

ViaVisión — Plataforma de Inteligencia de Riesgo Vial aplicado al municipio Calarcá 2021-2025

Autores: Elizabeth Garcés Isaza - elizabethgarcesisaza@gmail.com

Gabriel Garzón Henao - garzong2001@gmail.com

Jairo Acevedo Jaramillo - jairoacevedodj@gmail.com



MinTic – Datos Abiertos de Colombia

Proyecto de Caso

Resumen

La seguridad vial en Calarcá enfrenta retos significativos debido a los altos niveles de siniestralidad y a la falta de herramientas que integren información confiable, georreferenciada y actualizada. Actualmente, las decisiones institucionales se basan en reportes aislados y análisis manuales, dificultando la detección oportuna de zonas de riesgo y limitando la eficacia de las intervenciones.

ViaVisión surge como una solución integral orientada a consolidar múltiples fuentes de información —accidentalidad, parque automotor y puntos críticos— y aplicar metodologías avanzadas de análisis espacial, estadístico y de priorización territorial. El proyecto busca ofrecer una plataforma que permita identificar patrones de siniestros, caracterizar factores de riesgo, detectar zonas críticas y generar recomendaciones automáticas para la gestión preventiva de la seguridad vial.

Palabras claves: ViaVisión, siniestralidad, zonas, riesgo, puntos críticos, seguridad vial.

Tabla de contenido

1.	Descripción del problema	4
1.1.	Planteamiento del problema	4
2.	Justificación	5
3.	Objetivos	6
3.1.	Objetivo general.....	6
3.2.	Objetivos específicos	6
4.	Metodología	7
4.1.	Diseño metodológico	7
4.2.	Entendimiento del negocio.....	7
4.3.	Datos	8
5.	Fases de desarrollo del proyecto	8
5.1.	Limpieza, estandarización y descripción del dataset (ETL)	8
5.2.	Geocodificación territorial con H3	12
5.3.	Cálculo de indicadores por hexágono.....	12
5.4.	Modelo de score de Riesgo (0 – 100).....	13
5.5.	Integración con puntos críticos oficiales.....	13
5.6.	Sistema de recomendaciones por hexágono	14
5.7.	Mapa interactivo con Folium	14
5.8.	Generación del PDF del perfil de riesgo	15
6.	Limitaciones y recomendaciones	16
6.1.	Limitaciones detectadas en la calidad y disponibilidad de los datos	16
6.2.	Recomendaciones para mejorar la captura y utilidad analítica de los datos ...	17
7	Referencias bibliográficas.....	18

1. Descripción del problema

1.1. Planteamiento del problema

La seguridad vial en Colombia continúa siendo un desafío prioritario para las administraciones públicas, debido a los altos índices de siniestralidad y a la complejidad de los factores que influyen en su ocurrencia. La falta de herramientas que integren información confiable, actualizada y georeferenciada limita la capacidad institucional para anticipar riesgos, focalizar intervenciones y evaluar la efectividad de las acciones implementadas.

En el municipio de Calarcá, esta problemática se acentúa por la dispersión de las fuentes de datos —registros de accidentalidad, características del parque automotor y puntos críticos previamente identificados—, así como por la ausencia de mecanismos que permitan un análisis territorial sistemático. Actualmente, las decisiones en materia de seguridad vial suelen basarse en reportes aislados, procesos manuales o interpretaciones subjetivas, lo que retrasa la identificación de zonas de alto riesgo y disminuye la eficiencia de las intervenciones.

Frente a este panorama, ViaVisión se plantea como una solución integral que consolida múltiples fuentes de información y aplica metodologías avanzadas de análisis espacial, estadístico y de priorización territorial. Su objetivo es proporcionar un sistema de apoyo a la toma de decisiones que permite: Identificar patrones y tendencias de la siniestralidad, detectar zonas críticas con mayor precisión, caracterizar los factores asociados al riesgo vial, generar recomendaciones automáticas y adaptadas a las condiciones específicas de cada zona.

De esta manera, el proyecto busca transformar la gestión de la seguridad vial en Calarcá mediante un enfoque basado en evidencia, orientado a la anticipación del riesgo y a la intervención oportuna del territorio.

2. Justificación

La gestión de la seguridad vial requiere cada vez más enfoques basados en datos, capaces de anticipar riesgos y orientar intervenciones con precisión. En municipios como **Calarcá**, donde convergen dinámicas urbanas, rurales y de tránsito intermunicipal, la siniestralidad vial constituye un problema persistente que afecta la calidad de vida, la movilidad y el bienestar social. Sin embargo, la toma de decisiones suele verse limitada por la fragmentación de la información, la falta de análisis integrados y la ausencia de herramientas que permitan visualizar el riesgo de manera espacial y estratégica.

En este contexto, el desarrollo de ViaVisión se justifica por varias razones fundamentales, en primera instancia busca responder una necesidad institucional urgente, en donde las autoridades de tránsito requieren sistemas que integran datos dispersos y facilitan diagnósticos rápidos y confiables. ViaVisión permite consolidar información clave (accidentalidad, parque automotor, puntos críticos y variables territoriales), disminuyendo la dependencia en análisis de procesos manuales y mejorando la eficiencia operativa.

En segunda instancia, aporta evidencia para decisiones más acertadas, puesto que, la aplicación de análisis espacial, estadístico y modelos de priorización territorial permite comprender los patrones de riesgo desde una perspectiva multidimensional. Esto contribuye a que las intervenciones —señalización, control, infraestructura, educación vial— se fundamenten en criterios técnicos verificables. Por su parte, esta metodología también es permite optimizar los recursos públicos al identificar con precisión las zonas con mayor criticidad u riesgo, permitiendo la generación efectividad de capacidad de anticipación y respuesta en la detección de patrones temporales y espaciales de siniestralidad.

ViaVisión no solo atiende necesidades locales, sino que se constituye como un modelo escalable para otros municipios que enfrentan desafíos similares. Su diseño modular y basado en datos facilita su adaptación a diferentes contextos regionales.

3. Objetivos

3.1. Objetivo general

Analizar los datos de siniestralidad vial en Calarcá durante el periodo 2021 – 2025 a partir de los bases de datos abiertos suministrados por MinTic y la Alcaldía de Calarcá para generar una herramienta que permita mapear zonas e identificar los perfiles de riesgo en el municipio.

3.2. Objetivos específicos

- Explorar las bases de datos para identificar cruces e identificadores al consolidar la información.
- Aplicar un Factor de obsolescencia por tipo de vehículo para identificar la antigüedad del vehículo
- Generar perfiles de riesgo vial específicos que permitan focalizar intervenciones, optimizando recursos en infraestructura, control vehicular y cultura ciudadana.
- Mapear con figuras hexagonales las zonas de bajo, medio y alto riesgo de accidentalidad en el municipio.

4. Metodología

4.1. Diseño metodológico



Ilustración 1. Flujo CRISP-DM

4.2. Entendimiento del negocio

La gestión de la seguridad vial en el municipio de Calarcá se desarrolla en un entorno caracterizado por la coexistencia de zonas urbanas densas, áreas rurales y vías de conexión regional. Esta diversidad de escenarios genera dinámicas complejas de movilidad que incrementan la probabilidad de siniestros viales. Las entidades responsables — particularmente la Secretaría de Tránsito y la administración municipal — enfrentan el reto de tomar decisiones informadas en un contexto donde la información disponible es fragmentada, poco integrada y, en muchos casos, analizada de forma manual.

Actualmente, la identificación de puntos críticos, la priorización de intervenciones y la evaluación de tendencias de accidentalidad se realizan mediante procesos que presentan varias limitaciones:

- Falta de integración de fuentes de datos: los registros de accidentalidad, el parque automotor y la información geoespacial se gestionan en sistemas independientes, dificultando su análisis conjunto.

- Baja eficiencia en la interpretación del riesgo: la ausencia de herramientas analíticas automatizadas limita la capacidad para detectar patrones, anticipar escenarios o priorizar zonas de atención.
- Procesos reactivos en lugar de preventivos: las intervenciones suelen responder a hechos ya ocurridos, sin mecanismos que permitan prever riesgos emergentes.
- Escasez de soporte visual y territorial: la información rara vez se presenta en mapas, dashboards o modelos que faciliten la toma de decisiones estratégicas.

4.3. Datos

Además de las bases de datos abiertos suministradas por la página MinTic que corresponden a la Alcaldía de Calarcá, Vehículos matriculados 2020–2022 y Accidentes de tránsito de Calarcá, nos comunicamos directamente con funcionarios de la Alcaldía de Calarcá para solicitar mayor información sobre los registros, con ello obtuvimos datasets complementarios de Accidentalidad 2021–2025 y Puntos críticos de intervención vial.

5. Fases de desarrollo del proyecto

5.1. Limpieza, estandarización y descripción del dataset (ETL)

El proceso ETL (Extract, Transform, Load) inició con la integración de las diferentes fuentes de información disponibles: accidentalidad vial, parque automotor y puntos críticos suministrados por la administración municipal. Se unificaron estructuras, formatos de fecha y hora, y se estandarizaron los nombres de las variables siguiendo una sintaxis uniforme compatible con herramientas de análisis en Python.

Posteriormente, se ejecutaron las etapas de validación, limpieza y transformación, que incluyeron:

- **Depuración de registros incompletos y datos inconsistentes.**
 - **Corrección y normalización de direcciones**, con estandarización de abreviaturas (Cra, Cll, Av) y detección de ruido en texto.
 - **Conversión de tipos de variable** (numéricos, categóricos, fechas).
 - **Homologación de categorías** como gravedad, tipo de accidente, zona, tipo de servicio y tipo de vehículo.
 - **Detección de outliers estadísticos** en edad, número de involucrados y coordenadas.

Como resultado del proceso, se construyó una base de datos estructurada para análisis temporal, espacial y modelado posterior.

Diccionario de transformación del proceso ETL (Input → Output)

A continuación, se describe la transformación aplicada a las columnas originales para generar la estructura final del archivo accidentes.csv y accidentes.json:

Variables originales del dataset fuente

Id; Solo_danos; Herido; Muerto; Direccion; Latitud (sin información), Longitud (sin información); Dia; Mes; A_o; Choque; Atropello; Volcamiento; caida de ocupante; Otro; Rural; Urbana; Masculino; Femenino; <18; 18-30; 31-60; >60; Tipo_vehiculo; Publico; Particular; Oficial; Diplomatico

Variables resultantes y lógica aplicada

- **Id** → **Pi**

Se conserva como identificador principal.

- **Latitud,** **Longitud** → **Latitud,** **Longitud**

Campo sin información que se completa inicialmente a partir de ubicación con direcciones y validación con capa satelital de Google Earth.

- **Dia,** **Mes,** **A_o** → **day,** **month,** **year**

Renombradas para compatibilidad con librerías de análisis y creación de fechas completas.

- **Solo_danos,** **Herido,** **Muerto** → **gravedad**

Consolidación en una única categoría aplicando jerarquía de severidad:

Muerto > Herido > Solo daños.

- **Choque, Atropello, Volcamiento, Caída de ocupante, Otro** → **tipo_accidente**

Se selecciona la categoría activa; si varias son 1, se prioriza según reglas definidas.

- **<18,** **18-30,** **31-60,** **>60** → **categoria_edad**

Se convierte a una etiqueta homogénea, menor, joven, adulto y adulto mayor. Cuando se presentan diferentes rangos de edad en un siniestro se identifica con la categoría mixto.

- **Masculino,** **Femenino** → **cantidad_hombres,** **cantidad_mujeres**

A partir de estos valores se generan las columnas: cantidad_involucrados y genero_involucrados (Masculino, Femenino, Mixto)

- **Publico,** **Particular,** **Oficial,** **Diplomatico** → **Tipo_servicio**

Determinación del servicio predominante.

- **Tipo_vehiculo** → **tipo_vehiculo**

Normalización del nombre del vehículo.

- **Rural,** **Urbana** → **zona_accidente**

Transformación directa.

- **Direccion** → **Direccion_clean** y **direccion_tipo**

Normalización y clasificación del tipo de vía.

- **Nuevas variables derivadas del ETL:**

- fecha (AAA-MM-DD)
- day_name (nombre del día)
- month_name (nombre del mes)
- hora_siniestro (establecida a partir de datos sintéticos considerando franjas horarias típicas de mayor siniestralidad)
- Lngitud y latitud (se completan los campos a partir de las direcciones suministradas y validación de puntos a partir de visualización en Google Earth)

Exploración inicial

Con la base depurada se generaron análisis exploratorios mediante gráficos, entre ellos:

- Accidentes por mes según gravedad
- Accidentes por día de la semana según gravedad

Estos permitieron identificar patrones temporales, tendencias y ciclos que orientaron las siguientes etapas espaciales y de modelamiento.

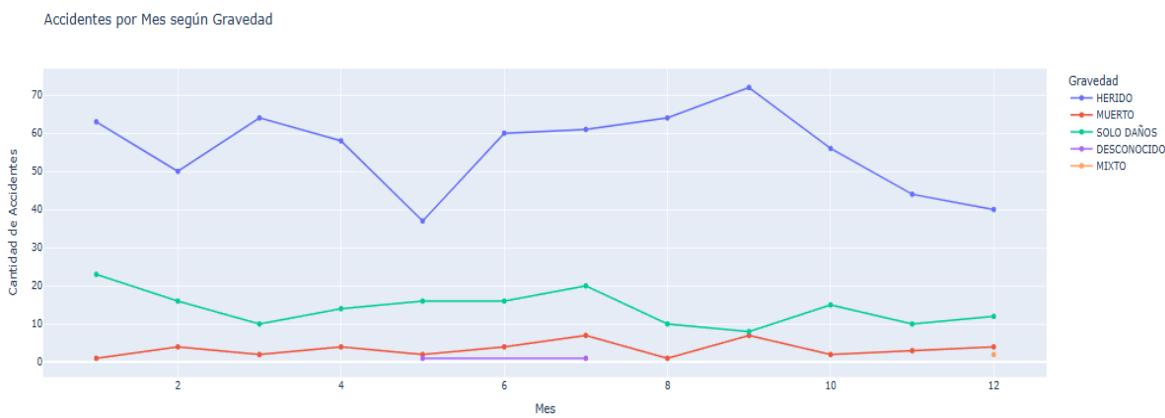


Ilustración 2. Accidentes por Mes según Gravedad

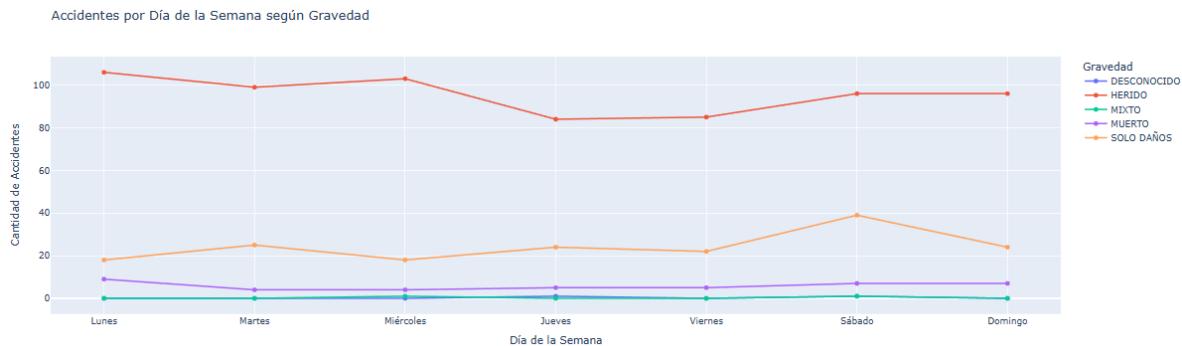


Ilustración 3. Accidentes por día de la semana según gravedad.

5.2. Geocodificación territorial con H3

Para garantizar un análisis espacial robusto se empleó **H3**, un sistema de indexación geoespacial desarrollado por Uber que divide el territorio en hexágonos de igual área. A cada siniestro, vehículo y punto crítico se le asignó un código H3 correspondiente a su ubicación.

Las ventajas del sistema H3 incluyen: Segmentación regular del territorio, reducción de distorsiones geométricas comunes en divisiones poligonales, capacidad de agregar información a múltiples resoluciones, identificación precisa de patrones espaciales de concentración y riesgo. Esta fase permite obtener un modelo territorial uniforme y comparable entre zonas.

5.3. Cálculo de indicadores por hexágono

Tras la asignación H3, se generaron métricas espaciales para evaluar el riesgo vial de cada hexágono, tales como:

- Número total de accidentes por hexágono.
- Índice de severidad (ponderación según heridos y fallecidos).
- Relación con el parque automotor según tipo de vehículo involucrado.
- Frecuencia y recurrencia histórica por períodos.

- Densidad de eventos por área hexagonal.

Estos indicadores componen la base cuantitativa del modelo de riesgo.

5.4. Modelo de score de Riesgo (0 – 100)

A partir de los indicadores previos se implementó una matriz de riesgo que pondera:

- Tipo de vehículo
- Severidad del siniestro.
- Frecuencia relativa.
- Cantidad de vehículos por tipo en el parque automotor.
- Obsolescencia promedio por tipo de vehículo.

El resultado es un score normalizado entre 0 y 100, donde 100 representa la mayor criticidad para los vehículos inscritos en el parque automotor del municipio.

Tipo de Vehículo	Nº Accidentes	Severidad Promedio	Parque Automotor	Obsolescencia Promedio	Índice de Riesgo (0-100)
MOTO	603	2.42	9107	0.65	100.00
BUS	18	2.16	292	0.85	48.25
CAMION	26	1.97	1081	0.69	30.88
CAMIONETA	38	1.88	1357	0.70	26.82
AUTOMOVIL	139	1.73	3965	0.71	26.41
VOLQUETA	4	1.67	134	0.85	14.33
BICICLETA	25	2.10	0	0.00	9.37
TRACTOCAMION	17	1.94	1	0.00	0.00

Ilustración 4. Matriz Riesgo por Tipo de Vehículo

5.5. Integración con puntos críticos oficiales

La plataforma integra la capa de puntos de intervención, suministrada por la Oficina TIC de la Alcaldía de Calarcá. Esto permite: Cruce entre áreas críticas históricas y resultados del modelo, priorización geográfica más precisa, identificación de zonas donde coinciden

evidencia empírica institucional y análisis espacial. Esta fase refuerza la validez territorial de los hallazgos.

5.6. Sistema de recomendaciones por hexágono

Para cada hexágono se generan recomendaciones operativas tales como: Mejoras en señalización, iluminación o demarcación. Ajustes en la operación vehicular (reducción de velocidad, priorización semafórica). Intervenciones de infraestructura. Acciones educativas y preventivas según tipo de riesgo predominante.

Estas recomendaciones se basan en el score, tipo de accidente y características viales de zona.



Ilustración 5. Hexágono del mapa

5.7. Mapa interactivo con Folium

Se desarrolló un mapa interactivo utilizando **Folium**, que permite:

- Visualizar hexágonos coloreados según su nivel de riesgo.
- Explorar puntos críticos e indicadores asociados.
- Examinar la distribución espacial de accidentes, vehículos y líneas de tendencia.

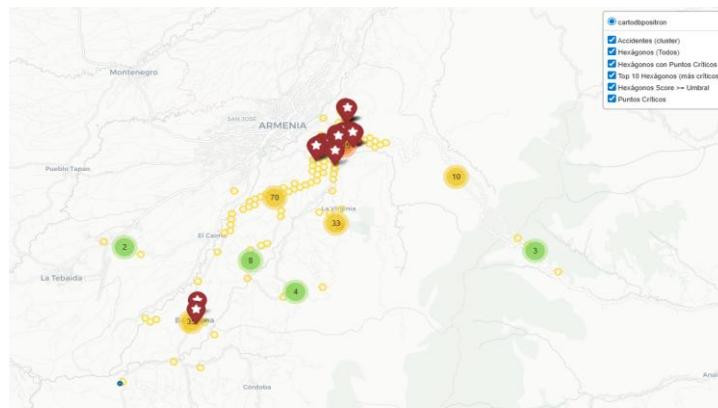


Ilustración 6. Mapa interactivo

5.8. Generación del PDF del perfil de riesgo

Cada hexágono cuenta con un **perfil de riesgo generado automáticamente en PDF**, que incluye:

- Indicadores principales.
- Históricos y tendencias.
- Recomendaciones específicas.
- Relación con puntos críticos o de intervención.

Esto facilita la toma de decisiones y la documentación institucional.



Ilustración 7. Perfil de riesgo por hexágono

6. Limitaciones y recomendaciones

El proceso de depuración, armonización y transformación de datos aplicados al dataset de siniestros viales del municipio de Calarcá permitió construir una base unificada y estructurada para análisis exploratorio, modelamiento estadístico y visualización geoespacial. No obstante, durante el proceso se identificaron diversas limitaciones y aspectos susceptibles de mejora, los cuales se detallan a continuación:

6.1. Limitaciones detectadas en la calidad y disponibilidad de los datos

6.1.1 Ausencia de coordenadas geográficas en múltiples registros

El dataset original no contaba con georreferenciación (latitud y longitud), solo con direcciones o indicaciones del lugar del siniestro, lo cual:

- Impide ubicar espacialmente algunos siniestros.
- Afeta la generación de mapas de calor, clustering espacial y análisis por zonas.
- Limita la precisión del análisis de puntos críticos.

Se realizaron procesos de corrección de direcciones, pero no fue posible asignar coordenadas confiables a todas las entradas ni tan precisas para la zona indicada del suceso.

6.1.2 Falta de información sobre condiciones físicas y operativas de la vía

El dataset no incluye variables clave como:

- Estado de la infraestructura (iluminación, señalización, superficie, peralte, pendientes).
- Condiciones climáticas al momento del siniestro.
- Tipo de control de tránsito en la intersección (semáforo, cebra, glorietas).
- Intensidad vehicular o peatonal.

La ausencia de estos factores afecta la capacidad explicativa de los modelos de riesgo y la posibilidad de generar otros insights valiosos para la toma de decisiones.

6.1.3 Dependencia de la periodicidad y consistencia institucional

Los registros provienen de reportes operativos que pueden presentar:

- Variabilidad en formatos según año y dependencia.
- Posibles errores de digitación o categorización.
- Dificultades para integrar flujos de información heterogéneos.

6.1.4 Inconsistencias en la categorización y valores mixtos por registros

En varias entradas se encontraron combinaciones incongruentes como:

- Múltiples tipos de accidente o edades en una misma fila.
- Registros marcados simultáneamente como "herido" y "solo daños".
- Valores faltantes en dirección o tipo de vehículo.

El proceso ETL aplicó reglas de prioridad y estandarización para corregir estos casos, se imputaron dos registros de un total de 886.

6.2. Recomendaciones para mejorar la captura y utilidad analítica de los datos

6.2.1 Recolección obligatoria de coordenadas geográficas en campo

Se recomienda estandarizar que todo siniestro incluya:

- Latitud y longitud obtenidas mediante GPS o aplicación móvil institucional.
- Validación automática de coordenadas fuera de rango o duplicadas.

Esto permitirá análisis espaciales más robustos y precisos.

6.2.2 Ampliar el registro de características del vehículo y del entorno

Incorporar variables adicionales como:

- Marca, modelo y antigüedad del vehículo (factor de obsolescencia del parque automotor).
- Hora exacta del siniestro.
- Condiciones de iluminación y clima.
- Tipo de control vial en el punto.

Estos atributos aumentan la capacidad predictiva para modelos de riesgo y para priorización de intervenciones.

6.2.3 Estandarización anual del formato institucional

Implementar una plantilla única de captura que reduzca errores de transcripción y mejore la integridad temporal del dataset.

7 Referencias bibliográficas

Ministerio de Tecnologías de la Información y las Comunicaciones – MinTIC. (s. f.). *Vehículos matriculados desde enero de 2020 hasta marzo de 2025* [Conjunto de datos]. Portal de Datos Abiertos Colombia. https://www.datos.gov.co/Transporte/Vehiculos-matriculados-desde-enero-de-2020-hasta-marzo-2025/bj7e-xc9g/about_data

MinTIC. (s. f.). *Sectores críticos – mortalidad 2022* [Conjunto de datos]. Portal de Datos Abiertos Colombia. https://www.datos.gov.co/Transporte/Sectores-críticos-mortalidad-2022/ybqk-8s42/about_data

MinTIC. (s. f.). *Accidentes de tránsito desde marzo 2017 a diciembre 2024* [Conjunto de datos]. Portal de Datos Abiertos Colombia. https://www.datos.gov.co/Transporte/ACCIDENTES-DE-TRANSITO-DESDE-MARZO-2017-A-DICIEMBRE/wacd-xkg8/about_data

OpenAI. (2025). *ChatGPT (noviembre 2025 versión GPT-5.1)* [Large language model]. <https://chat.openai.com/>