

Asignatura	Datos del alumno	Fecha
Diplomado en Analítica de Datos en la Gestión Empresarial	Apellidos:	
	Nombre:	

Fundación Universitaria Internacional de La Rioja

Pregrado en Ingeniería Informática

Diplomado - Analítica de Datos en la Gestión Empresarial



Actividad 2: Planteamiento de Proyecto (Pt 2)

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:	
	Nombre:	

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

1. Introducción

La siniestralidad vial representa uno de los desafíos más significativos para la salud pública y la seguridad en Colombia. El país enfrenta una problemática compleja en la gestión de su vasta red de carreteras, la cual está fragmentada administrativamente entre múltiples entidades como la Agencia Nacional de Infraestructura (ANI), INVIA y las gobernaciones departamentales. Esta división, sumada a la asignación de recursos limitados, genera un problema crítico de priorización

A pesar de que existen datos sobre los siniestros, falta un diagnóstico claro que permita a las entidades identificar con precisión qué sectores viales representan el riesgo estadístico más alto y quién es el responsable de su gestión.

Este proyecto aborda directamente esa necesidad. Su objetivo es transformar los datos brutos de siniestralidad en inteligencia accionable mediante el uso de la analítica de datos. Al aplicar técnicas estadísticas y geoespaciales, el análisis identificará los "puntos calientes" (*hotspots*) de mayor criticidad, responderá a las preguntas de "dónde" ocurre el riesgo y "quién" debe gestionarlo, proporcionando así un insumo directo para que las entidades de transporte puedan focalizar y priorizar sus intervenciones.

2. Revisión de la Literatura

El análisis de la siniestralidad vial ha evolucionado significativamente gracias a las herramientas de analítica y los Sistemas de Información Geográfica (SIG). La literatura en este campo se centra en tres áreas clave relevantes para este proyecto:

Asignatura	Datos del alumno	Fecha
Diplomado en Analítica de Datos en la Gestión Empresarial	Apellidos:	
	Nombre:	

- A. **Análisis de Hotspots (Puntos Calientes):** Tradicionalmente, la identificación de tramos peligrosos se basaba en el conteo absoluto de accidentes. Sin embargo, estudios modernos (como los que emplean el análisis Getis-Ord Gi*, el mismo que utiliza nuestro *dataset*) demuestran la importancia de la significancia estadística. Estos métodos permiten diferenciar si un alto número de siniestros en un sector es un evento aleatorio o parte de un *cluster* espacial estadísticamente significativo, lo cual es vital para una correcta priorización.
- B. **Identificación de Factores de Riesgo:** Más allá de la localización, la analítica de datos busca entender el "porqué". Se utilizan técnicas de correlación y, en estudios más avanzados, de regresión o clasificación para determinar qué factores (diseño de la vía, límites de velocidad, iluminación, etc.) contribuyen a la alta siniestralidad. Aunque nuestro proyecto se centra en la fase descriptiva ("dónde" y "quién"), valida la relación entre métricas clave (Fallecidos vs. GiZScore).
- C. **Modelos de Priorización para la Gestión:** La literatura coincide en que el fin último del análisis de datos en seguridad vial es optimizar la asignación de recursos. Los modelos de priorización, que combinan frecuencia, severidad (fallecidos) y riesgo estadístico, son herramientas esenciales para las entidades gestoras, permitiéndoles justificar intervenciones (mejora de infraestructura, controles) en las zonas de mayor impacto.

Asignatura	Datos del alumno	Fecha
Diplomado en Analítica de Datos en la Gestión Empresarial	Apellidos:	
	Nombre:	

3. Descripción del Problema

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.

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.

Nota: Contenido extraído directamente de la Actividad 2

Asignatura	Datos del alumno	Fecha
Diplomado en Analítica de Datos en la Gestión Empresarial	Apellidos:	
	Nombre:	

Preguntas de Investigación

- 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?

4. Diseño del Plan de Análisis

(Contenido extraído directamente de la Actividad 2, que ya cumple con los requisitos de técnicas, preparación de datos y justificación).

Asignatura	Datos del alumno	Fecha
Diplomado en Analítica de Datos en la Gestión Empresarial	Apellidos: Nombre:	

A. Técnicas de Análisis de Datos

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

- **Técnica:** Agregación de datos (Group By) y estadística descriptiva (Suma, Promedio).
- **Descripción:** Se agruparán los datos por Municipio, Departamento y ENTIDAD. Se sumará la variable Fallecidos y se calculará el GiZScore promedio para cada entidad. Se generarán rankings (Top 10).
- **Justificación:** Es la técnica más directa para cuantificar el problema y asignar responsabilidades, respondiendo a "dónde" y "quién".

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

- **Técnica:** Histograma y análisis de distribución.
- **Descripción:** Se analizará la variable GiZScore. 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 (si está sesgada) es vital para identificar los sectores *extremadamente* peligrosos que deben priorizarse.

Asignatura	Datos del alumno	Fecha
Diplomado en Analítica de Datos en la Gestión Empresarial	Apellidos:	
	Nombre:	

3. Análisis Geoespacial (Responde a Q4)

- **Técnica:** Mapeo de puntos (Point Map) o Mapa de Calor (Heatmap).
- **Descripción:** Utilizando las variables Latitud y Longitud, 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 Fallecidos.
- **Justificación:** Los rankings ocultan patrones espaciales. Un mapa es la única forma de visualizar *clusters* geográficos y corredores viales completos, permitiendo pasar de intervenciones puntuales a estrategias regionales.

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 visualizará esta relación.
- **Justificación:** Es crucial para validar si las dos métricas principales (cantidad de muertes y criticidad estadística) se mueven juntas. Si la correlación es baja, significa que algunos sectores con pocas muertes pueden ser estadísticamente muy peligrosos (riesgo emergente).

Asignatura	Datos del alumno	Fecha
Diplomado en Analítica de Datos en la Gestión Empresarial	Apellidos:	
	Nombre:	

B. Recolección y Preparación de Datos

- **Fuente de Datos:** [SECTORES CRÍTICOS DE SINIESTRALIDAD VIAL \(ID: rs3u-8r4q\) del Portal de Datos Abiertos de Colombia \(datos.gov.co\).](https://datos.gov.co/resource/rs3u-8r4q)
- **Formato:** Archivo CSV.
- **Variables Clave:** *ENTIDAD, GiZScore, Fallecidos, Latitud, Longitud, Municipio, Departamento.*
- **Proceso de Preparación de Datos (Limpieza):**
 1. **Carga de Datos:** Importar el archivo CSV.
 2. **Manejo de Valores Nulos:** Inspeccionar columnas críticas (Fallecidos, GiZScore, Latitud, Longitud, ENTIDAD) en busca de nulos. Las filas con nulos en coordenadas o métricas clave serán excluidas de los análisis específicos que las requieran.
 3. **Conversión de Tipos de Datos:** Asegurar que Fallecidos, GiZScore, Latitud, Longitud sean numéricos (float/integer).
 4. **Normalización Categórica:** Revisar la columna ENTIDAD para unificar valores (ej. "INVIAS" e "Invias" a "INVIAS").
 5. **Filtrado:** El dataset ya está pre-filtrado a 1000 sectores críticos, por lo que no se requiere filtrado adicional salvo para corregir errores de entrada evidentes

Asignatura	Datos del alumno	Fecha
Diplomado en Analítica de Datos en la Gestión Empresarial	Apellidos:	
	Nombre:	

5. Elección de Herramientas de Software

Para ejecutar el plan de análisis descrito, se ha seleccionado el lenguaje de programación **Python** junto con su ecosistema de librerías de ciencia de datos.

Justificación de la Elección:

1. Pandas: Es la herramienta fundamental para la preparación de datos.

Se utilizará para cargar el CSV, realizar la limpieza (manejo de nulos, conversión de tipos, la normalización categórica y, de manera crucial, para ejecutar el **Análisis Descriptivo Agregado** (las operaciones Group By y sum()).

2. Matplotlib y Seaborn: Estas librerías de visualización se emplearán para generar los gráficos requeridos:

- Los gráficos de barras del Top 10 (Q1).
- El histograma de distribución del GiZScore (Q3).
- El gráfico de dispersión (Scatter Plot) para el análisis de correlación (Q5).

3. SciPy / Statsmodels: Se utilizarán para calcular el Coeficiente de correlación de Pearson (Q5).

4. GeoPandas (Recomendado): Aunque se puede usar Matplotlib básico para el mapa de puntos, GeoPandas facilitará enormemente el **Análisis Geoespacial** (Q4), permitiendo plotear los puntos (Latitud, Longitud) sobre un mapa real de los departamentos de Colombia, mejorando la visualización de los *clusters*.

Asignatura	Datos del alumno	Fecha
Diplomado en Analítica de Datos en la Gestión Empresarial	Apellidos:	
	Nombre:	

Esta pila de herramientas (*stack*) ofrece una solución integral que cubre todo el flujo del proyecto, desde la limpieza de datos hasta el análisis estadístico y la visualización geoespacial, en un solo entorno.

Plan para la Presentación Breve del Proyecto

Para cumplir con el requisito de "Presentación del proyecto", esta es la estructura recomendada para una presentación en PowerPoint con los siguientes puntos:

Diapositiva 1: Título, Problema y Objetivo

- **Título:** Análisis y Priorización de Sectores Críticos de Siniestralidad Vial en Colombia.
- **El Problema:** Colombia tiene "puntos calientes" de siniestralidad. Los recursos son limitados y la gestión está dividida (ANI, INVIAS, etc.).
- **El Objetivo:** Usar analítica de datos para identificar los tramos más peligrosos y quién es responsable, para focalizar la inversión y salvar vidas.

Diapositiva 2: Preguntas de Investigación y Datos

- **Preguntas Clave:**
 - ¿*Dónde* están los municipios con más fallecidos?
 - ¿*Quién* (ANI, INVIAS) administra los tramos más letales?
 - ¿Existen sectores *extremadamente* peligrosos (GiZScore)?
 - ¿Cómo se ven estos *hotspots* en el mapa?
- **Datos:** Dataset de 1000 Sectores Críticos de Datos Abiertos Colombia.

Asignatura	Datos del alumno	Fecha
Diplomado en Analítica de Datos en la Gestión Empresarial	Apellidos:	
	Nombre:	

Diapositiva 3: Métodos (Plan de Análisis)

- Explicar brevemente los 4 análisis que se realizarán:
 1. **Análisis Descriptivo:** Rankings por Municipio y Entidad.
 2. **Análisis de Distribución:** Histograma del GiZScore.
 3. **Análisis Geoespacial:** Mapa de *hotspots*.
 4. **Análisis de Correlación:** (Fallecidos vs. Criticidad).

Diapositiva 4: Relevancia y Siguientes Pasos

- **Relevancia:** Este proyecto no es solo académico; entrega "inteligencia accionable" para priorizar intervenciones.
- **Siguientes Pasos:** Implementar el análisis en Python, generar los reportes y visualizaciones finales.
- **Abrir a Retroalimentación:** (Preguntar al grupo por sugerencias sobre el plan).

Conclusión y Enlace a la Visualización Interactiva

Como producto final del proyecto, la inteligencia accionable generada a partir del análisis de datos se consolida en un tablero de control interactivo. Este recurso permite a las entidades gestoras de la infraestructura vial de Colombia explorar y filtrar los resultados de las preguntas de investigación (Q1, Q2, Q4 y Q5) de manera visual y dinámica, facilitando la toma de decisiones basada en la evidencia geoespacial y estadística.

Tablero de Control y Resultados (Looker Studio):

[Tablero de Análisis y Priorización de Sectores Críticos de Siniestralidad Vial](#)