

Universidad Santo Tomás

Karen Suárez

Ricardo Vargas

---



# Análisis Exploratorio de Datos de Seguridad y Convivencia (2018-2024)



## 1. Descripción General y Objetivos del Proyecto

El proyecto tiene como finalidad principal **transformar datos brutos de seguridad y convivencia** (delitos) de Colombia para el período **2018-2024** en un **producto analítico interactivo** que facilite la toma de decisiones informadas. Se trata de un proyecto de *Business Intelligence* con enfoque en el **sector público de seguridad**.



### Objetivos Específicos

- Consolidación y Limpieza de Datos:** Integrar información de seguridad pública (delitos) proveniente de múltiples fuentes y años (2018 a 2024), aplicando procesos rigurosos de estandarización y normalización.
- Enriquecimiento del Dataset:** Cruzar la información de delitos con variables demográficas clave, específicamente la **población por municipio y año** y **Divipola**, para calcular métricas de impacto como la **tasa de delitos por cada 100,000 habitantes**.

3. **Desarrollo de una Aplicación Interactiva:** Crear un *dashboard* basado en **Streamlit** que permita a los usuarios explorar visualmente las tendencias temporales y geográficas (departamentos y municipios) de los delitos.
- 



## 2. Fase 1: Preparación y Transformación de Datos

(2018\_2024.py)

Este *script* de Python implementa el **ETL (Extract, Transform, Load)** del proyecto. Su enfoque está en la automatización de la ingesta y la garantía de calidad de los datos geográficos y temporales.



### Arquitectura y Flujo de Trabajo

El flujo se centra en la carga, la estandarización y la posterior fusión de los datos:

1. **Extracción Dinámica:** Define una función genérica (`cargar_delito`) para cargar datos de diferentes URL (presumiblemente archivos Excel por año), manejando variaciones en la estructura del archivo (ej. `header=10` o `header=9`).
2. **Normalización de Esquema:** Estandariza los nombres de las columnas, unificando variaciones comunes (ej. `ARMA_MEDIO`, `ARMAS/MEDIOS`, `ARMAS_Y_MEDIOS` a `ARMAS_MEDIOS`).
3. **Limpieza Geográfica (DIVIPOLA):** Implementa un filtro para eliminar registros asociados a **municipios no oficiales** según el código DIVIPOLA (ej. 'LA VICTORIA',

'PUERTO SANTANDER', 'PUERTO COLOMBIA'), asegurando la consistencia geográfica.

4. **Enriquecimiento con Población:** Realiza un `merge left_only` con la base de datos de población (`pob_2021`) utilizando **DEPARTAMENTO, MUNICIPIO y AÑO** como claves de unión.

## ✅ Hallazgos y Criterios de Calidad

- **Detección de Inconsistencias:** El código incluye un paso crucial (**Paso 8**) para identificar y reportar los municipios que, a pesar de estar en la base de datos de delitos, **no pudieron ser vinculados con los datos de población** (`sin_pob`).
  - *Métrica de Calidad:* El *script* imprime el mensaje: ⚠️ Municipios sin coincidencia de población: `{len(sin_pob)}` y lista los casos.
- **Preparación para Análisis:** El resultado final es la tabla `df_final_pob`, una base limpia y lista para el análisis que contiene la información de delitos junto a la población asociada.

---

## 💻 3. Fase 2: Desarrollo de la Aplicación Interactiva (`app2.py`)

Este *script* convierte la base de datos limpia en una herramienta de análisis dinámico utilizando **Streamlit**, orientada a la experiencia del usuario y la divulgación de los hallazgos.

## 💻 Tecnologías y Librerías Utilizadas

Categoría	Librerías	Propósito
<b>Aplicación Web</b>	<code>streamlit</code>	Construcción de la interfaz de usuario interactiva.
<b>Análisis de Datos</b>	<code>pandas, numpy</code>	Manejo y manipulación de la estructura de datos.
<b>Visualización</b>	<code>plotly.express, plotly.graph_objects, matplotlib, seaborn</code>	Generación de gráficos dinámicos y estáticos.
<b>Utilidades</b>	<code>datetime, os, warnings</code>	Manejo de fechas, rutas de archivo y supresión de alertas.
Export to Sheets		

## Estructura y Estilizado

- **Identidad Visual:** Se implementó un **CSS personalizado** para alinear la aplicación con la identidad corporativa de la Universidad Santo Tomás (USTA), utilizando las variables: `--usta-blue: #002D72` y `--usta-gold: #FDB813`.
- **Navegación:** La aplicación está diseñada con un *layout* `wide` y una barra lateral `expanded`, estructurada en secciones de análisis:
  1. **Análisis Temporal** (Tendencias a lo largo del tiempo).
  2. **Análisis Departamental** (Comparación entre regiones).
  3. **Análisis Municipal** (Vista granular de los *hotspots*).
  4. **Conclusiones y Recomendaciones**.

- **Fuente de Datos:** Espera cargar el resultado del ETL desde la ruta local:  
`delitos_con_poblacion_limpio.csv`.

## Manejo de Errores

El código incluye una sólida **validación de carga de archivos**. Si el archivo fuente no se encuentra, muestra un `st.error` detallado, indicando la **ruta esperada** y los posibles fallos (permisos, formato o columnas faltantes), lo cual es una buena práctica para la portabilidad y depuración de la aplicación.

---

## 4. Conclusiones y Hallazgos del Análisis (Según `app2.py`)

La sección de conclusiones en `app2.py` revela los hallazgos y recomendaciones clave que el *dashboard* está diseñado para soportar.

### Hallazgos Clave

- **Métrica Central:** El análisis se fundamenta en la **Tasa de Delitos por 100.000 Habitantes**, que es la métrica más precisa para comparar el riesgo entre poblaciones de diferente tamaño.
- **Identificación de Patrones:** Se hace énfasis en la detección de:
  - **Picos y Valles Históricos:** Identificación de eventos específicos o cambios de política que afectaron la tendencia de los delitos.

- **Hotspots Geográficos:** Municipios y departamentos con tasas de delitos consistentemente altas.
- **Factores Correlacionados:** Se sugiere la existencia de variables (no detalladas en el código visible) que se correlacionan con el aumento o disminución del delito.

## **Recomendaciones del Proyecto**

El informe concluye con una serie de recomendaciones estratégicas basadas en el análisis de datos:

1. **Asignación de Recursos:** La priorización de recursos de seguridad debe basarse en la **Tasa de Delitos** (riesgo real) y no solo en el **Número Absoluto de Casos**.
2. **Intervenciones Focalizadas:** Diseñar estrategias de prevención y disuasión dirigidas específicamente a los **municipios de alto riesgo (hotspots)** identificados.
3. **Análisis Multicausal:** Investigar las variables económicas, sociales y demográficas que explican el aumento de la criminalidad en las áreas identificadas.
4. **Optimización de Patrullaje:** Ajustar los esquemas de vigilancia según los **picos temporales y geográficos** revelados por el análisis.
5. **Monitoreo Continuo:** Mantener un sistema de seguimiento para detectar cambios en las tendencias a lo largo del tiempo, asegurando que la **toma de decisiones sea informada y dinámica**.

El proyecto, en su conjunto, cumple con el ciclo de **Big Data para Políticas Públicas**, desde la ingesta de datos heterogéneos hasta la presentación de *insights* actionable.