

ANALÍTICA Y CIENCIA DE DATOS

PROYECTO FINAL

**“ANÁLISIS SOBRE LA INCIDENCIA DELICTIVA EN LOS ESTADOS DE LA
REPÚBLICA MEXICANA”**

EQUIPO 5

INTEGRANTES:

JAVIER ISMAEL SAMPABLO CUEVAS

RIGOBERTO TORRES CISNEROS

ANA ISABEL HERNÁNDEZ DÍAZ

FELIPE DE JESÚS ARREOLA GARCÍA

MARIANNE ELVIRA HERNÁNDEZ HERNÁNDEZ

DELIA PAULINA GUTIÉRREZ ZAMORA

1. Selección y Justificación de la Base de Datos

Actualmente cómo población nos enfrentamos a una grave crisis de inseguridad que afecta a la totalidad de la República Mexicana generando la percepción colectiva de un continuo incremento de los niveles de violencia provocando intranquilidad y zozobra al recorrer espacios públicos. Cuando hablamos de inseguridad y violencia, nos referimos a diversas acciones que se clasifican en tipos y subtipos de delitos que al ser cometidos dejan una afectación o daño a derechos personales y colectivos, a los cuales se les conoce por el concepto de bien jurídico, pues representan un valor protegido por el orden jurídico.

Con el objetivo de conocer con mayor detalle los tipos de delitos cometidos en todas las entidades del país, se seleccionó una base de datos proporcionada por el Secretariado Ejecutivo del Sistema Nacional de Seguridad Pública que brinda una clasificación de delitos cometidos por estado desde el año 2015 al 2024, así mismo, nos aporta datos sobre los bienes jurídicos afectados en cada delito.

Con lo anterior, se pretende desarrollar un análisis exploratorio para resolver interrogantes como cuál es el estado con mayor incidencia delictiva o conocer cuál es el tipo de delito más frecuente en cada entidad, esto con la finalidad de tener más conocimiento respecto a los índices de criminalidad en el país, pues al dividirse por estados y tipo de delito, será posible saber las afectaciones particulares de cada uno.

2. Preparación y Limpieza de los Datos

2.1.1. Descripción de la Base de Datos

- Número de registros y variables/campos:

La base de datos está conformada por un total de 5 tablas, de las cuales, hay dos tablas importantes, que son las que tienen la información general sobre los delitos, en las que tenemos la tabla que tiene el dato por entidad, la cual tiene 31,360 registros, y otra tabla con el dato por municipio, la cual tiene 2,319,072 registros.

Estas dos tablas están conformadas por cerca de 20 columnas cada una, en las que tenemos información sobre los delitos, subtipos de delito, el bien jurídico afectado, la modalidad, es decir, la forma específica en que el delito fue cometido, la entidad o municipio, la clave de dicha entidad, y el año en que se cometieron los delitos, los 12 restantes campos corresponden a los 12 meses del año.

- Tipos de datos (numéricos, categóricos, fechas, etc):

Los datos presentados son numéricos, pues son la representación de las cifras de los delitos cometidos. Sin embargo, también son categóricos, ya que al existir distintos tipos de delitos, estos se dividen en subtipos para una mejor exposición del tema. Además, se utilizan fechas para ubicar la temporalidad en que fueron realizados, empleando un rango de tiempo desde 2015 a 2024.

Numérico	Categóricos	Fechas
Clave Ent	Bien jurídico afectado	Año
Enero	Tipo de delito	
Febrero	Subtipo de delito	
Marzo	Entidad	
Abril	Modalidad	
Mayo		
Junio		
Julio		
Agosto		
Septiembre		
Octubre		
Noviembre		
Diciembre		

2.1.2. Limpieza de Datos

- Identificación y manejo de valores faltantes

Al manipular la base de datos, se pudo observar que era necesario añadir columnas de identificación (ID) para un mejor manejo de los datos, debido a que, el análisis está centrado en los 32 estados de la República Mexicana junto con 7 categorías principales de delitos, dando un total de 20 columnas y 31,360 registros para las entidades, y 2,19,702 registros para la información por municipios.

- Corrección de errores en los datos

Los datos muestran coherencia entre sí y cada cifra es relevante con las otras, por lo que se optó por añadir los ID, y unas correcciones menores a los datos.

Un cambio que se aplicó para las tablas, fue el corregir datos nulos a numéricos en los meses de incidencia, para esto primero analizamos la información y observamos que era un detalle a corregir.

```
UPDATE aa_idafc_nm
SET enero = COALESCE(enero, 0),
    febrero = COALESCE(febrero, 0),
    marzo = COALESCE(marzo, 0),
    abril = COALESCE(abril, 0),
    mayo = COALESCE(mayo, 0),
    junio = COALESCE(junio, 0),
    julio = COALESCE(julio, 0),
    agosto = COALESCE(agosto, 0),
    septiembre = COALESCE(septiembre, 0),
    octubre = COALESCE(octubre, 0),
    noviembre = COALESCE(noviembre, 0),
    diciembre = COALESCE(diciembre, 0);
```

Otro punto a corregir, fue la parte de completar información para los estados, o corregir los nombres de los mismos.

```
-- Cambiar "Estado de México" a "México" (Hubo diferencias en los nombres, aqui esta la correccion)
UPDATE poblacion_estados
SET entidad = 'México'
WHERE entidad = 'Estado de México';

-- Cambiar "Michoacán" a "Michoacán de Ocampo"
UPDATE poblacion_estados
SET entidad = 'Michoacán de Ocampo'
WHERE entidad = 'Michoacán';

-- Cambiar "Veracruz" a "Veracruz de Ignacio de la Llave"
UPDATE poblacion_estados
SET entidad = 'Veracruz de Ignacio de la Llave'
WHERE entidad = 'Veracruz';

-- Cambiar "Coahuila" a "Coahuila de Zaragoza"
UPDATE poblacion_estados
SET entidad = 'Coahuila de Zaragoza'
WHERE entidad = 'Coahuila';
```

- Transformaciones necesarias

A cada estado y bien jurídico afectado se le asignó un ID, lo que nos permitió poder filtrar de forma más oportuna la información. Adicionalmente, la base de datos tuvo su origen en Excel, para ello, se añadieron pestañas para colocar el ID por identidad y por bien jurídico afectado. Por último, se añadió una columna que hace el recuento anual de las cifras, por lo que es posible conocer un total global por año.

3. Análisis Exploratorio de Datos (EDA)

3.1.1. Análisis Descriptivo:

El empleo de SQL como software de análisis nos permite realizar distintas consultas al combinar tablas y categorías, lo que da como resultado una exploración variada y detallada, pues al existir más de 30 mil registros la información es inmensa, por lo que de este modo se puede resumir y mostrar los datos más relevantes.

Tasa promedio del año 2023 por cada 100k habitantes.

Colima encabeza a las demás entidades en este sentido.

	entidad character varying (100)	num_años bigint	total_delitos bigint	poblacion integer	tasa_promedio_anual_por_100k numeric
1	Colima	1	28075	731391	3838.58
2	Baja California	1	121070	3769020	3212.24
3	Aguascalientes	1	42316	1466309	2885.89
4	Baja California Sur	1	22849	798447	2861.68
5	Quintana Roo	1	54327	1969957	2757.78

En los delitos separados por mes, existe un patrón donde se reducen la cantidad de estos en los meses más festivos del año, **en las fechas próximas a la navidad los meses anteriores y posteriores se denota una baja en la incidencia delictiva.**

	mes text	total_delitos bigint
1	Marzo	1720454
2	Mayo	1703792
3	Agosto	1694744
4	Junio	1683820
5	Julio	1652810
6	Abril	1611649
7	Octubre	1552611
8	Enero	1550730
9	Febrero	1548769
10	Septiembre	1487284
11	Noviembre	1449733
12	Diciembre	1390666

Los tipos de delito en representación porcentual son muy inconsistentes, más de la mitad de los tipos de delito (26) representan menos del 1% en la sumatoria de los delitos de todo el país. **El robo triplica la incidencia del segundo delito más común**, este sería el más grande problema que enfrenta la nación.

	tipo_delito character varying (255)	porcentaje_total_delitos numeric
1	Robo	34.0605968521549413
2	Violencia familiar	10.8212174665048079
3	Lesiones	10.4597233946106754
4	Otros delitos del Fuero Común	9.4310870621411323
5	Daño a la propiedad	6.7862802147648808
6	Amenazas	5.3669589567146891
7	Fraude	4.1511073991358877
8	Narcomenudeo	3.3295318721596013
9	Homicidio	2.0924329432014239
10	Despojo	1.3878413374199128
11	Abuso de confianza	1.3421755019225537
12	Incumplimiento de obligaciones de asistencia familiar	1.2258215991526672
13	Abuso sexual	1.1991980705475732
14	Otros delitos que atentan contra la libertad personal	1.0015455401993231
15	Delitos cometidos por servidores públicos	0.93718390794338780438
16	Falsificación	0.82279356259773817085
17	Allanamiento de morada	0.68495603153914236222
18	Violación simple	0.66919507060983998477
19	Otros delitos contra la familia	0.61670403550951847587
20	Otros delitos contra el patrimonio	0.58425283647420268806

Representación porcentual de la incidencia delictiva de cada entidad.

	entidad character varying (100)	porcentaje_total_delitos numeric
1	México	18.14
2	Ciudad de México	10.91
3	Jalisco	6.92
4	Guanajuato	6.51
5	Baja California	5.49
6	Nuevo León	4.39
7	Veracruz de Ignacio de la Llave	3.69
8	Chihuahua	3.52
9	Puebla	3.43
10	Coahuila de Zaragoza	2.80
11	Querétaro	2.71
12	Tabasco	2.65
13	Hidalgo	2.31
14	Morelos	2.27
15	San Luis Potosí	2.25
16	Michoacán de Ocampo	2.18
17	Tamaulipas	2.11
18	Quintana Roo	2.04
19	Oaxaca	1.82
20	Aguascalientes	1.77
21	Sonora	1.55
22	Guerrero	1.48
23	Durango	1.45
24	Sinaloa	1.34
25	Colima	1.17
26	Zacatecas	1.13
27	Baja California Sur	1.12
28	Chiapas	1.03
29	Yucatán	0.79
30	Campeche	0.42
31	Nayarit	0.32
32	Tlaxcala	0.27

Cálculo de media delictiva por Entidad del año 2023.

	media_delitos_2023 numeric
1	67922.156250000000

Cálculo de mediana delictiva por Entidad del año 2023.

	mediana_delitos_2023 double precision
1	45347.5

- Consultas:

```
-- Municipios con mas delitos porcentuales
SELECT
  municipio,
  ROUND(
    SUM(enero + febrero + marzo + abril + mayo + junio + julio + agosto + septiembre + octubre + noviembre + diciembre) * 100.0 / (
      SELECT SUM(enero + febrero + marzo + abril + mayo + junio + julio + agosto + septiembre + octubre + noviembre + diciembre)
      FROM delitos_municipales
    ), 2
  ) AS porcentaje_total_delitos
FROM delitos_municipales
GROUP BY municipio
ORDER BY porcentaje_total_delitos DESC;

-- Cálculo de la media de delitos por municipio
SELECT
  AVG(total_delitos) AS media_delitos
FROM (
  SELECT
    municipio,
    SUM(enero + febrero + marzo + abril + mayo + junio + julio + agosto + septiembre + octubre + noviembre + diciembre) AS total_delitos
  FROM delitos_municipales
  GROUP BY municipio
) AS subquery;

-- Cálculo de la mediana de delitos por municipio
SELECT
  PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY total_delitos) AS mediana_delitos
FROM (
  SELECT
    municipio,
    SUM(enero + febrero + marzo + abril + mayo + junio + julio + agosto + septiembre + octubre + noviembre + diciembre) AS total_delitos
  FROM delitos_municipales
  GROUP BY municipio
) AS subquery;

/* Años mas violentos (Exceptuando el año en curso, cada año consecutivo aumenta mas la cantidad de delitos a excepcion del año 2020
ya que se redujo la violencia aproximadamente a los del año 2016 Por lo tanto EL MAS VIOLENTO ES EL 2023 Y EL QUE MENOS 2015 */
SELECT
  año,
  SUM(enero + febrero + marzo + abril + mayo + junio + julio + agosto + septiembre + octubre + noviembre + diciembre) AS total_delitos
FROM delitos_municipales
GROUP BY año
ORDER BY total_delitos DESC;

/* Estados con mas delitos (Los 5 estados mas violentos son: Mexico, Ciudad de Mexico, Jalisco, Guanajuato y Baja California)
Los 5 estados que menos delitos tienen son: Tlaxcala, Nayarit, Campeche, Yucatan y Chiapas */
SELECT
  entidad,
  SUM(enero + febrero + marzo + abril + mayo + junio + julio + agosto + septiembre + octubre + noviembre + diciembre) AS total_delitos
FROM delitos_municipales
GROUP BY entidad
ORDER BY total_delitos DESC;

-- Proporción de delitos por entidad
SELECT
  entidad,
  ROUND(
    SUM(enero + febrero + marzo + abril + mayo + junio + julio + agosto + septiembre + octubre + noviembre + diciembre) * 100.0 / (
      SELECT SUM(enero + febrero + marzo + abril + mayo + junio + julio + agosto + septiembre + octubre + noviembre + diciembre)
      FROM delitos_municipales
    ), 2
  ) AS porcentaje_total_delitos
FROM delitos_municipales
GROUP BY entidad
ORDER BY porcentaje_total_delitos DESC;
```

```
-- Promedio de los datos en los 10 años
select round(avg(total_anual),0) from aa_idafc_nm ain ;

-- Informacion con el conteo total de incidentes por entidad, ordenando de mayor a menor por la cantidad de incidentes.
select e.entidad, count(*) as total_incidencias
from aa_idafc_nm ain
join entidades e ON e.id = ain.entidad_id
group by e.entidad
order by total_incidencias desc;
```

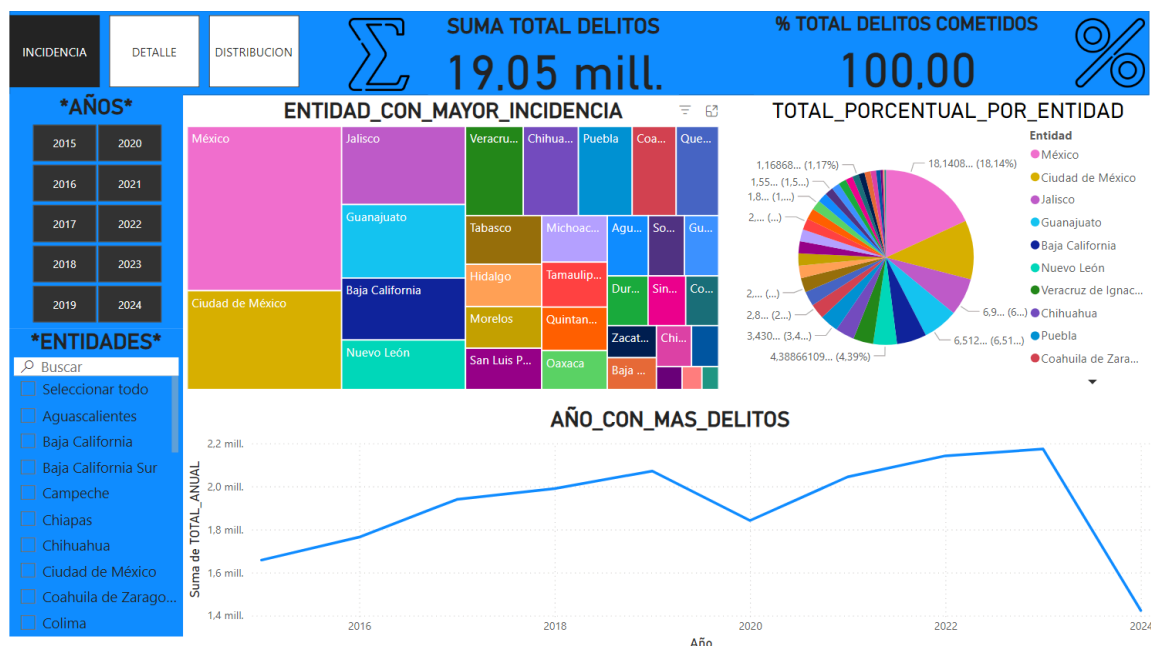
4. Visualización de Datos

4.1.1. Herramienta de Visualización:

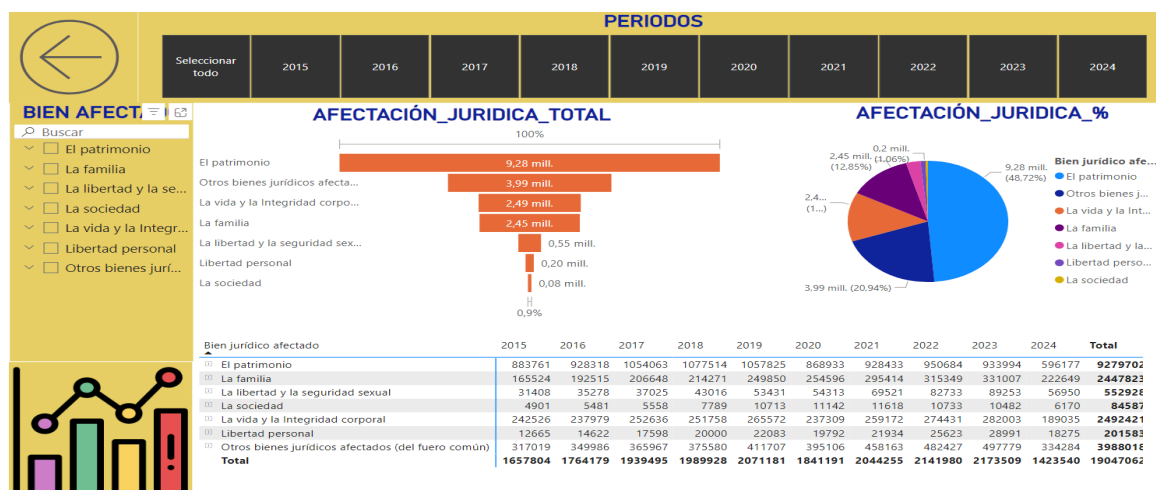
El software que se escogió fue Power BI. Esta decisión se tomó con base en la practicidad de la herramienta, además de poder visualizar de forma interactiva la exposición de los datos. Inicialmente, la base de datos se manejó en Excel, por lo que al pertenecer al mismo grupo de sistemas, fue muy fácil poder trasladar y plasmar la información.

4.1.2. Principales Visualizaciones:

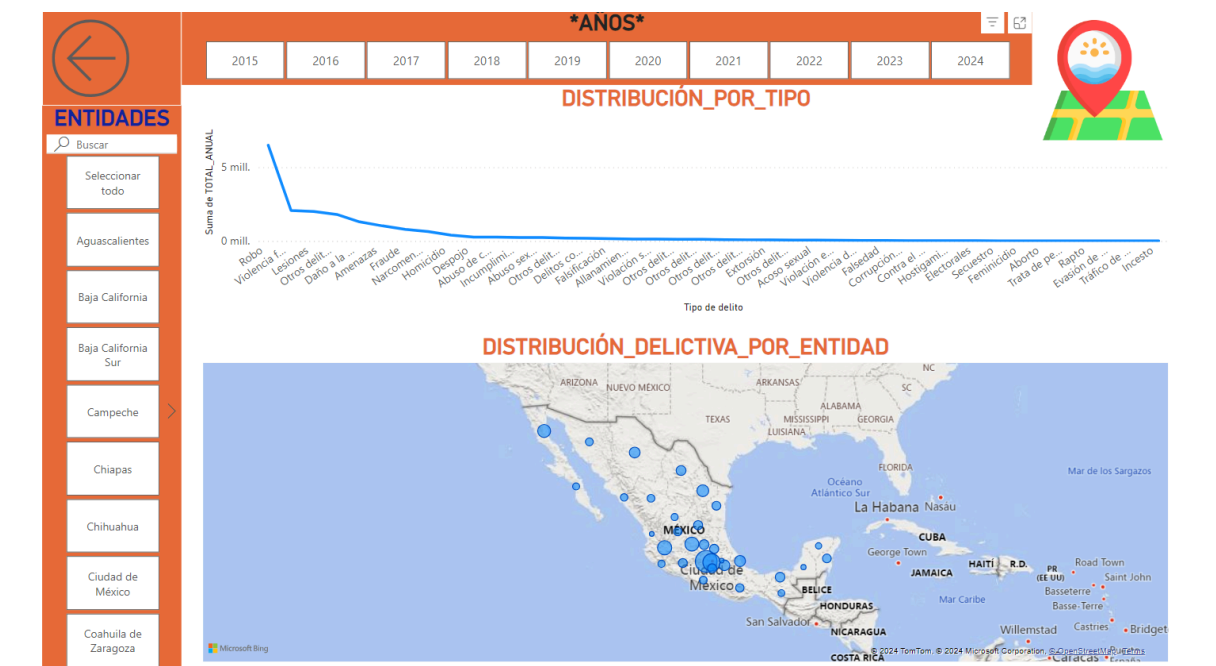
Tomando el período 2015-2024, se observa que la Ciudad de México es la entidad con mayor incidencia delictiva, seguido de Jalisco y posteriormente Guanajuato.



Tomando el período 2015-2024, se visualiza que, la principal afectación jurídica recae en el patrimonio con un 48.72%, en segundo lugar están otros bienes jurídicos (del fuero común) con un 20.94%, y en tercer lugar la vida e integridad corporal con un 13.09%.



De forma más particular, se encuentra un apartado que analiza los datos por entidad, dentro de ella se engloban los delitos y el porcentaje correspondiente al año en que se cometieron.



5. Interpretación y Conclusiones

5.1.1. Resultados del Análisis:

- Resumen de los hallazgos más importantes

La incidencia delictiva presenta diferencias marcadas entre estados, siendo algunas regiones más vulnerables a la violencia y otros mostrando niveles más bajos de criminalidad, dicha criminalidad en México es el resultado de una combinación de factores socioeconómicos, políticos y culturales. La alta incidencia delictiva afecta la percepción de seguridad y la calidad de vida de los ciudadanos, provocando miedo y desplazamientos forzados en las áreas más vulneradas.

- Interpretación de los resultados en el contexto del problema planteado

Los datos muestran que ciertos estados enfrentan niveles alarmantes de violencia, lo que sugiere una desigualdad en el acceso a recursos, oportunidades y seguridad. Esto refuerza la idea de que las soluciones deben ser adaptadas a las realidades locales. La percepción de inseguridad afecta el bienestar de la población y puede tener repercusiones económicas, como la reducción del turismo y la inversión. Esto resalta la necesidad de políticas que no solo aborden la delincuencia, sino que también mejoren la confianza de la comunidad en las instituciones.

5.2.1.Conclusiones:

- Conclusiones generales del proyecto

El análisis de datos desglosados por estado y tipo de delito revela patrones específicos que pueden guiar la formulación de políticas dirigidas y eficaces. Las estrategias deben ser personalizadas según las características delictivas de cada región.

El seguimiento de las tendencias en la incidencia delictiva a lo largo del tiempo permite identificar picos de criminalidad y evaluar la efectividad de las políticas implementadas. Esto es crucial para hacer ajustes oportunos en las estrategias de seguridad. La utilización de modelos predictivos basados en datos históricos puede ayudar a anticipar focos de criminalidad y permitir a las autoridades movilizar recursos de manera más eficiente en áreas de alto riesgo.

Evaluar el impacto de intervenciones pasadas a través de análisis de datos puede proporcionar información valiosa sobre qué estrategias han sido efectivas y cuáles no, permitiendo la mejora continua de las políticas de seguridad.

Los datos sugieren correlaciones significativas entre variables socioeconómicas y tasas de criminalidad, señalando que, mejorar indicadores como el ingreso, la educación y el empleo puede ser clave para reducir la delincuencia.

- Recomendaciones basadas en los resultados obtenidos

Realizar un análisis detallado y segmentado por estado y tipo de delito para identificar áreas críticas y desarrollar intervenciones específicas que aborden las particularidades de cada región.

Efectuar evaluaciones de impacto de las políticas y programas de seguridad a través del análisis de datos, identificando qué estrategias han sido más efectivas y cuáles requieren ajustes. Del mismo modo, invertir en tecnologías de información que faciliten la recopilación y análisis de datos en tiempo real, permitiendo una mejor gestión y respuesta ante situaciones de criminalidad.

Basar la formulación de políticas en evidencias derivadas del análisis de datos, garantizando que las decisiones se alineen con las necesidades reales y patrones delictivos identificados.

- Posibles limitaciones del análisis y sugerencias para trabajos futuros

La calidad y disponibilidad de datos pueden ser inconsistentes, lo que limita la precisión del análisis. Algunas regiones pueden tener registros incompletos o poco fiables; esto lo confirmamos en periodos donde hay registros en 0, o con datos muy bajos.

Los datos pueden tener categorías faltantes que ayuden a detallar los reportes, donde detallando estas, ayudará a tener estadísticas más amplias, que aporten a la lectura de la información.

Las diferencias en la clasificación y registro de delitos entre estados pueden dificultar comparaciones precisas y la interpretación de datos, además, la subestimación de delitos denunciados puede llevar a una percepción errónea de la incidencia delictiva, afectando la validez de los resultados, por ejemplo no podemos tener datos precisos de los delitos desde que existe una importante cifra negra de delitos que nunca son denunciados, así mismo en cuanto a la clasificación, podemos ver que hay estados donde hay muy pocos feminicidios y sin embargo hay una significativa cifra de homicidios dolosos de mujeres, lo cual puede abrir la posibilidad de que algunos de esos homicidios no se están clasificando correctamente.

Cómo recomendaciones se propone desarrollar un sistema estandarizado de recopilación y reporte de datos sobre criminalidad que facilite la comparación entre estados y asegure la calidad de la información así como combinar análisis cuantitativos con enfoques cualitativos, como entrevistas y encuestas a la comunidad, para obtener una perspectiva más amplia sobre la inseguridad y la criminalidad.