

CODERHOUSE

PROYECTO FINAL SQL

COMISIÓN: 53170

ALUMNO: FEDERICO EZEQUIEL MASSONE

DNI 35.674.907

PROFESOR: CAMILO ANDRÉS REDONDO

TUTOR ASIGNADO: NICOLÁS JAVIER OLIVERA

**TEMÁTICA DEL PROYECTO: BASE DE DATOS
HOMICIDIOS**



1. INTRODUCCIÓN

La enorme cantidad de asesinatos en el mundo entero obliga a reflexionar sobre la forma en la que vivimos. Sociedades cada vez más tecnificadas y globalizadas no parecieran constituir un avance a nivel social en cuanto a la violencia que siempre ha estado latente en la humanidad.

Sin embargo, poder determinar de forma fehaciente si el nivel de violencia ha ido mejorando o no, no es una tarea del todo sencilla. Es necesario poder cuantificar los hechos para luego sacar conclusiones en base a los datos recopilados. Es en este proceso de recopilación, almacenaje y correcto mantenimiento de la información obtenida donde se vuelve fundamental la puesta en funcionamiento y administración de una base de datos SQL.

Si se contara con suficiente información de calidad almacenada, a través de diversas consultas, se podría acceder a los datos de interés para desarrollar políticas públicas de prevención con el objetivo de mejorar los índices de asesinatos que asolan especialmente a las grandes ciudades.

Es por este motivo que para el desarrollo del presente trabajo se ha obtenido una base de datos sobre homicidios cometidos en Estados Unidos. Sobre esta se realizarán diversas consultas y acciones con el fin de obtener información filtrada que pueda resultar de interés para que el usuario pueda tomar decisiones. Estas decisiones deberán hacerse en base al análisis de los datos obtenidos, lo que permitirá concluir a partir de lo observado.

2. OBJETIVOS DEL PROYECTO

El objetivo general del proyecto es brindar al usuario la información y las herramientas necesarias, para que pueda realizar consultas eficientes y óptimas a una base de datos, que le permitan a posteriori, tomar decisiones de políticas públicas que tiendan a mejorar la calidad de vida de las personas. De este objetivo general se desprenden algunos objetivos específicos:

- Contribuir a la reducción de la criminalidad por homicidios a partir de políticas activas de prevención.
- Brindar un marco de referencia a las autoridades para la toma de decisiones.
- Exponer el grado de eficiencia de las agencias en cuanto a la resolución de casos.
- Detectar áreas de mayor incidencia en cuanto a cantidad de crímenes perpetrados.

3. TEMÁTICA DEL PROYECTO

La base de datos que se presenta contiene información sobre homicidios cometidos en Estados Unidos en el período 2006-2014. En la misma se pueden encontrar datos tales como la fecha de los homicidios, agencias involucradas en su resolución, rasgos de la víctima y el victimario, ubicación del lugar en el cual se perpetró el homicidio, arma utilizada en el crimen,

relación de la víctima al agresor y más campos que luego serán detallados en cada una de las tablas.

El interés en la elección del dataset se basa en la robustez de la información, además de ser potencialmente útil y atractivo para analizar en otra instancia, distinta a la del curso, los datos que contiene. Por otro lado, resulta especialmente adecuado ya que tanto, la cantidad de tablas como las instancias de cada una de ellas, proporcionan un escenario óptimo para aquello que se pretende lograr en el desarrollo del trabajo.

En otras palabras, la calidad de los datos, su cantidad, estructura e interés en el tema, justifican la elección de la base de datos a utilizar.

4. SITUACIÓN PROBLEMÁTICA

La cantidad de homicidios cometidos atentan contra la estabilidad del sistema social en cual estamos inmersos. Es importante poder resguardar cada vida humana para asegurar un estilo de vida con bajos índices de violencia, lo que permitirá mejorar la calidad de los años que nos quedan por delante. Asegurar esto es responsabilidad directa del Estado, pero este debe de disponer de las herramientas y la información adecuada para poder actuar.

A través de la presente base de datos se espera poder brindar la información necesaria a las autoridades para que detecten tendencias y comportamientos que le permitan tomar medidas de prevención para evitar futuros homicidios. Por medio de acciones tendientes a reducir los niveles de criminalidad, se mejorará la calidad de vida de las personas, logrando un entorno más seguro y apto para vivir.

5. MODELO DE NEGOCIO

Se espera poder ofrecer el siguiente proyecto a cualquier interesado que desee tener a disposición una base de datos consistente con los homicidios cometidos c en Estados Unidos en el período 2006-2014. A través de la misma quienes deseen invertir en el proyecto podrán tener acceso a información que les permitirá mejorar la gestión en seguridad del gobierno para el mandato que se encuentran desempeñando.

Si bien el destinatario principal es el sector público de Estados Unidos, los privados que trabajen con distintos niveles del gobierno, también podrían aprovecharse de la información para desarrollar proyectos que luego ellos puedan comercializar.

6. REPOSITORIO CON EL CONTENIDO DEL PROYECTO

La base de datos que se proporciona tiene el nombre de 'Homicidios'

Todos los archivos del proyecto pueden encontrarse en el repositorio "Curso_Coder_House_SQL_Homicidios". Link al repositorio:

https://github.com/Fedemass1/Proyecto_Final_Coder_House_SQL_Homicidios.git

7. DIAGRAMA ENTIDAD RELACIÓN DE LA BASE DE DATOS ‘HOMICIDIOS’

También conocido como Modelo E-R, es una herramienta que facilitará la comprensión y representación de la base de datos a través de las relaciones existentes entre las distintas entidades. Se presentará primero el diagrama completo (con todas sus tablas) y luego una versión sintética del mismo. Mientras que el primero es generado desde la herramienta MySQL Workbench 8.0 CE, el segundo es de elaboración propia.

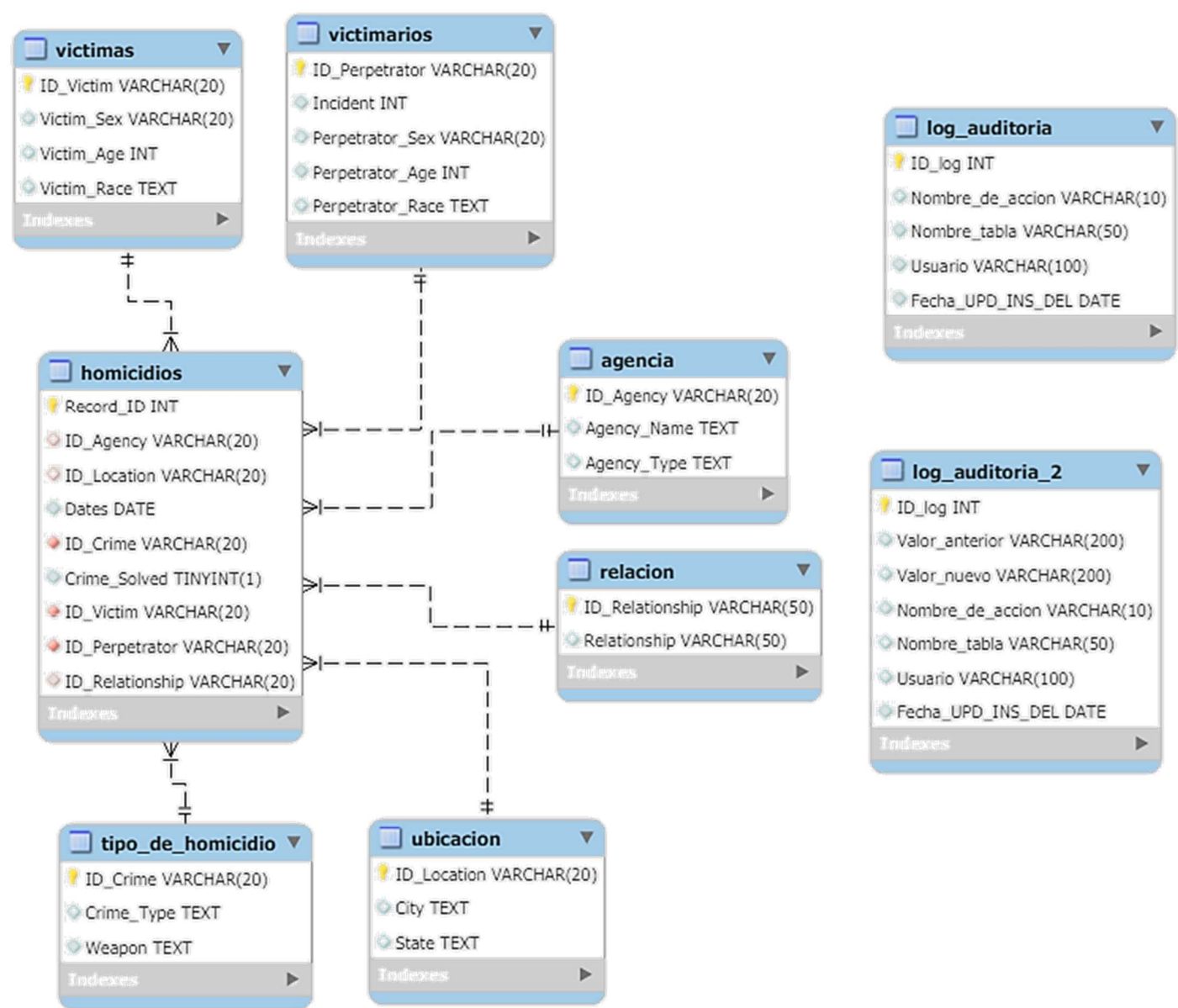


DIAGRAMA ENTIDAD RELACIÓN VERSIÓN SINTÉTICA

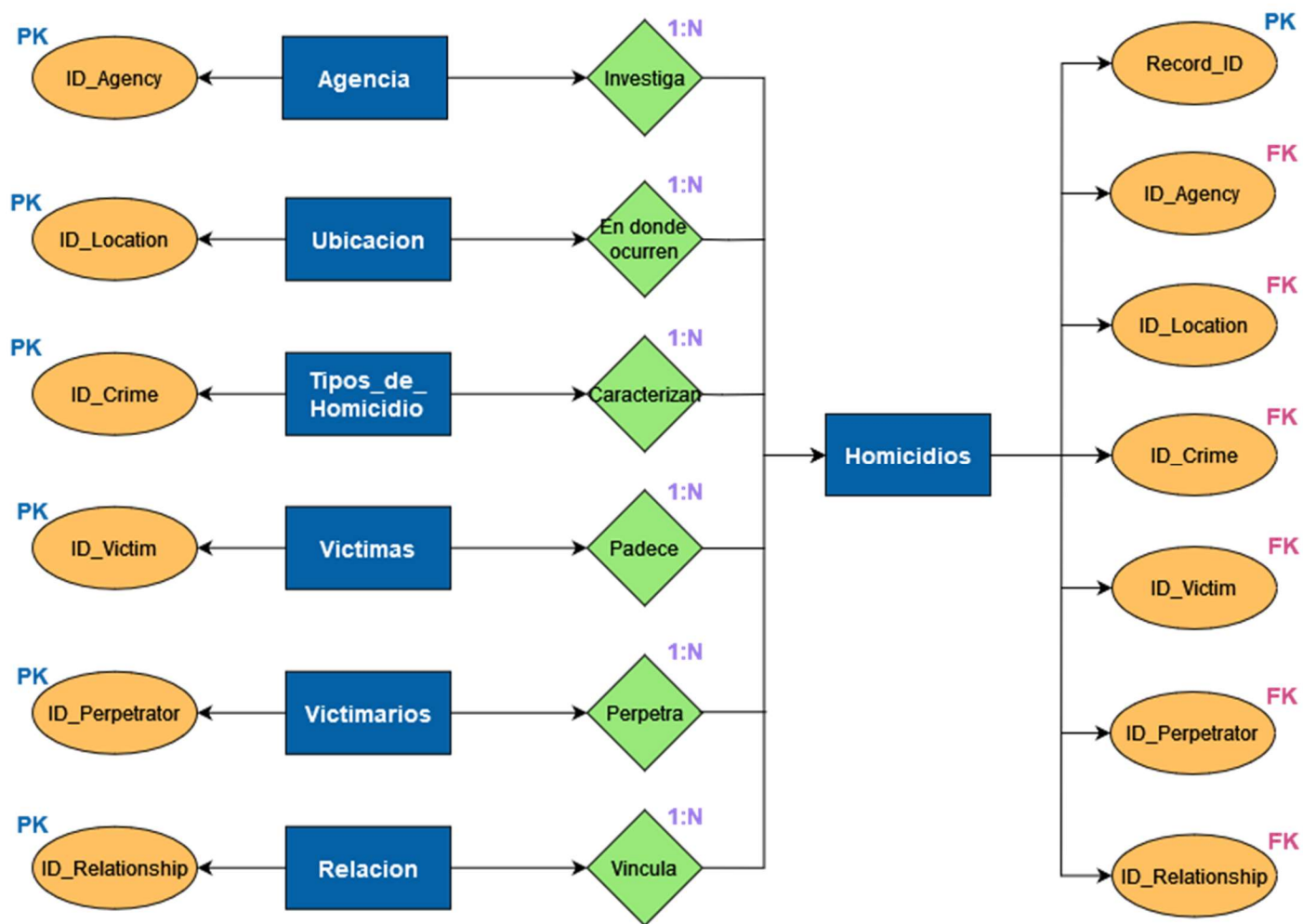


Ilustración 1 – DIAGRAMA E - R (MODELO ENTIDAD RELACIÓN)
Representa de forma esquemática los elementos de base de datos y sus relaciones que se analizarán. Los rectángulos representan las entidades o tablas, las elipses los atributos, campos o columnas y los rombos las acciones.

8. CAMPOS DE LAS TABLAS Y SUS TIPOS DE DATOS

Para lograr tener una comprensión más acabada y certera de la base de datos que se desea analizar, será necesario brindar información sobre los tipos de datos que tendrá cada campo de las tablas. Esto es de suma relevancia ya que establece las reglas de contenido, permitiendo dilucidar que tipo de dato será factible hallar en la base de datos que se pretende estudiar. Las tablas que se detallarán serán: Homicidios, Agencia, Ubicación, Tipo_de_Homicidio, Victimas, Victimarios y Relacion.

Tabla Homicidios
Cantidad de campos: 9

Homicidios		
NOMBRE DE CAMPO	ABREVIATURA	TIPO DE DATO
Record_ID	recid	NUMÉRICO
ID_Agency	ida	TEXTO
ID_Location	idl	TEXTO
Dates	date	FECHA
ID_Crime	idc	TEXTO
Crime_Solved	cs	BOOLEANO
ID_Victim	idv	TEXTO
ID_Perpetrator	idp	TEXTO
ID_Relationship	idr	TEXTO

La tabla homicidios se encarga de establecer la relación con las demás tablas de la base de datos, es decir recopila la información de las agencias de investigación, ubicación de los homicidios, tipos de crímenes, víctimas, victimarios y las relaciones entre ambos, a través de su ID correspondiente.

Tabla Agencia
Cantidad de campos: 3

Agencia		
NOMBRE DE CAMPO	ABREVIATURA	TIPO DE DATO
ID_Agency	ida	TEXTO
Agency_Name	an	TEXTO
Agency_Type	at	TEXTO

Contiene información sobre las agencias implicadas en la investigación de los homicidios perpetrados.

Tabla Ubicación
Cantidad de campos: 3

Ubicación		
NOMBRE DE CAMPO	ABREVIATURA	TIPO DE DATO
ID_Location	idl	TEXTO
City	city	TEXTO
State	ste	TEXTO

Expone datos de la localización de los crímenes cometidos.

Tabla Tipo de Homicidio

Cantidad de campos: 3

Tipo_de_Homicidio		
NOMBRE DE CAMPO	ABREVIATURA	TIPO DE DATO
ID_Crime	idc	TEXTO
Crime_Type	ct	TEXTO
Weapon	wpo	TEXTO

Muestra información de las armas utilizadas para cometer los homicidios.

Tabla Victimas

Cantidad de campos: 4

Victimas		
NOMBRE DE CAMPO	ABREVIATURA	TIPO DE DATO
ID_Victim	idc	TEXTO
Victim_Sex	vsex	TEXTO
Victim_Age	vage	NUMÉRICO
Victim_Race	vrace	TEXTO

Contiene información sobre las víctimas de los homicidios.

Tabla Victimarios

Cantidad de campos: 5

Victimarios		
NOMBRE DE CAMPO	ABREVIATURA	TIPO DE DATO
ID_Perpetrator	idp	TEXTO
Incident	inci	NUMÉRICO
Perpetrator_Sex	psex	TEXTO
Perpetrator_Age	page	NUMÉRICO
Perpetrator_Race	prace	TEXTO

Contiene información sobre los autores de los crímenes. El campo “Incident” se refiere a cuantos incidentes estuvo involucrado el perpetrador del homicidio.

Tabla Relacion
Cantidad de campos: 2

Relacion		
NOMBRE DE CAMPO	ABREVIATURA	TIPO DE DATO
ID_Relationship	ldr	TEXTO
Relationship	rel	NUMÉRICO

Contiene información sobre la relación entre la víctima y el victimario.

9. VISTAS

Esta sección incluye consultas embebidas en vistas, de tal forma que un usuario pueda realizar la consulta de forma sencilla y segura. Se incluye en el repositorio en github el archivo **“Script_consultas_vistas”**. Este script permite la creación de 5 vistas distintas para facilitar las consultas que se cree que serán más recurrentes por los futuros usuarios de la base de datos.

1. VW_CRIMENES_RESUELTOS_EN_ALASKA

Sirve para listar los crímenes resueltos en Alaska. Los campos que se podrán visualizar al ejecutar la consulta a la vista son: Fechas de los asesinatos, nombre de la agencia que intervino en la resolución, ciudad donde se cometió el crimen, arma involucrada y relación entre víctima y victimario.

2. VW_CRIMENES_NO_RESULETOS_MENORES_15_ANIOS

Trae información de crímenes no resueltos con víctimas menos de 15 años de edad. La relación entre víctima y victimario es conocida. La consulta traerá información de los campos Fecha del crimen, sexo de la víctima, edad de la víctima y relación con el asesino. Se establece un orden predeterminado ascendente por el campo edad de la víctima.

3. VW_CANTIDAD_ASESINOS_POR_SEXO_RAZA_Y_ARMA_HOMICIDA

Lista información relativa a la cantidad de asesinatos efectuados acorde al sexo de perpetrador, a su raza y al arma homicida. Todos estos campos se encuentran agrupados para el correcto funcionamiento de la consulta y que se pueda además realizar el conteo de cada combinación de los tres campos.

4. VW_TOP_10_AGENCIAS_CON_MAS_CASOS_RESUELTOS

Devuelve el top diez, en orden descendente, de las agencias que más crímenes han resuelto.

5. VW_RANKING_ASESINATOS_POR_ESTADO

Lista en orden ascendente la cantidad de crímenes cometidos por estado, tanto voluntarios como involuntarios que no sean producto de la negligencia humana.

10. FUNCIONES

La primera función que se presenta en el archivo “**Script_funciones**” se llama FN_ CONTAR_ HOMICIDIOS_ POR_ ESTADO. Esta función me permite contar por estado la cantidad de homicidios cometidos, resueltos, no resueltos o ambos. Recibe dos parámetros: **p_estado**, que es de tipo VARCHAR y se debe colocar entre comillas simples y **p_crimen_resuelto**, que es CHAR. Los únicos valores admitidos para este último parámetro son:

- 0 ---> Casos no resueltos.
- 1 ---> Casos resueltos.
- 2 ---> Todos los casos, es decir es la suma de casos no resueltos más resueltos.

El parámetro **p_crimen_resuelto** puede ser ingresado sin comillas.

Para el primer parámetro, **p_estado**, las opciones válidas son cada uno de los estados de Estados Unidos:

Alaska; Alabama; Arkansas; Arizona; California; Colorado; Connecticut; District of Columbia; Delaware; Florida; Georgia; Hawaii; Iowa; Idaho; Illinois; Indiana; Kansas; Kentucky; Louisiana; Massachusetts; Maryland; Maine; Michigan; Minnesota; Missouri; Mississippi; Montana; Nebraska; North Carolina; North Dakota; New Hampshire; New Jersey; New Mexico; Nevada; New York; Ohio; Oklahoma; Oregon; Pennsylvania; Rhode Island; South Carolina; South Dakota; Tennessee; Texas; Utah; Virginia; Vermont; Washington; Wisconsin; West Virginia; Wyoming

No importa si el nombre ingresado no respeta las mayúsculas y minúscula de las opciones brindadas ya que la función cumplirá su cometido de todas maneras.

La segunda función que se presenta es de nombre FN_ CONTAR_ HOMICIDIOS_ CON_ FILTROS. Esta se encarga de contar la cantidad de homicidios cometidos a partir de tres parámetros recibidos. Estos parámetros harán de filtros acorde a lo que se quiera contar. Todos ellos son del tipo VARCHAR, es decir que tienen características de texto. El primer parámetro es **p_arma** y admite los siguientes tipos de arma homicida:

Firearm	Shotgun	Gun	Drowning
Handgun	Rifle	Fire	Poison
Blunt Object	Unknown	Strangulation	Fall
Knife	Suffocation	Drugs	Explosives

El siguiente parámetro es p_sexo_victimario y admite tres opciones únicamente: Male, Female y Unknown.

El tercer y último se llama p_relacion (relación de la víctima con el victimario) y contiene las siguientes alternativas válidas:

Wife	Family	Employee	Stepson
Acquaintance	Mother	Stepfather	Ex-Husband
Brother	Husband	Common-Law Wife	Stepmother
Girlfriend	Boyfriend/Girlfriend	Son	
Father	Stepdaughter	Common-Law Husband	
Boyfriend	Sister	Ex-Wife	
Neighbor	In-Law	Employer	

Como punto importante, cabe mencionar que los tres parámetros aceptan por igual un valor extra que no fue mencionado precedentemente, el valor NULL. Lo que se logra con su utilización es obviar cierto/s parámetro/s como filtros. Por ejemplo, si solo se quisiera saber la cantidad de crímenes cometidos con armas de fuego, la sentencia sería:

```
SELECT FN_ CONTAR_ HOMICIDIOS_ CON_ FILTROS ('Firearm', NULL, NULL)
```

en donde, tanto el segundo como tercer parámetro son obviados mediante la manifestación del NULL. Este va sin comillas. También es válido el caso (NULL, NULL, NULL), que arrojará en este caso particular la totalidad de los homicidios cometidos, ya que no se aplicará ningún tipo de filtro a la consulta llamada a través de la función.

11. STORE PROCEDURES

Se han elaborado dos store procedures, los cuales cumplen objetivos muy diferentes en su funcionamiento.

SP_Ordenar_homicidios

Su finalidad consiste en que el usuario pueda ordenar el resultado de una consulta a través del ingreso de dos parámetros. Existe un tercer parámetro para la función que se detallará luego, por lo que los parámetros totales requeridos al usuario serán tres.

El primer parámetro es **P_nombre_columna**. Define la columna a través de la cual se desean ordenar los resultados. Las opciones son:

Dates	State	Victim_sex
Agency_name	Crime_Type	Victim_age
City	Weapon	Relationship

El segundo parámetro se denomina **P_orden** y solo acepta dos posibles valores:

ASC ---> Indica que el orden de la columna seleccionada en *P_nombre_columna* debe ser ascendente.

DESC ---> Indica que el orden de la columna seleccionada en *P_nombre_columna* debe ser descendente.

Como se mencionó anteriormente, existe un tercer parámetro, el cual no cumple función en el ordenamiento, pero sí en el filtrado de información. Este filtrado es para reflejar en la consulta que crímenes han sido resueltos y cuáles no. Las opciones a ingresar como parámetro son:

0 ---> Casos no resueltos

1 ---> Casos resueltos

2 ---> Todos los casos, es decir es la suma de casos no resueltos más resueltos.

SP_Ordenar_homicidios

Este SP recibe cinco parámetros. Se describirán a continuación cada uno de ellos en orden ascendente:

p_operacion: tiene dos posibilidades 'insert' o 'delete'. El primero sirve para agregar información a la tabla 'victimas' del esquema Homicidios en la base de datos. El segundo realiza justo lo opuesto, permite la eliminación de un registro.

p_id_victim: si el parámetro anterior es 'insert' entonces se debe ingresar el nuevo ID del registro a insertar. Por el contrario, si es 'delete' entonces el ID ingresado es del registro a eliminar.

p_victim_sex: si el primer parámetro *p_operacion* es 'delete' entonces su valor debe ser NULL, caso contrario se deberá agregar el sexo de la víctima al registro que se está adicionando.

p_victim_age: si el primer parámetro *p_operacion* es 'delete' entonces su valor debe ser NULL, caso contrario se deberá agregar la edad de la víctima al registro que se está adicionando.

p_victim_race: si el primer parámetro *p_operacion* es 'delete' entonces su valor debe ser NULL, caso contrario se deberá agregar la raza de la víctima al registro que se está adicionando.

Tanto cuando se ejecuta una operación de tipo insert o delete, aparecerá un mensaje (como si fuera el resultado una consulta) en donde indicará el éxito o fracaso de la operación realizada.

12. TRIGGERS

Se crearán una serie de disparadores de tal manera, que las operaciones que realicen inserción, actualización o borrado de datos, sean capturados en tablas de auditoria creadas para tal fin (log_auditoria y log_auditoria_2). Las tablas que se consideran importantes para realizarle el seguimiento son 'Victimarios' y 'Victimas'. Cualquiera de las operaciones anteriormente mencionadas será registrada en las tablas de log. La información capturada será:

Valor_anterior: constituye el valor antes de la modificación.

Valor_nuevo: valor luego de la modificación.

Nombre_de_accion: será acorde al tipo de operación realizada, es decir INSERT, UPDATE o DELETE.

Nombre_de_tabla: almacena la información sobre la cual se ha realizado la modificación. En este caso puntual, como las tablas que se consideran más sensibles son 'Victimarios' y 'Victimas', será alguno de estos dos valores.

Usuario: almacena el nombre de usuario que ha realizado la modificación.

Fecha_UPD_INS_DEL: fecha en la que se ha realizado la modificación a las tablas 'Victimarios' y 'Victimas'.

13. DCL – USUARIOS

Se proporcionan para su uso dos perfiles de usuarios distintos, acorde a los privilegios que se quieran proporcionar. El primero de ellos otorga únicamente permisos de lectura a la base de datos. El segundo es menos restrictivo ya que permite además realizar modificaciones en la estructura de las tablas y modificaciones en los datos ingresados. Cabe mencionar que en ningún caso se dan permisos de usuarios que permitan la eliminación de información, quedando esta tarea relegada únicamente al administrador de bases de datos, quien contará con todos los permisos.

Para obtener el acceso a los usuarios y las credenciales que brindan acceso a la base de datos, deben ser requeridas directamente al administrador, quien evaluará la solicitud y decidirá si es procedente o no tal requerimiento.

14. HERRAMIENTAS UTILIZADAS

Para la elaboración del presente proyecto fueron utilizadas las siguientes herramientas:

- MySQL Workbench
- Power BI
- Microsoft Excel
- Notepad ++

15. INFORMACIÓN ADICIONAL SOBRE TABLAS E INSTANCIAS

Las tablas de la base de datos están compuestas por la siguiente cantidad de registros:

NOMBRE DE TABLA	CANTIDAD DE REGISTROS
Homicidios	139.700
Agencia	6.917
Ubicación	2.572
Tipo_de_homicidio	32
Victimas	4.509
Victimarios	23.195
Relacion	28

16. REPORTE

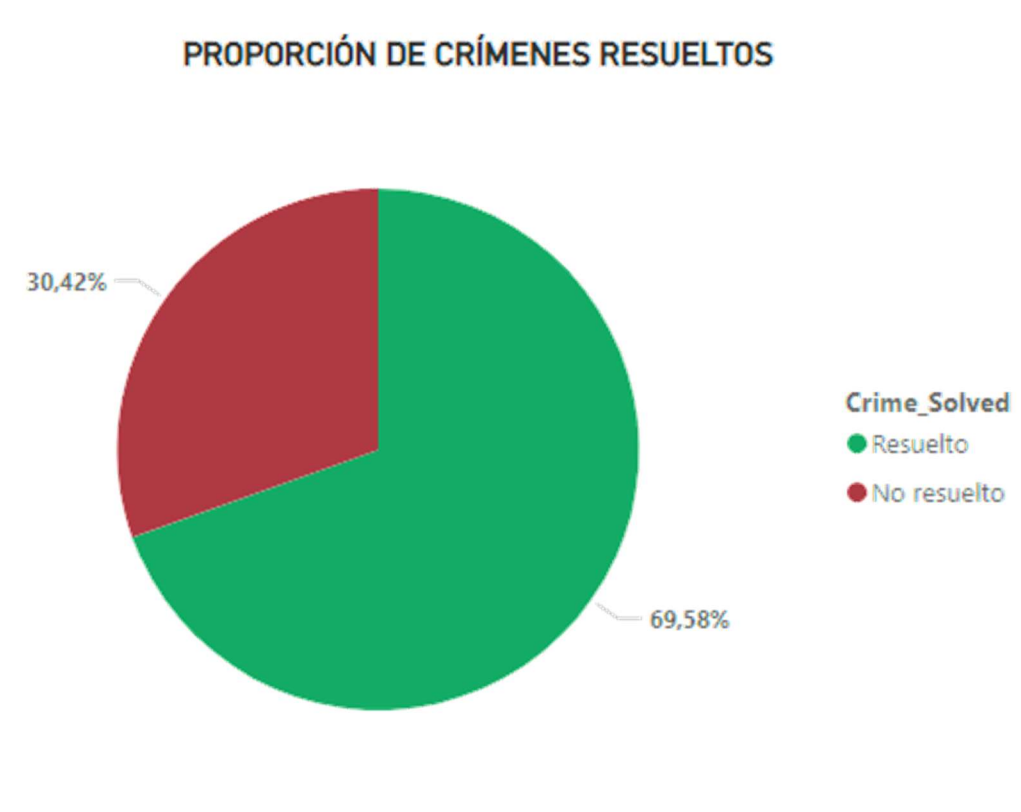


Ilustración 2 – Crímenes resueltos
De la totalidad de los casos de homicidios en la base de datos, se calcula la proporción de cuantos de aquellos fueron resueltos por las agencias gubernamentales. En este caso se puede apreciar como aproximadamente el 70% de los homicidios se han podido resolver exitosamente.

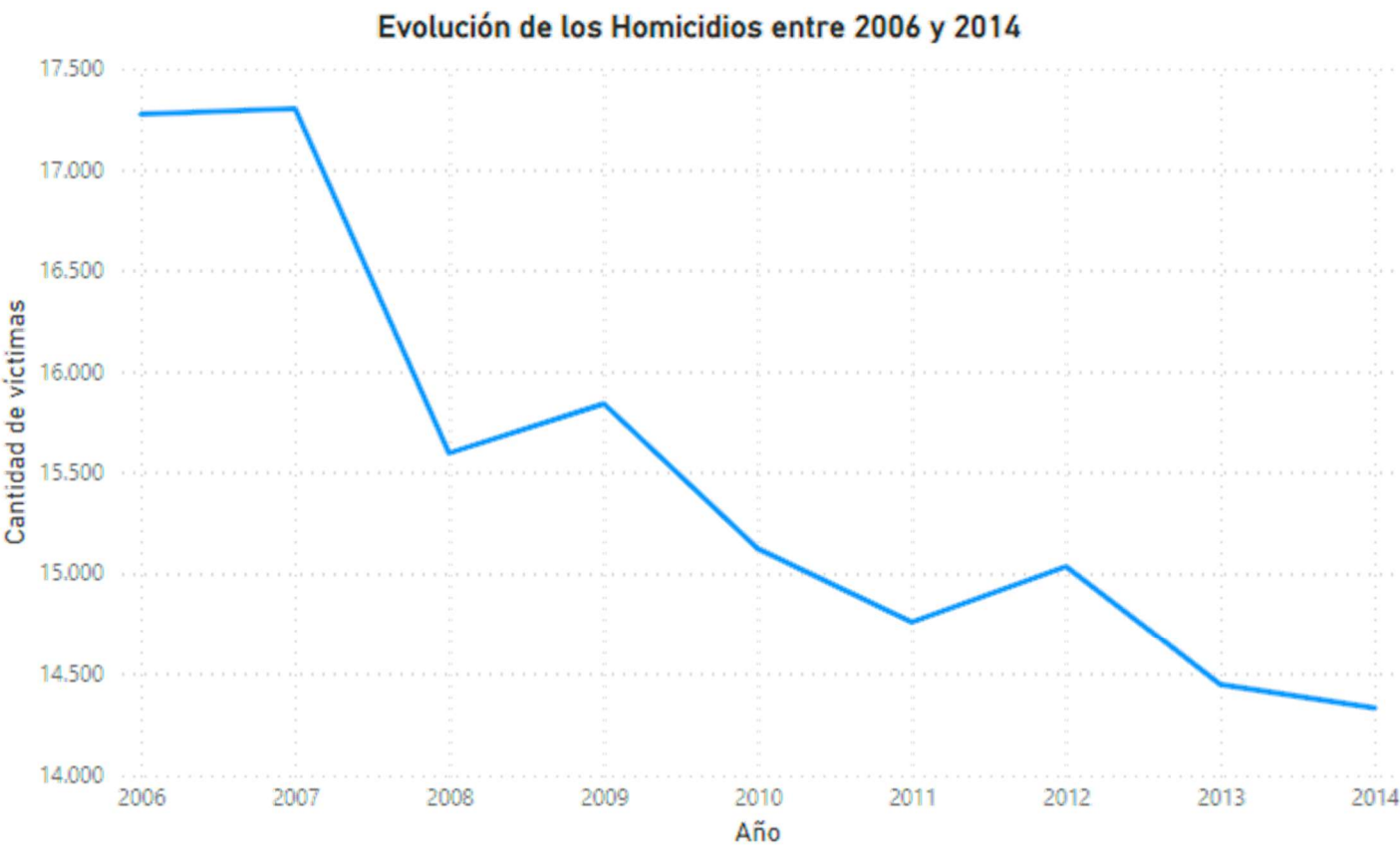


Ilustración 3 – Evolución de la cantidad de homicidios en el tiempo. Período 2006-2014
A través de presente gráfico de líneas se expone como las políticas públicas en materias de seguridad han contribuido fuertemente a disminuir los casos de homicidios por año. Se muestra que es posible reducir el nivel de violencia si se toman cartas en el asunto, por lo que la cantidad de casos puede disminuir aún más si las autoridades dan prioridad a la reducción de los crímenes en el país a través de políticas de prevención, detección y contención social en aquellos sectores más propensos o vulnerables.

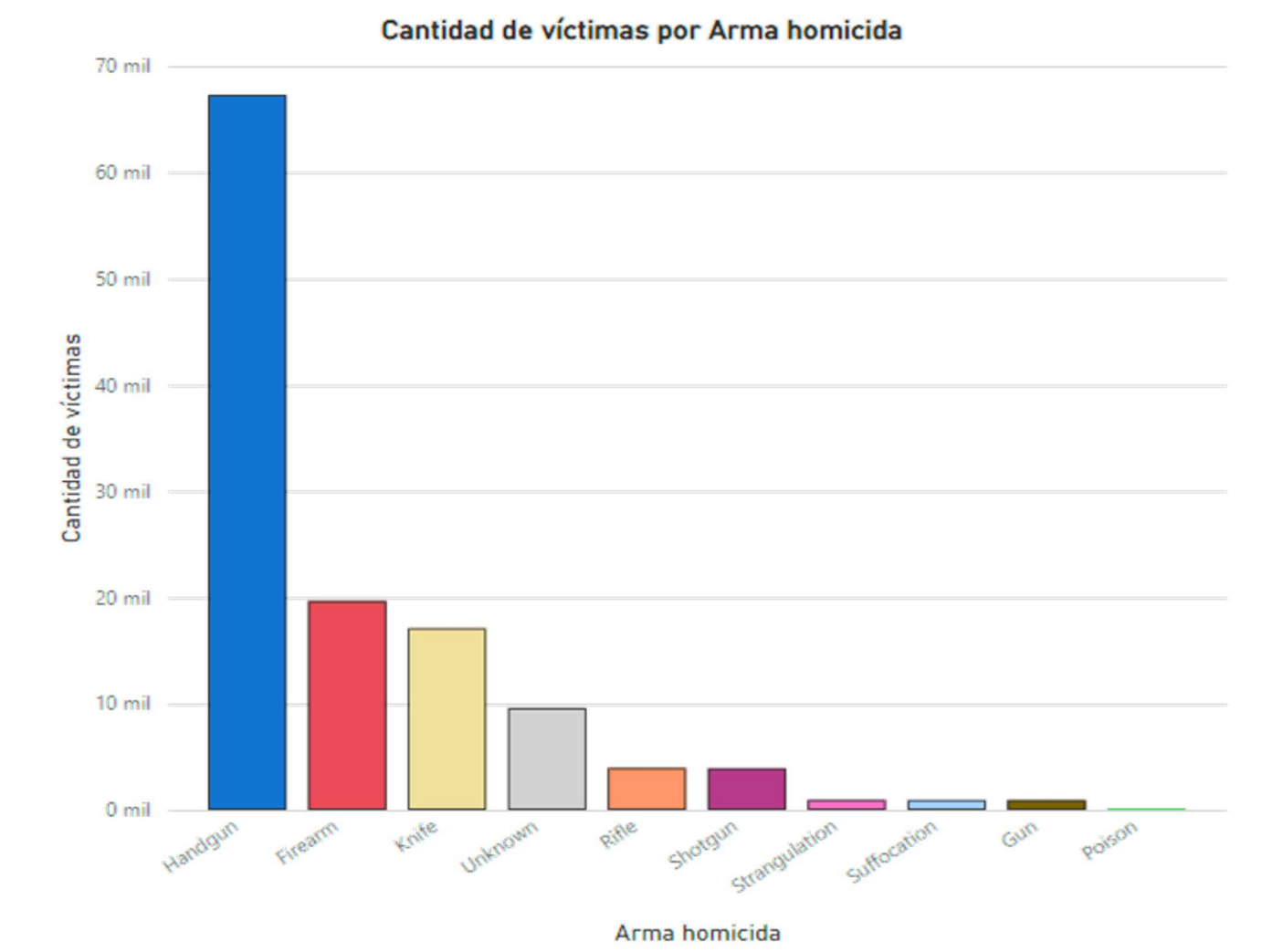


Ilustración 4 – Recuento de crímenes por arma homicida. Top 10.
Se establece el TOP 10 de armas con la que se han cometido mayor cantidad de homicidios en el período de estudio. Por diferencia el arma que mas vidas de ha cobrado en la pistola.

TOP 20 - Estados con más homicidios

Estado	Cantidad de víctimas
California	19076
Texas	12278
Florida	10738
New York	6970
Pennsylvania	6497
Michigan	5833
Georgia	5366
North Carolina	4676
Louisiana	4482
Illinois	4418
Ohio	4404
Maryland	4134
Tennessee	3822
Missouri	3755
Arizona	3638
New Jersey	3523
Virginia	3327
South Carolina	2979
Indiana	2904
Oklahoma	2024

Ilustración 5 – TOP 20 estados con mayor cantidad de homicidios
Se expone un ranking ordenado de manera descendente con los Estados que presentan una mayor cantidad de homicidios para el período de estudio.

TOP 20 - Homicidios por ciudad

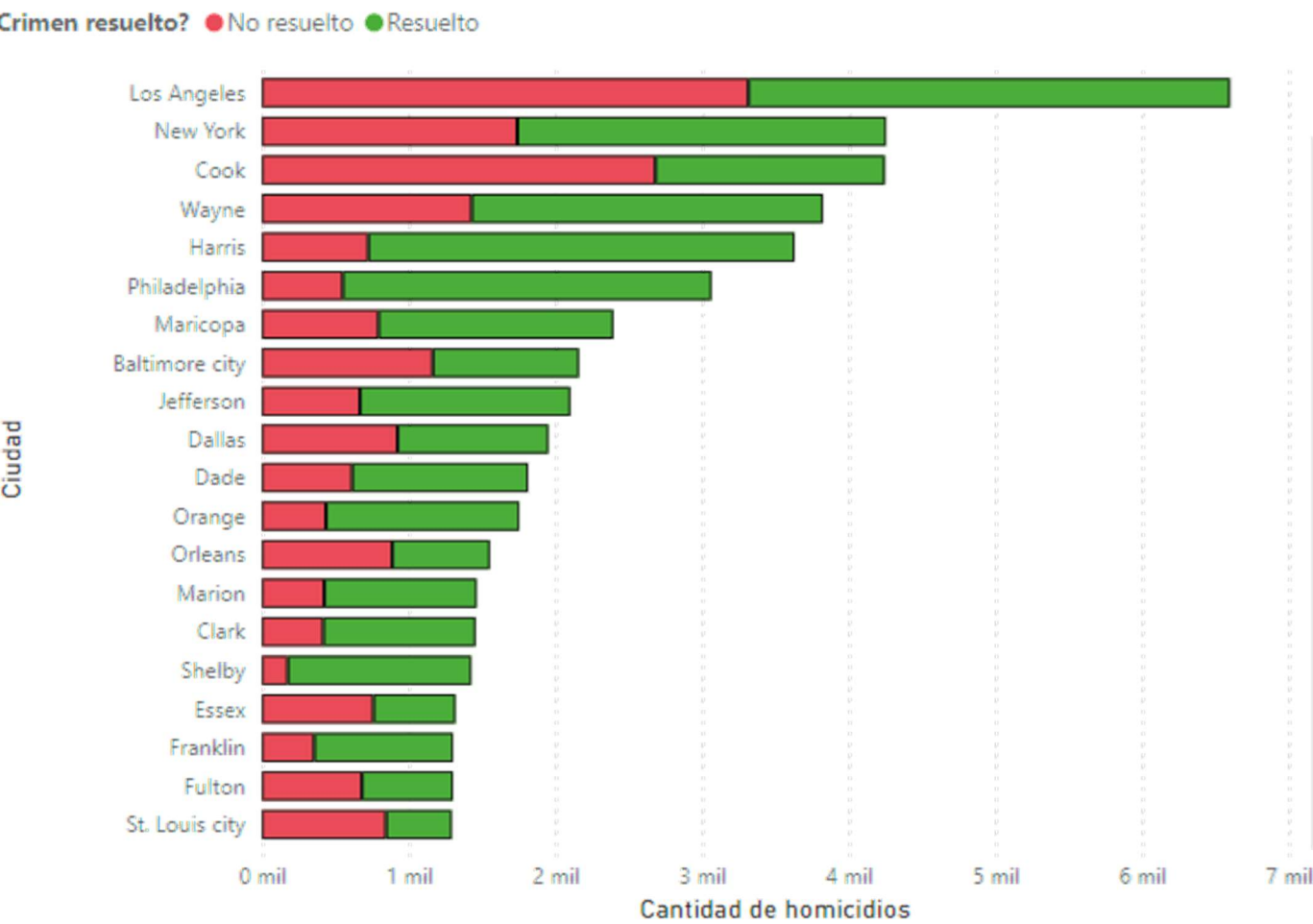


Ilustración 6 – TOP 20 ciudades con mayor cantidad de homicidios. Segmentado por resueltos o no resuelto. Se muestra el ranking de ciudades ordenado en forma descendente con las que tienen mayor cantidad de homicidios. Se realiza segmentación por colores acorde a si el crimen fue resuelto o no.

17. BACKUP DE BASE DE DATOS

Contiene una copia de respaldo de toda la base de datos. Ante cualquier eventual contingencia que pueda derivar en la pérdida de información, se podrá recuperar la información original, minimizando los daños. Con tal fin se provee, por un lado, el respaldo de todos los datos y, por otro, el de la estructura de todas las tablas.

18. COMENTARIOS FINALES

Una vez llegado a esta instancia del trabajo cabe decir que se espera que sea de utilidad para los futuros usuarios y que se le pueda sacar provecho a su funcionamiento. Los reportes elaborados precedentemente constituyen a modo de ejemplo todo lo que se puede lograr con las consultas efectuadas a la base de datos, ya que las mismas pueden ser exportadas y llevadas a visualizadores para elaborar reportes que permitan llegar a conclusiones.

Si a esta estructura de datos se le sigue ingresando información regularmente, incrementará la capacidad de prevención de delitos por homicidios. Es de esta manera como puede afirmarse que los datos constituyen una herramienta muy poderosa para interpretar y analizar la realidad. Si se disponen de métodos adecuados para relevar esa información, se podrá, mediante meticolosos análisis, poder lograr medidas que tengan un real impacto en nuestra sociedad.