

# PREDICIENDO EL CRIMEN EN CIUDADES INTERMEDIAS: UN MODELO DE “MACHINE LEARNING” EN BUCARAMANGA, COLOMBIA

Una presentación de Vera Andy

# INTRODUCCIÓN

La investigación sobre el uso de tecnología para prevenir el crimen se ha centrado principalmente en ciudades principales, que cuentan con una gran cantidad de datos y mejores capacidades locales. Sin embargo, en ciudades intermedias con poca información disponible, la aplicación de modelos de predicción del delito puede presentar limitaciones.

# OBJETIVO

El objetivo de esta investigación es presentar los resultados de un modelo de "machine learning" para predecir el delito en Bucaramanga, una ciudad intermedia de Colombia. Se describe el enfoque utilizado para desarrollar el modelo y se discuten los resultados obtenidos, incluyendo las dificultades encontradas en la aplicación del modelo en una ciudad intermedia con información limitada.

## SOPORTE TEORICO

Este estudio se basa en la efectividad demostrada de las estrategias de prevención del delito a lo largo de los años, con un enfoque en la reducción de la criminalidad. Se destaca que la concentración de recursos policiales en unidades geográficas más pequeñas ha demostrado maximizar la eficacia de estas estrategias.

1

la reducción de la delincuencia al duplicar el tiempo de patrullaje en áreas específicas en Bogotá, Colombia

2

la instalación de cámaras de vigilancia en Medellín, Colombia, que resultó en una reducción de los delitos contra la vida y la propiedad.

**LA ESTRATEGIA DE PREDICCIÓN DEL DELITO SE CONSIDERA EFICAZ PARA REDUCIR VARIOS TIPOS DE CRÍMENES, PERO NO NECESARIAMENTE TODOS, Y SE REQUIERE ADAPTARLA A LAS NECESIDADES ESPECÍFICAS DE CADA AUTORIDAD DE POLICÍA.**

**ESTE ESTUDIO SE CENTRA EN UN MODELO DE GRAFOS IMPLEMENTADO EN COLOMBIA Y SE RELACIONA CON INVESTIGACIONES PREVIAS EN OTROS LUGARES QUE APlicaron MODELOS ESPACIOTEMPORALES PARA PREDECIR EL DELITO CON RESULTADOS VARIADOS. SE DESTACA LA IMPORTANCIA DE CONTAR CON INFORMACIÓN ESPECÍFICA Y GEOGRÁFICAMENTE DETALLADA PARA IMPLEMENTAR CON ÉXITO ESTE TIPO DE MODELOS.**

# METODOLOGÍA

La metodología de esta investigación se basa en el uso de datos del Sistema de Información Estadístico, Delincuencial, Contravencional y Operativo de la Policía Nacional (SIEDCO)

Para obtener información georreferenciada a nivel de barrio o manzana, se utilizaron mapas (shapefiles) publicados en las páginas web de entidades públicas colombianas.

Se llevaron a cabo análisis descriptivos de la frecuencia de las variables utilizadas en el modelo de predicción del delito a escala nacional para identificar diferencias estructurales en el comportamiento y las modalidades del delito según el mes, la hora, la zona y el género.

La disponibilidad de datos detallados permitió interpretar los resultados del modelo de predicción del delito en niveles geográficos más específicos.

**Tabla 1. Matrices, ecuaciones y formulas utilizadas**

|            |                                                       |                                                                                                                                  |
|------------|-------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------|
| Matriz 1   | Matriz de delitos-tiempo                              | $\begin{pmatrix} d_{p_1,t-T} & \cdots & d_{p_1,t} \\ \vdots & \ddots & \vdots \\ d_{p_n,t-T} & \cdots & d_{p_n,t} \end{pmatrix}$ |
| Ecuación 1 | Cálculo de matriz de frecuencia delito-tiempo         | $f(t, p) = \frac{x_{p,t}}{\sum_{t' \in p} x_{t',p}}$                                                                             |
| Ecuación 2 | Cálculo de matriz inversa de frecuencia delito-tiempo | $if(t, P) = \left( \frac{ P }{\{ p \in P; t \in p \}} \right)$                                                                   |
| Ecuación 3 | Cálculo de matriz F                                   | $F = f * if$                                                                                                                     |
| Ecuación 4 | Cálculo del modelo de clasificación                   | $d_{i,t} = g(\sum d_{i,t-j} * F)$                                                                                                |

**SE CENTRA EN LA TRANSFORMACIÓN DE DATOS A PARTIR DE UNA SERIE DE ECUACIONES.**

# CLASIFICACIÓN

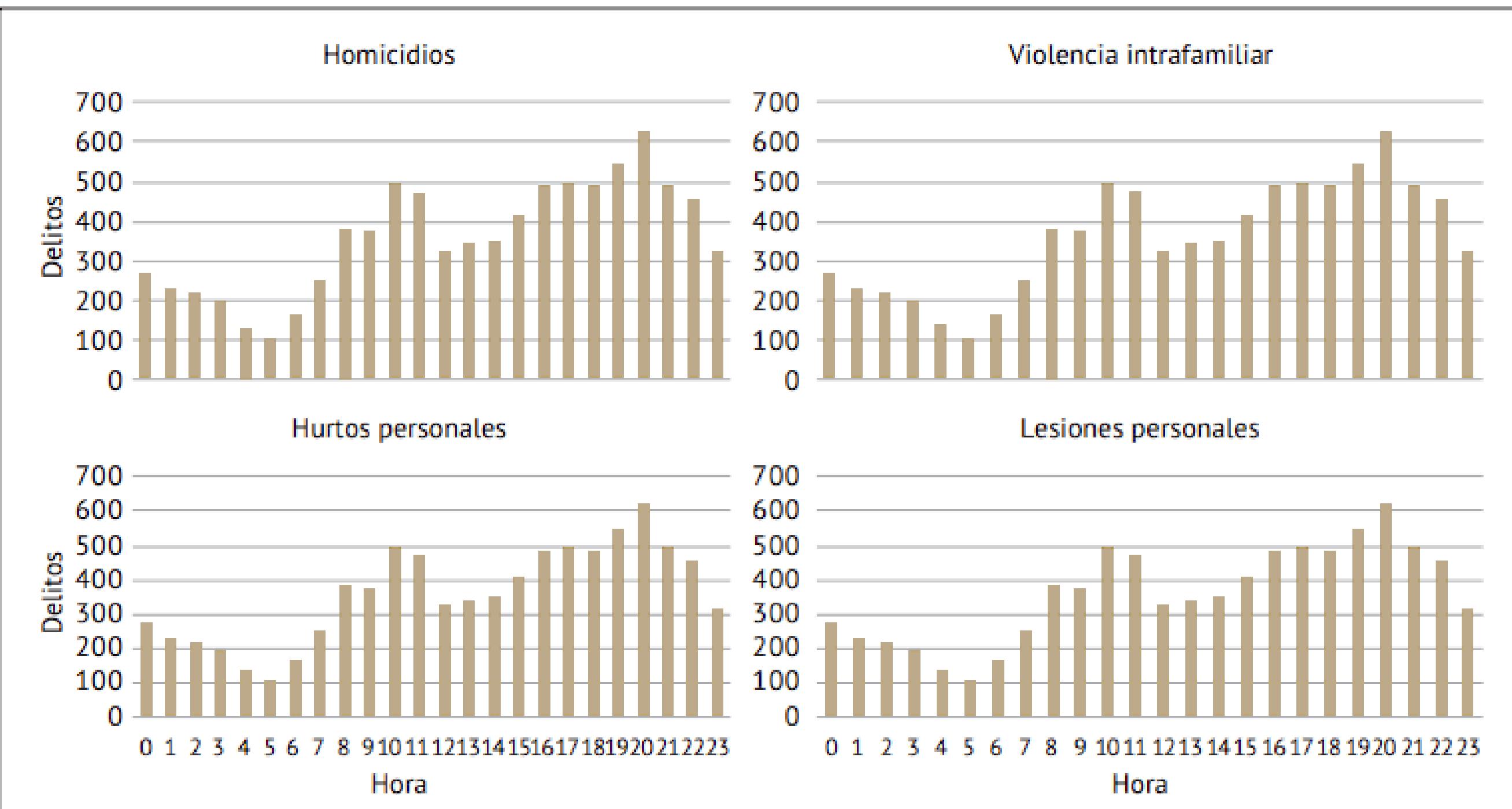
## CLASIFICACIÓN DE DELITOS:

- LOS DELITOS SE CLASIFICARON EN CUATRO CATEGORÍAS PRINCIPALES: HOMICIDIO, VIOLENCIA INTRAFAMILIAR, HURTOS Y LESIONES PERSONALES.

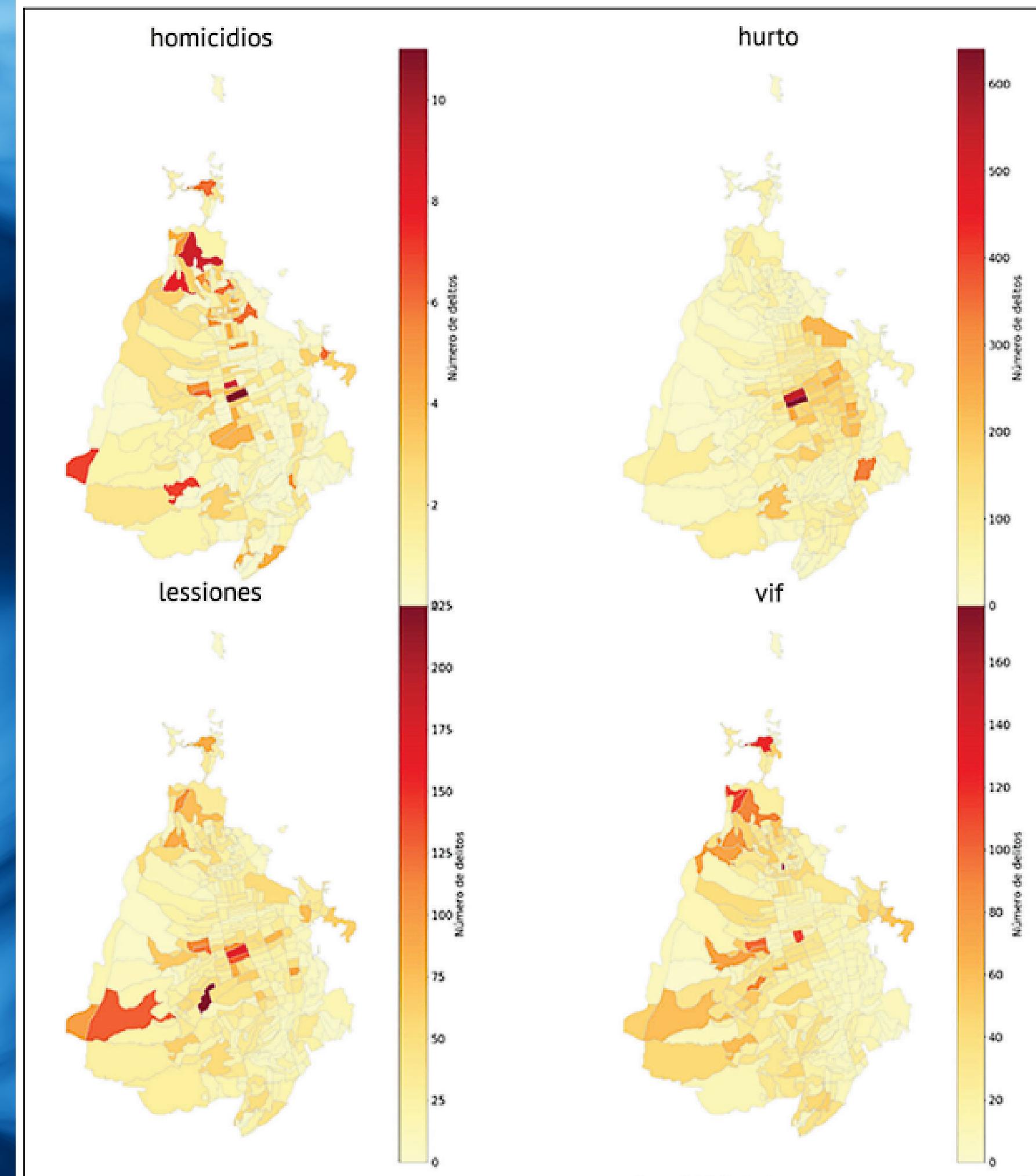
LA DISTRIBUCIÓN DE ESTOS DELITOS EN LA MUESTRA SE PRESENTA DE LA SIGUIENTE MANERA:

HURTOS A PERSONAS: 55.5%.  
LESIONES PERSONALES: 23.6%.  
VIOLENCIA INTRAFAMILIAR: 19.7%.  
HOMICIDIOS: 1.1%.

Figura 1. Distribución de los delitos de interés por hora del día



Mapa 1. Cantidad de delitos por tipo de delitos en Bucaramanga, 2016-2019



LOS MEJORES RESULTADOS EN LA PREDICCIÓN DEL CRIMEN EN BUCARAMANGA SE OBTUVIERON AL UTILIZAR MODELOS CON RELACIONES ESPACIALES DE GRAFOS POR SEMANAS. ESPECÍFICAMENTE, EL MODELO QUE LOGRÓ UNA MAYOR PRECISIÓN FUE EL MODELO KNN CON UNA FRECUENCIA SEMANAL, ALCANZANDO UN NIVEL DE PRECISIÓN EN LA PREDICCIÓN DE DELITOS ENTRE EL 50% Y EL 60%. ESTO SIGNIFICA QUE APROXIMADAMENTE LA MITAD DE LOS PRONÓSTICOS REALIZADOS POR EL MODELO ERAN VERDADEROS DELITOS.

LOS MODELOS DE FRECUENCIA DIARIA PRESENTARON MAYORES DIFICULTADES EN LA PREDICCIÓN, Y AUNQUE EL MODELO SVM FUE EL MEJOR EN ESTA CATEGORÍA, NO LOGRÓ SUPERAR EL 60% DE PRECISIÓN EN LA PREDICCIÓN.

# CONCLUSION

La predicción del delito en entidades territoriales con poca información estadística puede ser útil, pero a la vez compleja y costosa en su desarrollo e implementación. Esto se debe a la limitada disponibilidad de datos delictivos y georreferenciados en ciudades pequeñas e intermedias, así como a los desafíos procedimentales y logísticos en la recolección de información.