

Proyecto de Integración y Automatización de Datos para el seguimiento y monitoreo de siniestros viales en la ciudad de Bogotá D.C

Javier Antonio Gonzalez¹,

Universidad Central

Maestría en Analítica de Datos

Curso de Automatización e Integración de Datos para Inteligencia Artificial

Bogotá, Colombia

`jgonzalezg27@ucentral.edu.co`

27 de noviembre de 2023

Índice general

Introducción (Max 250 Palabras) - (<i>Primera entrega</i>)	3
Características del proyecto de investigación que hace uso de Integración y Automatiza- ción de Datos para IA (Max 500 Palabras) - (<i>Primera entrega</i>)	4
Titulo del proyecto de investigación	4
Objetivo general (Max 100 Palabras) - (<i>Primera entrega</i>)	4
Alcance (Max 200 Palabras) - (<i>Primera entrega</i>)	4
Pregunta de investigación (Max 100 Palabras) - (<i>Primera entrega</i>)	6
Hipotesis (Max 100 Palabras) - (<i>Primera entrega</i>)	6
Reflexiones sobre el origen de datos e información (Max 400 Palabras) - (<i>Primera entrega</i>) .	7
¿Cual es el origen de los datos e información ? (Max 100 Palabras) - (<i>Primera entrega</i>)	8
¿Cuales son las consideraciones legales o éticas del uso de la información? (Max 100 Palabras) - (<i>Primera entrega</i>)	8
¿Cuales son los retos de la información y los datos que utilizara en Integración y Au- tomatización de Datos para IA? (Max 100 Palabras) - (<i>Primera entrega</i>)	9
¿Que espera de la utilización de Integración y Automatización de Datos para IA pa- ra su proyecto? (Max 100 Palabras) - (<i>Primera entrega</i>)	9
Diseño de integración y Automatización de Datos para IA (Diagrama) (<i>Primera entrega</i>) .	10
Integración de Datos - (<i>segunda entrega</i>)	11
Recopilación de Datos Iniciales	11
Descripción de los datos	13
Verificación de la calidad de datos	18
Preparación de los Datos	19
Selección de Datos	19

Selección de Datos	19
Construcción de nuevos datos	20
Integración de datos	26
Análisis univariado	26
Análisis bivariado	27
IA - (<i>segunda entrega</i>)	28
Proximos Pasos - (<i>Tercera entrega</i>)	29
Propuesta de mejora al observatorio de movilidad para seguimiento de siniestros viales	31
Lecciones Aprendidas - (<i>Tercera entrega</i>)	34

Introducción (Max 250 Palabras) - (Primera entrega)

Uno de los principales efectos de la accidentalidad vial es su alto costo económico. Según datos de la Federación de Aseguradores Colombianos en 2016 (Federación de Aseguradores Colombianos [FASECOLDA], [2018](#)), esto ascendió a 2,3 billones de pesos, equivalente al 0,3 % del Producto Interno Bruto (PIB) de Colombia. La mayoría de este costo se destinó a la asistencia médica y pensiones para las víctimas.

Este problema se considera un asunto de salud pública, y las Naciones Unidas han instado a los gobiernos de todo el mundo a establecer metas para reducir los siniestros viales. Esto se reflejó en la Década de Acción para la Seguridad Vial 2021-2030, que busca reducir las muertes y lesiones por tráfico en un 50

Colombia también se ha comprometido con estas metas a través del Plan Nacional de Seguridad Vial 2022-2031. En Bogotá, se busca reducir las víctimas fatales y heridos graves en un 35 % para 2026.

Para monitorear el progreso, Bogotá cuenta con el Observatorio de Movilidad, que proporciona datos actualizados sobre siniestros viales. Además, organizaciones no gubernamentales como Bloomberg Philanthropies trabajan en conjunto con las autoridades para mejorar la seguridad vial.

La academia también desempeña un papel importante al contribuir con investigaciones y soluciones. Este estudio se enfoca en abordar la problemática de la accidentalidad vial en Bogotá y construir una herramienta para realizar el monitoreo , identificando los factores que inciden en la ocurrencia de los siniestros viales.

Características del proyecto de investigación que hace uso de Integración y Automatización de Datos para IA (Max 500 Palabras) - (Primera entrega)

Título del proyecto de investigación

Proyecto de Integración y Automatización de Datos para el seguimiento y monitoreo de siniestros viales en la ciudad de Bogotá D.C.

Objetivo general (Max 100 Palabras) - (Primera entrega)

Construir un tablero de control para monitorear las características de siniestros viales, que permita recomendar acciones para su seguimiento y prevención.

Objetivos específicos (Max 100 Palabras) - (Primera entrega)

- Realizar una integración de los datos recopilados por la Secretaría Distrital de movilidad que permita realizar análisis descriptivo.
- Determinar las características de los siniestros viales reportados por la Secretaría de Movilidad de la ciudad de Bogotá D.C. en el periodo 2018 a 2021.
- Proponer un modelo de machine learning, que sirva para determinar las causas principales asociadas al grado de severidad de los siniestros viales en Bogotá D.C.

Alcance (Max 200 Palabras) - (Primera entrega)

En las últimas décadas, la industria tecnológica ha experimentado un rápido avance en el desarrollo de capacidades computacionales para el procesamiento y almacenamiento de grandes volúmenes de datos. Este progreso tecnológico ha allanado el camino para la aplicación de técnicas estadísticas y de modelado de datos que son computacionalmente exigentes.

La revisión bibliográfica revela que investigadores han abordado la problemática de la accidentalidad vial desde diversos enfoques y han hecho recomendaciones significativas.

Tianyu et al. (2022) aplicaron modelos de series de tiempo y redes neuronales recurrentes para predecir la cantidad de lesionados en siniestros viales en el Noreste de China. Recomendaron el uso de redes neuronales recurrentes Tianyu et al. (2022).

Islam et al. (2022) utilizaron modelos Random Forest y Regresión Logística para predecir la severidad de siniestros viales en Al-Ahsa, Arabia Saudita, y también identificaron clusters espaciales de alta accidentalidad. Recomendaron la inclusión de variables adicionales en futuros estudios Islam et al. (2022).

Kumar y Melempat Kalapurayil (2022) exploraron modelos de clasificación multietiqueta para abordar la severidad y el tipo de colisión en siniestros viales en Texas, EE. UU. Concluyeron que estos modelos superan a los enfoques tradicionales.

Lia et al. (2022) adoptaron un enfoque innovador al combinar el análisis de redes sociales con modelos de ML para identificar y clasificar incidentes y siniestros viales en Panamá. El modelo Random Forest resultó ser el más efectivo Lia et al. (2022).

Cantillo et al. (2016) investigaron tramos críticos de accidentalidad en Cartagena, Colombia, utilizando técnicas como el Análisis de Componentes Principales y la Regresión Poisson. Identificaron relaciones empíricas entre la frecuencia de accidentes y variables como la densidad del tráfico y la geometría de las vías Cantillo et al. (2016).

Estos estudios resaltan la importancia de utilizar técnicas avanzadas de procesamiento de datos y la necesidad de considerar diversas variables para comprender y predecir la accidentalidad vial de manera más efectiva. Además, enfatizan la importancia de la integridad de los datos y la necesidad de explorar enfoques innovadores para abordar este desafío.

Pregunta de investigación (Max 100 Palabras) - (Primera entrega)

¿Cómo podemos detectar patrones que me permitan comprender y predecir la ocurrencia de siniestros viales y los factores que influyen en esta?

Hipotesis (Max 100 Palabras) - (Primera entrega)

El presente estudio tiene como objetivo abordar la problemática de la siniestralidad vial en Bogotá, analizando las características subyacentes a la gravedad de los accidentes. También busca desarrollar una herramienta actualizada y acorde a las dinámicas recientes de los siniestros viales, que contribuya a mitigar sus consecuencias, beneficiando a los diversos actores viales en la ciudad.

Se pretende responder a la pregunta de cuáles son los factores principales que influyen en la gravedad de los siniestros viales en Bogotá, utilizando los eventos documentados en el período de 2018 a 2022.

Es importante resaltar que con los resultados del presente estudio se identificaran los factores influyentes en la siniestralidad vial en Bogotá, lo cual permitirá formular, desde la academia, los cambios en el instrumento de seguimiento a los siniestros viales con que actualmente cuenta las autoridades distritales. Lo cual buscar redundar en la reducción de los siniestros viales y su gravedad.

Reflexiones sobre el origen de datos e información (Max 400 Palabras) - (Primera entrega)

Con la información capturada en el sitio de ocurrencia del siniestro, la autoridad de tránsito debe reportar la información al Registro Nacional de Accidentes de Tránsito (RNAT). Para el caso de la ciudad de Bogotá, la Secretaría de Movilidad de Distrital ha puesto a disposición de los ciudadanos estos datos desde el año 2015 a través del portal de datos abierto SIMUR.

Los datos publicados realizan una síntesis de las características inherentes a la ocurrencia del siniestro vial. Dentro de estos datos se encuentran: características demográficas de los actores viales involucrados, hipótesis o causas del siniestro vial, características de los vehículos vinculados y características de las vías donde ocurrió el siniestro.

Los datos obtenidos corresponden al corte del 3 de julio de 2023 y se dividen en 5 grupos. En la Tabla 0.2 se muestra el detalle.

Tabla 0.1: Datos de siniestralidad publicados por Secretaría Distrital de Movilidad - SIMUR -

Tabla	Columnas	Registros
Accidente	19	464699
Actor Vial	13	1657681
Causa	10	672815
Vehículo	9	1427204
Vía	19	483952

Datos geográficos La información recopilada en el formulario IPAT y puesta a disposición por la SDM, presenta una falencia en datos del entorno próximo al sitio de ocurrencia del siniestro. Por lo tanto, a partir de la información disponible en la Infraestructura de Datos Espaciales de Bogotá D.C. y su procesamiento por medio de Sistemas de Información Geográfica (SIG), se incluirán a la base de datos variables exógenas de los siniestros viales, con lo cual se espera a mejorar los resultados del presente estudio.

Las características a incluir son las siguientes:

- Cantidad de Comparendos impuestos entre 2015 - 2021 en la zona del siniestro.

- Si se cuenta con red de cicloinfraestructura a menos de 200 mts del siniestro.
- El estrato predominante en el sector del siniestro.
- Tipo de categoría de la vía donde ocurre el siniestro.
- El uso del suelo predominante en el sector donde ocurre el siniestro.

¿Cual es el origen de los datos e información ? (Max 100 Palabras) - (Primera entrega)

El Ministerio de transporte de Colombia por medio de la Resolución No. 11268 de 2012 adoptó el Informe Policial de Accidentes de Tránsito (IPAT), el cual es de obligatorio cumplimiento para la autoridad de tránsito.

El formato adoptado, asigna un número único para cada accidente de tránsito, recolecta información como ubicación, características del lugar, características de las vías y la identificación de los vehículos y sus conductores involucrados.

Esta información es recopilada y puesta a disposición por la Secretaría Distrital de Movilidad a través del portal de datos abierto SIMUR.

¿Cuales son las consideraciones legales o éticas del uso de la información? (Max 100 Palabras) - (Primera entrega)

”La Resolución No. 11268 de 2012 estipula que *‘Las autoridades de tránsito velarán por la seguridad de las personas y las cosas en la vía pública y privadas abiertas al público. Sus funciones serán de carácter regulatorio y sancionatorio y sus acciones deben ser orientadas a la prevención y la asistencia técnica y humana a los usuarios de las vías’* (Resolución No. 11268 de 2012).

Esto da lugar a la creación del Registro Único Nacional de Tránsito (RUNT) y el Registro Nacional de Accidentes de Tránsito (RNAT).

El marco técnico de dicha resolución dictamina *'Que con el objetivo que el RNAT sea la herramienta que permita identificar claramente las hipótesis de las causas de accidentalidad, es necesario adoptar un nuevo Informe Policial de Accidentes de Tránsito, incluyendo datos que permitan un mejor análisis para la elaboración de diagnósticos y la planeación en materia de seguridad vial'* (Resolución No. 11268 de 2012).

Así pues, al ser información que permite prevenir la frecuencia en siniestros viales, y generar hipótesis, su uso es público, y por tanto su análisis no constituye una razón para inducir a un uso irregular, que pudiera violentar la integridad de cualquier entidad pública o privada”(Resolución No. 11268 de 2012).

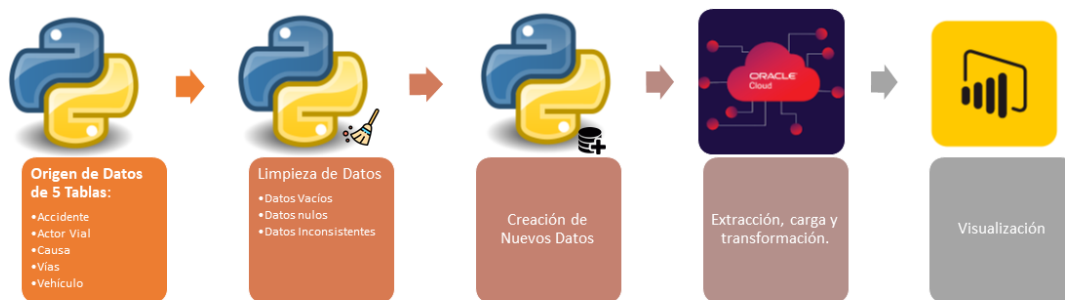
¿Cuales son los retos de la información y los datos que utilizara en Integración y Automatización de Datos para IA? (Max 100 Palabras) - (Primera entrega)

Es necesario señalar que durante la ejecución del proyecto se deben sortear tanto las dificultades que trae consigo la temática, como el reto que impone el tratamiento de los datos de diversas fuentes. De igual manera, se espera encontrar dificultades en la conciliación de las cifras entre los siniestros viales documentados y las cifras oficiales que se encuentran en los observatorios, anuarios y cifras del sector asegurador (Varela, 2018).

¿Que espera de la utilización de Integración y Automatización de Datos para IA para su proyecto? (Max 100 Palabras) - (Primera entrega)

Diseño de integración y Automatización de Datos para IA (Diagrama) (Primera entrega)

Figura 1: Procesamiento datos siniestros viales



Proceso de Automatización de Datos con Python, Oracle Cloud y Power BI

- **Origen:** El proceso comienza con la extracción de datos desde cinco tablas diferentes que actúan como origen de datos (Accidente, Actor Vial, Causa, Vehículo, Vía). Estas tablas pueden estar en una variedad de formatos, como bases de datos, hojas de cálculo o archivos CSV.
- **Limpieza de Datos con Python:** Los datos extraídos pasan por una fase de limpieza y pre-procesamiento utilizando scripts en Python. En esta etapa, se pueden realizar tareas como la eliminación de datos duplicados, la imputación de valores faltantes, la estandarización de formatos y la eliminación de valores atípicos utilizando bibliotecas como Pandas y NumPy.
- **Creación de Nuevos Datos con Python:** Después de la limpieza de datos, se pueden generar nuevos datos o características (features) utilizando técnicas de análisis y modelado dispo-

nibles en Python. Esto podría incluir la creación de variables derivadas, la segmentación de datos o la generación de características adicionales para el análisis.

- Aplicación de un modelo de Machine Learning para variables categóricas modelado de la gravedad de siniestros viales. Entre estos modelos se destacan Random Forest, SVM y Árboles de Decisión, estas técnicas serán las utilizadas para lograr el cumplimiento de los objetivos del presente estudio.
- ETL con Oracle Cloud: Los datos preparados y enriquecidos se transfieren a Oracle Cloud para el proceso de Extracción, Transformación y Carga (ETL). En esta etapa, los datos pueden ser almacenados en una base de datos en la nube, transformados según sea necesario y preparados para su análisis posterior.
- Visualización con la herramienta de visualización Power BI: Una vez que los datos se han cargado en Oracle Cloud y están listos para su análisis, se conectan a Power BI, una herramienta de visualización que permite crear paneles interactivos, informes y gráficos que permiten a los usuarios explorar y comprender los datos de manera efectiva.

Este proceso de automatización permite tomar datos crudos de múltiples fuentes, limpiarlos, enriquecerlos, almacenarlos en una plataforma segura y presentarlos de manera visualmente atractiva para tomar decisiones informadas. La automatización de este flujo de trabajo puede aumentar la eficiencia y la calidad del análisis de datos.

Integración de Datos - (*segunda entrega*)

Recopilación de Datos Iniciales

Es importante comprender con claridad de donde se obtienen los datos, realizar una descripción de los mismos y validar o verificar su calidad. En esta parte del proyecto, se informan cuáles fueron las fuentes donde se obtuvo la información a analizar.

Datos siniestralidad vial: Debido al manual con el cual cuenta el Ministerio de transporte de Colombia, IPAT (Informe Policial de Accidentes de Tránsito) , y en donde se puede encontrar un numero único para cada accidente de tránsito; información de su ubicación, características del lugar, características de las vías e identificación de los vehiculos y sus conductores involucrados (Ministerio de Transporte de Colombia, 2012), la autoridad de transito reporta esta información al Registro Nacional de Accidentes de Tránsito (RNAT) y en este caso para la ciudad de Bogotá, la Secretaría Distrital de Movilidad (SDM) nos permite tener a nosotros los ciudadanos estos datos desde el año 2015 a través de su portal SIMUR (Sistema Integrado de Información sobre Movilidad Urbana Regional).

Gracias a estos datos publicados, se obtiene una síntesis de características inherentes a la ocurrencia del siniestro vial en donde tenemos: características demográficas de los actores involucrados, hipótesis o causas del siniestro vial y características tanto del vehiculo como de la via donde ocurrió el siniestro.

A continuación, se presenta una tabla 0.2 en donde la SDM estructuró la información del siniestro vial en un periodo comprendido entre el año 2002 hasta el 3 de julio de 2023 ¹.

Tabla 0.2: Datos de siniestralidad publicados por SDM - Portal SIMUR

Tabla	Columnas	Registros
Accidente	19	464699
Actor Vial	13	1657681
Causa	10	672815
Vehículo	9	1427204
Vía	19	483952

Datos geográficos: Debido a una falencia presente en los datos del entorno proximo al sitio de ocurrencias del siniestro recopilada en el formulario IPAT, se incluyen a la base de datos variables exógenas de los siniestros viales en donde esta información se obtiene mediante procesamiento de sistemas de Información Geográfica (SIG) y en donde sus características a incluir son las siguientes:

¹Datos tomados de la WEB: <https://www.movilidadbogota.gov.co/web/simur>

Cantidad de Comparendos impuestos entre 2015 - 2021 en la zona del siniestro.

Si se cuenta con red de cicloinfraestructura a menos de 200 mts del siniestro.

El estrato predominante en el sector del siniestro.

Tipo de categoría de la vía donde ocurre el siniestro.

El uso del suelo predominante en el sector donde ocurrió el siniestro.

Descripción de los datos

A continuación describimos la información contenida en cada una de las tablas recopiladas:

Descripción causa. Dentro de la tabla causa [0.3](#) se encuentran relacionadas las posibles causas o hipótesis del siniestro vial.

Tabla 0.3: Diccionario de datos tabla causa de siniestros Viales en Bogotá D.C.

Campo	Descripción	Dominios
Objectid	Identificador en base de datos	
Formulario	Identificador del Formulario	
Codigo Accidente	Código del Accidente	
Codigo Vehiculo	Código del vehículo	
Codigo Causa	Código de la causa	
Nombre	Hipótesis del siniestro vial	120 Opciones según IPAT.
Tipo	A quien se le atribuye la causa o hipótesis	Conductor, Peaton, Vía, Ciclista-Motociclista, Vehículo, Pasajero-Acompañante.
Tipo Causa	Tipo de causa o hipótesis según el IPAT	Conductor, Peaton, Vía, Ciclista-Motociclista, Vehículo, Pasajero-Acompañante.

Descripción vía. En la tabla vía [0.4](#), se relacionan las características de las vías donde ocurrió el siniestro vial.

Tabla 0.4: Diccionario de datos tabla vía de siniestros viales en Bogotá D.C.

Campo	Descripción	Dominios
Codigo	Código Accidente	
Objectid	Identificador en base de datos	
Formulario	Identificador del Formulario	
Codigo Accidente	Código del Accidente	
Codigo Via	Identifica las vías que se cruzan, en caso de ser una intersección.	1, 2
Geometrica A	Corresponde a la característica geométrica de la vía.	Recta, Curva
Geometrica B	Corresponde a la característica geométrica de pendiente de la vía	Plano, Pendiente
Geometrica C	Corresponde a características anexas de la vía. Indica si tiene bahía estacionamiento, berma o andén.	Andén, Berma, Bahía estacionamiento.
Utilizacion	Define la utilización dada a la vía.	Doble sentido, Un Sentido, Reversible, Contraflujo, Ciclovia.
Calzadas	Cantidad De Calzadas de la Vía	Una, Dos, Tres, Cuatro o más, Variable.
Carriles	Cuantos Carriles tiene cada Calzada	Uno, Dos, Tres, Cuatro o más, Variable
Material	Indica el material principal en que esta construida la superficie de la vía.	Concreto, Asfalto, Afirmado, Adoquin, Tierra, Empedrado, Otro.
Estado	Indica el estado de la superficie de la vía. (Se considera la trayectoria pre-impacto)	Bueno, Derrumbes, Con huecos, Rizada, Hundimientos, En reparación, Parchada, Fisurada, Inundada.
Condiciones	Indica las condiciones en que se encontraba la superficie de la vía al momento del accidente, húmeda, seca, si hay material suelto o aceite o algún agente físico similar.	Seca, Material suelto, Húmeda, Alcantarilla destapada, Lodo, Aceite, Material Orgánico, Otra.
Iluminacion A	Si el siniestro ocurre en horas nocturnas o en condiciones de baja visibilidad, se determina si existe o no la iluminación artificial.	Sin, Con
Iluminacion B	Si se determina la existencia de iluminación artificial se indican las condiciones de la Iluminación.	Buena, Mala.
Agente Transito	Indica la presencia de un agente de tránsito direccionando o regulando el flujo vehicular, en el momento de ocurrir el accidente.	No, Si.
Semaforo	Indica el estado del semáforo en caso de existencia.	Operando, Oculto, Intermitente, Con daños, Apagado, Ninguno.
Visual	Indica si existían elementos que limitaron u obstruyeron la visibilidad de los conductores involucrados en el momento del accidente, que impidieron observar un riesgo o peligro.	Vehículo estacionado, Vallas, Encandilamiento, Poste, Construcción, Árbol, Vegetación.
Codigo	Código De La Vía	

Descripción Actor Vial. La información del actor vial se almacena en una tabla donde se enmarcan todas sus características (Tabla 0.5).

Tabla 0.5: Diccionario de datos tabla actor vial de siniestros Viales en Bogotá D.C.

Campo	Descripción	Dominios
Objectid	Identificador en base de datos	
Formulario	Identificador del formulario de accidente	
Codigo Accidentado	Identificador del accidentado	
Codigo Victima	Identificación de la víctima	
Codigo Vehiculo	Identificación del vehículo	
Condicion	Condición del actor vial involucrado en el siniestro	Conductor, Ciclista, Motociclista, Pasajero, Peatón, Sin Información.
“Estado	Indica la gravedad del siniestro.	Ileso, Herido, Muerto
“Muerte Posterior	Indica si hubo una muerte posterior del actor vial	N, S
Fecha Posterior Muerte	Fecha de la muerte posterior del actor vial	
“Genero	Género del actor vial	Sin información, Masculino, Femenino
Fecha Nacimiento	Fecha Nacimiento del actor vial	
Edad	Edad del actor vial	
Codigo	Código del actor vial	

Descripción Vehículo. En esta tabla (0.6) se encuentra la información de los vehículos involucrados en los siniestros viales.

Tabla 0.6: Diccionario de datos tabla vehículo de siniestros viales en Bogotá D.C.

Campo	Descripción	Dominios
Objectid	Identificador en base de datos	
Formulario	Identificador del formulario de accidente	
Placa	Indica la placa de identificación del vehículo	
Codigo Vehiculo	Código del Vehículo	
Clase	Indica la clase de vehículo según la licencia de tránsito.	Motocicleta, Automóvil, Bus, Tractocamión, Camión, Furgón, Camioneta, Microbús, Campero, Buseta, Volqueta, Motocarro, Bicicleta, No Identificado, Tracción animal, Otro, Motociclo, M. Industrial, M. Agrícola, Bus Articulado Bus Alimentador, Metro, Ambulancia, Tren, Bicitaxi, Cuatrimoto, Remolque, Semi-Remolque
Servicio	Indica la clase servicio de acuerdo a la licencia de tránsito.	Particular, Público, Oficial, Diplomático, Sin Información.
Modalidad	Se indica la modalidad de transporte público terrestre automotor a la que pertenece. En el caso de transporte público de pasajeros se identifica si es colectivo, individual, masivo, especial, turismo, especial escolar, especial. En el caso de carga normal se identifica si es extradiimensionada, extrapesada o mercancía peligrosa.	Pasajeros - Masivo, Carga - Mercancía Peligrosa, Carga-Pesada, Pasajeros - Individual, Pasajeros - Colectivo, Pasajeros - Especial Turismo, Pasajeros - Especial Ocasional, Mixto, Pasajeros - Especial Escolar, Carga - Extrapesada, Pasajeros - Especial Asalariad, Carga - Extra Dimensionada
Enfuga	Indica si el vehículo se fugo	N S
Codigo	Identificación del vehículo	

Descripción Accidente. En la siguiente table (0.7) encontramos toda la información relacionada al siniestro vial ocurrido.

Tabla 0.7: Diccionario de datos tabla accidentes de los siniestros viales de Bogotá D.C.

Campo	Descripción	Dominios
Objectid	Identificador en base de datos	
Formulario	Identificador del formulario de accidente	
Codigo Acc	Código del accidente	
Fecha Occur	Fecha de ocurrencia del accidente	
Hora Occur	Hora del accidente	
Ano Ocurre	Año del accidente	
Mes Ocurre	Mes del accidente	
Dia Ocurre	Día del accidente	
Direccion	Dirección donde ocurrió el accidente	
Gravedad	Gravedad del accidente	Solo Daños, Con Heridos, Con Muertos
Clase Acc	Indica la clase de accidente primario o inicial. <u>Choque</u> : Es el encuentro violento entre dos (2) o más vehículos, o entre vehículo y un objeto fijo. <u>Atropello</u> : Accidente en donde un peatón es objeto de un impacto por un vehículo. <u>Volcamiento</u> : Es el hecho primario en el cual el vehículo pierde su posición normal durante el accidente y puede quedar de manera lateral o longitudinal. <u>Caída de ocupante</u> : Se refiere a la caída de un conductor, acompañante o pasajero desde un vehículo hacia el exterior, interior o dentro del mismo. <u>Incendio</u> : Se refiere a aquellos casos en el que el Vehículo se incendia sin que exista accidente previo. <u>Otro</u> : Se refiere a aquel accidente no asimilable dentro de las cinco situaciones descritas.	Choque, Atropello, Volcamiento, Caída del ocupante, Otro, Autolesión, Incendio.
Localidad	Localidad donde ocurre el siniestro	20 Localidades de Bogotá D.C.
Municipio	Municipio donde ocurre el siniestro	Bogotá DC
Fecha Hora	Fecha y hora del accidente	
Latitud	Coordenada Latitud del sitio donde ocurre el accidente	
Longitud	Coordenada Longitud del sitio donde ocurre el accidente	
Civ	Ubicación del sitio de ocurrencia dentro de la vía, kilómetro	
Pk Calzada	Ubicación del sitio de ocurrencia dentro de la vía, kilómetro	
Geometry	Geometría del shapefile	

Descripción Factores Exógenos. Esta tabla contiene las características del entorno donde ha ocurrido el siniestro vial y fue obtenida a partir del análisis espacial (Tabla 0.8)

Tabla 0.8: Diccionario de datos tabla factores exógenos de siniestros viales en Bogotá D.C.

Campo	Descripción	Dominios
Formulario	Identificador del formulario de accidente	
Estratozon	Estrato predominante en la manzana más cercana al lugar donde ocurrió el siniestro	0, 1, 2, 3, 4, 5, 6
Tipovia	Identifica la clasificación de la vía donde ocurrió el siniestro	1- Malla vial Arterial, 2- Malla vial Intermedia, 3- Malla vial Local, 4- Malla vial Peatonal, 5- Malla vial Rural, 6- Sin definir, 7- Proyectada
Usozona	Identifica el uso del suelo predominante en la zona donde ocurrió el siniestro	01-Residencial, 03-Industrial, 04-Dotacional Público, 05-Recreacional público, 06-Dotacional Privado, 07-Mineros, 08-Recreacional privado, 21-Comercio corredor comercial, 22-Comercio en centros comerciales, 23-Comercio puntual, 24-Parqueaderos, 61-Urbanizado no edificado, 62-Urbanizable no urbanizado, 63-No urbanizables y suelo protegido, 64-Lotes del Estado, 65-Vías, 66-Espacio Público, 67-Predios con mejoras ajenas, 81-Agropecuarios
Compar200M	Cantidad de comparendos interpuestos en la zona donde se ubica el siniestro. Las zonas se definen para la ciudad de Bogotá en área de 200mX200m	
Redbici100	Indica si a menos de 100 metros del lugar donde ocurrió el siniestro existe cicloinfraestructura	Si, No

Verificación de la calidad de datos

Debido a posibles errores de codificación, valores perdidos, u otras incoherencias que pueden dificultar nuestro análisis, es importante realizar una verificación en la calidad de los datos. Para lograrlo, utilizamos las herramientas del software SPSS Modeler 18.12.1 en lo relacionado con el nodo auditoría de datos. Para cada campo individual, se evalúan aspectos como el

tipo de dato, detección de valores atípicos o extremos, acciones sugeridas en caso de detección de anomalías, imputación de valores perdidos, método de evaluación aplicado, porcentaje de completitud (Completo), número de registros válidos, valor nulo y cadenas vacías.

Preparación de los Datos

Para esta sección, preparamos los datos que fueron usados durante el entrenamiento de los modelos.

Selección de Datos

Datos fuera del área de estudio. Se revisa que cada siniestro vial se encuentre dentro de nuestra área de estudio (Bogotá). Con ayuda de un análisis SIG de superposición de mapas, se cruzaron los datos de los siniestros viales y el perímetro urbano de la ciudad de Bogotá, resultando que 86 siniestros no ocurrieron dentro de la ciudad o presentan error en su georreferenciación. Se eliminan estos registros de la base de datos.

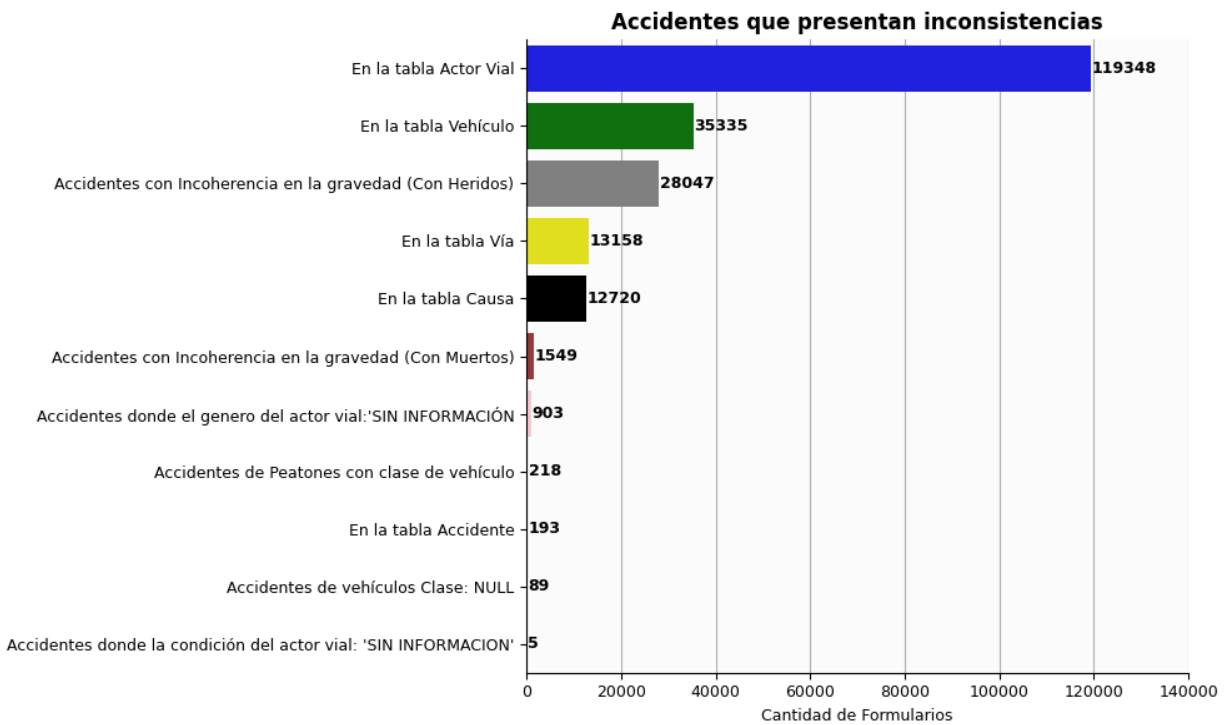
Ambigüedades. Se opta por eliminar la variable Estado del Actor vial, debido a que genera ambigüedad respecto a la variable Gravedad del siniestro vial. Es decir, el siniestro vial puede tener la máxima gravedad, con muertos, mientras uno de los actores viales involucrados puede salir ileso.

Selección de Datos

En la siguiente figura 2 se muestra la cantidad de accidentes que deben ser eliminados de la base de datos discriminados por la inconsistencia que presentan. En consonancia con las observaciones durante las auditorías de calidad de datos, se ha identificado que algunos datos cumplen con las características que se necesitan. Cabe aclarar que al eliminar un registro de cualquier ta-

bla, se puede fragmentar la información de un accidente, debido a que ésta contiene datos relacionados con el actor vial, el vehículo, la vía y la causa. Por ende, lo más apropiado es depurar por completo todos los datos asociados con el accidente, para mantener la integridad de la modelación.

Figura 2: Inconsistencias por tabla



Nota: La figura ilustra la cantidad de registros que presentan inconsistencias en cada una de las tablas insumo.

Construcción de nuevos datos

Algunas de las variables resultantes del proceso de limpieza de datos se someten a procedimientos como la normalización, el cálculo de intervalos según sus características o la categorización, con el fin de optimizar su utilidad y relevancia.

Transformación atributos. De acuerdo al análisis de los datos, se encontro la necesidad de transformación de algunas variables debido principalmente a la alta variabilidad que presentan. *Accidente.* Del campo "HORA" se establecen los siguientes intervalos de tiempo para determinar si un siniestro ocurrió en la mañana, en la tarde, o en la noche. También se crea el campo calculado tal y como se muestra en la tabla ??.

Tabla 0.9: Conversión a variable jornada

RECATEGORIZACIÓN JORNADA	HORA
MAÑANA	5:30 AM A 12:00 PM
TARDE	12:00 PM A 7:00 PM
NOCHE	7:00 PM A 6:00 AM

Otra de las transformaciones, con respecto al tiempo consistió en determinar si el siniestro ocurrió durante la hora valle o en la hora pico, ya sea por la mañana o por la tarde, tal y como se muestra en la tabla 0.10. Del campo "FECHA" establecen los siguientes intervalos de tiempo pa-

Tabla 0.10: Conversión a variable hora pico

RECATEGORIZACIÓN HORA PICO	HORA
PICO AM	5:30 AM – 9:00 AM
PICO PM	3:30 PM – 7:00 PM
HORA VALLE	RESTO

ra determinar si un siniestro ocurrió en un día festivo o domingo, un sábado o de lunes a viernes. 0.11.

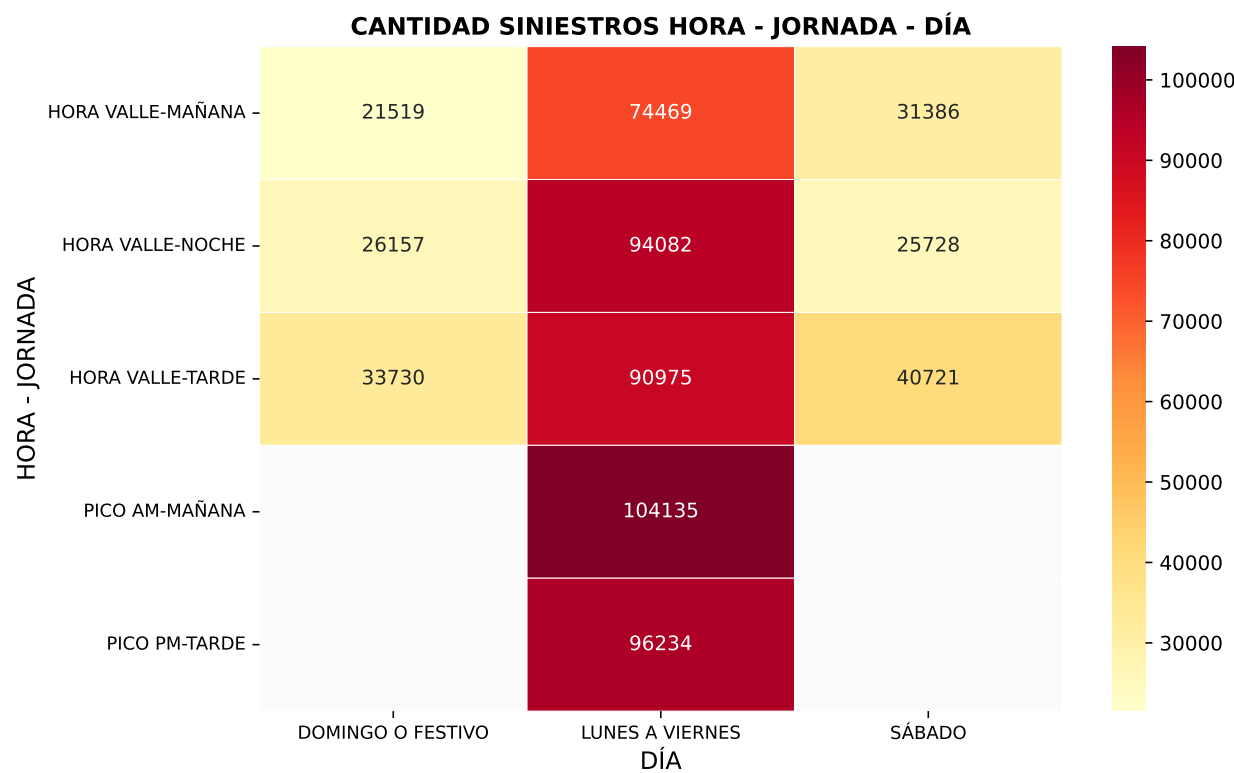
Tabla 0.11: Conversión variable día de la semana

DÍA	RECATEGORIZACIÓN DÍA
DOMINGO O FESTIVO	DOMINGO O FESTIVO
SÁBADO	SÁBADO
LUNES	LUNES A VIERNES
MARTES	LUNES A VIERNES
MIÉRCOLES	LUNES A VIERNES
JUEVES	LUNES A VIERNES
VIERNES	LUNES A VIERNES

En la Figura 3 se presenta un gráfico de frecuencias que ilustra la distribución de siniestros viales. Destacandose que la mayoría de estos siniestros se concentran en días laborales de

Lunes a viernes.

Figura 3: Siniestros por día, jornada y hora pico



Nota: La figura ilustra un mapa de calor donde se resaltan la cantidad de siniestros por jornada y por tipo de día de la semana.

Causa.De acuerdo con la información contenida en el manual de diligenciamiento del informe policial de accidente de transito, se asume que la causa del accidente puede ocurrir por tres factores principales:

- **Factor Humano:** Este factor se refiere a la responsabilidad del actor vial involucrado en el siniestro.
- **Factor Mecánico (Vehículo):** Este factor se relaciona con el estado mecánico del vehículo involucrado en el siniestro.
- **Factor del Entorno (Vía y Condiciones Ambientales):** Este factor engloba las condiciones del entorno en el que se produce el siniestro.

En la siguiente tabla 0.12 se muestra la clasificación de cada uno de los códigos en relación a un factor relacionado.

Tabla 0.12: Conversión variable causa del siniestro vial a factor que lo ocasionó

CÓDIGO CAUSA	FACTOR
090 - 099	FACTORES HUMANOS
101-157	FACTORES HUMANOS
118	FACTORES MECÁNICOS
126, 138	FACTORES DEL ENTORNO
136	FACTORES MECÁNICOS
201-217	FACTORES MECÁNICOS
301-308	FACTORES DEL ENTORNO
401-411	FACTORES HUMANOS
501-506	FACTORES HUMANOS

Vehículo. En la tabla Vehículo, se lleva a cabo la agrupación de las diversas clases de vehículos según su grado de exposición, considerando aspectos como si son motorizados o su tamaño 0.13.

Tabla 0.13: Conversión variable clase vehículo

CLASE VEHÍCULO	CLASE AGRUPADA
“MOTOCICLETA”, “CUATRIMOTO”, “MOTOCICLO”, “MOTOCARRO”	MOTOCICLETAS O SIMILARES
“BICICLETA”	BICICLETA
“CAMIONETA”, “CAMPERO”, “AUTOMOVIL”	AUTOMÓVILES O CAMIONETAS
“MICROBÚS”, “BUS”, “BUSETA”, “BUS ARTICULADO”, “BUS” “ALIMENTADOR”, “TREN”, “METRO”	TRANSPORTE PÚBLICO DE PASAJEROS
“CAMION”, “FURGON”, “VOLQUETA”, “TRACTOCAMION”	TRANSPORTE DE CARGA
, “OTRO”, “TRACCION ANIMAL”, “M. INDUSTRIAL”, “NO IDENTIFICADO”, “M. AGRICOLA”, “AMBULANCIA”,	OTROS

Vía. A partir de la tabla vía se ha realizado la categorización de los tipos de vías, como se detalla en la siguiente tabla 0.14

Uso del suelo. Considerando las diversas características del espacio público, se ha llevado a cabo una clasificación que busca simplificar la categorización de los usos del suelo y en donde se determinan cinco generales. Tabla 0.15

Tabla 0.14: Conversión tipo de vía

TIPO_VIA	TIPO VÍA AGRUPADO
1- Malla vial Arterial	Malla vial Arterial
2- Malla vial Intermedia	Malla vial Intermedia
3- Malla vial Local, 4-Malla vial Peatonal, 5- Malla vial Rural, 6- Sin definir, 7- Proyectada	Malla vial Local, Peatonal o Rural

Tabla 0.15: Conversión variable clase de uso

USO DE ZONA	USO DE ZONA AGRUPADO
01 Residencial	RESIDENCIAL
21 Comercio corredor comercial 22 Comercio en centros comerciales 23 Comercio puntual 24 Parqueaderos	COMERCIAL
03 Industrial	INDUSTRIAL
04 Dotacional Público 05 Recreacional público 06 Dotacional Privado 08 Recreacional privado 66 Espacio Público 65 Vias	DOTACIONAL O RECREACIONAL
61 Urbanizado no edificado 62 Urbanizable no urbanizado 63 No urbanizables y suelo protegido 64 Lotes del Estado	NO URBANIZADO O SUELO PROTEGIDO
07 Mineros 67 Predios con mejoras ajenas 81 Agropecuarios	OTROS

Factores Exógenos. Desde la tabla de información Exógena, y examinado el número de comparendos, se establece que su mediana es de 108 comparendos en las zonas donde ocurren siniestros viales, así como se ilustra en la tabla 0.16. En la tabla 0.17 se muestra la recategorización realizada.

Tabla 0.16: Estadísticas comparendos

Ítem	Valor
cantidad	639541.0
media	730.7
desviación estándar	1907.6
mínimo	0.0
Q1	108.0
Q2	277.0
Q3	669.0
máximo	89675.0

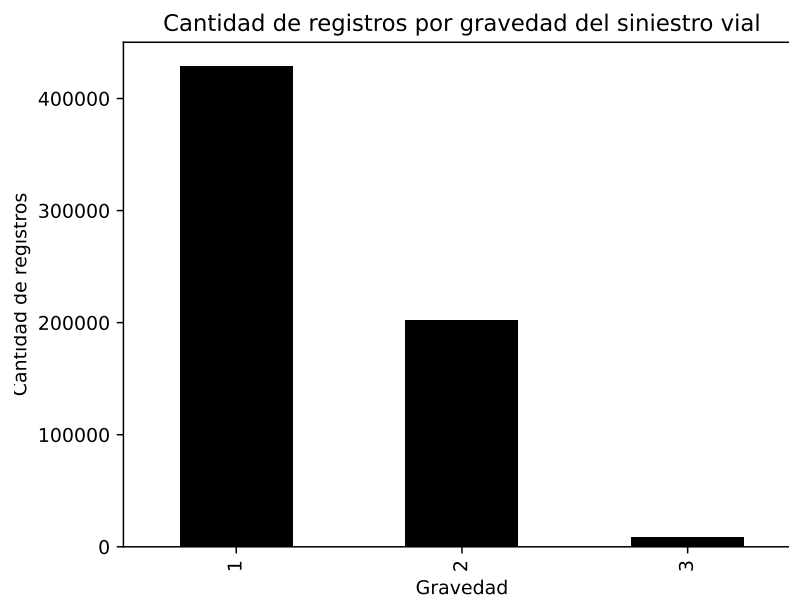
Balanceo de los datos. Una vez se realizan los procesos de limpieza y se crean los nuevos campos, se entrará a considerar la distribución de la variable objetivo respecto a las tres clases

Tabla 0.17: Conversión variable cantidad comparendos

ZONA COMPARENDOS	CANTIDAD DE COMPARENDOS EN EL PERIODO
BAJO	0-107
MEDIO	108-276
ALTO	277-668
MUY ALTO	> 669

de gravedad de los siniestros viales. En la Figura ?? se muestra el desbalanceo entre las distintas clases, siendo la clase 1 (Solo daños) la que presenta más casos, seguida por la clase 2 (Con heridos) y la clase 3 (Con muertos) en donde se tiene la menor cantidad de registros.

Figura 4: Distribución de los siniestros viales según su gravedad



Nota: Donde: 1 = Solo Daños, 2 = Con Heridos, 3 = Con Muertos..

García (2021) concluye en su estudio que, la técnica de resampling que se seleccione depende de la distribución de los grupos y los resultados dependerán del algoritmo de modelado, pues presentan resultados dispares. Sin embargo, se recomienda la utilización de métodos de resampling no heurísticos pues presentan mejores rendimientos. Las técnicas recomendadas son Undersampling Aleatorio y Oversampling Aleatorio.

Integración de datos

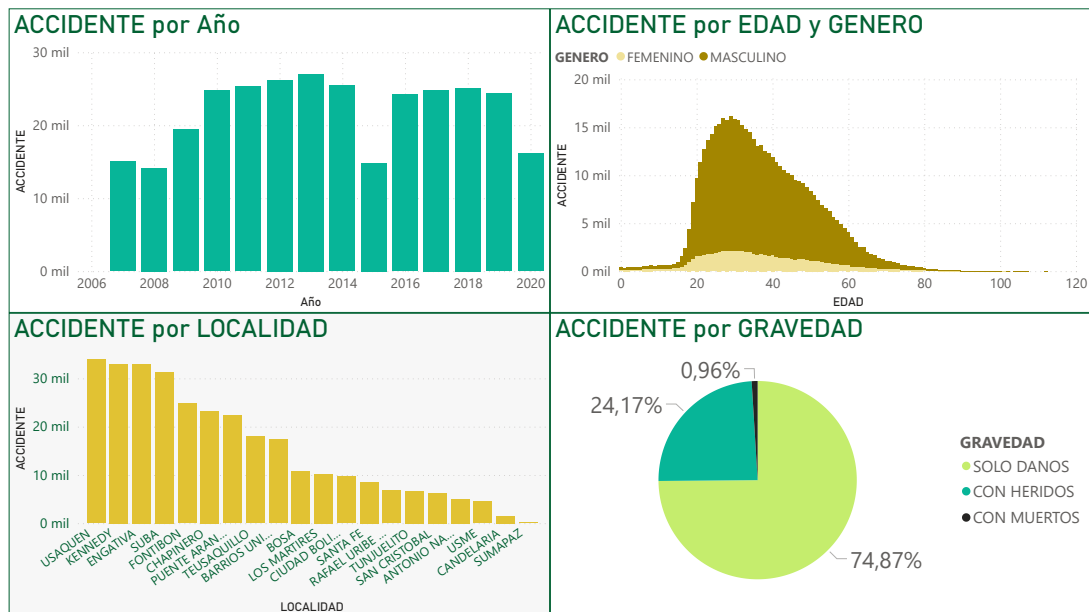
El orden de uniones fue Actor Vial, Vehículo, Causa, Accidente, Vía y por último los Factores Exógenos. Una vez realizada la integración de datos, se debe conllevar un proceso de limpieza, que garantiza que el conjunto de datos final tenga la calidad adecuada. A continuación se describen las depuraciones realizadas:

- Elimina las filas que no tienen causas válidas del actor o de la vía.
- Identifica y elimina registros de accidentes incompletos relacionados con peatones o condiciones incorrectas de clase o servicio.
- Elimina registros donde la condición es “SIN INFORMACIÓN”.
- Corrige la gravedad de algunos accidentes.
- Verifica y corrige la gravedad en función de las condiciones.
- Elimina registros con heridos pero sin registros correspondientes.
- Elimina actores donde se desconoce el género.
- Realiza transformaciones en la columna “CODIGO_CAUSA” para asegurar su formato.

Análisis univariado

En esta sección utilizamos la herramienta PowerBI para representar gráficamente las características de cada una de las variables empleadas en el modelo. Se observa que en la muestra recopilada durante los años 2007, 2008, 2015 y 2020, se registran índices bajos de accidentes, mientras que en los demás periodos se observa un comportamiento de más de 25 mil eventos completos. También se puede destacar que la mayoría de los actores viales son hombres en el rango de edad de 20 y 40 años.

Figura 5: Análisis univariado



Nota: Figura que ilustra la información mas relevante siniestralidad vial con información depurada. El detalle se presenta en el Anexo 10..

Análisis bivariado

Dado que el objetivo del trabajo es determinar el grado de severidad en los accidentes de tráfico, es importante aclarar que la variable objetivo en "GRAVEDAD", por lo que se realiza un análisis bivariado respecto a las variables independientes. Se emplean tablas cruzadas para para cumplir el objetivo utilizando el software SPSS Statistics donde se incluye el estadístico chi-cuadrado. Se plantea la prueba de hipótesis con un nivel de significancia (Grado de confianza de 95 %) de $\alpha = 0,05$.

Se define que:

H_0 : Existe asociación entre la variable "GRAVEDAD" y la variable Y explicativa. $Valorp < 0,05$

H_a No existe asociación entre la variable "GRAVEDAD" y la variable Y explicativa. $Valorp > 0,05$

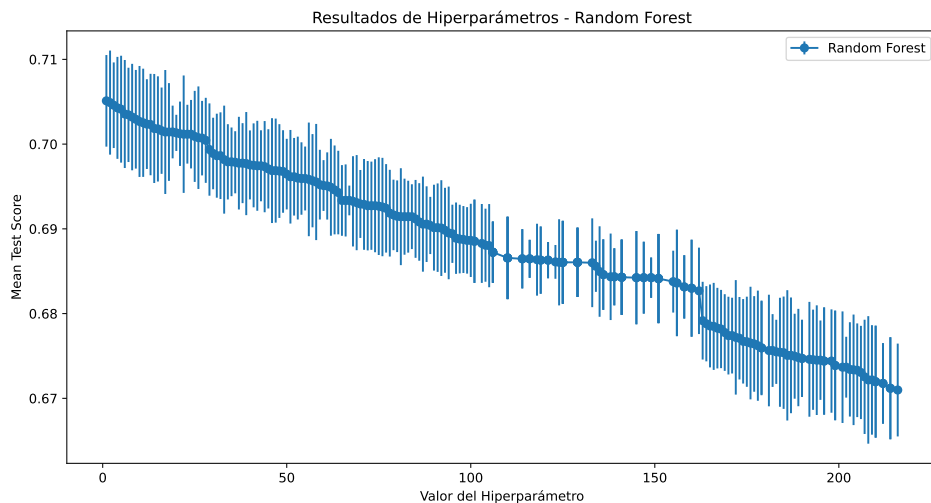
IA - (segunda entrega)

Bosques Aleatorios. La grilla propuesta para este modelo implica cinco parámetros, son:

- n estimators: 100, 200, 300
- max depth: None, 10,20,30
- min samples split: 2, 5, 10
- min samples leaf: 1, 2, 4
- max features: auto, sqrt, log2

Una vez ejecutado el algoritmo durante 444 minutos, se obtiene que la mejor configuración con exactitud de 0.70 corresponde con los parametros: max depth: 20, max features: sqrt, min samples leaf: 1, min samples split: 2, n estimators: 300. En la Figura 6 se detallan los resultados obtenidos para las 243 posibles configuraciones.

Figura 6: Búsqueda de hiperparámetros Bosques Aleatorios



Proximos Pasos - (Tercera entrega)

Figura 7: Matriz de confusión Modelo Random Forest

Matriz de Confusión

Valor Real	Predicción		
	Solo Daños	Con Heridos	Con Muertos
Solo Daños	1485	66	55
Con Heridos	426	832	348
Con Muertos	213	317	1087

De la matriz de confusión se pueden obtener las métricas por cada una de las clases, en la Tabla 0.18 se muestra el detalle. De los resultados obtenidos, se puede deducir que al modelo de bosques aleatorios entrenado, le va muy bien identificando los casos de gravedad “Solo Daños” con una Recall de 92 %, es decir que tan solo el 8 % de los casos que fueron siniestros con solo daños fueron categorizados equivocadamente por el modelo. En cuanto a la precisión de esta categoría, el modelo se equivoca en un 30 % de los casos al categorizarlos como Solo Daños cuando corresponden a con heridos o con muertos. Para la Categoría “Con Herido” el modelo presenta las métricas mas bajas, se equivocó en un 48 % de los casos que pertenecían a esta categoría, identificando un 27 % como Solo Daños y el 21 % restante como Con Muerto. En cuanto a la precisión de la categoría Con Heridos, el modelo logra identificar correctamente el 68 % de los casos con esta Gravedad y se equivoca al categorizar siniestros que realmente fueron Solo Daños 5 % y Con Muerto 27 %. Finalmente, la categoría “Con Muertos” tiene un Recall del 67 %, lo cual implica que el modelo erróneamente clasifico el 13 % de los siniestros del conjunto de prueba como Solo Daños y el 20 % como Con Heridos. La precisión de la Gravedad Con Muertos es del 73 %, el modelo clasificó en esta categoría erróneamente un 4 % de siniestros que eran Solo Daños y un

23 % que correspondían a la Gravedad Con Heridos. De este análisis se concluye que, el modelo para las tres clases de Gravedad de siniestros viales tiene una precisión en promedio del 70 %.

Ahora bien, se debe considerar que las tres categorías establecidas para la gravedad de un siniestro implicaran la existencia de personas heridas o muertas. Bajo este contexto, lo que busca este modelo es hacer predicciones acercándose lo más posible a la realidad con la identificación de aquellos casos que revisten mayor gravedad, por tanto se busca que la cantidad de falsos negativos sea lo menor posible para las categorías de “con muertos” y “con heridos”. En este orden de ideas, se concluye que el comportamiento del modelo se debe analizar con detalle la métrica de recall o sensibilidad.

En la tabla 0.18 se observa que, para el caso de los siniestros viales de gravedad con muertos la sensibilidad es del 67 % y para el caso de la gravedad con heridos es del 52 %.

Tabla 0.18: Métricas por clase modelo árboles aleatorios

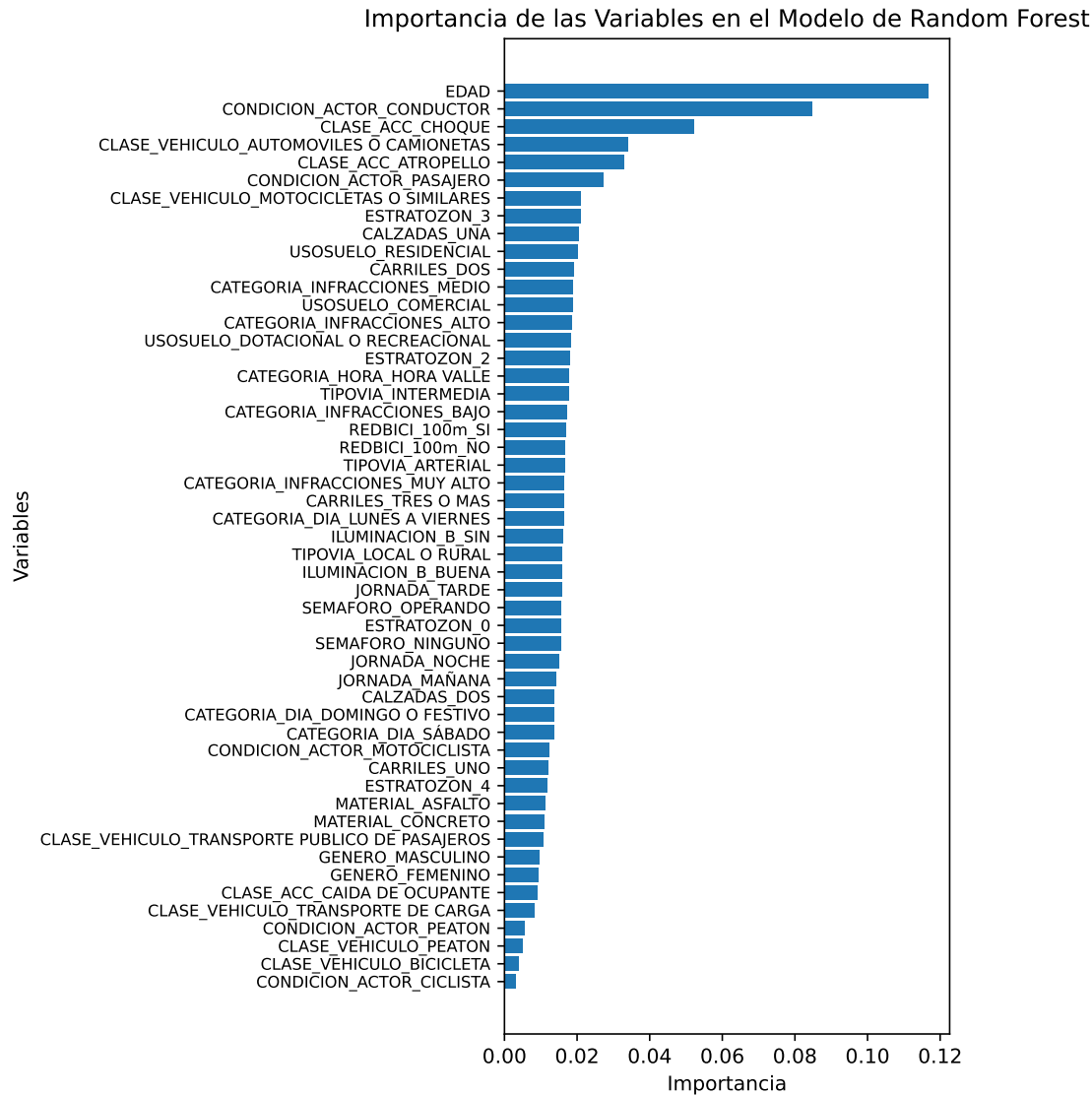
Categoría	Precisión	Recall	F1-Score
Solo daños	0.70	0.92	0.80
Con heridos	0.68	0.52	0.59
Con Muertos	0.73	0.67	0.70

Si bien existen las metas de reducción de la siniestralidad vial en general, las autoridades hacen grandes esfuerzos particularmente en la reducción de los casos más graves de siniestros viales, donde hay pérdidas de vidas o se generan lesiones a los actores involucrados. Se puede hablar entonces que sí; el modelo es capaz de discriminar estos dos casos entre “solo daños”, aun cuando se equivoque entre ellos, lograría impactar significativamente este fenómeno. Si se realiza este análisis conjunto entre clases la métrica de recall o especificidad sería del 80 %.

En síntesis, el modelo construido tiene buenos resultados con un accuracy o exactitud general del 70 % y puede ser utilizado para obtener predicciones bajo diferentes escenarios de siniestralidad vial, logrando con ello la toma de medidas oportunas para reducir la ocurrencia de siniestros viales que revistan la mayor gravedad.

Finalmente, en la Figura 8 se muestra la importancia de cada una de las variables según el modelo de bosques aleatorio.

Figura 8: Importancia de las variables dentro del modelo de bosques aleatorios

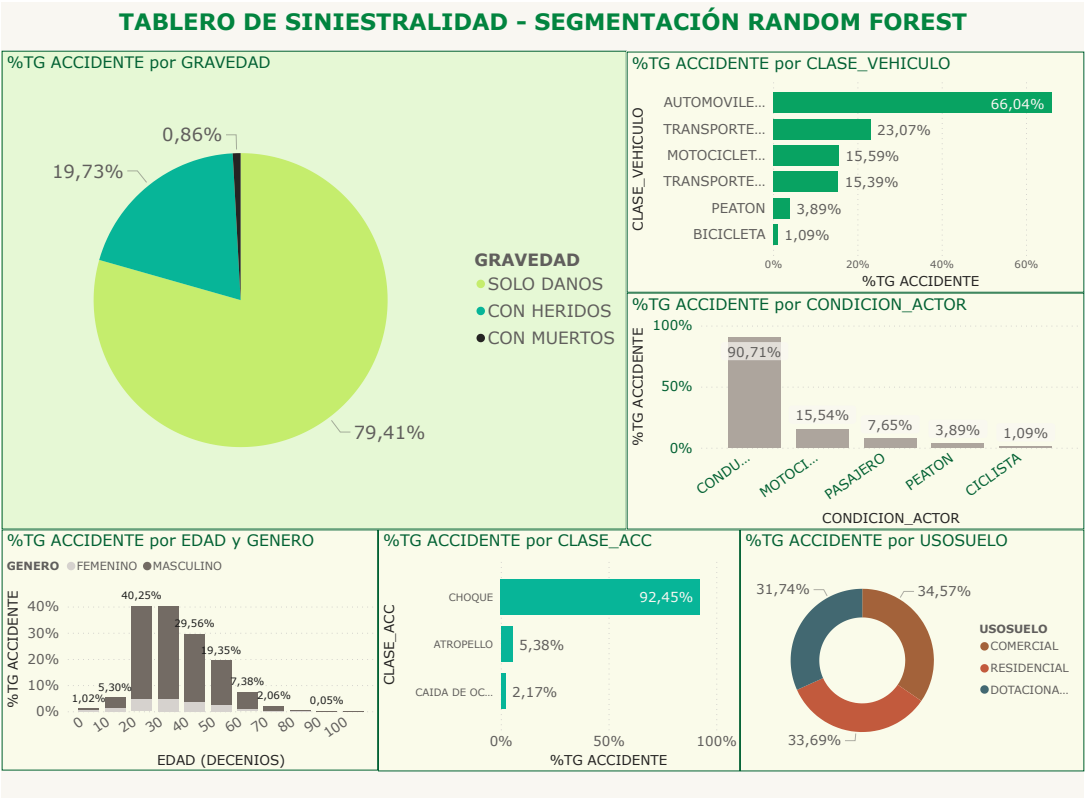


Propuesta de mejora al observatorio de movilidad para seguimiento de siniestros viales

Para dar cumplimiento al objetivo de proponer una mejora a la sección de Seguridad Vial del observatorio de movilidad de Bogotá D.C, se ha diseñado un tablero de visualización para resaltar la influencia de las variables independientes resaltadas en el modelo de Bosques aleatorios. Es importante aclarar que, dado que los datos han sido sometidos a un proceso de limpieza y transformación, es relevante realizar una comparación en términos relativos. En la siguiente figu-

ra 9 se ilustra el tablero propuesto, cuya visual más importante (GRAVEDAD) se encuentra en la parte superior izquierda. También podemos observar la proporcionalidad de variables como: clase de vehículo, condición del actor vial, edad, género, clase de accidente y el uso del suelo.

Figura 9: Propuesta de tablero segmentado



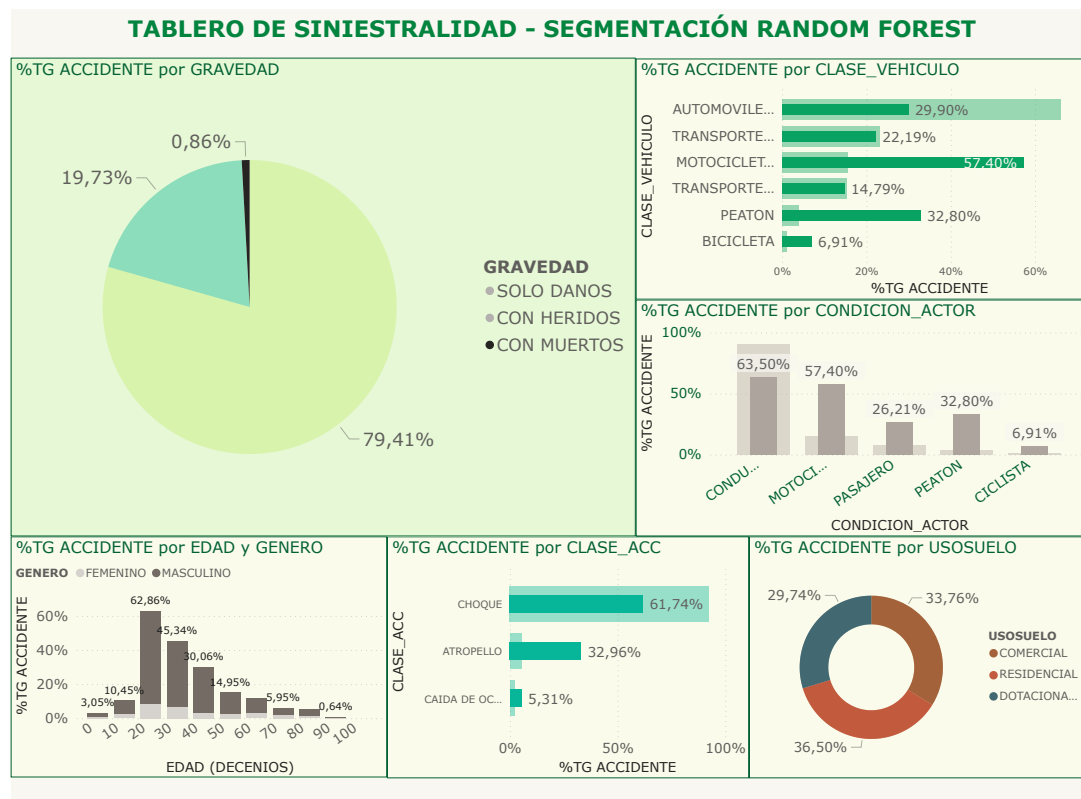
Nota: La figura ilustra el tablero propuesto para hacer seguimiento a la gravedad de los siniestros viales, sobre las cinco variables que más inciden.

Se observa que la muestra segmentada representa un 0,86 % de muertos, un 19,73 % de heridos y un 79,41 % de accidentes con solo daños. Dentro de las características más influyentes, se tiene una participación de automóviles 66 %, en condición de conductor de 90 %, un participación del 40 % similar en rangos de edad de 20-30 años y 30-40 años. Así mismo, una participación del 92,49 % en accidentes con choque, y una participación bastante equitativa en uso del suelo 34 % Comercial, 31 % Dotacional y 33 % Residencial.

Ahora bien, si se resalta un grado de severidad con muertos, los resultados cambian bastante para cada variable ilustrada.

En la figura 10 se muestra que la influencia de motocicletas involucradas en la gravedad de siniestros viales con categoría muerto, pasa a repuntar con un 57,4 %, la condición de conductor deja de ser tan relevante y pasa de 92 % a 63 %, dando lugar a una participación mayor por parte de los motociclistas (57 %) y peatones (7 %). De la misma manera, la edad entre 20-30 años toma protagonismo con un 62 %. Así mismo pierde relevancia la clase choque (61 %), por su parte toma fuerza la clase atropello (32 %), uso del suelo, y adicional aumenta la frecuencia relativa en zonas residenciales 36,5 %.

Figura 10: Propuesta de tablero segmentado, filtrado para la gravedad “Con Muertos”



Nota: La figura ilustra la variación del e seguimiento para el caso cuando la gravedad de siniestro vial es “Con Muerto”.

Esta comparación permite visualizar claramente las áreas de interés para aquellos que deseen abordar la severidad de los accidentes de tránsito.

Lecciones Aprendidas - (Tercera entrega)

El presente estudio ha construido un modelo para predecir la severidad de los siniestros viales a través del algoritmo de árboles aleatorios, su utilización puede ayudar en el seguimiento y reducción de fatalidades de los siniestros. Dado que las autoridades de Bogotá D.C. han puesto a disposición de la sociedad los registros de los siniestros viales dentro de nuestro rango de tiempo establecido (2002 - 2023) y una vez procesada esta información, el conjunto de datos se ha compuesto de 31 variables y 639,136 registros, con una ventana de tiempo final del año 2007 a 2020, la cual brinda una calidad apropiada en los datos, y siendo a su vez, resultado de un proceso de depuración, limpieza y transformación de los datos. Al realizar este proceso se detectaron incoherencias en las bases de datos de siniestros viales en la ciudad de Bogotá que requieren validación y depuración por parte de la administración distrital. Dentro de los errores más significativos se encuentran siniestros viales con gravedad “Con muertos” donde no se reportan actores viales fallecidos; siniestros viales con gravedad “Con heridos” donde solo reportan actores viales ilesos.

El análisis univariado revela que hubo una baja incidencia de accidentes en los años 2007, 2008, 2015 y 2020, mientras que en otros años se registró un aumento con más de 25,000 eventos en la muestra. La mayoría de los actores viales son hombres de 20 a 40 años. Las localidades con más accidentes son Usaquén, Kennedy, Engativá y Suba, mientras que Usme, Candelaria y Sumapaz tienen menos.

La gravedad de los accidentes se encuentra distribuida de la siguiente manera: 74.87 % con solo daños, 24.17 % con heridos y 0.96 % con muertos.

En el análisis estadístico bivariado, se concluye que las pruebas de hipótesis chi-cuadrado de las variables seleccionadas por el modelo, exhiben una significancia que permite evaluar la asociación de cada variable con “GRAVEDAD”. Esto sugiere que todas las variables analizadas tienen cierta influencia en la severidad de los siniestros viales.

Se ha construido un modelo de bosques aleatorios (random forest) para predecir la severi-

dad de siniestros viales en la ciudad de Bogotá D.C.

Se realizó una búsqueda de hiperparámetros en grilla con el fin de mejorar el rendimiento del modelo. El modelo presentó una exactitud del 70 %, lo cual representa un buen desempeño para la predicción de la severidad en siniestros viales. El modelo presenta resultados de sensibilidad o recall del 52 % y 67 % para las clases con heridos y con muertos respectivamente. Lo anterior implicaría la existencia de errores de tipo falsos negativos, sin embargo, si se considera una cifra global entre estas dos categorías el indicador ascendería al 80 %, lo cual demostraría que el modelo le va bien en la identificación de los siniestros viales que revisten alta gravedad.

Como complemento al tablero suministrado por el observatorio de Movilidad de Bogotá, se diseñó un tablero que fortalecerá esta información brindando datos importantes para reducir la gravedad en siniestros viales medido en términos relativos. Específicamente se debe centrar la atención, en la regulación del uso de la motocicleta, el comportamiento de los conductores al usar este tipo de vehículo, la movilidad en zonas residenciales y la educación de población joven principalmente masculina en cultura vial.