

**Análisis de modelos machine learning en las posibles causas que originan la siniestralidad  
vial en Bogotá**

William Hernando Martínez Rodríguez

Director

Edith Johana Morales Liberato

Universidad Nacional Abierta y a Distancia UNAD  
Escuela de Ciencias Básicas, Tecnología e Ingeniería ECBTI  
Especialización en Ciencia de Datos y Analítica

2024

### **Dedicatoria**

A Dios creador de la vida, quien me brinda todo lo que necesito para lograr lo que me propongo y permitirme cumplir mis metas.

A mis padres por su apoyo incondicional, para ellos que constantemente me han inculcado en surgir en la vida.

A mi esposa, mi coequipera, con quien Dios nos cruzó en el camino, gratitud por estar cuando siempre lo he necesitado.

A mi hijita hermosa, mi princesita, lo más lindo que me ha regalado Dios, el motor que mueve y revoluciona mi vida.

### **Agradecimientos**

A la UNAD, Universidad de la cual soy graduado de Ingeniería en Telecomunicaciones y Especialista en Seguridad Informática por los conocimientos obtenidos en todos los años de estar involucrado con ella.

Al profesor Edgar Andrés Villabon Aldana, por contribuir con su entendimiento y experiencia en el soporte y progreso de esta monografía, agradecido por la contribución y orientación que ha ofrecido en este trabajo.

A los profesores Nidia Danigza Lugo López y Luis Ángel Anillo Arrieta, los cuales me encauzaron en el contenido de la monografía y han contribuido con las indicaciones para el buen adelanto del presente documento.

## Resumen

Esta monografía se centra en el uso de modelos de aprendizaje automático para abordar la problemática de la siniestralidad vial en Bogotá, una ciudad que enfrenta altos índices de accidentes de tránsito debido a factores diversos y complejos. Mediante la aplicación de técnicas avanzadas de análisis de datos, se busca descubrir patrones y tendencias globales que puedan ser relevantes para el contexto local. La investigación se basa en una selección meticulosa de datos específicos de siniestros viales en Bogotá, abarcando variables críticas como el tipo de accidente, las condiciones climáticas, la ubicación geográfica y el tipo de vehículo involucrado.

El análisis de estos datos mediante modelos de aprendizaje automático permitirá evaluar el rendimiento y la efectividad de estos modelos en la identificación de las causas subyacentes de los accidentes de tránsito. La interpretación de los resultados proporcionará una comprensión profunda de los factores que contribuyen a la siniestralidad vial en Bogotá, ofreciendo una base sólida para el desarrollo de intervenciones y políticas públicas orientadas a mejorar la seguridad vial. Este estudio no solo busca contribuir al conocimiento académico, sino también proporcionar herramientas prácticas para la toma de decisiones informadas que puedan reducir significativamente la incidencia de accidentes y sus consecuencias en la ciudad.

***Palabras clave:*** Machine learning, analítica de datos, modelos de predicción, siniestralidad vial.

## **Abstract**

This monograph focuses on the use of machine learning models to address the problem of road accidents in Bogotá, a city that faces high rates of traffic accidents due to diverse and complex factors. By applying advanced data analysis techniques, we seek to discover global patterns and trends that may be relevant to the local context. The investigation is based on a meticulous selection of specific data from road accidents in Bogotá, covering critical variables such as the type of accident, weather conditions, geographical location and the type of vehicle involved.

The analysis of this data using machine learning models will allow the performance and effectiveness of these models to be evaluated in identifying the underlying causes of traffic accidents. The interpretation of the results will provide a deep understanding of the factors that contribute to road accidents in Bogotá, offering a solid basis for the development of interventions and public policies aimed at improving road safety. This study not only seeks to contribute to academic knowledge, but also to provide practical tools for making informed decisions that can significantly reduce the incidence of accidents and their consequences in the city.

***Keywords:*** Machine learning, data analytics, prediction models, road accidents.

## Tabla de Contenido

Introducción .....	10
Planteamiento del Problema .....	12
Justificación .....	14
Objetivos .....	16
Objetivo General .....	16
Objetivos Específicos .....	16
Marco de Referencia .....	17
Marco Conceptual .....	17
Marco Teórico .....	18
Identificación de Algunos Modelos de Machine Learning Relacionados con la Siniestralidad Vial y sus Resultados .....	23
Modelo de Análisis Correlacional para la Identificación de Patrones Asociados a los Accidentes de Tránsito de Yopal .....	23
Modelo Predictivo de Zonas de Riesgo Espacio – Temporal de Accidentes de Tráfico en la Ciudad de Manizales .....	26
Análisis de Siniestralidad Peatonal a Partir de la Metodología de Machine Learning .....	30
Selección de Datos de los Siniestros Viales Registrados en los Informes Policiales de Accidentes de Tránsito Ocurridos en Bogotá .....	35
Limpieza de Datos .....	36
Selección de Variables .....	38
Diseño de Modelos de Machine Learning .....	38
División de la Base de Datos .....	38

Entrenamiento y Predicciones.....	39
Métricas de Evaluación.....	40
Comparación del Desempeño .....	44
Interpretación del Rendimiento de los Modelos Diseñados y sus Métricas Verificando su Efectividad en la Identificación de Causas de Siniestralidad Vial en Bogotá .....	45
Métricas Regresión Logística.....	45
Métricas Árbol de Decisión .....	45
Métricas K-Nearest Neighbors (KNN) .....	45
Matriz de Confusión de Regresión Logística.....	46
Matriz de Confusión de Árbol de Decisión.....	46
Matriz de Confusión de K-Nearest Neighbors (KNN).....	47
Comparación y Selección del Mejor Modelo .....	47
Análisis Estadístico de las Variables que Podrían Influir en la Decisión de la Causalidad de la Siniestralidad Vial, de Acuerdo a los Datos Obtenidos .....	48
Top 5 de las Variables con Mayor Frecuencia de Ocurrencia .....	55
Conclusiones.....	58
Recomendaciones .....	60
Referencias.....	63

## Lista de Figuras

<b>Figura 1</b> <i>Regresión Lineal Simple</i> .....	24
<b>Figura 2</b> <i>Grafica de la Cantidad de Accidentes Anual</i> .....	27
<b>Figura 3</b> <i>Resumen Accidentes de Tránsito en Manizales – Dashboard</i> .....	30
<b>Figura 4</b> <i>Porcentaje de Peatones Heridos y Muertos, en un Siniestro Vial de la Ciudad de Bogotá D.C (2015-2019)</i> .....	33
<b>Figura 5</b> <i>Matriz de Confusión del Modelo de Penalización de la Clase Minoritaria</i> .....	34
<b>Figura 6</b> <i>Código Python Utilizado para Realizar la Limpieza de los Datos</i> .....	37
<b>Figura 7</b> <i>Resultado de la Limpieza</i> .....	37
<b>Figura 8</b> <i>Código Python Utilizado para la División (Entrenamiento - Prueba)</i> .....	39
<b>Figura 9</b> <i>Resultado División (Entrenamiento - Prueba)</i> .....	39
<b>Figura 10</b> <i>Código Python Utilizado para Entrenar los Modelos y Realizar las Predicciones</i> ....	40
<b>Figura 11</b> <i>Código Python para la Matriz de Confusión y Cálculo de Métricas</i> .....	41
<b>Figura 12</b> <i>Matriz de Confusión para el Modelo Regresión Logística</i> .....	42
<b>Figura 13</b> <i>Matriz de Confusión para el Modelo Árbol de Decisión</i> .....	43
<b>Figura 14</b> <i>Matriz de Confusión para el Modelo KNN</i> .....	43
<b>Figura 15</b> <i>Código Python para la Comparación de los Modelos Diseñados</i> .....	44
<b>Figura 16</b> <i>Distribución de la Gravedad de los Accidentes</i> .....	48
<b>Figura 17</b> <i>Distribución del Tipo de Accidente</i> .....	49
<b>Figura 18</b> <i>Distribución del Diseño del Lugar</i> .....	50
<b>Figura 19</b> <i>Distribución del Actor Vial en el Accidente</i> .....	51
<b>Figura 20</b> <i>Relación entre Tipo de Accidente y Gravedad</i> .....	52
<b>Figura 21</b> <i>Relación entre Actor Vial y Gravedad</i> .....	53

<b>Figura 22</b> <i>Relación entre Diseño del Lugar y Gravedad</i> .....	54
<b>Figura 23</b> <i>Top 5 Valores de Gravedad</i> .....	55
<b>Figura 24</b> <i>Top 5 Valores de Tipo de Accidente</i> .....	56
<b>Figura 25</b> <i>Top 5 Valores de Diseño de Lugar</i> .....	56
<b>Figura 26</b> <i>Top 5 Valores de Actor Vial</i> .....	57
<b>Figura 27</b> <i>Top 5 Valores de Código Causa</i> .....	57

## Introducción

La movilidad y la seguridad vial son aspectos fundamentales en el desarrollo de las ciudades modernas, especialmente en una ciudad como Bogotá, donde el crecimiento poblacional y el aumento de vehículos generan desafíos constantes en la gestión del tránsito. En este contexto, la siniestralidad vial representa uno de los problemas más críticos, afectando no solo la integridad física de los ciudadanos, sino también la eficiencia de la movilidad urbana y la calidad de vida en la ciudad.

Identificar las causas subyacentes de los accidentes de tránsito y comprender los factores que influyen en su gravedad son pasos esenciales para diseñar estrategias efectivas de prevención. En este sentido, las técnicas de aprendizaje automático ofrecen herramientas avanzadas para analizar grandes volúmenes de datos históricos y generar predicciones precisas sobre los patrones de siniestralidad. Estos modelos no solo permiten identificar puntos críticos y factores de riesgo, sino que también contribuyen al diseño de políticas públicas más informadas y orientadas a la reducción de accidentes.

La presente monografía se centra en evaluar el desempeño de tres modelos de aprendizaje automático —Regresión Logística, Árbol de Decisión y K-Nearest Neighbors (KNN)— para clasificar la gravedad de los accidentes viales en Bogotá. A partir de los resultados obtenidos, se identificaron las posibles causas más relevantes de la siniestralidad vial, así como patrones asociados a variables clave como el tipo de accidente, el diseño del lugar, los actores viales involucrados y las causas registradas.

Los hallazgos de este estudio no solo permiten comprender mejor la problemática de la siniestralidad vial, sino que también brindan recomendaciones prácticas para mejorar la seguridad vial en Bogotá. Además, este análisis refuerza la importancia de integrar herramientas

de aprendizaje automático en la planificación de la movilidad urbana, fomentando una gestión más proactiva y eficiente del tránsito en la ciudad.

## **Planteamiento del Problema**

La siniestralidad vial es un problema global que afecta a ciudades de todo el mundo, causando pérdidas humanas, económicas y sociales significativas. En el caso específico de Bogotá, la siniestralidad vial es una preocupación importante debido a la alta densidad de tráfico, la infraestructura vial compleja y la diversidad de factores que contribuyen a los accidentes de tránsito. En el año 2023, y de acuerdo a un informe presentado por la Agencia Nacional de Seguridad Vial, se registró un aumento en la tendencia de muertes con causa de accidentes en la infraestructuras viales de Colombia (Largo, 2024).

La problemática radica en la necesidad de identificar y comprender las causas subyacentes de la siniestralidad vial en la ciudad de Bogotá, con el fin de desarrollar estrategias efectivas para prevenir y reducir la incidencia de accidentes. De acuerdo al Anuario de siniestralidad vial 2022 presentado por la Secretaria de Movilidad de Bogotá, un total de 526 personas fallecieron como consecuencia de accidentes de tránsito (Gómez-Montoya & Cuartas-Arias, 2020).

Se pretende indagar qué papel juegan las tecnologías emergentes, como los modelos de aprendizaje automático, en la identificación y predicción de las causas de los accidentes de tránsito en las ciudades. Debido al impacto de la problemática, se ha buscado reducir las pérdidas humanas, económicas y materiales haciendo uso de la minería de datos y analítica, identificando e interpretando los resultados que puedan orientar y determinar situaciones comunes donde posiblemente haya una mayor exposición a un siniestro vial.

De igual forma, la complejidad de la problemática y la importancia de realizar un análisis, utilizando los datos espaciotemporales de los accidentes de tránsito y otras variables relacionadas, sean analizados y utilizados de manera efectiva y puedan apreciar la probabilidad

de que ocurran siniestros viales en alguna zona geográfica establecida, de manera que pronostique estos hechos y se puedan validar labores para su prevención, mejorando la seguridad vial y reduciendo la incidencia de accidentes.

## Justificación

Se fundamenta en la necesidad de abordar un problema de gran relevancia tanto a nivel global como local. Se considera la siniestralidad global y se pretende realizar un acercamiento específico a Bogotá. La siniestralidad vial es un problema de salud pública y seguridad ciudadana a nivel mundial, que causa millones de muertes y lesiones cada año. Los accidentes de tránsito son una de las principales causas de mortalidad y discapacidad en el mundo, especialmente entre la población joven y productiva (Martínez-Ramos et al., 2021).

Los accidentes de tránsito generan un impacto económico en términos de costos médicos, pérdida de productividad laboral y daños materiales. Además, afectan la calidad de vida y tienen repercusiones en el sistema de salud, el transporte y la infraestructura vial. Bogotá, una de las ciudades más grandes y pobladas de América Latina, enfrenta en materia de seguridad vial, alta densidad de tráfico, la complejidad de la infraestructura vial, el comportamiento de los conductores y peatones, y las condiciones climáticas variables que contribuyen a la siniestralidad vial.

En el contexto actual, el uso de tecnologías avanzadas como los modelos de aprendizaje automático ofrece nuevas oportunidades para analizar y comprender mejor las causas de los accidentes de tránsito. Estos modelos pueden procesar grandes cantidades de datos, identificar patrones ocultos y predecir tendencias futuras con mayor precisión que los enfoques tradicionales.

La identificación de las causas subyacentes de la siniestralidad vial en Bogotá es crucial para desarrollar intervenciones efectivas y políticas públicas orientadas a la prevención y reducción de accidentes. El análisis de modelos machine learning puede proporcionar información valiosa para tomar decisiones informadas y estratégicas en materia de seguridad vial.

(Cruz & Herrera, 2023).

## **Objetivos**

### **Objetivo General**

Analizar el desempeño de modelos de aprendizaje automático, identificando las posibles causas de la siniestralidad vial en Bogotá, contribuyendo a la mejora de la seguridad vial en la ciudad.

### **Objetivos Específicos**

Identificar los datos de la masa documental, indagando patrones y tendencias relacionadas con las posibles causas de los accidentes de tránsito a nivel global.

Seleccionar datos sobre siniestros viales en Bogotá, incluyendo variables como tipo de accidente, condiciones climáticas, ubicación geográfica, tipo de vehículo, entre otros.

Interpretar el rendimiento de los modelos utilizando métricas de evaluación adecuadas, determinando si son efectivos en la identificación de causas de siniestralidad vial en Bogotá.

## Marco de Referencia

### Marco Conceptual

**Analítica de Datos:** Indica el desarrollo de investigar y sondear acumulados de información (en tiempo efectivo, significativos, no distribuidos, distribuidos, cualitativos) para manifestar información apreciable y mediante métodos estadísticos e instrumentos computacionales, averiguando e identificando muestras, directrices y conexiones reservadas en la información, con el propósito de conseguir entendimientos que promuevan la determinación y forjen optimizaciones en diferentes ámbitos (TIBCO, s.f).

**Machine Learning:** Hace referencia a la sección de la inteligencia artificial que admite a los artefactos, instruirse instintivamente y optimar su ejecución mediante el estudio de datos. Los artefactos logran reseñar modelos y determinar decisiones fundadas en esto, empleando algoritmos y patrones (BBVA, 2019).

**Modelos de predicción:** Son instrumentos o algoritmos que se emplean para hacer apreciaciones o predicciones acerca de sucesos venideros o productos excluidos fundados en información histórica. Estos patrones se cimientan manejando métodos estadísticos o de aprendizaje automático y se emplean en diferentes sectores (edx.com, s.f).

**Observatorio de Movilidad:** Es un instrumento de administración de información que colecciona y examina acerca de los traslados y la movilidad en un sector determinado que admiten sobre la base de observadores, ingresar a los datos establecidos en mapas, imagenes y tablas con selecciones de intercambio, que preparan la indagación de los conjuntos de interés y segmentos afectados (Movilidad Bogotá, 2020).

**Regresión Lineal:** Es un método de analizar de información que admite pronosticar la utilidad de los datos inexplorados sobre la base de la utilización de diferentes valores de datos

con cualquier dependencia y distinguido. Forma exactamente la variable inexplorada o dependiente y la variable distinguida o independiente como una ecuación lineal (AWS, 2023).

Regresión logística: Es un método de análisis de datos que emplea las matemáticas para hallar las correspondencias entre dos elementos de datos, y maneja esta relación para pronosticar el valor de alguno de esos elementos apoyándose en el otro. Este pronóstico tiene un dígito definido de efectos, como un sí o un no (AWS, 2023).

Redes neuronales artificiales: Son estándares informáticos que se manejan en machine learning para solventar cuestiones complicadas. Están ajustados por nodos interrelacionados que se conocen como neuronas artificiales, que toman, ejecutan y transfieren datos por medio de enlaces (Walther, 2023).

Clustering o agrupamiento: Es un modelo machine learning no supervisado que se sirve para hallar factores y configuraciones escondidas en un grupo de datos. Radica en aglomerar objetos equivalentes en conjuntos, que se conocen como clústeres, conforme a su equivalencia o cercanía en tipologías o atributos determinados (Martínez, 2023).

### **Marco Teórico**

Indagando el grado de utilización de estándares de análisis de datos y aprendizaje automático (machine learning) en diferentes fuentes aplicadas por diversos autores, buscando examinar cuál es su estado de aplicabilidad y ejecución, crea la obligación de efectuar una exploración documental al grado de varios trabajos realizados en diferentes lugares dentro y fuera del país, determinando los primordiales modelos aplicados y los inconvenientes que reconocen para poder remediar con su implementación el problema identificado (Cruz & Herrera, 2023).

Luego de efectuar la investigación acerca de la puesta en funcionamiento de análisis de

información y aprendizaje automático, tanto a en el campo nacional como en el campo internacional, se hace necesario ahondar en esos trabajos que han contribuido concretamente en el terreno de siniestralidad con el objeto de estar al tanto cómo se han emprendido esa clase de dilemas, que modelos se han implementado y qué efectos se alcanzaron (Cruz & Herrera, 2023).

En la revisión de que instrumentos fueron empleados, diversas de las fuentes usaron regresión logística, en otras se utilizó el modelo de regresión lineal, en otras redes neuronales artificial y finalmente decision tree (CART). Los efectos alcanzados con los instrumentos previamente señalados, consiguieron determinar los horarios y los días en determinado periodo en que se mostraban con más reiteración los siniestros y de qué forma categorías como edad o la clase de vehículo, puedan ser elementos que más incurrieran en la siniestralidad, con lo que se consigue deducir pronósticos que admitan organizar tácticas para la disminución de accidentes viales o de tránsito.

Estas predicciones posibilitan relacionar que parte del tránsito está ejerciendo la utilización de análisis de datos por medio de instrumentos de aprendizaje automático (machine learning) para el pronóstico de sucesos de siniestralidad, con el propósito de puntualizar planes de movilidad segura y en diversas situaciones, resguardándose en monitoreo de tránsito.

El análisis de la siniestralidad es uno de los elementos cruciales para asegurar la prevención de accidentes de tránsito y optimizar la validez en la movilidad. Últimamente, el análisis de información y aprendizaje automático (machine learning), han surgido como instrumentos poderosos para indagar amplias dimensiones de información y conseguir discernimientos apreciables que contribuyen a advertir accidentes, reconociendo condiciones de amenaza vial y optimizando las estrategias de seguridad vial alcanzadas (Cruz & Herrera, 2023). Desde el año 2010, se ha contemplado un incremento en el valor en la implementación de estos

métodos en el análisis de siniestralidad vial.

En el trabajo “Mortalidad por accidentes de tráfico en Colombia. Estudio comparativo con otros países” concierne una indagación de causas de amenaza vial relacionados como causas habituales en la siniestralidad, en Colombia del mismo modo que en otros países con los que se plasmó el estudio cotejado (España y Estados Unidos). Los resultados más notables muestran la influencia que tiene el avance socio-cultural de los países estudiados, con relación a los factores que originan los siniestros viales y la periodicidad con la que estos suceden, precisando que los motivos más notables en nuestro país, están vinculados con el comportamiento y costumbres de los ciudadanos y el uso de las emociones que se forjan en las diferentes situaciones y en escenarios específicos. (Alarcón et al., 2018).

Con relación a ilustraciones vinculadas que hayan avanzado en nuestro país, en la investigación “Modelo predictivo de accidentes de tránsito”, cumple con un estudio de variables vinculadas a causas que alcanzan a incurrir en la aptitud de respuesta de los pilotos en Medellín, sobre la base de regresión lineal, infiriendo que el pronóstico de estos hechos no es eficaz puesto que el fallo aumenta en el tiempo del adiestramiento del modelo (Redondo, 2023).

En la revisión del documento “Análisis de Siniestralidad Peatonal A partir de la metodología de Machine Learning”, se destaca una modelación de datos correspondiente a la Secretaría de Movilidad Bogotá D.C en un rango desde enero del año 2015 y diciembre del año 2019, en donde hubo un registro de 157.499 accidentes viales y cuyas variables principales dentro del análisis fueron horario de ocurrencia, forma en que ocurrió, ubicación, tipo de persona afectada, género y edad de las personas que estuvieron involucradas en los siniestros reportados. Se trabajó con modelos como regresión logística, árbol de decisión, K-NN en los que se analizaron las métricas accuracy, recall, F1-score y precisión, los cuales tuvieron diferentes

desempeños en la predicción final, para lo que se sugiere incluir variables para analizar como rapidez del vehículo, la señalización vial e información del conjunto de implicados, que puedan aportar un alto desenvolvimiento en un patrón de aprendizaje automatizado (machine learning) (Martínez-Ramos et al., 2021).

Dentro de la masa documental, se encontró el trabajo “Modelo de pronóstico de incidentes en conducción de motocicletas basado en IOT e inteligencia artificial” en el cual se realizaron la comparación de tres modelos de machine learning para el conjunto de datos (Decisión Tree, KNN (k Nearest-Neighbor) y Random Forest). Los tres modelos efectuados suministraron resultados favorables en buenos índices con relación a la valoración de las clases, no obstante, el modelo de random forest (bosques aleatorios), muestra una ligera prelación en los resultados concebidos, por las métricas de estimación, el procedimiento de catalogar adecuadamente en categorías de riesgo la conducta al manejar moto (Cárdenas-Lancheros, 2021).

Otro de los antecedentes de aplicación de modelos machine learning en la siniestralidad vial fue “Modelo de análisis correlacional para la identificación de patrones asociados a los accidentes de tránsito de Yopal” se indago establecer un modelo correlacional que admita identificar las particularidades con gran ocurrencia en las siniestralidades viales en estipuladas franjas. Se adelantó para la obtención del objetivo planteado una regresión lineal simple, obteniendo como respuesta la detección de factores recurrentes relacionados a una propiedad establecida en los puntos con gran siniestralidad vial (Oregón-Fuentes & Castellom-Rodríguez, 2021).

Acorde a lo antepuesto, se demuestra que se han avanzado en diferentes observaciones en los que se instauran las causas más notables en la categoría vial para los sucesos o deducción de

sucesos de siniestralidad, no obstante, no se han perfeccionado modelos que admitan pronosticar la siniestralidad Bogotá.

## **Identificación de Algunos Modelos de Machine Learning Relacionados con la Siniestralidad Vial y sus Resultados**

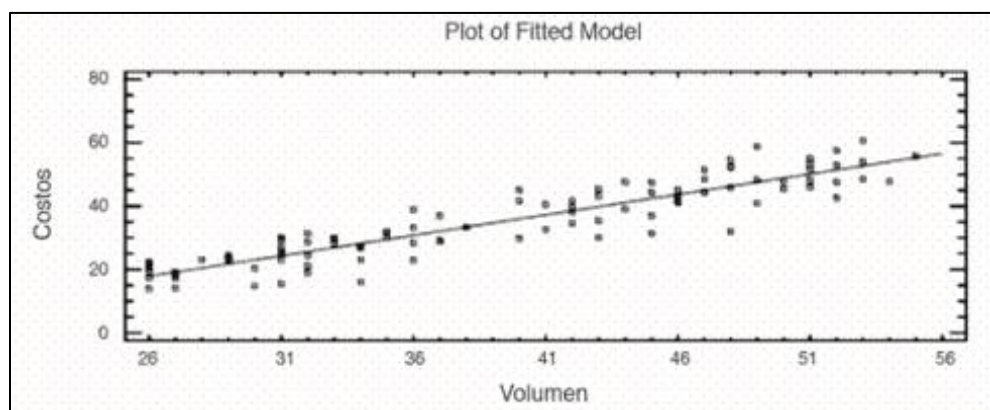
### **Modelo de Análisis Correlacional para la Identificación de Patrones Asociados a los Accidentes de Tránsito de Yopal**

Contexto: Este modelo pretende generar un modelo correlacional que pueda asemejar las singularidades con más ocasiones en los siniestros viales en ciertas franjas de la ciudad de Yopal en el departamento del Casanare - Colombia, y mostrar los esquemas coligados a los mismos. En la cimentación del modelo se emplean notaciones de regresión lineal simple de aprendizaje supervisado, montando la correlación entre variables dependientes e independientes. La meta es conseguir un modelo lineal que exponga la correspondencia entre las variables en la información de siniestros viales en la ciudad de Yopal y prever el coeficiente de determinación  $r^2$  y su significancia (p-value).

En la regresión simple el cómputo de la igualdad adecuada a la línea que más relata la correspondencia entre el objetivo y la variable que la expone. Tal igualdad simboliza la línea que más concuerda a los puntos en una representación de dispersión como se logra visualizar en la siguiente figura.

## Figura 1

### Regresión Lineal Simple



*Nota.* Tomada de Ángela Bibiana Oregón Fuentes, Juan David Castellom Rodríguez - Modelo de Análisis Correlacional para la Identificación de Patrones Asociados a los Accidentes de Tránsito de Yopal

Base de datos: La muestra que se tuvo en cuenta para la elaboración de la investigación, son los siniestros viales sucedidos en el rango de enero del 2017 a agosto del 2020, los cuales están suministrados por 1394 accidentes.

VARIABLES: Las variables que actúan en los siniestros viales que se tuvieron en cuenta para el desarrollo del modelo fueron:

- ✓ fecha
- ✓ gravedad
- ✓ edad
- ✓ día de la semana
- ✓ grado de alcoholimetría

Estas variables se corresponden con las variables independientes de precipitaciones meteorológicas. Además, se observan los puntos con más siniestralidad de la ciudad y los

elementos de mayor relevancia por cada punto, admitiendo a las entidades implicadas asemejar las causas que originan los sucesos, actuando y soslayar su repetición.

Descripción de la investigación: El desarrollo de la investigación se efectuó en fases, las cuales son las siguientes:

Fase 1. Selección de los datos a través del rediseño, actualización del aplicativo RECOATY digitalización de la información.

Fase 2. Pre procesamiento de la información y transformación de los datos.

Fase 3. Identificación de los puntos críticos de accidentalidad por georreferenciación.

Fase 4. Caracterización e identificación de los factores de riesgos asociados a los accidentes de tránsito.

Luego de las anteriores fases se procede a implementar el modelo correlacional para la identificación de patrones asociados teniendo en cuenta la información recolectada en cada punto crítico empleando regresión lineal simple.

Interpretación de Resultados: Una de las conjeturas enunciadas a partir de la identificación de los sucesos de siniestros viales en la ciudad de Yopal (Casanare), instauraba que bajo circunstancias de lluvia constaba una alta posibilidad de que pasara el siniestro vial, de igual forma, una alta magnitud en los perjuicios de los individuos implicados lo mismo que en la zona urbana así como en las vías, en virtud a las circunstancias estructurales de ellas y a los niveles de luminosidad, sin embargo, se analizó los datos, esta conjetura precisa de validez, toda vez que los siniestros en circunstancias de lluvia no son relevantes en relación a los incidentes viales mostrados en clima normal o soleado, del mismo modo los perjuicios graves nunca representan más ocurrencia ante estos sucesos.

Por otra parte, se reflexiona que las tres importantes razones de la siniestralidad de la

ciudad Yopal son la imprudencia, manejar en estado de embriaguez y el exceso de velocidad, logran suponerse como las más importantes razones de siniestralidad vial.

### **Modelo Predictivo de Zonas de Riesgo Espacio – Temporal de Accidentes de Tráfico en la Ciudad de Manizales**

Contexto: En la ciudad de Manizales del departamento de Caldas (Colombia), los siniestros viales son la tercer causa primordial de decesos, seguida por los homicidios y suicidios. El pronóstico del riesgo de siniestros con mayor determinación espacio-temporal es un reto característico, fundamentalmente a causa del complicado ambiente vial, la conducta humana y la inexactitud de información en tiempo real.

Pese a que las técnicas de machine learning han evidenciado renovaciones importantes en el pronóstico en balance con los modelos acostumbrados, aún constan dudas para usar estas técnicas en otras naciones como en la nuestra, donde la infraestructura y la existencia de información no siempre son superiores.

En este estudio, se propone utilizar múltiples técnicas de aprendizaje automático para mejorar la precisión de las predicciones. Estas técnicas incluyen enfoques basados en series temporales, como Facebook Prophet, así como algoritmos más avanzados, como Light Gradient Boosting y redes neuronales recurrentes, específicamente Long Short-Term Memory (LSTM). Estas técnicas se seleccionan en base a su capacidad demostrada en el estado del arte y su capacidad para procesar datos y aprender patrones.

Para llevar a cabo esta investigación, se construyen modelos para toda la ciudad y cada zona utilizando datos de 18 años en general y de 7 años por zonas, lo que permite obtener un panorama amplio y detallado del comportamiento de los accidentes de tráfico en la ciudad. Se analiza el desempeño de cada modelo en función de su capacidad para predecir accidentes, y se

selecciona el mejor modelo para cada zona. Con el fin de evaluar el rendimiento de los modelos y la precisión de las predicciones, se emplean las métricas MAPE y RMSE, que son ampliamente utilizadas en el campo de la predicción y permiten una comparación efectiva entre diferentes técnicas y enfoques.

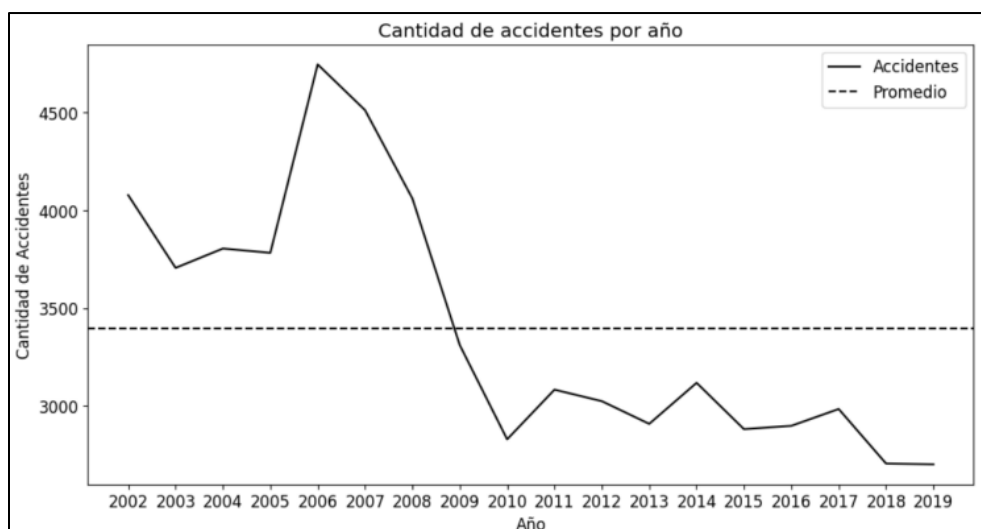
Base de Datos: El estudio analiza los datos de accidentes de tránsito ocurridos en la ciudad de Manizales desde 2002, los cuales están almacenados en una base de datos relacional.

Para este trabajo se seleccionó un conjunto de datos entre los años 2002 a 2019 de acuerdo a la relevancia que representan los accidentes más actuales sobre años anteriores y excluyendo el año 2020 debido a que contiene variaciones atípicas causadas por la pandemia del COVID.

La cantidad total de accidentes de tránsito para el periodo comprendido entre el 1 de enero de 2002 y el 31 de diciembre de 2019 es de 61.156 accidentes de tránsito.

## Figura 2

*Grafica de la Cantidad de Accidentes Anual*



*Nota.* Tomada de Juan Camilo Cardona Álvarez - Modelo Predictivo de Zonas de Riesgo Espacio – Temporal de Accidentes de Tráfico en la Ciudad de Manizales.

VARIABLES: Para el modelo se tuvieron en cuenta las siguientes variables que son contenidas en los informes de los agentes de tránsito de la ciudad:

- ✓ Lugar, fecha y hora del suceso.
- ✓ Clase de vehículo, placa y características.
- ✓ Nombre del conductor o conductores, documentos de identidad, número de la licencia o licencias de conducción, lugar y fecha de su expedición y número de la póliza de seguro y compañía aseguradora, dirección o residencia de los involucrados.
- ✓ Nombre del propietario o tenedor del vehículo o de los propietarios o tenedores de los vehículos.
- ✓ Nombre, documentos de identidad y dirección de los testigos.
- ✓ Estado de seguridad, en general, del vehículo o de los vehículos, de los frenos, de la dirección, de las luces, bocinas y llantas.
- ✓ Estado de la vía, huella de frenada, grado de visibilidad, colocación de los vehículos y distancia, la cual constará en el croquis levantado.
- ✓ Descripción de los daños y lesiones.
- ✓ Relación de los medios de prueba aportados por las partes.
- ✓ Descripción de las compañías de seguros y números de las pólizas de los seguros obligatorios exigidos por este código.

Descripción de la investigación: La preparación de los datos para el análisis de predicción de accidentes de tránsito en la ciudad de Manizales se llevó a cabo en dos fases principales:

La reestructuración y preprocesamiento de datos inicialmente, se obtuvieron datos de la Secretaría de Movilidad, en los que cada accidente contaba con un número de croquis asignado, así como una fecha y hora específicas. Estos datos necesitaban ser transformados en un formato

apropiado para su utilización con el algoritmo de predicción Prophet. La primera etapa de dicha transformación se detalla en la sección de tratamiento de datos.

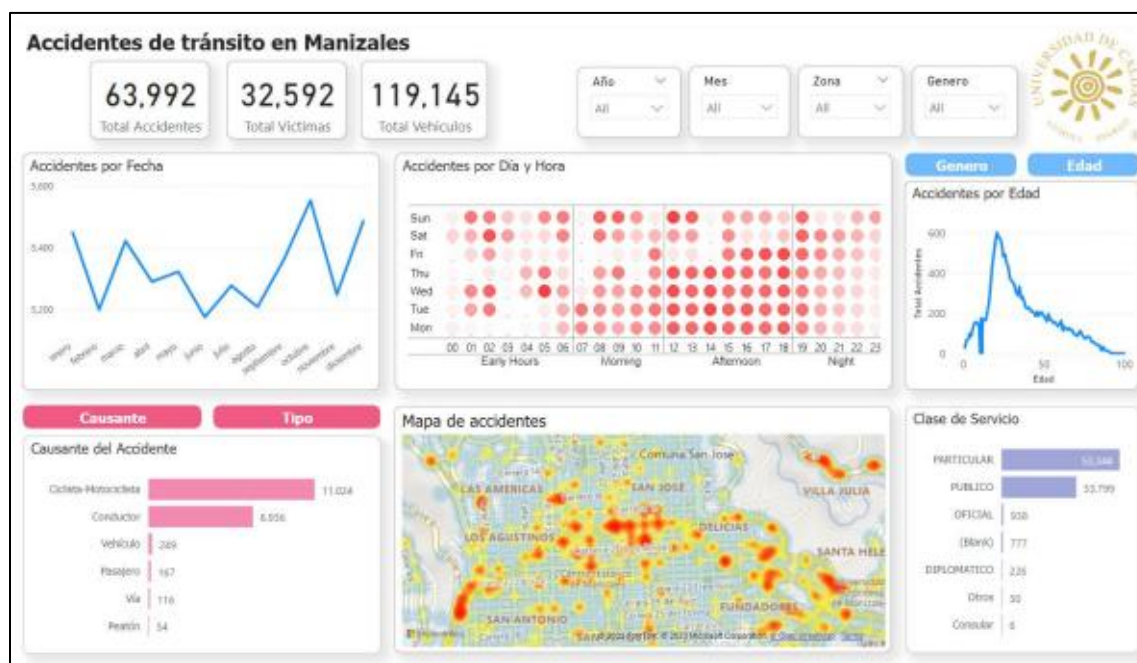
Posteriormente, se incorporaron los regresores, es decir, las variables predictivas adicionales a la fecha del accidente, en el formato específico que requiere Prophet.

La implementación Light Gradient Boosting (LGB) es un algoritmo de aprendizaje automático basado en árboles de decisión que se utiliza para la clasificación y regresión en grandes conjuntos de datos. Este algoritmo se basa en la técnica de impulso, que combina varios modelos débiles para crear un modelo fuerte. LGB es conocido por su capacidad para trabajar con grandes conjuntos de datos y su eficiencia en términos de tiempo y recursos.

Interpretación de Resultados: Los resultados mostraron que el algoritmo Prophet tuvo el mejor rendimiento en general para predecir accidentes de tráfico en Manizales, seguido por LSTM y LGB. El análisis de los resultados sugiere que la combinación de diferentes técnicas de aprendizaje automático puede mejorar la precisión de las predicciones y proporcionar información valiosa para las autoridades en la toma de decisiones, esto se da debido a que hay modelos que tienen mejor desempeño en una zonas o dimensiones temporales que otros, pero en otras zonas o dimensiones temporales son otros los modelos que tienen mejor desempeño, además, se observó que el rendimiento de los modelos variaba según la zona y el horizonte temporal, lo que indica que se pueden utilizar simultáneamente diferentes modelos en función de la zona y horizonte temporal en la que tienen mejor desempeño.

**Figura 3**

*Resumen Accidentes de Tránsito en Manizales – Dashboard*



*Nota.* Tomada de Juan Camilo Cardona Álvarez - Modelo Predictivo de Zonas de Riesgo Espacio – Temporal de Accidentes de Tráfico en la Ciudad De Manizales.

Los modelos de predicción de accidentes de tráfico, como Prophet, LGB y LSTM, muestran un mejor desempeño en horizontes temporales más amplios, siendo más efectivos en predicciones a medio y largo plazo.

### **Análisis de Siniestralidad Peatonal a Partir de la Metodología de Machine Learning**

**Contexto:** Se realiza un análisis en los accidentes de peatones, estando en los principales motivos de decesos a nivel mundial, razón por la cual la Organización Mundial de la Salud (OMS) la ha nombrado como una pandemia en sus documentaciones, esta disertación se hace sobre la base de la metodología Machine Learning.

El alcance de este estudio es pronosticar si un actor vial en un accidente de tránsito como

resultado sea muerto o herido. Se utilizó Oversampling para los datos de adiestramiento, cuyo resultado fue una matriz de confusión, en el que el estándar anuncia de modo ajustado a los heridos con una exactitud, precisión y confiabilidad del 90%, por el contrario, tan solo es competente de pronosticar la mitad de los fallecidos.

Base de Datos: La exploración de los siniestros pertenece a la base de datos de la Secretaría de Movilidad Bogotá D.C, durante los años 2015 y 2019. La observación de la base de datos fue provisto a través del sitio web Datos libres (Secretaría de Movilidad, 2021)

Se encontraron alrededor de 157.499 siniestros de tránsito, que estuvieron reconocidos por los agentes de tránsito. La información contiene datos con respecto de la temporalidad, tipo de accidente, localización, tipo de servicio, tipo de actor vial, género y edad de cada una de las personas involucrados en los sucesos.

En cuanto al análisis, se realizó una limpieza a la base de datos, ya que mostraba varios campos sin datos, a la par se puso como actor vial solamente a los peatones, resultando finalmente 3.632 accidentes viales. El grado de pronóstico para la gravedad del suceso fue catalogado en: Muertos y heridos.

La investigación se enfoca en un estudio supervisado de clasificación, en que se manejó una librería llamada PYCARET, valorando varios modelos entre los que se hallan: DecisionTreeClassifier, Logistic Regression, Ramdon Forest, entre otros, cada modelo es calculado con métricas como Accuracy, AUC, Recall, Precisión F1, que inmediatamente son cotejados con una matriz de confusión, con el objeto de conseguir un modelo con el principal desempeño.

Variables: Enfoque cuantitativo, en el cual se corresponden las variables que pronostican, con relación a la variable objetivo:

Variable Objetivo:

- ✓ Gravedad

Variables Predictoras:

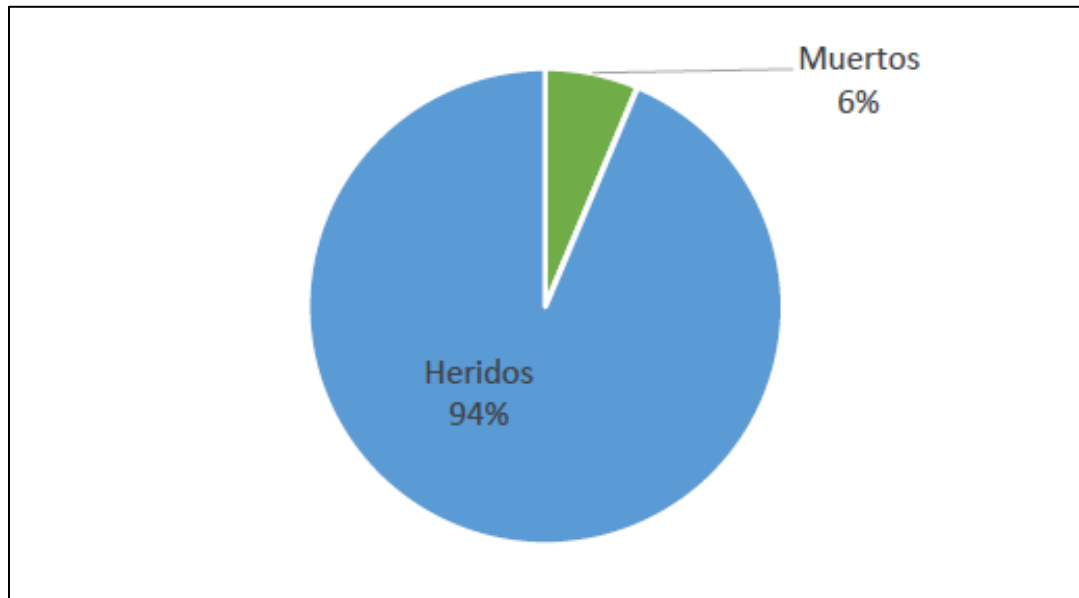
- ✓ Edad
- ✓ Sexo
- ✓ Tipo de Vehículo
- ✓ Hora
- ✓ Infraestructura Vial
- ✓ Ubicación del siniestro
- ✓ Número de carriles y ancho de carriles
- ✓ Vías principales, vías secundarias, Vías Locales
- ✓ Tiempo Meteorológico

Descripción de la investigación: Se realizó una etapa de datos de entrenamiento y datos de prueba, en esta fase fue primordial preparar la base de datos y realizar una comprobación de prueba con la información que concurre, con el objetivo que el modelo asimile y no repita, sin tener el inconveniente de underfitting o overfitting, por lo que se resolvió ocuparse con un modelo 80% con datos de adiestramiento y 20% de datos de prueba, de mismo modo, se utilizó la herramienta Pycaret.

Interpretación de Resultados: Se puede apreciar que el porcentaje de personas heridas en dependencia al porcentaje de personas muertas, es más alto, por lo que fue preciso balancear la información a través de herramientas para datos desbalanceado como Smote, respecto a los porcentajes de personas heridas y muertas, en concordancia con los días de la semana.

**Figura 4**

*Porcentaje de Peatones Heridos y Muertos, en un Siniestro Vial de la Ciudad de Bogotá D.C (2015-2019)*



*Nota.* Tomada de Magda Milena, Martínez Ramos, Leidy Johana Vásquez López, José John Fredy González Veloza - Análisis de Siniestralidad Peatonal a partir de la Metodología de Machine Learning.

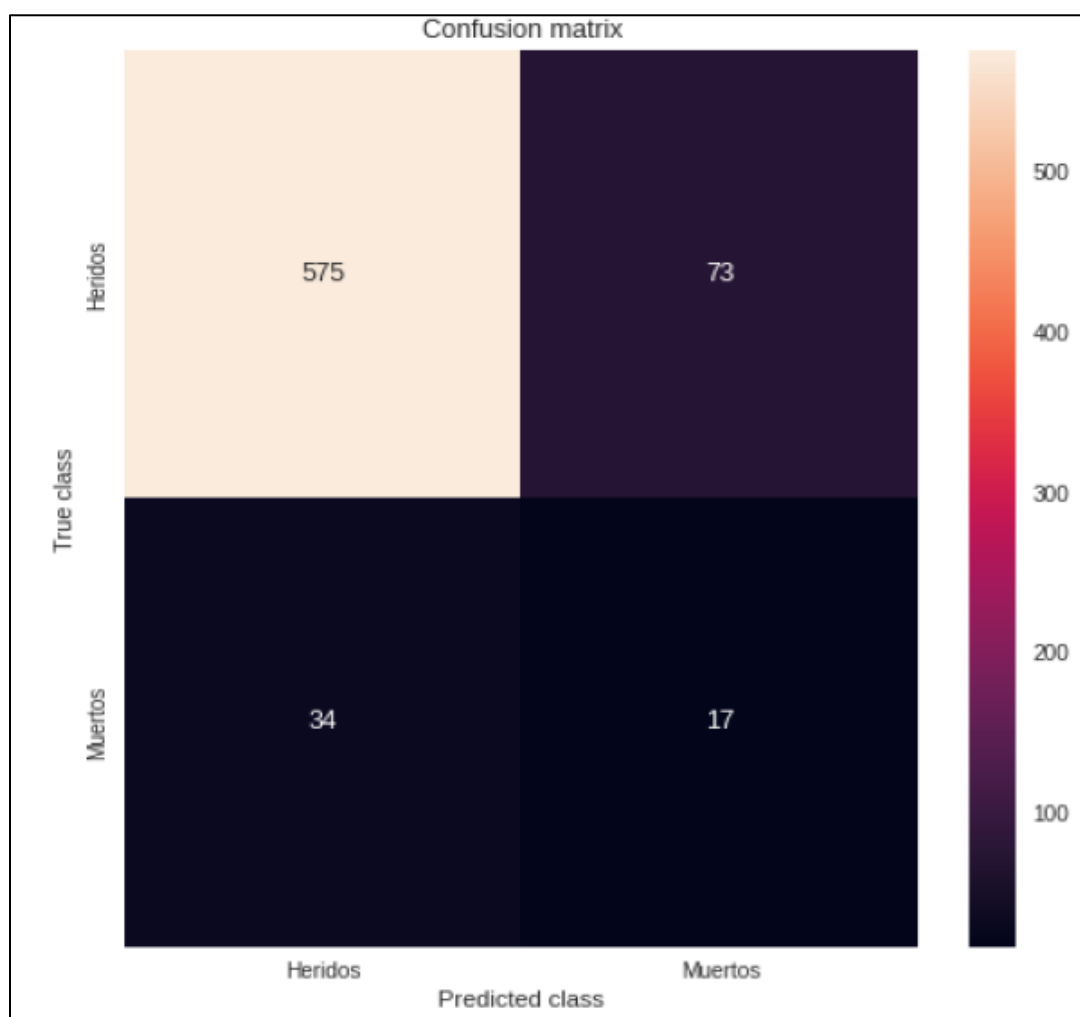
También en relación a los resultados, al poseer una base datos con la variable objetivo desbalanceada se proyectó una de las tácticas propuestas de acuerdo a especialistas, distinguida como Oversampling. Se halla correspondida la matriz de confusión con sus concernientes métricas, este modelo con el arreglo, pronostica de modo apropiado los heridos donde sus métricas para este tipo son del 89%, entre tanto que tan solo consigue a pronosticar el 50% de muertos.

El modelo que más se desempeñó logro proporcionando respuesta en parte al objetivo planteado, destacándose especialmente por su capacidad de manejo de desequilibrios en los

datos. Este rendimiento óptimo se obtuvo utilizando un enfoque penalizado, implementando la técnica de oversampling, que permitió aumentar artificialmente las muestras de las clases minoritarias. Esto mejoró la representatividad de estas clases en el modelo, mitigando los sesgos y favoreciendo un ajuste más equilibrado a los datos.

### Figura 5

*Matriz de Confusión del Modelo de Penalización de la Clase Minoritaria*



*Nota.* Tomada de Magda Milena, Martínez Ramos, Leidy Johana Vásquez López, José John Fredy González Veloza - Análisis de Siniestralidad Peatonal a partir de la Metodología de Machine Learning.

## **Selección de Datos de los Siniestros Viales Registrados en los Informes Policiales de Accidentes de Tránsito Ocurridos en Bogotá**

Se procedió a realizar una búsqueda de datos que pudiesen contener información acerca de los siniestros viales ocurridos en la ciudad de Bogotá, para lo cual se obtuvo una base de datos con el reporte alfanumérico de los incidentes viales registrados en los informes policiales de accidentes de tránsito ocurridos en Bogotá desde el año 2015. Estos datos se tomaron del sitio web de datos abiertos Bogotá. De esta base se adquirieron 459.799 registros de siniestros viales, lo cuales fueron seleccionados para realizar el diseño de modelos machine learning y la predicción de las posibles causas que puedan originarlos.

Inicialmente se hizo la planeación de lo que debía contener el análisis de los datos obtenidos, para lo cual se desarrollaron los siguientes ítems en búsqueda del diseño de modelos que puedan satisfacer el objetivo planteado:

- ✓ Limpieza de los datos, si es necesario.
- ✓ Seleccionar una o varias variables que sirvan en la predicción de posibles causas en los siniestros viales y su gravedad.
- ✓ Diseñar tres modelos de machine learning que permitan predecir las causas de los posibles siniestros viales en la ciudad.
- ✓ Seleccionar una base de entrenamiento (80%) y una base de prueba (20%) y con la base de entrenamiento aplicar los modelos diseñados.
- ✓ Realizar las predicciones para cada modelo.
- ✓ Diseñar una matriz de confusión por cada modelo y realizar el cálculo de las métricas de evaluación (Precision, Recall, accuracy y Specificity) para cada uno de los modelos, con los datos de prueba.

- ✓ Determinar cuál de los tres modelos es el de mejor desempeño.

La base de datos contiene 22 columnas, que incluyen información como la fecha, hora, gravedad del accidente, tipo de accidente, dirección, localidad, y características del vehículo y del conductor involucrado.

### **Limpieza de Datos**

Varias columnas, como CHOQUE\_CON, OBJETO\_FIJO y MODALIDAD\_VEHICULO, tienen valores nulos, por lo que se verificó la proporción de datos faltantes y se determinó si era necesario eliminarlas o imputarlas.

En la base de datos, algunas columnas tienen valores nulos significativos:

- OBJETO\_FIJO: 100% de valores nulos, por lo que es candidata a ser eliminada.
- CHOQUE\_CON: 60% de valores nulos, es posible eliminarla o imputar los valores si es relevante.
- MODALIDAD\_VEHICULO: 40% de valores nulos, podría ser útil si existe relación con el tipo de accidente.
- VEHICULO\_x: 10% de valores nulos, que podría imputarse o eliminarse según su relevancia.

Para la anterior actividad se decidió realizar las siguientes acciones:

- Eliminar las columnas OBJETO\_FIJO y VEHICULO\_x por su baja relevancia o falta de datos.
- Imputar valores en la columna CHOQUE\_CON.
- Mantener MODALIDAD\_VEHICULO para un análisis estadístico adicional.

## Figura 6

*Código Python Utilizado para Realizar la Limpieza de los Datos*

```
import pandas as pd

file_path = 'archivo_combinado - copia2.xlsx'
data = pd.read_excel(file_path)

# Eliminar columnas no necesarias
data_cleaned = data.drop(columns=['OBJETO_FIJO', 'VEHICULO_X'])

# Imputar valores en la columna `CHOQUE_CON`
data_cleaned['CHOQUE_CON'].fillna('Desconocido', inplace=True)

# Verificar los cambios realizados
print(data_cleaned.info())
print(data_cleaned.head())
```

## Figura 7

*Resultado de la Limpieza*

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 459799 entries, 0 to 459798
Data columns (total 20 columns):
#   Column                Non-Null Count  Dtype
---  -
0   CODIGO_ACCIDENTE      459799 non-null  int64
1   FECHA                 459799 non-null  object
2   HORA                  459799 non-null  object
3   GRAVEDAD              459799 non-null  object
4   TIPO_ACCIDENTE        459799 non-null  object
5   CHOQUE_CON            459799 non-null  object
6   DIRECCION             459799 non-null  object
7   LOCALIDAD             459799 non-null  object
8   DISEÑO_LUGAR          459799 non-null  object
9   CODIGO_ACCIDENTADO    459756 non-null  float64
10  ACTOR_VIAL            459756 non-null  object
11  ESTADO_SALUD          459756 non-null  object
12  EDAD                  459756 non-null  object
13  SEXO                  459756 non-null  object
14  VEHICULO_y            459756 non-null  object
15  CLASE_VEHICULO        459246 non-null  object
16  SERVICIO_VEHICULO     445902 non-null  object
17  MODALIDAD_VEHICULO    192628 non-null  object
18  ENFUGA                459756 non-null  object
19  CODIGO_CAUSA          459010 non-null  object
dtypes: float64(1), int64(1), object(18)
memory usage: 70.2+ MB
None
```

## Selección de Variables

Para predecir la gravedad de los siniestros viales, se consideran las variables que suelen influir en este aspecto:

Variables candidatas:

TIPO\_ACCIDENTE: indica el tipo de siniestro.

LOCALIDAD: puede dar información sobre zonas de mayor riesgo.

ACTOR\_VIAL: identifica el rol del actor involucrado (peatón, conductor, etc.).

ESTADO\_SALUD: estado físico de las personas involucradas.

CLASE\_VEHICULO: puede reflejar el tamaño y peso de los vehículos involucrados.

CODIGO\_CAUSA: describe la causa reportada del accidente.

## Diseño de Modelos de Machine Learning

Para un análisis más completo, se usaron tres modelos diferentes:

Regresión Logística: Un modelo adecuado para clasificar si un siniestro será leve, grave o muy grave.

Árbol de Decisión: Modelo que permite identificar las relaciones entre variables categóricas de forma intuitiva.

K-Nearest Neighbors (*KNN*): Para determinar similitudes entre siniestros y clasificar en función de los vecinos más cercanos.

## División de la Base de Datos

Se realizó la división de la base en conjuntos de entrenamiento y prueba:

## Figura 8

*Código Python Utilizado para la División (Entrenamiento - Prueba)*

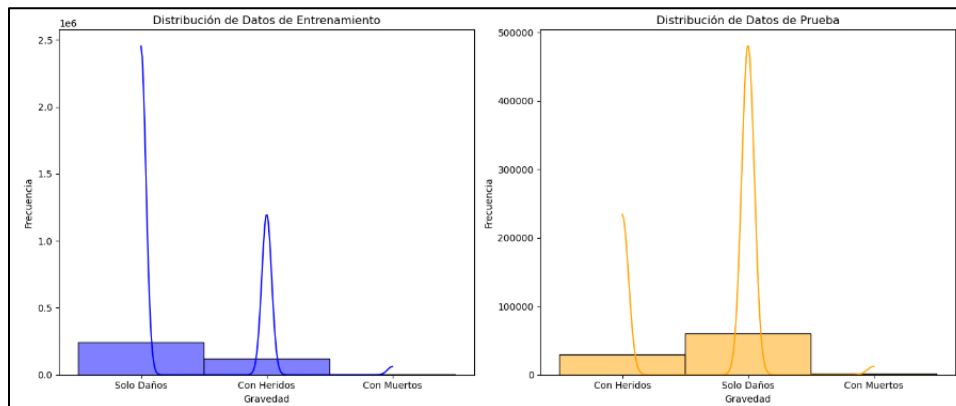
```
from sklearn.model_selection import train_test_split

# Separar las variables predictoras (X) y la variable objetivo (y)
X = data_cleaned[['TIPO_ACCIDENTE', 'LOCALIDAD', 'ACTOR_VIAL', 'ESTADO_SALUD', 'CLASE_VEHICULO', 'CODIGO_CAUSA']]
y = data_cleaned['GRAVEDAD'] # Asumiendo que 'GRAVEDAD' es la variable objetivo

# Dividir en conjunto de entrenamiento (80%) y de prueba (20%)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

## Figura 9

*Resultado División (Entrenamiento - Prueba)*



## Entrenamiento y Predicciones

Se procede a entrenar cada modelo con el conjunto de entrenamiento y hacer predicciones en el conjunto de prueba:

## Figura 10

### *Código Python Utilizado para Entrenar los Modelos y Realizar las Predicciones*

```

import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.tree import DecisionTreeClassifier
from sklearn.neighbors import KNeighborsClassifier

# Identificar las columnas categóricas
categorical_cols = X_train.select_dtypes(include=['object']).columns

# Aplicar One-Hot Encoding a las columnas categóricas en ambos conjuntos
X_train = pd.get_dummies(X_train, columns=categorical_cols, drop_first=True)
X_test = pd.get_dummies(X_test, columns=categorical_cols, drop_first=True)

# Asegurar de que las columnas de X_train y X_test sean las mismas
X_train, X_test = X_train.align(X_test, join='left', axis=1, fill_value=0)

import numpy as np

# Convertir X_train y X_test a arreglos de NumPy
X_train = X_train.values
X_test = X_test.values

# Entrenar los modelos
# Modelo 1: Regresión Logística
logreg = LogisticRegression(max_iter=1000)
logreg.fit(X_train, y_train)
y_pred_logreg = logreg.predict(X_test)

# Modelo 2: Árbol de Decisión
tree = DecisionTreeClassifier()
tree.fit(X_train, y_train)
y_pred_tree = tree.predict(X_test)

# Modelo 3: KNN
knn = KNeighborsClassifier()
knn.fit(X_train, y_train)
y_pred_knn = knn.predict(X_test)

print("Modelos entrenados y predicciones realizadas con éxito.")

```

## Métricas de Evaluación

Se realiza el diseño de la matriz de confusión y cálculo de las métricas de evaluación (Precision, Recall, accuracy and Specificity) para cada uno de los modelos, con los datos de prueba:

## Figura 11

### Código Python para la Matriz de Confusión y Cálculo de Métricas

```

from sklearn.metrics import confusion_matrix, accuracy_score, precision_score, recall_score, f1_score, ConfusionMatrixDisplay

# Función para mostrar métricas y matriz de confusión
def mostrar_metricas_y_matriz(y_true, y_pred, nombre_modelo):
    # Calcular métricas
    accuracy = accuracy_score(y_true, y_pred)
    precision = precision_score(y_true, y_pred, average='weighted')
    recall = recall_score(y_true, y_pred, average='weighted')
    f1 = f1_score(y_true, y_pred, average='weighted')

    # Mostrar métricas
    print(f"Métricas para el modelo {nombre_modelo}:")
    print(f"Precisión (Accuracy): {accuracy:.2f}")
    print(f"Precisión (Precision): {precision:.2f}")
    print(f"Sensibilidad (Recall): {recall:.2f}")
    print(f"Puntaje F1 (F1-Score): {f1:.2f}")
    print("\n")

    # Calcular matriz de confusión
    matriz_confusion = confusion_matrix(y_true, y_pred)

    # Graficar matriz de confusión
    plt.figure(figsize=(7, 5))
    disp = ConfusionMatrixDisplay(confusion_matrix=matriz_confusion)
    disp.plot(cmap="Blues", values_format="d")
    plt.title(f"Matriz de Confusión para {nombre_modelo}")
    plt.show()

# Evaluar cada modelo
mostrar_metricas_y_matriz(y_test, y_pred_logreg, "Regresión Logística")
mostrar_metricas_y_matriz(y_test, y_pred_tree, "Árbol de Decisión")
mostrar_metricas_y_matriz(y_test, y_pred_knn, "K-Nearest Neighbors")

```

Las métricas obtenidas y la matriz asociada para cada modelo diseñado son las siguientes:

Métricas para el modelo Regresión Logística:

Precisión (Accuracy): 0.90

Precisión (Precision): 0.90

Sensibilidad (Recall): 0.90

Puntaje F1 (F1-Score): 0.89

Métricas para el modelo Árbol de Decisión:

Precisión (Accuracy): 0.90

Precisión (Precision): 0.90

Sensibilidad (Recall): 0.90

Puntaje F1 (F1-Score): 0.90

Métricas para el modelo K-Nearest Neighbors:

Precisión (Accuracy): 0.89

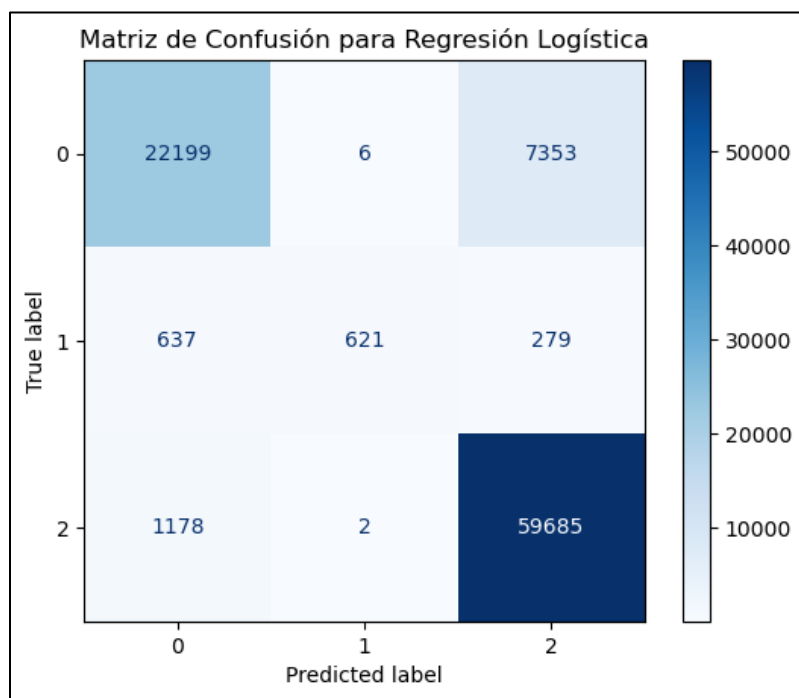
Precisión (Precision): 0.89

Sensibilidad (Recall): 0.89

Puntaje F1 (F1-Score): 0.88

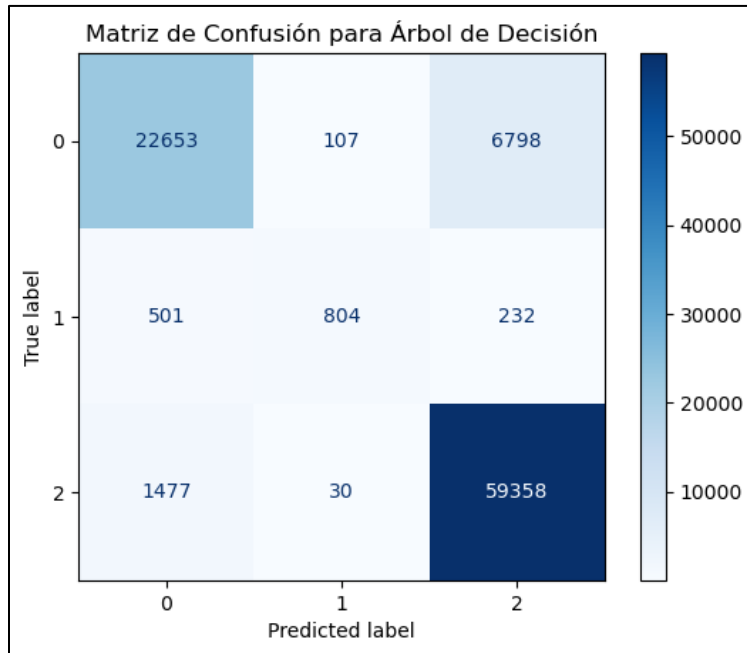
## Figura 12

*Matriz de Confusión para el Modelo Regresión Logística*

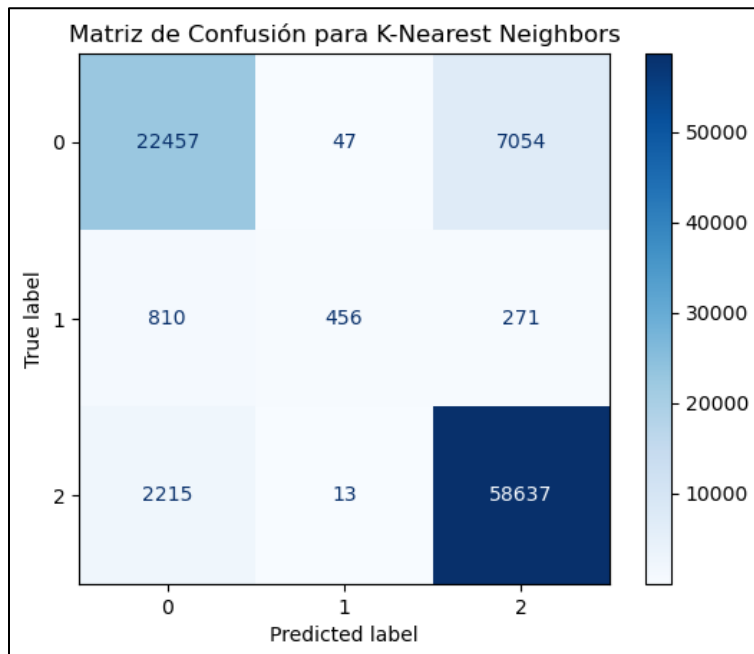


**Figura 13**

*Matriz de Confusión para el Modelo Árbol de Decisión*

**Figura 14**

*Matriz de Confusión para el Modelo KNN*



## Comparación del Desempeño

En esta etapa de actividades se efectúa una comparación para analizar cuál de los tres modelos tiene el mejor desempeño:

### Figura 15

*Código Python para la Comparación de los Modelos Diseñados*

```
# Comparar métricas y seleccionar el modelo con mayor precisión, recall y specificity.
models_metrics = {
    'Logistic Regression': metrics_logreg,
    'Decision Tree': metrics_tree,
    'K-Nearest Neighbors': metrics_knn
}

best_model = max(models_metrics, key=lambda model: models_metrics[model][0])
print("El modelo con mejor desempeño es:", best_model)

El modelo con mejor desempeño es: Decision Tree
```

## **Interpretación del Rendimiento de los Modelos Diseñados y sus Métricas Verificando su Efectividad en la Identificación de Causas de Siniestralidad Vial en Bogotá**

### **Métricas Regresión Logística**

La regresión logística logró una precisión, precisión positiva (precision) y sensibilidad (recall) del 90%, con un F1-Score de 0.89. Estos resultados indican que el modelo de regresión logística es consistente en su desempeño, capturando con éxito tanto los valores positivos como los negativos. Su F1-Score, ligeramente menor que el de las demás métricas, sugiere una ligera variabilidad en la precisión y el recall, aunque esta diferencia es mínima.

La regresión logística demuestra ser un modelo adecuado para predecir la gravedad de los accidentes con una precisión alta y consistente. Sin embargo, el leve descenso en el F1-Score indica que, aunque el modelo es robusto, podría beneficiarse de ajustes adicionales para igualar la precisión y la sensibilidad.

### **Métricas Árbol de Decisión**

El modelo de árbol de decisión obtuvo un rendimiento uniforme en todas las métricas, alcanzando un 90% en precisión, precisión positiva (precision), sensibilidad (recall) y F1-Score. Esto indica que el árbol de decisión logró un buen equilibrio entre la predicción de positivos y negativos sin mostrar preferencia por ninguna clase en particular.

Este modelo resulta efectivo y confiable para clasificar la gravedad de los accidentes, ya que proporciona predicciones balanceadas y coherentes en todas las métricas. Esto lo convierte en una opción sólida para la tarea y sugiere que puede capturar patrones importantes en los datos de manera eficaz.

### **Métricas K-Nearest Neighbors (KNN)**

El modelo KNN alcanzó una precisión, precisión positiva (precisión) y sensibilidad del

89%, y un F1-Score de 0.88. Estos resultados son comparables a los de los otros modelos, aunque son ligeramente inferiores. La baja diferencia en el F1-Score también refleja un buen equilibrio entre precisión y sensibilidad, aunque el rendimiento general es un poco inferior al de los otros dos modelos.

Aunque KNN es razonablemente efectivo, se queda ligeramente por debajo en comparación con la regresión logística y el árbol de decisión. Esto podría deberse a que KNN es más sensible a la distribución de los datos y sufre si no están bien representados en el espacio de características. Aun así, es un modelo competente, aunque podría no ser la opción óptima en comparación con los otros dos modelos en este caso.

De acuerdo con las métricas, tanto la regresión logística como el árbol de decisión presentan un rendimiento muy similar y consistente, con el árbol de decisión teniendo un leve margen en el F1-Score y siendo ideal por su balance.

Con relación al análisis de las matrices de confusión, se puede realizar el siguiente análisis:

### **Matriz de Confusión de Regresión Logística**

Aciertos principales: 22.199 (para clase 0) y 59.685 (para clase 2) indican una buena precisión en la clasificación correcta de estos tipos de siniestros.

Errores: La clase 0 tiene 7.353 instancias clasificadas incorrectamente como clase 2, lo que indica una confusión significativa.

La clase 1 también tiene algunos errores, especialmente hacia la clase 0 (637 casos).

### **Matriz de Confusión de Árbol de Decisión**

Aciertos: Tiene una clasificación correcta alta en la clase 0 (22.653) y clase 2 (59.358), similar a la regresión logística.

Errores: Mejora la precisión en la clase 1, con 804 instancias correctamente clasificadas, y reduce los errores hacia la clase 2 desde la clase 0 (6.798).

En general, el árbol de decisión parece manejar mejor la separación de la clase 1 que la regresión logística.

### **Matriz de Confusión de K-Nearest Neighbors (KNN)**

Aciertos: La cantidad de aciertos para la clase 2 es alta (58.637), pero menor que en los otros dos modelos.

Errores: KNN muestra un mayor número de errores, especialmente en la clase 1 y en la confusión de la clase 0 con la clase 2 (7.054 errores), además de confundir 2.215 casos de clase 2 como clase 0.

### **Comparación y Selección del Mejor Modelo**

Mejor rendimiento en clase 2 (mayor acierto): Los tres modelos presentan una alta precisión para la clase 2, pero Regresión Logística y Árbol de Decisión son más consistentes, con menos errores de clasificación hacia otras clases.

Confusión en clase 1: Árbol de Decisión reduce los errores en la clase 1 comparado con los otros modelos, clasificando correctamente 804 casos. Esto sugiere que maneja mejor la discriminación de la clase 1.

Confusión entre clases 0 y 2: Árbol de Decisión nuevamente parece manejar esta confusión un poco mejor, con 6.798 errores de la clase 0 hacia la clase 2, que es una mejora con respecto a los otros modelos.

Basándonos en estas observaciones, el Árbol de Decisión parece ser el mejor modelo entre los tres. Tiene una buena precisión en las clases 0 y 2 y logra una mejor precisión en la clase 1 que los otros modelos, lo que sugiere que es más equilibrado y efectivo en la separación

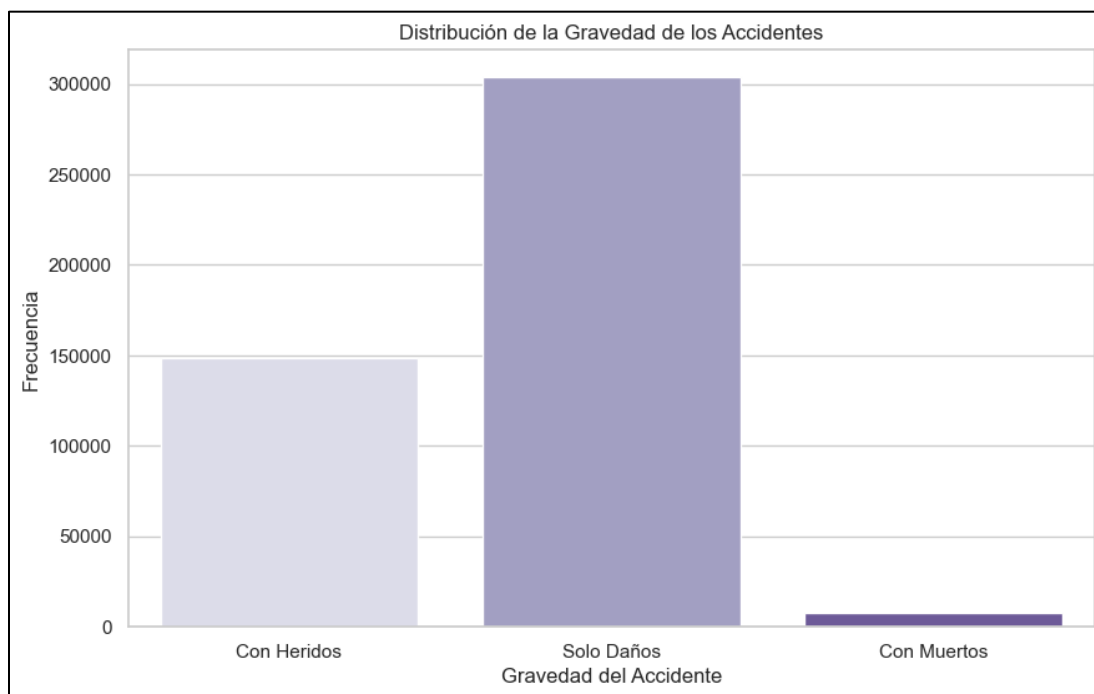
de todas las clases.

### **Análisis Estadístico de las Variables que Podrían Influir en la Decisión de la Causalidad de la Siniestralidad Vial, de Acuerdo a los Datos Obtenidos**

Distribuciones de frecuencia: Para este caso se revisan la distribución de cada variable categórica para observar cómo se distribuyen sus categorías. La frecuencia alta de ciertos valores en cada variable indica cuáles son los tipos de accidentes, diseños de lugares, actores viales y causas que más suelen estar involucrados en los siniestros. Esto es valioso para detectar factores críticos que requieren intervención o prevención en políticas de seguridad vial.

#### **Figura 16**

*Distribución de la Gravedad de los Accidentes*

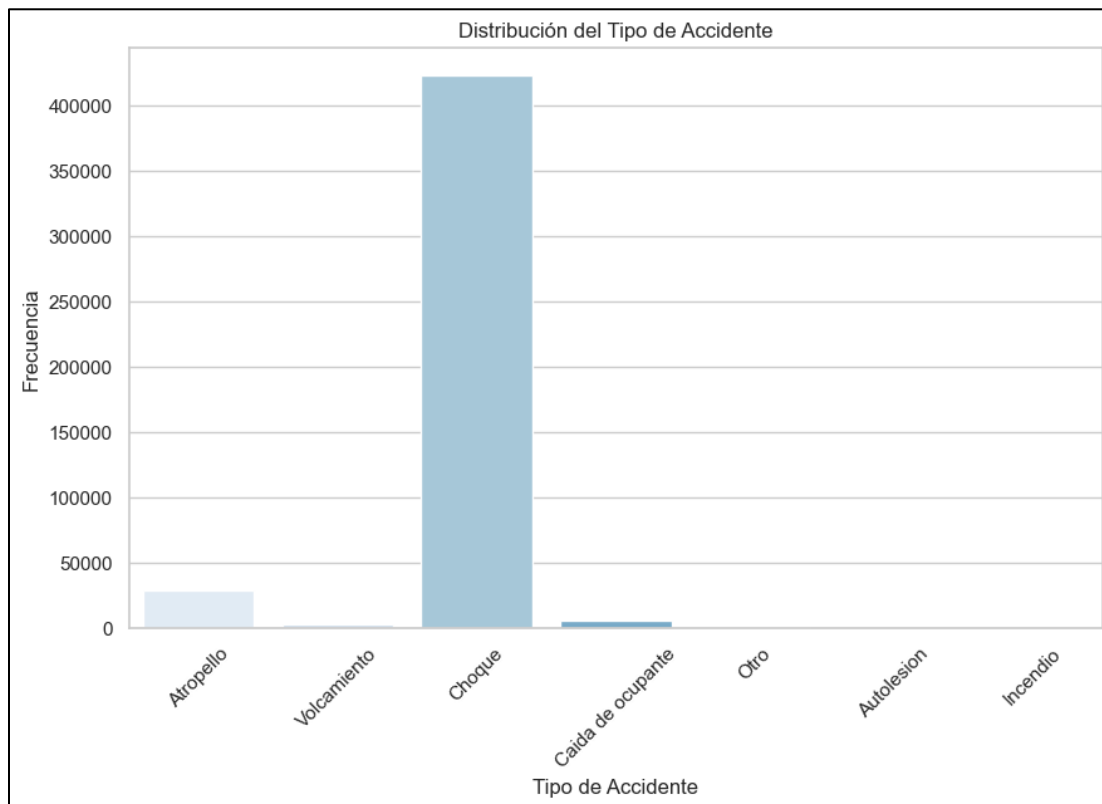


En la figura de la distribución de Gravedad se da a entender cuán graves suelen ser los accidentes en el conjunto de datos. Si se observa que ciertos tipos de accidentes o diseños de lugar tienen una alta frecuencia de gravedad elevada, se podría recomendar revisiones de

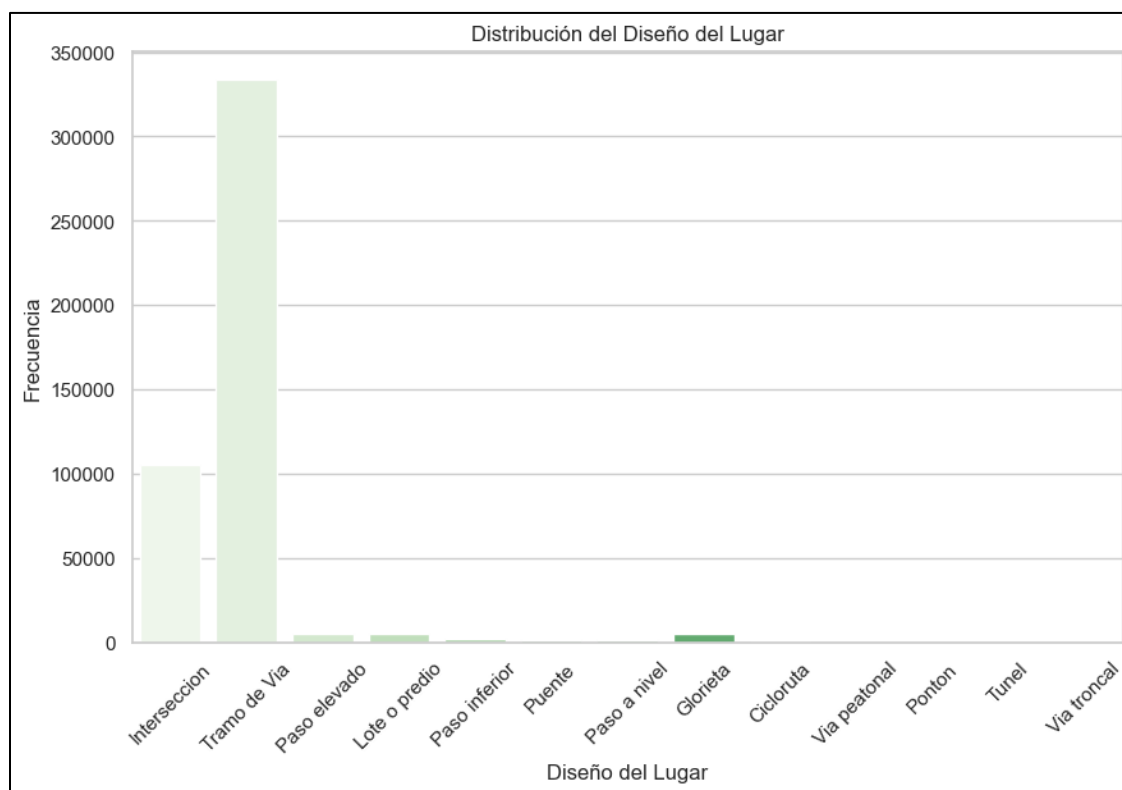
infraestructura en esos lugares específicos. En los accidentes leves que tienen una frecuencia mucho más alta, se podría interpretar que, aunque los accidentes son comunes, no siempre resultan en lesiones graves. Por el contrario, una frecuencia alta de accidentes graves podría indicar problemas serios en infraestructura o en comportamientos viales.

**Figura 17**

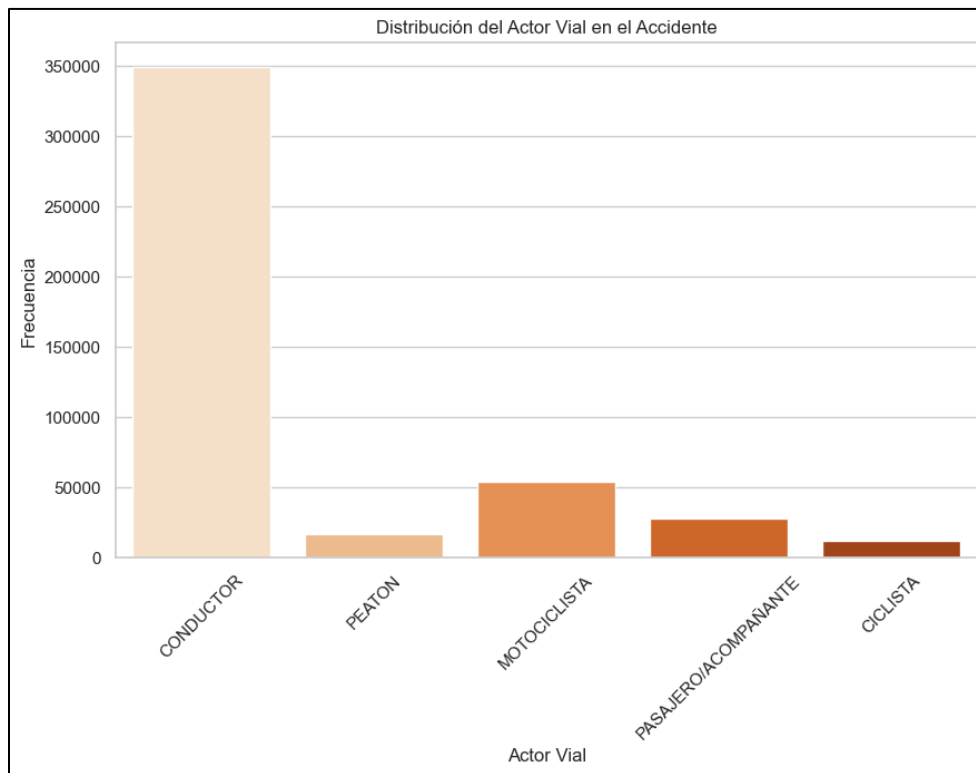
*Distribución del Tipo de Accidente*



En la figura de tipo de accidente, se evidencia que un tipo predominante, como colisiones laterales o choques por alcance, puede señalar problemas específicos de manejo, diseño vial o señalización.

**Figura 18***Distribución del Diseño del Lugar*

En la anterior figura, se evidencia que si los accidentes son más frecuentes en ciertos diseños de lugar (por ejemplo, tramo de vía), estos lugares podrían beneficiarse de intervenciones estructurales, ya que esto podría indicar que en esos lugares podrían carecer de ciertas características que la hagan segura para la movilidad.

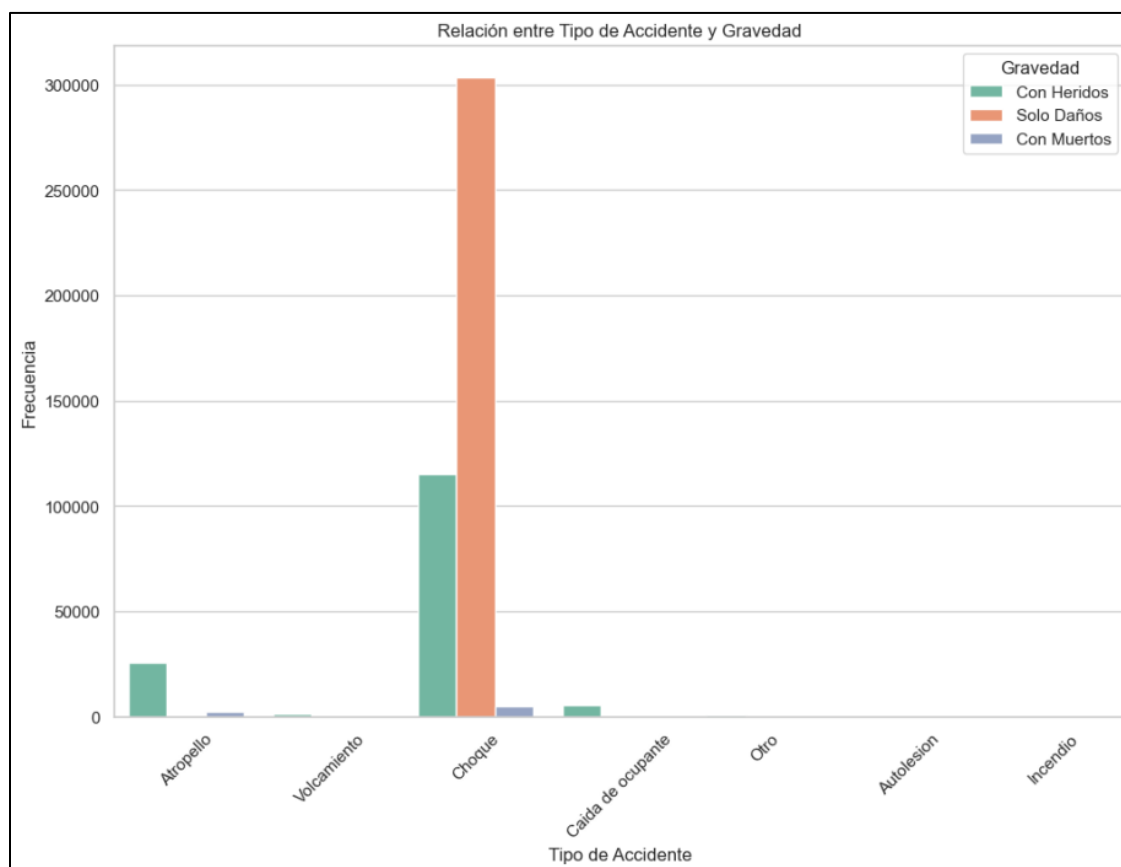
**Figura 19***Distribución del Actor Vial en el Accidente*

Para la figura del actor vial, una alta frecuencia de un cierto tipo de actor vial en accidentes podría resaltar la vulnerabilidad de estos grupos y la necesidad de mejorar la seguridad para ellos.

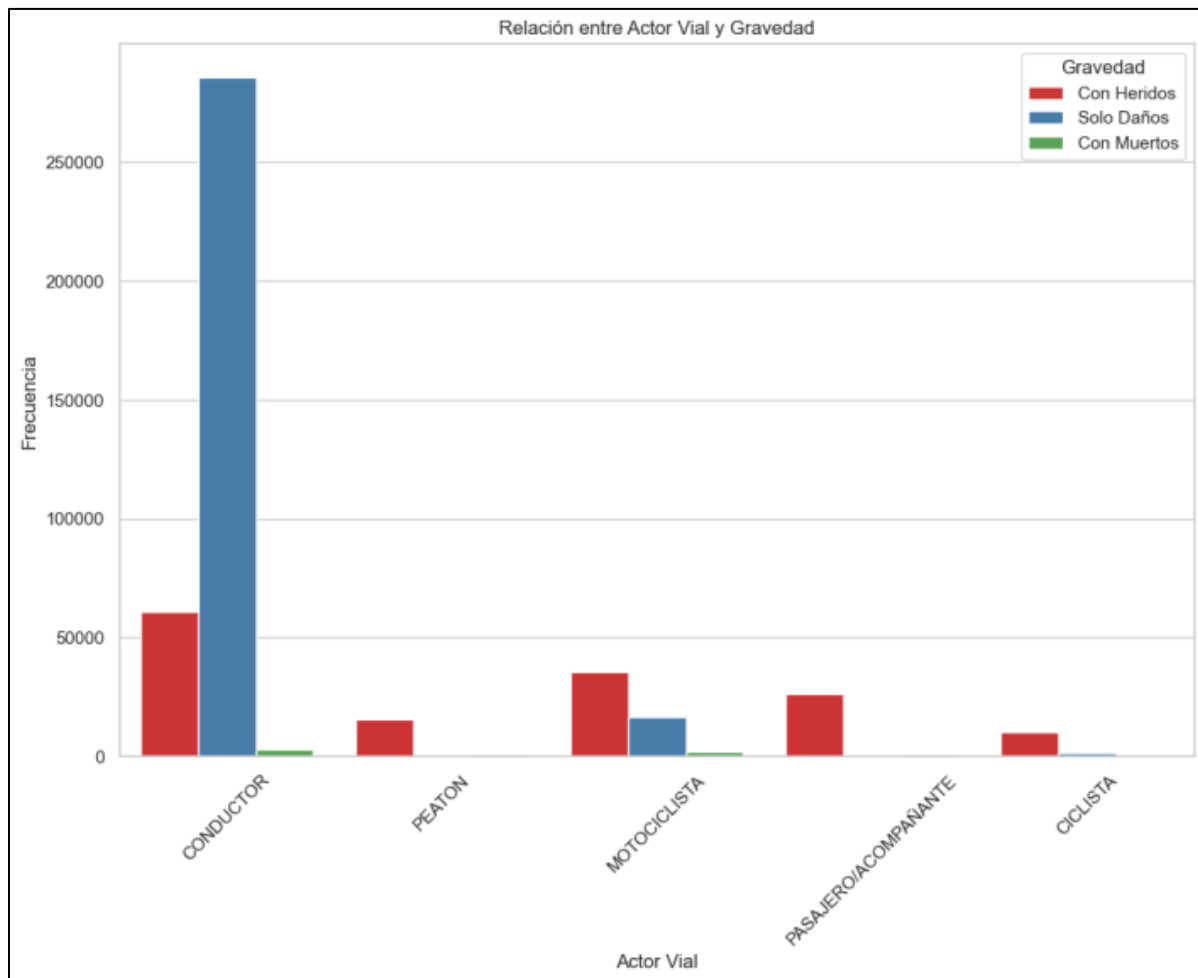
Cruces de tablas y análisis bivariado: Los cruces de tablas permiten observar cómo se relacionan dos variables categóricas entre sí, identificando patrones en función de la combinación de dos factores. Los cruces de tablas pueden revelar qué combinaciones de variables son más frecuentes en los accidentes, ayudando a identificar causas complejas y multidimensionales en la siniestralidad vial. Se realizó una observación de las relaciones entre cada par de variables relevantes, como GRAVEDAD con TIPO\_ACCIDENTE o GRAVEDAD con ACTOR\_VIAL.

**Figura 20**

*Relación entre Tipo de Accidente y Gravedad*



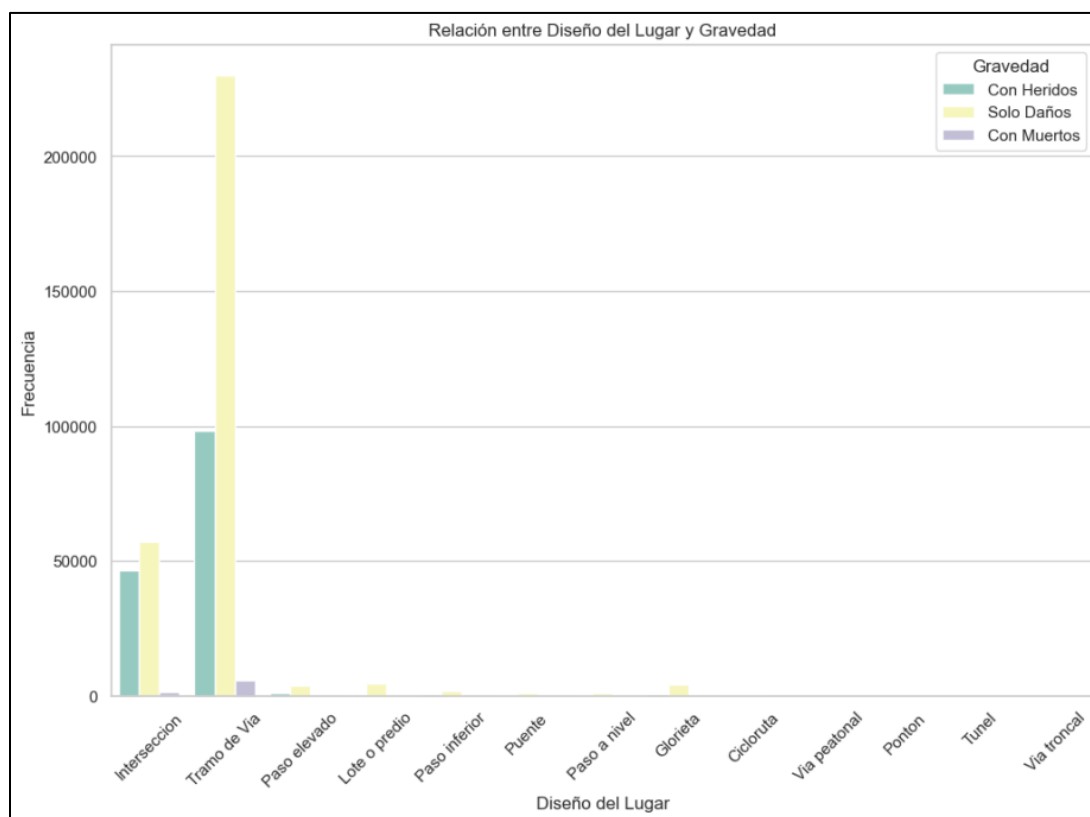
Como se evidencia en la figura anterior, un cruce de estas variables muestra que los accidentes de cierto tipo (como los atropellos) tienden a resultar en mayor gravedad, lo que puede guiar las intervenciones de seguridad. Esta nos muestra si ciertos tipos de accidentes tienden a ser más graves que otros.

**Figura 21***Relación entre Actor Vial y Gravedad*

En el figura actor vial y gravedad, nos ayuda a entender si hay actores (como peatones, ciclistas o conductores) que sufren accidentes más graves que otros. Al relacionar el tipo de accidente con la gravedad, permite identificar si ciertos tipos de accidentes ocurren frecuentemente con más gravedad en actores viales como motociclistas.

**Figura 22**

*Relación entre Diseño del Lugar y Gravedad*



En la anterior figura se puede indicar si ciertos diseños de lugar (por ejemplo, intersecciones, tramos de via) están más asociados a accidentes graves. Estos valores pueden resaltar la gravedad de los accidentes y los lugares específicos donde la intervención es más necesaria, como rediseñar intersecciones o implementar medidas de control de tráfico en avenidas específicas.

En términos generales, para el análisis estadístico anterior en cuanto a las distribuciones de frecuencias y las relaciones entre variables, estas interpretaciones pueden proporcionar una visión integral de los patrones y factores comunes en los accidentes viales en Bogotá. Al enfocarse en los valores más comunes y sus combinaciones, se pueden identificar áreas

prioritarias para intervenciones, tales como mejorar la infraestructura, implementar señalización adecuada, y desarrollar campañas educativas para reducir el riesgo de accidentes en los grupos y lugares identificados como de mayor incidencia. Este enfoque basado en los datos puede guiar decisiones informadas para reducir la siniestralidad vial de forma más efectiva.

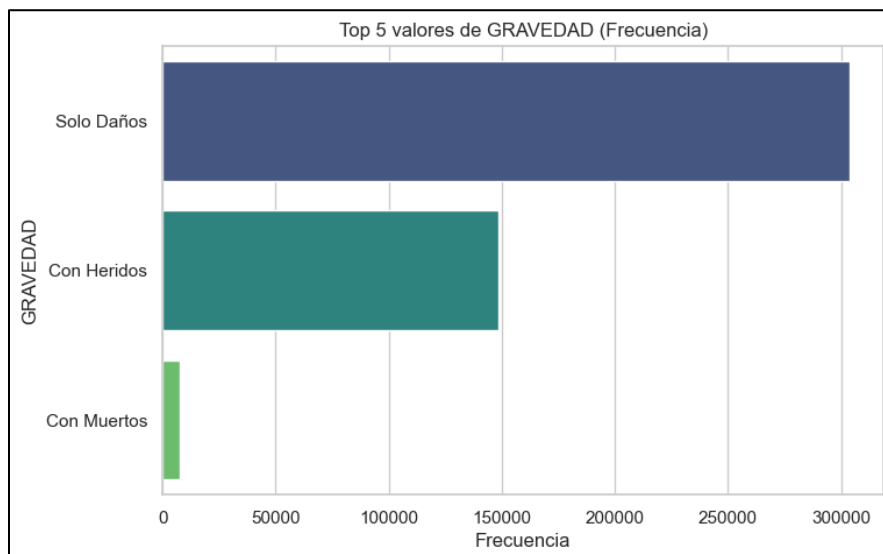
### **Top 5 de las Variables con Mayor Frecuencia de Ocurrencia**

Este análisis ofrece un panorama rápido sobre las características predominantes en la base de datos y puede ser útil para focalizar las causas y tipos de accidentes más comunes, lo que apoya la identificación de patrones en la siniestralidad vial en Bogotá. Los cinco valores más frecuentes en cada variable son útiles para centrarse en los factores más críticos de los accidentes. Pueden orientar políticas o recursos hacia los elementos que representan los mayores riesgos o que se presentan con mayor frecuencia. Cada gráfico de barras muestra los cinco valores más comunes para cada variable:

**Gravedad:** Muestra las categorías con mayor impacto en la salud pública y en los costos asociados con los siniestros.

### **Figura 23**

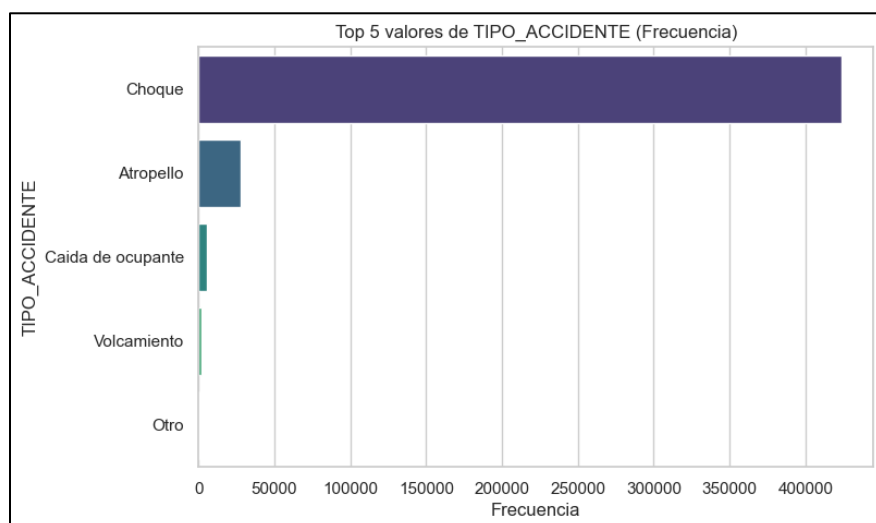
#### *Top 5 Valores de Gravedad*



Accidente y Diseño del Lugar: Esto permitirá ver los tipos de accidente predominantes e indica los diseños de lugar donde ocurren más siniestros. Estos valores pueden resaltar los tipos de accidentes y los lugares específicos donde la intervención es más necesaria, como rediseñar intersecciones o implementar medidas de control de tráfico en avenidas específicas.

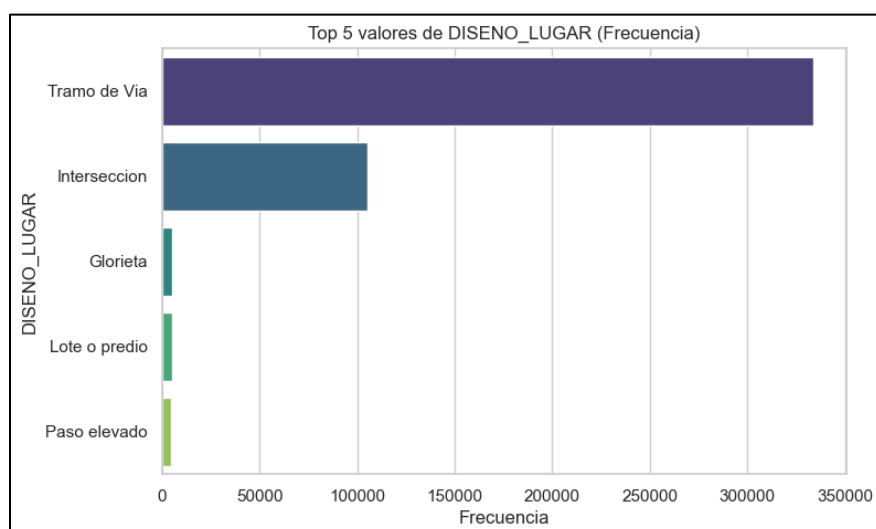
**Figura 24**

*Top 5 Valores de Tipo de Accidente*



**Figura 25**

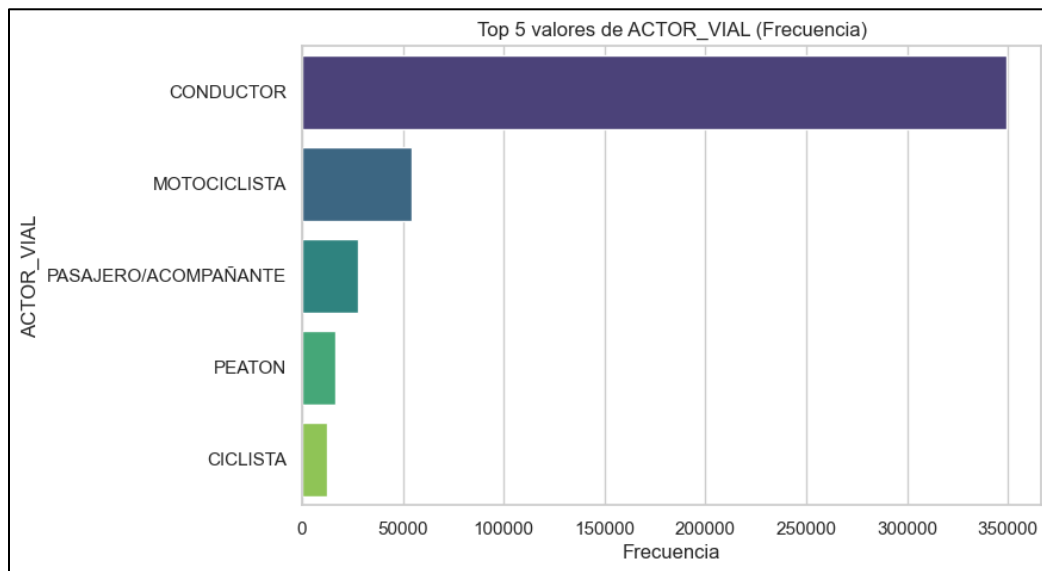
*Top 5 Valores de Diseño de Lugar*



Actor Vial y Código de Causa: Destaca los actores viales involucrados con mayor frecuencia y ayuda a identificar las causas más frecuentes de los accidentes. Identificar los actores viales y las causas más comunes en los accidentes puede guiar campañas de concientización y políticas de manejo para reducir conductas de riesgo.

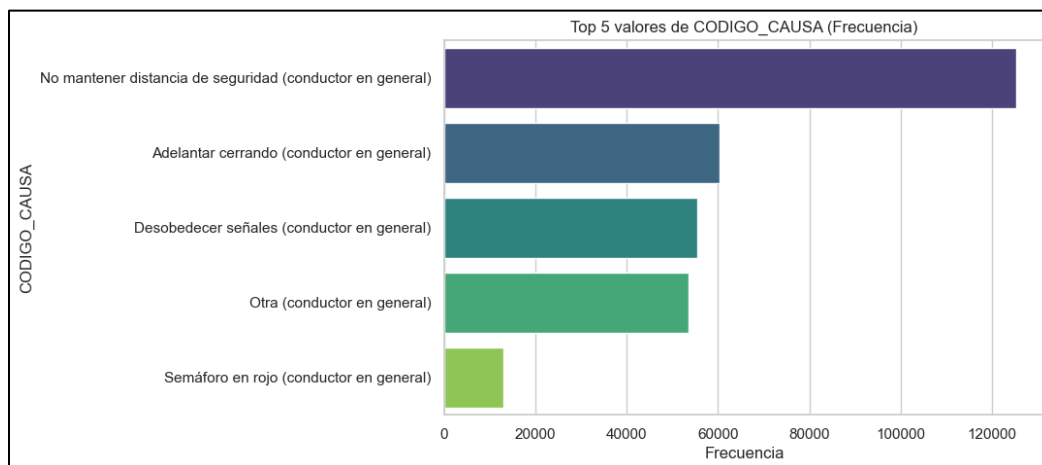
**Figura 26**

*Top 5 Valores de Actor Vial*



**Figura 27**

*Top 5 Valores de Código Causa*



## Conclusiones

Los tres modelos (Regresión Logística, Árbol de Decisión y KNN) mostraron una capacidad adecuada para clasificar los accidentes según su gravedad, aunque con variaciones en precisión y especificidad en cada clase. El modelo de Árbol de Decisión se destacó por su mayor equilibrio entre precisión y sensibilidad, especialmente al clasificar accidentes graves, lo que lo convierte en una herramienta potencialmente útil para identificar puntos críticos de siniestralidad en la ciudad.

Los modelos revelaron que ciertas características, como el diseño del lugar y el tipo de actor vial involucrado (ej. peatón, ciclista, motociclista), tienen un impacto considerable en la probabilidad y gravedad de los accidentes. Esto sugiere que intervenir en estos factores podría reducir significativamente la siniestralidad vial, recomendando mayor atención a la infraestructura de diseño urbano y al control de actores específicos en zonas de alto tráfico.

La variable código\_causa, que identifica las causas de los accidentes, mostró que ciertos comportamientos recurrentes, como el exceso de velocidad o la falta de señalización adecuada, se asocian frecuentemente con accidentes de gravedad alta. Esto subraya la necesidad de realizar campañas de educación vial y mejorar la señalización en áreas específicas, además de una vigilancia más estricta en la aplicación de normas de tránsito.

A pesar de la precisión obtenida, el análisis indicó que los modelos presentan dificultades para clasificar accidentes menos comunes (representados en la clase 1), lo cual podría atribuirse a una menor cantidad de datos representativos para estos eventos. Este hallazgo sugiere que, para mejorar el desempeño, sería útil recolectar y balancear datos adicionales de estas clases menos frecuentes y explorar técnicas avanzadas como el ensamblaje de modelos.

Los modelos pueden ser utilizados para predecir y, por ende, prevenir la siniestralidad en

ciertas ubicaciones y condiciones específicas en Bogotá. Por ejemplo, el Árbol de Decisión podría ayudar a la Secretaría de Movilidad de Bogotá a identificar áreas de alto riesgo en tiempo real, permitiendo aplicar medidas correctivas como controles de velocidad o mejoras en la infraestructura vial en esas zonas críticas.

Los resultados del análisis refuerzan la necesidad de integrar los hallazgos de estos modelos en las políticas de movilidad urbana. Las autoridades podrían emplear estos modelos para hacer un monitoreo proactivo de factores de riesgo, lo cual facilita una intervención temprana y dirigida, contribuyendo a la disminución de los accidentes y a la mejora de la seguridad en vías clave.

La utilización de modelos de aprendizaje automático ha permitido no solo analizar las causas más relevantes de la siniestralidad vial en Bogotá, sino también identificar áreas y factores críticos en los cuales las autoridades pueden enfocar sus esfuerzos de mejora. La incorporación de estos hallazgos en el diseño de políticas públicas podría aumentar la eficiencia de los programas de seguridad vial y reducir la frecuencia y gravedad de los accidentes de tránsito en la ciudad, contribuyendo a una movilidad más segura y eficiente.

## Recomendaciones

A partir de los resultados en las predicciones, métricas de evaluación de modelos y datos estadísticos de la siniestralidad vial, estas recomendaciones están orientadas a reducir los accidentes y mejorar la movilidad en Bogotá:

Implementar medidas preventivas en zonas de alta siniestralidad:

Con base en los datos de diseño de lugar y tipo de accidente, priorizar la implementación de infraestructura preventiva en intersecciones o vías donde se concentran ciertos tipos de accidentes (por ejemplo, colisiones traseras en avenidas de alta velocidad).

Impacto Esperado: Reducir la frecuencia de accidentes al eliminar condiciones viales que fomentan comportamientos de riesgo.

Optimización de señalización en áreas críticas:

La frecuencia de algunos tipos de accidentes y actores viales involucrados sugiere que mejorar la señalización en zonas con peatones y ciclistas podría reducir su involucramiento en siniestros. Colocar más semáforos y señalización visible en áreas donde la gravedad de los accidentes es alta, ayudaría a reducir la mortalidad.

Impacto Esperado: Aumentar la seguridad de los actores viales más vulnerables y reducir accidentes graves en zonas de tránsito peatonal.

Refuerzo de la vigilancia en comportamientos de riesgo:

Según el análisis de código de causa, es crucial reforzar la vigilancia en torno a comportamientos como exceso de velocidad y desobediencia de señales. Se sugiere implementar tecnologías de detección de velocidad y cámaras en puntos estratégicos para desincentivar el exceso de velocidad y la falta de atención.

Impacto Esperado: Reducción de accidentes graves relacionados con exceso de velocidad

y comportamientos irresponsables.

Educación vial focalizada:

Al observar los cruces de tablas entre actor vial y causa del accidente, se recomienda desarrollar campañas educativas focalizadas, especialmente para conductores de motocicletas y peatones, quienes suelen ser actores de mayor riesgo. Estas campañas pueden enfocarse en fomentar comportamientos seguros y la importancia de la atención en la vía.

Impacto Esperado: Mayor conciencia de los actores viales más vulnerables, ayudando a reducir la siniestralidad.

Evaluación y rediseño de vías para mejorar la movilidad y la seguridad:

Rediseñar las vías con alto flujo vehicular o donde ocurren frecuentemente ciertos tipos de accidentes. Esto podría incluir la creación de pasos elevados o rediseños de cruces para vehículos y peatones.

Impacto Esperado: Mejora en la movilidad al reducir los puntos de conflicto y aumentar la seguridad para todos los usuarios de la vía.

Análisis continuo y modelos predictivos en tiempo real:

Fortalecer el sistema de monitoreo de siniestralidad mediante el uso de modelos predictivos en tiempo real, que pueden identificar patrones emergentes y ayudar a la toma de decisiones rápida para evitar la acumulación de siniestros en momentos o zonas específicas.

Impacto Esperado: Mayor adaptabilidad en la gestión del tráfico y la reducción de siniestros a partir de datos en tiempo real.

Promoción de la cultura de movilidad compartida:

Incentivar el uso de transporte compartido y público, reduciendo así el número de vehículos en las vías y, en consecuencia, el riesgo de colisiones. Esto puede incluir incentivos

fiscales o descuentos en servicios de movilidad compartida.

Impacto Esperado: Disminución del volumen de tráfico, mayor fluidez vehicular y reducción de accidentes por alta densidad de vehículos.

Estas estrategias, fundamentadas en los hallazgos del análisis de datos, pueden ser eficaces para reducir la siniestralidad vial y mejorar el flujo de tráfico en Bogotá. La integración de medidas preventivas, tecnología, educación y rediseño de infraestructura puede crear un sistema de movilidad más seguro y eficiente.

## Referencias

- Alarcón, J.D., Gich-Saladich, I., Vallejo-Cuellar, L., Ríos-Gallardo, A.M., Montalvo-Arce, C. y Bonfill-Cosp, X. (2018) Mortalidad por accidentes de tráfico en Colombia. Estudio comparativo con otros países. *Revista Española de Salud Pública*, 92 (5).  
[http://scielo.isciii.es/scielo.php?script=sci\\_arttext&pid=S1135-57272018000100415&lng=es&tlng=es](http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1135-57272018000100415&lng=es&tlng=es)
- Amazon Web Services - AWS (s.f.). *¿Qué es la regresión lineal?*. Recuperado el 20 de mayo 2024. <https://aws.amazon.com/es/what-is/linear-regression/>.
- Amazon Web Services - AWS (s.f.). *¿Qué es la regresión logística?*. Recuperado el 20 de mayo 2024. <https://aws.amazon.com/es/what-is/logistic-regression/>.
- BBVA (8 de noviembre 2019). *Machine learning: qué es y cómo funciona*. BBVA.  
<https://www.bbva.com/es/innovacion/machine-learning-que-es-y-como-funciona/>
- Cárdenas-Lancheros, E.A. (2021) *Modelo de pronóstico de incidentes en conducción de motocicletas basado en iot e inteligencia artificial*. [tesis de maestría, Universidad Distrital Francisco José De Caldas]. Repositorio Institucional Universidad Distrital Francisco José De Caldas.  
<https://repository.udistrital.edu.co/bitstream/handle/11349/30384/C%C3%A1rdenasLancherosEstebanAlejandro2021.pdf?sequence=3&isAllowed=y>
- Cardona-Álvarez, J.C. (2023). *Modelo predictivo de zonas de riesgo espacio – temporal de accidentes de tráfico en la ciudad de Manizales*. [tesis de maestría, Universidad de Caldas]. Repositorio Institucional Universidad de Caldas.  
<https://repositorio.ucaldas.edu.co/bitstream/handle/ucaldas/19537/MODELO%20PREDICTIVO%20DE%20ZONAS%20DE%20RIESGO%20ESPACIO%20TEMPORAL%20>

E%20ACCIDENTES%20DE%20TR%3%81FICO%20EN%20LA%20CIUDAD%20D  
E%20MANIZALES.pdf?sequence=1&isAllowed=y

Cruz Beltrán, J.S., y Herrera Martín, L.J. (2023) *Modelo de analítica de datos para definición de estrategias de circulación segura de los vehículos de carga en el corredor vial de la calle 13 de la ciudad de Bogotá. D.C.* [tesis de maestría, Universidad Libre]. Repositorio Institucional Unilibre.

<https://repository.unilibre.edu.co/bitstream/handle/10901/25704/Modelo%20de%20Anal%3%ADtica%20de%20Datos%20para%20Definici%3%B3n%20de%20Estrategias%20de%20Circulaci%3%B3n%20Segura%20de%20los%20Veh%3%ADculos%20de%20Carga%20en%20el%20Corredor%20Vial%20de%20la%20Calle%2013%20de%20la%20Ciudad%20de%20Bogot%3%A1%20DC.pdf?sequence=1&isAllowed=y>

Edx (s.f). *¿Qué es un modelo predictivo?.* Recuperado el 8 de mayo 2024.

<https://www.edx.org/es/aprende/modelo-predictivo>.

Ferro-Briceño, P.V. (2021) *Uso de redes neuronales para determinar la influencia del estado del pavimento en siniestros viales de la ciudad de Bogotá.* [tesis de maestría, Universidad de los Andes]. Repositorio Institucional Uniandes.

<https://repositorio.uniandes.edu.co/server/api/core/bitstreams/1e7ef712-dfaa-42d0-bf90-fe652361aa33/content>

Gómez Montoya, J.E., y Cuartas-Arias, M. (2020). Accidentalidad vial: efectos de la calidad del sueño en el funcionamiento ejecutivo de conductores de transporte público urbano.

*Revista Investigacion e Innovacion en Ciencias de la Salud*, 2(1, )41-55.

<https://doi.org/10.46634/riics.43>

Henao Pereira, J.P., Tovar León, A.E., Castillo Landinez, S.P., y Caicedo Rodriguez, P.E.

- (2020). Los accidentes de tránsito desde la perspectiva de la minería de datos. Una revisión de la literatura. *Aibi revista de investigación, administración e ingeniería*. 8(2), 133-141. <https://doi.org/10.15649/2346030X.743>
- Herrera-Briones, J. (2021) *Análisis y predicción de la lesividad en accidentes de tráfico mediante la aplicación de random forest*. [tesis de grado, Universidad Politecnica de Madrid]. Repositorio Institucional Universidad Politecnica de Madrid. [https://oa.upm.es/67548/1/TFG\\_JUAN\\_HERRERA\\_BRIONES.pdf](https://oa.upm.es/67548/1/TFG_JUAN_HERRERA_BRIONES.pdf)
- Largo, J.M. (15 de Febrero de 2024). *Bogotá fue la ciudad que registró más siniestros viales en 2023: este es el panorama en otras capitales*. INFOBAE. <https://www.infobae.com/colombia/2024/02/15/bogota-fue-la-ciudad-que-registro-mas-siniestros-viales-en-2023-este-es-el-panorama-en-otras-capitales/>
- Martinez, J. (16 de septiembre de 2023). *Clustering (Agrupamiento), K-Means con ejemplos en Python*. IARTIFICIAL. <https://www.iartificial.net/clustering-agrupamiento-kmeans-ejemplos-en-python/>.
- Martínez-Ramos, M.M., Vásquez-López, L.J., y González-Veloz, J.J. (2021) *Análisis de Siniestralidad Peatonal a partir de la metodología de Machine Learning*. Repositorio Institucional Universidad Libertadores. <https://repository.libertadores.edu.co/server/api/core/bitstreams/81fbc9f3-d847-415f-ba55-c829649d791c/content>
- Observatorio de Movilidad de Bogotá (s.f.). *Definición de Observatorio de Movilidad*. Recuperado el 22 de abril 2024. <https://observatorio.movilidadbogota.gov.co/#:~:text=El%20Observatorio%20de%20Movilidad%20de,de%20inter%C3%A9s%20y%20partes%20interesadas.>

- Oregón-Fuentes, A.B., y Castellom-Rodríguez, J.D. (2021) Modelo de análisis correlacional para la identificación de patrones asociados a los accidentes de tránsito de Yopal. *Encuentro Internacional De Educación En Ingeniería*. <https://doi.org/10.26507/ponencia.1787>
- Redondo-Morelo, J.F. (2023) *Modelo predictivo de accidentes de tránsito*. [tesis de grado, Universidad de Antioquia]. Repositorio Institucional UDEA. [https://bibliotecadigital.udea.edu.co/bitstream/10495/34367/1/RedondoJohn\\_2023\\_ModeloPredictivoAccidentes.docx.pdf](https://bibliotecadigital.udea.edu.co/bitstream/10495/34367/1/RedondoJohn_2023_ModeloPredictivoAccidentes.docx.pdf)
- TIBCO, (s.f). *¿Qué es la analítica de datos?*. Recuperado el 22 de abril 2024. [https://docs.tibco.com/pub/sfire-cloud/14.3.0/doc/html/es-ES/TIB\\_sfire\\_client/client/topics/es-ES/analyze\\_data.html](https://docs.tibco.com/pub/sfire-cloud/14.3.0/doc/html/es-ES/TIB_sfire_client/client/topics/es-ES/analyze_data.html)
- Walther (26 de abril 2023). *¿Qué son los modelos de inteligencia artificial y cuáles son los más usados?*. *DONGEE*. <https://www.dongee.com/tutoriales/que-son-los-modelos-de-inteligencia-artificial-y-cuales-son-los-mas-usados/>.