

Asignatura	Datos del alumno	Fecha
Diplomado en Analítica de Datos en la Gestión Empresarial	Apellidos: López Fernández Nombre: Douglas	22/10/2025

Fundación Universitaria Internacional de La Rioja

Pregrado en Ingeniería Informática

Diplomado - Analítica de Datos en la Gestión Empresarial



Actividad 2: Planteamiento de Proyecto (Pt 1)

Douglas López Fernández

Laura Marcela Barona Marmolejo

Deisy Johanna Villarraga Cuaycan

Oscar David Bocanegra Capera

GLENN ELMER HERNANDEZ CAMELO

2025

Asignatura	Datos del alumno	Fecha
Diplomado en Analítica de Datos en la Gestión Empresarial	Apellidos: López Fernández Nombre: Douglas	22/10/2025

Título del Proyecto: *Análisis y Priorización de Sectores Críticos de Siniestralidad Vial en Colombia para la Toma de Decisiones.*

Criterio 1: Definición de un problema relevante y específico, con preguntas de investigación claramente planteadas.

Problema

La siniestralidad vial es uno de los principales problemas de salud pública y seguridad en Colombia. A pesar de los esfuerzos nacionales, existen "puntos calientes" o "sectores críticos" en la red vial que concentran una cantidad desproporcionada de siniestros graves, resultando en un alto número de fallecidos.

La gestión de la vasta red de carreteras del país se divide entre múltiples entidades (Agencia Nacional de Infraestructura - ANI, INVIA, Departamentos). Esta división administrativa, combinada con recursos limitados, crea un **problema de asignación y priorización**: las entidades responsables necesitan identificar con precisión no solo *dónde* ocurren los siniestros, sino *qué sectores* representan el riesgo estadístico más significativo y *quién* es el responsable de administrarlos.

La falta de un diagnóstico claro que combine la frecuencia absoluta (Nº de fallecidos) con la significancia estadística (GiZScore) y la ubicación geográfica dificulta la focalización de intervenciones efectivas (mejora de infraestructura, señalización, controles de velocidad, etc.).

Asignatura	Datos del alumno	Fecha
Diplomado en Analítica de Datos en la Gestión Empresarial	Apellidos: López Fernández Nombre: Douglas	22/10/2025

Relevancia e Importancia

Este proyecto es relevante porque busca transformar los datos brutos de siniestralidad en inteligencia accionable. Al identificar y caracterizar los tramos viales más peligrosos, se pueden salvar vidas y reducir los costos económicos y sociales asociados a los accidentes. Proporcionará a las entidades de transporte un insumo directo para priorizar la inversión y las estrategias de seguridad vial en las zonas que más lo necesitan.

Preguntas de Investigación

Para abordar este problema, el proyecto responderá a las siguientes preguntas:

1. **Pregunta de Priorización Geográfica:** ¿Cuáles son los municipios y departamentos que concentran el mayor número de fallecidos en los sectores viales críticos identificados?
2. **Pregunta de Responsabilidad Administrativa:** ¿Qué entidades (ANI, INVIAS, Departamentos, etc.) gestionan los tramos viales con (a) el mayor número acumulado de fallecidos y (b) los índices de criticidad (GiZScore) más altos?
3. **Pregunta de Distribución Estadística:** ¿Cómo se distribuye el índice de criticidad (GiZScore) entre los 1000 sectores? ¿Se trata de un problema homogéneo o existen valores atípicos extremos que requieren atención inmediata?
4. **Pregunta de Visualización Espacial:** Geográficamente, ¿dónde se localizan los principales "puntos calientes" (hotspots) de siniestralidad vial y existe algún patrón de *clusters* o corredores viales evidentes?
5. **Pregunta de Validación Métrica:** ¿Existe una correlación estadística directa entre el número absoluto de fallecidos y el índice de criticidad (GiZScore) en un sector?

Asignatura	Datos del alumno	Fecha
Diplomado en Analítica de Datos en la Gestión Empresarial	Apellidos: López Fernández Nombre: Douglas	22/10/2025

Criterio 2: Diseño de un plan de análisis detallado, con técnicas de análisis de datos apropiadas y una justificación sólida.

Para responder a las preguntas de investigación, se ejecutará el siguiente plan de análisis:

1. Análisis Descriptivo Agregado (Responde a Q1 y Q2)

- **Técnica:** Agregación de datos (Group By) y estadística descriptiva (Suma, Promedio).
- **Descripción:**
 1. Se agruparán los datos por Municipio y Departamento y se sumará la variable Fallecidos para cada uno. Se generará un ranking (Top 10) para identificar las zonas administrativas con mayor carga de mortalidad.
 2. Se agruparán los datos por ENTIDAD (el administrador del tramo). Se calculará la suma total de Fallecidos y el GiZScore promedio para cada entidad.
- **Justificación:** Esta técnica es la más directa para cuantificar el problema y asignar responsabilidades. Permite saber qué municipios necesitan apoyo urgente y qué entidades administran los tramos más letales, respondiendo directamente a las preguntas de "dónde" y "quién".

Asignatura	Datos del alumno	Fecha
Diplomado en Analítica de Datos en la Gestión Empresarial	Apellidos: López Fernández Nombre: Douglas	22/10/2025

- **Visualización Esperada:** Gráficos de barras, como los ya generados:

Figura 1: Top 10 Municipios por Fallecidos.

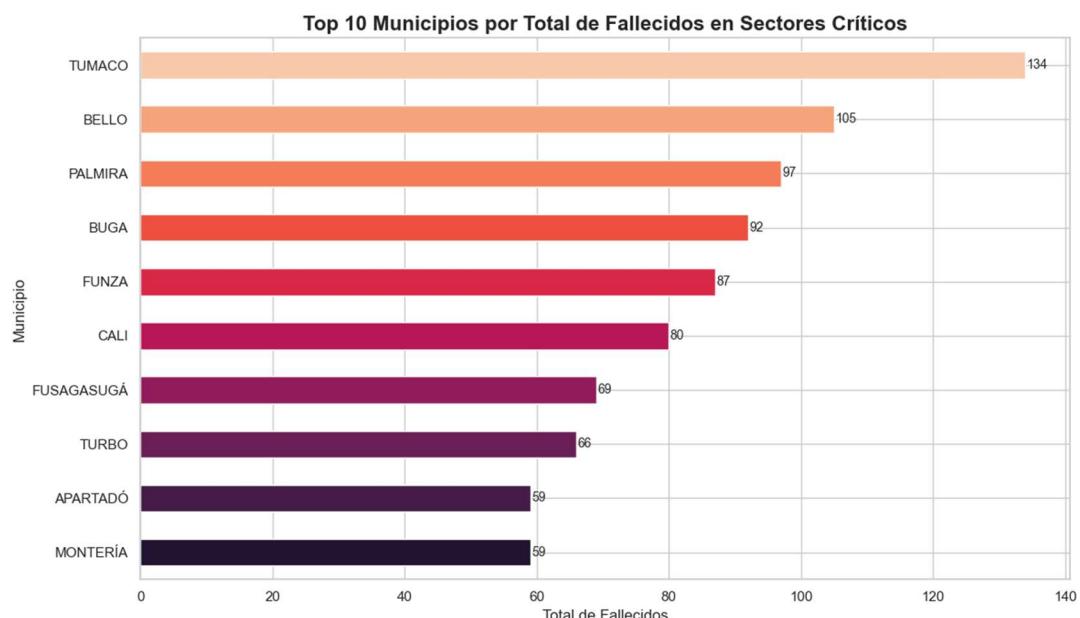
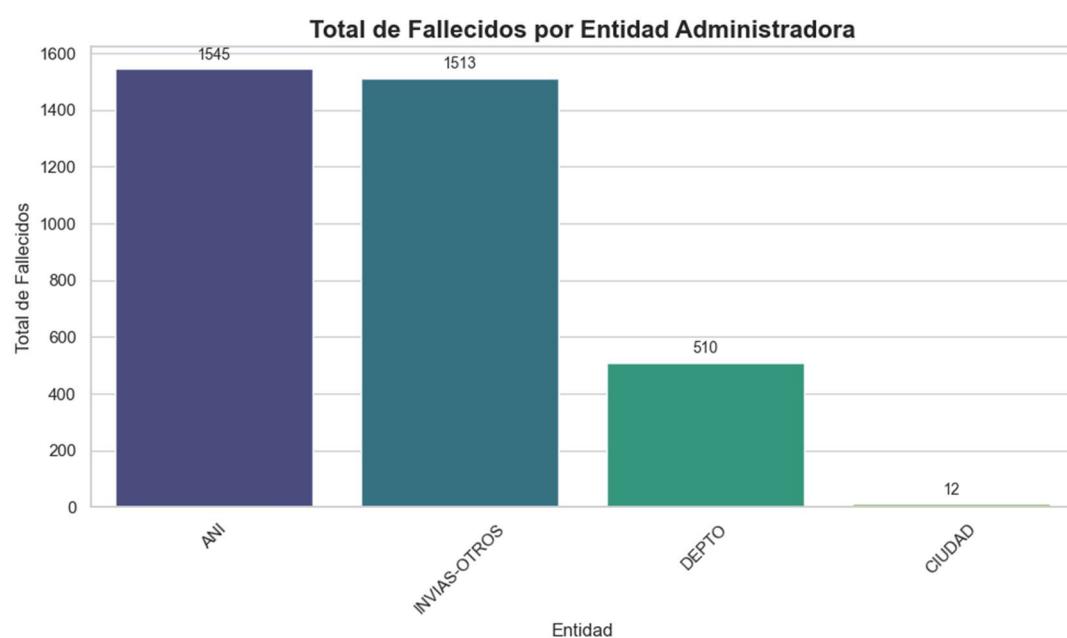


Figura 2: Fallecidos por Entidad Administradora.

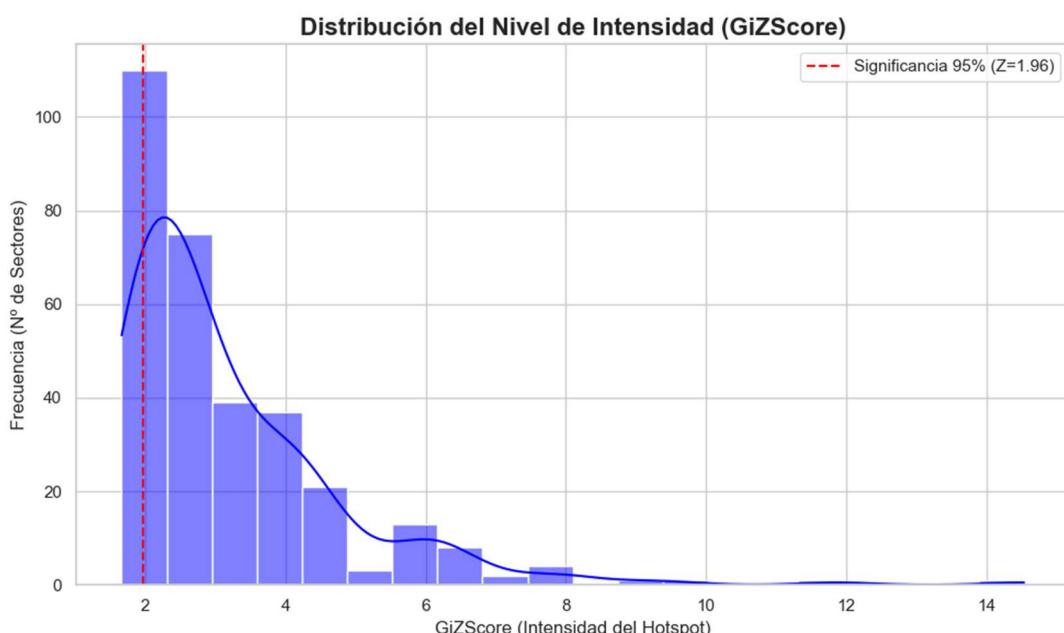


Asignatura	Datos del alumno	Fecha
Diplomado en Analítica de Datos en la Gestión Empresarial	Apellidos: López Fernández Nombre: Douglas	22/10/2025

2. Análisis de Distribución de Frecuencias (Responde a Q3)

- **Técnica:** Histograma y análisis de distribución.
- **Descripción:** Se analizará la variable GiZScore. El GiZScore (un puntaje Z estadístico del análisis Getis-Ord Gi*) mide la intensidad de la agrupación espacial de los siniestros. Se generará un histograma para observar la forma de la distribución, su tendencia central y la presencia de valores atípicos.
- **Justificación:** Un GiZScore alto indica un "punto caliente" estadísticamente significativo. Comprender su distribución es vital. Si la distribución está muy sesgada a la derecha (como sugiere la Figura 3), significa que la mayoría de los sectores tienen una criticidad baja o moderada, pero existe una "larga cola" de sectores *extremadamente* peligrosos que son los que deben priorizarse por encima de todos los demás.
- **Visualización Esperada:** Histograma de la variable GiZScore.

Figura 3: Distribución del Índice de Criticidad (GiZScore).



Asignatura	Datos del alumno	Fecha
Diplomado en Analítica de Datos en la Gestión Empresarial	Apellidos: López Fernández Nombre: Douglas	22/10/2025

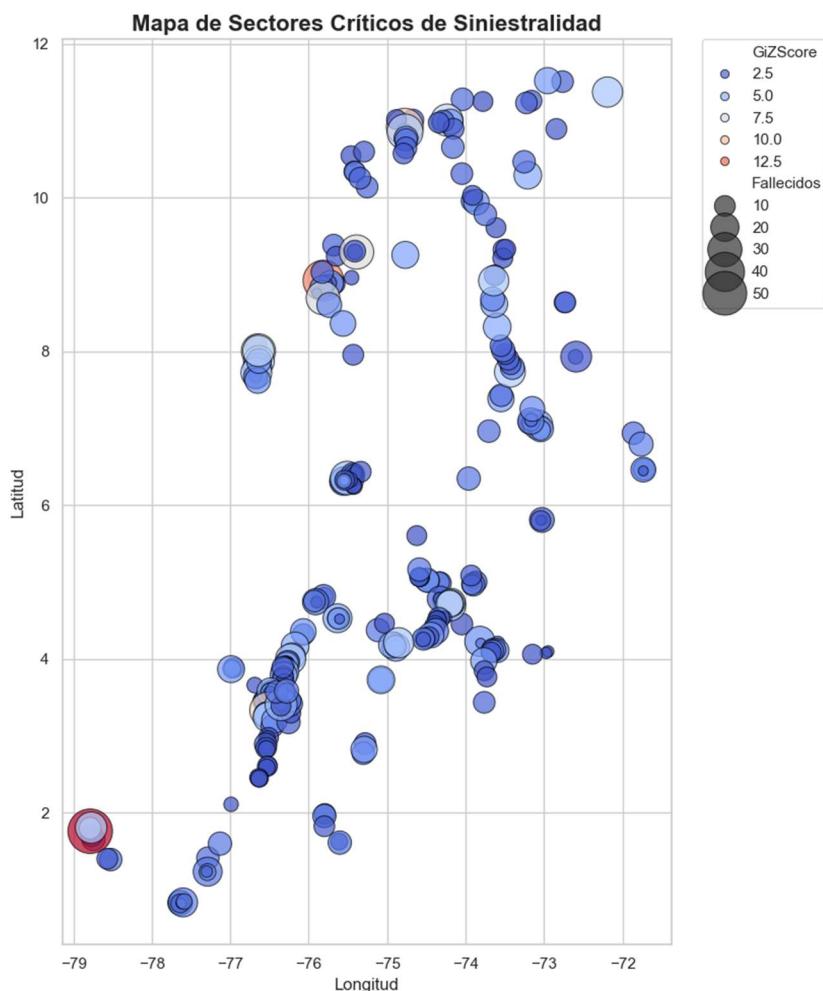
3. Análisis Geoespacial (Responde a Q4)

- **Técnica:** Mapeo de puntos (Point Map) o Mapa de Calor (Heatmap).
- **Descripción:** Utilizando las variables Latitud y Longitud de cada sector, se plotearán los 1000 puntos críticos en un mapa base de Colombia. Los puntos se pueden colorear o dimensionar en función de su GiZScore o del número de Fallecidos para visualizar la intensidad.
- **Justificación:** Los rankings y las tablas ocultan patrones espaciales. Un mapa es la única forma de visualizar *clusters* geográficos y corredores viales completos (por ejemplo, una ruta comercial completa que atraviesa varios municipios). Esto permite pasar de intervenciones puntuales a estrategias regionales integradas.

Asignatura	Datos del alumno	Fecha
Diplomado en Analítica de Datos en la Gestión Empresarial	Apellidos: López Fernández Nombre: Douglas	22/10/2025

- **Visualización Esperada:** Mapa de Colombia con los puntos críticos.

Figura 4: Mapa de Hotspots de Siniestralidad Vial.



Asignatura	Datos del alumno	Fecha
Diplomado en Analítica de Datos en la Gestión Empresarial	Apellidos: López Fernández Nombre: Douglas	22/10/2025

4. Análisis de Correlación (Responde a Q5)

- **Técnica:** Coeficiente de correlación de Pearson y Gráfico de Dispersión (Scatter Plot).
- **Descripción:** Se calculará la correlación entre las variables numéricas Fallecidos y GiZScore. Un gráfico de dispersión (scatter plot) visualizará esta relación (eje X: Fallecidos, eje Y: GiZScore).
- **Justificación:** Esta técnica es crucial para validar si las dos métricas principales del estudio (cantidad de muertes y criticidad estadística) se mueven juntas. Si la correlación es alta, cualquiera de las dos métricas sirve para priorizar. Si es baja, significa que algunos sectores pueden tener pocas muertes, pero ser estadísticamente muy peligrosos (quizás un riesgo nuevo o emergente), requiriendo una estrategia de priorización más matizada.

Criterio 3: Recolección de datos de diversas fuentes y preparación de datos eficiente.

Fuente de Datos Primaria

- **Nombre del Conjunto:** SECTORES CRÍTICOS DE SINIESTRALIDAD VIAL (ID: rs3u-8r4q).
- **Fuente:** Portal de Datos Abiertos de Colombia (datos.gov.co).
- **Acceso:** [SECTORES CRÍTICOS DE SINIESTRALIDAD VIAL](#)
- **Formato:** Archivo CSV
(SECTORES_CRITICOS_DE_SINIESTRALIDAD_VIAL_20251022.csv).

Asignatura	Datos del alumno	Fecha
Diplomado en Analítica de Datos en la Gestión Empresarial	Apellidos: López Fernández Nombre: Douglas	22/10/2025

Descripción del Dataset

El conjunto de datos consta de 1000 filas (representando los 1000 sectores más críticos) y 13 columnas. Las variables clave para el análisis son:

- **ENTIDAD:** (Categórica) Administrador del tramo (ANI, INVIAS, DEPTO, etc.).
- **GiZScore:** (Numérica) Índice estadístico de criticidad (Z-score de Getis-Ord Gi*).
- **Fallecidos:** (Numérica) Número de fallecidos en el sector.
- **Latitud**, **Longitud**: (Numérica) Coordenadas geográficas del sector.
- **Municipio, Departamento:** (Categórica) Ubicación administrativa.
- **divipola:** (Categórica/ID) Código de identificación geográfica.

Proceso de Preparación de Datos (Limpieza)

1. **Carga de Datos:** Importar el archivo CSV a la herramienta de análisis (Python/Pandas, R, Power BI, etc.).
2. **Manejo de Valores Nulos:** Inspeccionar el dataset en busca de nulos. El snippet del CSV muestra valores "<Null>" en la columna ID_MT. Se debe verificar si columnas críticas para el análisis (como Fallecidos, GiZScore, Latitud, Longitud, ENTIDAD) contienen nulos. Si un sector tiene datos nulos en sus coordenadas o en la variable Fallecidos, deberá ser excluido de los análisis específicos que requieran esa variable.
3. **Conversión de Tipos de Datos:** Asegurar que las columnas numéricas (Fallecidos, GiZScore, Latitud, Longitud) sean interpretadas como tipo integer o float y no como string. Las columnas categóricas (ENTIDAD, Municipio, Departamento) serán tratadas como tal.

Asignatura	Datos del alumno	Fecha
Diplomado en Analítica de Datos en la Gestión Empresarial	Apellidos: López Fernández Nombre: Douglas	22/10/2025

4. **Normalización Categórica:** Revisar la columna ENTIDAD para asegurar consistencia. Por ejemplo, si existen "INVIAS" e "Invias", se unificarán a un solo valor ("INVIAS").
5. **Filtrado:** Dado que el dataset ya está pre-filtrado para incluir solo los 1000 sectores críticos, no se requiere un filtrado adicional de filas, a menos que se encuentren valores atípicos irreales (ej. 9999 fallecidos) que indiquen un error de entrada.
6. **Dataset Limpio:** El resultado es un *dataset* limpio y estructurado, listo para aplicar el plan de análisis detallado en el Criterio 2.