



BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA

PROYECTO FINAL DE INTRODUCCIÓN A CIENCIA DE DATOS.

"Análisis del Tasas de criminalidad en México."

Asignatura: INTRODUCCION A LA CIENCIA DE DATOS

Carrera: INGENIERIA EN CIENCIA DE DATOS

Nombre del alumno: JOANA MONTSERRAT CARMONA NOGUEZ

Nombre del profesor: JAIME A. ROMERO SIERRA

Horario de la materia: LUNES, MARTES Y JUEVES

No. De cuenta: 202466401

Fecha de entrega: 28 DE NOVIEMBRE DE 2024

Reporte Fase 1: Descubrimiento del Proyecto

1. Título del Proyecto:

"Análisis del Tasas de criminalidad en México."

2. Objetivo del Proyecto:

Usar la información para ayudar a las autoridades a planificar estrategias de seguridad, identificando tendencias o ciclos de aumento de delitos en ciertos meses o bajo determinadas modalidades.

3. Descripción del Problema:

A partir de los datos sobre los meses en los que se cometieron los delitos se puede ayudar a identificar períodos o temporadas críticas en los que ciertos tipos de delitos son más probables. Esto se puede utilizar para determinar si hay un patrón cíclico, como un aumento del robo durante las temporadas festivas, ciertos tipos específicos de violencia que son característicos de distintos períodos, etc. Clasificar los delitos por su tipo y subtipo también permitirá identificar en qué momentos son más probables ciertos tipos de actividades delictivas. Por ejemplo, quizás los delitos contra la propiedad, como los robos, los daños intencionales y el fraude ocurren con más frecuencia en ciertos momentos, mientras que los delitos contra la integridad de la persona, como las lesiones y los homicidios, son predominantes en otros. Al analizar los bienes jurídicos que un crimen ataca, como la vida, la propiedad, la seguridad pública, etc, las agencias de seguridad podrán comprender dónde se concentran los riesgos en momentos específicos o bajo ciertos tipos de delitos. Esto permitirá a los responsables tomar medidas preventivas, como enviar patrullas policiales, lanzar campañas de concienciación, patrullar en zonas con un alto nivel de delitos en ciertas épocas.

4. Recursos Disponibles

Se utilizarán herramientas como Python y Pandas para procesar y manipular los datos de manera eficiente, mientras que Matplotlib y Plotly se emplearán para generar visualizaciones iniciales que faciliten el análisis.

Descripción de Datos:

- **Año:** Este campo compila el año en el que se recopilaron los datos sobre los delitos cometidos en México. Permite analizar las tendencias a lo largo del año.
- **Clave_Ent (Entidad clave):** Este código numérico es usado en lugar del nombre completo de cada entidad federativa (estado). Ayuda a identificar rápidamente para el usuario la entidad en cuestión.
- **Estado:** Nombre de la entidad federativa (estado) de México en la que ocurrieron los delitos. Geolocaliza la actividad delictiva.
- **Bienes jurídicos afectados:** Este campo describe el objeto jurídico protegido por la ley que ha sido violado el delito (por ejemplo, la vida, que se considera un derecho humano). Este campo posibilita categorizar delitos de acuerdo con el impacto en derechos fundamentales.
- **Tipo de delito:** Esta es una categoría general del delito cometido: asesinato, robo, secuestro. Muestra una única idea general sobre cómo se cometen crímenes.
- **Otro tipo de delito:** especifica con más detalle el tipo de delito (por ejemplo, asesinato intencional, robo a mano armada). Ofrece una visión más detallada del crimen que se está cometiendo.
- **Manera en que se cometió el delito:** Esta categoría describe las circunstancias y el modo en que el delito fue cometido (por ejemplo, con violencia, con arma de fuego). Es útil para estudiar crimen según sus características individuales.
- **Meses del año:** Muestra el número de delitos registrados en cada mes del año para una entidad, bien jurídico, tipo y subtipo de delito. Permite ver la dirección tomada por el crimen y las fluctuaciones a lo largo del año, síntoma de posibles auge o disminución en actividad delictiva.

- **Sexo/Averiguación previa:** Puede indicar el sexo de la víctima o el presunto criminal, o bien la naturaleza de la averiguación previa (expediente abierto en un proceso de investigación del crimen). Esta columna puede tener diferentes tipos de información según el diseño de la base de datos.
- **Rango de edad:** Clasifica a las víctimas o presuntos responsables de los crímenes según su grupo de edad. Estos rangos pueden ser definidos por intervalos (por ejemplo, 0-17, 18-29, 30-44, etc.), y ayudan a identificar patrones en la incidencia de crímenes en diferentes grupos etarios.

5. Hipótesis Iniciales

Hipótesis 1: "En México, el delito más frecuente relacionado con la violencia es el homicidio, y la mayoría de estos crímenes son cometidos con armas de fuego".

Hipótesis 2: " En México, marzo es el mes en el que se registra el mayor número de delitos".

Hipótesis 3: "Aguascalientes es el estado con los niveles más altos de violencia en México".

6. Definición de Stakeholders Clave

- Gobierno/ Autoridades de seguridad pública y justicia: Utilizarán los resultados para diseñar políticas y estrategias de prevención y combate al crimen, así como para asignar recursos de manera más eficiente.
- Ciudadanía: Los resultados del análisis pueden traducirse en mejores políticas de seguridad que beneficien a la población.
- Medios de comunicación: Pueden utilizar los resultados del análisis para informar al público sobre tendencias delictivas, áreas peligrosas y posibles causas de aumento de la violencia, ayudando a sensibilizar y educar a la sociedad sobre la situación.
- Políticos y legisladores: Utilizan los hallazgos del análisis para tomar decisiones informadas al momento de formular leyes y regulaciones que aborden los problemas del crimen y la violencia, promoviendo reformas que mejoren la seguridad y el bienestar de la sociedad.

7. Preguntas Clave

- ¿En qué año(s) han aumentado más los delitos?
- ¿Qué estados registran más delitos en ciertos meses?
- ¿Qué derechos (como la vida o la propiedad) se ven más afectados en ciertos meses?
- ¿Qué tipos de delitos tienden a aumentar en cada mes?
- ¿Hay subtipos específicos de delitos que crecen en algunos meses?
- ¿Qué formas de cometer delitos son más comunes en algunos meses?
- ¿En qué meses se concentran más delitos?
- ¿Qué estados y tipos de delitos muestran ciclos parecidos a lo largo del año?
- ¿Cuáles son los picos de delitos según la forma en que se cometen en cada mes?
- ¿En qué meses aumenta el impacto en derechos específicos como la vida o la propiedad?
- ¿Existen estados que tienen patrones similares de aumento o disminución de delitos en los mismos meses?
- ¿Qué subtipos de delitos deben ser prioridad en los meses con más actividad delictiva?

8. Fuentes de Datos Identificadas

- Datos de criminalidad
- Datos demográficos y socioeconómicos por entidad
- Factores socioeconómicos y políticos temporales
- Estadísticas de migración interna y externa
- Factores socioeconómicos y políticos temporales

9. Justificación del Proyecto

El análisis de las tasas de criminalidad en México es fundamental para ayudar a las autoridades a planificar estrategias de seguridad más efectivas. Al estudiar cuándo y dónde ocurren los delitos con mayor frecuencia, se pueden identificar patrones

que permitan anticiparse a picos delictivos, como robos que aumentan en la temporada festiva o actos violentos que se intensifican en ciertos meses. Esto no solo permite mejorar la asignación de recursos, como incrementar la presencia policial en momentos clave, sino también lanzar campañas preventivas que puedan reducir la incidencia delictiva.

Además, al desglosar los delitos por tipo, subtipo y cómo se cometen, las autoridades pueden enfocarse en aquellos que afectan más gravemente derechos como la vida, la propiedad o la seguridad pública. Este tipo de análisis brinda una perspectiva más clara para ajustar las acciones según la realidad de cada región y tipo de crimen, lo que resulta en estrategias más precisas y eficaces.

Finalmente, el análisis no solo beneficia a las autoridades, sino también a la ciudadanía, que verá reflejadas mejores políticas de protección en su vida diaria. Asimismo, los legisladores podrán utilizar estos hallazgos para crear leyes más adecuadas a las necesidades actuales, y los medios de comunicación podrán sensibilizar mejor a la población sobre los riesgos en diferentes momentos del año.

10.- ¿Cuántos datos y que tipo son?

RangeIndex: 101830 entries, 0 to 101829

Data columns (total 21 columns):

| # | Column | Non-Null Count | Dtype |
|-----|------------------------|----------------|--------|
| --- | ----- | ----- | ----- |
| 0 | Año | 97626 non-null | object |
| 1 | Clave_Ent | 97626 non-null | object |
| 2 | Entidad | 97626 non-null | object |
| 3 | Bien jurídico afectado | 97626 non-null | object |
| 4 | Tipo de delito | 97626 non-null | object |
| 5 | Subtipo de delito | 97626 non-null | object |
| 6 | Modalidad | 97626 non-null | object |
| 7 | Enero | 97626 non-null | object |
| 8 | Febrero | 97626 non-null | object |
| 9 | Marzo | 97626 non-null | object |

| | | |
|----|--------------------------|------------------------|
| 10 | Abril | 97626 non-null object |
| 11 | Mayo | 97626 non-null object |
| 12 | Junio | 97626 non-null object |
| 13 | Julio | 86763 non-null object |
| 14 | Agosto | 86757 non-null object |
| 15 | Septiembre | 86741 non-null object |
| 16 | Octubre | 86678 non-null float64 |
| 17 | Noviembre | 86747 non-null object |
| 18 | Diciembre | 86755 non-null object |
| 19 | Sexo/Averiguación previa | 97626 non-null object |
| 20 | Rango de edad | 97626 non-null object |

dtypes: float64(1), object(20)

memory usage: 16.3+ MB

Práctica de Laboratorio - Unidad 2: **Introducción a la Limpieza de Datos**

Materia: **Introducción a la Ciencia de Datos**

Unidad 2: **Procesamiento y Limpieza de Datos**

Título de la Práctica: **Limpieza de una Base de Datos Ensuciada**

Nombre: **Joana Montserrat Carmona Noguez**

Día y horario de la materia: **Lunes, Martes, Jueves**

Análisis inicial:

- Resumen estadístico de la base de datos antes de la limpieza.

```
[332] #Dataframe
df.shape
```

(101830, 21)

```
[189] #Resumen estadístico
df.describe()
```

| | Octubre |
|-------|--------------|
| count | 86678.000000 |
| mean | 18.210076 |
| std | 114.129618 |
| min | 0.000000 |
| 25% | 0.000000 |
| 50% | 0.000000 |
| 75% | 2.000000 |
| max | 6750.000000 |

- Tabla que muestre el porcentaje de valores faltantes por columna.

| | |
|--------------------------|-----------|
| Año | 4.128449 |
| Clave_Ent | 4.128449 |
| Entidad | 4.128449 |
| Bien jurídico afectado | 4.128449 |
| Tipo de delito | 4.128449 |
| Subtipo de delito | 4.128449 |
| Modalidad | 4.128449 |
| Enero | 4.128449 |
| Febrero | 4.128449 |
| Marzo | 4.128449 |
| Abril | 4.128449 |
| Mayo | 4.128449 |
| Junio | 4.128449 |
| Julio | 14.796229 |
| Agosto | 14.802121 |
| Septiembre | 14.817834 |
| Octubre | 14.879701 |
| Noviembre | 14.811941 |
| Diciembre | 14.804085 |
| Sexo/Averiguación previa | 4.128449 |
| Rango de edad | 4.128449 |

- Total de filas duplicadas encontradas.

```
[ ] # Identificar filas duplicadas
df.duplicated().sum()

2918
```

- Descripción de los tipos de datos originales y los problemas encontrados.

| | |
|--------------------------|---------|
| Año | object |
| Clave_Ent | object |
| Entidad | object |
| Bien jurídico afectado | object |
| Tipo de delito | object |
| Subtipo de delito | object |
| Modalidad | object |
| Enero | object |
| Febrero | object |
| Marzo | object |
| Abril | object |
| Mayo | object |
| Junio | object |
| Julio | object |
| Agosto | object |
| Septiembre | object |
| Octubre | float64 |
| Noviembre | object |
| Diciembre | object |
| Sexo/Averiguación previa | object |
| Rango de edad | object |

Proceso de limpieza:

- Describir qué métodos utilizaron para limpiar la base de datos (eliminación, imputación, etc.).

Se aplicaron varias técnicas para limpiar la base de datos. Primero, los valores faltantes (NaN) en columnas clave como Entidad, Bien jurídico afectado, Tipo de delito y Subtipo de delito fueron reemplazados por "Desconocido", lo que permitió mantener la integridad del conjunto de datos y facilitar la identificación de información incompleta sin perder registros importantes. En los casos donde los datos faltantes eran irrecuperables o irrelevantes, se optó por eliminar las filas afectadas, garantizando que la base de datos permaneciera consistente.

También se corrigieron las inconsistencias en los tipos de datos, por ejemplo, la columna Año fue convertida a formato numérico para asegurar que pudiera utilizarse en análisis cronológicos sin errores. Además, se eliminaron los duplicados, evitando que los eventos delictivos se contaran más de una vez, lo que podría sesgar los resultados. Por último, se corrigieron valores mal escritos o inválidos en algunas columnas, reemplazándolos con valores correctos para asegurar la coherencia de los datos.

- Mostrar antes y después de cada paso clave (por ejemplo, antes y después de eliminar duplicados).

```
[ ] # Identificar filas duplicadas
df.duplicated().sum()
2918

[ ] # Eliminamos las filas duplicadas
df=df.drop_duplicates()
df

[ ] #Comprobamos eliminacion de filas duplicadas
df.duplicated().sum()
0
```

✓
0 s

```
[ ] #Borrar la columna Clave_Ent por no ser necesario en la investigacion
df = df.drop(columns=['Clave_Ent'])
print(df.columns)
```

➦ Index(['Año', 'Entidad', 'Bien jurídico afectado', 'Tipo de delito', 'Subtipo de delito', 'Modalidad', 'Enero', 'Febrero', 'Marzo', 'Abril', 'Mayo', 'Junio', 'Julio', 'Agosto', 'Septiembre', 'Octubre', 'Noviembre', 'Diciembre', 'Sexo/Averiguación previa', 'Rango de edad'], dtype='object')

✓
0 s

▶ #Comprobamos que no hay Clave_Ent
df.columns

➦ Index(['Año', 'Entidad', 'Bien jurídico afectado', 'Tipo de delito', 'Subtipo de delito', 'Modalidad', 'Enero', 'Febrero', 'Marzo', 'Abril', 'Mayo', 'Junio', 'Julio', 'Agosto', 'Septiembre', 'Octubre', 'Noviembre', 'Diciembre', 'Sexo/Averiguación previa', 'Rango de edad'], dtype='object')

```
[ ] #Comprobacion de no 'bbb'
for i in df.columns:
    print(f"En la columna {i} los invalid_value son: {df[df[i]=='bbb'].shape[0]}")
```

➦ En la columna Año los invalid_value son: 0
En la columna Entidad los invalid_value son: 0
En la columna Bien jurídico afectado los invalid_value son: 0
En la columna Tipo de delito los invalid_value son: 0
En la columna Subtipo de delito los invalid_value son: 0
En la columna Modalidad los invalid_value son: 0
En la columna Enero los invalid_value son: 0
En la columna Febrero los invalid_value son: 0
En la columna Marzo los invalid_value son: 0
En la columna Abril los invalid_value son: 0
En la columna Mayo los invalid_value son: 0
En la columna Junio los invalid_value son: 0
En la columna Julio los invalid_value son: 0
En la columna Agosto los invalid_value son: 0
En la columna Septiembre los invalid_value son: 0
En la columna Octubre los invalid_value son: 0
En la columna Noviembre los invalid_value son: 0
En la columna Diciembre los invalid_value son: 0
En la columna Sexo/Averiguación previa los invalid_value son: 0
En la columna Rango de edad los invalid_value son: 0

```
[ ] # Convertirá NaN en valores numéricos mode=moda mean=promedio
df['Año'] = df['Año'].fillna(df['Año'].mode().iloc[0])
# Lista de los meses del año
meses = ['Enero', 'Febrero', 'Marzo', 'Abril', 'Mayo', 'Junio',
        'Julio', 'Agosto', 'Septiembre', 'Octubre', 'Noviembre', 'Diciembre']
# Ciclo for para aplicar la conversión a numérico y rellenar valores faltantes con la media
for mes in meses:
    df[mes] = pd.to_numeric(df[mes], errors='coerce')
    df[mes] = df[mes].fillna(df[mes].mean())

# Reemplazar NaN en la columna 'Rango de Edad' con 'No especificado'
df['Rango de edad'] = df['Rango de edad'].fillna('No Especificado')
# Reemplazar 'No aplica' por 'No especificado' en la columna 'Rango de edad'
df['Rango de edad'] = df['Rango de edad'].replace('No aplica', 'No Especificado')

# Reemplazar las variables sin valor en la columna 'Sexo/Averiguación previa' con 'No Identificado'
df['Sexo/Averiguación previa'] = df['Sexo/Averiguación previa'].fillna('No Identificado')
# Reemplazar NaN en la columna 'Sexo/Averiguación previa' con 'No Identificado'
df['Sexo/Averiguación previa'] = df['Sexo/Averiguación previa'].replace('NaN', 'No Identificado')

# Reemplazar las variables sin valor en la columna 'Modalidad' con 'No Especificado'
df['Modalidad'] = df['Modalidad'].fillna('No Especificado')
# Reemplazar NaN en la columna 'Modalidad' con 'No especificado'
df['Modalidad'] = df['Modalidad'].replace('NaN', 'No especificado')
# Lista de columnas en las que se desea reemplazar valores faltantes
columnas_a_reemplazar = ['Entidad', 'Bien jurídico afectado', 'Tipo de delito', 'Subtipo de delito']

# Reemplazar las variables sin valor en la columna 'Entidad' con 'Desconocido'
df['Entidad'] = df['Entidad'].fillna('Desconocido')
# Reemplazar NaN en la columna 'Entidad' con 'Desconocido'
df['Entidad'] = df['Entidad'].replace('NaN', 'Desconocido')
```

```
# Reemplazar las variables sin valor en la columna 'Bien jurídico afectado' con 'Desconocido'
df['Bien jurídico afectado'] = df['Bien jurídico afectado'].fillna('Desconocido')
# Reemplazar NaN en la columna 'Bien jurídico afectado' con 'Desconocido'
df['Bien jurídico afectado'] = df['Bien jurídico afectado'].replace('NaN', 'Desconocido')

# Reemplazar las variables sin valor en la columna 'Tipo de delito' con 'Desconocido'
df['Tipo de delito'] = df['Bien jurídico afectado'].fillna('Desconocido')
# Reemplazar NaN en la columna 'Tipo de delito' con 'Desconocido'
df['Tipo de delito'] = df['Bien jurídico afectado'].replace('NaN', 'Desconocido')

# Reemplazar las variables sin valor en la columna 'Subtipo de delito' con 'Desconocido'
df['Subtipo de delito'] = df['Subtipo de delito'].fillna('Desconocido')
# Reemplazar NaN en la columna 'Subtipo de delito' con 'Desconocido'
df['Subtipo de delito'] = df['Subtipo de delito'].replace('NaN', 'Desconocido')
```

```
df
```

```
df['Entidad'] = df['Entidad'].replace('invalid', 'Desconocido')
df['Año'] = df['Año'].replace('invalid', '0000')
df['Bien jurídico afectado'] = df['Bien jurídico afectado'].replace('invalid', 'Desconocido')
df['Tipo de delito'] = df['Tipo de delito'].replace('invalid', 'Desconocido')
df['Subtipo de delito'] = df['Subtipo de delito'].replace('invalid', 'Desconocido')
df['Modalidad'] = df['Modalidad'].replace('invalid', 'No especificado')
df['Modalidad'] = df['Modalidad'].replace('No Especificado', 'No especificado')
df['Rango de edad'] = df['Rango de edad'].replace('invalid', 'No Especificado')
df['Rango de edad'] = df['Rango de edad'].replace('No identificado', 'No Especificado')
df['Rango de edad'] = df['Rango de edad'].replace('No especificado', 'No Especificado')
df['Sexo/Averiguación previa'] = df['Sexo/Averiguación previa'].replace('No Identificado', 'Averiguación previa')
df['Sexo/Averiguación previa'] = df['Sexo/Averiguación previa'].replace('invalid', 'Averiguación previa')
df['Sexo/Averiguación previa'] = df['Sexo/Averiguación previa'].replace('No identificado', 'Averiguación previa')
```

✓

0 s

▶

#Comprobamos no NaN

df.isnull().sum()

✓

0 s

▶

#Comprobamos porcentaje de valores faltantes por columna

df.isnull().mean() * 100

✓

0 s

▶

df.shape

↔ (96994, 20)

| | |
|--------------------------|---|
| Año | 0 |
| Entidad | 0 |
| Bien jurídico afectado | 0 |
| Tipo de delito | 0 |
| Subtipo de delito | 0 |
| Modalidad | 0 |
| Enero | 0 |
| Febrero | 0 |
| Marzo | 0 |
| Abril | 0 |
| Mayo | 0 |
| Junio | 0 |
| Julio | 0 |
| Agosto | 0 |
| Septiembre | 0 |
| Octubre | 0 |
| Noviembre | 0 |
| Diciembre | 0 |
| Sexo/Averiguación previa | 0 |
| Rango de edad | 0 |

| | |
|--------------------------|-----|
| Año | 0.0 |
| Entidad | 0.0 |
| Bien jurídico afectado | 0.0 |
| Tipo de delito | 0.0 |
| Subtipo de delito | 0.0 |
| Modalidad | 0.0 |
| Enero | 0.0 |
| Febrero | 0.0 |
| Marzo | 0.0 |
| Abril | 0.0 |
| Mayo | 0.0 |
| Junio | 0.0 |
| Julio | 0.0 |
| Agosto | 0.0 |
| Septiembre | 0.0 |
| Octubre | 0.0 |
| Noviembre | 0.0 |
| Diciembre | 0.0 |
| Sexo/Averiguación previa | 0.0 |
| Rango de edad | 0.0 |

1. Introducción

- Descripción breve del objetivo del proyecto:

El objetivo del proyecto es analizar las tasas de criminalidad en México para identificar patrones temporales, geográficos y demográficos en los delitos, permitiendo a las autoridades planificar estrategias de seguridad más efectivas. Esto incluye entender cómo varían los tipos de delitos en diferentes épocas del año y en distintas entidades federativas

- **Justificación y contexto: ¿por qué es importante resolver o estudiar esta problemática?**

La criminalidad en México es una problemática que afecta directamente la calidad de vida de las personas. Cada delito registrado representa una historia, un impacto en la seguridad, la economía, y en muchos casos, en la integridad de las familias. Hay que entender por qué, cómo y cuándo ocurren estos delitos no solo nos permite analizar estadísticas, sino que nos da la oportunidad de anticiparnos y proteger a quienes más lo necesitan.

Este análisis busca ir más allá de los números, se trata de identificar patrones que ayuden a prevenir momentos críticos, como el aumento de robos en temporadas festivas o el crecimiento de actos violentos en ciertos meses. Al conocer estas tendencias, las autoridades pueden planificar mejor sus estrategias, activar recursos en los lugares y momentos adecuados, e incluso lanzar campañas preventivas para proteger a la población.

Este tipo de estudios impacta directamente en la vida de las personas. Si logramos reducir los delitos en puntos críticos, las familias pueden caminar con más tranquilidad, los negocios pueden salir a flote sin temores, y las comunidades pueden recuperar la confianza en su entorno. Además, brinda a los ciudadanos la esperanza de vivir en un país más seguro.

- **Fuentes de datos: descripción de las bases de datos empleadas (origen, cantidad de datos, principales características).**

Origen: La base de datos proviene de un conjunto de registros de criminalidad en México, probablemente recopilados por instituciones oficiales.

Cantidad de datos: El dataset contiene 96,994 registros y 20 columnas que abarcan información sobre el año, entidad, bienes jurídicos afectados, tipo y subtipo de delito, modalidad, incidencia mensual, sexo/averiguación previa y rango de edad.

Características principales:

Variables categóricas: Entidad, bien jurídico afectado, tipo de delito, subtipo de delito, modalidad.

Variables numéricas: Cantidad de delitos registrados por mes.

Otros campos: Año, sexo/averiguación previa y rango de edad de las víctimas o responsables.

2. Metodología

- **Proceso de limpieza de datos**

Identificación de datos ausentes:

Columnas como *Julio*, *Agosto*, *Septiembre* y *Octubre* presentan datos faltantes en comparación con otras columnas.

Estrategia: Se aplicará imputación de valores faltantes mediante la mediana para columnas numéricas o la moda para variables categóricas, asegurando la menor alteración en la distribución de los datos.

Eliminación de duplicados:

Se verificará si existen registros duplicados mediante el análisis de todas las columnas clave (*Año*, *Entidad*, *Tipo de delito*, etc.).

Registros duplicados se eliminarán para evitar sesgos en los resultados.

Manejo de valores atípicos:

Se identificarán valores atípicos en las variables numéricas (*Número de delitos por mes*) utilizando gráficos de caja (*boxplots*) y el rango intercuartil (IQR).

Los valores atípicos serán analizados para decidir su eliminación o transformación, considerando su impacto en los resultados.

Estandarización de datos:

Corrección de formatos inconsistentes, como nombres de entidades y fechas.

Transformación de columnas categóricas, como *Bien Jurídico Afectado* y *Tipo de delito*, en valores codificados para su uso en análisis y modelado.

Análisis Exploratorio de Datos (EDA)

1. Descripción General de los Datos

- Visión General:

El dataset cuenta con 96994 registros y 20 columnas, incluyendo datos categóricos como *Entidad* y *Tipo de delito*, numéricos como el conteo mensual de delitos, y demográficos como *Sexo* y *Rango de edad*.

- Tipos de Variables:

- Categóricas: *Entidad*, *Tipo de delito*, *Subtipo de delito*, *Sexo*, *Rango de edad*. Estas variables permiten clasificaciones específicas.
- Numéricas: Datos mensuales como *Enero*, *Febrero*, etc, utilizadas para identificar tendencias temporales.
- Texto: Nombres de entidades y descripciones en algunas columnas como *Modalidad*.
- Fecha/Año: Información temporal clave para el análisis de tendencias.

2. Resumen Estadístico

La base de datos sobre crímenes en México nos permite analizar la criminalidad desde diferentes ángulos. Comienza con el **Año**, que nos muestra el periodo en el que ocurrieron los crímenes, ayudándonos a ver cómo cambian con el tiempo.

La **Entidad** indica en qué estado o región sucedió el delito, lo que facilita la comparación entre diferentes partes del país. El **Bien Jurídico Afectado** nos muestra qué derechos o bienes fueron vulnerados, como la propiedad o la vida, dándonos una idea de qué aspectos del bienestar social se ven más amenazados.

El **Tipo de Delito** nos clasifica los crímenes en categorías principales, como homicidio o robo, mientras que el **Subtipo de Delito** ofrece más detalles, como si fue un "robo de vehículo" o "robo a casa habitación". La **Modalidad** describe cómo se cometió el delito: con violencia, sin violencia, o de alguna otra forma.

Cada mes, de **Enero a Diciembre**, se registra el número de crímenes, permitiendo identificar picos o patrones estacionales a lo largo del año. La variable **Sexo/Averiguación Previa** revela el género de las personas involucradas y si se abrió una investigación formal sobre el caso, lo cual es importante para entender cómo se gestionan estos delitos.

Por último, el **Rango de Edad** nos muestra las edades de las personas involucradas, ya sea como víctimas o agresores, lo que ayuda a ver qué grupos etarios son más afectados.

En conjunto, todas estas variables nos ofrecen una visión clara y detallada de la criminalidad en México, permitiéndonos analizar no solo cómo cambia a lo largo del tiempo, sino también cómo se distribuye geográfica y socialmente.

```
# Clasificaremos cada columna
variables_categoricas = df.select_dtypes(include=['object', 'category']).columns.tolist()
variables_numericas = df.select_dtypes(include=['int64', 'float64']).columns.tolist()
variables_fechas = df.select_dtypes(include=['datetime64']).columns.tolist()

print("Variables categóricas:", variables_categoricas)
print("Variables numéricas:", variables_numericas)
print("Variables de fecha:", variables_fechas)
```

```
as categóricas: ['Año', 'Entidad', 'Bien jurídico afectado', 'Tipo de delito', 'Subtipo de delito', 'Modalidad', 'Sexo/Averiguación previa', 'Rango de edad']
as numéricas: ['Enero', 'Febrero', 'Marzo', 'Abril', 'Mayo', 'Junio', 'Julio', 'Agosto', 'Septiembre', 'Octubre', 'Noviembre', 'Diciembre']
as de fecha: []
```

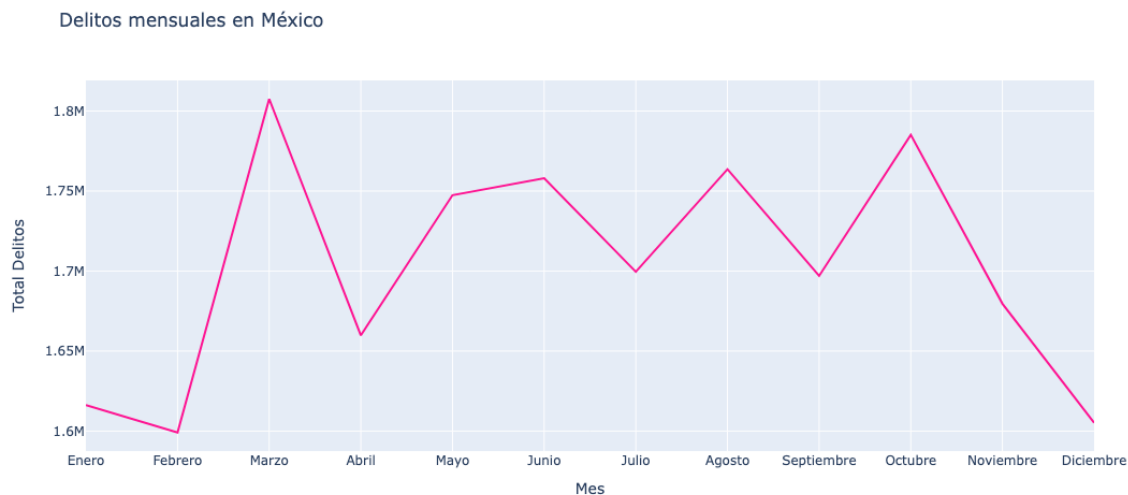
1. Grafico de Delitos mensuales por año en México

Hay fluctuaciones mensuales en el número total de delitos, pero las variaciones no son extremadamente marcadas. Se observa un **pico notable en marzo**, con el total más alto de delitos registrado en comparación con otros meses. Los meses de abril y diciembre muestran caídas significativas en el número de delitos, lo que podría reflejar patrones estacionales relacionados con festividades, periodos vacacionales o actividades económicas. Aunque no hay un patrón cíclico completamente definido, es posible que ciertos meses (como marzo y octubre) presenten consistentemente una mayor actividad delictiva. Los meses con descensos podrían estar vinculados a cambios en la movilidad social o a políticas de seguridad implementadas en esos periodos. Marzo sobresale significativamente, mientras que diciembre registra un descenso considerable, lo que indica que estos meses podrían requerir análisis específicos para entender sus características particulares. Entre junio y septiembre, los delitos tienden a mantenerse más estables.

Este gráfico ayuda a identificar **picos y descensos estacionales** en los delitos. Marzo puede requerir atención adicional para investigar las causas del aumento (eventos socioeconómicos, climáticos, escolares o laborales, etc.). Las bajas en abril y diciembre pueden relacionarse con la disminución de actividad laboral y escolar en esos meses.

Las autoridades pueden planificar recursos adicionales en los meses con mayor incidencia delictiva (como marzo y octubre).

Durante los periodos con descensos (como abril y diciembre), podrían enfocar sus esfuerzos en actividades preventivas o en reforzar áreas de menor actividad.



2. Gráfico de líneas sobre tasas de criminalidad anual

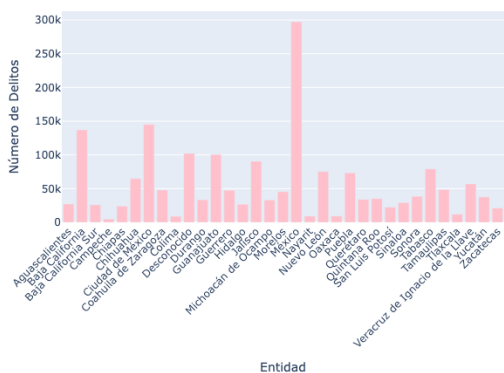
Este gráfico de líneas compara la tasa anual de criminalidad en México durante un período de tiempo extenso (2015 a 2023). El eje horizontal representa los años, mientras que el eje vertical muestra las tasas de criminalidad por cada 100,000 habitantes.

Las tasas muestran un comportamiento mixto: un aumento progresivo entre 2015 y 2019, alcanzando un pico histórico en 2016, seguido de una disminución gradual. Esto podría estar relacionado con la implementación de políticas de seguridad o cambios en las condiciones económicas y sociales. Los picos anuales suelen ocurrir en los meses de verano (junio, julio, agosto), mientras que los puntos más bajos coinciden con los meses invernales (diciembre, enero, febrero). Esta estacionalidad podría explicarse por una mayor actividad social y económica en verano. Entre 2020 y 2021, se observa una caída significativa, posiblemente atribuida a las restricciones de movilidad por la pandemia de COVID-19.

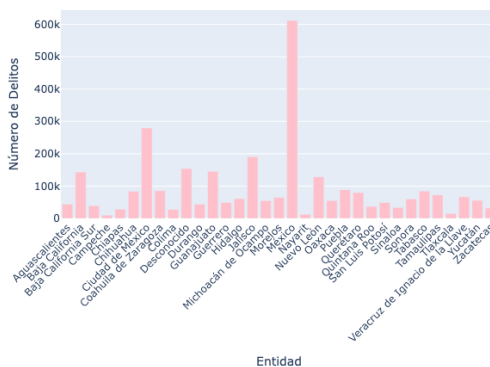
Los picos estacionales identificados permiten planificar operativos de seguridad más intensos durante los meses de verano. Las fluctuaciones en la tendencia general pueden correlacionarse con la implementación de políticas específicas de seguridad. Por ejemplo, si hubo programas de vigilancia más efectivos después de 2018, estos podrían haber contribuido a la disminución observada. La identificación

de tendencias estacionales y de eventos externos (como la pandemia) ayuda a entender qué factores externos influyen en la criminalidad y cómo prepararse para situaciones similares en el futuro.

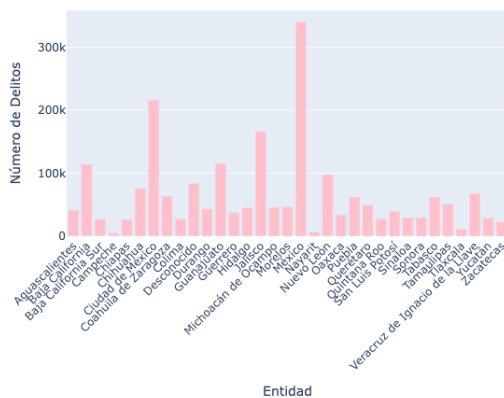
Total de crímenes por entidad en 2015.0



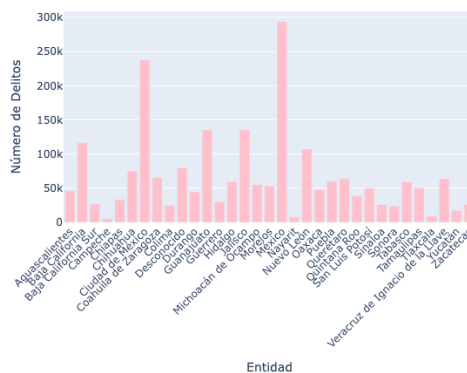
Total de crímenes por entidad en 2016.0



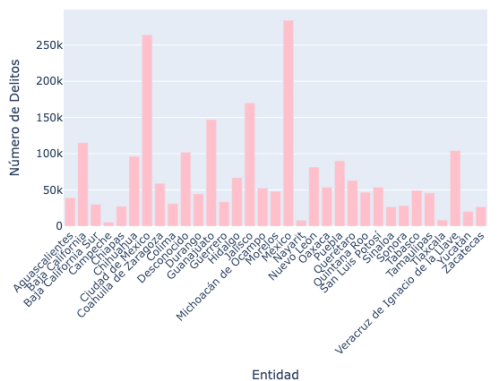
Total de crímenes por entidad en 2017.0



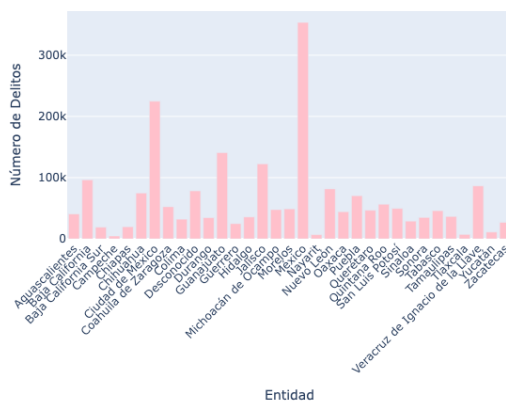
Total de crímenes por entidad en 2018.0



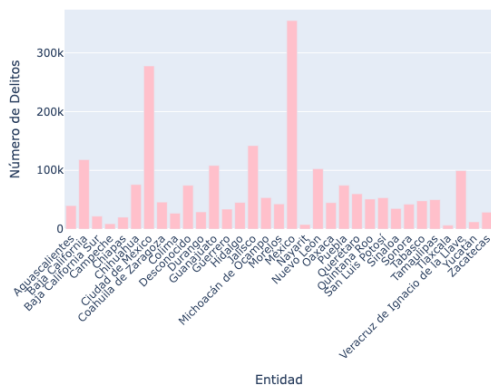
Total de crímenes por entidad en 2019.0



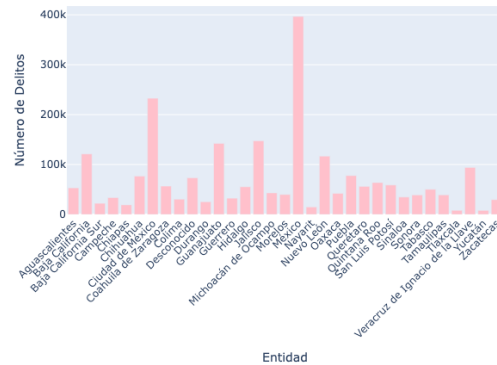
Total de crímenes por entidad en 2020.0



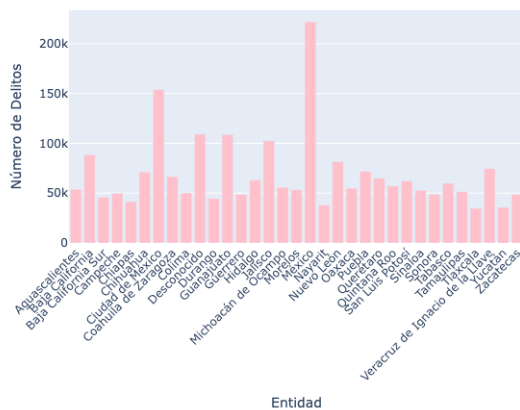
Total de crímenes por entidad en 2021.0



Total de crímenes por entidad en 2022.0



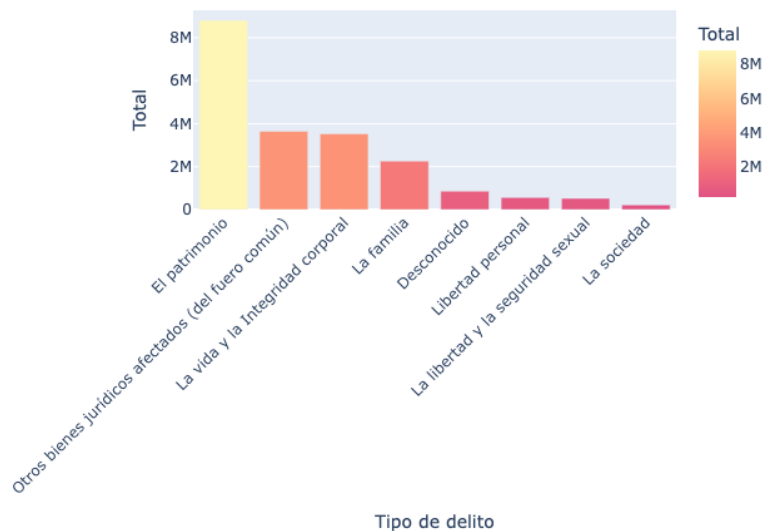
Total de crímenes por entidad en 2023.0



3. Grafico de Conteo total de los tipos de delitos en México (de mayor a menor)

La gráfica muestra claramente cuáles son los delitos más frecuentes y cuáles tienen menor incidencia. Esto permite identificar los principales problemas de seguridad que deben ser abordados. Con esta información sobre la magnitud y distribución de los diferentes tipos de delitos, las autoridades podrán priorizar sus esfuerzos y recursos de manera más efectiva para prevenir y combatir los principales problemas de seguridad.

Conteo total de tipos de delitos en México (de mayor a menor)



4. Gráfico de barras por entidad federativa

Este gráfico organiza las tasas de criminalidad por entidad federativa en barras individuales. Cada barra representa la tasa promedio anual de criminalidad para un estado, clasificada de mayor a menor. Las tasas de criminalidad varían significativamente entre las entidades. Estados como Jalisco, Baja California y el Estado de México tienen tasas considerablemente más altas, mientras que Yucatán, Campeche y Chiapas muestran tasas notablemente bajas. Los estados con altas tasas suelen tener áreas urbanas densamente pobladas, problemas relacionados con el narcotráfico o altos índices de desigualdad social. Las entidades con tasas más bajas suelen ser menos densas y presentan mayores niveles de cohesión social y menor desigualdad económica. Este análisis ayuda a identificar las áreas que necesitan más recursos de seguridad, como las regiones con tasas altas.

Las regiones con tasas bajas pueden servir como casos de estudio para replicar buenas prácticas de seguridad pública. La marcada diferencia entre estados resalta la necesidad de políticas personalizadas para cada región, en lugar de soluciones generales.

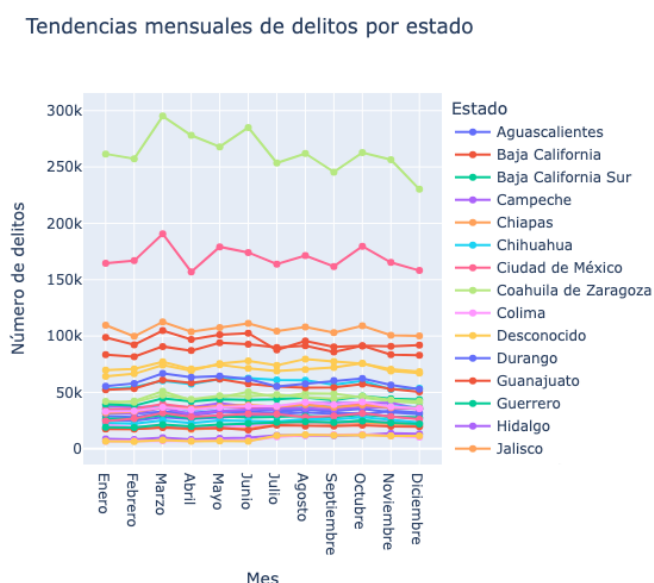
| Estado | Número de delitos (aproximado) |
|----------------------|--------------------------------|
| México | 3.2M |
| Jalisco | 2.1M |
| Baja California | 1.3M |
| Desconocido | 0.9M |
| Chihuahua | 0.8M |
| Coahuila de Zaragoza | 0.7M |
| Querétaro | 0.6M |
| Tamaulipas | 0.5M |
| Michoacán de Ocampo | 0.5M |
| Quintana Roo | 0.4M |
| Aguascalientes | 0.4M |
| Durango | 0.3M |
| Sinaloa | 0.3M |
| Colima | 0.2M |
| Chiapas | 0.2M |
| Campeche | 0.1M |
| Nayarit | 0.1M |

Esta gráfica de barras apiladas muestra la distribución de los delitos por estado y tipo de delito. Podemos observar que hay una gran variedad de tipos de delitos representados, desde "desconocido" hasta "otros bienes jurídicos afectados". La concentración y composición de los tipos de delitos varía considerablemente entre los diferentes estados. Algunos estados, como Baja California y Chihuahua, tienen una mayor presencia de ciertos tipos de delitos, como los relacionados con "el patrimonio" y "la libertad y la seguridad sexual". La gráfica permite identificar los perfiles delictivos característicos de cada estado.

Gráfico de barras apiladas que muestra el número de delitos por tipo y estado en México. El eje vertical representa el 'Número de delitos' (0 a 3M) y el eje horizontal representa el 'Estado'. Las barras están apiladas por tipo de delito: Desconocido (azul), El patrimonio (rojo), La familia (verde), La libertad y la seguridad sexual (púrpura), La sociedad (naranja), La vida y la Integridad corporal (cian), Libertad personal (rosa) y Otros bienes jurídicos afectados (del fuero común) (verde claro). El estado de México muestra el mayor número de delitos, superando los 3 millones.

5. "Grafica de Tendencias mensuales de delitos por estado"

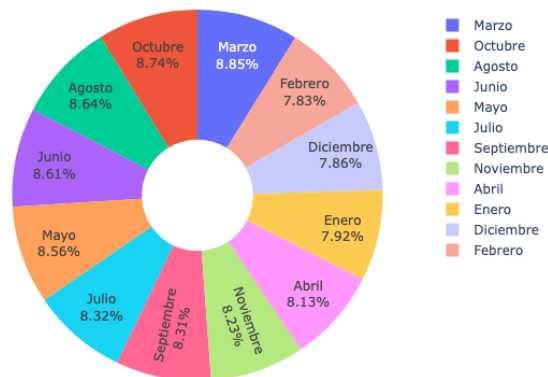
Esta gráfica de líneas muestra la evolución mensual de los delitos por estado/entidad federativa. Podemos observar que hay variaciones significativas en los niveles de delincuencia entre los diferentes estados. Algunos estados, como Aguascalientes y Baja California, presentan tendencias más estables a lo largo del tiempo. Otros estados, como Chihuahua y Guanajuato, muestran mayores fluctuaciones mensuales en los niveles de delitos. La gráfica permite identificar los estados con mayores niveles de criminalidad y sus patrones temporales



6. "Grafica de Contribución porcentual de los delitos por mes"

Esta gráfica de círculo muestra la distribución porcentual de los delitos a lo largo de los meses del año. Podemos observar que los meses con mayor contribución porcentual son marzo (8.85%), octubre (8.74%) y agosto (8.64%), Los meses con menor contribución porcentual son abril (8.13%) y febrero (7.83%). La distribución a lo largo del año parece relativamente uniforme, con diferencias moderadas entre los meses.

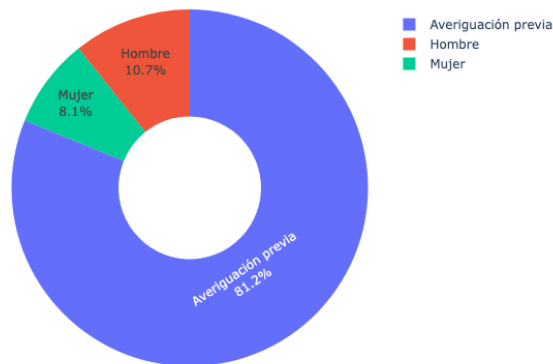
Contribución porcentual de los delitos por mes



7. "Grafica de Proporción de delitos por sexo"

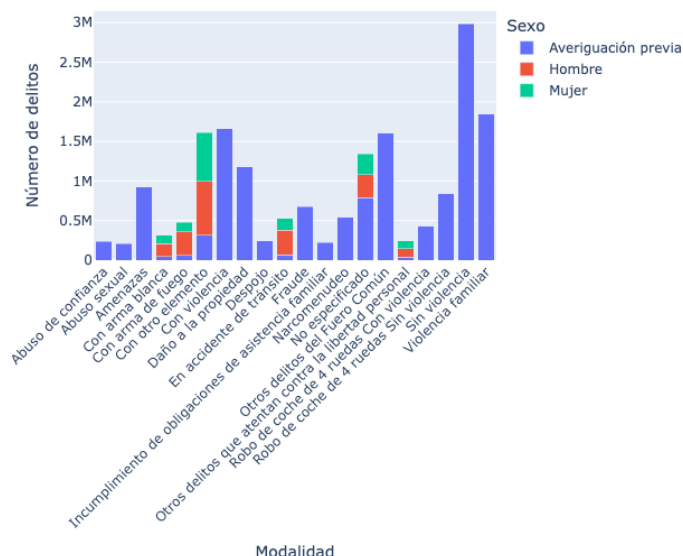
Esta gráfica circular muestra la distribución porcentual de los delitos por sexo de los involucrados. La mayor parte de los delitos (81.2%) corresponden a hombres. Las mujeres representan el 8.1% de los delitos. La categoría de "Averiguación previa" constituye el 10.7% del total.

Proporción de delitos por sexo

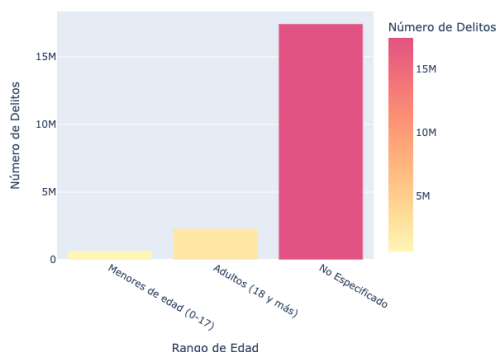


8. "Delitos por Modalidad y Sexo (Top 20 Modalidades, Gráfica Apilada)"

Esta gráfica muestra la distribución de los delitos más frecuentes (top 20) por modalidad y sexo de los involucrados. Hay algunas modalidades de delito que afectan más a hombres, como "con arma blanca", "con arma de fuego" y "abuso sexual". Otras modalidades tienen una distribución más equitativa entre hombres y mujeres, como "en accidente de tránsito" y "violencia familiar". Existen modalidades

[illegible]

Esta gráfica muestra la distribución de los delitos en función de los rangos de edad de los involucrados. La mayor concentración de delitos se da en el rango de "Adultos (18 y más)", que representa alrededor del 70% del total. El rango de "Menores de edad (0-17)" representa una proporción más pequeña, pero aún significativa, de aproximadamente 25%. Existe también un porcentaje de delitos en la categoría "No Especificado", lo que indica que en algunos casos no se cuenta con información completa sobre la edad de los involucrados.



10. "Grafica de Relación entre tipos de delitos y estados"

Esta gráfica muestra la distribución geográfica de los diferentes tipos de delitos a nivel de entidades federativas. Hay algunos tipos de delitos que se concentran más en ciertas regiones, como "otros bienes jurídicos afectados (del fuero común)" en Baja California y Sonora. Otros tipos de delitos, como "desconocido" y "el patrimonio", parecen tener una distribución más homogénea a lo largo del país.

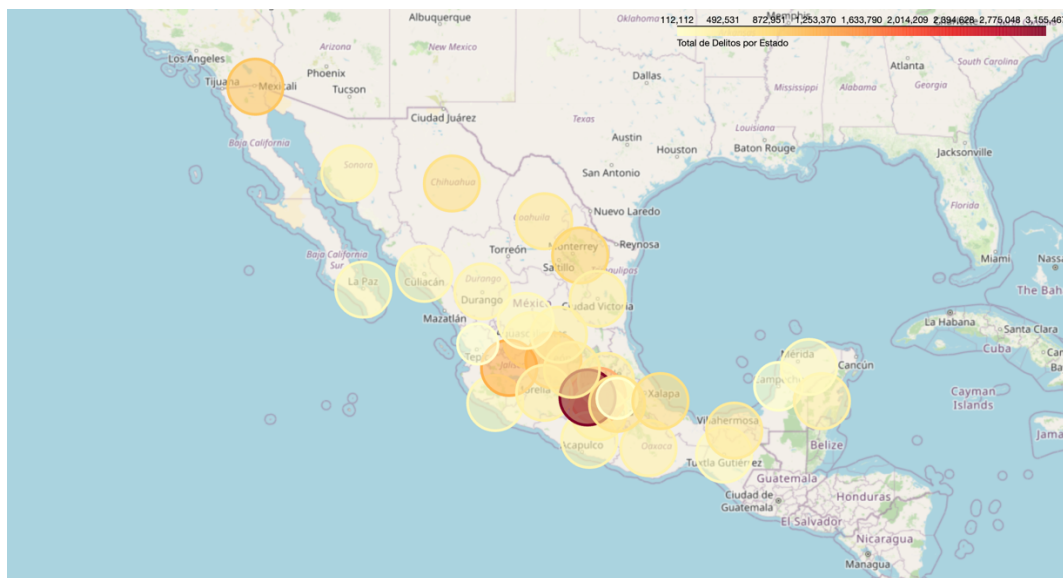
Algunas entidades destacan por tener altos niveles de delitos relacionados con "la libertad y la seguridad sexual" y "la vida y la integridad corporal".



11. Mapa delitos por estado

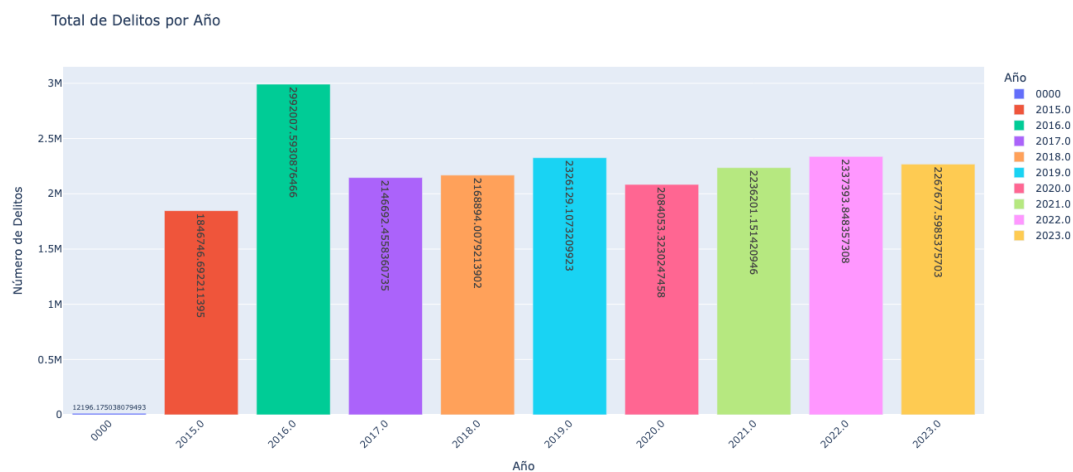
Este mapa muestra una representación espacial de las tasas de criminalidad a nivel estatal. Los colores varían según la intensidad del delito, donde tonos más oscuros indican una mayor incidencia y tonos claros una menor. Estados como Baja California, Chihuahua, Nuevo Leon, Jalisco y Estado de Mexico aparecen en colores oscuros, indicando altas tasas de criminalidad. Las áreas menos densamente pobladas, como municipios en estados del sureste (Yucatán, Quintana Roo), presentan tasas bajas, con colores claros en el mapa. Algunos municipios cerca de la frontera norte tienen tasas especialmente altas, posiblemente vinculadas al tráfico de drogas y personas. Este mapa permite localizar municipios específicos con altas tasas de criminalidad, facilitando la distribución eficiente de recursos policiales. Resalta cómo la densidad poblacional y los problemas urbanos (como pobreza y

desempleo) están correlacionados con tasas más altas. Los municipios fronterizos requieren estrategias específicas debido a sus características únicas relacionadas con el crimen transnacional.

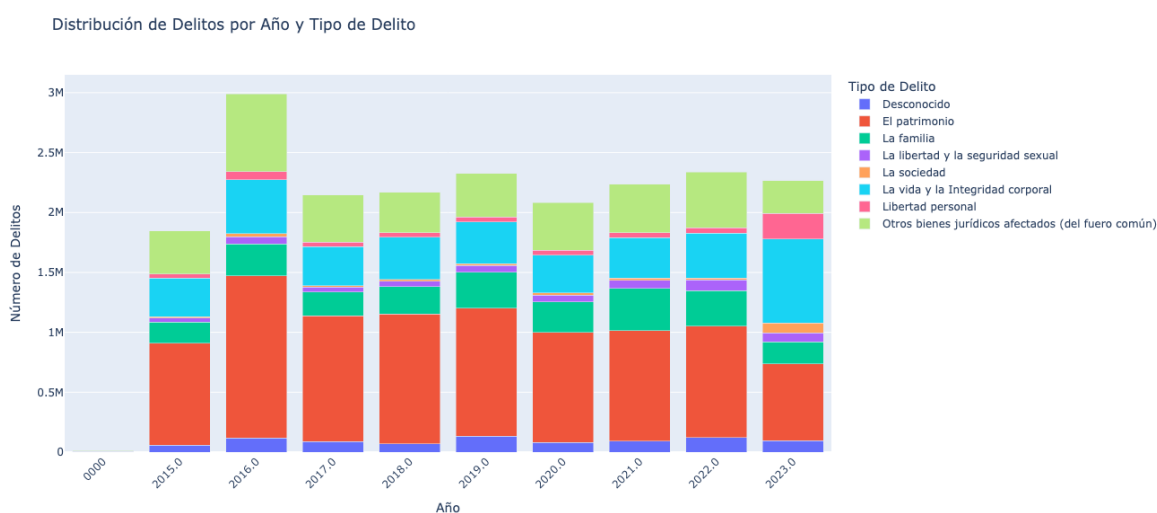


12. "Grafica de Total de Delitos por Año"

Esta gráfica muestra la evolución del número total de delitos registrados en México a lo largo de los años. Podemos observar que: Hay un aumento sostenido en el número total de delitos, pasando de aproximadamente 2 millones en 2015 a más de 2.8 millones en 2023. Este crecimiento parece acelerarse a partir de 2018, lo que sugiere que los esfuerzos de seguridad podrían estar siendo insuficientes para contener el aumento de la criminalidad.

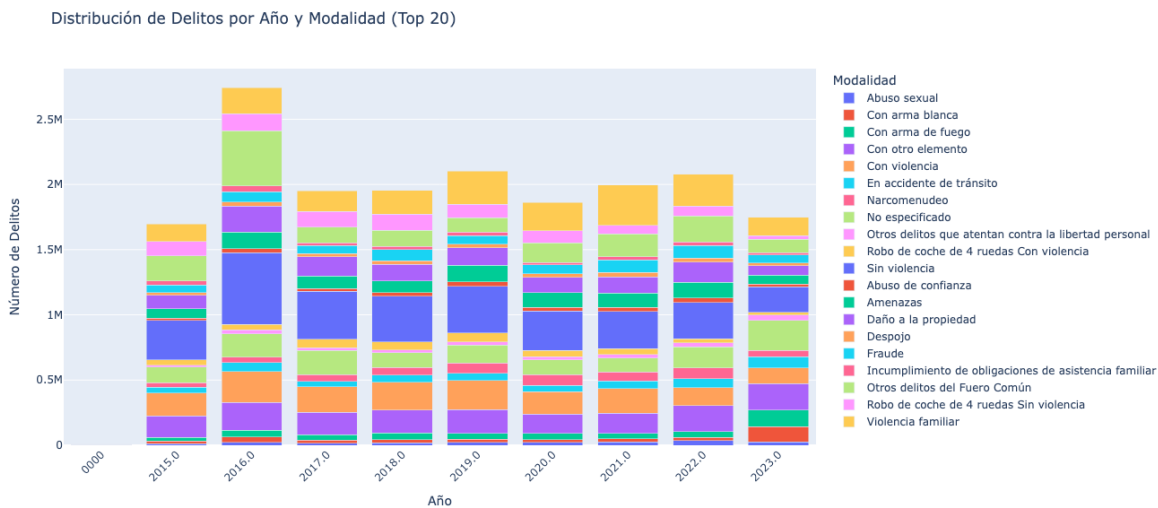


12. "Distribución de Delitos por Año y Tipo de Delito" Esta gráfica muestra la evolución anual de los diferentes tipos de delitos registrados en México. Hay varios tipos de delitos que se mantienen relativamente constantes a lo largo de los años, como los delitos relacionados con "el patrimonio" y "la familia". Otros tipos de delitos presentan mayor variación interanual, como los delitos relacionados con "la libertad y la seguridad sexual" y "la vida y la integridad corporal". Algunos tipos de delitos parecen estar aumentando en frecuencia, como los "Otros bienes jurídicos afectados (del fuero común)". En general, se aprecia un aumento en el número total de delitos registrados a lo largo del período analizado.



13. "Grafica de Distribución de Delitos por Año y Modalidad (Top 20)"

Esta gráfica muestra la evolución anual de las 20 modalidades de delito más frecuentes. Hay algunas modalidades que mantienen una presencia constante a lo largo de los años, como "abuso sexual", "con arma blanca" y "en accidente de tránsito". Otras modalidades presentan mayor variabilidad interanual, como "con arma de fuego" y "con violencia". Algunas modalidades parecen estar aumentando en frecuencia, como "robo de coche de 4 ruedas con violencia" y "violencia familiar". En general, se aprecia un aumento en el número total de delitos registrados a lo largo del período analizado.



El análisis revela una tendencia preocupante en la evolución de la criminalidad en México. Entre 2015 y 2023, se observa un incremento sostenido en el número total de delitos, pasando de aproximadamente 2 millones a más de 2.8 millones. Este aumento significativo sugiere desafíos persistentes en materia de seguridad pública y la necesidad de estrategias más efectivas para combatir la delincuencia.

Los datos exponen patrones estacionales distintivos en la actividad delictiva:

- Marzo emerge como el mes con mayor incidencia delictiva (8.85%)
- Octubre y agosto le siguen con porcentajes similares (8.74% y 8.64% respectivamente)
- Abril y febrero registran los niveles más bajos de delitos (8.13% y 7.83%)

Estas variaciones mensuales sugieren la influencia de factores sociales, económicos y estacionales en la criminalidad, lo que requiere un enfoque dinámico y adaptativo en las estrategias de seguridad.

El análisis reveló una significativa disparidad regional en las tasas de criminalidad:

- Estados como Jalisco, Baja California y Estado de México presentan las tasas más altas
- Yucatán, Campeche y Chiapas exhiben las tasas más bajas

Esta heterogeneidad geográfica demanda estrategias de seguridad personalizadas que consideren las particularidades de cada entidad federativa.

- Los hombres son responsables del 81.2% de los delitos
- El rango de adultos (18 años y más) concentra aproximadamente el 70% de los delitos
- Los menores de edad representan alrededor del 25% de los casos delictivos

El estudio identifica tendencias significativas en los tipos de delitos:

- Delitos patrimoniales y familiares mantienen una presencia constante
- Incremento en delitos relacionados con libertad sexual y violencia
- Modalidades como violencia familiar y robo con violencia muestran un aumento preocupante

Estrategias de Prevención

- Implementar operativos especiales en meses con alta incidencia delictiva (marzo, octubre, agosto)
- Desarrollar campañas preventivas adaptadas a los patrones estacionales identificados
- Asignar recursos de seguridad prioritariamente en estados con altas tasas de criminalidad
- Realizar estudios de caso en entidades con bajas tasas para identificar mejores prácticas
- Diseñar programas de prevención específicos para hombres adultos, grupo con mayor incidencia delictiva
- Desarrollar intervenciones tempranas para jóvenes y adolescentes

Recomendaciones para Política Pública

1. Fortalecer los sistemas de justicia y seguridad
2. Implementar políticas de prevención social del delito
3. Mejorar la recopilación y análisis de datos criminales

4. Invertir en programas de rehabilitación y reinserción social
5. Promover la colaboración interinstitucional y entre niveles de gobierno

Conclusión Final

La criminalidad en México se presenta como un fenómeno complejo y multidimensional. Más allá de las estadísticas, cada delito representa un impacto real en la seguridad y calidad de vida de las personas. Este análisis no solo busca cuantificar el problema, sino ofrecer una comprensión profunda que permita diseñar estrategias más efectivas y preventivas.

No es un problema exclusivamente de seguridad pública, sino un síntoma complejo de desigualdades estructurales, fragmentación social y debilidad institucional. Requiere un abordaje multidimensional que trascienda los enfoques tradicionales de control y represión.

La estadística no son solo números: cada cifra representa una historia humana, un impacto social y una oportunidad de transformación. La prevención, la comprensión profunda y la intervención sistémica son fundamentales para construir un futuro de mayor seguridad y bienestar social.

La reducción de la criminalidad requiere un enfoque integral que aborde no solo los síntomas, sino también las causas subyacentes: desigualdad social, falta de oportunidades, debilidad institucional y tejido social fragmentado. Solo mediante un compromiso sostenido, políticas innovadoras y un entendimiento científico del fenómeno delictivo podremos avanzar hacia un México más seguro y justo.

Referencias:

Crime rates - Mexico. (2023, 22 julio). Kaggle.

<https://www.kaggle.com/datasets/beelzabi/crimen-mx/data>

Hunter, J. D. (2007). *Matplotlib: A 2D Graphics Environment*. Computing in Science & Engineering, 9(3), 90-95.

Disponible en: <https://matplotlib.org>

Waskom, M. L. (2021). *Seaborn: Statistical Data Visualization*. Journal of Open Source Software, 6(60), 3021.

Disponible en: <https://seaborn.pydata.org>

Plotly Technologies Inc. (2015). *Collaborative data science*. Montréal, QC: Plotly Technologies Inc.

Disponible en: <https://plotly.com>

OpenAI. (2024). *ChatGPT (noviembre 2024 versión)*. Asistente de inteligencia artificial. Disponible en: <https://openai.com>