**Fundación Universitaria Internacional de La Rioja**

Pregrado en Ingeniería Informática

Diplomado - Analítica de Datos en la Gestión Empresarial



**Actividad 2: Planteamiento de Proyecto (Parte 1)**

Douglas López Fernández  
Laura Marcela Barona Marmolejo  
Deisy Johanna Villarraga Cuaycan  
Oscar David Bocanegra Capera

**GLENN ELMER HERNANDEZ CAMELO**

**2025**

**Propuesta de Proyecto de Analítica de Datos**

**Título del Proyecto:** *Análisis de la Siniestralidad Vial en Cali para la Optimización de Estrategias de Seguridad Ciudadana (2022-2024).*

1. **Definición del Problema y Preguntas de Investigación**

**(Criterio 1: Define un Problema Relevante y Específico)**

**Problema Relevante y Específico:** La ciudad de Cali enfrenta un desafío constante en materia de seguridad vial. Los accidentes de tránsito no solo representan un alto costo económico en daños materiales y atención médica, sino que también tienen un impacto social devastador debido a las lesiones y pérdidas de vidas humanas. A pesar de los esfuerzos de las autoridades, a menudo las estrategias de prevención no están suficientemente focalizadas porque carecen de un análisis profundo y detallado de los datos. Este proyecto busca utilizar la analítica de datos para transformar la información cruda de los siniestros en conocimiento accionable que permita a la Secretaría de Movilidad de Cali diseñar e implementar intervenciones más eficientes y localizadas.

**Preguntas de Investigación Clave:**

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

1. **Pregunta Principal:**

¿Cuáles son los patrones y factores de mayor incidencia en los accidentes de tránsito en Cali que pueden ser utilizados para proponer medidas de prevención focalizadas?

1. **Preguntas Secundarias:**

* **Geográficas:** ¿En qué comunas, barrios y vías específicas de Cali se concentra la mayor cantidad de accidentes? ¿Existen "puntos calientes" de siniestralidad?
* **Temporales:** ¿Qué días de la semana y franjas horarias registran la mayor frecuencia de accidentes? ¿Existen patrones estacionales a lo largo del año?
* **Causales:** ¿Qué clase de accidente (choque, atropello, volcamiento) es más común y cuál está asociado a una mayor gravedad (heridos o fallecidos)
* **Demográficas**: ¿Qué tipo de vehículos (motocicletas, automóviles, transporte público) están involucrados con más frecuencia en los siniestros viales?

1. **Recolección y Preparación de Datos**

**(Criterio 3: Recopila Datos y Realiza una Preparación Eficiente)**

**Fuente de Datos:**

El conjunto de datos principal para este análisis se obtendrá de la plataforma de Datos Abiertos de Colombia. Específicamente, se utilizará el dataset:

* **Nombre del Dataset:** "Siniestros viales en el Municipio de Santiago de Cali".
* **Entidad:** Alcaldía de Santiago de Cali.
* **Enlace (Ejemplo):** https://www.datos.gov.co/d/y34d-e4sx (Este es un ejemplo, se debe buscar el dataset más actualizado).
* **Acceso:** Los datos son públicos y se pueden descargar en formato CSV.

**Proceso de Preparación de Datos (ETL - Extract, Transform, Load):**

1. **Extracción:** Descarga del archivo .csv desde el portal de datos abiertos.
2. **Transformación y Limpieza:**

* **Manejo de Valores Faltantes:** Inspeccionar columnas como barrio o diseño\_via. Si los datos faltantes son pocos, se podrían eliminar esas filas; si son muchos, se podrían agrupar bajo una categoría "No especificado".
* **Corrección de Tipos de Datos:** Asegurar que las columnas de fecha (FECHA\_ACCIDENTE) sean reconocidas como formato datetime, y las coordenadas (LATITUD, LONGITUD) como numéricas.
* **Estandarización:** Unificar valores en columnas categóricas. Por ejemplo, si aparece "Moto" y "MOTOCICLETA", se unifican en una sola categoría.
* **Creación de Nuevas Variables (Feature Engineering):** A partir de la columna FECHA\_ACCIDENTE, se crearán nuevas columnas que serán muy útiles para el análisis:
* Dia\_Semana (Lunes, Martes, etc.)
* Franja\_Horaria (Madrugada, Mañana, Tarde, Noche)
* Mes\_Año

1. **Carga:** El conjunto de datos limpio y transformado se cargará en la herramienta de análisis seleccionada (por ejemplo, en un DataFrame de Pandas para análisis con Python, o directamente en Power BI para visualización).
2. **Plan de Análisis Detallado**

**(Criterio 2: Diseña un Plan de Análisis Detallado con Técnicas Apropiadas)**

Se utilizará una combinación de técnicas de análisis descriptivo y geoespacial para responder a las preguntas de investigación.

**Fase 1: Análisis Descriptivo Univariado**

Este análisis inicial permitirá entender la distribución de cada variable de interés de forma individual.

* **Técnica:** Cálculo de frecuencias y porcentajes.
* **Justificación:** Es fundamental para obtener un panorama general de la situación. Permite identificar las categorías más comunes en variables clave.
* **Visualización Propuesta:**
* **Gráfico de Barras - Accidentes por Clase:** Para ver qué tipo de accidente (ej. choque, atropello) es más frecuente.
* **Gráfico de Barras - Accidentes por Gravedad:** Para comparar la cantidad de accidentes con solo daños, con heridos o con fallecidos.

**Fase 2: Análisis Descriptivo Bivariado y Temporal** Aquí se cruzarán variables para encontrar relaciones y patrones.

* **Técnica:** Tablas de contingencia y gráficos de series de tiempo.
* **Justificación:** Permite responder preguntas sobre cuándo ocurren los accidentes y si existe una relación entre el tiempo y la gravedad.
* **Visualización Propuesta:**
* **Gráfico de Barras - Accidentes por Día de la Semana:** Para identificar los días más críticos. Es muy probable que los fines de semana (viernes, sábado) muestren picos.
* **Gráfico de Líneas - Accidentes por Hora del Día:** Para visualizar las horas con mayor concentración de siniestros.

**Fase 3: Análisis Geoespacial**

El componente geográfico es crucial en este problema para focalizar las intervenciones.

* **Técnica:** Mapeo de coordenadas (latitud y longitud).
* **Justificación:** Una tabla de datos no permite ver la concentración espacial. Un mapa es la forma más eficaz de identificar "puntos calientes" de alta siniestralidad.
* **Visualización Propuesta:**
* **Mapa de Calor (Heatmap) de Cali:** Usando las coordenadas de cada accidente para visualizar las zonas de mayor concentración. Las áreas en rojo intenso indicarán dónde se deben priorizar las medidas de seguridad (ej. más señalización, reductores de velocidad, presencia de agentes de tránsito).
* **Mapa de Puntos por Comuna:** Un mapa coroplético donde cada comuna se colorea según la cantidad de accidentes, permitiendo una comparación rápida entre zonas administrativas.

**Herramientas Propuestas:**

* **Análisis y Manipulación de Datos:** **Python** con las librerías **Pandas** (para la manipulación de datos) y **Matplotlib/Seaborn** (para gráficos iniciales).
* **Visualización y Dashboard Interactivo:** **Power BI** o **Tableau**. Estas herramientas son ideales para crear un informe interactivo donde un usuario (ej. un funcionario de la alcaldía) pueda filtrar los datos por fecha, comuna o tipo de accidente y ver cómo cambian los gráficos y el mapa de calor en tiempo real.

1. **Análisis Profundo: Hora, Vehículo y Gravedad**

Estos exploran nuevas dimensiones de los datos: ***la distribución horaria, el tipo de vehículo involucrado y la relación entre el vehículo y la gravedad del siniestro***. Estas visualizaciones te permitirán formular recomendaciones mucho más específicas y detalladas.

**Análisis Profundo: Hora, Vehículo y Gravedad**

**1. Siniestros Viales por Hora del Día**

Este gráfico responde a la pregunta: ***"¿En qué franjas horarias se concentran los siniestros?"***

* **Observación:** El gráfico muestra dos picos principales de siniestralidad a lo largo del día. El primero es un pico matutino entre las 7 a. m. y 8 a. m., coincidiendo con el inicio de la jornada laboral y académica. El segundo, y más pronunciado, es un pico vespertino que se extiende desde las 5 p. m. hasta las 7 p. m., correspondiente a la hora pico de regreso a casa. También se observa una actividad considerable al mediodía.
* **Interpretación para el proyecto:** La congestión vehicular durante las horas pico es un factor determinante. Las autoridades deberían enfocar la gestión del tráfico y la presencia de agentes de movilidad específicamente durante estas ventanas de tiempo críticas. Medidas como la sincronización de semáforos y la promoción de horarios laborales flexibles podrían ayudar a mitigar la concentración de accidentes en estos periodos.

**2. Vehículos Más Involucrados en Siniestros**

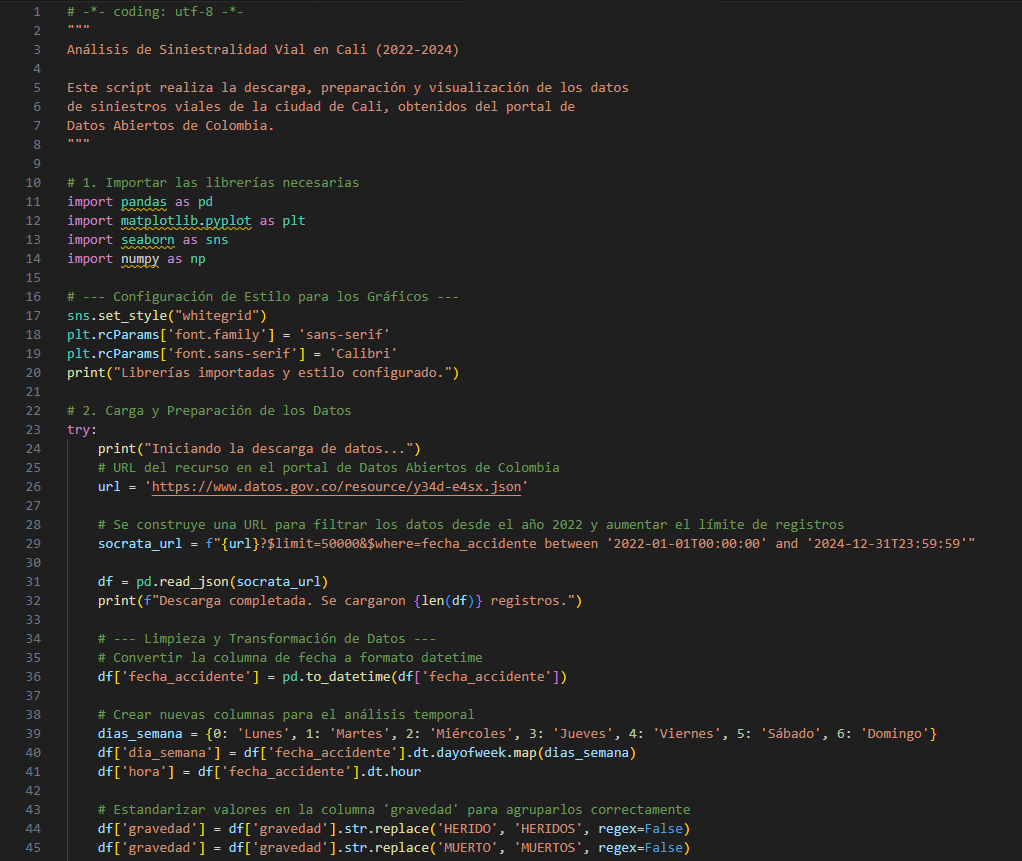
Este gráfico responde a la pregunta: **"*¿Qué tipo de vehículos están más comúnmente involucrados en los accidentes?"***

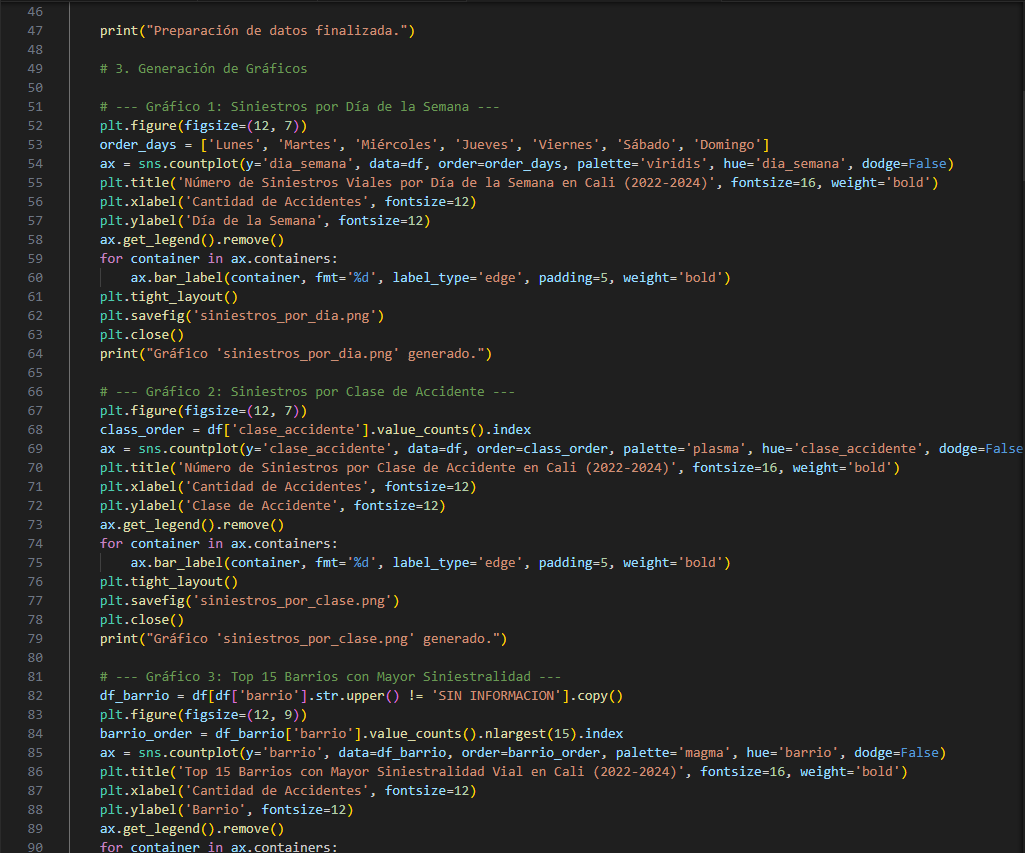
* **Observación:** La **motocicleta** es, con una diferencia muy significativa, el tipo de vehículo más involucrado en siniestros viales en Cali. Le sigue el automóvil particular. El transporte de carga y los autobuses tienen una participación mucho menor en comparación.
* **Interpretación para el proyecto:** Este hallazgo es crucial. Demuestra que cualquier estrategia de seguridad vial en Cali debe tener un ***enfoque prioritario en los motociclistas***. Esto no solo implica fiscalización (controles de velocidad, uso del casco), sino también campañas de educación vial dirigidas a este grupo, revisión de la infraestructura para hacerla más segura para vehículos de dos ruedas y fomento de la conducción defensiva tanto para motociclistas como para otros conductores.

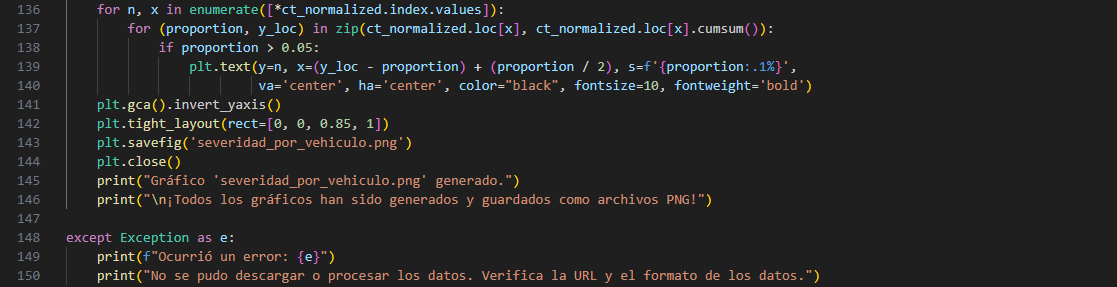
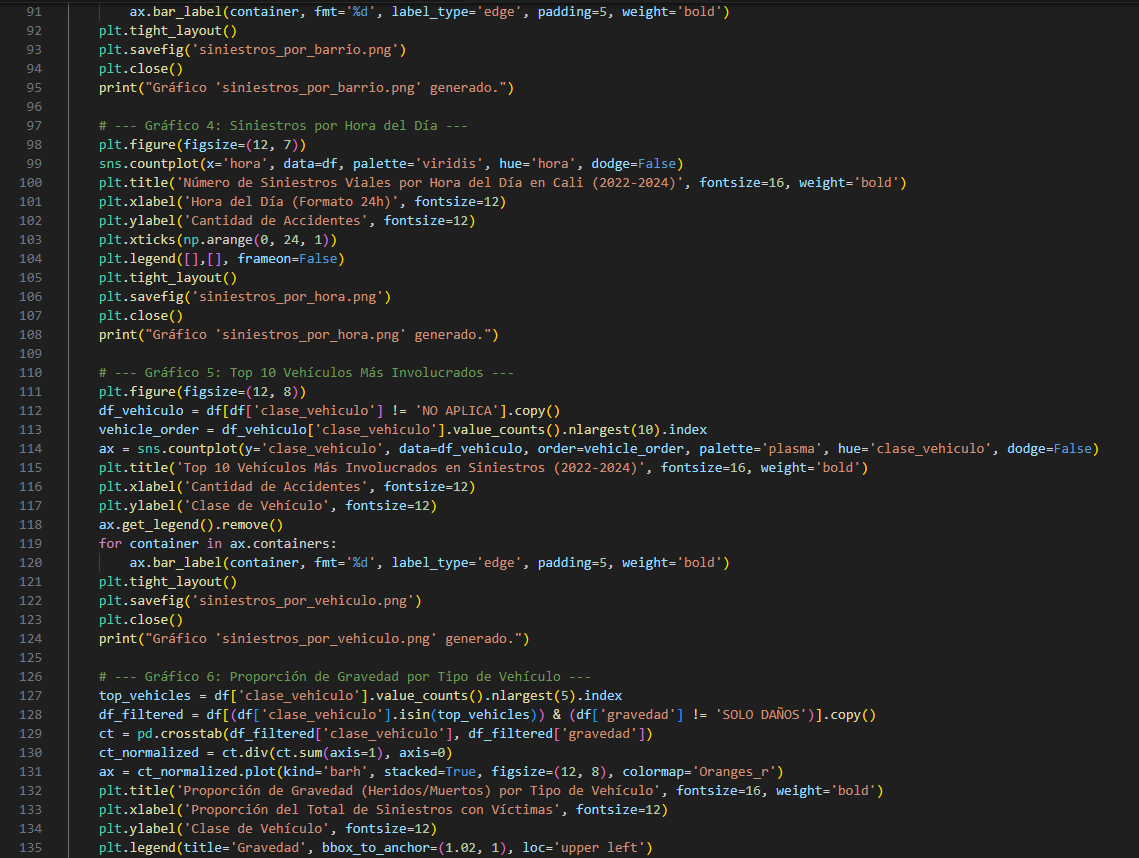
**3. Relación entre Tipo de Vehículo y Gravedad del Siniestro**

Este gráfico aborda una pregunta más compleja: **"*¿Existe una relación entre el tipo de vehículo y la probabilidad de que un accidente resulte en heridos o fallecidos?"***

* **Observación:** Este gráfico de barras 100% apiladas revela una realidad alarmante. Mientras que en los accidentes que involucran solo automóviles un alto porcentaje resulta en "Solo Daños", en los siniestros que involucran motocicletas y bicicletas, la proporción de accidentes con ***"Heridos" o "Muertos" es drásticamente mayor***. Los motociclistas y ciclistas son, por mucho, los actores viales más vulnerables.
* **Interpretación para el proyecto:** Este es el argumento más poderoso para justificar políticas de protección a los usuarios vulnerables. La data demuestra que, aunque un choque entre dos automóviles es común, un siniestro que involucra una motocicleta tiene una probabilidad mucho más alta de terminar en tragedia. Las recomendaciones deben incluir la mejora de la infraestructura ciclista, la implementación de límites de velocidad más estrictos en zonas con alto flujo de motocicletas y campañas de concientización sobre la "fragilidad" de estos actores en la vía.

1. **Código en Python para el Análisis y Visualización de Datos**





**Bibliografía y Referencias:**

Esta bibliografía está dividida en: la fuente de datos principal, libros y recursos metodológicos clave, y documentación de las herramientas sugeridas.

1. **Fuente de Datos Principal**

* **Alcaldía de Santiago de Cali. (2024). *Siniestros viales en el Municipio de Santiago de Cali*. Datos Abiertos Colombia.**
* **Descripción:** Este es el conjunto de datos principal que se utilizará para el análisis. Contiene registros detallados de los accidentes de tránsito en Cali, incluyendo variables como fecha, hora, ubicación geográfica (latitud y longitud), clase de accidente, gravedad y tipo de vehículos involucrados.
* **Enlace de acceso:** <https://www.datos.gov.co/Transporte/Siniestros-viales-en-el-Municipio-de-Santiago-de/y34d-e4sx>

1. **Libros y Recursos Metodológicos sobre Analítica y Visualización**

* **Cairo, A. (2016). *The Truthful Art: Data, Charts, and Maps for Communication*. New Riders.**
* **Descripción:** Este libro es una referencia esencial sobre cómo diseñar visualizaciones de datos que sean precisas, claras y éticas. Sus principios son clave para justificar la elección de los gráficos propuestos (barras, líneas, mapas de calor) y asegurar que comuniquen los hallazgos de manera efectiva.
* **Knaflic, C. N. (2015). *Storytelling with Data: A Data Visualization Guide for Business Professionals*. Wiley.**
* **Descripción:** Aporta el marco metodológico para no solo mostrar datos, sino contar una historia convincente con ellos. Es fundamental para estructurar la presentación final del proyecto, enfocándose en comunicar los *insights* más relevantes para la toma de decisiones en seguridad vial.
* **McKinney, W. (2017). *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. O'Reilly Media**.
* **Descripción:** Es la guía de referencia para el uso de la librería Pandas en Python. Justifica las técnicas de preparación de datos mencionadas en el plan de análisis, como la limpieza, transformación y creación de nuevas variables (feature engineering), que son pasos críticos para asegurar la calidad del análisis.
* **Provost, F., & Fawcett, T. (2013). *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*. O'Reilly Media.**
* **Descripción:** Ofrece un marco conceptual para abordar problemas de negocio (en este caso, un problema de política pública) desde una perspectiva de ciencia de datos. Ayuda a justificar la estructura del proyecto, desde la definición del problema hasta el plan de análisis, asegurando que el enfoque esté orientado a generar valor y soluciones prácticas.

1. **Documentación de Herramientas de Software**
   * **Microsoft Corporation. (2024). *Documentación de Power BI*. Microsoft Docs.**

* **Descripción:** Recurso oficial para la implementación de las visualizaciones y dashboards interactivos. Se utilizará como guía técnica para la creación de los mapas de calor y los gráficos dinámicos que permitirán a los usuarios explorar los datos de siniestralidad.
  + **Enlace de acceso:** [https://docs.microsoft.com/es-es/power-bi/](https://www.google.com/search?q=https://docs.microsoft.com/es-es/power-bi/)
* **Tableau Software. (2024). *Soporte y Recursos de Tableau*. Tableau.**
  + **Descripción:** Alternativa a Power BI, su documentación oficial es la fuente principal para justificar las capacidades de análisis geoespacial y la creación de informes interactivos.
  + **Enlace de acceso:** <https://www.tableau.com/es-es/support>