

APRENDIZAJE 1 - CREACIÓN DE UNA BASE DE DATOS ANALÍTICA

Carlos Arturo Gómez Jiménez

Ingeniería de Software y Datos
Facultad de Ingeniería y Ciencias Agropecuarias
Institución Universitaria Digital de Antioquia

Big Data - 020109

Ing. Andrés Felipe Callejas

18 de mayo del 2025



TABLA DE CONTENIDO

| | |
|---------------------------------------------------------------------------|----------|
| 1. INTRODUCCIÓN | 3 |
| 2. DESCRIPCIÓN DEL PROBLEMA | 3 |
| 3. OBJETIVOS..... | 3 |
| OBJETIVO GENERAL..... | 3 |
| OBJETIVO ESPECÍFICO..... | 3 |
| 4. DESCRIPCIÓN DE LOS DATOS DISPONIBLES..... | 4 |
| 5. SOLUCIÓN PROPUESTA (ELECCIÓN DEL SGBD Y ESQUEMA DISEÑADO) | 4 |
| CAPTURA DE DATOS | 5 |
| ALMACENAMIENTO Y PROCESAMIENTO | 5 |
| ANÁLITICA DE DATOS..... | 5 |
| 6. METODOLOGÍA EMPLEADA | 6 |
| 7. RESULTADOS Y CONCLUSIONES | 7 |
| 8. ANEXO..... | 7 |
| 9. BIBLIOGRAFÍA..... | 7 |

1. INTRODUCCIÓN

En Colombia, la presencia de minas antipersonales y áreas con sospecha de contaminación por artefactos explosivos representa un desafío humanitario, logístico y de planificación territorial. Esta situación afecta la movilidad de comunidades rurales, limita el desarrollo de proyectos productivos y constituye una amenaza constante para la vida. Frente a este contexto, el acceso a datos abiertos y estructurados permite desarrollar herramientas de análisis y gestión que fortalezcan la toma de decisiones y la acción estatal. Este informe presenta el diseño, implementación y resultados de un sistema de gestión de base de datos (SGBD) que permite estructurar, almacenar y consultar eficientemente información pública sobre eventos mineros reportados en el país.

2. DESCRIPCIÓN DEL PROBLEMA

Colombia ha sido históricamente uno de los países más afectados por minas antipersonales en el mundo. A pesar de los esfuerzos de desminado humanitario, persisten regiones donde los eventos de sospecha de campo minado, accidentes por artefactos explosivos improvisados y operaciones de desminado militar se siguen reportando. La fragmentación de la información, la ausencia de bases de datos organizadas y la baja accesibilidad a fuentes actualizadas limitan el análisis sistemático y dificultan la identificación de patrones, priorización de intervenciones y evaluación de riesgos.

3. OBJETIVOS

Objetivo general

Diseñar e implementar un sistema de gestión de base de datos que permita almacenar, consultar y analizar datos públicos sobre eventos relacionados con minas en Colombia.

Objetivo Específico

- Integrar datos abiertos del portal oficial del gobierno colombiano en un sistema relacional.
- Normalizar y estructurar los datos, incorporando campos de auditoría.
- Diseñar un esquema eficiente en SQLite para almacenamiento local.
- Exportar los datos en formatos interoperables (CSV y SQLite).
- Desarrollar una metodología replicable para futuras actualizaciones o migraciones a otros SGBD.

4. DESCRIPCIÓN DE LOS DATOS DISPONIBLES

La fuente principal utilizada en este proyecto es el conjunto de datos abiertos titulado “Eventos Minas Antipersonal en Colombia”, publicado por la Oficina del Alto Comisionado para la Paz y disponible a través del portal oficial del Estado colombiano datos.gov.co. Este recurso contiene registros históricos de eventos relacionados con minas antipersonal (MAP), municiones sin explosionar (MUSE) y actividades de desminado militar en diferentes regiones del país.

A continuación, se presenta una descripción estructurada de los campos que componen esta base de datos:

| Campo (API) | Descripción | Tipo de dato |
|--------------------------|-------------------------------------------------------------------------------------------|---------------|
| tipo_evento | Clasificación del tipo de evento: Sospecha de campo minado, desminado militar o incidente | Texto |
| departamento | Nombre del departamento donde ocurrió el evento | Texto |
| codigo_dane_departamento | Código DANE del departamento | Numérico |
| municipio | Nombre del municipio donde se reportó el evento | Texto |
| codigo_dane_municipio | Código DANE del municipio | Numérico |
| a_o | Año en el que ocurrió el evento | Numérico |
| mes | Mes en el que ocurrió el evento | Numérico |
| evento | Clasificación del suceso como accidente o incidente | Texto |
| tipo_lugar | Tipo de lugar geográfico (por ejemplo: finca, camino, escuela, etc.) | Texto |
| tipo_area | Área donde ocurrió el evento: rural o urbana | Texto |
| latitud_cabecera | Latitud de la cabecera municipal correspondiente al evento | Numérico |
| longitud_cabecera | Longitud de la cabecera municipal correspondiente al evento | Numérico |
| sitio | Nombre común del lugar exacto donde ocurrió el evento | Texto |
| geocoded_column | Representación geoespacial en formato Point (coordenadas geográficas) | Punto (texto) |

5. SOLUCIÓN PROPUESTA (ELECCIÓN DEL SGBD Y ESQUEMA DISEÑADO)

Para la implementación de la solución se propuso un flujo de trabajo compuesto por tres fases secuenciales: captura de datos, almacenamiento y procesamiento, y analítica de datos, como se muestra en el siguiente esquema funcional:

Captura de datos

La información se obtiene desde el conjunto de datos oficial Eventos Minas Antipersonal en Colombia, a través de su interfaz API proporcionada por el portal datos.gov.co. Esta etapa corresponde a la descarga y transformación inicial, convirtiendo el JSON en un DataFrame manipulable con Python.

Almacenamiento y procesamiento

Se eligió SQLite como sistema de gestión de base de datos (SGBD) debido a su portabilidad, facilidad de integración con Python y capacidad para proyectos de escala local sin requerimientos de servidor. En esta etapa se estructuró un esquema relacional con tres entidades principales:

TablaCruda: contiene los campos originales del conjunto de datos, con tipos de datos explícitos y trazabilidad temporal (`fecha_creacion`, `fecha_update`).

DatosProcesados: tabla que proyecta la generación futura de indicadores clave de desempeño (KPIs) como densidad de eventos, frecuencia relativa, índice de riesgo, entre otros.

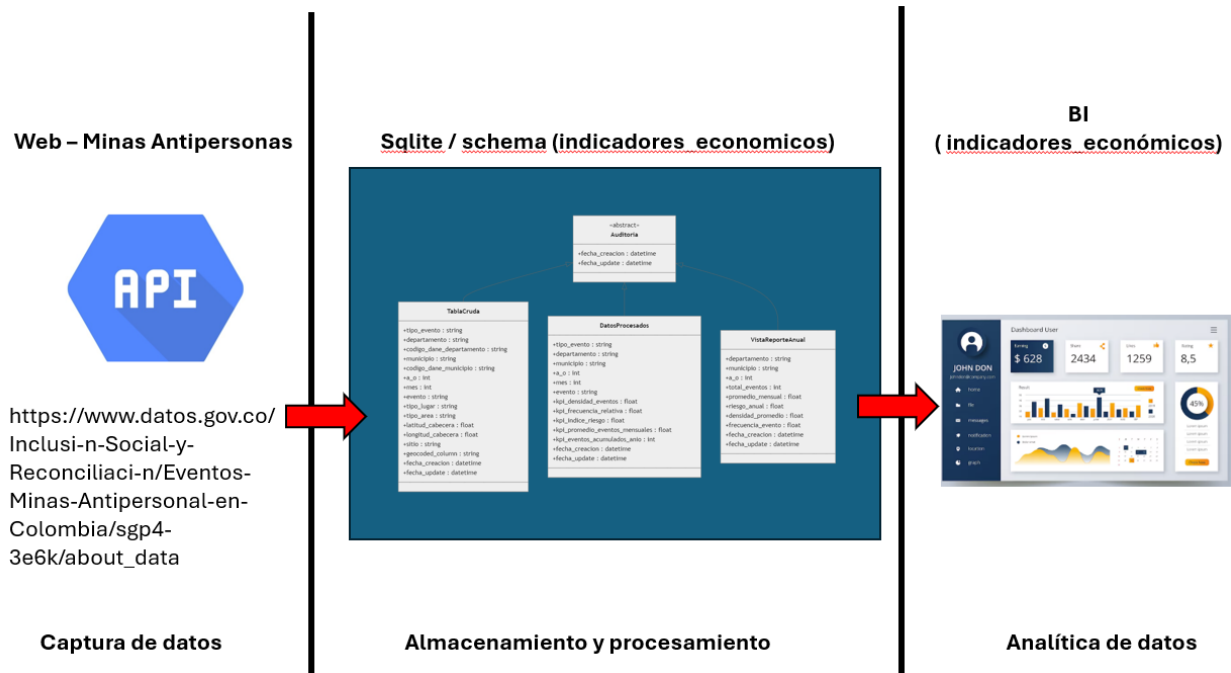
VistaReporteAnual: vista o tabla de consolidación para reportes anuales agregados, orientada al análisis por municipio, año y clasificación del evento.

El diseño integra una clase abstracta común denominada Auditoria, que estandariza la trazabilidad temporal para todas las tablas mediante los campos `fecha_creacion` y `fecha_update`.

Analítica de datos

Aunque no se implementó una solución completa de inteligencia de negocios (BI), el esquema quedó preparado para futuras integraciones con herramientas como Power BI, Tableau o dashboards desarrollados en Python. La base de datos relacional puede servir como backend para reportes interactivos que permitan visualizar el comportamiento espacial y temporal de los eventos relacionados con minas antipersonal en el país.

Esta solución modular permite escalar el sistema hacia estructuras más robustas, incorporar fuentes adicionales o enlazarse con plataformas de análisis geoespacial.



6. METODOLOGÍA EMPLEADA

La metodología seguida se estructuró conforme a los pasos definidos en la evidencia de aprendizaje:

- Identificación de la problemática: Se seleccionó como caso de análisis la situación de las minas antipersonal en Colombia, a partir de datos abiertos provistos por el gobierno nacional.
- Recolección de información relevante: Se identificó como fuente oficial el conjunto de datos “Eventos Minas Antipersonal en Colombia”, disponible en el portal [datos.gov.co](https://www.datos.gov.co).
- Selección del sistema de gestión de base de datos: Se eligió SQLite como motor liviano, embebido y fácilmente integrable con Python, adecuado para una solución local de bajo costo computacional.
- Instalación del SGBD: SQLite fue implementado mediante la librería `sqlite3` incluida en el entorno estándar de Python.
- Diseño del esquema de la base de datos: Se definió una tabla denominada `minas_analisis`, estructurada a partir de los campos del conjunto de datos original, con la adición de columnas de auditoría (`fecha_creacion`, `fecha_update`).
- Carga de datos: Los datos fueron extraídos desde el archivo JSON en línea, transformados a DataFrame con pandas, y cargados en la base de datos utilizando el método `to_sql`.

7. RESULTADOS Y CONCLUSIONES

Se logró satisfactoriamente la extracción de datos desde la fuente pública, su procesamiento y posterior almacenamiento en una base de datos analítica implementada en SQLite, cumpliendo con cada uno de los pasos establecidos en la evidencia de aprendizaje. La solución desarrollada constituye una base técnica sólida sobre la cual pueden realizarse análisis posteriores o integrarse nuevos módulos de visualización o consulta.

8. ANEXO

Link github: https://github.com/cargeo95/bigdata_2025_1_2.git

9. BIBLIOGRAFÍA

Gobierno de Colombia. (s.f.). Eventos Minas Antipersonal en Colombia [Conjunto de datos]. Datos Abiertos Colombia. <https://www.datos.gov.co/Inclusi-n-Social-y-Reconciliaci-n/Eventos-Minas-Antipersonal-en-Colombia/sgp4-3e6k>

McKinney, W. (2018). Python for data analysis: Data wrangling with pandas, NumPy, and IPython (2.ª ed.). O'Reilly Media.

SQLite Consortium. (s.f.). SQLite Documentation. <https://www.sqlite.org/docs.html>

pandas development team. (2024). pandas: powerful Python data analysis toolkit. <https://pandas.pydata.org/>

Requests Development Team. (2024). Requests: HTTP for Humans. <https://requests.readthedocs.io/>

OpenAI. (2025). Asistencia técnica generada con inteligencia artificial para proyectos académicos y de análisis de datos.