variables relevantes para la creación de una base de datos óptima. Posteriormente, se aplican los modelos de clustering definidos para sintetizar comunas con perfiles delictivos similares. Como complemento, se desarrolla una herramienta visual basada en Streamlit que permite la exploración geoespacial de los clústeres por parte de los tomadores de decisiones. Esta herramienta incluye un módulo

basado en modelos de lenguaje (Large Language Models, LLM), que facilita la generación de insights y la interpretación asistida de los patrones criminales detectados.

El documento se organiza de la siguiente manera: primero se presenta el marco teórico; luego, la metodología utilizada; posteriormente, se exponen los resultados del análisis de clústeres; a continuación, se discuten los hallazgos en relación con el contexto de la ciudad; y, por último, se ofrecen las conclusiones y recomendaciones derivadas del estudio.

Planteamiento del Problema

Bucaramanga ha venido enfrentando un grave problema de seguridad ciudadana, evidenciado en el incremento sostenido de la criminalidad y la violencia urbana. La ciudad ha experimentado, en los últimos años, un deterioro en la seguridad derivado de la creciente influencia del microtráfico y la consolidación de grupos criminales dedicados al hurto. Esta evolución histórica delictiva ha generado un ambiente de inseguridad que afecta la calidad de vida de los habitantes y repercute negativamente en la economía local, la inversión y el desarrollo urbano (Alcaldía de Bucaramanga, 2020).

Aunque este fenómeno afecta a muchas ciudades latinoamericanas, su manifestación varía significativamente entre países y contextos locales. Según el Informe Regional de Desarrollo Humano del PNUD (2021), en México algunos estados presentaron tasas de homicidio superiores a 200 por cada 100.000 habitantes, mientras que otros no superaron los 2. En El Salvador, el 5 % de los municipios concentró la mitad de los homicidios del país, y en Brasil, ciudades como Belo Horizonte concentraron la violencia en solo 6 de sus 81 distritos. En Río de Janeiro, el 50 % de los homicidios se registró en apenas el 1,1 % del espacio urbano. En Colombia, además de una de las tasas más altas de homicidios en adolescentes del mundo, entre el 10 % y el 20 % de la población reportó haber sido víctima de algún delito (Programa de las Naciones Unidas para el Desarrollo (PNUD), 2021)

Estas cifras evidencian que el crimen es un fenómeno multidimensional, que se concentra territorialmente y responde a dinámicas locales. Por tanto, el análisis de ciudades intermedias como Bucaramanga exige enfoques diferenciados que reconozcan estas particularidades.

Diversos factores estructurales agravan esta problemática. Entre ellos se destacan la impunidad y la corrupción, así como la carencia de recursos y herramientas tecnológicas en el

sistema judicial, lo que limita la capacidad de las autoridades para prevenir y combatir eficazmente el crimen. Por ejemplo, la incidencia de hurtos a personas se ha disparado de 761.9 a 1,466.7 incidentes por cada 100,000 habitantes en 2023, según datos de la Alcaldía de Bucaramanga (2024). Este alarmante aumento evidencia no solo la intensificación del delito, sino también la ineficacia de las estrategias tradicionales de seguridad.

Además, (Gelvez Ferreira et al., 2022) señalan que la limitada disponibilidad y actualización de la información estadística y georreferenciada impide el desarrollo de modelos predictivos precisos, lo que obstaculiza la identificación y análisis de patrones complejos en las dinámicas criminales.

Frente a este escenario, se hace evidente la necesidad de explorar enfoques innovadores que permitan comprender de manera integral la distribución y el comportamiento delictivo. En particular, el análisis basado en técnicas de Machine Learning no supervisado, como el clustering, ofrece una alternativa prometedora para descubrir patrones ocultos en los datos, sin depender de una variable objetivo predefinida. Su utilidad ha sido ampliamente demostrada en investigaciones recientes desarrolladas en Colombia. Por ejemplo, (Fontalvo Herrera et al., 2023) aplicaron técnicas de clustering jerárquico para segmentar los 32 departamentos del país en cuatro conglomerados diferenciados según la incidencia de delitos violentos como homicidios, secuestro, terrorismo y extorsión, logrando identificar zonas con mayor impacto delictivo y proponiendo una red neuronal capaz de predecir la pertenencia futura de un territorio a uno de estos perfiles con una precisión del 97,7 %. Estos resultados reafirman el potencial del enfoque no supervisado para optimizar la planificación, focalizar estrategias de prevención del delito y orientar políticas públicas basadas en evidencia, lo cual resulta especialmente pertinente para ciudades intermedias como Bucaramanga.

Con lo cual se plantea la siguiente interrogante:

¿Qué patrones espaciales y demográficos emergen al aplicar técnicas de clustering sobre delitos registrados en Bucaramanga?

Justificación

En el contexto actual de creciente preocupación por la seguridad ciudadana, se hace indispensable explorar soluciones innovadoras que aborden los retos que enfrentan las ciudades. En Bucaramanga, la grave situación de seguridad, marcada por el incremento de delitos y la influencia de redes de microtráfico, pone de manifiesto la necesidad urgente de estrategias más efectivas. En este sentido, la aplicación de algoritmos de clustering, como K-Means, se presenta como una herramienta para la prevención y reducción del crimen. Estas técnicas permiten identificar patrones y delimitar zonas de alta criminalidad, lo que a su vez optimiza el despliegue de recursos y el diseño de estrategias preventivas basadas en datos (A. Alkhaibari Ping-Tsai Chung, 2017).

Si bien métodos tradicionales como los mapas de calor y la estadística descriptiva proporcionan una visión general de la distribución geográfica de los delitos, estos enfoques suelen limitarse a representar la frecuencia y la dispersión de los eventos sin profundizar en las relaciones subyacentes entre múltiples variables. En contraste, el análisis basado en clustering revela patrones ocultos y relaciones complejas, permitiendo segmentar áreas en función de características delictivas y económicas similares, lo que fortalece la toma de decisiones en seguridad pública (Sinaga & Yang, 2020).Otros autores, como (Torres, 2023) destacan el uso de algoritmos jerárquicos para caracterizar la criminalidad y analizar su relación con variables económicas. Por ejemplo, en Londres, la Policía Metropolitana ha empleado K-Means para identificar “puntos calientes” de criminalidad, optimizando la asignación de recursos; de manera similar, en Vancouver, el uso de clustering ha permitido intervenciones policiales más efectivas, y en Chicago, la aplicación de aprendizaje automático ha contribuido a reducir las tasas de delitos en zonas críticas (Ratra et al., 2023).

Además, la integración de estos análisis en un dashboard público añade un valor práctico significativo. Al ofrecer acceso abierto y visualizaciones interactivas, este tipo de herramienta facilita la interpretación de los datos por parte de los tomadores de decisiones, permitiendo ajustar estrategias de seguridad en tiempo real, mejorar la transparencia y fomentar la participación ciudadana. La incorporación de un módulo basado en LLM potencia aún más este sistema al procesar y sintetizar información no estructurada para generar insights en lenguaje natural, lo que permite a los usuarios interactuar de manera intuitiva y obtener explicaciones detalladas que complementan el análisis cuantitativo.

Objetivos

Objetivo General

Evaluar los patrones de comportamiento criminal en Bucaramanga mediante técnicas de Machine Learning no supervisado para la segmentación de comunas según sus características delictivas y demográficas.

Objetivos Específicos

Identificar las variables relevantes de criminalidad, organizando los datos en función de su importancia, para la estructuración de una base de datos óptima para el análisis.

Sintetizar los clústeres para la descripción de zonas con características delictivas similares mediante K-MEANS y Agrupamiento Jerárquico.

Desarrollar una aplicación interactiva con Streamlit que permita la visualización geoespacial de la distribución de delitos y los clústeres identificados en Bucaramanga.

Integrar un módulo basado en modelos de lenguaje de gran tamaño (LLM) que apoye la generación de *insights* y facilite la interpretación de los patrones delictivos por parte de los tomadores de decisiones.

Marco Teórico

El estudio de los patrones delictivos ha evolucionado significativamente a nivel mundial, dando lugar a metodologías avanzadas que permiten identificar y predecir comportamientos criminales con mayor precisión. Si bien los métodos tradicionales—como los mapas de calor y la estadística descriptiva—han sido útiles para representar la distribución de los delitos, técnicas más recientes basadas en el aprendizaje no supervisado, como el clustering, permiten revelar patrones ocultos y relaciones complejas en los datos. En el contexto latinoamericano, y especialmente en Colombia, se han propuesto múltiples enfoques teóricos para comprender la criminalidad. Sin embargo, limitar el análisis a los registros delictivos ignora la complejidad estructural del fenómeno. Estudios recientes, como el realizado por el Departamento Nacional de Planeación y la Universidad Nacional (Pinto, 2022), destacan que el crimen en Colombia debe entenderse como parte de un sistema dinámico, influenciado por variables sociales, económicas, institucionales y demográficas. Esta visión sistémica justifica el uso de técnicas de segmentación territorial que integren variables socioeconómicas, con el fin de comprender y abordar de forma integral la distribución espacial de la criminalidad.

Una forma de abordar comparaciones entre territorios con condiciones estructurales distintas ha sido el uso de métricas ajustadas, como la Razón de Mortalidad Estandarizada (RME), ampliamente utilizada en epidemiología. Este indicador compara el número de muertes observadas con las que se esperarían si la población tuviera la misma estructura etaria que una población estándar, eliminando así el sesgo demográfico (Martínez-González, 1993).

Aunque originalmente aplicada al estudio de la mortalidad, su lógica de estandarización territorial resulta conceptualmente útil como inspiración para construir indicadores análogos que

ajusten la carga delictiva en función de variables estructurales. Esto permite ir más allá de los simples conteos de delitos y considerar el contexto que los produce.

En Bucaramanga, donde el incremento de delitos y la influencia de redes de microtráfico han generado un escenario de creciente inseguridad, resulta fundamental aplicar enfoques innovadores que trasciendan las representaciones descriptivas tradicionales. El aprendizaje no supervisado permite analizar datos sin etiquetas, descubriendo estructuras y relaciones intrínsecas que no se evidencian mediante análisis convencionales (Ezugwu et al., 2022). Por ejemplo, algoritmos como K-Means, reconocidos por su simplicidad y eficiencia (A. Alkhaibari Ping-Tsai Chung, 2017), agrupan los datos en clusters minimizando la distancia entre puntos y sus respectivos centroides. Sin embargo, estos métodos presentan limitaciones, como la necesidad de definir previamente el número de clusters, la sensibilidad a valores atípicos y la interpretación ambigua cuando los clusters tienen formas no esféricas (Huang, 1998).

Además del enfoque particional de K-Means, se ha explorado el clustering jerárquico, el cual organiza los datos en una estructura representada mediante un dendrograma que permite observar relaciones a diferentes niveles. Una ventaja destacada de este método es que no requiere predefinir el número de clusters, lo que le otorga flexibilidad para adaptarse a distintos tipos de datos, aunque su costo computacional puede ser elevado en conjuntos de datos grandes (Ezugwu et al., 2022).

Estudios internacionales han evidenciado el potencial de estas técnicas para la toma de decisiones en seguridad pública. Por ejemplo, en Londres, la Policía Metropolitana ha empleado K-Means para identificar “puntos calientes” de criminalidad, optimizando la asignación de recursos; en Vancouver y Chicago, el uso de clustering ha permitido intervenciones policiales más efectivas y la reducción de tasas delictivas (Ratra et al., 2023).

En Ecuador, investigaciones han destacado la utilidad de algoritmos jerárquicos para relacionar la criminalidad con variables económicas, lo que refuerza la aplicabilidad de estos enfoques en contextos con características particulares de un país (Torres, 2023). En el caso de Colombia, se han empleado técnicas de *clustering* e inteligencia artificial para clasificar y proyectar delitos violentos a nivel departamental, logrando una agrupación territorial basada en factores de violencia y una alta precisión predictiva mediante redes neuronales e inteligencia artificial para clasificar y proyectar delitos violentos a nivel departamental, logrando una agrupación territorial basada en factores de violencia y una alta precisión predictiva mediante redes neuronales (Fontalvo Herrera et al., 2023). Por su parte, en Bucaramanga se ha aplicado el machine learning principalmente con fines de predicción del delito, utilizando modelos espaciales basados en grafos y técnicas como TF-IDF adaptadas al contexto urbano. Aunque los resultados más prometedores se lograron con modelos semanales, como el KNN, que alcanzó un 59 % de recall y más del 60 % de accuracy, la investigación también evidenció limitaciones derivadas de la escasez de datos y la capacidad institucional reducida en ciudades intermedias (Gelvez Ferreira et al., 2022).

En este sentido, explorar enfoques no supervisados como el *clustering* puede representar una alternativa metodológica adecuada, al no requerir etiquetas previas y permitir la identificación de patrones emergentes que orienten estrategias de prevención más focalizadas.

Dentro de este tipo de enfoques, la validación de la calidad de los clústeres constituye un aspecto fundamental. Uno de los indicadores más utilizados es el índice de silueta, una métrica que evalúa simultáneamente la cohesión interna y la separación entre grupos, donde valores cercanos a uno indican una segmentación eficiente (Gao et al., 2023).No obstante, esta evaluación debe complementarse con criterios de interpretabilidad, esenciales para garantizar

que los clústeres identificados sean comprensibles y útiles para la formulación de políticas públicas (Han, 2012).

En conjunto, esta revisión teórica articula las bases conceptuales y metodológicas necesarias para comprender y analizar los patrones delictivos en Bucaramanga, resaltando tanto las potencialidades como las limitaciones de las técnicas de *clustering*. La integración de estos enfoques ofrece un camino prometedor para enriquecer la interpretación de datos complejos y fortalecer los procesos de toma de decisiones en materia de seguridad ciudadana.

Metodología

El proyecto sigue la metodología CRISP-DM por su enfoque estructurado, iterativo y adaptable a diferentes contextos de minería de datos. Consta de las siguientes etapas: Comprensión del negocio, Comprensión de los datos, preparación de los datos, Modelado, Evaluación y Despliegue (Chapman, 2000).

En línea con la metodología y con el objetivo de facilitar la interacción con los modelos desarrollados y promover el uso práctico de los resultados, se desarrolló una aplicación web con la biblioteca Streamlit, desplegada en Streamlit Cloud para acceso público. Esta herramienta permite a los usuarios visualizar los clústeres generados en mapas interactivos, seleccionar el número de clústeres y el tipo de modelo (K-Means o jerárquico), incluir variables contextuales (como población, manzanas o densidad), comparar modelos mediante métricas como Silhouette, Davies-Bouldin e Inercia, y enviar los resultados a un modelo de lenguaje (LLM) para obtener recomendaciones automáticas.

Una vez listadas las etapas metodológicas y descrita la aplicación en Streamlit para facilitar el análisis, se hace necesario situar el estudio dentro del marco de investigaciones previas en clustering criminal. Para ello, se realizó una revisión de estudios relevantes, lo que permitió identificar estrategias y enfoques aplicados en contextos similares. La siguiente tabla resume investigaciones clave, detallando la muestra, las variables explicativas y las técnicas de clustering utilizadas, lo que orientó la selección y aplicación de los métodos en el presente estudio.

Tabla 1

Investigaciones Previas de Clustering Criminal

Referencia	Muestra	Variables	Técnicas
(Fontalvo Herrera et al., 2023)	727k	Amenazas, delitos sexuales, extorsión, homicidio, secuestro y terrorismo.	Clustering: K-Means, Jerárquico.
(Jerin et al., 2023)	319k	Coordenadas, tipo de crimen, fecha, periodos del día.	K-Means, Agglomerative Clustering, DBSCAN.
(Ceballos, 2023)	319kk	Tipo de crimen, ubicación, fecha, nivel de ofensa.	K-Means, K-Modes.
(Sagala & Gunawan, 2022)	179k	Tipo crimen, fecha, ubicación	K-Means
(De Vasconcelos Dos Santos Junior et al., 2023)	18.4k	Macrocausas criminales, factores socioeconómicos y demográficos	K-Means, Agglomerative (Ward y Gower), MDS
(Gouri Jha, 2020)	-	Fraudes serios a la propiedad (India), pérdida económica por estado	K-Means, visualización de clusters, análisis de centroides

Nota. Artículos de clustering criminal con diferentes variables explicativas y modelos.

Si bien la revisión de estudios y la rigurosa aplicación de la metodología adoptada aportan una base sólida al análisis, es importante reconocer que el estudio también enfrenta

ciertos desafíos. En lo que sigue, se presentan brevemente las limitaciones metodológicas que podrían haber influido en los resultados, destacando su relevancia para contextualizar el análisis y orientar investigaciones futuras.

Limitaciones Metodológicas

Al consolidar los datos por comuna se pueden ocultar variaciones y patrones locales (intra-comuna) importantes para identificar áreas específicas de riesgo o particularidades en la incidencia delictiva.

Se utiliza la población proyectada para 2021 y registros hasta 2023, lo cual puede no reflejar cambios demográficos recientes o dinámicas delictivas emergentes, afectando la precisión comparativa en el tiempo.

Al utilizar la población de 2021 para calcular tasas de criminalidad en distintos años, se asume una estabilidad demográfica que puede no reflejar cambios reales, afectando la precisión y comparabilidad de los resultados.

Los modelos de lenguaje son gratuitos con un límite de 100 mil billones de parámetros para procesar y generar textos.

A continuación, se describen brevemente las etapas metodológicas implementadas en este estudio. Se parte del análisis y preparación de los datos, seguido por el modelado mediante técnicas de clustering y la evaluación de sus resultados, para concluir con el despliegue de una aplicación web en Streamlit.

Comprensión de los Datos

Se exploraron y analizaron los datos disponibles para comprender su estructura y calidad. La fuente de datos es Datos Abiertos Colombia, se manejan tres tablas:

Tabla de crímenes en Bucaramanga entre 2016 y 2023, con 101k filas y 26 columnas.

Tabla con la población proyectada para el 2021 en Bucaramanga por comuna. En el análisis se toma la población de este año como referencia.

Tabla con polígonos de las comunas de Bucaramanga, contiene 17 filas. Se obtiene de obtenida desde ArcGIS.

La siguiente tabla presenta las variables utilizadas para el análisis. Cabe aclarar que estas variables ya fueron seleccionadas y transformadas a partir del conjunto de datos original. El proceso metodológico mediante el cual se depuraron y estandarizaron estas variables se explica en la sección de Preparación de los Datos.

Tabla 2

Variables de Interés de los Datos de Crímenes

Variable	Descripción
fecha_hecho	Fecha cuando se generan los hechos del delito según la Policía
hora_hecho	Hora cuando se generan los hechos del delito según la Policía
sexo	Sexo de la víctima del delito según la Policía.
movil_victima	Movil en el que se desplazaba la víctima en el momento del delito según la Policía.
movil_agresor	Movil en el que se desplazaba el agresor en el momento del delito según la Policía.
curso_vida	Etapa de la vida, se personaliza edad según ciclo de vida de minsalud.
num_com	Número de la comuna (17 comunas)

Variable	Descripción
nom_com	Nombre de la comuna según planeación
categoria_delito	Robos y Hurtos, Violencia Familiar, Delitos Sexuales,Crimen Organizado y Delitos Violentos (Se creó a partir de los delitos).
momento_del_dia	Madrugada, Mañana, tarde o noche.
tipo_amenaza	Se categoriza de acuerdo con el arma del delito en: Violencia letal, Violencia no letal, Robo, Incendios o sin empleo de armas.

Nota. La variable categoría delito categoriza todos los delitos.

De la tabla de población se toma:

Tabla 3

Formato Población y Número de Manzanas por Comuna Año 2021

Variable	Descripción
Cod_comuna	Número de la comuna
Personas	Población comuna año 2021
Manzanas	Numero de manzanas comuna año 2021

De la tabla de polígonos se toma:

Tabla 4

Formato Datos de Polígonos Comunas Bucaramanga

Variable	Descripción
Nombre_com	Nombre de la comuna
cod_comuna	Código de la comuna
geometry	Polígono de la comuna para incluir en mapa.
área	Área en km2 de la comuna

Nota. Los datos de geometry son para realizar el mapa.

Adicionalmente se consulta en fuentes de la alcaldía de Bucaramanga para organizar una tabla con descripciones de cada comuna:

Tabla 5

Formato Datos Contextuales Acerca de las Comunas de Bucaramanga

Variable	Descripción
id	Número de la comuna
comuna	Nombre de la comuna
descripción	Contexto de cada comuna. Por ejemplo, si es centro es zona comercial.

Nota. Los datos de las descripciones se recopilan de documentos de la alcaldía.

Preparación de los Datos

Para el análisis se preparó un dataset consolidado a nivel de comuna, transformando la base de datos original para cumplir con los requisitos de los modelos de clustering (K-Means y jerárquico), que requieren variables numéricas.

Se pivotó la tabla utilizando las comunas como índice y se agregaron los datos de criminalidad por tipo de delito y año, calculando el promedio anual por comuna. Este enfoque normaliza la información y permite comparaciones equitativas entre territorios, a pesar de las variaciones en el número de registros históricos.

Una de las principales transformaciones consistió en la agregación de los datos de criminalidad por tipo de delito y año, seguida del cálculo del promedio anual por comuna. Esta transformación permitió normalizar la información y compararla de forma equitativa entre territorios, independientemente del número de registros por año o de posibles variaciones en los datos históricos. Además, este enfoque se justifica dado que los datos de población utilizados para ajustar las tasas corresponden únicamente al año 2021.

Figura 1

Ejemplo Transformación Datos Crímenes en Base a Comunas

Se realizó una depuración y selección manual de variables, evaluando su completitud, correlación y pertinencia. Así, se eligieron aquellas que mejor representan la criminalidad y se excluyeron otras (como fecha, móviles o curso de vida) para evitar la introducción de ruido en los modelos, destinándolas en cambio a filtros en la aplicación en Streamlit.

Además del promedio anual de crímenes por comuna, se incorporaron variables adicionales (número de manzanas, población total, área) y la variable cualitativa "descripción" para aportar contexto sobre las características de cada comuna, facilitando así la interpretación de los patrones delictivos. Esta preparación integral garantiza que los modelos trabajen con datos comparables y contextualizados, fortaleciendo la utilidad práctica de los resultados para la formulación de estrategias de intervención territorial.

Figura 2

Ejemplo de Integración de Otras Variables por Comuna.

comuna	crimen 1	crimen 2	manzanas	personas	area	descripción
A	0,5	1	3	6000	700	zona comercial
B	1	0	4	7000	800	zona alto tráfico

Se generaron variables nuevas, como la tasa de criminalidad por 1,000 habitantes, la densidad delictiva y la Razón de Morbilidad Estándar (RME), calculada con la población de 2021 para ajustar la incidencia delictiva según el tamaño poblacional de cada comuna.

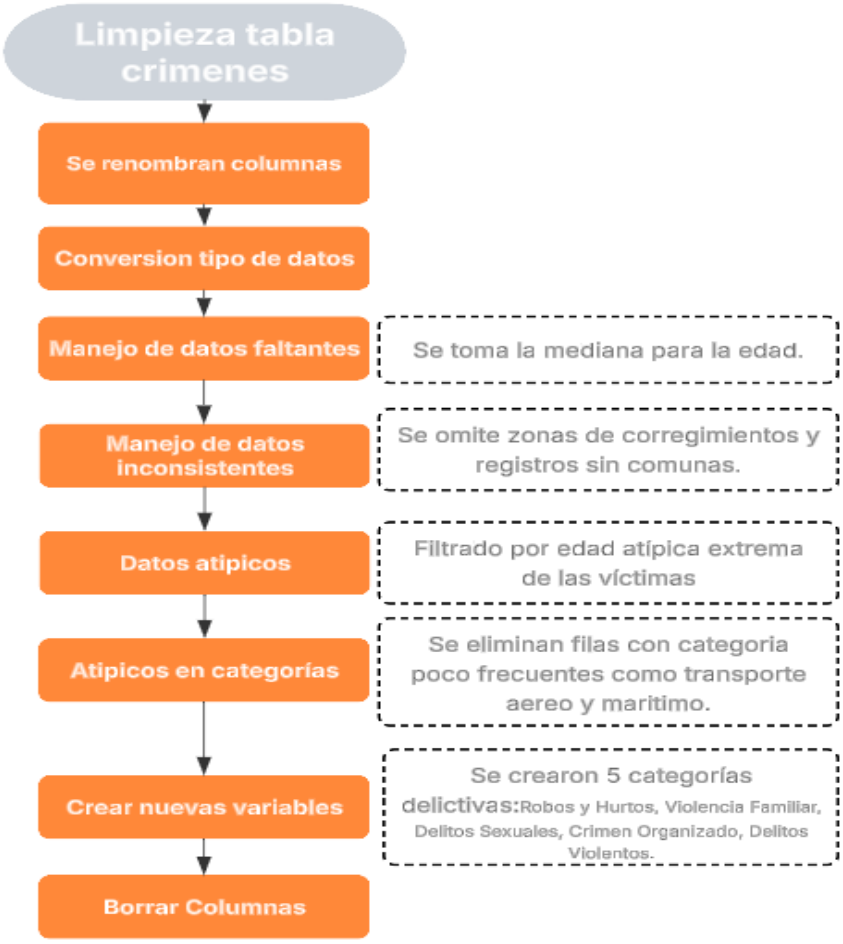
El cálculo de la RME se realizó en tres etapas: (1) se calculó la tasa global por tipo de delito, dividiendo el total de casos registrados entre la población total de la ciudad; (2) se estimaron los casos esperados por comuna, multiplicando la población local por dicha tasa global; y (3) se obtuvo la RME como la razón entre los casos observados y los esperados. Una

RME superior a 1 indica que la comuna presenta una carga delictiva mayor a la esperada, mientras que un valor inferior a 1 refleja una incidencia por debajo de lo estimado.

Para ilustrar de manera general el flujo de limpieza y preparación de la tabla de crímenes, en la siguiente figura se muestran los pasos clave seguidos antes de la fase de análisis. Estos abarcan desde la conversión de tipos de datos, hasta la detección de valores atípicos y la creación de nuevas variables, asegurando así la calidad y consistencia de la información utilizada en etapas posteriores. La figura 3 resumen los pasos que se siguieron para la limpieza de la tabla de hechos o registros de crímenes.

Figura 3

Limpieza Tabla de Crímenes.



Modelado

Se implementaron técnicas de clustering para segmentar las zonas de Bucaramanga, con el objetivo de obtener una comprensión detallada de la distribución espacial. Se utilizó el algoritmo K-Means por su simplicidad y eficiencia en la identificación de grupos en datos estructurados, junto con Agrupamiento Jerárquico para comparar los resultados obtenidos.

El número óptimo de clusters se determinó mediante análisis visual y métricas de rendimiento específicas de cada algoritmo.

Las herramientas empleadas para la implementación de los modelos fueron Python y bibliotecas especializadas como Scikit-learn.

Evaluación

Se utilizaron varias métricas clave para evaluar los resultados del modelo y garantizar su alineación con los objetivos. Inercia, aplicable a K-means, mide la compacidad de los grupos, siendo preferible un valor menor. Silhouette Score, que indica qué tan bien se ajustan los datos a sus clústeres, con valores cercanos a 1 como indicativo de un buen ajuste. Calinski-Harabasz, que evalúa la relación entre la dispersión interna y entre clústeres, con valores más altos siendo mejores. Davies-Bouldin, que mide la similitud entre grupos, donde un valor menor es mejor.

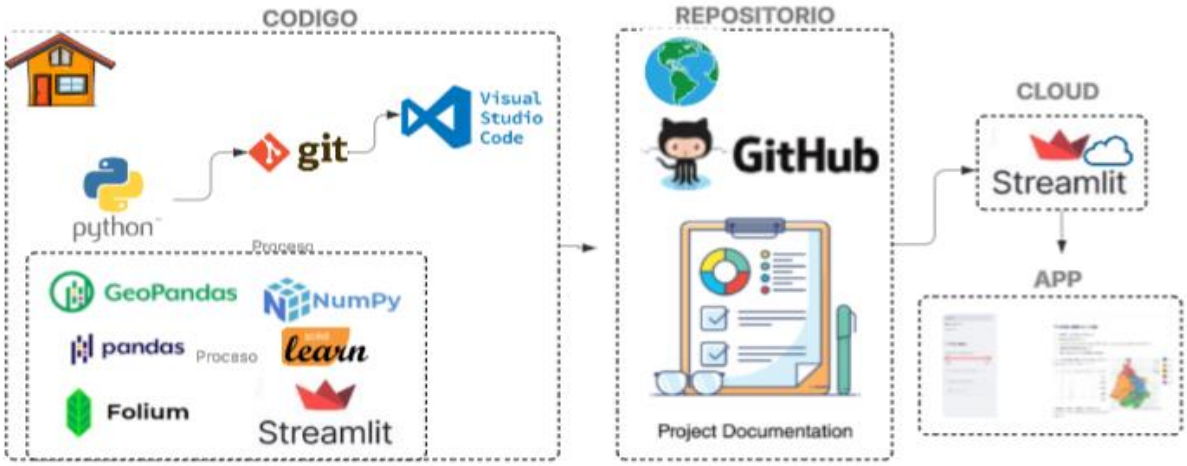
Finalmente, se validaron los clústeres en términos de su utilidad para identificar patrones delictivos, realizando ajustes cuando los resultados no fueron significativos.

Despliegue

Para el desarrollo del proyecto se codifica en Python llevando un control de versión de manera local con Git. El proyecto se empuja a un repositorio en Github para salvar el control y la documentación. Después se configura la cuenta de Streamlit cloud y se integra al repositorio para desplegar finalmente la App. La figura 4 muestra un resumen del proceso.

Figura 4

Esquema App



Aplicación en Streamlit

Streamlit es una plataforma de desarrollo web de código abierto, diseñada especialmente para científicos de datos, que permite transformar scripts de Python en aplicaciones interactivas de forma rápida y sencilla. Gracias a esta herramienta, los analistas pueden crear interfaces dinámicas sin necesidad de conocimientos avanzados en desarrollo web, superando las limitaciones de las visualizaciones estáticas y documentos tradicionales (Richards, 2021, p. 4)

En el contexto del presente estudio, Streamlit se utiliza para visualizar la distribución de los crímenes registrados en la ciudad, permitiendo representar espacialmente los resultados del agrupamiento por comunas. La plataforma facilita la interacción con los datos mediante filtros, mapas y modelos ajustables, lo que posibilita un análisis exploratorio más profundo de los patrones delictivos. Esta interactividad no solo mejora la comprensión de las tendencias identificadas, sino que también apoya la toma de decisiones con visualizaciones claras y salidas interpretativas generadas por modelos de lenguaje natural.

La aplicación desarrollada se publicó en Streamlit Cloud, garantizando su acceso público desde cualquier dispositivo con conexión a internet. Esto permite que autoridades, analistas y otros actores interesados consulten y utilicen en tiempo real los resultados del análisis, sin requerir conocimientos técnicos o instalaciones adicionales.

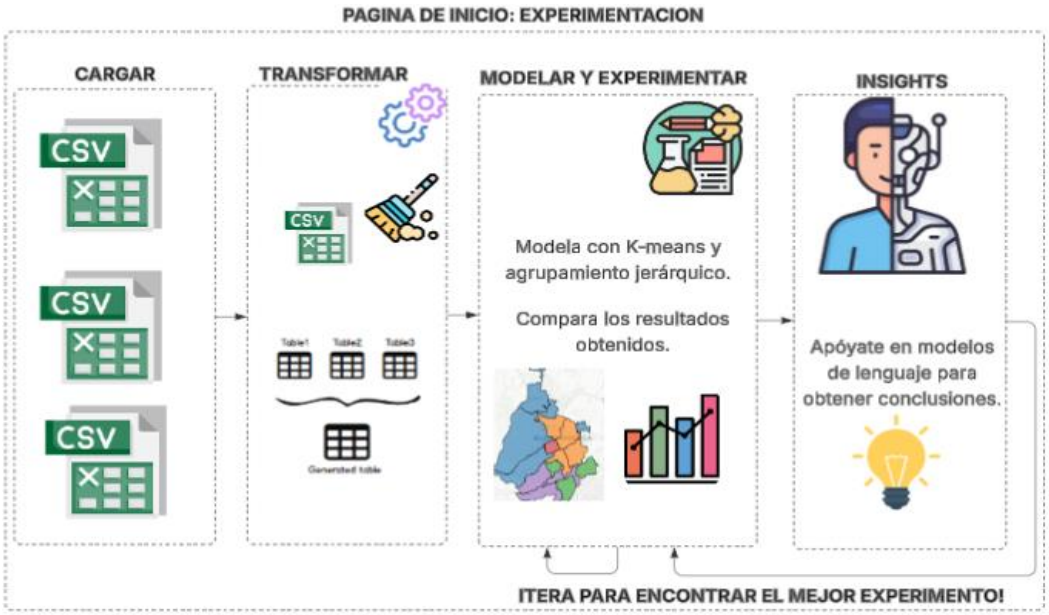
Diseño e Implementación de la Aplicación

La aplicación se estructura en tres secciones: Experimentos, Ajustes LLM y Ayuda, cada una orientada a explorar y analizar los patrones espaciales de criminalidad en Bucaramanga.

Las funcionalidades básicas de la página de experimentación las resume el grafico 5.

Figura 5

Flujo Aplicación Streamlit



Por otro lado, la personalización del LLM se resume en la figura 6.

Figura 6

Ajustes LLM en App

En la página de ajustes LLM se puede cambiar el modelo de lenguaje, el prompt y las descripciones contextuales de las comunas.

En la página de ayuda hay una guía rápida de cómo utilizar la aplicación y cuál es su alcance.

Modelos en Aplicación

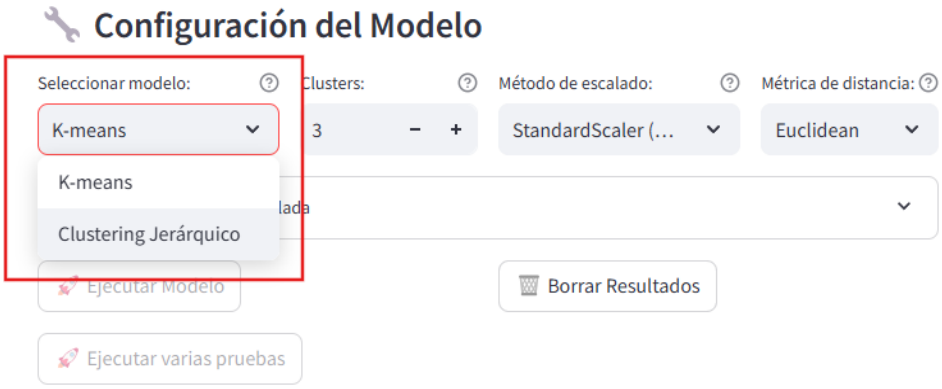
Una vez preparado el conjunto de datos y habilitadas las funcionalidades interactivas en la aplicación, se procedió a la implementación de los modelos de agrupamiento.

La interfaz de la aplicación permite al usuario personalizar la configuración del modelo, seleccionando el algoritmo de agrupamiento, el número de clústeres, el método de escalado de variables y la métrica de distancia para calcular similitudes entre comunas. Esta configuración

puede utilizarse para realizar análisis individuales. No obstante, al presionar el botón “Ejecutar varias pruebas”, se generan todas las combinaciones posibles entre los filtros seleccionados.

Figura 7

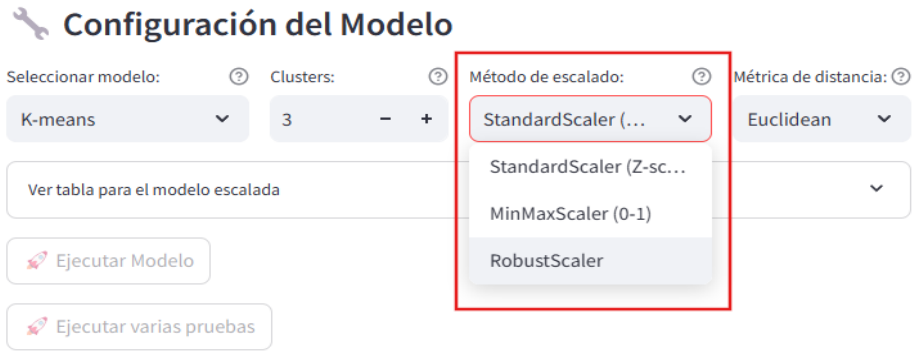
Configuración del Modelo, Selección de Modelos



Se pueden utilizar distintos métodos de escalado. No obstante, al emplear la tasa RME para medir los crímenes, el escalado se limita únicamente a las variables no delictivas, como el número de manzanas, el área, entre otras.

Figura 8

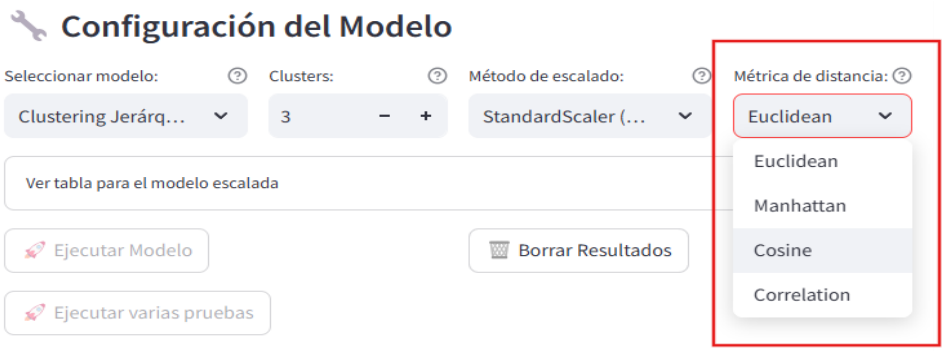
Configuración del Modelo, Selección de Método de Escalado



Además, cuando se selecciona el modelo de clustering jerárquico, la interfaz permite elegir la métrica de distancia a utilizar para calcular la similitud entre comunas.

Figura 9

Configuración del Modelo, Selección de Método de Escalado



Evaluación en Aplicación

Una vez ejecutados los modelos de agrupamiento bajo distintas configuraciones, fue necesario comparar su rendimiento para identificar aquellas combinaciones que generan clústeres más coherentes y útiles para el análisis. Para ello, se implementó un panel de evaluación comparativa dentro de la aplicación, el cual permite visualizar y ordenar los resultados de cada prueba de manera estructurada. La evaluación se basó en las cuatro métricas: Inercia, Silhouette, Calinski-Harabasz y Davies-Bouldin.

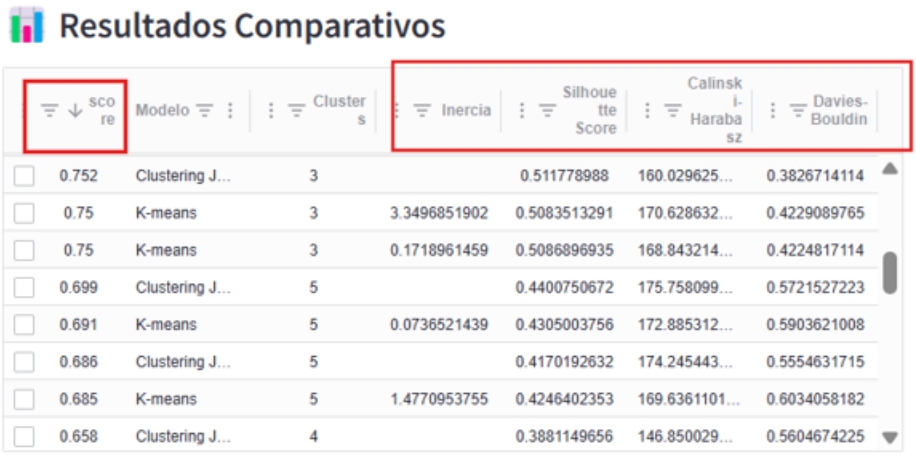
Adicionalmente, se incluyó una columna denominada score, que resume el desempeño global de cada modelo. Esta puntuación fue calculada combinando las tres métricas comparables entre todos los modelos (Silhouette, Calinski-Harabasz y Davies-Bouldin), luego de normalizarlas y asignarles un peso relativo. Las métricas donde un valor más alto es mejor se estandarizaron directamente, mientras que aquellas donde un valor más bajo es deseable se transformaron para que mantuvieran la misma lógica interpretativa. Finalmente, se calculó un

promedio ponderado con mayor peso asignado al Silhouette Score (0.5), seguido de Calinski-Harabasz (0.3) y Davies-Bouldin (0.2), generando un puntaje único que permite ordenar fácilmente los resultados.

La siguiente figura muestra un fragmento de la tabla resultante, ordenada de mayor a menor según el score, lo cual facilita la identificación de las configuraciones con mejor desempeño relativo:

Figura 10

Resultados de Varios Experimentos en App



score	Modelo	Clusters	Inercia	Silhouette Score	Calinski-Harabasz	Davies-Bouldin
0.752	Clustering J...	3		0.511778988	160.029625...	0.3826714114
0.75	K-means	3	3.3496851902	0.5083513291	170.628632...	0.4229089765
0.75	K-means	3	0.1718961459	0.5086896935	168.843214...	0.4224817114
0.699	Clustering J...	5		0.4400750672	175.758099...	0.5721527223
0.691	K-means	5	0.0736521439	0.4305003756	172.885312...	0.5903621008
0.686	Clustering J...	5		0.4170192632	174.245443...	0.5554631715
0.685	K-means	5	1.4770953755	0.4246402353	169.6361101...	0.6034058182
0.658	Clustering J...	4		0.3881149656	146.850029...	0.5604674225

Experimentación en Aplicación

En esta sección se presentan los resultados obtenidos a partir de la aplicación desarrollada. La herramienta permite ajustar parámetros del modelo, explorar diferentes combinaciones de variables y visualizar la segmentación espacial de las comunas en función de sus características delictivas y contextuales. El objetivo principal de esta fase es evaluar el impacto de distintas configuraciones del modelo, como el número de clústeres, el tipo de variables utilizadas, el algoritmo de agrupamiento seleccionado y el método de escalado aplicado.

Las pruebas realizadas permiten evaluar la estabilidad de los resultados, su claridad operativa y su utilidad para la toma de decisiones territoriales. La herramienta desarrollada facilita la selección de una configuración específica desde la tabla comparativa de evaluación, para explorar sus resultados en mayor detalle.

Al seleccionar un modelo, se muestran tres elementos clave: una tabla con la proporción de comunas por clúster, un mapa interactivo con la distribución espacial de los grupos y, según el modelo, un mapa de calor con los centroides (para K-means) o un dendograma (para clustering jerárquico). Estos recursos permiten analizar la segmentación de manera visual y comprensible. Para contextualizar las funcionalidades de la aplicación se describe a continuación dos casos.

Experimento 1: Clustering con Modelo Jerárquico y Escala RME

Se llevan a cabo diversas combinaciones de filtros y se selecciona un modelo de clustering jerárquico que muestra métricas aceptables.

Figura 11

Experimentos Modelo Jerárquico

score	Modelo	Clusters	Inercia	Silhouette Score	Calinski-Harabasz	Davies-Bouldin
0.904	Clustering Je…	4		0.3904445267	862.4885957…	0.6123528653
0.904	Clustering Je…	4		0.3904445267	862.4885957…	0.6123528653
0.897	K-means	3	14.62656500…	0.3793968862	865.5057149…	0.6200784584
0.86	Clustering Je…	5		0.3475781835	810.1599171…	0.5956190429
0.823	Clustering Je…	3		0.3375635885	729.1540702…	0.6338505911
0.82	Clustering Je…	3		0.3257417597	750.6603833…	0.6750696042
0.812	Clustering Je…	5		0.3244953523	719.8554634…	0.6351086122
0.795	K-means	4	10.365047411	0.2890828365	757.8569193…	0.751645618

index	score	Modelo	Clusters	Inercia	Silhouette Score	Calinski-Harabasz	Davies-Bouldin
41	0.86	Clustering Jerárquico	5	None	0.3476	810.1599	0.5956

El modelo de clustering jerárquico muestras métricas que son consideradas aceptables para la evaluación de la calidad de la segmentación. Con un Silhouette Score de 0.34, que indica un buen ajuste del modelo a los datos, y un Calinski-Harabasz Index de 810.15, que refleja una separación adecuada entre los clusters, se concluye que el modelo ofrece resultados razonables. Además, el Davies-Bouldin Index de 0.59 es lo suficientemente bajo como para sugerir que los clusters son distintos y bien definidos.

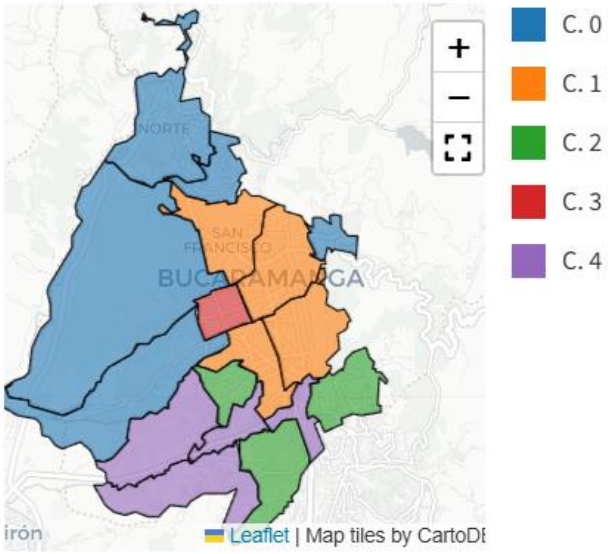
Figura 12

Clusters Modelo Jerárquico

Esta tabla muestra el tamaño de cada clúster y su peso porcentual dentro del total.

	Cluster	Tamaño del Clúster	Peso (%)
0	0	5	29.41%
1	1	4	23.53%
2	2	3	17.65%
3	3	1	5.88%
4	4	4	23.53%

La proporción de los clusters indica si la segmentación es equilibrada y su impacto en el análisis.



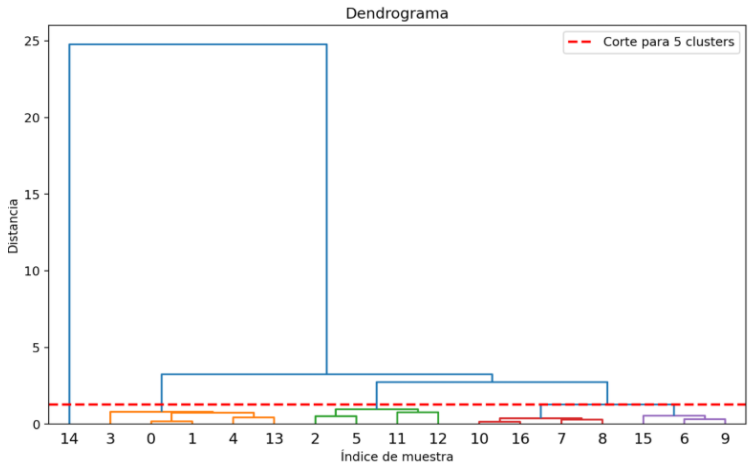
La distribución de los clusters muestra que la segmentación es relativamente equilibrada, con un tamaño de cluster que varía entre 1 y 5, y los porcentajes de peso de cada uno dentro del total. El cluster más grande representa el 29.41% del total, mientras que el más pequeño tiene un peso del 5.88%. Esto sugiere que, aunque hay una ligera variabilidad, la distribución no está

completamente sesgada hacia un solo grupo, lo que indica que el modelo ha logrado una división adecuada en términos de tamaño relativo de los clusters.

Al analizar el dendrograma generado por el modelo de agrupamiento jerárquico, se evidencia la presencia de una comuna con características significativamente distintas en comparación con las demás. Esta comuna, correspondiente al centro de la ciudad, presenta un perfil marcadamente comercial, lo cual explica su distancia considerable en el esquema de agrupamiento. Esta separación sugiere que las dinámicas delictivas en dicha zona podrían estar influenciadas por factores asociados al comercio, la afluencia poblacional y la actividad económica.

Figura 13

Dendograma

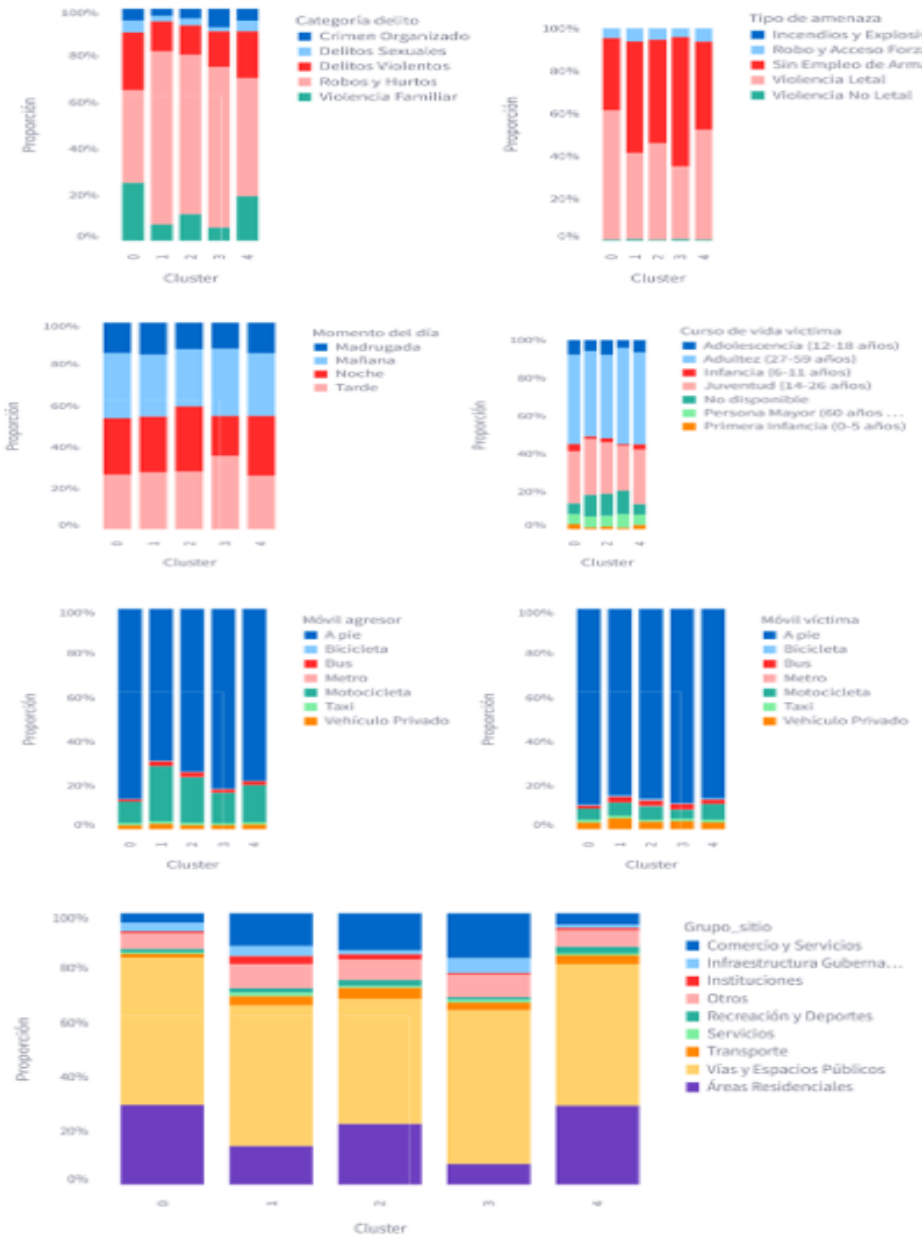


A continuación, se presentan las gráficas de barras apiladas que permiten caracterizar cada clúster a partir de distintas variables relevantes en el análisis criminológico. Estas visualizaciones muestran la distribución proporcional de cada categoría dentro de los clústeres formados, lo cual facilita la comparación entre ellos en términos de tipo de delito, modalidad de agresión y victimización, momento del día, ciclo de vida de la víctima, tipo de amenaza y

ubicación geográfica del hecho. Este enfoque gráfico resulta fundamental para interpretar los perfiles delictivos predominantes en cada conglomerado y orientar recomendaciones diferenciadas según el contexto.

Figura 14

Caracterización de Clusters



El clúster 0, denominado "Familia Vulnerable", agrupa a las comunas Norte, Nororiental, Occidental, García Rovira y Morrorico. Este grupo se caracteriza por presentar la proporción más alta de casos de violencia intrafamiliar, así como una alta incidencia de crímenes con violencia letal. También concentra la mayor proporción de víctimas en primera infancia (entre 0 y 11 años), lo que evidencia un patrón de vulnerabilidad social crítica. En contraste, los robos y hurtos tienen una presencia significativamente baja. Otro rasgo distintivo es la movilidad: las víctimas y agresores en este clúster se desplazan mayoritariamente a pie, mientras que el uso de motocicleta o transporte público por parte de los agresores es mínimo. La mayoría de los delitos ocurren en residencias, y son poco frecuentes en vehículos de transporte o en zonas comerciales.

El clúster 1, identificado como "Hurto Motorizado", incluye las comunas Oriental, Cabecera del Llano, La Concordia y San Francisco. Su característica principal es la alta proporción de robos y hurtos, con un patrón claro de uso de motocicleta por parte de los agresores. A su vez, registra la proporción más baja de víctimas a pie y una incidencia reducida de robos en zonas residenciales, lo cual sugiere que estos delitos tienden a ocurrir en espacios públicos o de tránsito.

El clúster 2, denominado "Juventud Nocturna", está conformado por Ciudadela Real, Tejar y Provenza. En este grupo se observan las mayores proporciones de víctimas adolescentes, así como una concentración significativa de incidentes ocurridos durante la tarde y la noche. Además, se presenta una alta incidencia de delitos cometidos en medios de transporte, así como en establecimientos comerciales y de servicios, lo cual sugiere un perfil vinculado a movilidad y vida urbana activa.

El clúster 3, clasificado como "Red Organizada", corresponde exclusivamente al Centro de la ciudad. Este grupo presenta la mayor proporción de crímenes organizados y de robos o

hurtos cometidos sin uso de armas. Los delitos se concentran casi exclusivamente en la tarde, con una baja ocurrencia durante la noche. Se destaca, además, por la alta proporción de incidentes en zonas comerciales y edificios gubernamentales. En contraste, este clúster muestra las proporciones más bajas de violencia intrafamiliar y delitos sexuales.

El clúster 4, llamado "Residencial Activo", incluye las comunas Mutis, Suroccidente, Sur y Pedregosa. Presenta la segunda proporción más alta de violencia intrafamiliar y delitos sexuales. Los delitos tienden a concentrarse en zonas residenciales y en espacios de recreación o actividad deportiva. También se observa una elevada incidencia de hechos delictivos en medios de transporte, mientras que la ocurrencia en comercios es baja. Los incidentes están distribuidos de manera equilibrada a lo largo del día, sin picos horarios definidos.

Entre todos los grupos, el clúster "Familia Vulnerable" (clúster 0) destaca como el más crítico y debe ser el foco prioritario en términos de intervención pública. En primer lugar, impacta de forma directa a la población infantil, registrando la mayor proporción de víctimas en la primera infancia, lo que representa una alarma grave en términos de protección de derechos. En segundo lugar, la naturaleza de la violencia —principalmente intrafamiliar— genera secuelas psicológicas profundas y sostenidas, que requieren intervenciones inmediatas de tipo sociofamiliar. Por último, el alcance comunitario de esta problemática es especialmente preocupante, ya que tiende a reproducirse y mantenerse en el tiempo si no se implementan programas de protección y apoyo a las víctimas.

Por estas razones, el clúster identificado como “Familia Vulnerable” debe recibir prioridad en la asignación de recursos institucionales y en el diseño de políticas públicas orientadas a la prevención, atención psicosocial y fortalecimiento del entorno familiar en las comunas afectadas.

Experimento 2: Clustering con K-Means y Escala REM

Aunque el agrupamiento jerárquico ha dado buenos resultados, se prueba un segundo experimento utilizando K-means.

Figura 15

Resumen Resultados K-Means

index	score	Modelo	Clusters	Inercia	Silhouette Score	Calinski-Harabasz	Davies-Bouldin
7	0.691	K-means	5	0.0737	0.4305	172.8853	0.5904

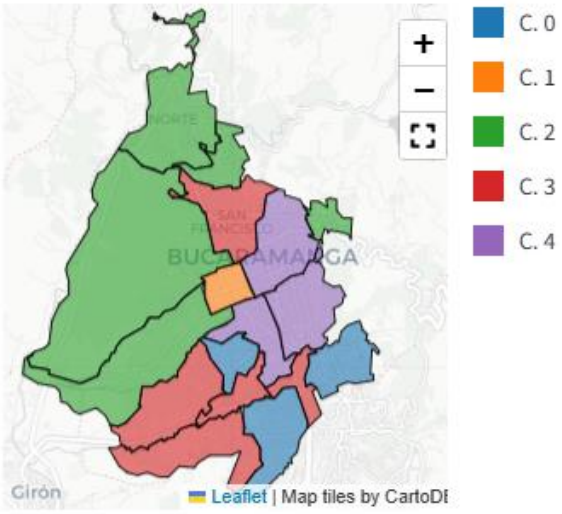
Figura 16

Clusters K-Means

Esta tabla muestra el tamaño de cada clúster y su peso porcentual dentro del total.

	Cluster	Tamaño del Clúster	Peso (%)
0	0	3	17.65%
1	1	1	5.88%
2	2	5	29.41%
3	3	5	29.41%
4	4	3	17.65%

La proporción de los clusters indica si la segmentación es equilibrada y su impacto en el análisis.

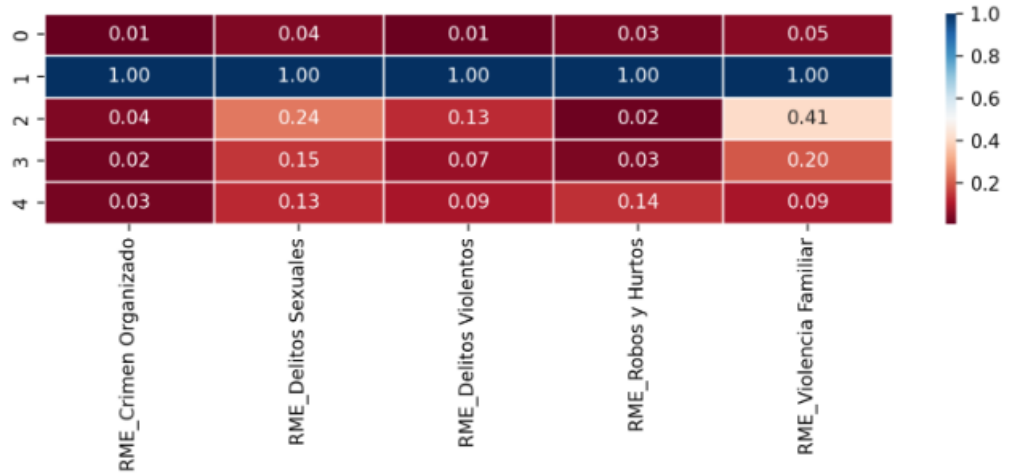


La tabla de distribución muestra que el modelo generó cinco clústeres con una segmentación relativamente equilibrada. El Cluster 2 y el Cluster 3 representan los grupos más numerosos, cada uno con un peso del 29.41%, lo que indica que casi el 60% de las comunas presentan patrones delictivos similares a los de estos dos grupos. Por su parte, el Cluster 1 se

destaca por ser el menos representado, con solo el 5.88% del total, lo cual sugiere un comportamiento atípico o altamente especializado.

Figura 17

Mapa de Calor Centroides



La matriz de calor construida a partir de los resultados del modelo permite visualizar la distribución relativa de cinco tipos de delito en cada uno de los clústeres identificados. Los valores han sido normalizados en un rango entre 0 y 1, lo que facilita la comparación de la intensidad relativa de cada dimensión delictiva dentro de cada grupo. Esta representación gráfica permite identificar perfiles delictivos diferenciados por zona, apoyando la interpretación cualitativa de los resultados del clustering.

El clúster 0, correspondiente a las comunas Ciudadela Real, Tejar y Provenza, representa zonas de baja incidencia delictiva. En este grupo, todos los delitos presentan proporciones iguales o inferiores a 0.05, lo que sugiere un entorno con presencia delictiva moderada o dispersa. Este perfil puede estar asociado a condiciones urbanas más controladas o a menor exposición a factores de riesgo.

En contraste, el clúster 1, conformado únicamente por la comuna Centro, presenta la máxima proporción (1.00) en todos los tipos de delito analizados. Este resultado revela una alta complejidad delictiva y acumulación de problemáticas, posicionando a esta comuna como el principal foco multidelictivo crítico de la ciudad. La concentración de delitos puede estar asociada a una alta densidad poblacional, flujo comercial constante y deficiencias estructurales en control urbano.

El clúster 2 agrupa a las comunas Norte, Nororiental, Occidental, García Rovira y Morrorico. Este grupo muestra valores moderadamente altos en delitos sexuales (0.24) y violencia familiar (0.41), lo que sugiere la presencia de dinámicas de violencia interpersonal con énfasis en problemáticas de género o condiciones de vulnerabilidad social. Este patrón requiere atención especial por parte de las autoridades, dado su impacto en la integridad física y emocional de las víctimas.

El clúster 3, denominado "Conflictos interpersonales esporádicos", incluye las comunas San Francisco, Mutis, Suroccidente, Sur y Pedregosa. Este grupo presenta un perfil bajo en términos absolutos, aunque registra máximos relativos en delitos violentos (0.15) y violencia familiar (0.20). Si bien las cifras no son alarmantes, podrían indicar focos localizados de tensión interpersonal que ameritan monitoreo preventivo.

El clúster 4, compuesto por las comunas Oriental, Cabecera del Llano y La Concordia, muestra un perfil general bajo, pero con valores destacados en robos y hurtos (0.14) y delitos sexuales (0.13), solo superados por el clúster del centro. Esta configuración sugiere la presencia de delitos de oportunidad o conductas contra la integridad sexual en contextos urbanos con menor vigilancia o infraestructura preventiva limitada.

Finalmente, se observa una alta similitud entre las agrupaciones generadas por los algoritmos K-means y jerárquico, lo cual refuerza la robustez de los modelos aplicados. La consistencia entre ambas segmentaciones indica que los patrones subyacentes de criminalidad han sido capturados de forma estable y confiable, lo que valida el enfoque metodológico adoptado. Esta coherencia fortalece la utilidad de los resultados como insumo para el diseño de estrategias diferenciadas de prevención y control del delito en Bucaramanga.

Análisis Asistido con Modelos de Lenguaje.

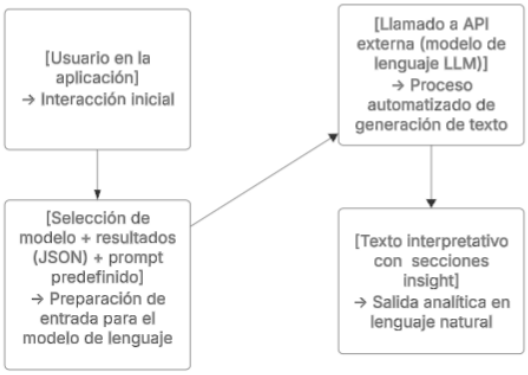
Como complemento a la evaluación cuantitativa de los modelos de agrupamiento, se integró una herramienta de asistencia basada en modelos de lenguaje (LLM), con el objetivo de generar observaciones e interpretaciones automáticas a partir de los resultados obtenidos. Esta funcionalidad permitió traducir salidas numéricas y estructuras de agrupamiento en posibles recomendaciones o descripciones cualitativas, facilitando su análisis por parte de usuarios no técnicos.

El modelo fue alimentado con la información de los clústeres generados, incluyendo estadísticas por grupo, variables utilizadas, y descripciones contextuales de las comunas. A partir de esta entrada estructurada, el LLM generó respuestas en lenguaje natural con posibles interpretaciones, comparaciones entre grupos y sugerencias preliminares de intervención.

Este enfoque permitió enriquecer la evaluación tradicional al incorporar una capa interpretativa basada en IA, alineada con los objetivos del proyecto de facilitar la toma de decisiones y comunicar los resultados de forma accesible.

Figura 18

Pasos App para Ejecutar API LLM



Para generar interpretaciones automáticas a partir de los resultados del análisis, se estructuró un mensaje de entrada al modelo de lenguaje (LLM) compuesto por dos elementos principales:

Prompt predefinido: Un texto que instruye al modelo a actuar como un experto en criminología, orientado a producir recomendaciones claras y operativas. Este prompt guía la respuesta siguiendo cinco secciones: resumen general, análisis por clúster, recomendaciones estratégicas, acciones prioritarias y plan táctico por zona.

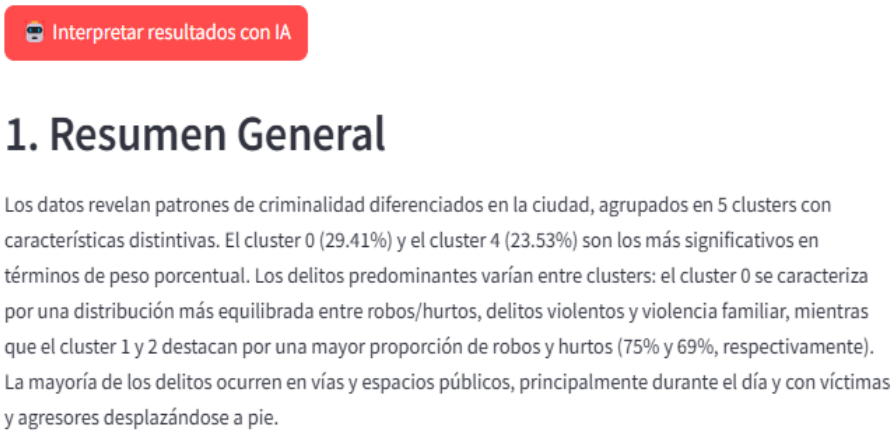
JSON estructurado: Se generó un objeto JSON estructurado que resume los resultados del modelo de agrupamiento, incluyendo el tipo de algoritmo, el número de clústeres y el método de escalado aplicado. También se integran las métricas de evaluación (Silhouette, Davies-Bouldin y Calinski-Harabasz), el tamaño y proporción de cada clúster, y la distribución de comunas por grupo con sus valores estandarizados de criminalidad, densidad poblacional, número de manzanas y descripciones cualitativas.

Esta combinación permite al LLM generar un texto interpretativo en lenguaje natural, estructurado y útil para usuarios no técnicos, con enfoque práctico en seguridad territorial.

A continuación, un ejemplo de la respuesta del LLM, considerando el resumen.

Figura 19

Ejecución LLM y Fragmento de Texto que Genera



En resumen, el uso de un modelo de lenguaje (LLM) permitió transformar resultados técnicos complejos en interpretaciones claras, estructuradas y accionables. Esta herramienta no solo facilitó la comprensión de los patrones delictivos identificados por los modelos de agrupamiento, sino que además aportó recomendaciones tácticas realistas, adaptadas a contextos de recursos limitados. Su integración en la aplicación fortalece la utilidad operativa del análisis, ofreciendo una capa adicional de valor para la planificación estratégica y la toma de decisiones en seguridad pública.

Resultados

A continuación, se presentan los resultados obtenidos a partir de la aplicación de dos técnicas de agrupamiento no supervisado: clustering jerárquico y K-means, cada una configurada para segmentar las comunas de Bucaramanga en cinco clústeres. Ambos modelos permitieron identificar grupos territoriales con características delictivas diferenciadas, aportando perspectivas complementarias sobre la distribución espacial de los delitos.

Tabla 6

Resultados Agrupamiento Jerárquico

Cluster	Nombre	Comunas
0	Familia Vulnerable	Norte, Nororiental, Occidental, García Rovira, Morrorico.
1	Hurto Motorizado	Oriental, Cabecera del Llano, Concordia, San Francisco
2	Juventud Nocturna	Ciudadela Real, Tejar, Provenza
3	Red Organizada	Centro
4	Residencial Activo	Mutis, Suroccidente, Sur, Pedregosa

Tabla 7

Resultados K-Means

Clusters	Nombre	Comunas
0	Robos y Hurtos Moderados	Ciudadela Real, Tejar, Provenza
1	Crimen Organizado y Violencia Extrema	Centro
2	Violencia Familiar y Delitos Violentos	Norte, Nororiental, Occieental, Garcia Rovira, Morrorico
3	Robos y Hurtos con Menor Violencia	San Francisco, Mutis, Suroccidente, Sur, Pedregosa
4	Zonas Comerciales con Robos	Oriental, Cabecera del Llano, La Concordia

Análisis Comparativo de Resultados

Una observación destacada es la coincidencia entre ambos modelos al identificar el Centro de Bucaramanga como una zona crítica, caracterizada por la concentración de crimen

organizado. En el modelo jerárquico se clasifica como Red Organizada, mientras que en K-means se le denomina Crimen Organizado y Violencia Extrema, reafirmando su condición de foco prioritario dentro del mapa delictivo urbano.

Del mismo modo, las comunas del norte, nororiente y occidente fueron agrupadas por ambos modelos como zonas asociadas a violencia familiar y delitos interpersonales. Estas comunas, etiquetadas como Familia Vulnerable en el modelo jerárquico y Violencia Familiar y Delitos Violentos en K-means, comparten condiciones estructurales que las posicionan como áreas de atención social prioritaria.

Por otra parte, Ciudadela Real, Tejar y Provenza fueron clasificadas como zonas con una dinámica delictiva de menor intensidad, principalmente asociada a hurtos no violentos. Mientras que el modelo jerárquico las denomina *Juventud Nocturna*, K-means las agrupa bajo la categoría *Robos y Hurtos Moderados*, lo cual podría reflejar patrones asociados a población joven y alta circulación peatonal o comercial.

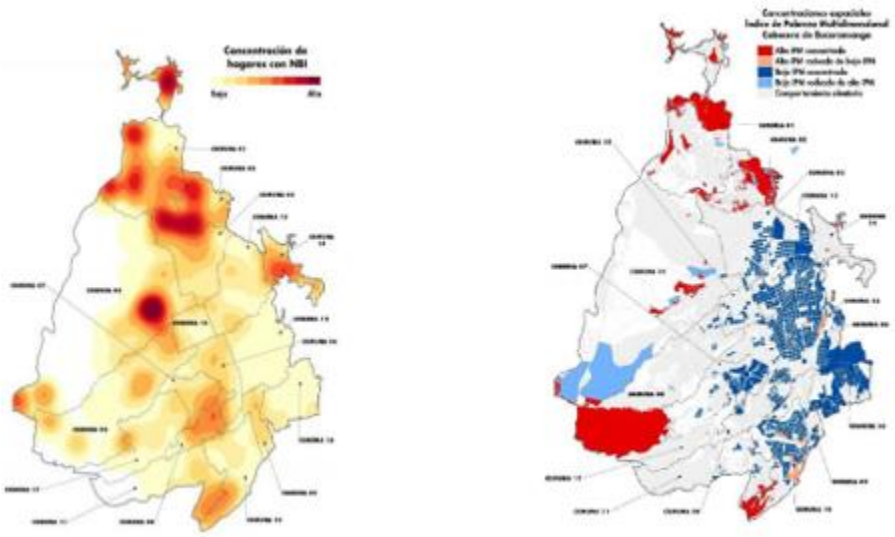
Las zonas comerciales, como Oriental, Cabecera del Llano y La Concordia, fueron identificadas en ambos modelos como focos de delitos patrimoniales. El modelo jerárquico las clasifica como Hurto Motorizado, mientras que K-means las ubica en el clúster Zonas Comerciales con Robos, evidenciando una tipología delictiva vinculada a la actividad económica y al tránsito vehicular.

Por último, Mutis, Suroccidente, Sur y Pedregosa fueron agrupadas como comunas de menor intensidad delictiva. En el modelo jerárquico aparecen como Residencial Activo y en K-means como Robos y Hurtos con Menor Violencia, lo que sugiere un entorno más controlado o menos conflictivo desde el punto de vista criminal.

Un aspecto relevante adicional surge al relacionar los resultados con el índice de Necesidades Básicas Insatisfechas (NBI). Según el Plan de Desarrollo de Bucaramanga 2020, las comunas 1, 2, 14, 4 y 5, así como el sur de las comunas 8 y 10, representan zonas prioritarias para la intervención social. Esta superposición con clústeres de alta conflictividad refuerza el vínculo entre pobreza y criminalidad, y evidencia la necesidad de políticas públicas integrales que aborden ambos fenómenos de manera conjunta, ver figura 20.

Figura 20

Concentración de Hogares con NBI



Nota. DANE 2020

Al superponer los resultados del clúster *Familia Vulnerable* con los mapas de NBI e IPM del DANE, se evidencia una fuerte coincidencia espacial. Esta convergencia respalda el vínculo entre condiciones socioeconómicas precarias y presencia de patrones delictivos, y resalta la necesidad de intervenciones sociales integrales en estas zonas (ver Figura 20).

Conclusiones

Se consolidó una base de datos robusta para el análisis de agrupamiento, priorizando variables representativas de la dinámica criminal en Bucaramanga. Entre las dimensiones seleccionadas se incluyeron tipo de delito, edad de las víctimas, uso de armas, momento del día, ubicación del hecho y medios de movilidad del agresor y la víctima. Esta estructuración permitió capturar de manera adecuada las características relevantes del fenómeno delictivo.

La incorporación de tasas ajustadas por población, mediante la Razón de Morbilidad Estandarizada (RME), permitió mitigar los sesgos asociados al tamaño poblacional de cada comuna. Esta transformación fortaleció la validez comparativa del modelo y garantizó una representación más precisa de la carga delictiva relativa en el territorio.

A partir del procesamiento de estos datos, se identificaron cinco clústeres con características delictivas diferenciadas, tanto mediante el modelo jerárquico como con el algoritmo K-means. La coincidencia general entre ambos métodos confirma la utilidad del *clustering* no supervisado para detectar patrones espaciales de criminalidad en contextos urbanos complejos.

Los clústeres presentaron una alta coherencia temática entre ambos algoritmos, lo que refuerza la consistencia de los resultados obtenidos. Aunque se observaron ligeras diferencias en la asignación de ciertas comunas —como el caso particular de San Francisco— la estructura general de los grupos se mantuvo estable, lo que evidencia la robustez del enfoque metodológico empleado.

Se definieron escenarios delictivos claramente diferenciados por tipo y densidad de crimen. Entre ellos se identificaron: (i) crimen organizado y violencia extrema en el Centro; (ii) violencia estructural y familiar en comunas periféricas del norte y occidente; (iii) robos y hurtos

moderados en zonas con alta movilidad juvenil; (iv) delitos patrimoniales en sectores comerciales como Cabecera del Llano, Oriental y La Concordia; y (v) zonas residenciales con delitos menos violentos, como Mutis, Suroccidente y Sur.

Como producto aplicado del análisis, se desarrolló una aplicación interactiva en Streamlit que permite visualizar la segmentación espacial, explorar los clústeres obtenidos y acceder a interpretaciones automáticas mediante un módulo basado en lenguaje natural (LLM). Esta herramienta facilita la apropiación de los resultados por parte de actores institucionales y usuarios no técnicos, promoviendo el uso práctico de los hallazgos.

Finalmente, la diferencia observada en la asignación de San Francisco entre ambos modelos pone de manifiesto la existencia de zonas con perfiles mixtos o transicionales. Este hallazgo subraya la importancia de complementar el análisis cuantitativo con conocimiento contextual y criterios cualitativos para lograr una lectura más integral y estratégica del territorio.

Discusión

Una línea de investigación prometedora consiste en la aplicación de modelos supervisados utilizando los clústeres como etiquetas. A partir de los agrupamientos obtenidos, es posible entrenar algoritmos de clasificación como Random Forest, Support Vector Machines (SVM) o redes neuronales, con el objetivo de predecir la pertenencia de nuevas zonas o comunas a un perfil delictivo específico, según sus características actuales. Esta estrategia permitiría convertir los resultados del análisis exploratorio en un sistema predictivo aplicable a contextos similares o en evolución.

Asimismo, se sugiere avanzar en un análisis temporal multiclúster, incorporando series de tiempo que permitan observar la estabilidad o variabilidad de los patrones delictivos a lo largo de los años. Esta perspectiva facilitaría la identificación de comunas que cambian de perfil, así como el reconocimiento temprano de zonas en transición hacia escenarios de mayor o menor riesgo delictivo. De este modo, sería posible anticipar cambios críticos y orientar acciones preventivas con mayor oportunidad.

Otra dirección relevante es la integración de variables socioeconómicas más detalladas, tales como índices de escolaridad, tasas de desempleo, patrones de movilidad urbana o niveles de percepción de inseguridad. Estos datos permitirían enriquecer el contexto explicativo de cada clúster, identificando posibles factores asociados a las dinámicas delictivas. Su incorporación podría mejorar la capacidad de diagnóstico del modelo y fortalecer la relación entre factores sociales y comportamientos criminales.

Finalmente, se recomienda realizar una validación externa de los resultados mediante datos cualitativos, como encuestas de percepción ciudadana o análisis de denuncias. Esta comparación permitiría verificar si la segmentación generada por los algoritmos refleja

efectivamente la experiencia y percepción de la población en cada territorio. Este paso sería clave para aumentar la legitimidad de los resultados y fortalecer su aplicabilidad en el diseño de políticas públicas.

Recomendaciones

Con base en los patrones delictivos identificados mediante los modelos de clustering, se proponen una serie de acciones orientadas a optimizar la gestión de la seguridad en Bucaramanga. Estas recomendaciones responden a las características específicas de cada clúster, permitiendo intervenciones diferenciadas que se ajustan al perfil delictivo y contexto territorial de cada zona.

Para el clúster asociado a crimen organizado y violencia extrema, correspondiente a la comuna Centro, se sugiere reforzar la presencia policial permanente en puntos críticos, con especial atención en horarios nocturnos. Además, se recomienda implementar inteligencia urbana que permita identificar redes organizadas, patrones delictivos persistentes y nodos de operación. Esta estrategia debe ir acompañada de una articulación institucional con fiscalía y entes de control, con el fin de fortalecer las rutas judiciales de respuesta rápida ante delitos complejos.

En el caso del clúster de violencia familiar y delitos violentos, que abarca comunas del norte, nororiente y occidente, se propone desplegar campañas interinstitucionales de prevención de violencia intrafamiliar, orientadas a comunidades vulnerables. Se recomienda también articular el trabajo con líderes sociales y organizaciones de base para fomentar la denuncia y el acompañamiento comunitario. Igualmente, es fundamental garantizar la disponibilidad de rutas de atención psicosocial en centros de salud, comisarías de familia y espacios comunitarios clave.

Para las zonas comerciales con robos, como Cabecera del Llano, Oriental y La Concordia, se plantea la aplicación de estrategias de vigilancia inteligente, como la instalación de cámaras en corredores comerciales y áreas de alta circulación. A esto se suma la necesidad de campañas de prevención situacional del delito dirigidas a comerciantes y clientes, así como la promoción de protocolos de seguridad colaborativa con participación del sector privado.

En el clúster caracterizado por robos moderados y alta movilidad juvenil, que incluye comunas como Ciudadela Real, Tejar y Provenza, se recomienda fomentar actividades culturales, deportivas y formativas para jóvenes, especialmente en espacios públicos. Adicionalmente, se deben fortalecer los patrullajes preventivos durante horarios de mayor riesgo (tardes y noches de fines de semana) y mejorar las condiciones urbanas, como iluminación pública, que contribuyan a la vigilancia natural y el control informal.

Respecto a las zonas residenciales con baja violencia, como Mutis, Suroccidente, Sur y Pedregosa, se sugiere mantener una presencia policial disuasiva mediante patrullajes aleatorios, sin llegar a una militarización del espacio. También se propone incentivar la creación de frentes de seguridad ciudadana y redes vecinales, particularmente en barrios en expansión o con crecimiento reciente. Por último, se recomienda implementar estrategias de prevención del delito de oportunidad, como talleres comunitarios sobre seguridad residencial y la instalación de cámaras compartidas.

A nivel transversal, se deben priorizar las zonas identificadas como más vulnerables según el índice de Necesidades Básicas Insatisfechas (NBI), en donde confluyen factores de exclusión social y alta carga delictiva. Los resultados del análisis de clústeres deben integrarse en los planes operativos de la Policía Metropolitana, orientando la focalización de recursos y patrullajes. Finalmente, se recomienda el uso de la aplicación desarrollada como una herramienta práctica para la consulta y planificación táctica, en espacios como los comités locales de seguridad y convivencia.

Referencias Bibliográficas

A. Alkhaibari Ping-Tsai Chung. (2017). *Cluster Analysis for Reducing City Crime Rates* (A. and T. C. (LISAT) Long Island Systems, Ed.; pp. 1–6). IEEE. https://doi.org/10.1109/LISAT.2017.8001983

Alcaldía de Bucaramanga. (2020). *Plan de Desarrollo 2020-2023: Bucaramanga, Ciudad Comprometida*.

Alcaldía de Bucaramanga. (2024). *Plan de Desarrollo Municipal de Bucaramanga 2024-2027: Bucaramanga Avanza Segura*.

Ceballos, J. D. (2023). Clasificación de crímenes por zonas en la ciudad de Nueva York utilizando técnicas de aprendizaje automático no supervisado. In *Google Scholar*. Universidad de Antioquia.

De Vasconcelos Dos Santos Junior, R., Venceslau Coelho, J. V., Azevedo Cacho, N. A., & Amorim De Araujo, D. S. (2023). Analyzing Criminal Macrocauses on Intentional Lethal Violent Crimes: An Unsupervised Learning Approach for Smart City Initiatives. *Proceedings of 2023 IEEE International Smart Cities Conference, ISC2 2023*. https://doi.org/10.1109/ISC257844.2023.10293658

Ezugwu, A. E., Ikotun, A. M., Oyelade, O. O., Abualigah, L., Agushaka, J. O., Eke, C. I., & Akinyelu, A. A. (2022). A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects. In *Engineering Applications of Artificial Intelligence* (Vol. 110). Elsevier Ltd. https://doi.org/10.1016/j.engappai.2022.104743

Fontalvo Herrera, T. J., Vega Hernández, M. A., & Mejía Zambrano, F. (2023). Método de clustering e inteligencia artificial para clasificar y proyectar delitos violentos en Colombia.

Revista Cientifica General Jose Maria Cordova, *21*(42), 551–572. https://doi.org/10.21830/19006586.1117

Gao, C. X., Dwyer, D., Zhu, Y., Smith, C. L., Du, L., Filia, K. M., Bayer, J., Menssink, J. M., Wang, T., Bergmeir, C., Wood, S., & Cotton, S. M. (2023). An overview of clustering methods with guidelines for application in mental health research. In *Psychiatry Research* (Vol. 327). Elsevier Ireland Ltd. https://doi.org/10.1016/j.psychres.2023.115265

Gelvez Ferreira, J. D., Nieto-Rodríguez, M.-P., & Rocha-Ruiz, C.-A. (2022). Prediciendo el crimen en ciudades intermedias: un modelo de "machine learning" en Bucaramanga, Colombia. *URVIO. Revista Latinoamericana de Estudios de Seguridad*, *34*, 83–98. https://doi.org/10.17141/urvio.34.2022.5395

Gouri Jha, L. A. A. R. (2020). *Criminal Behaviour Analysis and Segmentation using K-Means Clustering.* IEEE.

Han, J. (2012). *Data Mining Concept and Techniques* (ELSEVIER, Ed.; Third Edition).

Huang, Z. (1998). *A Fast Clustering Algorithm to Cluster Very Large Categorical Data Sets in Data Mining*.

Jerin, J. Q., Khan, N. K., Biswas, S., & Sharmin, N. (2023). Comparative Study of Clustering Algorithms: Scenario Based on Boston Crime Dataset. *2023 26th International Conference on Computer and Information Technology, ICCIT 2023*. https://doi.org/10.1109/ICCIT60459.2023.10441641

Martínez-González, M. A. , O. A. , D.-R. M. , & F.-C. J. (1993). *Evaluación de la significación estadística y cálculo del intervalo de confianza de la razón de mortalidad estandarizada*. *67*, 65–74.

Núñez, L. C., Tolentino, F., & y Rodríguez, H. (2023). Factores sociodemográficos en la dinámica del comportamiento delictivo: análisis descriptivo de criminalidad en Colombia, año 2022. *Revista Criminalidad*, *65*(3), 161–280. https://doi.org/10.47741/17943108.525

Pinto, O. O. (2022). *Comprensión dinámica del fenómeno de la criminalidad en Colombia* [Universidad Jorge Tadeo Lozano]. https://doi.org/20.500.12010/28052

Programa de las Naciones Unidas para el Desarrollo (PNUD). (2021). *Atrapados: Alta desigualdad y bajo crecimiento en América Latina y el Caribe.*

Ratra, A., Agarwal, A., Vats, S., Sharma, V., Kukreja, V., & Yadav, S. P. (2023). A Comprehensive Review on Crime Patterns and Trends Analysis using Machine Learning. *Proceedings of the 2023 2nd International Conference on Augmented Intelligence and Sustainable Systems, ICAISS 2023*, 732–736. https://doi.org/10.1109/ICAISS58487.2023.10250664

Richards, T. (2021). *Getting Started with Streamlit for Data Science : Create and Deploy Streamlit Web Applications From Scratch in Python*.

Sagala, N. T. M., & Gunawan, A. A. S. (2022). Discovering the Optimal Number of Crime Cluster Using Elbow, Silhouette, Gap Statistics, and NbClust Methods. *ComTech: Computer, Mathematics and Engineering Applications*, *13*(1), 1–10. https://doi.org/10.21512/comtech.v13i1.7270

Sinaga, K. P., & Yang, M. S. (2020). Unsupervised K-means clustering algorithm. *IEEE Access*, *8*, 80716–80727. https://doi.org/10.1109/ACCESS.2020.2988796

Torres, B. (2023). Caracterización de la Criminalidad en las Provincias del Ecuador y su Incidencia en las Variables Macroeconómicas Empleando Clustering Jerarquizado. In *Google Schoolar*. Escuela Superior Politécnico Del Litoral

Apéndices

Apéndice A

Repositorio en Github

https://github.com/jcamilobm/streamlit_cluster_crimes

Apéndice B

Demo

https://appclustercrimes-hps7f8ushwdyf2bvx9b6k8.streamlit.app

Apéndice C

Prompt Modelo de Lenguaje

Para la interpretación automatizada de los resultados del modelo de clustering aplicado a datos de criminalidad urbana, se empleó un modelo de lenguaje de gran escala (LLM) al que se le suministró un conjunto estructurado de instrucciones. El propósito fue traducir los patrones estadísticos en estrategias operativas comprensibles y accionables para cuerpos de policía con recursos limitados. El LLM fue instruido para actuar como un experto en criminología y análisis de datos aplicados a la seguridad pública operativa.

El objeto de entrada consistía en un archivo JSON con cuatro secciones principales: (1) un esquema explicativo de cada clave; (2) metadatos del modelo como configuración, fechas y parámetros de entrenamiento (informacion_modelo); (3) resultados técnicos del modelo, incluyendo etiquetas asignadas (labels_), métricas de separación y proporciones de cada cluster (resultados_modelo); y (4) proporciones normalizadas por dimensión delictiva para cada cluster (proporciones_por_dimension). Estas dimensiones incluían tipo de delito, uso de armas,

momento del día, edad de la víctima, movilidad y ubicación. También se incluía una lista de comunas con su Riesgo Multivariado Estimado (RME), descripción contextual y cluster asignado.

Se establecieron reglas explícitas para garantizar rigor metodológico: no se podían inferir ni estimar valores que no estuvieran explícitamente contenidos en los datos; todos los porcentajes debían citarse textualmente; y cada cluster debía ser nombrado de forma autónoma por el modelo, con base en las proporciones dominantes de cada dimensión.

La estructura esperada de la respuesta incluía cinco secciones: (1) un Resumen General de las tendencias y anomalías encontradas en los datos; (2) un Análisis por Cluster con nombre representativo, tipos de delito predominantes y nivel de riesgo; (3) Recomendaciones Estratégicas por cluster, adaptadas a sus características internas y contexto socioeconómico; (4) Acciones Prioritarias organizadas por impacto, factibilidad y urgencia; y (5) un Plan Táctico por Cluster que indicara objetivos operativos, recursos mínimos requeridos, acciones concretas, tiempos estimados e indicadores esperados. Todo debía expresarse en lenguaje claro, directo y sin tecnicismos, considerando que el insumo final sería utilizado por personal operativo en terreno.

Apendice D

Ejemplo Respuesta LLM con Modelo Meta-llama/llama-4-maverick:free

El análisis de criminalidad en la ciudad, basado en técnicas de agrupamiento no supervisado, permitió identificar cinco clusters diferenciados con características delictivas particulares. Los clusters con mayor peso porcentual fueron el Cluster 0 (29,41%) y el Cluster 4 (23,53%). Se evidencian patrones reiterados como la prevalencia del delito de “Robos y Hurtos”

en todos los grupos, aunque con distintas intensidades (por ejemplo, 40% en el Cluster 0 frente a un 75% en el Cluster 1). La violencia letal es un componente relevante en todos los segmentos, destacándose en el Cluster 0 con un 61%. Asimismo, las comunas con los mayores índices delictivos presentan altos niveles de necesidades básicas insatisfechas y elevada informalidad laboral, lo que sugiere una relación entre condiciones socioeconómicas y criminalidad.

Cluster 0: Violencia y Desigualdad. Este grupo concentra delitos de Robos y Hurtos (40%), Delitos Violentos (25%) y Violencia Familiar (25%). Las víctimas y agresores se desplazan principalmente a pie (89% y 86% respectivamente), y los hechos ocurren en gran medida en vías públicas (54%) y zonas residenciales (29%). La violencia letal es el tipo de amenaza predominante (61%), lo que sitúa a este cluster en un nivel de riesgo alto.

Cluster 1: Robos y Hurtos. Este segmento presenta un predominio significativo de Robos y Hurtos (75%). Tanto víctimas como agresores se movilizan mayoritariamente a pie (85% y 69%), aunque se observa una participación creciente de agresores en motocicleta (25%). Los delitos se concentran en vías públicas (52%) y en establecimientos comerciales o de servicios (12%). Su nivel de riesgo es moderado-alto.

Cluster 2: Delincuencia Moderada. También marcado por Robos y Hurtos (69%), este cluster comparte rasgos con el anterior, pero con menor incidencia general. Resalta una mayor proporción de agresores en motocicleta (21%), lo cual puede indicar una tendencia emergente hacia delitos más rápidos y móviles. El riesgo se clasifica como moderado.

Cluster 3: Crimen Organizado. Aquí, Robos y Hurtos alcanzan el 69%, junto con una proporción no despreciable de Crimen Organizado (8%). Los delitos se concentran en vías públicas (57%) y sectores comerciales (17%). Predominan víctimas adultas. Aunque representa

solo el 5,88% del total, su nivel de riesgo es alto debido a la sofisticación de las redes delictivas involucradas.

Cluster 4: Violencia Diversa. Este grupo muestra una combinación de Robos y Hurtos (51%), Delitos Violentos (20%) y Violencia Familiar (19%). Las dinámicas de movilidad reflejan que tanto víctimas como victimarios se movilizan mayoritariamente a pie, y las escenas del delito se sitúan principalmente en vías públicas (52%) y residencias (29%). Se considera de riesgo moderado-alto.

Recomendaciones Estratégicas

Para el Cluster 0, se recomienda un patrullaje intensivo en vías públicas y zonas residenciales, así como programas comunitarios orientados a reducir la violencia intrafamiliar. En los Clusters 1 y 2, es crucial incrementar la vigilancia en áreas comerciales y fortalecer las capacidades de inteligencia para anticipar y neutralizar robos y hurtos. En el caso del Cluster 3, se debe focalizar la acción operativa en el desmantelamiento de redes criminales, con una coordinación efectiva entre policía judicial, fiscalía y organismos de inteligencia. Finalmente, el Cluster 4 amerita estrategias similares al Cluster 0, con un enfoque especial en la atención integral de los distintos tipos de violencia coexistentes.

Acciones Prioritarias

Entre las acciones de mayor impacto destacan: el fortalecimiento del patrullaje en vías públicas, la implementación de programas comunitarios preventivos enfocados en violencia familiar, el aumento de la vigilancia en zonas de alta circulación comercial y el desarrollo de capacidades de inteligencia para enfrentar redes organizadas de crimen.