

Análisis de la Incidencia Delictiva en México para la Toma de Decisiones en Seguridad Pública

Objetivo del proyecto

El principal objetivo de este proyecto es encontrar patrones, tendencias y relaciones en los delitos en México para crear información útil que ayude a tomar decisiones sobre seguridad pública. Desde el punto de vista de gestión pública y social, se busca reducir los delitos en un 10% en las zonas más peligrosas en los próximos 24 meses. Esto se logrará mejorando el uso de recursos, enfocando las estrategias de prevención y mejorando cómo los ciudadanos perciben su seguridad.

Este objetivo no solo busca bajar la cantidad de delitos, sino también mejorar la eficiencia del gobierno: que los gobiernos estatales y municipales usen mejor sus presupuestos, que las fuerzas de seguridad sean más preventivas que reaccionarias, y que la ciudadanía note un cambio real en su entorno.

Descripción del problema

La inseguridad en México es uno de los retos más grandes del país. La delincuencia en sus diferentes formas golpea directamente a la gente, empeorando la calidad de vida, limitando dónde y cuándo podemos ir, y creando desconfianza en las autoridades. En la economía, la violencia y el crimen dañan la inversión extranjera, el turismo, y la competitividad; a nivel social, rompen la sociedad, causan desplazamientos dentro del país y hacen creer que la violencia es cosa de todos los días. Aunque hay informes oficiales de delitos, como los del Secretariado Ejecutivo del Sistema Nacional de Seguridad Pública SESNSP, no siempre se usan bien estos datos. Las políticas públicas son usualmente reactivas, solo actuando cuando las cosas ya van mal en lugar de prever y evitar problemas. Esto desperdicia recursos, porque las fuerzas de seguridad se mueven por situaciones urgentes y no por problemas mayores que siguen un patrón.

Otro aspecto vital, es la fragmentación de la información. Los registros de crímenes se presentan en series temporales y clasifican por tipo de delito, aunque, rara vez se integran con variables socioeconómicas, demográficas o territoriales, esto impide comprender las causas profundas de la criminalidad. Ejemplo, delitos patrimoniales, como robo a transeúnte o el robo de vehículos, pueden estar vinculados al desempleo o a la marginación urbana, pero, estas correlaciones no suelen explorarse en los análisis oficiales.

La desigualdad territorial es, a su vez, un factor clave. Mientras, algunos estados muestran tasas de criminalidad relativamente bajas, otros concentran delitos de alto impacto, como homicidios dolosos, secuestros o extorsiones. Esta disparidad exige estrategias distintas, no obstante, la falta de análisis comparativos limita la capacidad de diseñar políticas focalizadas.

Por último, la brecha entre cifras oficiales y percepción ciudadana es otro problemilla. Encuestas, como la ENVIPE, muestran que, aun en lugares donde los datos indican bajas en ciertos delitos, la gente sigue sintiéndose insegura. Esta desconexión debilita la confianza en las instituciones, dificultando la colaboración ciudadana en programas de prevención.

Entendiendo el meollo del asunto

El problema fundamental por solucionar es éste: aunque poseemos exhaustivos datos sobre la delincuencia en México, estos no se traducen en una inteligencia estratégica eficaz. Esto impide prever tendencias, focalizar recursos correctamente, y crear políticas públicas preventivas verdaderamente útiles. En consecuencia, las autoridades reaccionan en vez de actuar proactivamente, la gente siente una constante sensación de inseguridad, y los pocos recursos disponibles se despilfarran inútilmente.

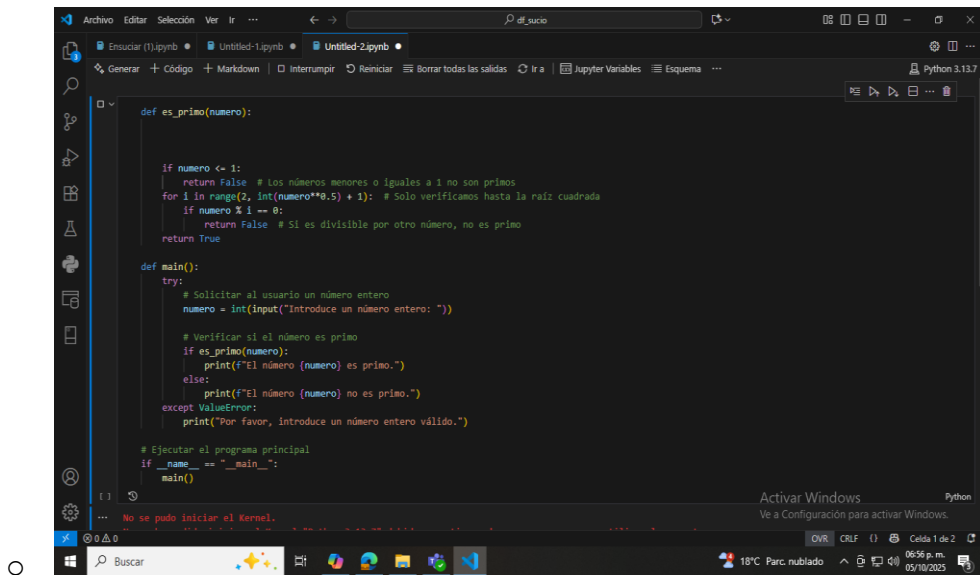
Así las cosas, el proyecto tiene una meta clave: transformar los archivos históricos criminales en un conocimiento que se pueda usar para actuar. Queremos que responda a preguntas como: ¿Qué delitos son más comunes?, ¿Dónde ocurren con más frecuencia?, ¿Cuándo suelen subir?, y ¿Qué factores sociales o económicos influyen en su aumento o disminución? Resolver este dilema nos permitirá dejar la improvisación atrás, y basar las decisiones en evidencia, mejorando así la eficiencia gubernamental y afianzando la confianza de la sociedad.

Recursos Disponibles

- Tecnologías y Herramientas

Para la elaboración de este análisis se usan los siguientes recursos:

- Lenguaje de programación Python: Python es conocido por su sintaxis clara y fácil de entender, lo que permite a los científicos de datos concentrarse en resolver problemas en lugar de lidiar con la complejidad del lenguaje de programación.



○ Pandas: Para la manipulación y análisis de datos tabulares.

The screenshot shows a Jupyter Notebook window displaying a Pandas DataFrame. The DataFrame has 11 columns: Año, Clave_Ent, Entidad, Bien jurídico afectado, Tipo de delito, Subtipo de delito, Modalidad, Enero, Febrero, Marzo, Mayo, Junio, Julio, Agosto, Septiembre, and Octubre. The data is as follows:

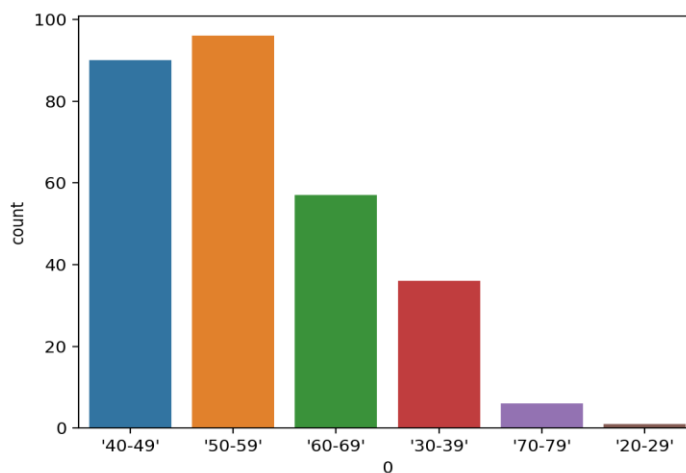
	Año	Clave_Ent	Entidad	Bien jurídico afectado	Tipo de delito	Subtipo de delito	Modalidad	Enero	Febrero	Marzo	Mayo	Junio	Julio	Agosto	Septiembre	Octubre
0	2015.0	1.0	NaN	La libertad y la seguridad sexual	Abuso sexual	Abuso sexual	Abuso sexual	6.0	4	0.0	NaN	1.0	NaN	1.0	0.0	0.
1	2015.0	1.0	NaN	La libertad y la seguridad sexual	Delito sexual	Acoso sexual	NaN	0.0	0	0.0	0.0	NaN	0.0	NaN	0.0	0.
2	NaN	1.0	Aguascalientes	La libertad y la seguridad sexual	Delito sexual	Hostigamiento sexual	Hostigamiento sexual	1.0	1	0.0	3.0	0.0	1.0	0.0	0.0	1.
3	2015.0	1.0	Aguascalientes	bbb	Violación	Violación simple	Violación simple	5.0	5	3.0	3.0	15.0	8.0	NaN	6.0	9.
4	2015.0	1.0	Aguascalientes	La libertad y la seguridad sexual	Violación	Violación equiparada	Violación equiparada	0.0	5	3.0	3.0	4.0	4.0	4.0	3.0	Na
...

The status bar at the bottom indicates "Activar Windows" and "19°C Parc. nublado".

○ NumPy: Para cálculos matemáticos y operaciones numéricas eficientes.

```
import numpy as np.py 1
import numpy as np
2
3 # crear un arreglo
4 a = np.array([1, 2, 3, 4])
5
6 # Sumar todos los elementos
7 print(np.sum(a))
8
9 # calcular promedio
10 print(np.mean(a))
11
```

- Matplotlib y Seaborn: Para la visualización de datos de manera comprensible y atractiva



- Datos

El proyecto cuenta con una base de datos de incidencia delictiva en México, recopilada y organizada a partir de registros oficiales. Cada fila representa la ocurrencia de un delito en un periodo de tiempo (mes y año) y en una entidad federativa determinada. Las columnas disponibles son las siguientes:

- Año: Variable numérica que indica el año en el que se registró el delito. Permite analizar tendencias históricas y comparaciones interanuales.
- Mes: Variable categórica/temporal que señala el mes del registro. Es útil para identificar patrones estacionales o picos de criminalidad en determinadas épocas del año.

- Entidad: Variable categórica que corresponde al estado de la República Mexicana donde ocurrió el delito. Permite realizar análisis geográficos y comparativos entre regiones.
- Tipo de delito: Clasificación general del delito (ejemplo: homicidio, secuestro, robo, extorsión, fraude, etc.). Es la categoría principal de análisis.
- Subtipo de delito: Desglose más específico dentro del tipo de delito (ejemplo: dentro de “robo” se incluyen robo a casa habitación, robo de vehículo, robo a transeúnte, etc.). Ayuda a entender la naturaleza particular de los crímenes.
- Modalidad: Describe la forma en que se cometió el delito, por ejemplo, con violencia o sin violencia. Esta variable es clave para diferenciar delitos de alto impacto social.
- Número de casos: Variable numérica que indica la cantidad de delitos registrados en ese mes, entidad y categoría. Es la métrica principal para medir la incidencia.
- Tasa por cada 100,000 habitantes (en algunos registros): Indicador estandarizado que permite comparar la incidencia delictiva entre entidades con diferentes tamaños de población.

Hipótesis Iniciales

1. Concentración geográfica: Los delitos graves, como homicidios, secuestros y extorsiones, se concentran en estados con más población y actividad económica. Esto se debe a la presencia de mercados ilegales y al mayor anonimato en las ciudades.
2. Factores socioeconómicos: Existe una relación clara entre el desempleo y un aumento en delitos como robos, fraudes y asaltos. La falta de oportunidades puede impulsar la delincuencia.
3. Estacionalidad del crimen: La cantidad de delitos sube en ciertos meses del año, especialmente durante las vacaciones y festividades, cuando aumenta la movilidad y el consumo.

Estas ideas ayudarán a analizar y verificar si los patrones observados corresponden a intuiciones iniciales o si hay aspectos más complejos en juego.

Definición de Stakeholders Clave

- **Gobiernos Estatales y Municipales:** Son responsables de poner en marcha políticas públicas y de asignar recursos para la seguridad.
- **Secretaría de Seguridad y Protección Ciudadana:** Encargada de coordinar estrategias nacionales y de identificar las zonas más peligrosas.
- **Organizaciones Civiles y ONGs:** Son actores importantes en prevenir delitos y en vigilar la comunidad.
- **Ciudadanía:** Es la principal beneficiada cuando hay menos crímenes y la seguridad mejora.
- **Medios de Comunicación:** Son quienes difunden información y pueden ayudar a que la gente confíe más en las instituciones.

Cada grupo tiene un papel importante, pero todos buscan lo mismo: reducir la criminalidad y hacer que México sea más seguro.

Preguntas Clave

1. ¿Qué tipos de delitos son más comunes en México?
2. ¿Cómo varía la cantidad de delitos entre estados y municipios?
3. ¿Hay patrones en el tiempo (cada mes, cada año) en la ocurrencia de Delitos?
4. ¿Qué delitos afectan más la percepción de inseguridad entre los Ciudadanos?
5. ¿Qué relación hay entre factores socioeconómicos y la cantidad de delitos?
6. ¿Cuáles estados tienen más delitos violentos?
7. ¿Qué delitos han aumentado más en los últimos 5 años?
8. ¿Existen conexiones entre delitos relacionados con bienes y la tasa de desempleo?
9. ¿En qué zonas se necesita más inversión para prevenir delitos?
10. ¿Cómo se comparan los patrones de delitos en México con otros países de la región?

Estas preguntas ayudarán a guiar la investigación y a enfocar el análisis en patrones, causas y posibles soluciones.

Fuentes de Datos Identificadas

- Dataset Kaggle “Crime rates - Mexico”: Base principal con registros de delitos clasificados por tipo y entidad.
- INEGI (Instituto Nacional de Estadística y Geografía): Encuestas de victimización y percepción de seguridad.
- Secretariado Ejecutivo del Sistema Nacional de Seguridad Pública (SESNSP): Reportes oficiales de incidencia delictiva.
- Datos socioeconómicos (CONAPO, Banco de México): Indicadores de empleo, pobreza y desarrollo.
- Datos de percepción ciudadana (ENCVI, ENVIPE): Encuestas sobre confianza en instituciones y percepción de inseguridad.

La combinación de estas fuentes permitirá un análisis integral que no solo mida delitos, sino también contexto social y económico.

Justificación del Proyecto

La inseguridad en México es uno de los principales retos para el desarrollo social, económico y político del país. La delincuencia no solo afecta a las víctimas, sino que también influye en cómo las personas perciben la seguridad, confían en las instituciones y la competitividad del país. Por eso, analizar los datos delictivos es importante, ya que ayuda a entender las cifras y a tomar mejores decisiones.

Primero, la criminalidad tiene un gran costo económico. Diversos estudios estiman que la violencia y la inseguridad causan pérdidas millonarias cada año en inversión extranjera, turismo y productividad laboral. Las empresas gastan más en seguridad privada, seguros y protocolos, lo que aumenta sus costos y reduce su ventaja competitiva. A nivel general, la inseguridad desalienta la llegada de inversiones extranjeras y limita el crecimiento de sectores como el turismo, que depende mucho de sentirse seguro.

En segundo lugar, la inseguridad tiene un gran impacto social. Sentir que siempre hay riesgo afecta la calidad de vida de las personas, limita la movilidad en espacios públicos y deteriora las comunidades. Muchas regiones del país han tenido desplazamientos internos, fragmentación familiar y pérdida de confianza en las instituciones que deben protegerlos. Esto daña las relaciones sociales y crea un

ambiente de miedo, que evita que la gente participe en actividades comunitarias y políticas.

En tercer lugar, hay un problema en el uso de datos. Aunque México tiene registros oficiales de delitos, estos suelen usarse solo para responder a emergencias, no para prever tendencias. Esto hace que las políticas públicas se basen en la urgencia, en lugar de una planificación a largo plazo. El proyecto busca cambiar esto, proponiendo un análisis sistemático que identifique patrones de crimen según lugar, tiempo y condiciones sociales.

El proyecto también tiene una dimensión institucional. Usar datos confiables ayuda a que las autoridades tomen decisiones acertadas y a que la ciudadanía confíe más en las instituciones de seguridad. Ser transparente con los datos y comunicar bien los resultados puede ayudar a reducir la brecha entre lo que la gente piensa y las cifras oficiales, generando más confianza y colaboración.

Finalmente, este análisis ayuda a optimizar recursos. Con presupuestos limitados, es clave usar el dinero en seguridad de manera eficiente, priorizando las zonas y delitos más importantes. Basar las acciones en datos mejora las políticas públicas y también ayuda a reducir los gastos en estrategias que no funcionan o están mal enfocadas.

En resumen, este proyecto es importante porque analizar la delincuencia no solo es un ejercicio académico, sino una herramienta práctica para mejorar la realidad social. Convertir los datos en información útil busca no solo bajar los niveles de criminalidad, sino también fortalecer la comunidad, mejorar cómo la gente percibe su seguridad y apoyar el desarrollo sostenible del país.

¿Cuántos datos y que tipo son?

Cantidad de Datos

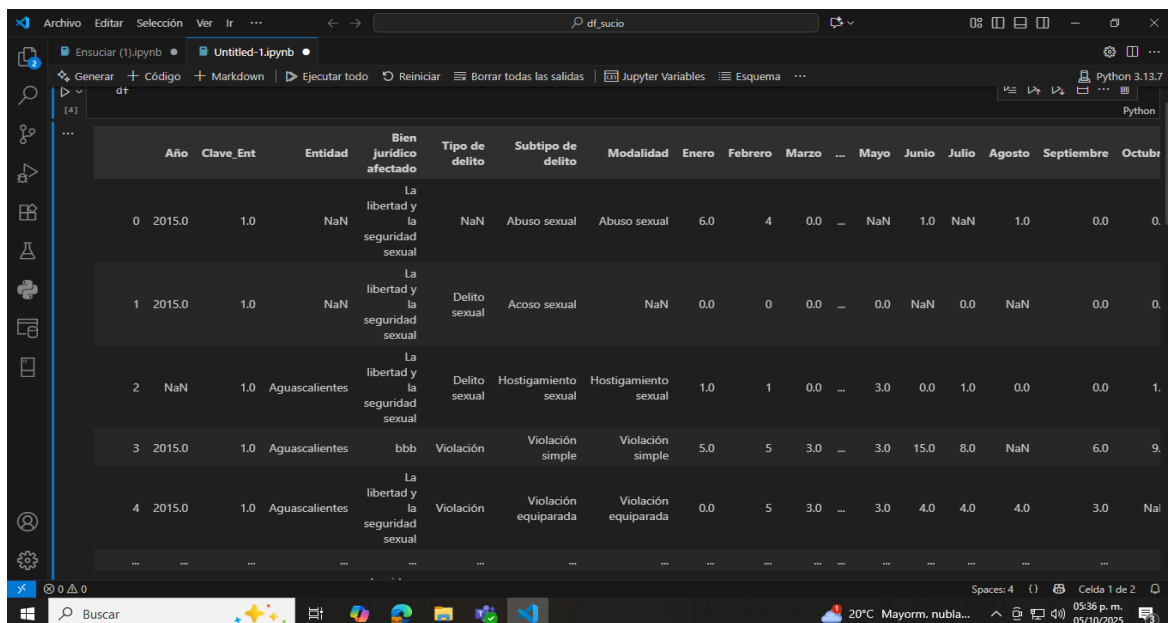
- El dataset contiene más de 300,000 registros (filas), correspondientes a reportes de delitos en México.
- La cobertura temporal va de enero de 2015 hasta 2023 (dependiendo de la última actualización).
- Cada fila representa un registro mensual de incidencia delictiva por entidad federativa y tipo de delito.

☐Tipos de Datos Incluidos

El dataset está estructurado en columnas que combinan variables categóricas, numéricas y temporales:

- Variables de identificación y contexto:
 - Año (numérica, entera) → año del registro.
 - Mes (categórica/temporal) → mes del registro.
 - Entidad (categórica) → estado de la república mexicana.
- Variables de clasificación del delito:
 - Tipo de delito (categórica) → homicidio, secuestro, robo, extorsión, etc.
 - Subtipo (categórica) → por ejemplo, dentro de “robo”: robo a casa habitación, robo de vehículo, robo a transeúnte.
 - Modalidad (categórica) → con violencia, sin violencia.
- Número de casos (numérica, entera) → cantidad de delitos registrados en ese mes, entidad y categoría.
- En algunos casos, se incluyen tasas normalizadas (por cada 100,000 habitantes).

Ejemplo de Estructura de un Registro



The screenshot shows a Jupyter Notebook interface with a dataset named 'df_sucio'. The dataset has columns for identification and context (Año, Clave_Ent, Entidad, Bien jurídico afectado), classification (Tipo de delito, Subtipo de delito, Modalidad), and temporal data (Enero, Febrero, Marzo, Mayo, Junio, Julio, Agosto, Septiembre, Octubre). The data is displayed in a table format with rows representing individual records.

	Año	Clave_Ent	Entidad	Bien jurídico afectado	Tipo de delito	Subtipo de delito	Modalidad	Enero	Febrero	Marzo	...	Mayo	Junio	Julio	Agosto	Septiembre	Octubre
0	2015.0	1.0	NaN	La libertad y la seguridad sexual	NaN	Abuso sexual	Abuso sexual	6.0	4	0.0	...	NaN	1.0	NaN	1.0	0.0	0.
1	2015.0	1.0	NaN	La libertad y la seguridad sexual	Delito sexual	Acoso sexual	NaN	0.0	0	0.0	...	0.0	NaN	0.0	NaN	0.0	0.
2	NaN	1.0	Aguascalientes	La libertad y la seguridad sexual	Delito sexual	Hostigamiento sexual	Hostigamiento sexual	1.0	1	0.0	...	3.0	0.0	1.0	0.0	0.0	1.
3	2015.0	1.0	Aguascalientes	bbb	Violación	Violación simple	Violación simple	5.0	5	3.0	...	3.0	15.0	8.0	NaN	6.0	9.
4	2015.0	1.0	Aguascalientes	La libertad y la seguridad sexual	Violación	Violación equiparada	Violación equiparada	0.0	5	3.0	...	3.0	4.0	4.0	4.0	3.0	NaN

En resumen

- Volumen: cientos de miles de registros (gran base para análisis estadístico y predictivo).
- Tipos de datos:
 - Temporales (año, mes).
 - Catégoricos (estado, tipo de delito, subtipo, modalidad).

Numéricos (número de casos, tasas).

Con esta estructura, el dataset permite hacer análisis espaciales (por estado), temporales (por año/mes), tipológicos (por delito) y socioeconómicos (cruzando con otras fuentes).