GENERACIÓN DE ALERTAS TEMPRANAS PARA REDUCIR LA MORTALIDAD EN ACCIDENTES DE TRÁNSITO EN CALI CON DATOS HISTÓRICOS, MEDIANTE UN MODELO DE PATRONES PUNTUALES

Carlos Camilo Cuchumbé Escandón CC: 1.144.031.185
Gabriel Alejandro Gordillo Alvarez CC: 1.014.299.990

Proyecto Aplicado para optar al título de
Magister en Ciencia de Datos

Director(a)
David Arango Londoño
CC: 1130586950

FACULTAD DE INGENIERÍA Y CIENCIAS
MAESTRÍA EN CIENCIA DE DATOS

SANTIAGO DE CALI, JUNIO 9 DE 2025

FICHA RESUMEN

TRABAJO DE GRADO DE MAESTRÍA

TÍTULO: Generación de alertas tempranas para reducir la mortalidad en accidentes de tránsito en cali con datos históricos, mediante un modelo de patrones puntuales

1. TIPO DE PROYECTO (Aplicado, Innovación, Investigación): Aplicado
2. ÁREA DE TRABAJO: Sector público
3. ESTUDIANTE(S): Carlos Camilo Cuchumbé Escandón y Gabriel Alejandro Gordillo Alvarez
4. CORREO ELECTRÓNICO: camilo2289@javerianacali.edu.co, gordillogabriel@javerianacali.edu.co.
5. DIRECCIÓN Y TELÉFONO: Cll 152 #72-85 Bogotá Colombia TEL. 3016463892
6. DIRECTOR: David Arango Londoño
7. VINCULACIÓN DEL DIRECTOR: Docente Catedrático
8. CORREO ELECTRÓNICO DEL DIRECTOR: david.arango@javerianacali.edu.co
9. GRUPO O EMPRESA QUE LO AVALA (Si aplica): Secretaría de movilidad de Cali
10. PALABRAS CLAVE (al menos 5): Mortalidad en accidentes de tránsito, prevención de accidentes, seguridad vial y sistemas de alertas tempranas, infracciones y accidentes de tránsito, análisis espacial.
11. FECHA DE INICIO: 15/05/2024
12. DURACIÓN ESTIMADA (En meses): 11 meses
13. RESUMEN: Este estudio desarrolló un modelo de alerta temprana para prevenir desenlaces fatales de accidentes de tránsito en la ciudad de Cali, un problema crítico de salud pública con altos costos asociados a la pérdida de vidas humanas. Se preveía que la mortalidad por este tipo de sucesos sería una de las principales causas de muerte en los países en vías de desarrollo, y Cali no era ajena a esta tendencia mundial, con cifras que superan la media nacional. El objetivo principal fue identificar la correlación entre infracciones de tránsito y accidentes mortales mediante técnicas de patrones puntuales y modelos de aprendizaje automático explicable, con el fin de desarrollar un sistema de alertas tempranas para informar a autoridades y ciudadanos sobre conductas y zonas de riesgo. Los resultados obtenidos

incluyeron un modelo predictivo del riesgo de accidentes de tránsito mortales a partir de las infracciones registradas,buscando una mejora en la movilidad urbana y en la eficiencia de la gestión de recursos. Las aplicaciones derivadas de este proyecto abarcan desde la implementación de políticas públicas y estrategias de intervención hasta la integración con sistemas de gestión de tráfico y planificación urbana. La disponibilidad de datos de infracciones y el uso de tecnologías avanzadas aseguraron la viabilidad del proyecto y su potencial para ofrecer soluciones efectivas y sostenibles para la mejora de la seguridad vial en Cali.

TABLA DE CONTENIDO

LISTA DE TABLAS

LISTA DE FIGURAS

INTRODUCCIÓN

Las muertes por accidentes de tránsito constituyen una constante preocupación en la ciudad de Cali, con una media aproximada de 19 fallecimientos por cada cien mil habitantes desde que se tienen registros del año 2005 a 2023 , esto según datos del Departamento Administrativo Nacional de Estadística (DANE) [1]. Esta problemática no es exclusiva de la ciudad de Cali, ya que a nivel nacional se observan patrones similares, con una media de 14 muertes por cada cien mil habitantes [2]. La planificación inadecuada, la señalización deficiente y la irresponsabilidad de los conductores incrementan la probabilidad de muertes por accidentes, generando elevados costos para los presupuestos municipales y limitando la capacidad de inversión en áreas con retornos a largo plazo. En este contexto, la seguridad vial se convierte así en una variable estratégica crucial para prevenir el aumento del gasto en salud a nivel municipal.

La capital del Valle del Cauca presenta recurrentes deficiencias en materia de tránsito. La seguridad vial implica el cumplimiento adecuado de las normas de tránsito, lo cual es esencial para prevenir accidentes. Según datos de la Secretaría de Movilidad de la ciudad, en 2023 se impusieron 387,458 comparendos, de los cuales el 70% se atribuyen a conducir sin portar el Seguro Obligatorio (37.3%) y a no realizar la revisión técnico-mecánica (33.2%) [3]. Estas infracciones reflejan un incumplimiento significativo de las normas básicas de seguridad vial, subrayando la necesidad urgente de intervenir en estos aspectos para reducir la accidentalidad y sus consecuencias fatales.

En este marco, los modelos de alertas tempranas emergen como una herramienta prometedora para abordar las variables estructurales que afectan la seguridad vial en una ciudad, Este proyecto tiene como objetivo formalizar un modelo que logre identificar zonas rojas con mayor probabilidad de ocurrencia de siniestros viales. Factores como la planificación urbana y el estado de las calzadas son variables críticas que pueden explicar gran parte de los siniestros viales. Este enfoque busca identificar y analizar los determinantes de los accidentes de tránsito, proporcionando una base sólida para la implementación de medidas preventivas y de mitigación más efectivas. Los resultados pretenden a través de la implementación de la herramienta de alertas tempranas reducir la probabilidad de ocurrencia de siniestros con desenlaces fatales.

El sistema de alertas tempranas funciona como una herramienta para los hacedores de políticas y planificadores urbanos, permitiendo la implementación de soluciones eficientes mediante el uso de herramientas estadísticas, analítica espacial y sistemas de información geográfica (SIG). Este estudio se compone de la identificación de las tendencias mundiales en seguridad vial, la concentración de siniestros en grupos de edad específicos según los últimos reportes de los organismos dedicados a velar por el bienestar general. Además de la reflexión de la importancia de concentrar esfuerzos en pensar soluciones profundas. Finalmente se describe la noción matemática detrás de la medición utilizada como base de la inferencia estadística.

1. DEFINICIÓN DEL PROBLEMA

1.1. PLANTEAMIENTO DEL PROBLEMA

La mortalidad en accidentes de tránsito es catalogada por la Organización Mundial de la Salud (OMS) como la principal causa de muerte entre las personas de 5 a 29 años, con una media aproximada de 1.9 millones de muertes al año, donde las tendencias actuales indican que, si no se toman medidas urgentes, los accidentes de tránsito se convertirán en el 2030 en la quinta causa de muertes. A pesar de que el 60% de los vehículos se concentran en los países de ingresos medianos y bajos, estos países representan el 92% de las muertes por accidentes de tránsito en un año. Los principales factores de riesgo asociados a la mortalidad por accidentes de tránsito son la velocidad, la conducción bajo los efectos del alcohol u otras sustancias psicoactivas, la no utilización de cascos, cinturones de seguridad y sistemas de sujeción para niños, las distracciones durante la conducción, la falta de seguridad de la infraestructura vial y de los vehículos, la atención insuficiente tras las colisiones y el cumplimiento insuficiente de las normas de tránsito [4].

Según la OMS el 28% de las defunciones por accidentes de tránsito se produjeron en la Región de Asia Sudoriental, el 25% en la Región del Pacífico Occidental, el 19% en la Región de África, el 12% en la Región de las Américas, el 11% en la Región del Mediterráneo Oriental y el 5% en la Región de Europa [5].

En la Unión Europea, las víctimas mortales en carretera en 2022 aumentaron un 3% con respecto al 2021, donde en los últimos tres años, el número de muertes en carretera en países como Irlanda, España, Francia, Italia, los Países Bajos y Suecia se han mantenido estable o incluso han aumentado. Según datos disponibles relativos a 2021, en el conjunto de la Unión Europea, el 52% de las víctimas mortales en accidentes de tráfico tuvieron lugar en carreteras rurales, frente a 39% en zonas urbanas y en 9% en autopistas, donde los ocupantes de automóviles representaron el 45% de todas las muertes en carreteras, mientras que los peatones representaron el 18%, los motociclistas el 19% y los ciclistas el 9% del total de las víctimas mortales [6].

En África las muertes por accidentes de tráfico representan aproximadamente una cuarta parte del número mundial de víctimas, a pesar de que el continente apenas cuenta con el 2% del parque automovilístico mundial. El África subsahariana, la región del mundo más afectada por los accidentes de tráfico, tiene una tasa de 27 víctimas mortales por cada 100.000 habitantes. Esta cifra es tres veces superior a la media europea de nueve por cada 100.000, y muy por encima de la media mundial de 18 por cada 100.000 [7].

Entre 2010 y 2019, más de la mitad de los países de América Latina y el Caribe registraron un aumento de sus elevadas tasas de muertes por accidentes de tránsito. Según datos de la OMS, en promedio, la región pasó en ese período de 15,68 a 17,66 fallecidos por cada cien mil habitantes, siendo superada solo por África. En números absolutos, según la OMS, en la región de las Américas murieron 144.090 personas en 2021 por esta causa, lo que corresponde al 12% del total mundial [8].

En Colombia, la accidentalidad vial constituye una de las principales causas de mortalidad. Las ciudades que más incrementaron fallecidos entre enero y junio de 2023 son: Bogotá, D.C., Cali, Sincelejo, Santa Marta, Yopal, Neiva, Villavicencio, Medellín, Montería y Pereira. Durante los años 2022 y 2023, se registraron 16,669 muertes por accidentes de tránsito. Del 2021 al 2022, se observó un incremento porcentual en las muertes por accidentes de tránsito de un 13.67%. Los departamentos con mayor representación en esta mortalidad son Antioquia y Valle del Cauca, con un 12.8% y 10.6% respectivamente. En el Valle del Cauca, la ciudad de Cali representa el 36.4% de las muertes en siniestros viales del departamento siendo los usuarios de motocicleta y peatones los más afectados con el 52% y el 37.1% de las muertes representativas en el 2023 [2].

Si se sostiene la situación actual por el número de muertes y lesiones por siniestros viales, no solo se podría tener un impacto en la salud pública de la ciudad de Cali sino también en su economía y la del país en general ya que según los datos aportados por la OMS la mayoría de los países gastan el 3% PIB en colisiones debidas a tránsito [4].

1.2. FORMULACIÓN DEL PROBLEMA

Para abordar de manera correcta el reto de generar un sistema de alertas tempranas a través de un modelo de predicción de mortalidad por accidentes de tránsito en la ciudad de Cali, es necesario responder a los siguientes interrogantes: ¿Cómo se puede utilizar la ciencia de datos para desarrollar un modelo de alertas tempranas que prediga las zonas de mayor riesgo de mortalidad por accidentes de tránsito en la ciudad de Cali? ¿Es posible identificar las zonas de mayor riesgo? ¿Qué variables explican la mortalidad por siniestros viales en la ciudad de Cali utilizando un modelo de ciencia de datos que logre relacionar las infracciones de tránsito con los accidentes fatales? ¿Cómo se pueden utilizar los datos históricos de infracciones de tránsito y accidentes para prevenir estos accidentes fatales una vez identificados los mecanismos de transmisión y sus posibles causas de muertes? ¿Qué tipo de infracciones de tránsito están más correlacionadas con el riesgo de un accidente fatal? ¿Cuáles son las áreas de la ciudad de Cali con mayor ocurrencia de infracciones y accidentes de tránsito? ¿Qué factores demográficos y temporales pueden influir en el riesgo de un accidente fatal? ¿Cuáles son los grupos de riesgo para accidentes de tránsito en la ciudad y cómo se pueden identificar de manera efectiva?

2. OBJETIVOS DEL PROYECTO

2.1. OBJETIVO GENERAL

Desarrollar un modelo de ciencia de datos utilizando metodología de patrones puntuales para identificar y analizar datos históricos de infracciones de tránsito en la ciudad de Cali, con el fin de generar alertas tempranas que permitan identificar patrones, factores y zonas de riesgo, facilitando la implementación de medidas preventivas para reducir la mortalidad en siniestros viales.

2.2. OBJETIVOS ESPECÍFICOS

1) Identificar los grupos vulnerables bajo los siniestros viales con mayor frecuencia de mortalidad.

2) Estimar la influencia de los factores demográficos y temporales en la mortalidad por accidentes de tránsito.

3) Identificar las correlaciones cruzadas examinando la relación entre mortalidad e infracciones de tránsito.

4) Desarrollar, evaluar y validar los modelos predictivos utilizando técnicas de análisis de patrones puntuales para determinar las áreas con mayor riesgo de accidentes de tránsito.

5) Generar alertas tempranas basadas en el modelo predictivo, integrando la visualización de los datos obtenidos durante la exploración y modelado para prevenir accidentes de tránsito en zonas de alto riesgo.

3. MARCO TEÓRICO DE REFERENCIAS Y ANTECEDENTES

3.1. MARCO CONCEPTUAL

El presente capítulo expone las bases conceptuales y metodológicas que tienen relación con la predicción de la mortalidad en accidentes de tránsito y el desarrollo de sistemas de alertas tempranas. Los siguientes conceptos son esenciales para comprender el contexto y la justificación de este estudio.

3.1.1. SINIESTRO VIAL

Es el que permite vincular causas, consecuencias y responsabilidades de la persona en un evento de tránsito. Incluso, la palabra "siniestro" tiene un significado de catástrofe y se asocia con circunstancias dolorosas, como las lesiones o la pérdida de una vida, las cuales se pudieron haber prevenido en el marco de la responsabilidad y la autorregulación. En este sentido, en seguridad vial se opta por siniestro vial y no accidente vial, ya que éste es un suceso imprevisible e inevitable asociado al azar donde se exonera a la persona de toda responsabilidad [9].

3.1.2. CHOQUE

Este tipo de siniestro vial se produce por el impacto violento entre dos o más vehículos en movimiento, o entre un vehículo y un objeto fijo [9] (como postes, barreras o árboles). A diferencia de otras modalidades de accidentes, los choques destacan por su dinamismo, ya que la gravedad de las consecuencias depende directamente de factores como la velocidad al momento del impacto, el tipo de vehículos involucrados y las condiciones de la vía.

3.1.3. ATROPELLAMIENTO

Este tipo de siniestro vial ocurre cuando un vehículo impacta a un peatón [9], ya sea en cruces, calles o incluso aceras, convirtiéndose en uno de los accidentes más graves debido a la vulnerabilidad de las personas a pie. A diferencia de otras colisiones, aquí el riesgo de lesiones mortales o incapacitantes es especialmente alto, pues el cuerpo humano queda expuesto directamente a la fuerza del impacto.

3.1.4. VOLCAMIENTO

Es el evento primario en el cual el vehículo pierde su posición normal durante el siniestro vial y puede quedar de manera lateral o longitudinal; sus llantas pierden el contacto con la superficie de la vía [9].

3.1.5. ACCIDENTE DE TRÁNSITO

Evento generalmente involuntario, generado al menos por un vehículo en movimiento, que causa daños a personas y bienes involucrados en él e igualmente afecta la normal circulación de los vehículos que se movilizan por la vía o vías comprendidas en el lugar o dentro de la zona de influencia del hecho.

3.1.6. FACTORES DE RIESGO EN ACCIDENTES DE TRÁNSITO

Un factor de riesgo en accidentes de tránsito es cualquier actividad, acción o condición que incrementa la probabilidad de que ocurra un desenlace negativo, como una lesión o muerte.

3.1.6.1. FACTORES HUMANOS

Esta es la causa de un gran número de accidentes de tráfico, que se deben a la irresponsabilidad del conductor o del peatón [10].

3.1.6.2. FACTORES VEHICULARES

Estos factores engloban tanto las condiciones técnicas del vehículo como sus modificaciones, las cuales pueden convertirse en determinantes críticos para la ocurrencia o severidad de un siniestro vial. A diferencia de otros tipos de factores (humanos o ambientales), los vehiculares tienen la particularidad de ser, en muchos casos, prevenibles mediante mantenimiento adecuado y regulaciones técnicas [11].

3.1.6.3. FACTORES AMBIENTALES Y VIALES

Este factor se debe a condiciones como la época del año, hielo, la niebla, los materiales que alteran las condiciones normales de la calzada, el alumbrado público y las condiciones estructurales de la arteria vial [11].

3.2. MARCO TEÓRICO

3.2.1 MODELOS DE PREDICCIÓN EN SEGURIDAD VIAL

Un modelo estadístico de predicción de accidentes fiable capaz de analizar grandes volúmenes de datos y aplicar técnicas estadísticas avanzadas para el análisis, garantizando la precisión de sus predicciones que permita generar factores de alerta sobre puntos estratégicos de la estructura de la ciudad [12].

3.2.2 MODELOS ESTADÍSTICOS

Son las estimaciones empíricas que permiten inferir variables en un modelo establecido con validez estadística.

3.2.3 MODELOS DE PATRONES PUNTUALES

Son los modelos que identifican una zona determinada refiriéndose a la distribución y organización de los eventos, describiendo la fiabilidad y la seguridad de la red ante interacción de diferentes factores que concentran alta accidentalidad [13].

Diggle define un proceso puntual como los datos que forman un conjunto de puntos y que se distribuyen irregularmente dentro de una región del espacio. Las coordenadas espaciales (longitud, latitud) son denominadas eventos [14].

El objetivo general es comprobar si los puntos exhiben algún tipo de patrón

- Patrón aleatorio: los puntos se distribuyen aleatoriamente en el espacio
- Patrón regular: existe una distancia media entre los puntos que tiende a ser constante.
- Patrón agregado: existen aglomeraciones de puntos en el espacio.

3.2.4 ESTADÍSTICAS DE RESUMEN DE PRIMER ORDEN

3.2.4.1 INTENSIDAD

El primer paso para el análisis exploratorio es la intensidad del patrón, esta medida describe el número esperado de eventos por unidad de área. La densidad empírica de los puntos está dada por la siguiente expresión:

$$\bar{\lambda} = \frac{n(x)}{|W|}$$

donde n(x) corresponde al número de puntos en la región de análisis y |W| su respectiva área [14].

3.2.4.2 ESTIMACIÓN DE LA FUNCIÓN DE INTENSIDAD POR SUAVIZAMIENTO DE KERNEL

La estimación no paramétrica de la función de intensidad se realiza a través del estimador de kernel propuesto por Diggle y dado por la siguiente expresión:

$$\bar{\lambda} = \sum_{i=1}^{n} \frac{1}{e(x_i)} k(u - x_i)$$

donde su localización dentro de la ventana de observación; W representa la Ventana de observación; k(u) representa la función kernel que debe ser una densidad de probabilidad k(u) ≥ 0 para todas las localizaciones u [14].

3.2.4.3 MÉTODO DE CONTEO POR CUADRANTES PARA ALEATORIEDAD ESPACIAL COMPLETA (CSR)

Este método consiste en dividir la ventana de observación W en subregiones B_1,….,B_m llamadas cuadrantes que, en principio, se consideran de igual área. A partir de esto, se realiza el conteo de los puntos que caen en cada subregión nj = n(X ∩ Bj) n_j=n(X∩B_j) para j = 1, ..., m. Al tratarse de intensidad homogénea, los conteos deberían ser iguales o cercanos en cada subregión, por el contrario, los cambios fuertes en los conteos pueden sugerir tendencias espaciales [14].

Con el objetivo de contrastar la hipótesis de homogeneidad

H_0 = La intensidad es homogénea (CRS)

H_1 = La intensidad no es homogénea

Se aplica la prueba X^2 (Chi-cuadrado), dado el número total de puntos $n = \sum_j n_j$, el área total de la ventana de observación $a = \sum_j a_j$, la intensidad estimada $\bar{\lambda} = \frac{n}{a}$ y el número esperado de conteos en el cuadrante $\overline{B}_j$ es $e_j = \bar{\lambda}_{aj} = na_j/a$. El estadístico está dado por:

$$X^2_{calculado} = \sum_j \frac{(observados - esperados)^2}{esperados} = \sum_j \frac{(n_j - e_j)^2}{e_j} = \sum_j \frac{(n_j - \bar{\lambda}_{aj})}{(\bar{\lambda}_{aj})^2}$$

El contraste se realiza a través del valor crítico donde se rechaza la hipótesis nula si

$$X^2_{calculado} > X^2_{critico,gl}$$

donde g representa los grados de libertad dados por el número de cuadrantes de igual área resultado de la partición menos uno (n − 1) [14].

3.2.5 PROPIEDADES DE PRIMER Y SEGUNDO ORDEN

Las propiedades de primer orden son descritas por la intensidad del proceso, su función está dada por:

$$\lambda(x) = \left(\frac{E[N(dx)]}{|dx|}\right)$$

Donde dx define una mínima región espacial alrededor de x, |dx| el área de la región, y N(dx) es el número de eventos ocurridos en dx [14].

Las propiedades de segundo orden dan a conocer el tipo de las interacciones entre los eventos del proceso. La intensidad de segundo orden de dos puntos x e y refleja la probabilidad de que ocurra cualquier par de eventos en las proximidades de x e y, respectivamente, su función está dada por:

$$\lambda_2(x, y) = \left(\frac{E[N(dx,dy)]}{|dx||dy|}\right)$$

Donde dy es un intervalo que contiene el tiempo y, |dy| representa la longitud del intervalo y N(dx,dy) es el número de eventos ocurridos en dx*dy [14].

En estas propiedades se incluye la densidad de covarianza la cual está dada por:

$$\gamma(x,y) = \lambda_2(x,y) - \lambda(x)\lambda(y)$$

La densidad de covarianza es el proceso puntual similar a la función de covarianza de un proceso estocástico evaluado en los reales, la densidad de covarianza será idénticamente cero [14].

Además, se encuentra en estas propiedades la función de correlación de pares de puntos, la cual está dada por:

$$g(x,y) = \frac{\lambda_2(x,y)}{\lambda(x)\lambda(y)}$$

La función de correlación de pares se puede interpretar como la función de densidad de probabilidad estándar aplicada a que un evento ocurra en cada uno de las observaciones centradas en los puntos (x, y) [14].

3.2.6 FUNCIÓN K DE RIPLEY

La función K de Ripley es una función de λ2 para procesos estacionarios e isotrópicos. También es conocida como la medida reducida del segundo momento (Cressie(1993)), y es la función reducida del segundo momento. Mide el número de eventos encontrados hasta una distancia dada de cualquier evento en particular [14].

Está dada por:

$$k(d) = \frac{E(N_e)}{\lambda}$$

Donde:

λ = Intensidad

N_e = Número de eventos adicionales a una distancia d de un evento elegido aleatoriamente

3.2.7 PROCESOS DE POISSON

El proceso Poisson da un punto de inicio natural para un desarrollo estadístico de un patrón puntual observado.

Suponiendo $\{T_1, T_2, \ldots, T_n\}$ variables aleatorias distribuidas representan los tiempos que transcurren entre una ocurrencia del evento y la siguiente ocurrencia, además estos tiempos son independientes y cada uno sigue una distribución exp(λ) Por lo tanto, se puede definir el proceso de Poisson al tiempo t como el número de ocurrencias del evento que se han observado hasta ese instante t.

De esta manera se puede definir de manera espacio temporal un proceso de Poisson así:

Sea $\{T_1, T_2, \ldots, T_n\}$ una sucesión de variables aleatorias independientes cada una con distribución exp(λ), el proceso de Poisson de parámetro λ para un proceso continuo $\{X_t: t \geq 0\}$, estará definido de la siguiente manera

$$X_t = máx\{n \geq 1 : T_1 + T_2 + \ldots + T_n \leq t\}$$

Esto significa que, para un valor de n ≥ 1, todos los tiempos de Inter arribo a partir de n, siguen teniendo distribución exp(λ), y por lo tanto el proceso de conteo de eventos a partir del tiempo n es un proceso de Poisson.

3.2.8 PROCESOS DE POISSON NO HOMOGÉNEOS

Se originan cuando el proceso no es estacionario y el parámetro λ del proceso de Poisson no es necesariamente una constante sino una función del tiempo y se obtiene sustituyendo la intensidad constante λ de un proceso Poisson por una función de intensidad variable λ(x).

Estos procesos están definidos por:

- · N(A): Número de eventos en la región A.
- · |A|: Área de la región A.
- E(N(A)) = λ(A).
- λ(A)=∫_a λ(s)ds.

Estos procesos de Poisson no homogéneos tienen las siguientes características:

- · N(0) = 0
- N(A) ~ Poisson (λ(A))
- Si y son dos regiones no traslapadas, N(A_1) y N(A_2) son independientes.

Dado N(A) = n, las localizaciones $\{s_1, s_2, \ldots, s_n\}$ son realizaciones independientes distribuidas uniformemente.

$$f_A(n) = \frac{\lambda(s)}{\int_A \lambda(s).ds}$$

En este caso los intervalos de tiempo entre eventos sucesivos T_n, n ≥ 1 ya no son independientes ni tienen distribución exponencial.

3.2.9 PATRONES POISSON

En Huang (2011) se estudia la relación de correlación entre procesos de dos puntos. Cuando el proceso puntual es Poisson, un solo coeficiente es suficiente para describir la relación de correlación.

Dado el teorema:

sea $\{N_t: > 0\}$ el proceso de Poisson. El número esperado y la varianza de eventos en (0, t] son

$$E(N_t) = \lambda t$$
$$V(N_t) = \lambda t$$

Donde $N_1(b)$ y $N_2(b)$ son procesos de dos puntos, ρ(b) denotando los coeficientes de correlación entre y $N_1(b)$ y $N_2(b)$

$$\{N_1(b) = \overline{N_1(b)} + N(b)\}$$
$$\{N_2(b) = \overline{N_2(b)} + N(b)\}$$

Donde $N_1(b)$ y $N_2(b)$ y $N(b)$, son procesos de Poisson independientes con tasa λ_1,λ_2 y λ_i entonces $N_1(b)$ y $N_2(b)$ son procesos de dos puntos, por lo tanto, el coeficiente de correlación entre $N_1(b)$ y $N_2(b)$ es:

$$\rho(b) = \frac{\lambda}{\sqrt{(\lambda_1+\lambda)(\lambda_2+\lambda)}}$$

3.3. ANTECEDENTES

3.3.1 ARQUITECTURA DE INFRAESTRUCTURA Y MODELO DE OPTIMIZACIÓN PARA LA REDUCCIÓN DE LA PROBABILIDAD DE ACCIDENTES DE USUARIOS VULNERABLES EN LA VÍA PÚBLICA

Con el desarrollo de las ciudades que tratan de implantar conexiones entre las oportunidades que ofrecen, han surgido diversos problemas derivados de su expansión. Entre ellos, cómo llegar a puntos concretos de la ciudad y cómo garantizar una mayor y más eficiente cobertura de su población. La movilidad urbana es un componente vital que oxigena las ciudades y permite su desarrollo. Esta accesibilidad depende de una planificación, un uso del suelo y una coordinación adecuada. Sin embargo, las conexiones también aumentan la probabilidad de accidentes de tráfico, un tema de gran relevancia en la literatura debido a los costes sanitarios. Se estima que, en 2016, en Pakistán, el coste total ocasionado por los accidentes de tráfico supuso aproximadamente el 0.0074% del PIB. [13].

En Colombia, entre el año 2011 y 2016, se realizó un estudio en donde se analizaron diez ciudades del país, destacándose entre ellas Bucaramanga, Neiva y Popayán, las cuales se encuentran entre las ciudades con las tasas de mortalidad por accidentes de tráfico más altas del país. Para el año 2011, el costo económico de las muertes por accidentes de tránsito en estas ciudades representó entre el 4% y el 21% del PIB de cada ciudad [15].

En consecuencia, se han llevado a cabo algunas investigaciones, entre los que se destaca «desarrollo de herramientas para la visualización e identificación de accidentes de tránsito a una distancia específica de puntos de interés en Bogotá, como parte del proyecto VISIR. Esta investigación busca ayudar a la secretaría de movilidad en la toma de decisiones con el fin de reducir los accidentes de tránsito en Bogotá. Adicionalmente, busca identificar puntos críticos de infraestructura donde ocurran accidentes a una distancia cercana [16].

Otra investigación destacable es la titulada «Arquitectura de infraestructuras y modelo de optimización para la reducción de la probabilidad de accidentes de usuarios vulnerables de la vía[17]. Este estudio se centró en las altas tasas de mortalidad por accidentes de tránsito en Colombia, destacando que los más vulnerables son los peatones, motociclistas y ciclistas.

Bonilla, propone una arquitectura de infraestructuras que integra vehículos conectados e informática perimetral para reducir la probabilidad de accidentes entre usuarios vulnerables de la vía pública. Además, desarrolla un modelo matemático de optimización diseñado para reducir la probabilidad de accidentes entre peatones, motociclistas y ciclistas. Debido a la ausencia de datos disponibles, se generó un conjunto de datos personalizados mediante simulaciones.

Por último, los resultados de la investigación indicaron que el modelo optimizado presentaba resultados prometedores, mostrando una tendencia general a la reducción de accidentes en comparación con las simulaciones sin apoyo de red [17].

3.3.2 IDENTIFICATION OF HOTSPOT AREAS FOR TRAFFIC ACCIDENTS AND ANALYZING DRIVERS' BEHAVIORS AND ROAD ACCIDENTS

Este estudio evalúa las diversas causas de accidentes de tránsito en la ciudad de Abu Dhabi, considerando factores circunstanciales, características demográficas e infracciones de tránsito entre otras. Para realizar el análisis, se emplea un análisis de autocorrelación espacial, con el objetivo de identificar los puntos críticos de accidentes en el año 2014. Además, se llevó a cabo una encuesta en 2017 a 1,072 conductores involucrados en accidentes de tránsito en la ciudad. Los resultados de esta encuesta, analizados mediante modelos de regresión logística, revelan que la conducción descuidada es uno de los factores más significativos en la ocurrencia de accidentes de tránsito, seguida por la experiencia en conducción y la edad del conductor. El estudio concluye con sugerencias preventivas para mejorar la seguridad vial en Abu Dhabi [18].

Este trabajo aporta de manera significativa al proyecto actual al utilizar un enfoque metodológico que utiliza análisis de autocorrelación espacial y modelos de regresión logística para identificar y analizar puntos críticos de accidentes, sin embargo aunque el proyecto plantea gran similitud en el enfoque metodológico la ciudad de cali tiene contextos demográficos y urbanos diferentes esto influye en la naturaleza y factores de riesgo de los accidentes de tránsito, además en el presente proyecto no solo realiza un análisis predictivo de accidentes de transito mortales basado en una base de datos de infracciones de tránsito sino que adicionalmente realiza un sistema de alertas tempranas.

3.3.3 EVALUATING EXPRESSWAY TRAFFIC CRASH SEVERITY BY USING LOGISTIC REGRESSION AND EXPLAINABLE & SUPERVISED MACHINE LEARNING CLASSIFIERS

Este estudio aborda el incremento significativo de accidentes de tránsito en las autopistas de Sri Lanka, atribuible a la expansión de la red de transporte y al alto volumen de tráfico. Este estudio utiliza el aprendizaje automático explicable para investigar los factores que afectan la gravedad de los accidentes de tránsito en autopistas. Evalúa dos grupos de accidentes: los fatales o graves, y otros que incluyen lesiones no graves o solo daños a la propiedad. Cinco factores contribuyentes fueron analizados: condición de la superficie de la carretera, alineación de la carretera, ubicación, condición climática y efecto de la iluminación. Se desarrollaron y compararon cuatro modelos de aprendizaje automático (Random Forest, Decision Tree, XGBoost, K-Nearest Neighbor) con un modelo de regresión logística utilizando 223 datos de entrenamiento y 56 de prueba. Los modelos de aprendizaje automático mostraron predicciones más precisas que el modelo de regresión logística. Para explicar los modelos, se usaron SHAP y LIME, revelando que todos los factores analizados disminuyeron la posibilidad de accidentes fatales. El estudio concluyó que los métodos de aprendizaje automático explicables son más efectivos que los análisis de regresión tradicionales para evaluar el rendimiento en seguridad. Los resultados del estudio pueden mejorar la seguridad vial proporcionando explicaciones precisas para la toma de decisiones en modelos complejos [19].

Este trabajo aporta significativamente al proyecto actual al demostrar la efectividad del aprendizaje automático explicable para analizar la gravedad de los accidentes de tránsito. La metodología utilizada, que incluye tanto modelos predictivos como técnicas de explicación de modelos, proporciona una base para entender los factores que contribuyen a la gravedad de los accidentes.

Aunque ambos estudios comparten la utilización de métodos avanzados de análisis de datos para mejorar la seguridad vial, existen diferencias notables. El estudio de Sri Lanka se centra en las autopistas y utiliza datos de accidentes en un contexto específico de expansión de la red vial y alto volumen de tráfico. En contraste, el proyecto propuesto en Cali se enfoca en un entorno urbano con diferentes características demográficas y de infraestructura. Además, se distingue por su intención de desarrollar un modelo predictivo basado en una base de datos de infracciones de tránsito y la implementación de un sistema de alertas tempranas específico para zonas de alto riesgo.

4. ANÁLISIS DESCRIPTIVO DE LOS DATOS

Para el desarrollo del modelo de patrones puntuales que busca relacionar los accidentes de tránsito con las infracciones, se utilizaron dos bases de datos proporcionadas por la Secretaría de Movilidad de la ciudad de Cali. La primera base de datos corresponde a los

comparendos impuestos en la ciudad durante los años 2021 y 2022, mientras que la segunda contiene los registros de accidentes de tránsito ocurridos en el mismo período.

4.1 BASE DE DATOS DE INFRACCIONES

La base de datos de multas inicialmente cuenta con 510.694 registros, en los cuales se identificaron 95 tipos de infracciones diferentes. Con el objetivo de focalizar el análisis en las infracciones con mayor incidencia y riesgo de accidentalidad, se realizó un proceso de limpieza y selección de datos. Este proceso consistió en reducir los 95 tipos de infracciones a 11 categorías, las cuales se seleccionaron por ser las que presentan un mayor número de registros y, al mismo tiempo, están asociadas a un mayor riesgo de provocar accidentes. Las infracciones finalmente seleccionadas se detallan en la *Tabla 1*.

Tipo Multa o Infracción	cantidad	Código multa
No realizar la revisión técnico-mecánica y de emisiones contaminantes en los plazos establecidos, o a conducir un vehículo que, a pesar de tener los certificados correspondientes, no cumple con las condiciones técnico-mecánicas o de emisiones	**161556**	**C35**
No detenerse ante una luz roja o amarilla de semáforo, una señal de "PARE" o un semáforo intermitente en rojo.	**35937**	**D04**
Conducir motocicleta sin observar las siguientes normas: **a) Transitar ocupando un carril, observando lo dispuesto en los artículos 60 y 68 del presente Código, así:** **1. Ocupar el carril dentro de las líneas de demarcación, y atravesarlos solamente para efectuar maniobras de adelantamiento o de cruce.** **2. En una vía de sentido único de tránsito, con velocidad reglamentada para sus carriles, los vehículos utilizarán el carril de acuerdo con su velocidad de marcha.** **3. En una vía de sentido único de tránsito, donde los carriles no tengan reglamentada su velocidad, los vehículos transitarán por el carril derecho y los demás carriles se emplearán para maniobras de adelantamiento.** **4. En una vía de doble sentido de tránsito con dos (2) carriles, el vehículo deberá transitar por el carril de su derecha y utilizar con precaución el carril de su izquierda para maniobras de adelantamiento y respetar siempre la señalización respectiva.** **5. En una vía de doble sentido de tránsito con tres (3) carriles los vehículos deberán transitar por los carriles extremos que queden a su derecha; el carril central sólo se utilizará en el sentido que señale la autoridad competente.** **6. En una vía de doble sentido de tránsito con cuatro (4) carriles, los carriles exteriores se utilizarán para el tránsito ordinario de vehículos, y los interiores, para maniobras de adelantamiento o para circular a**	**21630**	**C24**

mayores velocidades dentro de los límites establecidos. **7. Transitar en motocicletas y motociclos por las ciclorrutas o ciclovías. Además el vehículo será inmovilizado;** **b) Podrán llevar un acompañante en su vehículo, el cual también deberá utilizar casco y la prenda reflectiva exigida para el conductor;** **c) Deberán usar de acuerdo con lo estipulado para vehículos automotores, las luces direccionales. De igual forma utilizar, en todo momento, los espejos retrovisores;** **d) Todo el tiempo que transiten por las vías de uso público, deberán hacerlo con las luces delanteras y traseras encendidas;** **e) Los conductores y los acompañantes cuando hubieren, deberán utilizar casco de seguridad y en él, conforme a la reglamentación que expida el Ministerio de Transporte, el número de la placa del vehículo en que se transite, con excepción de los pertenecientes a la fuerza pública, que se identificarán con el número interno asignado por la respectiva institución. La no utilización del casco de seguridad cuando corresponda dará lugar a la inmovilización del vehículo;** **f) No se podrán transportar objetos que disminuyan la visibilidad, que incomoden al conductor o acompañante o que ofrezcan peligro para los demás usuarios de las vías;** **g) Los conductores de estos tipos de vehículos y sus acompañantes deben vestir chalecos o chaquetas reflectivas de identificación que deben ser visibles cuando se conduzca entre las 18:00 y las 6:00 horas del día siguiente, y siempre que la visibilidad sea escasa;** **h) No deben sujetarse de otro vehículo o viajar cerca de otro carruaje de mayor tamaño que lo oculte de la vista de los conductores que transiten en sentido contrario;** **i) No deben transitar sobre las aceras, lugares destinados al tránsito de peatones y por aquellas vías en donde las autoridades competentes lo prohíban. Deben conducir en las vías públicas permitidas o, donde existan, en aquellas especialmente diseñadas para ello;** **j) Deben respetar las señales, normas de tránsito y límites de velocidad;** **k) No deben adelantar a otros vehículos por la derecha o entre vehículos que transiten por sus respectivos carriles. Siempre utilizarán el carril libre a la izquierda del vehículo a sobrepasar;** **l) Deben usar las señales manuales detalladas en este código.**		
No acatar las señales de tránsito o requerimientos impartidos por los agentes de tránsito.	**19893**	**C31**
Conducir un vehículo a velocidad superior a la máxima permitida, la cual	**7559**	**C29**

deberá estar señalizada en forma sectorizada, no obstante esta no podrá ser superior a: **a) En vías urbanas del Distrito o Municipio respectivo a una velocidad superior a los 80 kilómetros por hora;** **b) En las vías urbanas, los vehículos de servicio público, de carga y de transporte escolar, a una velocidad superior a sesenta (60) kilómetros por hora;** **c) En las carreteras nacionales y departamentales las velocidades autorizadas, en ningún caso podrá sobrepasar los 120 kilómetros por hora;** **d) En las carreteras nacionales y departamentales para el servicio público, de carga y de transporte escolar, el límite de velocidad en ningún caso podrá exceder los ochenta (80) kilómetros por hora.**		
No respetar el paso de peatones que cruzan una vía en sitio permitido para ellos, o no darles la prelación en las franjas para ello establecidas.	**2524**	**C32**
Transitar en sentido contrario al estipulado para la vía, calzada o carril. En el caso de motocicletas se procederá a su inmovilización hasta tanto no se pague el valor de la multa o la autoridad competente decida sobre su imposición en los términos de los artículos 135 y 136 del Código Nacional de Tránsito.	**2332**	**D03**
No utilizar el cinturón de seguridad por parte de los ocupantes del vehículo y los cinturones de seguridad en los asientos traseros en los vehículos fabricados a partir del año 2004.	**1540**	**C06**
Conducir realizando maniobras altamente peligrosas, siempre y cuando la maniobra viole las normas de tránsito que pongan en peligro a las personas o las cosas y que constituyan conductas dolosas o altamente imprudentes. En el caso de motocicletas se procederá a su inmovilización hasta tanto no se pague el valor de la multa o la autoridad competente decida sobre su imposición en los términos de los artículos 135 y 136 del Código Nacional de Tránsito.	**926**	**D07**
Conducir un vehículo sobre aceras, plazas, vías peatonales, separadores, bermas, demarcaciones de canalización, zonas verdes o vías especiales para vehículos no motorizados. En el caso de motocicletas se procederá a su inmovilización hasta tanto no se pague el valor de la multa o la autoridad competente decida sobre su imposición en los términos de los artículos 135 y 136 del Código Nacional de Tránsito.	**662**	**D05**

Tabla 1. *tabla de las infracciones seleccionadas.*

Tras el proceso de depuración, la base de datos final contiene 254,559 registros que representan las infracciones con mayor potencial de causar accidentes. Esta depuración permite un análisis más eficiente, centrándonos en los patrones más relevantes para identificar relaciones entre infracciones y siniestros viales.

Los datos revelan que las motocicletas son el tipo de vehículo con mayor participación en infracciones, representando 20.07% del total en 2021, 33.88% en 2022. Este incremento del 68.81% interanual es alarmante. Los automóviles ocupan el segundo lugar, con un 16.31% en 2021 y 28.34% en 2022. Esto representa un aumento aún mayor (73.75%) entre ambos años. Estos resultados demuestran claramente que motocicletas y automóviles son los vehículos con mayor reincidencia en infracciones durante el periodo analizado. Los datos fueron reclasificados para obtener solo 4 posibles tipos de vehículos las frecuencias de infracciones para cada vehículo se muestran en la *Tabla 2*.

Clase automóvil	Años				Variabilidad
	2021		2022		
	Cantidad	Porcentaje	Cantidad	Porcentaje	
MOTOCICLETA	51086	20,07%	86239	33,88%	68,81%
AUTOMÓVIL	41431	16,31%	72151	28,34%	74,15%
CAMION	791	0,32%	1006	0,39%	27,18%
BUS	587	0,23%	916	0,36%	56,05%

Tabla 2. *Tabla de registro de infracciones por vehículo y año.*

Al analizar el comportamiento de las infracciones por día de la semana, se observa que, tanto en 2021 como en 2022, los días de lunes a sábado presentan una frecuencia de infracciones similar. Para el año 2021, el número de infracciones durante estos días osciló entre 12,500 y 15,000, mientras que en 2022 se registró un aumento significativo, con frecuencias que variaron entre 22,000 y 26,000 infracciones. Por otro lado, los domingos mostraron una menor incidencia de infracciones en comparación con los demás días de la semana, tanto en 2021 como en 2022 esto se puede observar en la *Figura 1*.

Este análisis tiene como objetivo identificar patrones temporales en la ocurrencia de infracciones, lo que permitirá comprender mejor los días de mayor riesgo y contribuirá a la implementación de estrategias preventivas por parte de las autoridades de movilidad. Además, estos hallazgos serán fundamentales para establecer correlaciones con los datos de accidentes de tránsito.

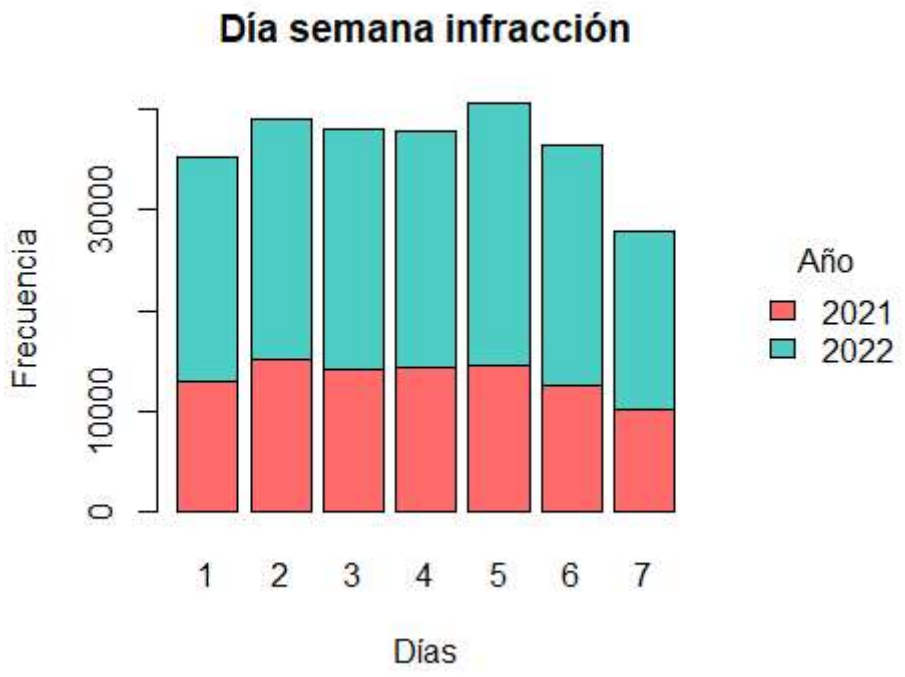

Figura 1. *Figura de distribución de las infracciones por día de semana y año.*

Continuando con el análisis de la temporalidad de los comparendos, se procedió a examinar la distribución de las infracciones por hora del día, con el fin de identificar los periodos de mayor riesgo de accidentalidad. Los resultados mostraron tendencias consistentes en ambos años analizados (2021 y 2022). Se evidenció que las horas con mayor frecuencia de infracciones se concentran entre las 7:00 a.m. y las 5:00 p.m., lo que coincide con los horarios de mayor movilidad vehicular, como las horas pico de la mañana y la tarde.

Por otro lado, se observó una frecuencia media de infracciones en los intervalos de 6:00 a.m. a 7:00 a.m. y de 6:00 p.m. a 11:00 p.m., correspondientes a los períodos de transición entre las horas de mayor y menor actividad. Finalmente, las horas con menor incidencia de infracciones se registraron entre la medianoche (12:00 a.m.) y las 5:00 a.m., lo que puede atribuirse a la reducción del flujo vehicular durante la madrugada, los datos mencionados se presentan en la *Figura 2*.

Este análisis horario resulta fundamental para comprender los patrones de comportamiento de los conductores y su relación con los riesgos de accidentalidad.

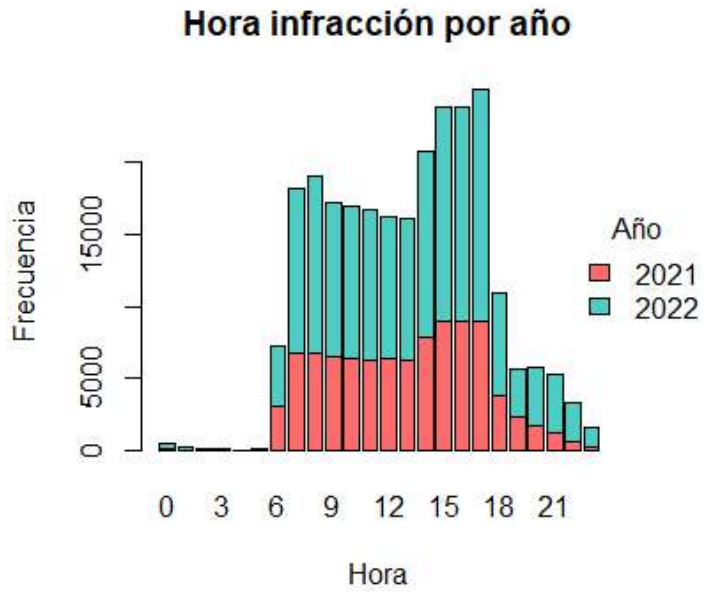

Figura 2. *distribución de las infracciones por hora y año*

Al desglosar el análisis por día de la semana, se observa un comportamiento diferenciado en los fines de semana (sábado y domingo). Durante estos días, se registra un incremento notable en la frecuencia de infracciones en el horario comprendido entre las 6:00 p.m. y las 11:00 p.m., en comparación con los días hábiles (de lunes a viernes). Este aumento podría estar asociado a factores como una mayor circulación vehicular en horas nocturnas, actividades sociales o recreativas, y una posible relajación en el cumplimiento de las normas de tránsito durante los fines de semana como se evidencia en la *Figura 3*.

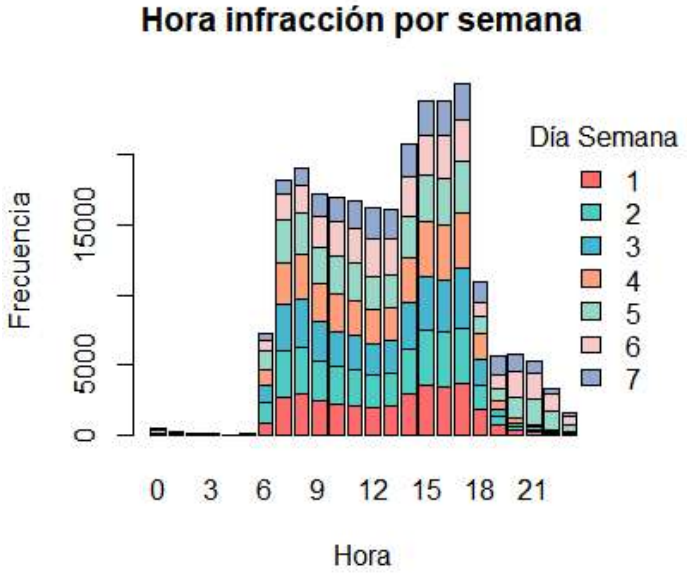

Figura 3. *Infracciones por hora y día de la semana.*

4.2 BASE DE DATOS DE SINIESTROS VIALES

La base de datos de accidentes o siniestros cuenta con un total de 55.092 registros, correspondientes a eventos ocurridos en el período comprendido entre 2018 y 2022. Durante el proceso de preparación de los datos, se identificaron 9 registros que no contaban con información de dirección, lo que imposibilita su geocodificación. Dado que la localización espacial es un aspecto fundamental para el análisis propuesto en este trabajo, estos registros fueron eliminados, obteniendo una base de datos final con 55.083 registros.

Además, debido a que la información fue ingresada de manera manual, se presentaron diversos errores de digitación y problemas relacionados con la inconsistencia en la entrada de datos. Esto requirió un extenso proceso de estandarización y depuración para garantizar la calidad y confiabilidad de la información. Dentro de la base de datos, se encuentran registrados diferentes tipos de eventos, clasificados en lesiones, daños materiales y muertes.

Al analizar la distribución de estos eventos por año, se evidencia que los de mayor frecuencia son aquellos que involucran daños materiales o lesiones, mientras que los eventos mortales representan una proporción significativamente menor como se observa en la *Figura 4*. Asimismo, se observa un comportamiento similar en la distribución de los tipos de eventos a lo largo de los años, lo que sugiere patrones consistentes en la ocurrencia de accidentes durante el período analizado.

Este análisis preliminar permite identificar las características principales de la base de datos de accidentes, sentando las bases para su integración con la base de datos de infracciones y la posterior identificación de patrones espaciales y temporales.

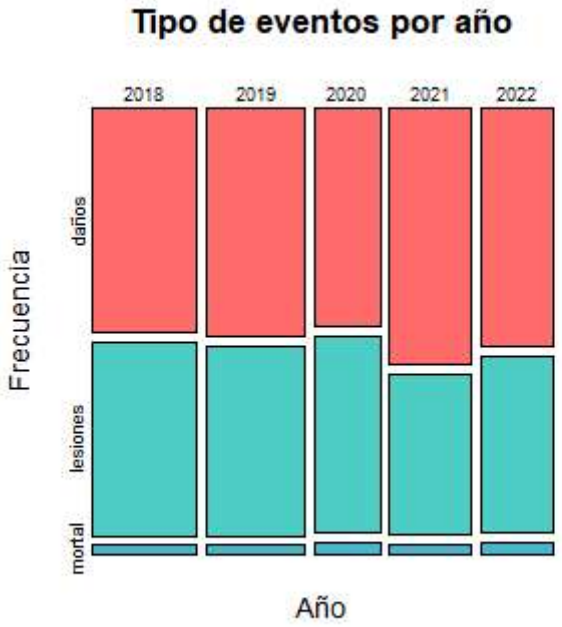

Figura 4. *Distribución del tipo de accidente por año.*

Al analizar la distribución de los accidentes de tránsito por semana y año, se observa que los años 2018 y 2019 presentaron las mayores frecuencias de siniestros viales. Sin

embargo, en el año 2020, se registró una disminución significativa en el número de accidentes, lo cual puede atribuirse a las restricciones de movilidad implementadas durante la pandemia del COVID-19, que redujeron drásticamente el flujo vehicular y, por ende, la probabilidad de ocurrencia de estos eventos.

Para los años 2021 y 2022, se evidencia un incremento en la cantidad de accidentes en comparación con 2020, lo que coincide con la reactivación gradual de las actividades económicas y sociales. No obstante, al contrastar estos años con los niveles previos a la pandemia (2018 y 2019), se observa que la frecuencia de accidentes continúa en una tendencia a la disminución. Este comportamiento sugiere que, aunque hubo una recuperación en la movilidad, los niveles de siniestralidad no han regresado a los registrados antes de la pandemia, lo que podría estar relacionado con cambios en los patrones de desplazamiento, la implementación de medidas de seguridad vial o otros factores asociados al contexto post pandemia como el incremento en los trabajos de modalidad virtual, estas dinámicas se muestran en la *Figura 5* y la *Tabla 3*.

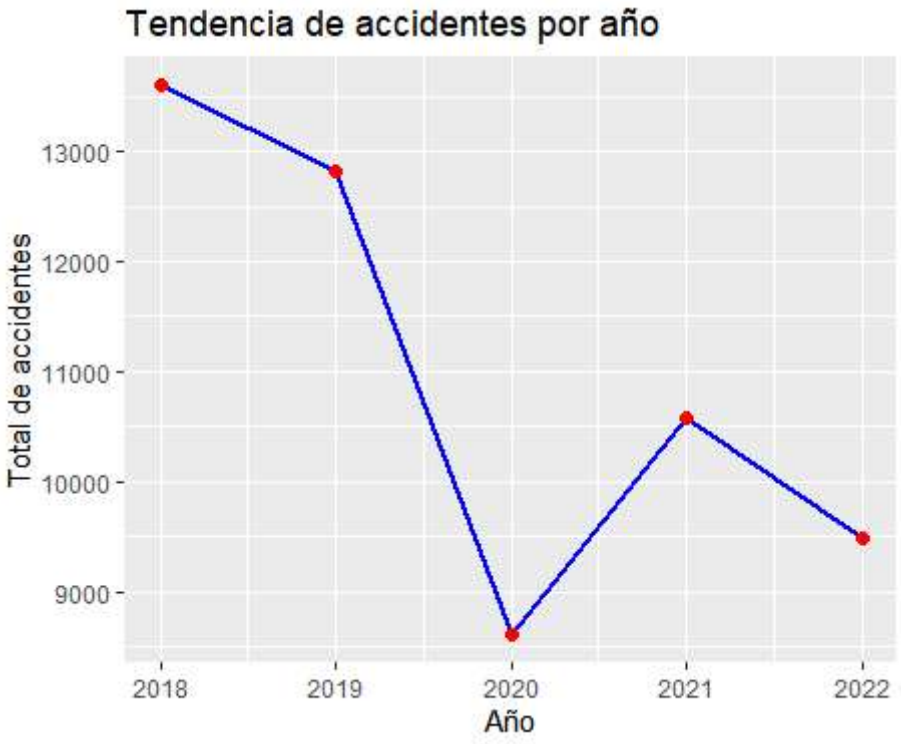

Figura 5. *Cantidad de accidentes por año*

DIA EVENTO	AÑO					
	2018	2019	2020	2021	2022	TOTAL
Lunes	1949	1798	1177	1564	1358	7846
Martes	1951	1970	1306	1557	1388	8172
Miércoles	1970	1865	1334	1534	1361	8064
Jueves	2035	1832	1285	1586	1350	8088
Viernes	2180	1943	1302	1631	1460	8516
Sábado	1983	1975	1233	1550	1453	8194
Domingo	1541	1446	967	1151	1104	6209
TOTAL	13609	12829	8604	10573	9474	55089

Tabla 3. *Accidentes por año y día de la semana.*

Con el fin de identificar los vehículos más presentes en accidentes de tránsito, se calculó una tabla de frecuencias que muestra un evidente predominio de automóviles, los cuales representan el 67.03% de los vehículos reportados. Le siguen las motocicletas con un 18.25%. Es importante destacar que estos porcentajes no representan la cantidad de registros individuales, ya que un mismo accidente puede involucrar varios vehículos del mismo tipo o diferentes. A continuación, se presenta el detalle de la distribución mediante la tabla 4.

vehículo	Frecuencia	Porcentaje
automóvil	37879	67,03%
motocicleta	10312	18,25%
peatón	593	1,05%
bicicleta	277	0,49%
bus	6851	12,12%
camión	393	0,70%
objeto fijo	11	0,02%
ambulancia	191	0,34%

Tabla 4. *Tabla de accidentes por vehículo.*

En cuanto a los vehículos involucrados en accidentes mortales, se observa un incremento significativo en la participación de peatones, motociclistas y ciclistas. Estos actores viales representan porcentajes del 34.05%, 46.00% y 9.80%, respectivamente. Esta tendencia refleja la vulnerabilidad de estos grupos en situaciones de tránsito. A continuación, se detalla la distribución mediante la *Tabla 5*.

vehículo	frecuencia	porcentaje
motociclista	532	41,47%
peatón	488	38,04%
ciclista	137	10,68%
parrillero	78	6,08%
pasajero de auto	25	1,95%
conductor	18	1,40%
sin dato	3	0,23%
peatón	2	0,16%

Tabla 5. *Tabla de accidentes mortales por vehículo.*

Al comparar los días de la semana entre los accidentes con lesiones y daños, y los accidentes mortales, se evidencia un valor más alto en los domingos para los accidentes mortales, mientras que aquellos que no resultan en fatalidad disminuyen como se muestra en la *Figura 6*.

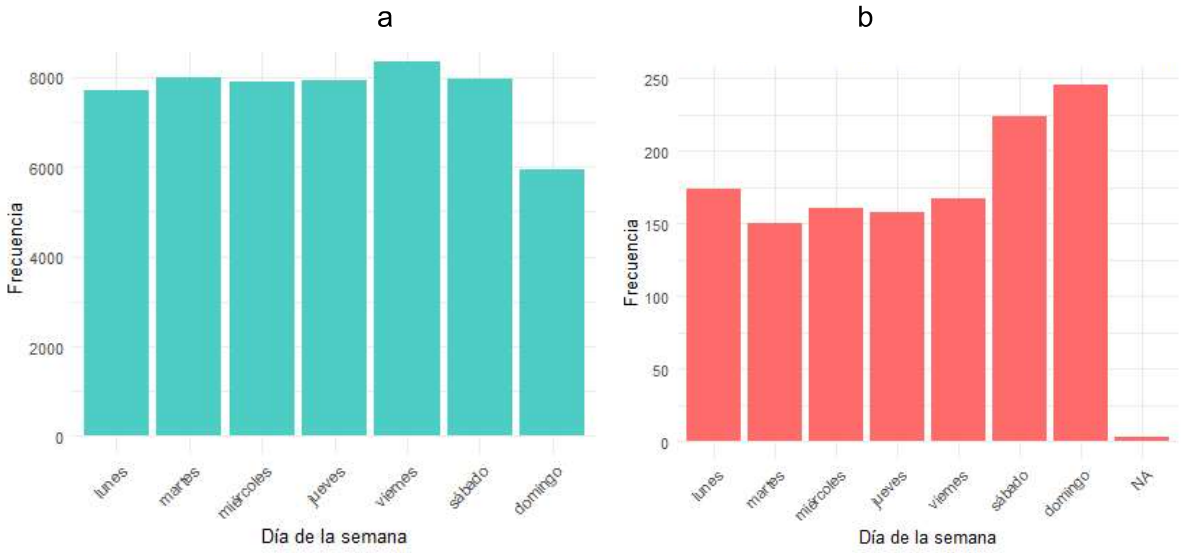

Figura 6. (a) *accidentes no mortales por día de la semana* ***(b)*** *accidentes mortales por día de la semana*.

Finalmente, al analizar la frecuencia de eventos mortales por edades y género, se observa que la mayor incidencia ocurre en los grupos etarios de 18 a 25 años y de 46 a 60 años. Además, la mayoría de las muertes corresponden al género masculino como se evidencia en la *Figura 7*.

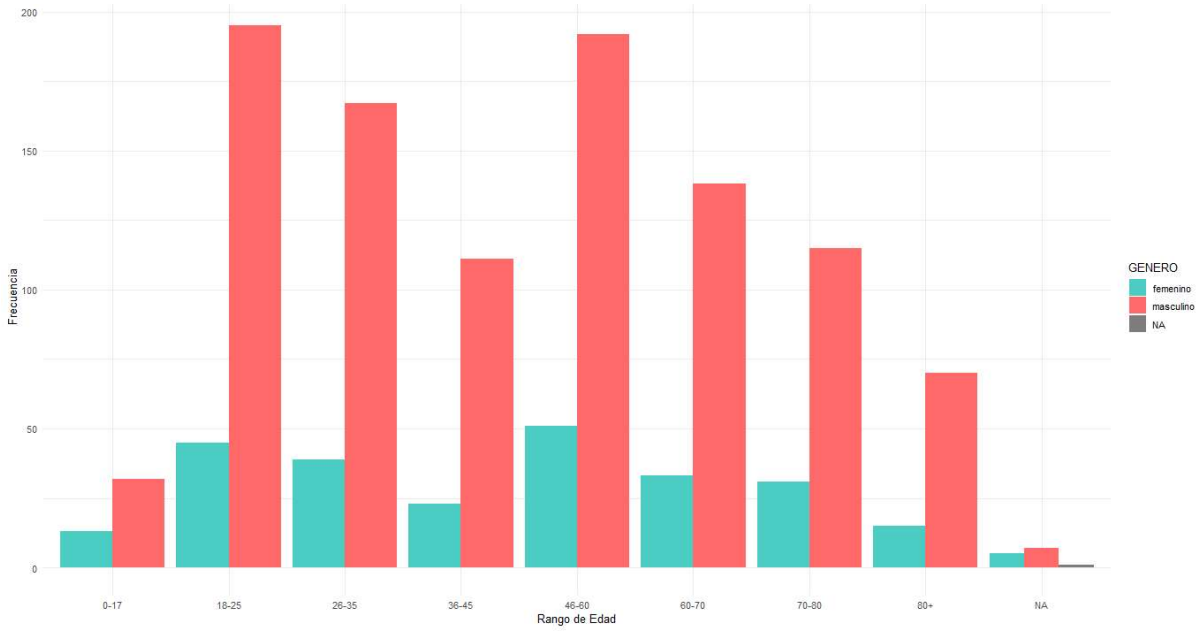

Figura 7. *Densidad de accidentes por edad y género.*

El análisis descriptivo de las bases de datos de infracciones y accidentes permitió identificar patrones en las características de infracciones y accidentes relevantes. Se observó un incremento significativo en infracciones asociadas a motocicletas y automóviles entre 2021 y 2022, así como una concentración de eventos en horarios diurnos y días laborales. Por otro lado, los accidentes mostraron una mayor incidencia en automóviles, aunque con mayor letalidad en usuarios vulnerables como peatones y motociclistas. Estos resultados sientan las bases para profundizar en la relación espacial entre infracciones y accidentes, tema que se abordará en el capítulo 5 mediante técnicas de geocodificación y análisis de correlación.

5. ANÁLISIS ESPACIAL DE INFRACCIONES Y ACCIDENTES

Para estudiar la relación espacial entre las infracciones de tránsito y los accidentes viales en la ciudad de Cali, fue necesario transformar las direcciones textuales de las bases de datos en coordenadas geográficas (latitud y longitud). Este proceso, conocido como geocodificación, permite ubicar los eventos en un sistema de referencia espacial durante este capítulo se detalla el proceso para llevar a cabo esta geocodificación y las posibles relaciones existentes entre accidentes de tránsito mortales e infracciones.

5.1. GEOCODIFICACIÓN DE LAS DIRECCIONES DE LAS BASES DATOS DE INFRACCIONES Y ACCIDENTES

Se realizó la geocodificación de las direcciones en las bases de datos de infracciones y accidentes. Para ello, primero se estandarizaron las direcciones, convirtiéndolas en minúsculas. Luego, se utilizaron los servicios de Google Geocode para obtener la latitud y longitud de cada registro.

El proceso de geocodificación se automatizó en R mediante un bucle for, que permitió enviar todas las direcciones a la API de Google y recuperar las coordenadas correspondientes. Finalmente, estas coordenadas fueron incorporadas en ambas bases de datos, asegurando su disponibilidad para análisis posteriores el codigo utilizado se muestra en la *Figura 8*.

```
##estanzarizar direcciones

lat=array(NA,913731)
long=array(NA,913731)

for(i in 86:913731){
dirs=str_replace_all(string = inf19_20$DIRECCION[i],pattern = "con",replacement = " ")
geo_dirs=geocode_OSM(q = paste(dirs,", Cali"))
if(length(geo_dirs)>0){
lat[i]=geo_dirs$coords[1]
long[i]=geo_dirs$coords[2]
}
print(i)
}

##pegar las coordenadas en la BD
inf19_20$lat=lat
inf19_20$long=long
```

Figura 8. *Código de la geocodificación de las direcciones en las bases de datos.*

Los casos con coordenadas fuera de Colombia fueron identificados y corregidos manualmente, dado que eran pocos, aproximadamente 30 registros. Entre estos se encontraron ubicaciones erróneas en países como Estados Unidos y Costa Rica. La limpieza de estos datos se realizó asegurando que todas las coordenadas correspondieran a ubicaciones dentro del territorio colombiano.

5.2 CORRELACIÓN ENTRE ACCIDENTES E INFRACCIONES

En esta sección se busca identificar patrones geográficos que relacionan la ubicación de accidentes mortales con zonas de alta densidad de infracciones específicas a través de la realización de mapas, evaluando así la viabilidad de un modelo predictivo de riesgo vial.

5.2.1 RELACIÓN ENTRE INFRACCIONES Y ACCIDENTES MORTALES

Primero se busca una relación general entre todas las infracciones y los accidentes mortales para lo cual se realiza una división de la ciudad de cali en diferentes zonas

5.2.1.1 ZONA SUR

En la zona sur se identificó que las infracciones y accidentes de tránsito comparten ubicaciones específicas, como la Avenida Simón Bolívar y un tramo de la calle 48, donde se encuentran nuevos conjuntos residenciales en el sector de Valle del Lili como se observa en los mapas de la *Figura 9*.

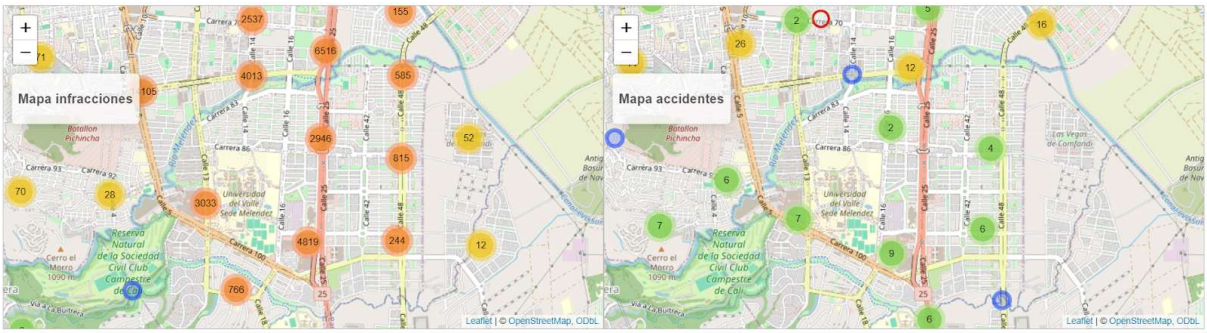

Figura 9. *Mapa de la zona sur mostrando puntos en común entre infracciones y accidentes de tránsito.*

5.2.1.2 ZONA NORTE

En la zona norte, la calle 70, que conecta con la Avenida Simón Bolívar, es otro punto donde tanto infracciones como accidentes de tránsito coinciden en ubicación como se observa en la *Figura 10*.

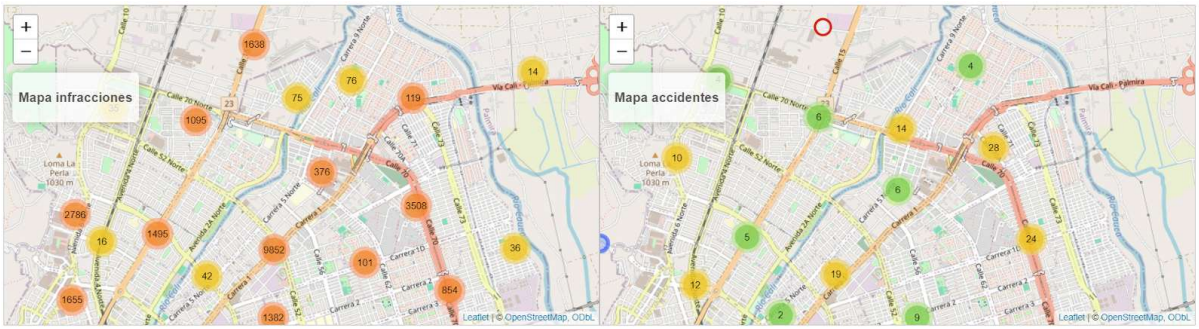

Figura 10. *Figura representativa de la zona norte relacionando accidentes con infracciones.*

5.2.1.3 ZONA CENTRO

En el centro de la ciudad, se evidencia que la carrera 23 (Autopista Sur) presenta una alta concentración de infracciones y accidentes de tránsito, compartiendo en gran parte las mismas ubicaciones *Figura 11*.

Figura 11. *Correlación de la zona centro de la ciudad entre accidentes e infracciones.*

Dado que se ha identificado una correlación espacial entre infracciones y accidentes de tránsito, se llevará a cabo un análisis detallado para determinar qué tipo de infracción tiene mayor relación con los accidentes.

5.2.2 RELACIÓN ENTRE LA INFRACCIÓN DE EXCEDER LA VELOCIDAD PERMITIDA (C29) Y LOS ACCIDENTES

El exceso de velocidad (infracción C29, según Tabla 1) es un factor crítico asociado a la severidad de los accidentes, especialmente en zonas urbanas. Estudios previos indican que un aumento del 5% en la velocidad promedio eleva un 15% el riesgo de accidentes mortales [20].

5.2.2.1 ZONA SUR

En la zona sur no se encuentra mayor correlación entre los datos de infracciones y accidentes a excepción de los encontrados sobre la calle 25 como se observa en la *Figura 12*.

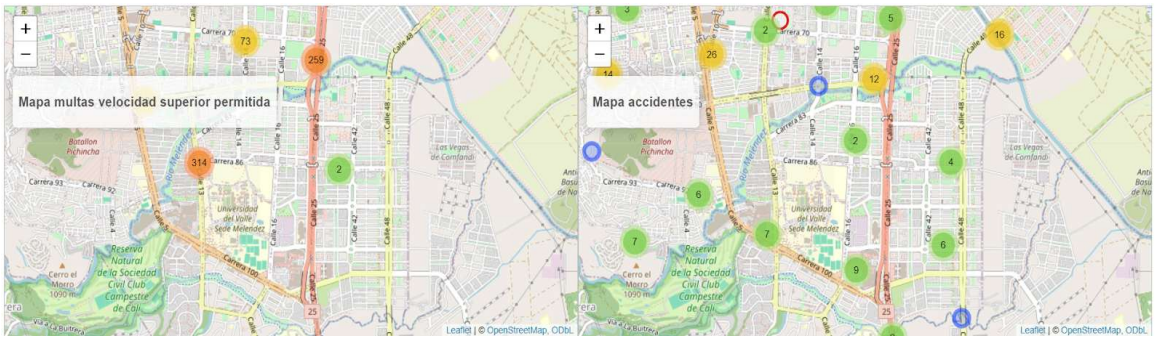

Figura 12. *Figura de accidentes de tránsito e infracciones por exceso de velocidad zona sur.*

5.2.2.2 ZONA NORTE

En la zona norte se encuentra una fuerte correlación sobre la calle 70 este comportamiento se puede observar en las *Figura 13.*

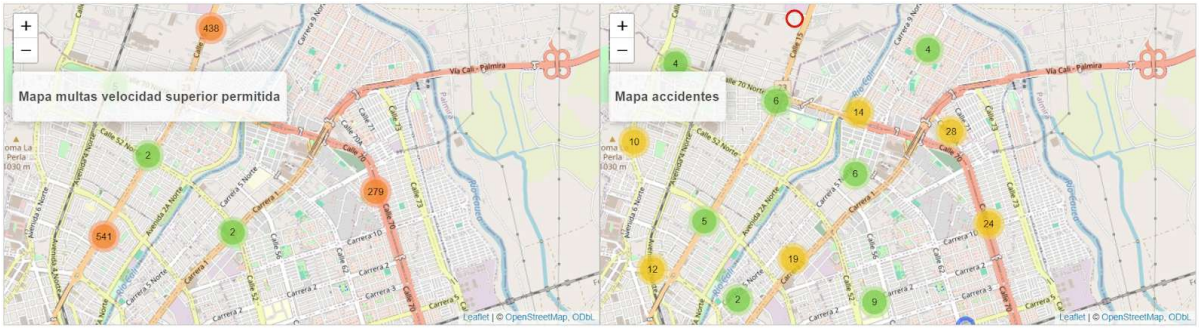

Figura 13. *Figura de accidentes de tránsito e infracciones por exceso de velocidad zona norte.*

5.2.2.3 ZONA CENTRO

En la *Figura 14* se evidencian las relaciones en la zona centro y aunque se nota un patron mas grande de accidentes mortales las relaciones con las infracciones se ven distribuidas sobre la autopista sur.

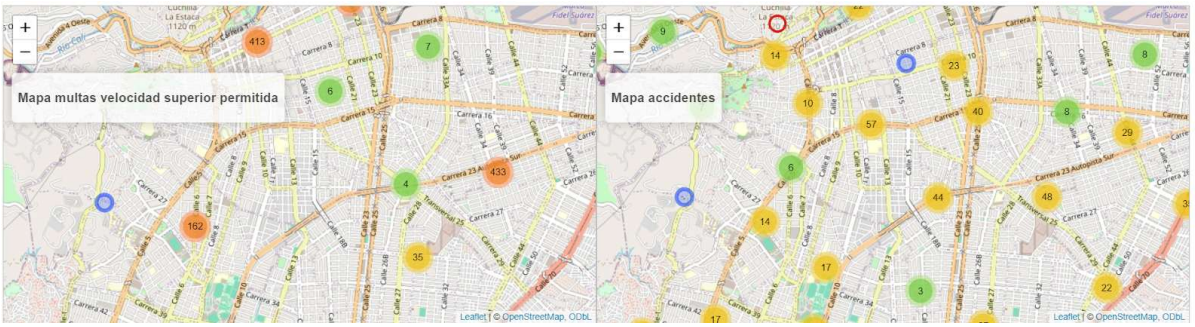

Figura 14. *Relación entre accidentes de tránsito y multas por exceso de velocidad zona centro.*

5.2.3 RELACIÓN ENTRE LA INFRACCIÓN POR NO REALIZAR LA REVISIÓN TÉCNICO-MECÁNICA (C35) Y LOS ACCIDENTES

La falta de revisión técnico-mecánica (infracción C35, según Tabla 1) es una de las infracciones más frecuentes en el dataset, con 161,556 registros, y está asociada a fallas mecánicas que pueden provocar accidentes.

5.2.3.1 ZONA SUR

En la zona sur se encontró una fuerte relación entre esta infracción y los accidentes de tránsito. Puntos críticos incluyen la intersección de la Avenida 5 con la Carrera 100, sectores de Valle del Lili y un tramo de la Carrera 80 como se observa en la *Figura 15*.

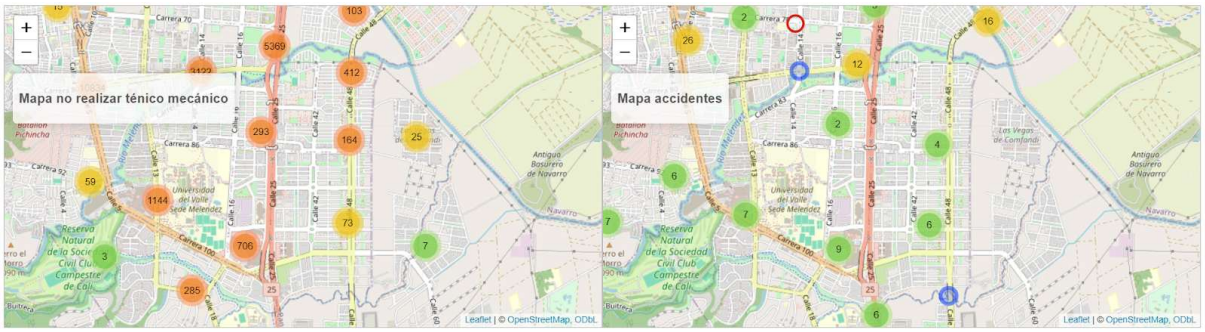

Figura 15. *Relación entre no realizar revisión técnico-mecánica y accidentes zona sur*.

5.2.3.2 ZONA NORTE

En la zona norte, la calle 70 es una vía donde se identificó una alta coincidencia entre infracciones por no realizar la revisión técnico-mecánica y accidentes de tránsito, evidenciando una relación significativa en esta parte de la ciudad como se ve en la *Figura 16*.

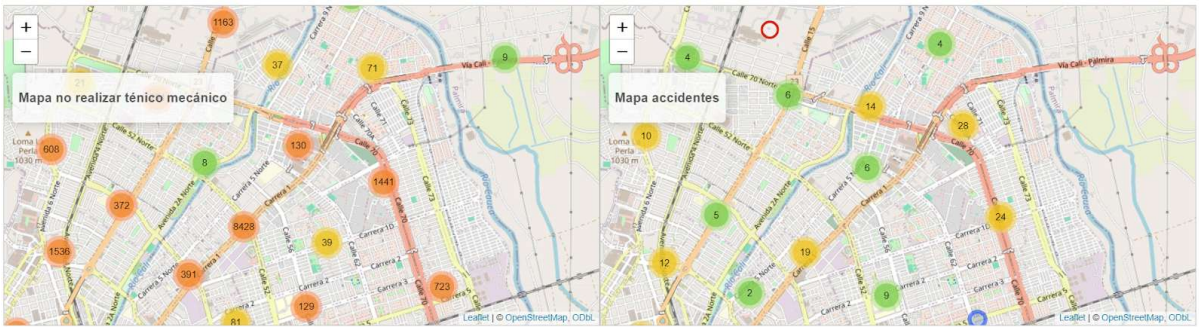

Figura 16. *Relación entre no realizar revisión técnico-mecánica y accidentes zona norte.*

5.2.3.3 ZONA CENTRO

En el centro de Cali, se observó una alta relación entre accidentes de tránsito y la falta de revisión técnico-mecánica en puntos como la Carrera 8, Carrera 15, Carrera 23 (Autopista Sur), y la intersección entre la Carrera 29 y la Calle 7. Esto sugiere que el incumplimiento de los mantenimientos preventivos en los vehículos podría estar contribuyendo a la ocurrencia de siniestros viales en estas áreas, comportamiento visible en la *Figura 17*.

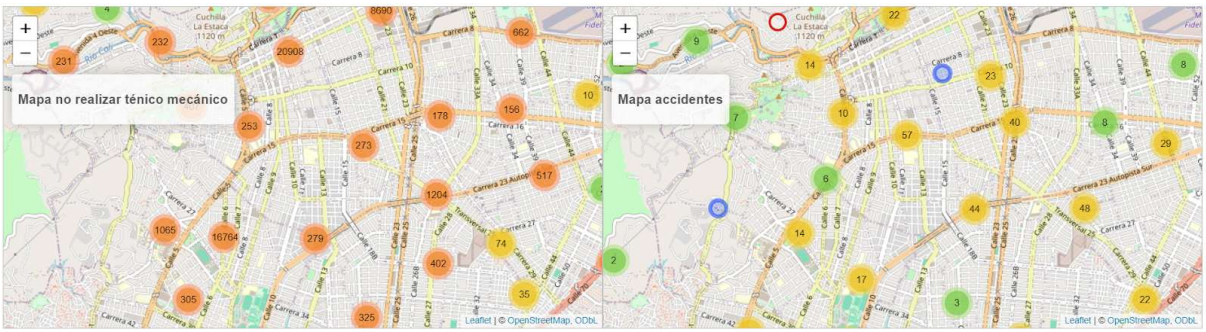

Figura 17. *Relación entre no realizar revisión técnico-mecánica y accidentes zona centro.*

5.2.4 RELACIÓN ENTRE LA INFRACCIÓN DE NO RESPETAR SEMÁFOROS O SEÑALES DE PARE (D04) Y LOS ACCIDENTES

La infracción por no respetar semáforos o señales de PARE (código D04, según Tabla 1) representa una de las transgresiones más peligrosas, con 35,937 registros en el período de estudio.

5.2.4.1 ZONA SUR

Al analizar la relación entre los accidentes y la infracción de cruzar semáforos en rojo o no detenerse en señales de pare, se identificaron puntos críticos en la intersección de la Calle 5 con Carrera 100 y en la Calle 48, en el sector de Valle del Lili. En estas ubicaciones, se evidencia una alta coincidencia entre los siniestros viales y este tipo de infracción cometida por los conductores como se ve en la *Figura 18.*

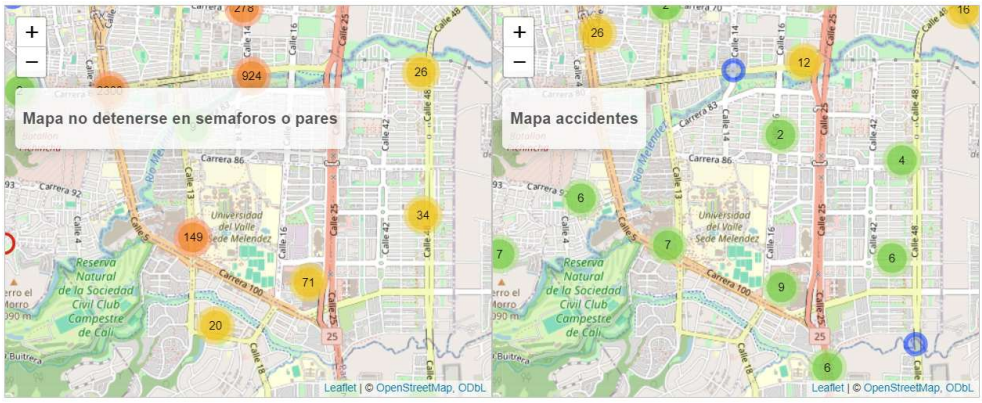

Figura 18. *Relación entre no respetar semáforos o señales de pare y accidentes zona sur.*

5.2.4.2 ZONA NORTE

En la zona norte, específicamente en la calle 70, se identificó una alta coincidencia entre las infracciones por no respetar los semáforos y los señalamientos de PARE, lo que evidencia un patrón de comportamiento significativo en esta área de la ciudad como se observa en la *Figura 19*.

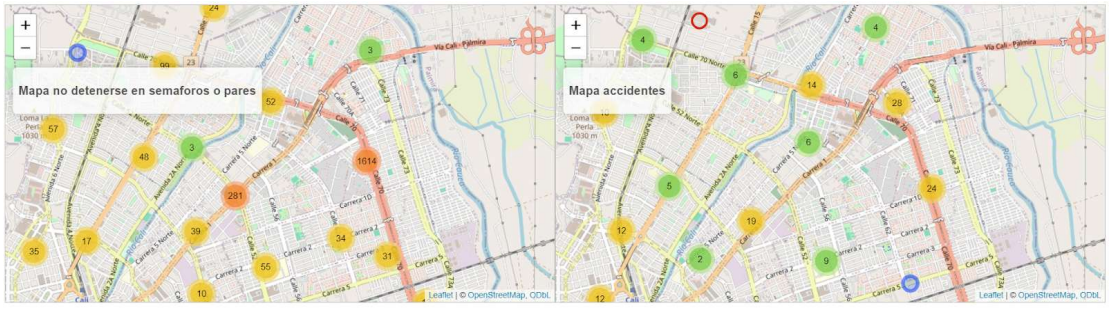

Figura 19. *Relación entre no respetar semáforos o señales de pare y accidentes zona norte.*

5.2.4.3 ZONA CENTRO

En la zona centro se observa una clara correlación entre los puntos de infracción por no respetar semáforos y señales de pare, y la ubicación de accidentes de tránsito. Este patrón es particularmente evidente a lo largo de la Carrera 23 (Autopista 23), la Carrera 15 y un segmento de la Calle 70, donde se concentran ambos tipos de eventos esto se puede observar en la *Figura 20*.

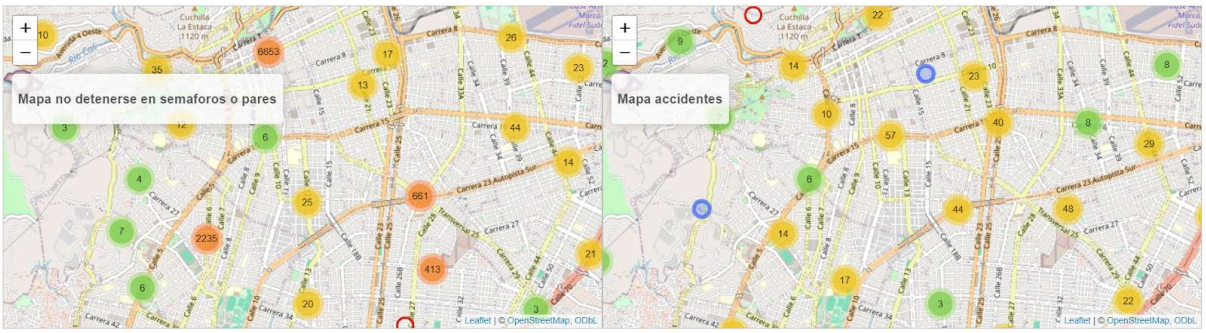

Figura 20. *Relación entre no respetar semáforos o señales de pare y accidentes zona centro.*

5.2.5 RELACIÓN ENTRE LA INFRACCIÓN CONDUCIR MOTO SIN OBSERVAR NORMAS (C24) Y LOS ACCIDENTES.

La infracción de conducir motocicleta sin observar las normas muestra una relación significativa con los accidentes reportados, particularmente en zonas donde los motociclistas han sido identificados como grupo vulnerable (Capítulo 4.2). Este vínculo resulta especialmente relevante para el análisis de seguridad vial.

5.2.5.1 ZONA SUR

En la zona sur, esta correlación se hace evidente a lo largo de la Calle 48, específicamente en el sector del Valle del Lili, y en un tramo de la Calle 25 (Avenida Simón Bolívar). Estos sectores concentran tanto las infracciones por mal manejo de motos como los accidentes registrados, lo que refuerza la necesidad de intervenciones focalizadas en dichas vías como se ve en la *Figura 21*.

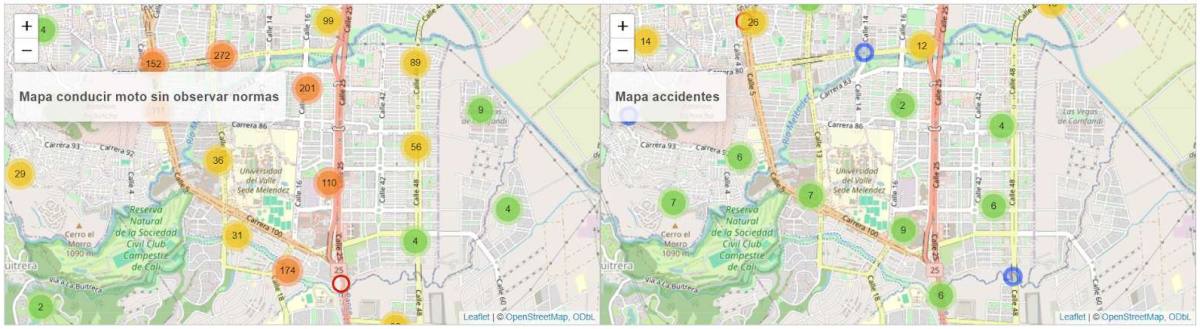

Figura 21. *Relación entre conducir moto sin observar las normas y accidentes zona sur.*

5.2.5.2 ZONA NORTE

En la zona norte se identifican puntos críticos donde coinciden las infracciones por conducir motocicleta sin observar las normas y los accidentes reportados. Esta relación se hace evidente en sectores específicos de la Calle 70, la Carrera 1 y la Calle 15, mostrando una correlación significativa entre ambos tipos de eventos como se observa en la *Figura 22*.

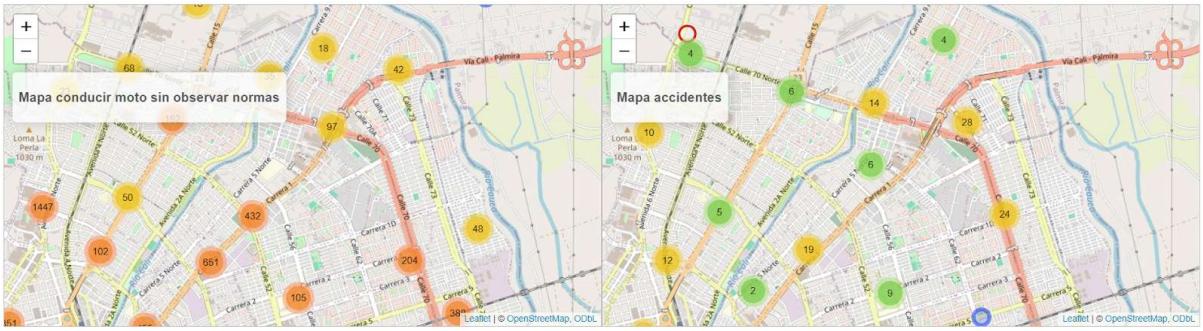

Figura 22. *Relación entre conducir moto sin observar las normas y accidentes zona norte.*

5.2.5.3 ZONA CENTRO

En la zona centro se identifican puntos críticos donde coinciden las infracciones por conducir motocicleta sin observar las normas y los accidentes de tránsito. Esta relación es particularmente evidente en sectores específicos de la Carrera 23 (Autopista Sur), la Carrera 15 y la Calle 5, donde se concentran ambos tipos de incidentes como se ve en la *Figura 23*.

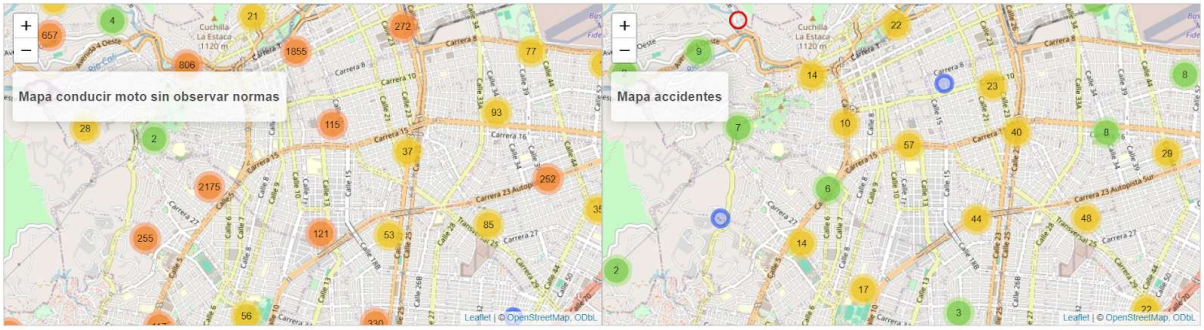

Figura 23. *Relación entre conducir moto sin observar las normas y accidentes zona centro.*

5.2.6 RELACIÓN ENTRE LA INFRACCIÓN NO ACATAR LAS SEÑALES DE TRÁNSITO (C31) Y LOS ACCIDENTES.

Esta infracción representa un riesgo significativo para la seguridad vial, ya que el incumplimiento de señales de tránsito (semáforos, pare, ceda el paso, entre otras) incrementa la probabilidad de colisiones y atropellos, especialmente en zonas con alta densidad vehicular y peatonal. Su análisis por zonas revela patrones claros de correlación con accidentes graves.

5.2.6.1 ZONA SUR

Se observa una clara relación entre las infracciones por no acatar las señales de tránsito y los accidentes reportados en varios sectores estratégicos de la ciudad. Este patrón es particularmente evidente a lo largo de la Calle 48 (Valle del Lili), la Calle 25, la Carrera 100 y la Calle 5, donde se concentran ambos tipos de incidentes como se observa en la *Figura 24*.

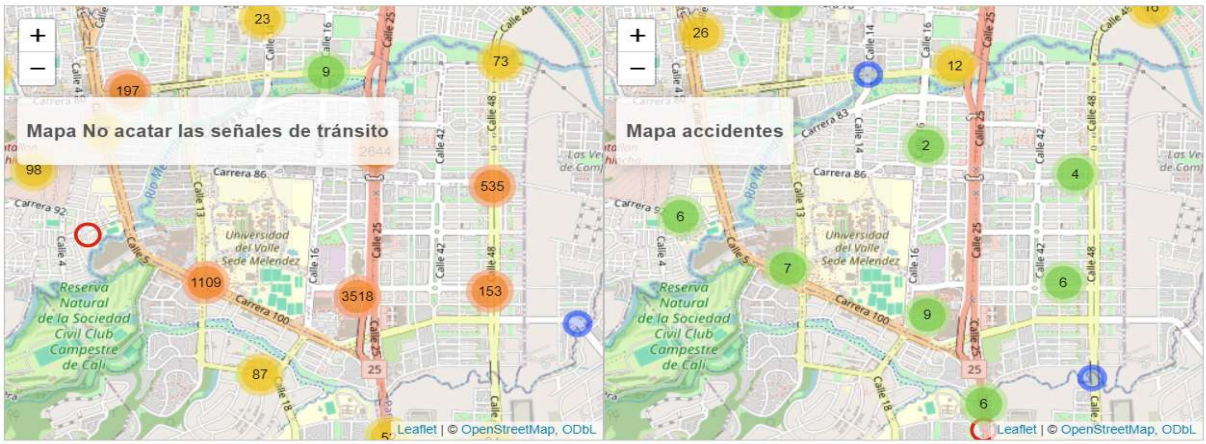

Figura 24. *Relación entre no acatar las señales de tránsito y accidentes zona sur.*

5.2.6.2 ZONA NORTE

En la zona norte se ha identificado una relación directa entre las infracciones por no acatar las señales de tránsito y la ocurrencia de accidentes viales. Este patrón se manifiesta con mayor intensidad en sectores específicos de la Calle 70 (Valle del Lili), la Carrera 1, la Avenida 6 Norte y la Calle 15 y se visualiza en la *Figura 25*.

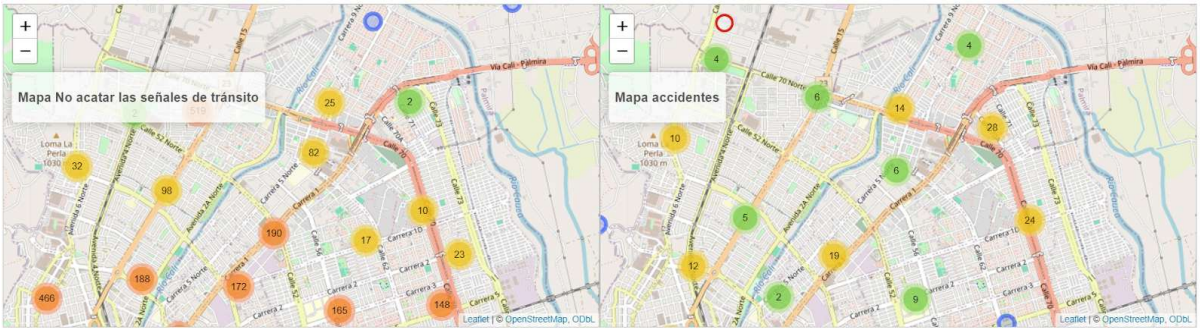

Figura 25. *Relación entre no acatar las señales de tránsito y accidentes zona norte.*

5.2.6.3 ZONA CENTRO

En el análisis realizado se observa una relación directa entre las infracciones por no respetar las señales de tránsito y los accidentes reportados en la zona. Esta correlación se hace especialmente evidente en puntos específicos ubicados a lo largo de la Carrera 23 (Autopista Sur), la Carrera 15 y la Calle 5 esto se muestra en la *Figura 26*.

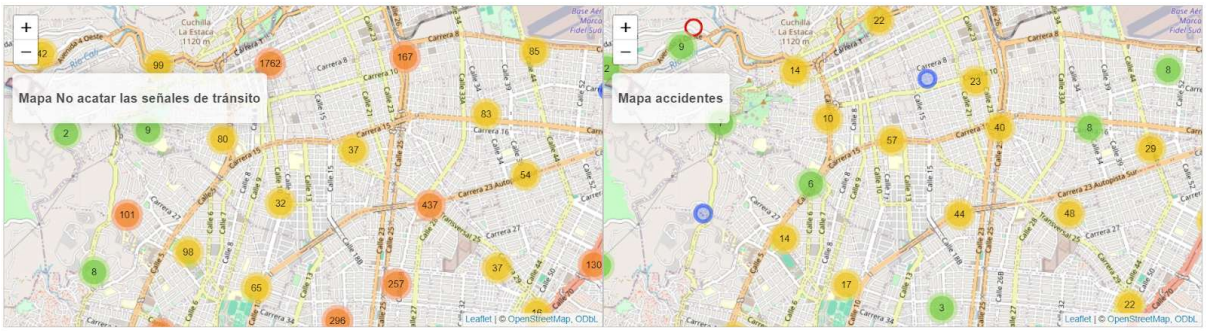

Figura 26. *Relación entre no acatar las señales de tránsito y accidentes zona centro.*

5.2.7 RELACIÓN ENTRE LA INFRACCIÓN NO RESPETAR EL PASO DE PEATONES QUE CRUZAN UNA VÍA EN SITIO PERMITIDO PARA ELLOS (C32) Y LOS ACCIDENTES.

Esta infracción resulta particularmente relevante, ya que en el capítulo 4.2 se identificó a los peatones como uno de los actores viales más vulnerables. Este hallazgo permite relacionar las zonas donde se comete frecuentemente esta infracción con áreas potencialmente peligrosas para este grupo de usuarios de la vía.

5.2.7.1 ZONA SUR

En la zona sur, el análisis muestra que esta infracción no presenta una alta coincidencia espacial con los accidentes reportados. Sin embargo, se registran algunos puntos críticos, específicamente 7 muertes en los alrededores de la Universidad del Valle y 2 muertes en el sector comprendido entre la Calle 25, Calle 26 y Carrera 85, donde sí se evidencia esta relación como se ve en la *Figura 27*.

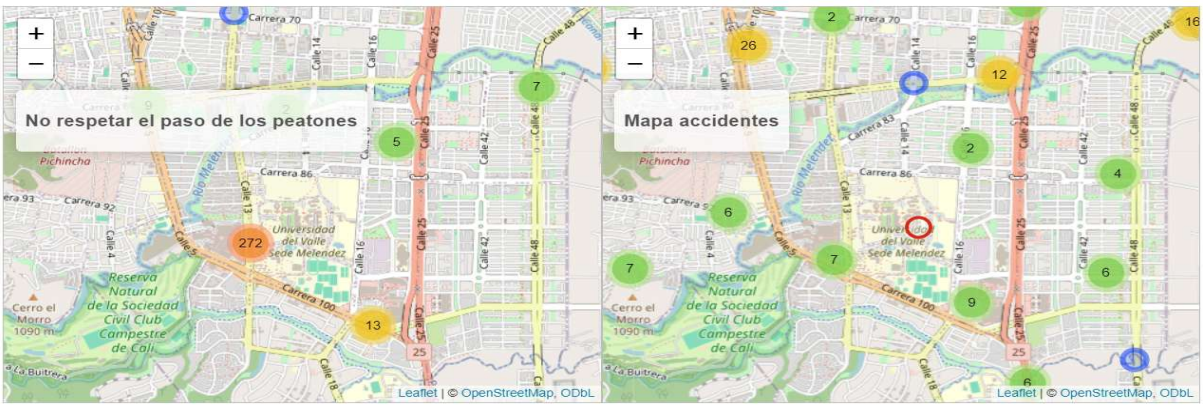

Figura 27. *Relación entre* no respetar el paso de peatones que cruzan una vía en sitio permitido para ellos *y accidentes zona sur.*

5.2.7.2 ZONA NORTE

En la zona norte se observa un comportamiento similar al detectado en la zona sur, donde la relación entre infracciones y accidentes muestra una coincidencia limitada. El análisis revela que solo algunos puntos específicos presentan esta correlación, particularmente a lo

largo de la Calle 70, donde se concentran ambos tipos de incidentes como se evidencia en el mapa de la *Figura 28*.

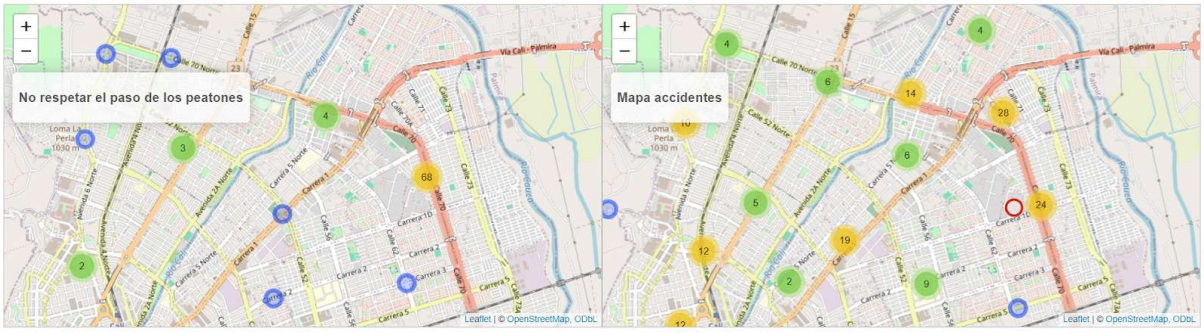

Figura 28. *Relación entre* no respetar el paso de peatones que cruzan una vía en sitio permitido para ellos *y accidentes zona norte.*

5.2.7.3 ZONA CENTRO

En esta zona particular, el análisis muestra una escasa coincidencia entre los puntos de infracción y los accidentes reportados. Solo se logra identificar un punto específico donde ambos fenómenos convergen, ubicado en el sector de la Carrera 23 (Autopista Sur) con la Calle 33A como se ve en la *Figura 29*.

Figura 29. *Relación entre* no respetar el paso de peatones que cruzan una vía en sitio permitido para ellos *y accidentes zona centro.*

5.2.8 RELACIÓN ENTRE LA INFRACCIÓN TRANSITAR EN SENTIDO CONTRARIO AL ESTIPULADO PARA LA VÍA (D03) Y LOS ACCIDENTES.

Esta infracción reviste especial importancia por los riesgos que genera la circulación indebida en carretera, tanto para otros conductores como para peatones. Al analizar su distribución por zonas, se observan patrones específicos de coincidencia entre los puntos de infracción y los accidentes registrados.

5.2.8.1 ZONA SUR

En la zona sur, particularmente, se identifican sectores críticos donde convergen las infracciones por transitar en sentido contrario y los accidentes de tránsito. Estos puntos

problemáticos se localizan principalmente en la Calle 48 (Valle del Lili), la Calle 25 (Avenida Simón Bolívar) y en el cruce de la Carrera 70 con Calle 13 (Avenida Pasoancho), donde se evidencia una clara correlación espacial entre ambos fenómenos como se observa en la *Figura 30*.

Figura 30. *Relación entre transitar en sentido contrario al estipulado para la vía y accidentes zona sur.*

5.2.8.2 ZONA NORTE

En la zona norte del estudio, se ha identificado una relación significativa entre las infracciones por transitar en sentido contrario y la ocurrencia de accidentes de tránsito. Este patrón se manifiesta con mayor claridad en dos ejes viales principales: la Calle 70 y la Carrera 1, donde se concentran tanto las infracciones registradas como los incidentes reportados como se observa en la *Figura 31*.

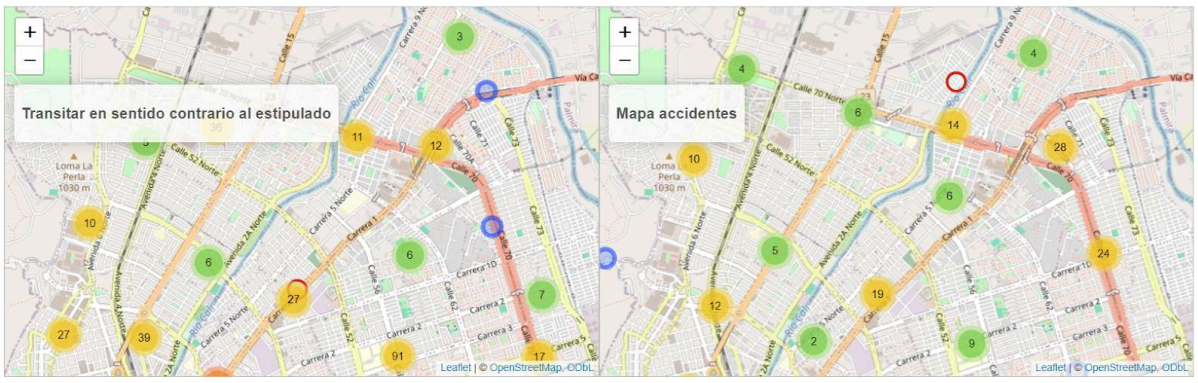

Figura 31. *Relación entre transitar en sentido contrario al estipulado para la vía y accidentes zona norte.*

5.2.8.3 ZONA CENTRO

En la zona centro del análisis se identifican puntos críticos donde coinciden las infracciones y los accidentes de tránsito. Esta relación se hace evidente principalmente a lo largo de la Carrera 23 (Autopista Sur), donde se concentra la mayor parte de los incidentes. Adicionalmente, se observan algunos puntos aislados de coincidencia en el sector de la Carrera 15 con Calle 23 comportamiento visto a través de la *Figura 32*.

Figura 32. *Relación entre transitar en sentido contrario al estipulado para la vía y accidentes zona centro.*

5.2.9 RELACIÓN ENTRE LA INFRACCIÓN NO UTILIZAR EL CINTURÓN DE SEGURIDAD POR PARTE DE LOS OCUPANTES DEL VEHÍCULO (C06) Y LOS ACCIDENTES.

El no uso o uso incorrecto del cinturón de seguridad incrementa significativamente el riesgo para los ocupantes del vehículo, pudiendo derivar en consecuencias fatales. A continuación se presenta la distribución de esta problemática según las zonas analizadas.

5.2.9.1 ZONA SUR

En la zona sur se ha identificado una relación significativa entre las infracciones reportadas y los accidentes mortales registrados. Esta correlación se manifiesta con mayor intensidad a lo largo de la Calle 25 (Avenida Simón Bolívar), donde se concentran ambos tipos de incidentes como se muestra en la *Figura 33*.

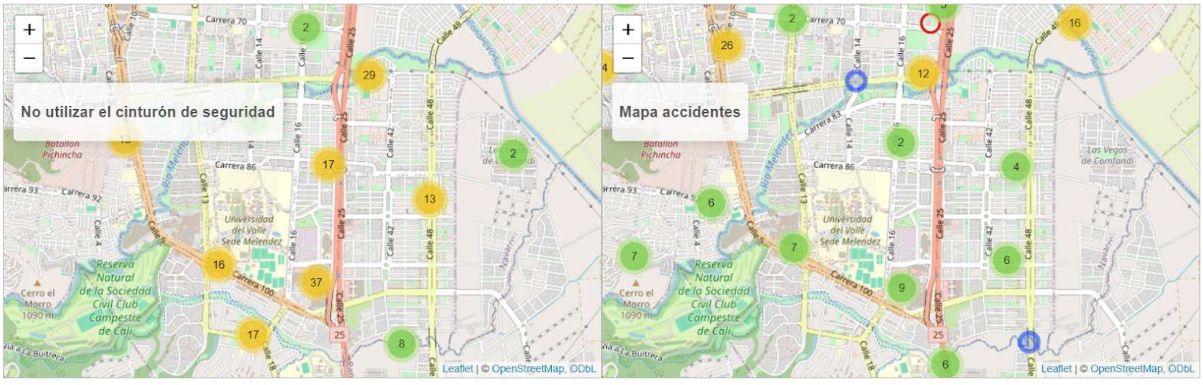

Figura 33. *Relación entre no utilizar el cinturón de seguridad y accidentes zona sur.*

5.2.9.2 ZONA NORTE

En esta zona se ha identificado una correlación espacial entre la comisión de estas infracciones y la ocurrencia de accidentes mortales. Este patrón se manifiesta principalmente en dos ejes viales críticos: la Calle 70 y la Carrera 1, donde convergen tanto los reportes de infracción como los incidentes con víctimas fatales esto se observa en la

Figura 34.

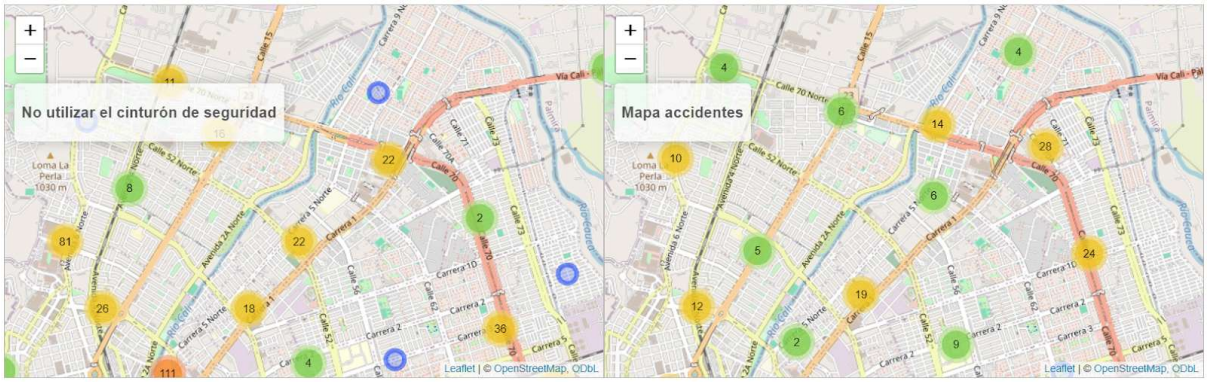

Figura 34. *Relación entre no utilizar el cinturón de seguridad y accidentes zona norte.*

5.2.9.3 ZONA CENTRO

En la zona centro del estudio se ha detectado una relación significativa entre este tipo de infracciones y la ocurrencia de accidentes mortales. Los puntos de mayor coincidencia se concentran principalmente en la Carrera 23 (Autopista Sur) y la Carrera 15, donde se evidencia una clara correlación espacial entre ambos fenómenos como se observa en la *Figura 35*.

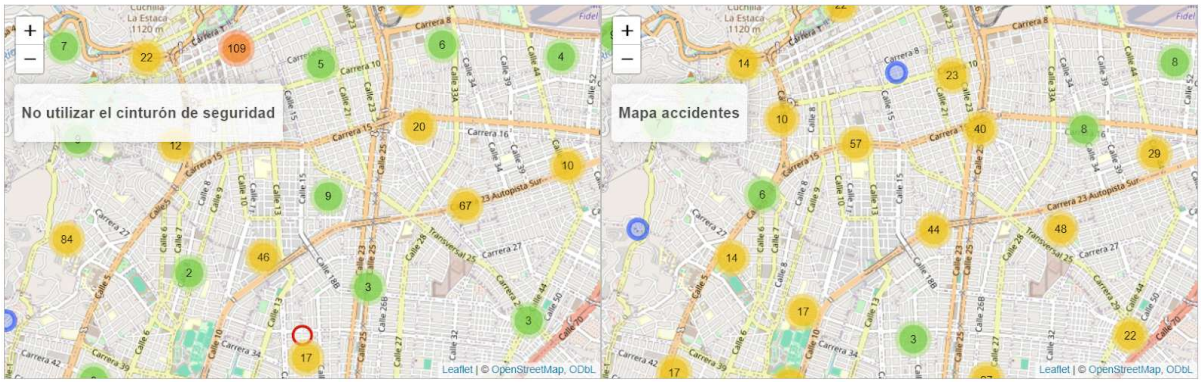

Figura 35. *Relación entre no utilizar el cinturón de seguridad y accidentes zona centro.*

5.2.10 RELACIÓN ENTRE LA INFRACCIÓN CONDUCIR REALIZANDO MANIOBRAS ALTAMENTE PELIGROSAS (D07) Y LOS ACCIDENTES.

Esta infracción adquiere especial relevancia debido a que la realización de maniobras peligrosas incrementa considerablemente la probabilidad de ocurrencia de accidentes de tránsito. El análisis por zonas permite identificar sectores críticos donde este comportamiento se asocia directamente con incidentes viales.

5.2.10.1 ZONA SUR

En el caso específico de la zona sur, se observa una clara correlación entre estos puntos de infracción y los accidentes registrados, particularmente en la Calle 48 (Valle del Lili), la Calle 25 (Avenida Simón Bolívar) y en algunos tramos de la Carrera 100. Estos sectores concentran tanto las maniobras peligrosas reportadas como los accidentes derivados de dichas prácticas como se observa en la *Figura 36*.

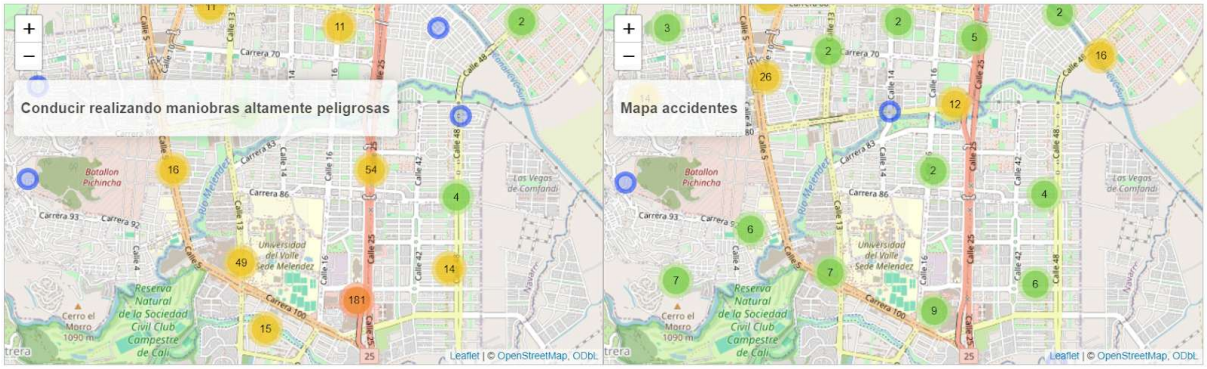

Figura 36. *Relación entre* conducir realizando maniobras altamente peligrosas *y accidentes zona sur.*

5.2.10.2 ZONA NORTE

En esta zona en particular, se ha identificado una coincidencia espacial entre las infracciones reportadas y los accidentes mortales registrados. Este patrón se manifiesta principalmente a lo largo de dos ejes viales críticos: la Calle 70 y la Carrera 1, donde convergen ambos tipos de eventos *Figura 37*.

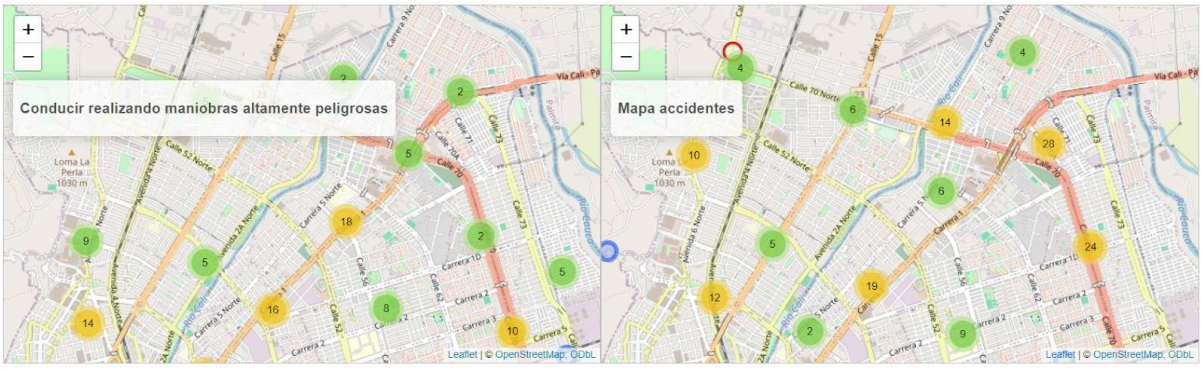

Figura 37. *Relación entre* conducir realizando maniobras altamente peligrosas *y accidentes zona norte.*

5.2.10.3 ZONA CENTRO

En la zona centro del estudio se ha detectado una relación significativa entre este tipo de infracciones y la ocurrencia de accidentes mortales. Esta correlación se concentra principalmente en el sector de la Carrera 23 (Autopista Sur), donde se observa una coincidencia espacial entre los puntos de infracción reportados y los incidentes con víctimas fatales como se ve en la *Figura 38*.

Figura 38. *Relación entre* conducir realizando maniobras altamente peligrosas *y accidentes zona centro.*

5.2.11 RELACIÓN ENTRE LA INFRACCIÓN CONDUCIR UN VEHÍCULO SOBRE ACERAS, PLAZAS, VÍAS PEATONALES, SEPARADORES, BERMAS (D05) Y LOS ACCIDENTES.

Como se detalló en el numeral 5.2.7, esta infracción involucra situaciones que comprometen la seguridad peatonal, siendo este grupo particularmente vulnerable según lo establecido en el capítulo 4.2. Su impacto en la movilidad urbana requiere especial atención.

5.2.11.1 ZONA SUR

En la zona sur, el análisis muestra una escasa coincidencia entre los puntos de infracción y los accidentes registrados. Solo se identifican algunas correlaciones aisladas, principalmente en sectores de la Calle 25, con presencia menor en la Carrera 100 y la Calle 5, donde se registran contados casos que vinculan ambos fenómenos esto se puede evidenciar en la *Figura 39.*

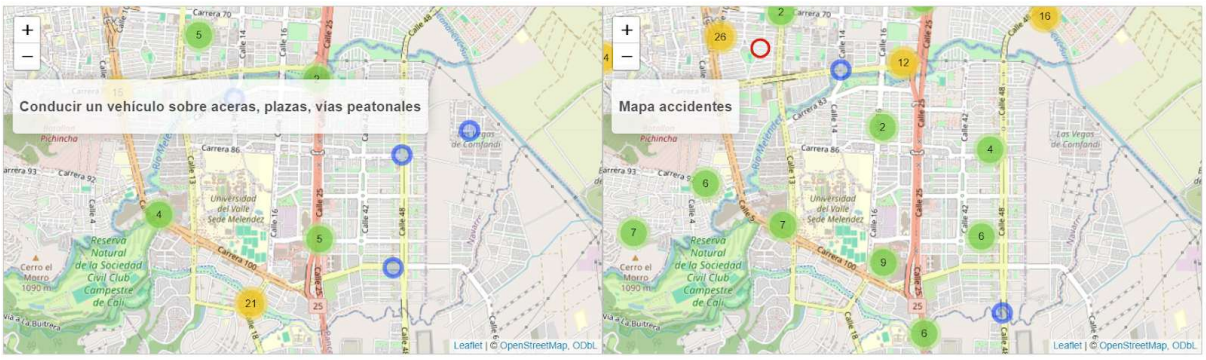

Figura 39. *Relación entre* conducir un vehículo sobre aceras, plazas, vías peatonales, separadores, bermas *y accidentes zona sur.*

5.2.11.2 ZONA NORTE

En la zona norte se observa una baja coincidencia entre esta infracción y los accidentes reportados. El análisis revela que únicamente se presentan algunos puntos de correlación

en sectores específicos de la Carrera 1 y la Calle 15, donde se registra simultáneamente la ocurrencia de ambas situaciones como se ve en la *Figura 40.*

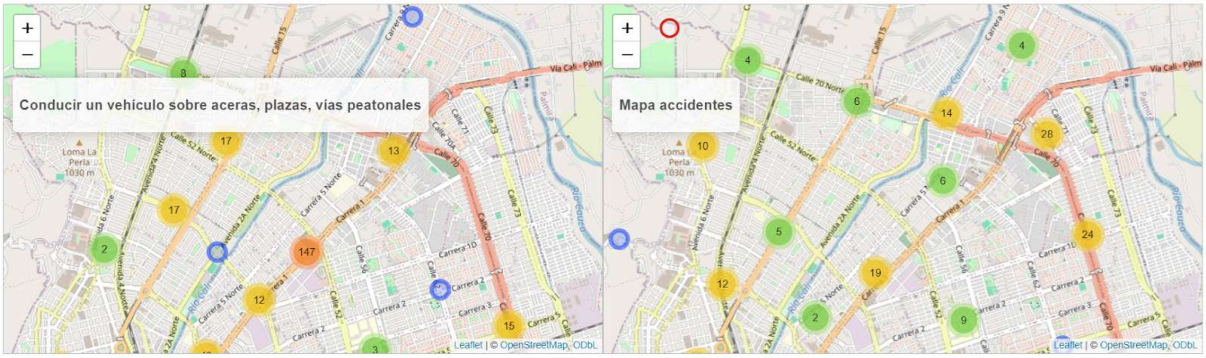

Figura 40. *Relación entre* conducir un vehículo sobre aceras, plazas, vías peatonales, separadores, bermas *y accidentes zona norte.*

5.2.11.3 ZONA CENTRO

En la zona centro, el análisis muestra una limitada correlación entre esta infracción y los accidentes de tránsito reportados. Se identifican algunos puntos de coincidencia aislados, principalmente en sectores de la Carrera 23 (Autopista Sur) y la Carrera 15, donde convergen ambos tipos de incidentes como se ve en la *Figura 41*.

Figura 41. *Relación entre* conducir un vehículo sobre aceras, plazas, vías peatonales, separadores, bermas *y accidentes zona centro.*

Los mapas analizados revelan una clara relación espacial entre los puntos donde se cometen diversas infracciones de tránsito y los lugares donde ocurren accidentes mortales en la ciudad. Específicamente, se observa que la mayoría de estas infracciones coinciden geográficamente en ciertos corredores críticos: la Calle 48 (Valle del Lili), la Calle 25 (Avenida Simón Bolívar), la Carrera 100, la Calle 70, la Carrera 1, la Carrera 23 (Autopista Sur) y la Carrera 15. Esta coincidencia espacial sugiere que el análisis de las infracciones puede ser fundamental para comprender los patrones de ocurrencia de accidentes mortales en Cali.

Ante estos hallazgos, se realizó un cálculo de correlación espacial para cada tipo de infracción en relación con los accidentes mortales. Los resultados de este análisis cuantitativo, que confirman las observaciones iniciales, se presentan detalladamente en la

Tabla 6. Este proceso permitió validar estadísticamente lo observado en los mapas y establecer con mayor precisión el grado de asociación entre las variables estudiadas.

INFRACCIONES	CORRELACIÓN	P VALUE
D04	0,352	<0,001
C29	0,227	<0,001
C35	0,373	<0,001
C06	0,736	<0,001
C24	0,593	<0,001
C31	0,552	<0,001
C32	0,359	<0,001
D03	0,523	<0,001
D05	0,536	<0,001
D07	0,492	<0,001

Tabla 6. *Tabla de correlación espacial entre las infracciones y los accidentes mortales.*

En el capítulo 6 se utilizarán los datos recopilados en los capítulos 4 y 5 como base para el desarrollo de un modelo predictivo. Este modelo tendrá como objetivo principal estimar la probabilidad de ocurrencia de accidentes mortales en la ciudad, a partir de los patrones identificados previamente.

La información obtenida sobre infracciones de tránsito, características viales y distribución geográfica de accidentes servirá como insumo fundamental para entrenar el algoritmo. Este proceso permitirá establecer relaciones cuantitativas entre los factores de riesgo analizados y los accidentes con desenlace fatal.

6. ELABORACIÓN DEL MODELO

A partir de los análisis descriptivos (Capítulo 4) y geoespaciales (Capítulo 5) que buscaban demostrar relaciones entre las diferentes variables y los accidentes mortales en la ciudad de Cali, en este capítulo se desarrolla un modelo predictivo del riesgo vial con el objetivo de desarrollar la herramienta de alertas tempranas que funcione como insumo para proteger zonas de riesgo.

6.1 GENERACIÓN DE VARIABLES ESPACIALES (KERNEL DENSITY)

Para estimar la distribución espacial de los eventos, se generaron mapas de densidad kernel utilizando múltiples anchos de banda (σ) [21], los cuales determinan el área de influencia de cada punto en el cálculo de intensidad. Se evaluaron tres métodos para optimizar el ancho de banda primero el Método de Diggle (adecuado para patrones espaciales no homogéneos), Pseudo Verosimilitud Puntual (robusto frente a valores atípicos) y la regla de Scott (eficiente en grandes volúmenes de datos) [22].

Adicionalmente, se implementó un kernel adaptativo para ajustar dinámicamente el ancho de banda en zonas con dispersión irregular de datos. Estos mapas permitieron cuantificar la concentración espacial de los accidentes mortales y las infracciones para la inclusión de las variables en el modelo, un ejemplo de los mapas obtenidos se muestra en la *Figura 42*.

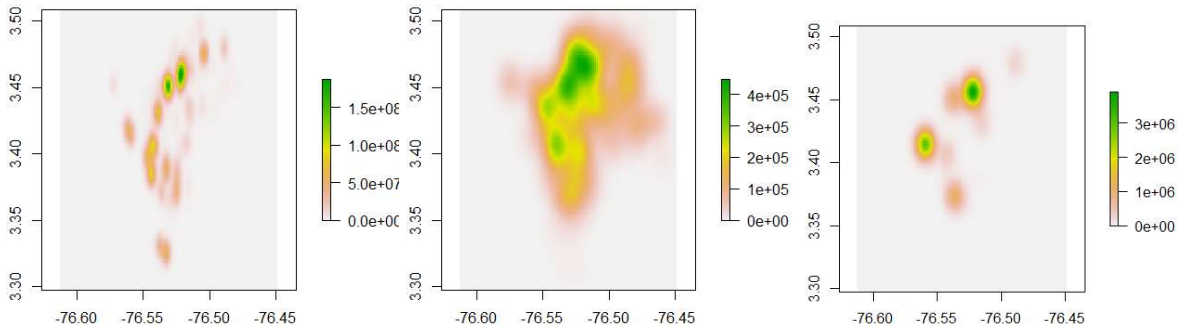

Figura 42. *Ejemplo de los mapas de densidad generados para algunas infracciones.*

6.2 VARIABLE DEPENDIENTE (ACCIDENTES MORTALES)

Una vez realizado el mapa de densidad de la variable dependiente, se procedió a graficar su densidad y su representación sobre la geografía de la ciudad, comprobando que coincide con los puntos de accidentes mortales registrados en la base de datos. Los mapas obtenidos se muestran en la *Figura 43*.

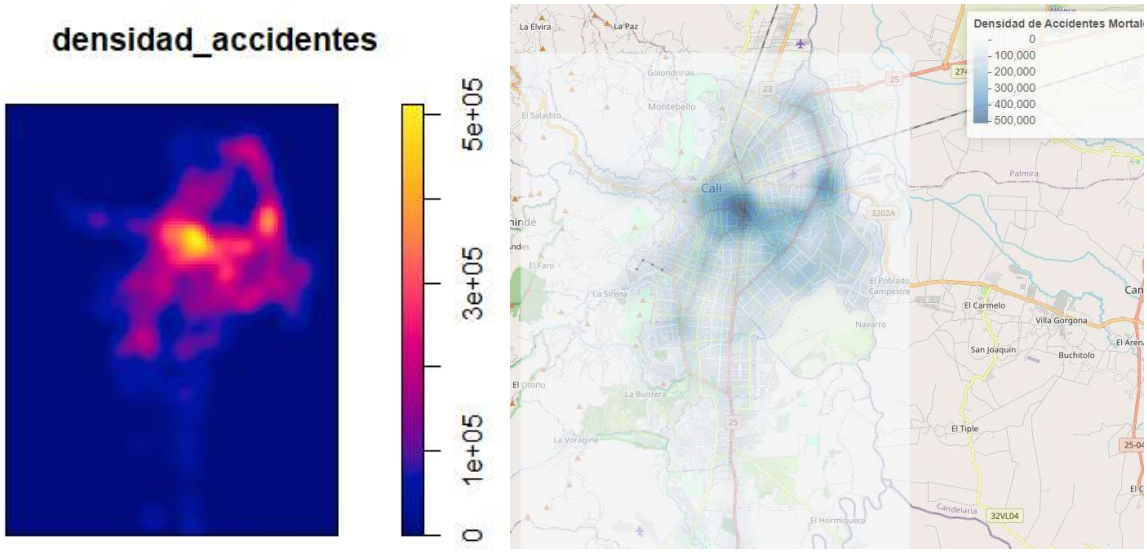

Figura 43. a) *densidad de accidentes mortales obtenida,* ***b)*** *densidad superpuesta sobre mapa de cali.*

Con esta variable, se procedió a realizar la función K y su respectiva linealización (función L), con el fin de identificar si la variable es aleatoria o no con respecto al espacio. Los resultados obtenidos se presentan en la *Figura 44*.

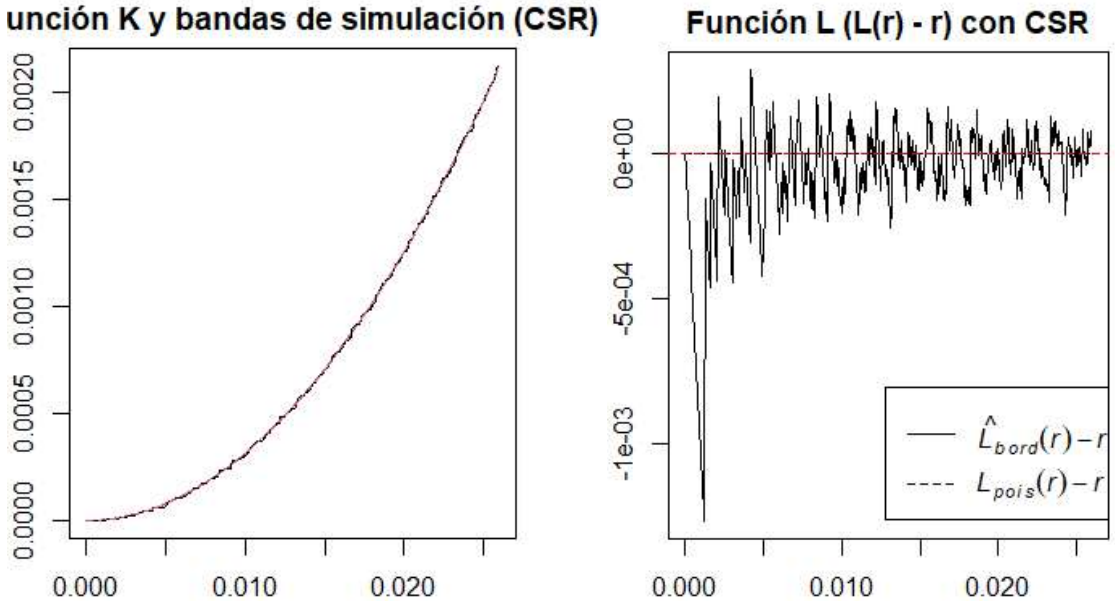

Figura 44. *Función K y función L para determinar la aleatoriedad espacial de los datos.*

De la gráfica de la *Figura 44*, es posible concluir que los accidentes mortales siguen un patrón aleatorio en el espacio, por lo que un modelo espacial no aportaría resultados significativos para describir su comportamiento. Este hallazgo resulta positivo, ya que una alta correlación espacial hubiera indicado que la ocurrencia de accidentes está altamente concentrada en zonas específicas, lo que requeriría intervenciones localizadas (como

mejoramiento de vías o mayor control normativo en esos puntos críticos) por lo que una y herramienta de predicción de accidentes mortales resulta menos significativa.

6.3 SELECCIÓN DE VARIABLES INDEPENDIENTES

Para la selección de las variables independientes se consideraron todas las infracciones de tránsito disponibles y los distintos tipos de vehículos, manteniéndose solo aquellas con sustento empírico y bajo nivel de colinealidad. Tras la normalización de los datos y la verificación de multicolinealidad (Figura 45), se descartaron las variables con correlación superior a 0.7 y aquellas redundantes en su representación espacial. El conjunto final se seleccionó en función de tres criterios: evidencia teórica, representatividad en los datos y relevancia en los modelos de predicción.

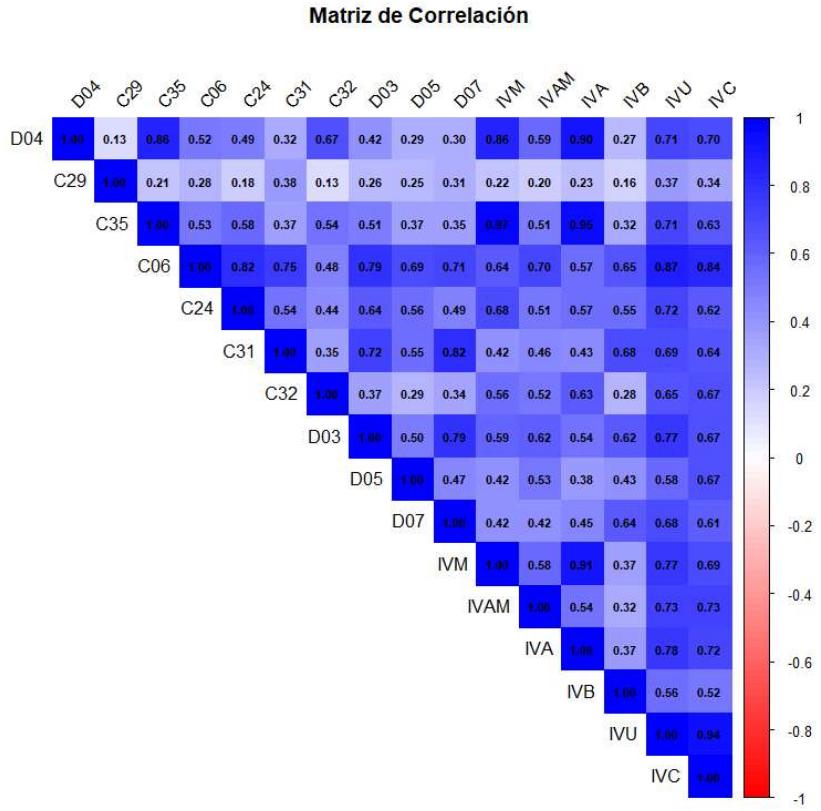

Figura 45. *Correlación entre todas las posibles variables a incluir en el modelo.*

Donde:
ivm - infracciones de vehículo moto.
ivam -infracciones de vehículo ambulancia.
iva infracciones de vehículo automóvil.
ivb - infracciones de vehículo bicicleta.
ivu - infracciones de vehículo bus.
ivc - infracciones de vehículo camión.

Para validar esta selección, se entrenaron modelos de Random Forest y XGBoost. La importancia de las variables se evaluó mediante la ganancia de Gini (en ambos modelos) y,

adicionalmente, con valores SHAP para el modelo de XGBoost, lo cual permite interpretar el aporte marginal de cada variable a las predicciones del modelo, los resultados para la importancia de las variables a partir del modelo random forest se observan en la *Figura 46* en la *Figura 47* podemos encontrar la importancia para el modelo XGboost y finalmente en la *Figura 48* se observan los valores de SHAP obtenidos.

```
> print(importancia_rf)
rf variable importance

            Overall
c06        100.0000
c24         54.8273
d05         41.4446
X35         17.8323
d07         10.6473
d03          7.2542
Ambulancia   6.2398
c31          5.9665
Bicicleta    3.9708
X29          3.5743
Camion       1.8683
X04          1.7105
c32          1.1672
Bus          0.4614
Automovil    0.0000
```

Figura 46. *Importancia de variables según la ganancia de Gini en el modelo Random Forest.*

```
> print(importance_matrix)
      Feature        Gain       Cover   Frequency
       <char>       <num>       <num>       <num>
 1:       c31 0.4567041501 0.132794580 0.213923133
 2:       c24 0.1541479838 0.050526885 0.079767948
 3:       c06 0.0985464808 0.080307965 0.048585932
 4:       X29 0.0757390445 0.226286574 0.264684554
 5:    Camion 0.0455961572 0.108429077 0.029731690
 6:       d07 0.0420015029 0.012852332 0.033357505
 7:       X35 0.0315625769 0.129724878 0.080493111
 8:       d03 0.0266570478 0.107859055 0.030456853
 9:       d05 0.0246011357 0.004010074 0.021754895
10:       Bus 0.0141678859 0.064343654 0.042784627
11: Bicicleta 0.0105840996 0.004462166 0.029006526
12: Automovil 0.0097652679 0.019580960 0.034807832
13:       c32 0.0059878210 0.001622301 0.008701958
14: Ambulancia 0.0030011140 0.001108492 0.006526468
15:       X04 0.0009377318 0.056091007 0.075416969
```

Figura 47. *Importancia de variables según la ganancia de Gini en el modelo XGboost.*

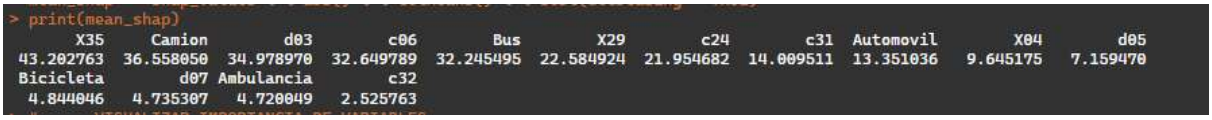

Figura 48. *Importancia de variables según valores SHAP en el modelo XGBoost.*

En Random Forest, la variable con mayor importancia fue C06 (no uso del cinturón de seguridad), seguida de C24 (infracción común en motociclistas) y D05 (conducir por zonas peatonales). En XGBoost, según la ganancia, se destacaron C31, C24 y C06, mientras que con SHAP las más influyentes fueron C35 (revisión técnico-mecánica) con un valor promedio de 43.2, Camión (36.5), D03 (34.9) y C06 (32.6).

A partir de los análisis realizados anteriormente se optó por eliminar todas las variables que presentaran un valor de colinealidad mayor a 0.7, manteniendo aquellas que resultan más significativas para cada modelo y conservando aquellas con mayor correlación según la *Tabla 6* además de las que tuvieran más datos en la base según la *Tabla 1* y la *Tabla 2*. El conjunto final de variables seleccionadas fue:

Infracciones: C29, C35, C06, C32, D05

Vehículos: Ambulancias y bicicletas

Esta selección es consistente con diversos factores asociados a los accidentes de tránsito:

Infracción C06: No uso del cinturón de seguridad: Fue la variable más importante en Random Forest (100% de importancia relativa) y una de las más influyentes en XGBoost (SHAP: 32.6). Su relación directa con la mortalidad está ampliamente respaldada por estudios como el de Yang et al., 2023 [23], que confirma su fuerte impacto en la severidad de las lesiones.

Infracción C29: Conducir un vehículo a velocidad superior a la máxima permitida. Como se mencionó en el Capítulo 5.2.2, esta infracción muestra una gran correlación con la mortalidad en accidentes de tránsito.

Infracción C35: No realizar la revisión técnico-mecánica y de emisiones contaminantes. Esta infracción está asociada a posibles fallas mecánicas en los vehículos que pueden resultar en accidentes, como problemas con el sistema de frenado.

Infracción C32: No respetar el paso de peatones que cruzan una vía en sitio permitido. Representa una gran problemática, ya que, como se mencionó en el Capítulo 4.2, los peatones son uno de los actores viales más vulnerables.

Infracción D05: Conducir por espacios peatonales: Importante tanto en Random Forest (41.4%) como en XGBoost. Este comportamiento pone en riesgo a peatones, otro grupo altamente vulnerable según Myhrmann et al., 2020 [24].

Vehículo ambulancia: Asociado a las altas velocidades y maniobras peligrosas que estos vehículos realizan en su operación diaria.

Vehículo bicicleta: Muestra gran relevancia, ya que los ciclistas también fueron identificados como uno de los grupos más vulnerables en el Capítulo 4.2.

Los resultados de la matriz de correlación resultante después de la selección de variables se observan en la *Figura 49*.

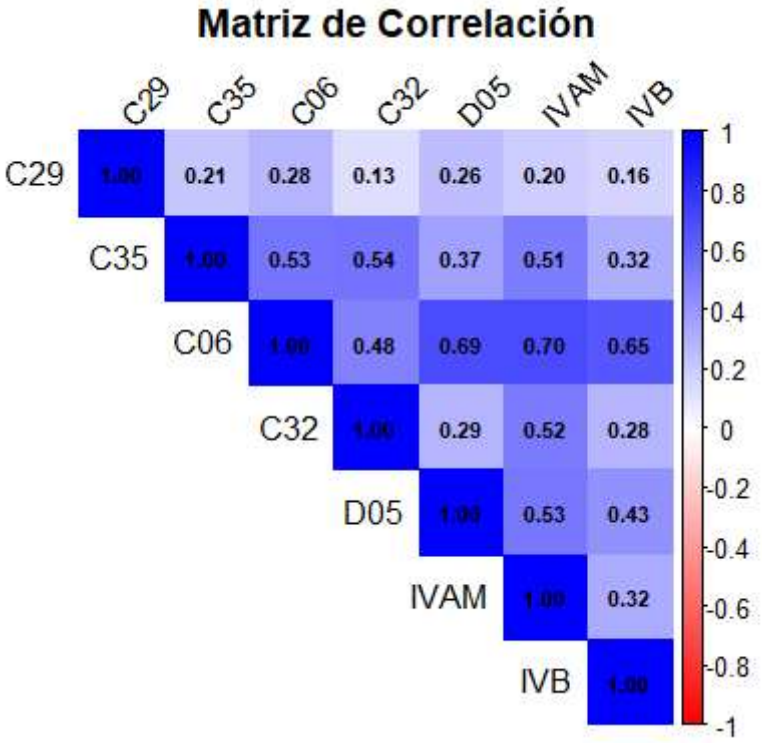

Figura 49. *Correlación de las variables que se eligieron para el modelo.*

6.4 PLANTEAMIENTO DE MODELO ESPACIAL

A pesar de los resultados obtenidos en el Capítulo 6.2, se decidió implementar un modelo espacial con el objetivo de evaluar si la aparente aleatoriedad en la distribución de accidentes era global o si existían patrones locales concentrados en zonas específicas. Sin embargo, los resultados no fueron favorables, como se muestra en la *Figura 50*.

```
> summary(result)
Time used:
    Pre = 4.89, Running = 125, Post = 3.86, Total = 134
Fixed effects:
             mean    sd 0.025quant 0.5quant 0.975quant mode   kld
(Intercept)    0 0.017     -0.008        0      0.008    0 0.746
C29            0 0.008     -0.016        0      0.016    0 0.000
C35            0 0.010     -0.020        0      0.020    0 0.000
C06            0 0.016     -0.032        0      0.032    0 0.000
C32            0 0.010     -0.020        0      0.019    0 0.000
D05            0 0.011     -0.022        0      0.021    0 0.000
IVAM           0 0.012     -0.024        0      0.023    0 0.000
IVB            0 0.011     -0.021        0      0.021    0 0.000

Random effects:
  Name    Model
    spatial SPDE2 model

Model hyperparameters:
                    mean   sd 0.025quant 0.5quant 0.975quant  mode
Theta1 for spatial -2.10 3.16     -8.184    -2.14       4.24 -2.32
Theta2 for spatial  6.62 2.96      0.822     6.61      12.46  6.57

Deviance Information Criterion (DIC) ...............: 32785.01
Deviance Information Criterion (DIC, saturated) ....: 17.01
Effective number of parameters .....................: 8.50

Watanabe-Akaike information criterion (WAIC) ...: 32776.52
Effective number of parameters .................: 0.011

Marginal log-Likelihood:  -16447.55
 is computed
Posterior summaries for the linear predictor and the fitted values are computed
(Posterior marginals needs also 'control.compute=list(return.marginals.predictor=TRUE)')
```

Figura 50. *Resultados del modelo espacial planteado.*

De esta Figura cabe resaltar:

- Los valores extremadamente altos obtenidos para DIC de 32,785.01 y WAIC de 32,776.52, que sugieren un ajuste deficiente de los datos.

- Los valores theta (θ1 = -2.10, θ2 = 6.62) que indican un comportamiento no estacionario del campo aleatorio espacial.

- Todos los coeficientes beta (β) de las variables independientes (C29, C35, etc.) mostraron medias posteriores centradas en 0 con intervalos de credibilidad que incluyen el 0 (-0.016 a 0.016 para C29, -0.020 a 0.020 para C35, etc.), lo que sugiere que ninguna variable tuvo un efecto estadísticamente significativo en el modelo espacial.

6.5 MODELOS NO ESPACIALES

Dado que el modelo espacial no mostró significancia estadística (β ≈ 0, DIC/WAIC altos) y se confirmó la aleatoriedad espacial mediante la función L (*Figura 44*), se implementaron modelos no espaciales basados en densidades kernel. Este enfoque permitió relacionar

accidentes mortales con infracciones mediante rejillas territoriales (con el uso de datos tipo raster) evitando el supuesto de dependencia espacial que no se cumple en los datos.

6.5.1 METODOLOGÍA

Para la elaboración de los distintos modelos, se crearon cinco particiones de los datos. El ochenta por ciento de estos se destinó al entrenamiento, mientras que el veinte por ciento restante se utilizó para el testeo. Cada modelo se desarrolló con su respectivo conjunto de entrenamiento y se sometió a una validación cruzada. La configuración tanto del modelo como de la validación cruzada se detalla en la *Figura 51*. Con estas configuraciones, se procedió a probar los siguientes modelos, empleando variables derivadas de los rasters de densidad:

```
set.seed(123)
particion ← createDataPartition(df$y, p = 0.8, list = FALSE)
train_data ← df[particion, ]
test_data  ← df[-particion, ]
```

```
# ═══ CONTROL DE ENTRENAMIENTO ═══

ctrl ← trainControl(method = "cv", number = 5)
```

Figura 51. *Configuraciones para la validación cruzada.*

Para todos los modelos utilizados en este trabajo (Regresión Lineal, Random Forest, SVM, Redes Neuronales, GBM y XGBoost), se empleó la librería caret de R para gestionar el proceso de entrenamiento y optimización de hiperparámetros. caret permite automatizar la búsqueda de los mejores hiperparámetros mediante el argumento tuneLength, que indica cuántas combinaciones probar; en este caso se definió en 5 para modelos como Random Forest, SVM, NNET, GBM y XGBoost. Internamente, caret construye una rejilla de hiperparámetros (grid search) con valores generados automáticamente en rangos típicos recomendados para cada método (por ejemplo, para Random Forest se exploran valores de mtry; para SVM el costo y el parámetro gamma; para XGBoost, parámetros como max_depth y eta). Cada combinación se evalúa mediante validación cruzada (en este caso ctrl, configurado con method = "cv", number = 5), lo que significa que el conjunto de entrenamiento se divide en 5 particiones, entrenando y validando iterativamente el modelo para estimar su desempeño medio y reducir el riesgo de sobreajuste. Finalmente, se verifica el modelo optimizado aplicándolo a un conjunto de prueba independiente (20% de los datos) para evaluar la capacidad real de generalización. Esta metodología garantiza la selección rigurosa de hiperparámetros, evitando sobreajuste y maximizando la robustez de las predicciones.

a) Modelo Poisson

Aunque el modelo no es muy adecuado para estos ejemplos debido a que no le es posible identificar interacciones complejas entre las variables además de asumir igualdad entre media y varianza (que no es el caso de los accidentes de tránsito). Se implementa porque permite inferencia estadística sobre la importancia de las variables por lo que resulta útil para el estudio del caso [27]. La ecuación de este modelo es la mostrada en la *Figura 52* y su respectiva implementación se observa en la *Figura 53*.

$$log(\lambda_i) = \beta_0 + \beta_1 C29 + \beta_2 C35 + \beta_3 D06 + \beta_4 C32 + \beta_5 D05 + \beta_6 IVAM + \beta_7 IVB$$

Figura 52. *Fórmula del modelo poisson.*

```
modelo_glm ← glm(y ~ C29 + C35 + C06 + C32 + D05 + IVAM + IVB,
                 family = "poisson", data = df)
```

Figura 53. *Implementación del modelo poisson.*

b) Random Forest (RF)

Útil para identificar relaciones complejas en las variables la implementación se muestra en la *Figura 54,* los hiper parámetros seleccionados por la función Caret para este modelo fueron:
mtry = 5 .

```
rf ← train(y ~ ., data = train_data, method = "rf", tuneLength = 5, trControl = ctrl)
```

Figura 54. *Figura de la implementación del modelo random forest.*

c) XGBoost (XGB)

Eficiencia con datos desbalanceados (útil para eventos raros como accidentes mortales) sin embargo se requiere un ajuste muy específico de los hiperparametros.

Hiperparámetros ajustados a través de la función Caret:

nrounds (número de iteraciones) = 250.

max_depth (profundidad máxima de cada árbol) = 5.

eta (tasa de aprendizaje) = 0.3.

gamma (reducción mínima de pérdida para realizar división) = 0.

colsample_bytree (proporción de columnas muestreadas por árbol) = 0.8

min_child_weight (peso mínimo de hijos) = 1.

subsample (proporción de datos muestreados en cada iteración) = 0.75.

El modelo ajustado se muestra en la *Figura 55.*

```
xlb ← train(y ~ ., data = train_data, method = "xgbTree" , tuneLength = 5, trControl = ctrl)
```

Figura 55. *Figura del planteamiento del modelo XGBoost.*

d) Otros modelos evaluados

Se evalúan además otros modelos para ver su comportamiento y poder realizar una comparación de resultados tales como modelos de redes neuronales (NNet), SVM/GBM y modelos lineales.

Teniendo en cuenta que el modelo espacial no mostró significancia estadística (betas ≈ 0, DIC/WAIC altos), lo que confirma la aleatoriedad espacial de los accidentes mortales. Por ello, se optó por modelos no espaciales basados en densidades kernel, destacando Random Forest y XGBoost por su capacidad para capturar patrones complejos en los datos. Los resultados detallados de estos modelos, incluyendo su desempeño y capacidad predictiva, se presentan en el Capítulo 7, los hiper parámetros calculados para cada uno de los modelos restantes se listan a continuación:

SVM radial:
C = 4
Sigma = 39.81223

NEET:
Size = 9
Decay = 0.01

GBM:
n.trees = 250
interaction.depth = 5
shrinkage = 0.1
n.minobsinnode = 10

7. RESULTADOS DEL MODELO

Para realizar cada uno de los modelos mencionados en el Capítulo 6.5 se llevó a cabo una normalización de los datos para que aquellos modelos que se ven altamente afectados por las escalas de las variables obtengan un mejor rendimiento.

7.1 SIGNIFICANCIA ESTADÍSTICA (POISSON)

Para calcular el impacto real de cada variable en el número de infracciones, se optó por implementar el modelo Poisson sin normalizar los datos. Esto garantiza que los coeficientes estimados reflejan directamente el cambio en la variable de respuesta por cada evento de infracción registrado. Los resultados obtenidos, presentados en la *Figura 56*, muestran los efectos brutos de las variables predictoras, lo que permite una interpretación más intuitiva de su influencia en el fenómeno estudiado.

```
> summary(modelo_glm)

Call:
glm(formula = y ~ C29 + C35 + C06 + C32 + D05 + IVAM + IVB, family = "poisson",
    data = df)

Coefficients:
             Estimate Std. Error  z value Pr(>|z|)
(Intercept)  9.457e+00  8.299e-05 113956.4   <2e-16 ***
C29          5.192e-07  1.987e-10   2612.9   <2e-16 ***
C35         -1.530e-09  1.546e-12   -989.9   <2e-16 ***
C06          8.273e-06  5.880e-10  14071.0   <2e-16 ***
C32         -1.375e-08  1.101e-10   -124.9   <2e-16 ***
D05          1.329e-06  6.649e-10   1998.1   <2e-16 ***
IVAM        -1.764e-04  3.967e-08  -4447.1   <2e-16 ***
IVB          3.223e-04  4.644e-08   6939.7   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 1545817392  on 16383  degrees of freedom
Residual deviance:  767917044  on 16376  degrees of freedom
AIC: Inf

Number of Fisher Scoring iterations: 7
```

Figura 56. *Resultados del modelo poisson.*

Un primer análisis de la *Figura 56* revela que todas las variables incluidas en el modelo presentan significancia estadística. Sin embargo, para comprender con mayor profundidad el peso específico de cada predictor, se realizaron dos análisis complementarios:

Importancia relativa de las variables: Calculada mediante la función varImp, que evalúa la contribución de cada factor en el modelo.

Efectos marginales: Obtenidos al modificar incrementalmente los valores de cada variable y registrar los cambios correspondientes en las predicciones del modelo.

El código empleado para estos análisis se detalla en la *Figura 57*, mientras que los resultados gráficos de importancia variable y efectos marginales se presentan en las *Figura 58* y *Figura 59* respectivamente. La síntesis numérica de estos hallazgos puede consultarse en la *Tabla 7*.

```r
efectos_marginales <- sapply(names(df)[names(df) != "y"], function(var) {
  nuevo_dato <- df[1, ]
  nuevo_dato[[var]] <- nuevo_dato[[var]] + 1
  pred_nueva <- predict(modelo_glm, newdata = nuevo_dato, type = "response")
  pred_original <- predict(modelo_glm, newdata = df[1, ], type = "response")
  diferencia <- pred_nueva - pred_original
  return(diferencia)
})

# Convertir a data.frame para graficar
efectos_df <- data.frame(
  Variable = names(efectos_marginales),
  Cambio_en_lambda = as.numeric(efectos_marginales)
)
```

Figura 57. *Código implementado para medir efectos marginales.*

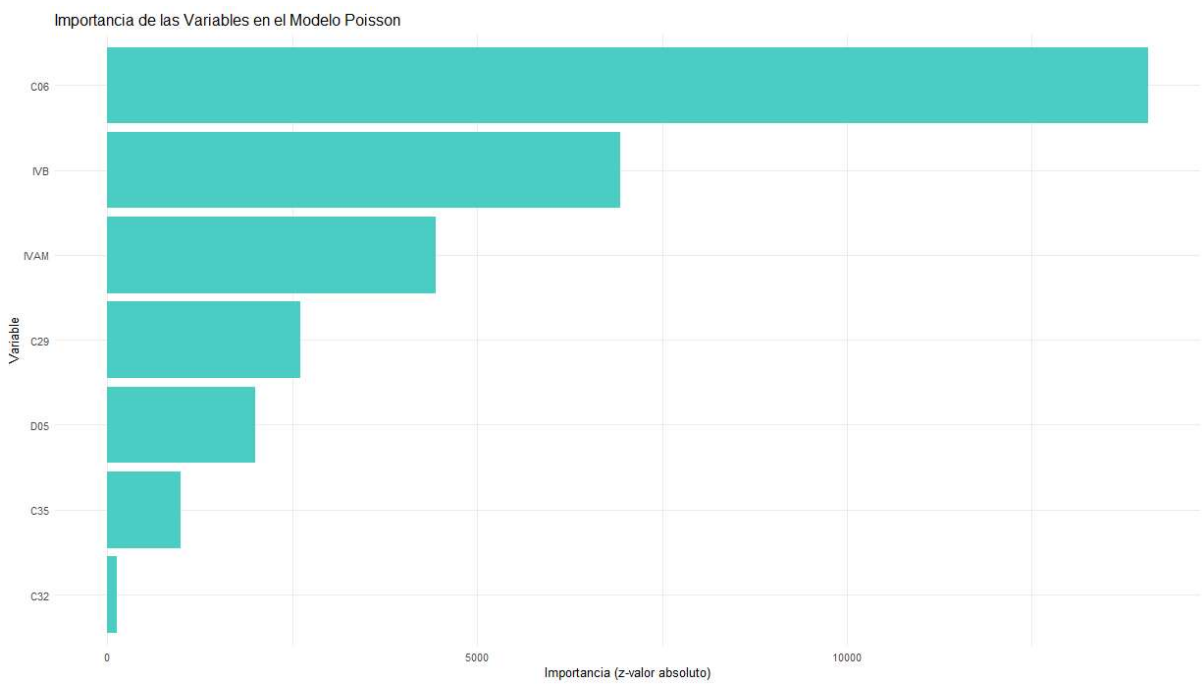

Figura 58. *Gráfica de importancia relativa de las variables.*

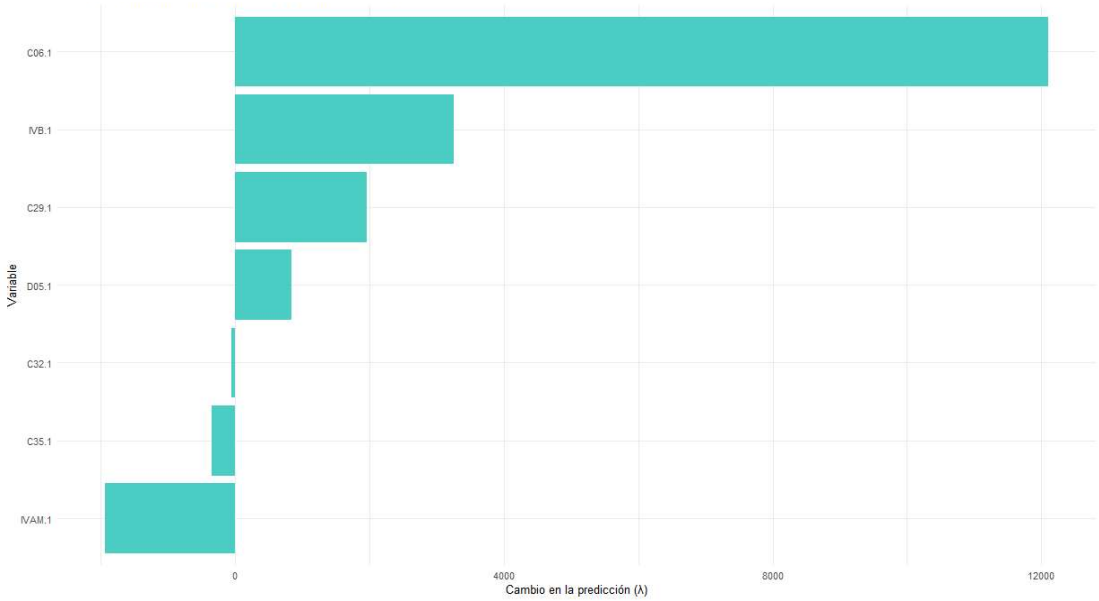

Figura 59. *Figura de los efectos marginales de las variables.*

VAR	β	INTERPRETACIÓN exp(β)	IMPORTANCIA	EFECTO MARGINAL	DESCRIPCIÓN EM(efecto marginal)/SE(significancia estadística)
C29	5.192e-7	1.0000005	2612.91	6.6486e-3	Incrementos en C29 aumentan λ en 0.00005% EM:Bajo SE:Alta
C35	-1.530e-7	0.999999	989.93	-1.9598e-5	Incrementos en C35 disminuyen λ en 0.0001% EM:Muy bajo SE:Alta
C06	8.273e-6	1.0000083	14071.03	1.0595e-1	Incrementos en C06 aumentan λ en 0.00083% EM:Moderado SE:Más Influyente
C32	-1.375e-8	0.999999	124.93	-1.7608e-4	incrementos en C32 disminuye λ en 0.00001% EM:Moderado SE:Baja
D05	1.329e-6	1.0000013	1998.10	1.7015e-2	Incrementos en D05 aumentan λ en 0.00013% EM:Bajo SE:Alta
IVAM	-1.764e-4	0.99982	4447.12	-2.2592	Incrementos en IVAM disminuye λ en 0.018% EM:Bajo SE:Moderada
IVB	3.223e-4	1.00033	6939.73	4.1284	Incrementos en IVB aumentan λ en 0.033% EM:Alto SE:Alta

Tabla 7. *Tabla resumen de interpretación de los coeficientes beta.*

Tras presentar los resultados del modelo de regresión de Poisson, es necesario evaluar uno de los supuestos fundamentales de este tipo de modelos: que la media y la varianza de la variable dependiente sean iguales (equidispersión). Cuando este supuesto no se cumple y la varianza excede la media, se produce lo que se conoce como sobredispersión, lo que puede invalidar las inferencias del modelo de Poisson. Por este motivo, se procedió a calcular el coeficiente de sobredispersión y realizar pruebas formales, con el fin de confirmar este fenómeno y, en caso de encontrarlo, ajustar un modelo alternativo que controle adecuadamente la dispersión de los datos o mostrarnos como la importancia y significado de las variables se mantienen.

Para ello se calcula el coeficiente de sobredispersión como se muestra en la *Figura 60*.

```
> coef_sobredisp ← sum(residuals(modelo_glm, type="pearson")^2) / modelo_glm$df.residual
> print(paste("Coeficiente de sobredispersión:", round(coef_sobredisp, 3)))
[1] "Coeficiente de sobredispersión: 54773.506"
```

Figura 60. *Cálculo de coeficiente de sobre dispersión.*

Dado que el coeficiente de sobredispersión obtenido es mucho mayor que 1, lo que indica presencia de sobredispersión en el modelo de Poisson se procede a realizar una prueba más formal como se muestra en la *Figura 61*.

```
> prueba_sobredisp ← dispersiontest(modelo_glm)
> print(prueba_sobredisp)

        Overdispersion test

data:  modelo_glm
z = 42.985, p-value < 2.2e-16
alternative hypothesis: true dispersion is greater than 1
sample estimates:
dispersion
  54740.34
```

Figura 61. *Prueba de Cameron y Trivedi.*

La prueba de Cameron y Trivedi realizada en la figura z confirma la sobredispersión de forma estadísticamente significativa ($p < 0.05$), justificando el uso de un modelo alternativo.

Por ello se decidió reestimar el modelo mediante regresión binomial negativa. Los resultados muestran que los signos y la significancia estadística de las variables permanecen estables, validando la robustez de las relaciones encontradas esto se observa en la *Figura 62*.

```
> modelo_nb ← glm.nb(y ~ ., data = train_data)
There were 50 or more warnings (use warnings() to see the first 50)
> summary(modelo_nb)

Call:
glm.nb(formula = y ~ ., data = train_data, init.theta = 117787262.8,
    link = log)

Coefficients:
             Estimate Std. Error   z value Pr(>|z|)
(Intercept)  1.002e+01  6.273e-05 159744.53   <2e-16 ***
C29          1.414e-01  6.120e-05   2310.45   <2e-16 ***
C35         -2.958e-02  3.168e-05   -933.89   <2e-16 ***
C06          6.662e-01  5.281e-05  12615.49   <2e-16 ***
C32         -3.198e-03  3.722e-05    -85.92   <2e-16 ***
D05          5.796e-02  3.574e-05   1621.73   <2e-16 ***
IVAM        -1.618e-01  4.084e-05  -3961.16   <2e-16 ***
IVB          2.274e-01  3.640e-05   6248.05   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for Negative Binomial(117787263) family taken to be 1)

    Null deviance: 1235358136  on 13107  degrees of freedom
Residual deviance:  610726750  on 13100  degrees of freedom
AIC: 610804792
```

Figura 62. *importancia y coeficiente de variables seleccionadas.*

7.2 DESEMPEÑO COMPARATIVO (RF VS. XGB VS. OTROS)

Tras completar el entrenamiento de los modelos, se evaluó su desempeño mediante métricas cuantitativas y representaciones gráficas de las predicciones. Los resultados numéricos de estas evaluaciones se presentan en la *Figura 63*.

```
> print(tabla_modelos)
         Modelo      RMSE       MAE        R2
1 Poisson (GLM) 62064.068 33436.453 0.5032291
2            LM 46096.962 25349.251 0.5670000
3           SVM 23240.258  8951.997 0.9140000
4          NNET 44958.903 23780.908 0.5890000
5            RF  3154.879  1320.297 0.9980000
6           GBM 15674.646  8133.078 0.9510000
7       XGBoost  4148.586  2363.253 0.9970000
```

Figura 63. *Métricas de desempeño para los modelos realizados.*

De la *Figura 63* es importante realizar un análisis de las diferentes métricas.

7.2.1 RAÍZ DEL ERROR CUADRÁTICO MEDIO (RMSE)

Los resultados de la Raíz del Error Cuadrático Medio (RMSE) muestran diferencias importantes en la precisión de los modelos, siendo particularmente útil para identificar su sensibilidad a errores grandes:

Poisson (GLM) y Modelo Lineal (LM) presentan los valores más altos de RMSE (62,064 y 46,097 respectivamente), lo que evidencia notables discrepancias entre sus predicciones y los valores reales. Esto refuerza su limitación para modelar relaciones complejas en los datos.

Red Neuronal (NNET) muestra un RMSE de 44,959, inferior al GLM pero aún elevado, lo que sugiere que, a pesar de su capacidad para aprender patrones no lineales, el modelo no logró converger a una solución óptima, posiblemente por falta de datos. Modelos avanzados como SVM, RF, GBM, XGBoost destacan por su bajo RMSE en especial Random Forest (RF): RMSE = 3,154.9, XGBoost: RMSE = 4,148.6

7.2.2 ERROR ABSOLUTO MEDIO (MAE)

Los valores del Error Absoluto Medio (MAE) revelan diferencias notables en la precisión predictiva de los modelos:

Poisson (GLM), Modelo Lineal (LM) y Red Neuronal (NNET) presentan errores elevados (MAE > 23,700), lo que indica una mayor desviación promedio entre las predicciones y los valores reales. Este comportamiento coincide con sus limitaciones para modelar relaciones complejas (Poisson y LM) o con problemas de ajuste en el caso de la red neuronal (NNET).

Los modelos restantes (SVM, RF, GBM, XGBoost) muestran un MAE significativamente menor (< 9,000), destacando: Random Forest (RF): MAE = 1,320.3 modelo con la mejor precisión absoluta, XGBoost: MAE = 2,363.3.

7.2.3 COEFICIENTE DE DETERMINACIÓN (R2)

Los valores de R^2 muestran un desempeño diferenciado entre los modelos evaluados:

Poisson (GLM), Modelo Lineal (LM) y Red Neuronal (NNET) presentan coeficientes de determinación bajos ($R^2 \leq 0.59$), lo que indica una capacidad limitada para explicar la varianza de los datos. Esto se debe a que los modelos Poisson y lineal tienen dificultades para captar relaciones complejas/no lineales entre las variables y el modelo de red neuronal (NNET) no contó con suficientes datos o ajustes óptimos durante el entrenamiento.

El resto de los modelos (SVM, RF, GBM, XGBoost) muestran un R^2 excelente (> 0.91), destacando los modelos mencionados a continuación: Random Forest (RF): R^2 = 99.8% (mejor desempeño global), XGBoost: R^2 = 99.7% (equivalente a RF en precisión).

7.2.4 GRÁFICOS DE DENSIDAD PREDICHOS

Con base en los resultados obtenidos, se generaron las gráficas de predicción para cada modelo. Al analizarlas, se observa que los modelos Poisson (GLM), Lineal (LM) y la Red Neuronal (NNET) presentan dificultades para ajustarse adecuadamente a los valores reales, mostrando discrepancias evidentes en sus predicciones. Por otro lado, los mapas predictivos de los modelos Random Forest (RF), XGBoost y GBM muestran una mayor similitud con los datos originales. A simple vista, estos modelos capturan mejor los patrones y relaciones presentes en los datos, destacando especialmente RF y XGBoost por su precisión las gráficas se muestran en la *Figura 64*.

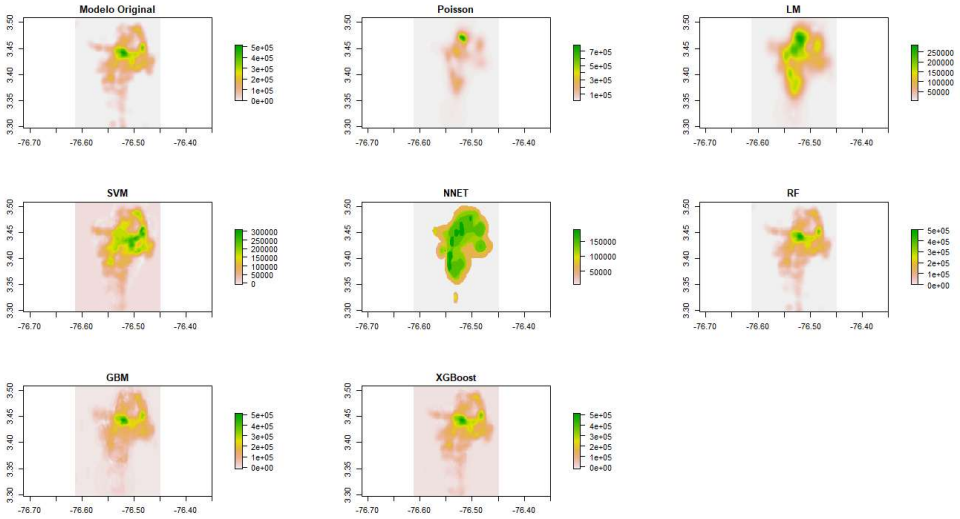

Figura 64. *Mapas generados con la predicción de los modelos con variables normalizadas.*

7.3 VALIDACIÓN ESPACIAL (MAPAS PREDICTIVOS VS. REALES)

Dado que a la simple observacion de los graficos de densidad no es posible discernir el nivel de ajuste en algunos casos se procede a realizar graficos de densidad restando la calculada mediante los modelos y los datos de accidentes mortales, los resultados obtenidos para este ejercicio se presentan en la *Figura 65.*

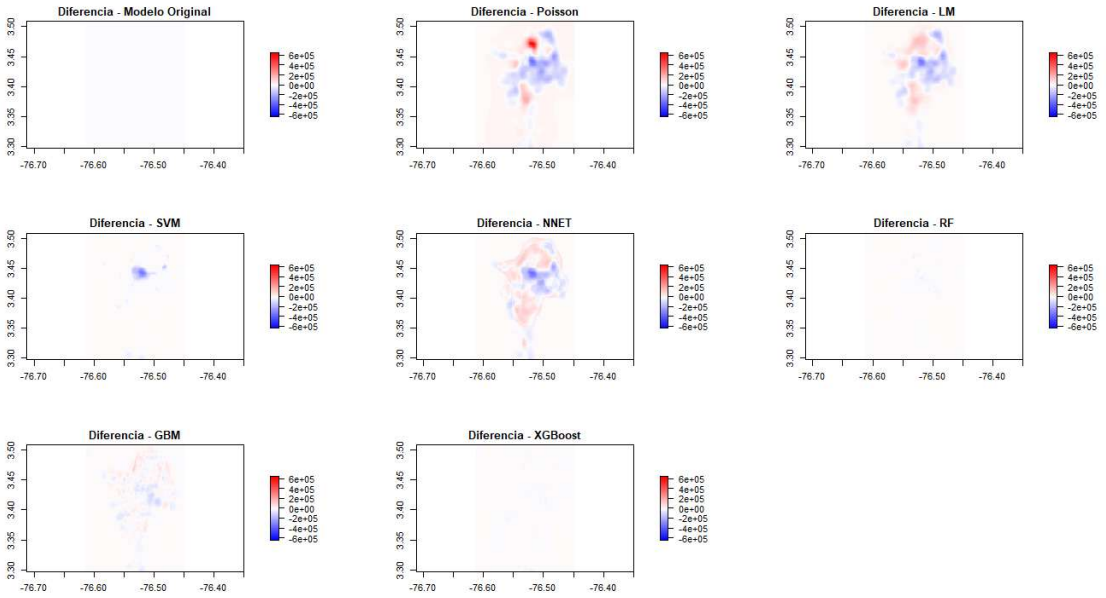

Figura 65. *diferencia entre la predicción del modelo y los accidentes mortales con variables normalizadas.*

De la Figura a se puede ver que los dos modelos que ajustan mejor son el Random forest y el XGboost y se presentan mayores diferencias en el gbm.

7.4 IMPACTO DE LA NORMALIZACIÓN

Una vez validado el desempeño de cada uno de los modelos se procede a realizar el entrenamiento esta vez sin las variables normalizadas esto debido a que presenta una mayor facilidad para implementar en un sistema de alertas tempranas.

Los resultados relacionados a las variables de desempeño se pueden observar en la *Figura 66*, donde se observa que el modelo SVM sufren un desajuste notorio debido a que es sensibles a las escalas de las variables mientras que otros modelos no parecen verse mayormente afectados.

	Modelo	RMSE	MAE	R2
1	Poisson (GLM)	62064.068	33436.453	0.5032291
2	LM	46096.962	25349.251	0.5670000
3	SVM	23240.258	8951.997	0.9140000
4	NNET	43759.026	21943.809	0.6100000
5	RF	3155.730	1326.992	0.9980000
6	GBM	15628.402	8152.093	0.9520000
7	XGBoost	3644.849	2048.139	0.9970000

Figura 66. *Resultados de los diferentes modelos probados.*

De igual manera para estos modelos con las variables sin normalizar se presentan los

gráficos de densidad predichos en la *Figura 67* y las diferencias con la densidad de accidentes mortales en la *Figura 68*.

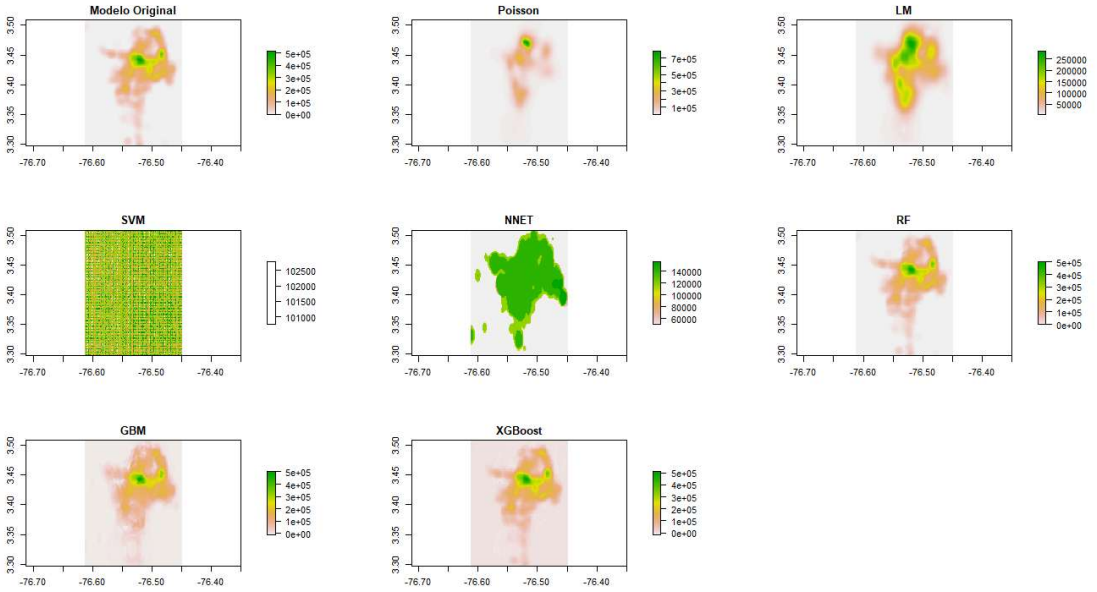

Figura 67. *Mapas de predicción de los modelos con variables sin normalizar*

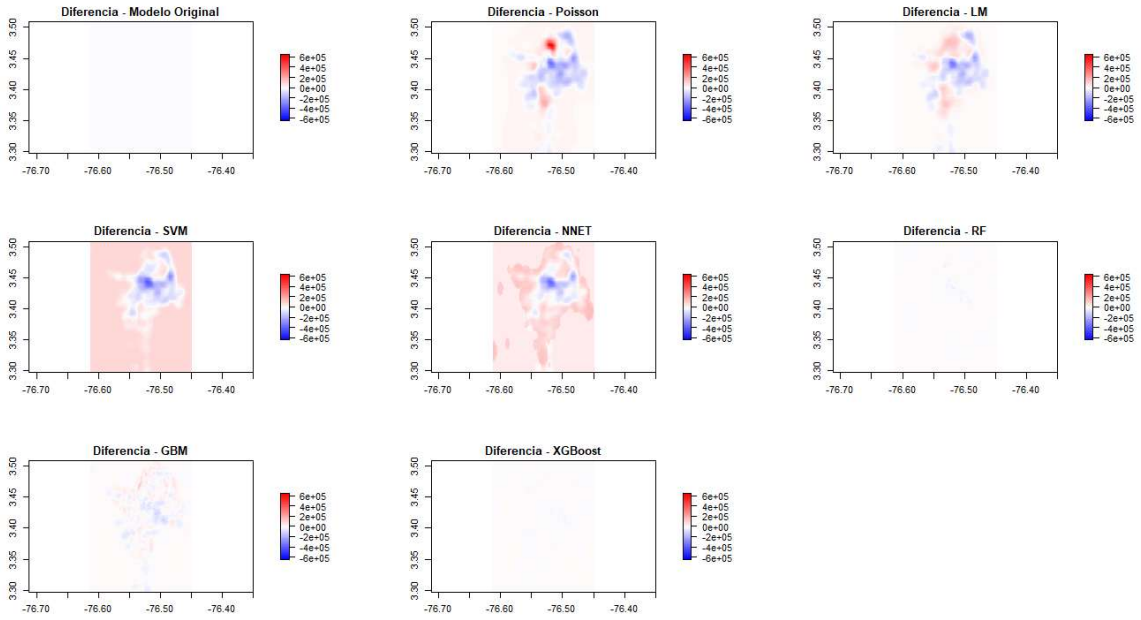

Figura 68. *Diferencia entre la predicción del modelo y los accidentes mortales con variables sin normalizar.*

Como se puede observar los modelos con los mejores resultados en las variables normalizadas tambien lo son con las variables normalizadas lo que nos permite usarlos sin mayor inconveniente para predecir los accidente.

7.5 CONCLUSIONES Y MODELO SELECCIONADO

Para la implementación final se optó por utilizar el modelo con las variables en su escala original, sin normalización. Esta decisión se tomó considerando dos factores clave: el buen desempeño que mostró el modelo en esta configuración y la mayor practicidad para su implementación en análisis de tiempo real. Como estrategia de modelado, se empleó un promedio de las predicciones generadas por los modelos Random Forest (RF) y XGBoost, cuya configuración específica se detalla en la *Figura 69*.

```
pred_promedio_simple ← (xgb_model + rf_model) / 2
plot(pred_promedio_simple, main = "Promedio Simple de Predicciones")
```

Figura 69. *Código para promediar modelos.*

Los resultados de este enfoque combinado se presentan en la *Figura 70*, donde puede observarse el comportamiento predictivo del modelo integrado. Una vez definida esta arquitectura, tanto los modelos individuales como el modelo promedio fueron guardados en archivos con formato .rds. Este formato permite su posterior carga y utilización en la herramienta desarrollada específicamente para el análisis en tiempo real de los datos.

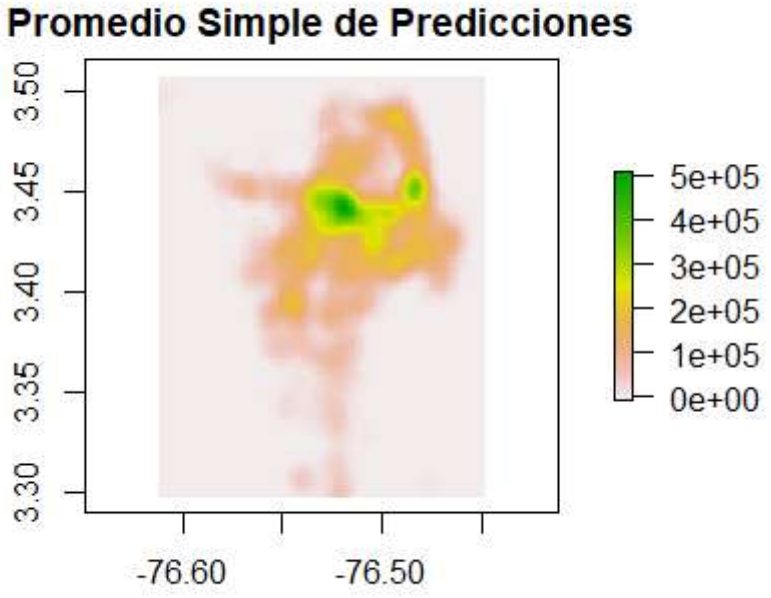

Figura 70. a) *densidad del modelo realizado con el promedio.*

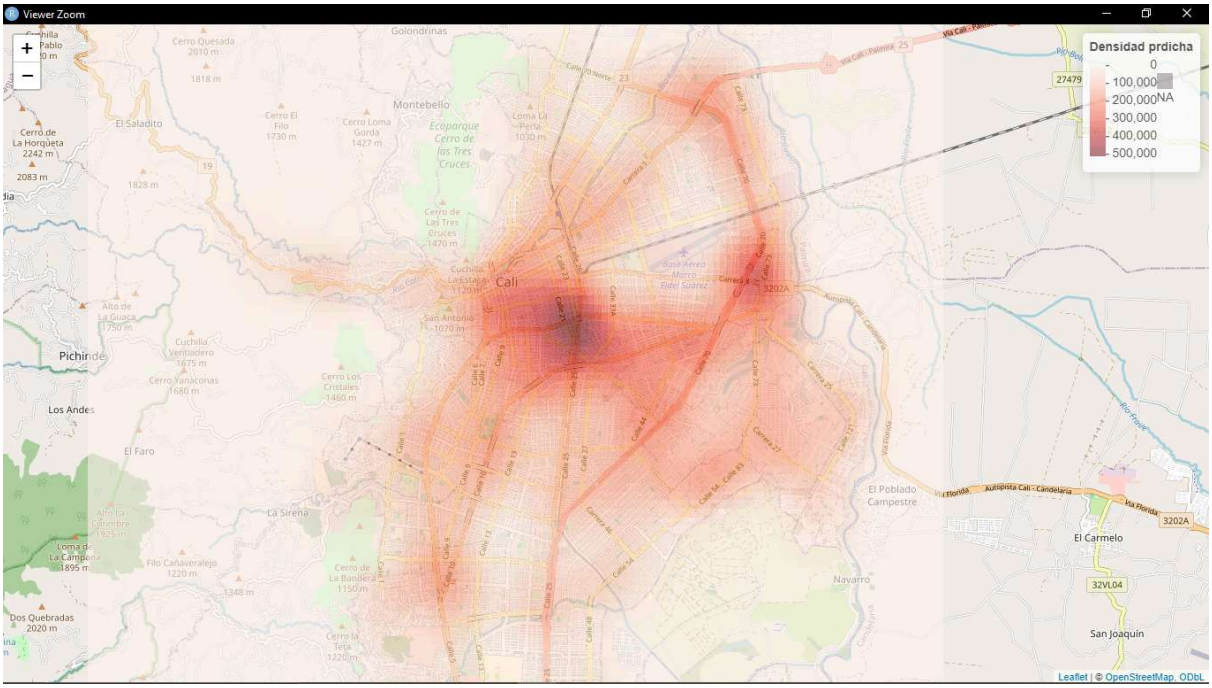

Figura 70. b) *Densidad superpuesta en el mapa de Cali del modelo realizado con el promedio.*

8. HERRAMIENTA DE ALERTAS TEMPRANAS

Para el desarrollo de la herramienta de alertas tempranas, se establece una conexión directa con la base de datos de infracciones en tiempo real. De esta base se extraen tres campos clave: las coordenadas geográficas (representadas en las columnas CORD_X y CORD_Y) y el tipo de evento registrado (capturado en la columna TIPO, que incluye categorías como infracciones, bicicletas o ambulancias). Como ejemplo ilustrativo, la *Figura 71* muestra un archivo Excel estructurado con estas tres columnas, donde puede observarse el formato y tipo de datos manejados.

Tipo	CordX	CordY
D05	-76,489677	3,47976984
C29	-76,511123	3.445.100
C35	-76,511123	3.445.100
C06	-76,511123	3.445.100
C32	-76,511123	3.445.100
C05	-76,511123	3.445.100
IVAM	-76,511123	3.445.100
IVB	-76,511123	3.445.100
IVAM	-76,511123	3.445.100

Figura 71. *Formato de la data de entrada.*

Una vez obtenidos los datos, se procede a su procesamiento para integrarlos al modelo predictivo. El primer paso consiste en transformar esta información en archivos raster, los cuales servirán como variables de entrada para el sistema. Es importante destacar que, previo a su incorporación al modelo, estos datos raster NO pasan por un proceso de normalización para garantizar los requerimientos técnicos del algoritmo. El código empleado para llevar a cabo esta normalización se presenta detalladamente en la *Figura 72*.

```
rut = "C:/Users/vicod/Documents/Maestria/tesis/carp_tif/ALERTA_TEMPRANA/" # poner ruta a carpeta
raster_base ← raster(paste(rut,"densidad_accidentes_mortales.tif",sep=""))
rf ← readRDS(paste(rut,"modelo_rf.rds",sep=""))
xlb ← readRDS(paste(rut,"modelo_xgb.rds",sep=""))
gb ← readRDS(paste(rut,"modelo_gb.rds",sep=""))

infracciones_nuevas ← read_excel(paste(rut,"INFRACCIONES.xlsx",sep=""))

tipos ← unique(infracciones_nuevas$Tipo)

# Crear stack vacío
stack_nuevo ← stack()

for (tipo in tipos) {
  r ← raster(raster_base)  # Mismo tamaño, resolución y proyección
  values(r) ← 0            # Inicializa todo en 0
  names(r) ← tipo
  stack_nuevo ← addLayer(stack_nuevo, r)
}


for (i in 1:nrow(infracciones_nuevas)) {
  tipo ← infracciones_nuevas$Tipo[i]
  x ← infracciones_nuevas$CordX[i]
  y ← infracciones_nuevas$CordY[i]

  celda ← cellFromXY(stack_nuevo[[tipo]], cbind(x, y))
  stack_nuevo[[tipo]][celda] ← stack_nuevo[[tipo]][celda] + 1
}
```

Figura 72. a) *Primera parte del código del sistema de alertas tempranas.*

```
r_pred_1 ← raster::predict(stack_nuevo, rf, type = "raw")
r_pred_2 ← raster::predict(stack_nuevo, xlb, type = "raw")
#plot(r_pred_2)

pred_promedio_simple ← (r_pred_1 + r_pred_2) / 2
plot(pred_promedio_simple, main = "Promedio Simple de Predicciones")

# Crear un mapa base de la ciudad de Cali
mapa_base ← leaflet() %>%
  addTiles() %>%  # Añadir capa de OpenStreetMap
  setView(lng = -76.5225, lat = 3.4372, zoom = 12)  # Centrar en Cali

mapa_densidad ← mapa_base %>%
  addRasterImage(pred_promedio_simple, colors = "Reds", opacity = 0.6) %>%
  addLegend(pal = colorNumeric("Reds", values(pred_promedio_simple)),
            values = values(pred_promedio_simple),
            title = "Densidad prdicha")
mapa_densidad
```

Figura 72. b) *Segunda parte del codigo del sistema de alertas tempranas.*

Como complemento fundamental del sistema, se desarrolló un dashboard interactivo que permite visualizar en tiempo real tanto las infracciones registradas como los accidentes mortales reportados en la base de datos. Esta plataforma integra múltiples capas de información estratégica las capas se listan a continuación:

Mapa de infracciones de tránsito ***Figura 73***

Mapa de accidentes mortales ***Figura 74***
Mapa de densidad de accidentes mortales ***Figura 75***
Mapa de densidad predicha por los modeles ***Figura 76***

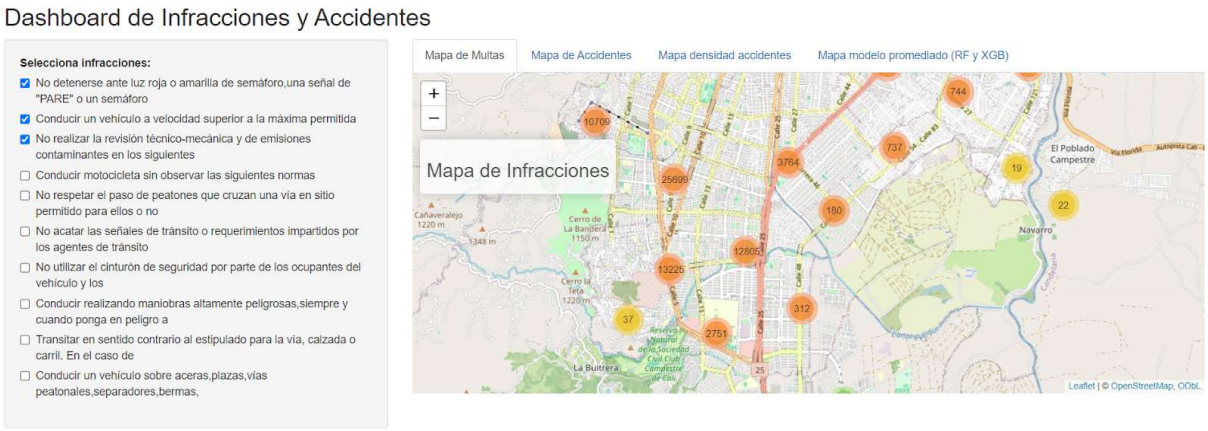

Figura 73. *Mapa de infracciones sobre dashboard realizado.*

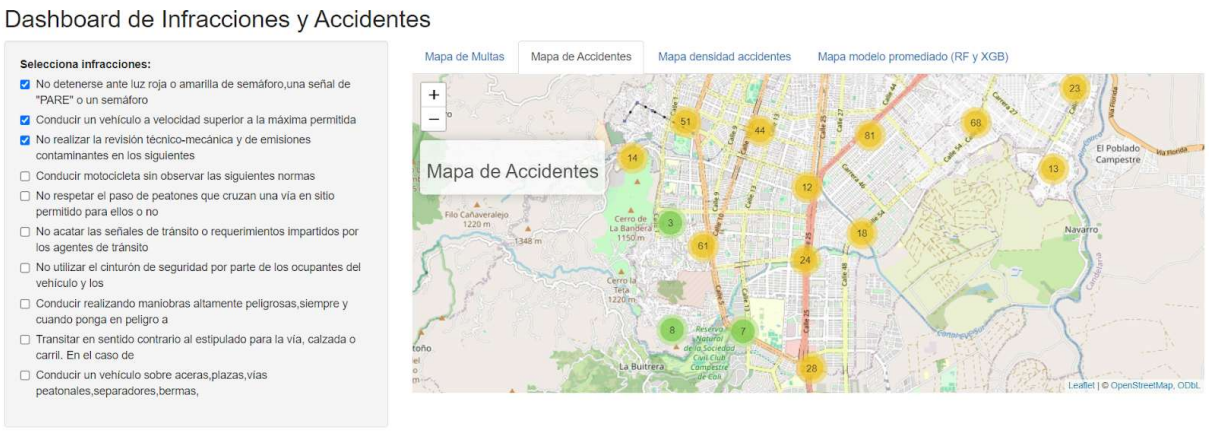

Figura 74. *Mapa de accidentes mortales sobre dashboard realizado.*

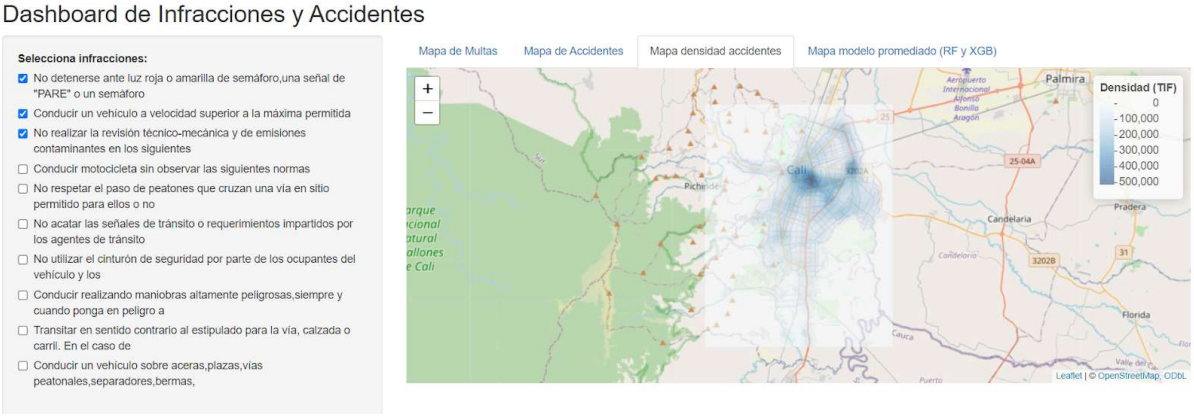

Figura 75. *Mapa de densidad de accidentes mortales sobre dashboard realizado.*

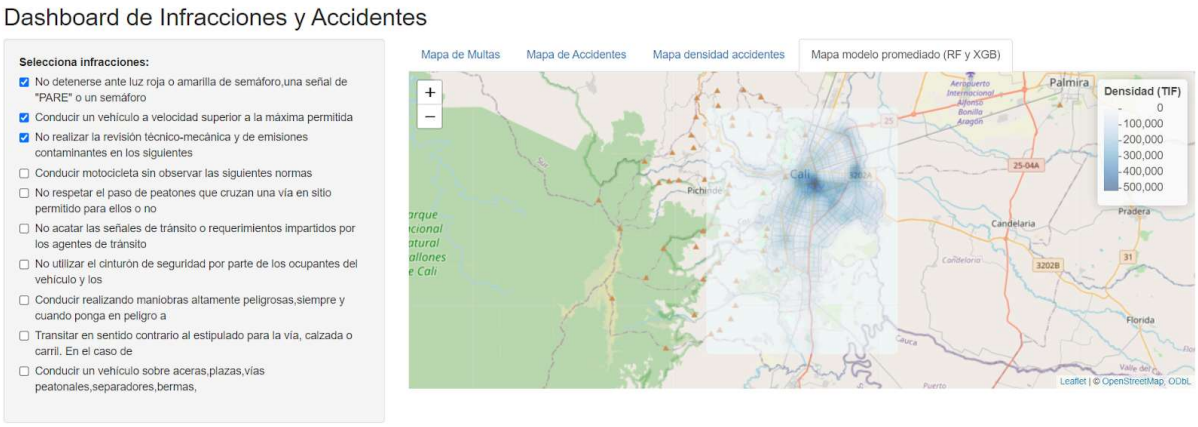

Figura 76. *Mapa de densidad de accidentes sobre dashboard realizado.*

Para implementar el sistema de alertas tempranas mencionado previamente, se analizan las infracciones registradas en un día, identificando pequeñas regiones de densidad que indican una mayor probabilidad de accidentes mortales en puntos específicos del mapa donde ocurrieron dichas infracciones.

En la Figura 77 se observa la distribución de las infracciones analizadas, mientras que en la Figura 78 se visualiza la densidad de posibles accidentes mortales, calculada a partir de los datos de entrada.

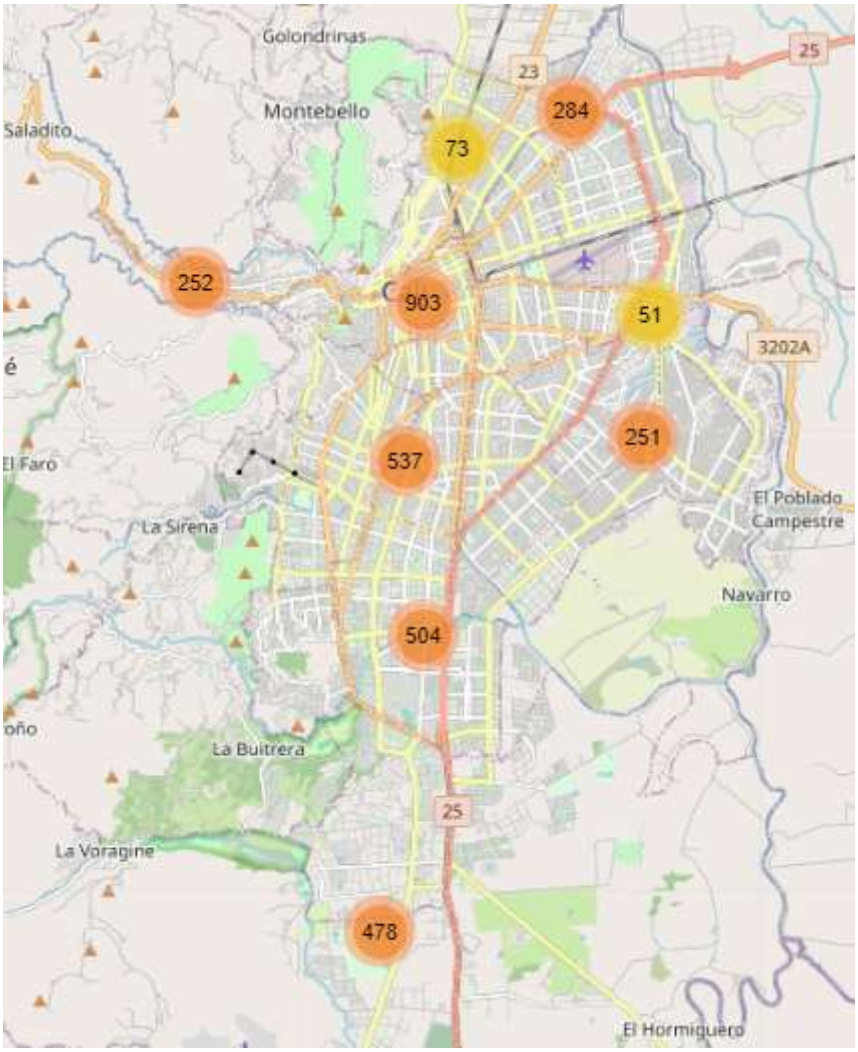

Figura 77. *Mapa de infracciones utilizadas para verificar el modelo.*

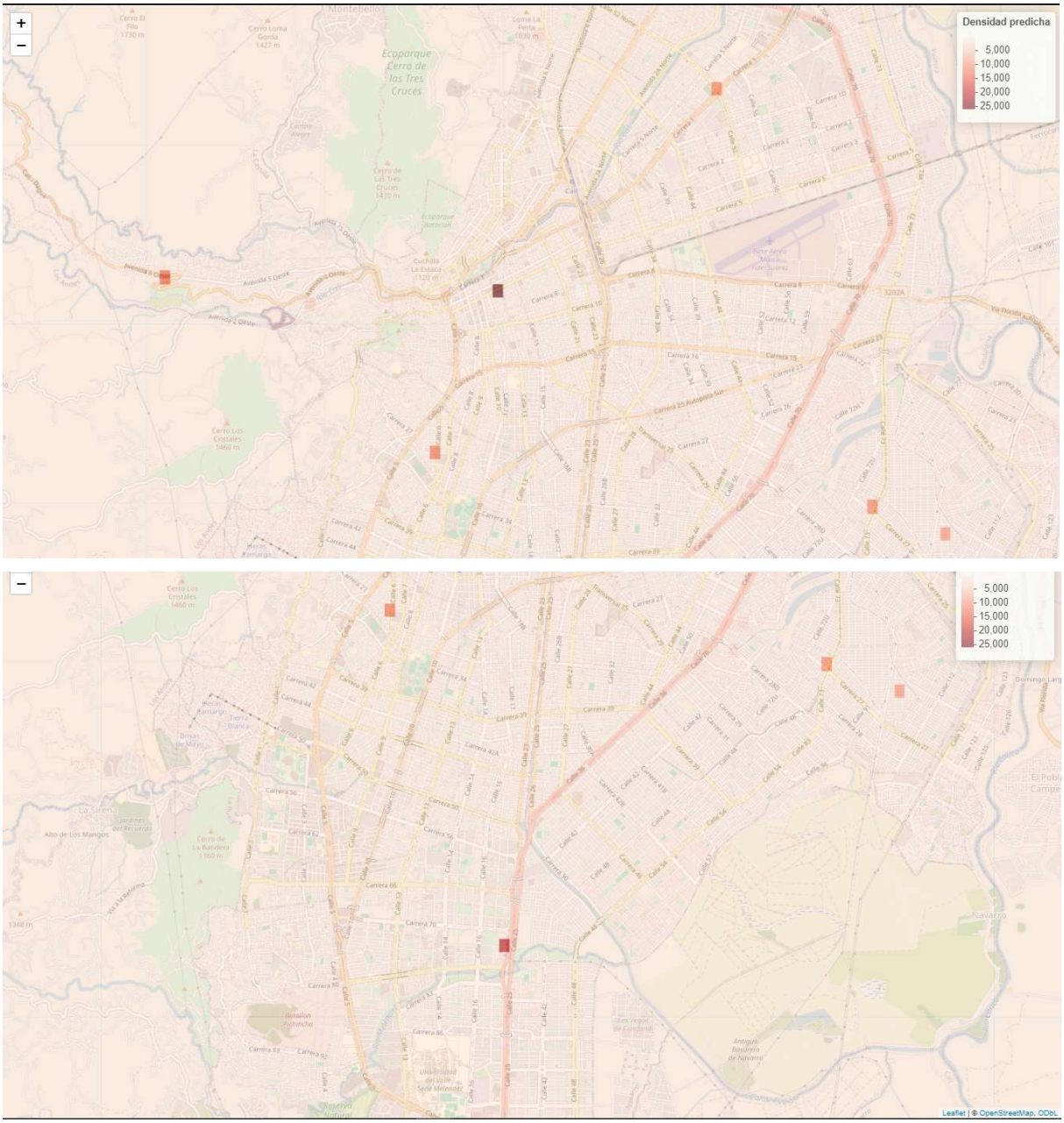

Figura 78. *Mapa de densidad de accidentes predichos con las infracciones de entrada.*

Finalmente, se integraron las infracciones de entrada y la densidad de salida en el dashboard, permitiendo que los stakeholders visualicen las zonas de mayor riesgo y tomen acciones basadas en los resultados.

En la Figura 79 se muestran las infracciones ingresadas, mientras que en la Figura 80 se presenta la densidad de accidentes mortales predichos.

Figura 79. *Mapa de densidad de accidentes sobre dashboard realizado.*

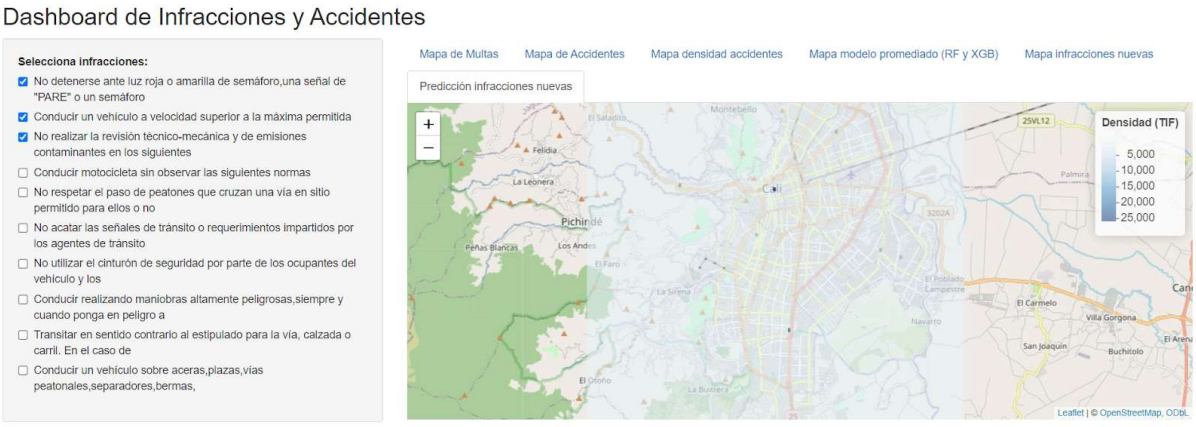

Figura 80. *Mapa de densidad de accidentes sobre dashboard realizado.*

La herramienta de alertas tempranas desarrollada representa un avance significativo en la gestión proactiva de la seguridad vial en Cali. Al integrar modelos predictivos de alta precisión (RF + XGBoost) con un dashboard interactivo, se logra no solo identificar zonas de riesgo en tiempo real, sino también priorizar intervenciones basadas en evidencia.

9. CONCLUSIONES Y RECOMENDACIONES

Mediante el estudio se pudo evidenciar que los actores viales más vulnerables son los peatones, motociclistas y ciclistas con porcentajes del 34.05%, 46.00% y 9.80%, respectivamente.

Al analizar la frecuencia de eventos mortales por edades y género, se observa que la mayor incidencia ocurre en los grupos etarios de 18 a 25 años y de 46 a 60 años. Además, la mayoría de las muertes corresponden al género masculino.

Al comparar los puntos de los accidentes con cada uno de los puntos de las infracciones se observa que la mayoría de estas coinciden geográficamente en ciertos corredores críticos: la Calle 48 (Valle del Lili), la Calle 25 (Avenida Simón Bolívar), la Carrera 100, la Calle 70, la Carrera 1, la Carrera 23 (Autopista Sur) y la Carrera 15, lo que indicaría que podría presentarse relación entres los accidentes y las infracciones en estos puntos o zonas de la ciudad de Cali.

Se realizó un cálculo de correlación espacial para cada tipo de infracción en relación con los accidentes mortales. Este proceso permitió validar estadísticamente lo observado en los mapas y establecer con mayor precisión el grado de asociación entre las variables estudiadas.

Las Infracciones:

- **C29** (exceso de velocidad),
- **C35** (falla técnico-mecánica),
- **C06** (no uso de cinturón),
- **C32** (no respeto al paso peatonal),
- **D05** (uso indebido de zonas peatonales).

Estas infracciones representan **comportamientos de alto riesgo** directamente vinculados a la ocurrencia o gravedad de accidentes.

El modelo espacial fue descartado por falta de significancia estadística ($\beta \approx 0$, DIC/WAIC elevados) y por evidencia de aleatoriedad espacial en los datos de accidentes mortales (confirmada con la función L). Esto justificó el uso de modelos no espaciales basados en densidades de kernel.

Aunque útil para análisis de significancia estadística, el modelo Poisson presentó limitaciones importantes para predecir eventos complejos como accidentes de tránsito (RMSE y MAE altos, baja capacidad explicativa – $R^2 < 0.6$). Se utilizó principalmente para interpretar el impacto de variables específicas.

Los modelos Random Forest (RF) y **XGBoost (XGB)** demostraron ser los modelos más robustos y precisos:

- **RF**: mejor desempeño general con R^2 de 99.8%, RMSE de 3,154.9 y MAE de 1,320.3.
- **XGB**: desempeño muy similar (R^2 de 99.7%), aunque más exigente en ajuste de hiperparámetros.

Ambos modelos capturaron adecuadamente patrones complejos y se adaptan bien tanto con variables normalizadas como sin normalizar.

Se seleccionó el **promedio entre RF y XGBoost con variables no normalizadas** como modelo final. Esta decisión se basó en:

- Su excelente desempeño.
- Facilidad de integración con sistemas en tiempo real.
- Mayor eficiencia operativa.

El modelo final se integró a una **herramienta de alertas tempranas** conectada a una base de datos en tiempo real. Se diseñó un **dashboard interactivo** que incluye:

- Mapas de infracciones, accidentes y densidades.
- Predicciones del modelo.

Esta herramienta es clave para la gestión proactiva de riesgos viales, con el fin de que se pueda mejorar la movilidad de la ciudad de Cali, observando las zonas con mayor presencia de accidentes e infracciones y así poder tomar decisiones para prevenir accidentes fatales.

Recomendaciones:

Aunque variables como **género y edad** son conocidas por influir en la ocurrencia y severidad de accidentes, no pudieron ser incluidas debido a la ausencia de estos datos en la base disponible. Esta limitación representa una **restricción importante** para el análisis y la capacidad explicativa del modelo, por lo cual, esta es una recomendación vital para un próximo análisis.

Otra recomendación es la **Integración de datos externos**: cruzar la base actual con fuentes como Waze, Google Maps (tráfico), meteorología o cámaras urbanas puede enriquecer futuros análisis.

10. REFERENCIAS BIBLIOGRÁFICAS

[1] Departamento Administrativo Nacional de Estadística (DANE), "Estadísticas Vitales - Defunciones: III Trimestre 2023," [Online]. Available: https://www.dane.gov.co/files/operaciones/EEVV/pres-EEVV-Defunciones-IIItrim2023.pdf DIC 2023 [Accessed: 2024].

[2] ANSV, "Estadísticas históricas de víctimas," [Online]. Available: https://www.ansv.gov.co/es/observatorio/estad%C3%ADsticas/historico-victimas. ENE 2024 [Accessed: 2024].

[3] Secretaría de Movilidad de Cali, "Datos abiertos de movilidad," [Online]. Available: https://datos.cali.gov.co/organization/secretaria-de-movilidad . [Accessed: 2024].

[4] Organización Mundial de la Salud, "Road traffic injuries," [Online]. Available: https://www.who.int/es/news-room/fact-sheets/detail/road-traffic-injuries. DIC 2023 [Accessed: 2024].

[5] Organización Mundial de la Salud, "Despite notable progress, road safety remains an urgent global issue," [Online]. Available: https://www.who.int/es/news/item/13-12-2023-despite-notable-progress-road-safety-remains-urgent-global-issue. DIC 2023 [Accessed: 2024].

[6] European Commission, "Road traffic safety: EU continues to improve but more efforts needed," [Online]. Available: https://ec.europa.eu/commission/presscorner/detail/es/ip_23_953. FEB 2023 [Accessed: 2024].

[7] ONU News, "UN road safety progress report: Urgent action needed to meet targets," [Online]. Available: https://news.un.org/es/story/2023/05/1521097. MAY 2023 [Accessed: 2024].

[8] Deutsche Welle, "El flagelo de los accidentes de tránsito en América Latina," [Online]. Available: https://www.dw.com/es/el-flagelo-de-los-accidentes-de-tr%C3%A1nsito-en-am%C3%A9rica-latina/a-68693381. MAR 2024 [Accessed: 2024].

[9] Secretaría de Educación de Bogotá, "Protocolo de atención de siniestros viales en establecimientos educativos," [Online]. Available: https://www.educacionbogota.edu.co/portal_institucional/sites/default/files/inline-files/Anexo%2011%20protocolo_atencion_siniestros_viales_Establecimientos_edu.pdf. [Accessed: 2024].

[10] "Applied Mechanics, Behavior of Materials, and Engineering systems," Springer, 2017. doi: 10.1007/978-3-319-41468-3.

[11] S. Cociu, O. Ioncu, C. Cazacu-Stratu, S. Cebanu, and C. Hamann, "Major behavioral risk factors for road traffic injuries," One Health & Risk Management, vol. 2, no. 4, pp. 28-34, 2021. doi: 10.38045/ohrm.2021.4.02.

[12] D. Brčić, K. Tepeš, and D. Šojat, "Accident Prediction Models in the Scope or Road Safety Improvements," presented at International Conference "Road safety strategic management," Berane, 2014, pp. 41-53.

[13] Zhongping Li, Lirong Cui, Jianhui Chen, Traffic accident modelling via self-exciting point processes, Reliability Engineering & System Safety, Volume 180, 2018, Pages 312-320, ISSN 0951-8320, [Online]. Available: https://doi.org/10.1016/j.ress.2018.07.035. (https://www.sciencedirect.com/science/article/pii/S0951832018301078)

[14] P. J. Diggle, "Statistical analysis of spatial and spatio-temporal point patterns," CRC Press, 2015.

[15] Victoria, I. C., & Galvis, O. (2014). Road Safety Conditions and Estimated Economic Cost of Traffic Fatalities in Medium-Size Colombian Cities. Transportation Research Record, 2465(1), 40-47. [Online]. Available: https://doi.org/10.3141/2465-06

[16] Suárez Parra, Á. (2021). Herramienta de visualización e identificación de siniestros viales sucedidos a una distancia específica de un punto de interés en Bogotá para el proyecto VISIR. Universidad de los Andes. [Online]. Available: http://hdl.handle.net/1992/53430

[17] Bonilla Verdugo, D. (2020). Arquitectura de infraestructura y modelo de optimización para la reducción de la probabilidad de accidentes de usuarios vulnerables en la vía pública. Universidad de los Andes. [Online]. Available: http://hdl.handle.net/1992/43900

[18] Khaula Alkaabi, Identification of hotspot areas for traffic accidents and analyzing drivers' behaviors and road accidents, Transportation Research Interdisciplinary Perspectives, Volume 22, 2023, 100929, ISSN 2590-1982,B [Online]. Available: https://doi.org/10.1016/j.trip.2023.100929 (https://www.sciencedirect.com/science/article/pii/S2590198223001768)

[19] J.P.S. Shashiprabha Madushani, R.M. Kelum Sandamal, D.P.P. Meddage, H.R. Pasindu, P.I. Ayantha Gomes, Evaluating expressway traffic crash severity by using logistic regression and explainable & supervised machine learning classifiers, Transportation Engineering, Volume 13, 2023, 100190, ISSN 2666-691X, [Online]. Available: https://doi.org/10.1016/j.treng.2023.100190 . (https://www.sciencedirect.com/science/article/pii/S2666691X23000301)

[20] International Transport Forum (ITF), "Speed and Crash Risk," OECD, [Online]. Available: https://www.itf-oecd.org/speed-crash-risk . 2018 [Accessed: 2024].

[21] P. Moraga, "Spatial Data Science with applications in R," [Online]. Available: https://www.paulamoraga.com/book-spatial/intensity-estimation.html . 2023 [Accessed: 2024].

[22] D. Arango-Londoño, "Tarea 2 - Estadística Espacial Avanzada: Datos de Sismos para Colombia," RPubs, [Online]. Available: https://rpubs.com/darango/833142 . AGO 2021 [Accessed: 2024].

[23] S. Schepers et al., "Factors influencing the injury severity of single-bicycle crashes," Accident Analysis & Prevention, vol. 149, Nov. 2020. [Online]. Available: https://www.researchgate.net/publication/346583986_Factors_influencing_the_injury_severity_of_single-bicycle_crashes [Accessed: 2024].

[24] M. A. Tavakol et al., "Identifying Key Factors in Traffic Accident Severity: A SHAP and Machine Learning Approach in North Carolina," in 10th International Conference on Industrial Engineering and Systems, Ferdowsi University of Mashhad, Sep. 2024. [Online]. Available: https://www.researchgate.net/publication/384661508_Identifying_Key_Factors_in_Traffic_Accident_Severity_A_SHAP_and_Machine_Learning_Approach_in_North_Carolina [Accessed: 2024].

[25] A. Baddeley et al., Spatial Point Patterns: Methodology and Applications with R. CRC Press, 2015.

[26] Organización Mundial de la Salud, "Global status report on road safety 2023," [Online]. Available: https://www.who.int/teams/social-determinants-of-health/safety-and-mobility/global-status-report-on-road-safety-2023. DIC 2023 [Accessed: 2024].

[27] "Universidad Estatal de Milagro," [Online]. Available: https://dialnet.unirioja.es/servlet/articulo?codigo=7878153. MAR 2021 [Accessed: 2024].

[28] K. Chaudhry and M. Iqbal, "Economic Cost of Road Traffic Accidents in Twin Cities, Pakistan," European Scientific Journal, vol. 14, 2018. doi: 10.19044/esj.2018.v14n25p142.

[29] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785-794.

[30] L. Anselin, Spatial Econometrics: Methods and Models. Springer, 1988.

[31] N. Jean et al., "Combining Satellite Imagery and Deep Learning," Proceedings of the National Academy of Sciences (PNAS), vol. 116, no. 14, pp. 7921-7926, 2019.

[32] H. Meyer et al., "Comparing Machine Learning Models for Spatial Data," Environmental Modelling & Software, vol. 118, pp. 172-181, 2019.