

2023

DATA ANALYTICS

PROYECTO FINAL

AUTOR: GONZALO BURGA ZAPATA

DOCENTE: LUCIANO JULIAN GOMEZ OLIVERA

CODERHOUSE | DATA ANALYTICS - COMISION 42290

CONTENIDO

DESCRIPCIÓN DE LA TEMÁTICA.....	2
HIPÓTESIS.	2
HERRAMIENTAS TECNOLÓGICAS.....	2
DIAGRAMA ENTIDAD – RELACIÓN.....	3
DEFINICIÓN DE TABLAS.	6
GENERACIÓN DE FONDOS.....	8
GENERACIÓN DE HIPERVÍNCULOS.....	8
APLICACIÓN DE FONDOS.	9
IMPORTACIÓN DE TABLAS.....	9
TABLA DE FECHAS.....	10
RELACIONES ENTRE TABLAS.	10
TABLA DE MEDIDAS.....	11
VISUALIZACIONES.	16
CONCLUSIÓN.	29

DESCRIPCIÓN DE LA TEMÁTICA.

El dataset elegido, cuenta con más de 100,000 homicidios, perpetrados en distintos estados y localidades de USA, en un periodo de años entre 2006 hasta 2014. Cada uno de los homicidios cuenta con información sobre la agencia policial que lo cubrió, dónde ocurrió, fecha, el tipo de homicidio, si el homicidio fue resuelto o no, información, así como de la víctima como del victimario y el tipo de relación de estos últimos.

HIPÓTESIS.

Con un Analytic Dashboard o Tablero Analítico en Power BI se podría realizar un análisis geográfico para resaltar en que estados ha habido más homicidios y diferenciarlos con los estados en que ha habido menos.

También se podría analizar los homicidios a lo largo del tiempo y si existe alguna tendencia dependiendo de algún mes específico.

Otro análisis que se examinaría es cuáles son los tipos de homicidios más comunes.

Dependiendo de las características de las víctimas y victimarios (edad, raza, género) se podría observar algún tipo de tendencia o patrón.

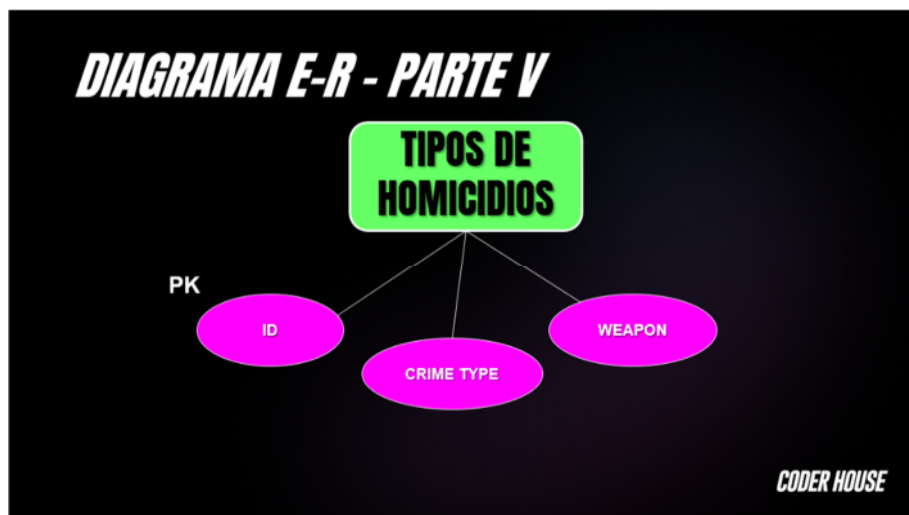
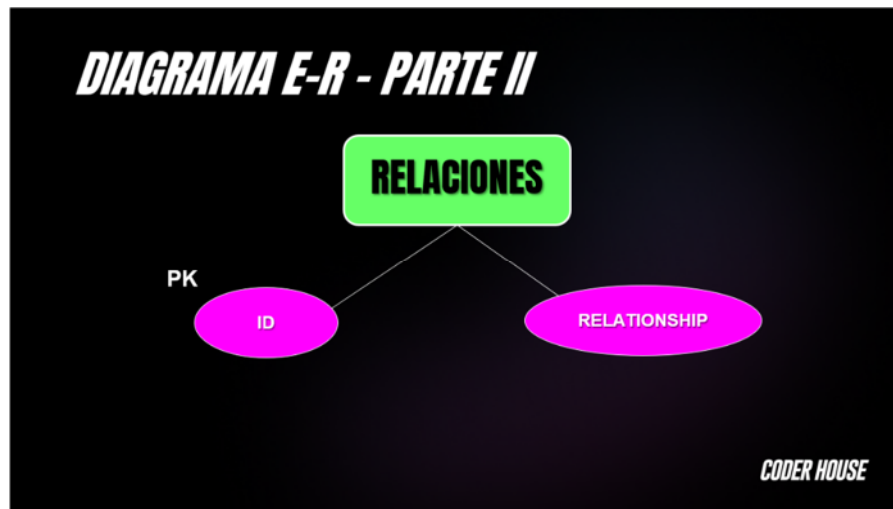
Así como lo mencionado anteriormente, se irá observando más tipos de análisis con el objetivo de realizar un tablero analítico lo más completo posible.

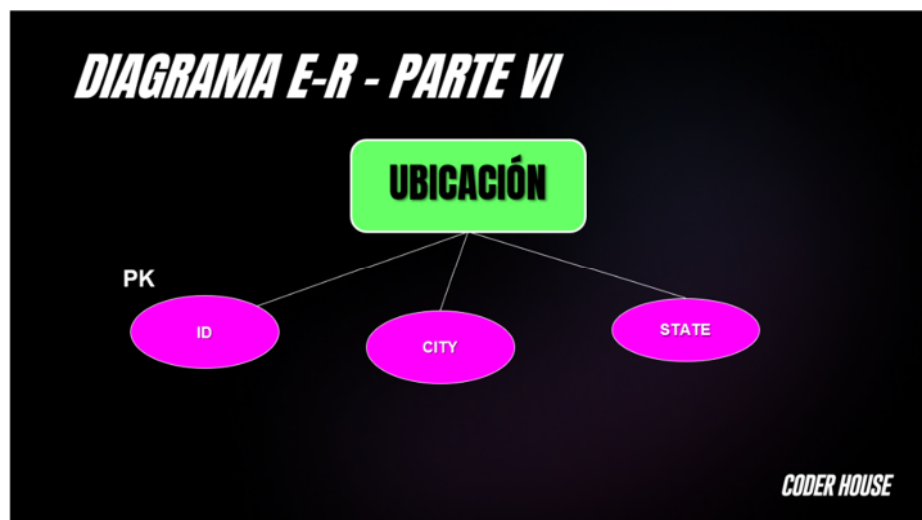
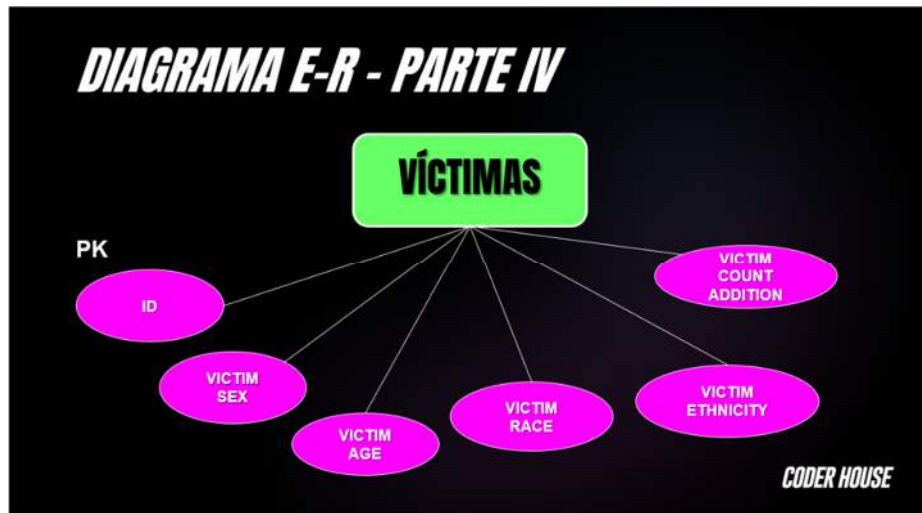
HERRAMIENTAS TECNOLÓGICAS.

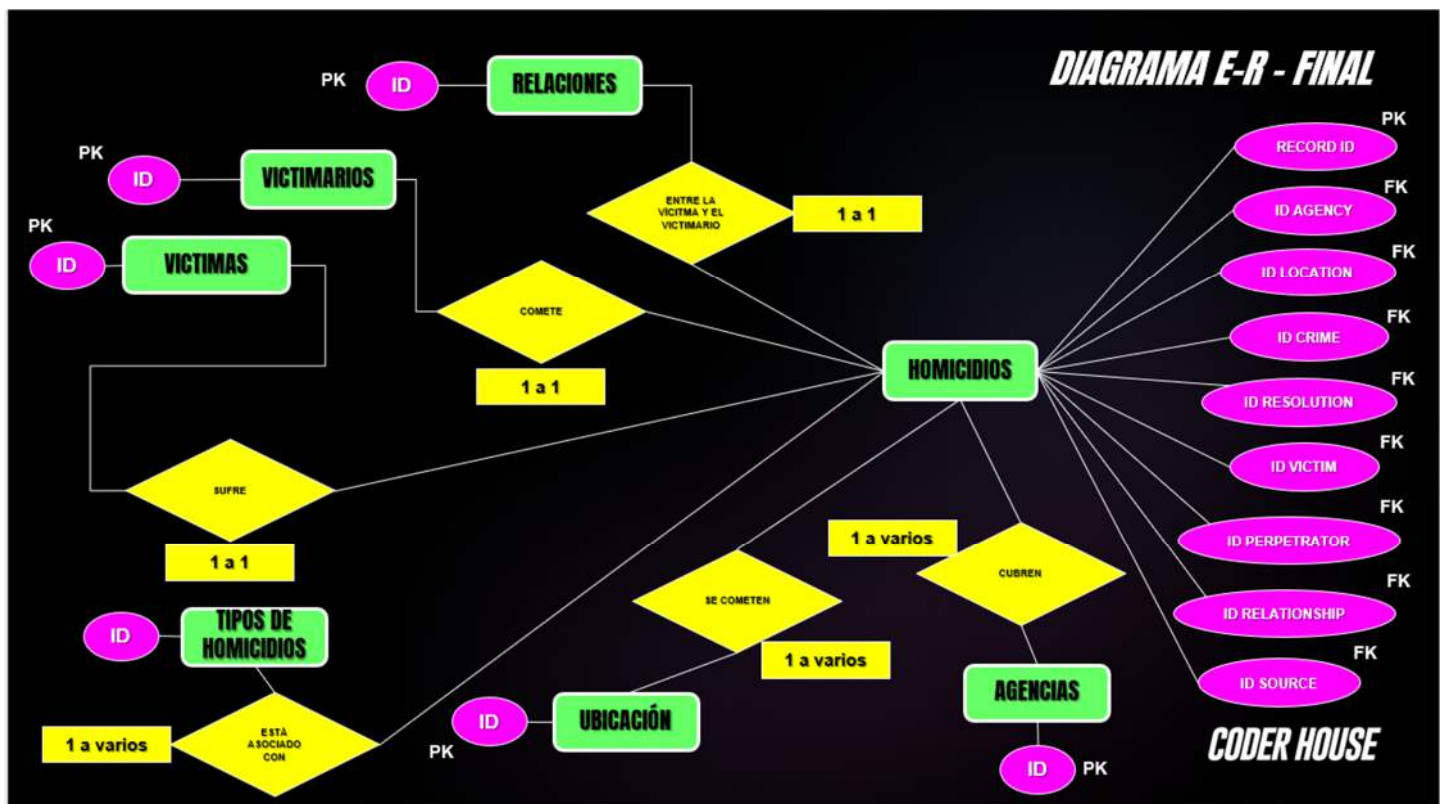
En el presente trabajo se han utilizado distintas tecnologías como:

- ✓ Excel (Lectura de Datasets).
- ✓ Power BI (Elaboración del Dashboard).
- ✓ PowerPoint (Elaboración del Diagrama E-R y Fondos para el Dashboard).
- ✓ FlatIcon (Web de descarga de íconos).
- ✓ Google Imágenes (Descarga de imágenes).

DIAGRAMA ENTIDAD – RELACIÓN.







DEFINICIÓN DE TABLAS.**Relaciones.**

La tabla Relación muestra la relación entre la víctima y el victimario.

CAMPO	TIPO DE DATO	TIPO DE CLAVE
ID	VARCHAR	PK
RELATIONSHIP	VARCHAR	

Victimarios.

La tabla Victimario muestra datos relevantes sobre el agresor, tales como si es reincidente, sexo, edad, entre otros.

CAMPO	TIPO DE DATO	TIPO DE CLAVE
ID	VARCHAR	PK
INCIDENT	SMALLINT	
PERPETRATOR SEX	VARCHAR	
PERPETRATOR AGE	TINYINT	
PERPETRATOR RACE	VARCHAR	
PERPETRATOR ETHNICITY	VARCHAR	
PERPETRATOR COUNT	TINYINT	

Víctimas.

Así como en la anterior tabla se muestran los datos del agresor, en la tabla Víctimas se muestran datos de las víctimas.

CAMPO	TIPO DE DATO	TIPO DE CLAVE
ID	VARCHAR	PK
VICTIM SEX	VARCHAR	
VICTIM AGE	TINYINT	
VICTIM RACE	VARCHAR	
VICTIM ETHNICITY	VARCHAR	
VICTIM COUNT ADDITION	TINYINT	

Tipos de Homicidios.

La tabla Homicidios indica si el crimen ha sido un asesinato u homicidio por negligencia, así como también si se ha usado algún tipo de arma.

CAMPO	TIPO DE DATO	TIPO DE CLAVE
ID	VARCHAR	PK
CRIME TYPE	VARCHAR	
WEAPON	VARCHAR	

Ubicación.

Como su propio nombre indica, esta tabla, denominada Ubicación, presenta la ubicación (Estado y Ciudad) donde se cometió el crimen.

CAMPO	TIPO DE DATO	TIPO DE CLAVE
ID	VARCHAR	PK
CITY	VARCHAR	
STATE	VARCHAR	

Agencias.

La tabla Agencias muestra el nombre y el tipo de la agencia que asistió el homicidio.

CAMPO	TIPO DE DATO	TIPO DE CLAVE
ID	VARCHAR	PK
AGENCY NAME	VARCHAR	
AGENCY TYPE	VARCHAR	

Homicidios.

La tabla Homicidios, es la tabla a analizar. Básicamente muestra todos los ID que se han visto en las anteriores tablas y adicionalmente la fecha del homicidio y si el crimen fue resuelto o no.

CAMPO	TIPO DE DATO	TIPO DE CLAVE
ID	INT	PK
ID AGENCY	VARCHAR	FK
ID LOCATION	VARCHAR	FK
DATE	DATE	
ID CRIME	VARCHAR	FK
RESOLUTION	VARCHAR	
ID VICTIM	VARCHAR	FK
ID PERPETRATOR	VARCHAR	FK
ID RELATIONSHIP	VARCHAR	FK
SOURCE	VARCHAR	

GENERACIÓN DE FONDOS.

Utilizando PowerPoint se procedió a crear el fondo para la portada y para las páginas del informe de Power BI.



Portada



Resto de páginas

GENERACIÓN DE HIPERVÍNCULOS.

En la primera página del informe, se colocó el logo de WhatsApp. A través del logo el usuario podrá acceder directamente al chat con el diseñador del informe.

APLICACIÓN DE FONDOS.

Se aplican los fondos generados previamente a las distintas páginas de Power BI.

IMPORTACIÓN DE TABLAS.

Se importa la información del dataset Homicidios USA, que tiene formato de Excel al Power BI.

- Se usa la opción “Usar la primera fila como encabezado” en las tablas Agencias, Homicidios, Tipos de Homicidios y Ubicación.
- En la tabla Relaciones se elimina una columna cargada erróneamente donde no hay ningún dato.
- De la tabla Ubicación, en la columna “State”, se cambia de nombre al estado “Rhodes Island” por “Rhode Island”.
- De la tabla Homicidios se cambia el nombre de la columna “Record ID” por “ID”.
- De la tabla Víctimas, en la columna “Victim Age” se cambia el valor 998 por 98.
- De la misma tabla Víctima, se eliminan los campos “Victim Ethnicity” y “Victim Count Additional”.
- De la tabla Victimarios, se elimina el campo “Perpetrator Ethnicity” y “Incident”.
- De la misma tabla Victimarios, se cambia el nombre del campo “Perpetrator Count” a “Incident”.
- Desde la Vista de modelo se oculta el ID de algunas tablas.
- En la tabla Víctimas, se crea una columna personalizada para agrupar las edades por rango:
Rango = if [Victim Age] >=0 and [Victim Age] <=11 then "Niño (0 a 11 años)" else if [Victim Age] >=12 and [Victim Age] <=18 then "Adolescente (12 a 18 años)" else if [Victim Age] >=19 and [Victim Age] <=27 then "Joven (19 a 27 años)" else if [Victim Age] >=28 and [Victim Age] <=64 then "Adulto (28 a 64 años)" else "Adulto Mayor (65 años a más)"
- En la tabla Victimarios se hace lo mismo que se hizo con la tabla Víctimas:
Rango = if [Perpetrator Age] >=0 and [Perpetrator Age] <=11 then "Niño (0 a 11 años)" else if [Perpetrator Age] >=12 and [Perpetrator Age] <=18 then "Adolescente (12 a 18 años)" else if [Perpetrator Age] >=19 and [Perpetrator Age] <=27 then "Joven (19 a 27 años)" else if [Perpetrator Age] >=28 and [Perpetrator Age] <=64 then "Adulto (28 a 64 años)" else "Adulto Mayor (65 años a más)"

TABLA DE FECHAS.

Con la función CALENDARAUTO, se genera una tabla de fechas en Power Bi:

Fechas = CALENDARAUTO().

Posteriormente, se le da el formato de día/mes/año a la fecha para una mejor visualización. Después se marca toda la tabla como Tabla de Fechas con la opción "Marcar como tabla de fechas".

Lo siguiente, es crear nuevas columnas para generar distintos tipos de gráficos utilizando filtros de tiempo y cálculos de inteligencia de tiempo.

Cálculo de Años.

Año = YEAR(Fechas[Fecha])

Cálculo de Meses (Número).

Mes (En números) = MONTH((Fechas[Fecha]))

Cálculo de Meses (Letras).

Mes (En letras) = UPPER(LEFT(FORMAT(Fechas[Fecha], "MMMM"), 1)) & MID(FORMAT(Fechas[Fecha], "MMMM"), 2, 15)

Cálculo de Semestres.

Trimestre = QUARTER(Fechas[Fecha]) & "º trimestre"

RELACIONES ENTRE TABLAS.

A continuación, se presenta todas las relaciones del conjunto de datos.

TABLA	TABLA	CAMPO EN COMÚN
Relaciones	Homicidios	ID/ID Relationship
Victimarios	Homicidios	ID/ID Perpetrator
Víctimas	Homicidios	ID/ID Victim
Tipos de Homicidios	Homicidios	ID/ID Crime
Ubicación	Homicidios	ID/ID Location
Agencias	Homicidios	ID/ID Agency
Calendario	Homicidios	Fecha/Date

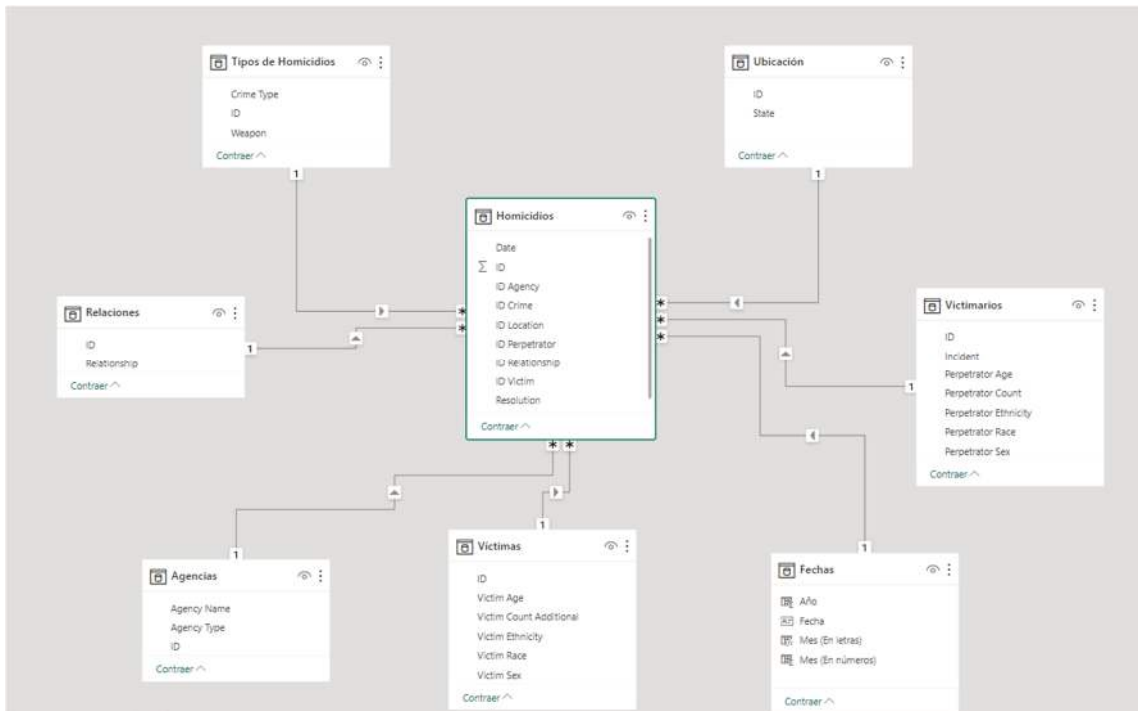


TABLA DE MEDIDAS.

Con la opción “Especificar datos” se crea una nueva tabla llamada “Medidas”. Aquí se incluirán todas las medidas a usar en el informe.

La primera es una “Medida rápida” que nos muestra la variación de homicidios de año a año:

Tasa de Variación por Año =

```
VAR __PREV_YEAR = CALCULATE([Total de Homicidios], DATEADD('Fechas'[Fecha], -1, YEAR))
```

RETURN

```
DIVIDE([Total de Homicidios] - __PREV_YEAR, __PREV_YEAR)
```

Esta misma medida es corregida para que no muestre la tasa de variación en el total:

Tasa de Variación por Año =

```
VAR __PREV_YEAR = CALCULATE([Total de Homicidios], DATEADD('Fechas'[Fecha], -1, YEAR))
```

RETURN

```
IF(SELECTEDVALUE(Fechas[Año]) IN VALUES(Fechas[Año]), DIVIDE([Total de Homicidios] - __PREV_YEAR, __PREV_YEAR), BLANK())
```

Se crea una “Medida rápida” más, esta vez para que calcule la variación de homicidios trimestre a trimestre:

Tasa de Variación por Trimestre =

```
VAR __PREV_QUARTER = CALCULATE([Total de Homicidios],  
DATEADD('Fechas'[Fecha], -1, QUARTER))
```

RETURN

```
DIVIDE([Total de Homicidios] - __PREV_QUARTER, __PREV_QUARTER)
```

Esta misma medida se modifica para que no muestre la tasa en el primer trimestre:

Tasa de Variación por Trimestre =

```
VAR __PREV_QUARTER = CALCULATE([Total de Homicidios],  
DATEADD('Fechas'[Fecha], -1, QUARTER))
```

RETURN

```
IF(SELECTEDVALUE(Fechas[Trimestre]) = "1º Trimestre", BLANK(), DIVIDE([Total de  
Homicidios] - __PREV_QUARTER, __PREV_QUARTER))
```

Se crea otra “Medida Rápida” para calcular la variación mensual:

Tasa de Variación por Mes =

```
VAR __PREV_MONTH = CALCULATE([Total de Homicidios], DATEADD('Fechas'[Fecha],  
-1, MONTH))
```

RETURN

```
DIVIDE([Total de Homicidios] - __PREV_MONTH, __PREV_MONTH)
```

Se modifica la medida para que no muestre la tasa en el primer mes:

Tasa de Variación por Mes =

```
VAR __PREV_MONTH = CALCULATE([Total de Homicidios], DATEADD('Fechas'[Fecha],  
-1, MONTH))
```

RETURN

```
IF(SELECTEDVALUE(Fechas[Mes (En letras)]) = "Enero", BLANK(), DIVIDE([Total de  
Homicidios] - __PREV_MONTH, __PREV_MONTH))
```

La tercera medida a crear cuenta la cantidad total de homicidios:

Total de Homicidios = COUNTROWS(Homicidios).

Otra medida más, contará la cantidad de homicidios resueltos:

Homicidios Resueltos =

CALCULATE(COUNT(Homicidios[Resolution]), Homicidios[Resolution] = "Yes")

La siguiente medida, cuenta la cantidad de homicidios no resueltos:

Homicidios no Resueltos =

CALCULATE(COUNT(Homicidios[Resolution]), Homicidios[Resolution] = "No")

Se genera otra medida que calcule el porcentaje de homicidios resueltos con respecto al total de homicidios:

% Homicidios Resueltos =

DIVIDE(CALCULATE(COUNT(Homicidios[Resolution]), Homicidios[Resolution] = "Yes"), [Total de Homicidios])

Ahora se crea una media que calcule el porcentaje de homicidios no resueltos con respecto al total de homicidios:

% Homicidios no Resueltos =

DIVIDE(CALCULATE(COUNT(Homicidios[Resolution]), Homicidios[Resolution] = "No"), [Total de Homicidios])

Se crea una medida que indica el año seleccionado:

Año Seleccionado = SELECTEDVALUE(Fechas[Año])

Se crea una medida para calcular la moda de la edad de los victimarios, es decir la edad más frecuente:

Moda Edad Victimarios =

VAR Grupos = SUMMARIZE(Victimarios, Victimarios[Perpetrator Age], "Total", COUNTROWS(Victimarios))

VAR Maximo = MAXX(Grupos,[Total])

VAR Moda = FILTER(Grupos, [Total] = Maximo)

RETURN

CONCATENATEX(Moda,Victimarios[Perpetrator Age], " – ")

Se crea otra medida para calcular el porcentaje de victimarios que ha sido reincidentes:

Reincidentes =

DIVIDE(CALCULATE(COUNTROWS(Victimarios), Victimarios[Perpetrator Count] > 0), COUNTROWS(Victimarios))

Se genera otra medida que muestra el trimestre con más homicidios en el año:

Trimestre con Mayor Homicidios =

VAR ResumenTrimestres =SUMMARIZE(Homicidios, Fechas[Trimestre], "Total Homicidios", COUNT(Homicidios[ID]))

VAR MaxHomicidios = MAXX(ResumenTrimestres, [Total Homicidios])

RETURN

MAXX(FILTER(ResumenTrimestres, [Total Homicidios] = MaxHomicidios), Fechas[Trimestre])

Se genera una medida que calcula el promedio de homicidios anual desde el 2006 hasta el 2014:

Promedio Homicidios por Año = AVERAGEX(VALUES(Fechas[Año]), [Total de Homicidios])

Ahora se crea una medida que calcula la resolución promedio de homicidios al año:

Promedio Casos Resueltos por Año =

AVERAGEX(SUMMARIZE('Homicidios', Fechas[Año]), [% Homicidios Resueltos])

Se crea otra medida que calcula el total de homicidios por trimestre:

Total de Homicidios por Trimestre =

SUMX(SUMMARIZE('Homicidios', Fechas[Trimestre]),[Total de Homicidios])

Se genera otra medida que muestra el trimestre donde se suele generar el mayor índice de homicidios en un año:

Trimestre con Mayor Frecuencia Homicidios =

VAR Grupo = SUMMARIZE (Homicidios, Fechas[Trimestre], "Frecuencia", [Total de Homicidios por Trimestre])

VAR Maximo = MAXX(Grupo, [Frecuencia])

VAR MaxTrim = FILTER(Grupo, [Frecuencia] = Maximo)

RETURN

CONCATENATEX(MaxTrim, Fechas[Trimestre])

Finalmente, se crean 3 medidas más para mostrar las conclusiones:

**Conclusion1 = VAR Texto1 = "El promedio anual es de " VAR PromedioAnual =
FORMAT([Promedio Homicidios por Año], "#,###") VAR Texto2 = " casos."**

RETURN Texto1 & PromedioAnual & Texto2

**Conclusion2 = VAR Texto1 = "El " VAR Texto2 = " de cada año es donde se suele
registrar la mayor cantidad de homicidios."**

RETURN Texto1 & [Trimestre con Mayor Frecuencia Homicidios] & Texto2

**Conclusion3 = VAR Texto1 = "La resolución promedio de homicidios al año es del "
VAR ResolucionAnual = FORMAT([Promedio Casos Resueltos por Año], "0.00%")**

RETURN Texto1 & ResolucionAnual

VISUALIZACIONES.

Antes que todo, se terminó de diseñar la página principal con la portada, y en la segunda página se agregó una hipótesis.



DATA ANALYTICS - COMISION 42290







HOMICIDIOS EN USA




ALUMNO: Gonzalo Burga




PROFESOR: Luciano Julián Gómez Olivera

Para cualquier consulta, comunicarse directamente conmigo, via WhatsApp





HOMICIDIOS EN USA

DESCRIPCIÓN

El dataset elegido, cuenta con más de 100,000 homicidios, perpetrados en distintos estados y localidades de USA, en un periodo de años entre 2006 hasta 2014. Cada uno de los homicidios cuenta con información sobre la agencia policial que lo cubrió, dónde ocurrió, fecha, el tipo de homicidio, si el homicidio fue resuelto o no, información, así como de la víctima como del victimario y el tipo de relación de estos últimos.


HIPÓTESIS

El presente análisis busca revelar patrones ya sea geográficos, temporales, etc. que ayuden a identificar zonas donde exista alta incidencia de homicidios, así como también factores sociales y demográficos asociados. Todo lo mencionado puede ayudar a elaborar medidas preventivas para poder reducir los índices de homicidio.

PROGRAMAS UTILIZADOS


MICROSOFT EXCEL

Los datos que se analizarán provienen de un archivo Excel, donde se puede leer el mismo.



MICROSOFT POWER POINT

Para la generación de fondos de las páginas del informe, se utilizó Power Point

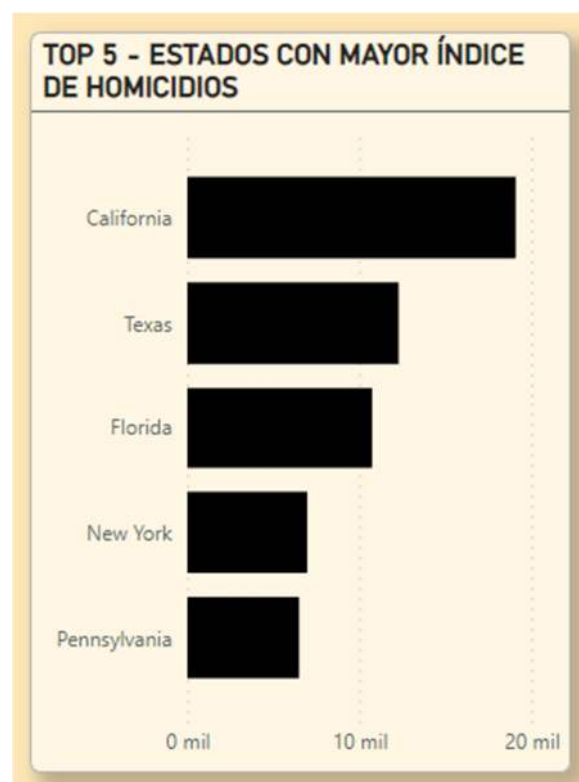


En la página “General 1” se generan las siguientes visualizaciones:

- Matriz con el total de homicidios, tasa de variación y porcentaje por años y con el tipo de homicidio.

Tipo de Crimen Año	Manslaughter by Negligence			Murder or Manslaughter			Total		
	Total	Variación	Participación	Total	Variación	Participación	Total	Variación	Participación
2006	212		0.15%	17,063		12.21%	17,275		12.37%
2007	307	44.81 % ↑	0.22%	16,996	-0.39 % ↓	12.17%	17,303	0.16 % ↑	12.39%
2008	91	-70.36 % ↓	0.07%	15,504	-8.78 % ↓	11.10%	15,595	-9.87 % ↓	11.16%
2009	347	281.32 % ↑	0.25%	15,493	-0.07 % ↓	11.09%	15,840	1.57 % ↑	11.34%
2010	279	-19.60 % ↓	0.20%	14,842	-4.20 % ↓	10.62%	15,121	-4.54 % ↓	10.82%
2011	293	5.02 % ↑	0.21%	14,463	-2.55 % ↓	10.35%	14,756	-2.41 % ↓	10.56%
2012	327	11.60 % ↑	0.23%	14,706	1.68 % ↑	10.53%	15,033	1.88 % ↑	10.76%
2013	300	-8.26 % ↓	0.21%	14,145	-3.81 % ↓	10.13%	14,445	-3.91 % ↓	10.34%
2014	317	5.67 % ↑	0.23%	14,014	-0.93 % ↓	10.03%	14,331	-0.79 % ↓	10.26%
Total	2,473		1.77%	137,226		98.23%	139,699		100.00%

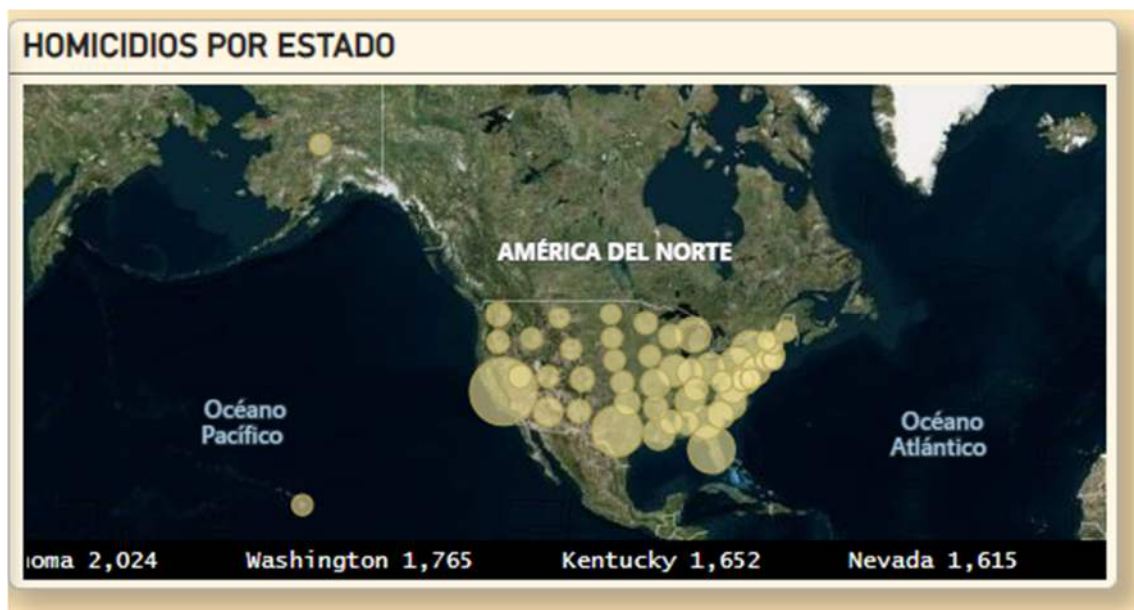
- Gráfico de columnas con el Top 5 de Estados con mayor cantidad de homicidios



- Top 3 de armas más usadas en los homicidios.



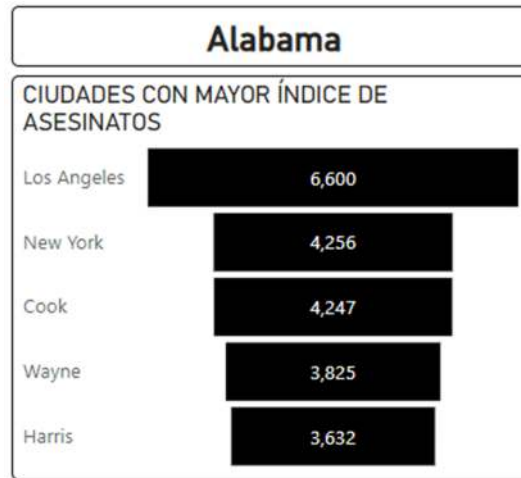
- Mapa con burbujas representando el total de homicidios por estado y un scroller en la parte inferior con el mismo dato.



- Tarjeta con el total de homicidios.



- Adicionalmente en esta página, se utiliza un tooltip asociado al mapa, que muestra las ciudades con mayor índice de homicidios, filtradas por estado.



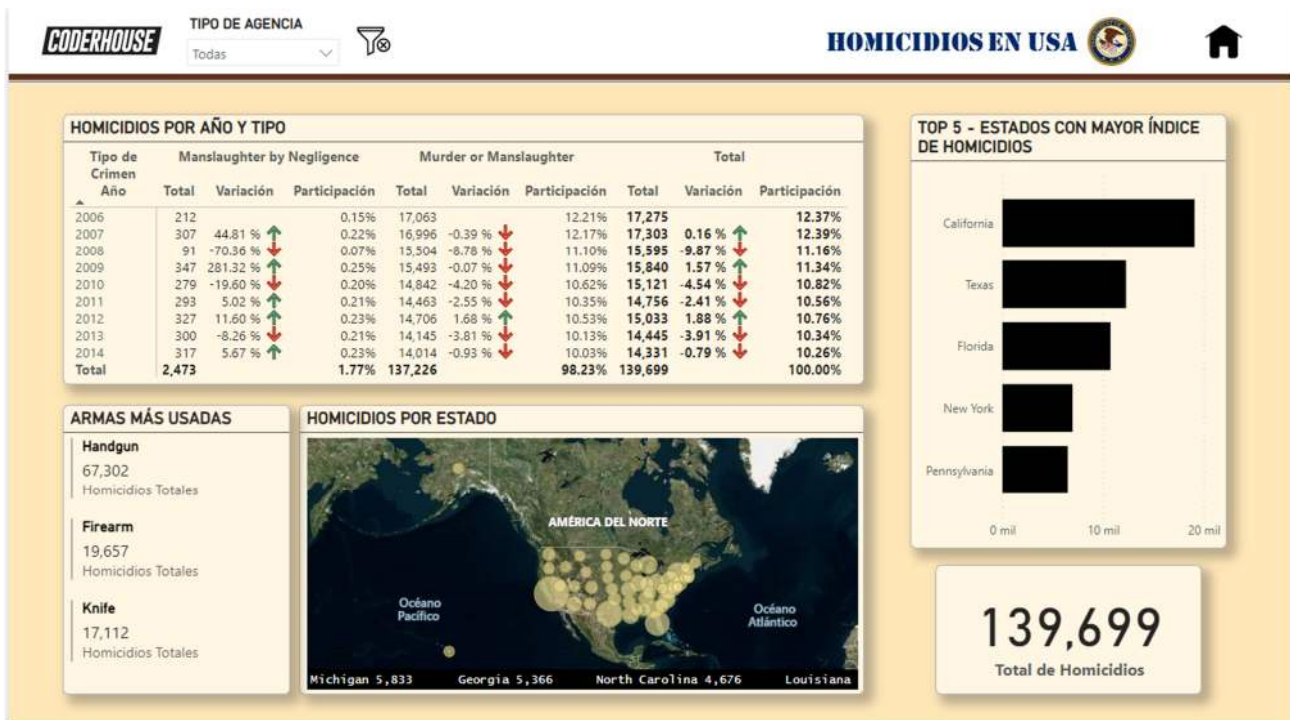
- Segmentación por tipo de agencia y un marcador para eliminar filtros

TIPO DE AGENCIA ▼

Todas ▼

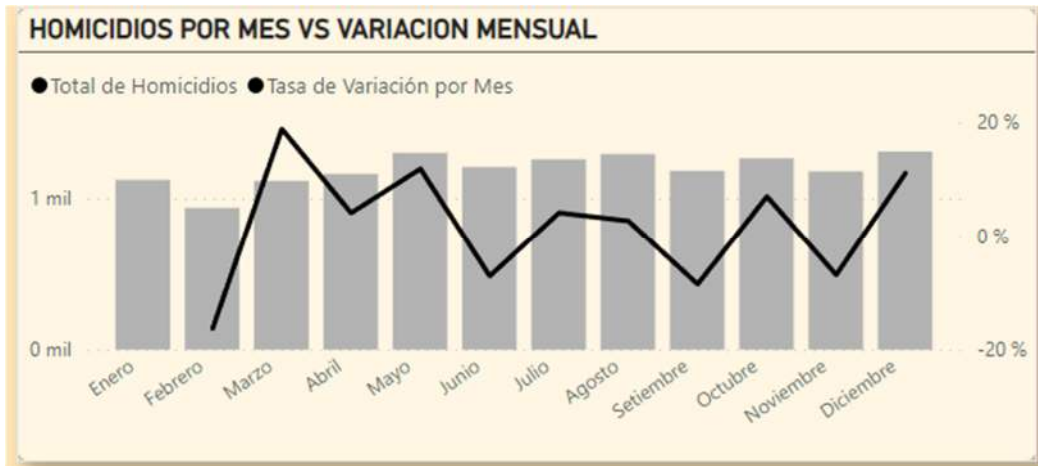
🗑️

De esta manera quedaría la primera página del informe:

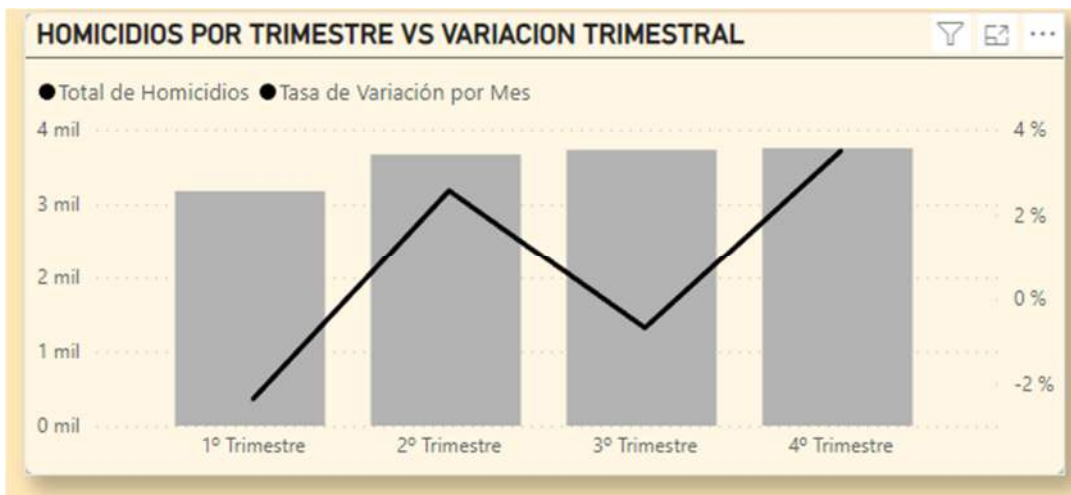


Si seleccionamos un año de la página “General 1”, y pulsamos en la opción Obtener detalles, nos dirigirá a la página “Detalle de Años”, donde están las siguientes visualizaciones:

- Gráfico de columnas agrupadas y de líneas con el total de homicidios por mes y su tasa de variación mensual.



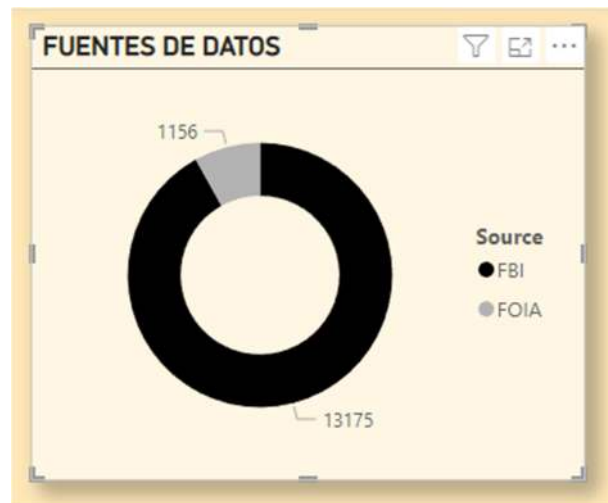
- Gráfico de columnas agrupadas y de líneas con el total de homicidios por trimestre y su tasa de variación trimestral.



- Tarjeta que muestra el año seleccionado y el total de homicidios en ese año.



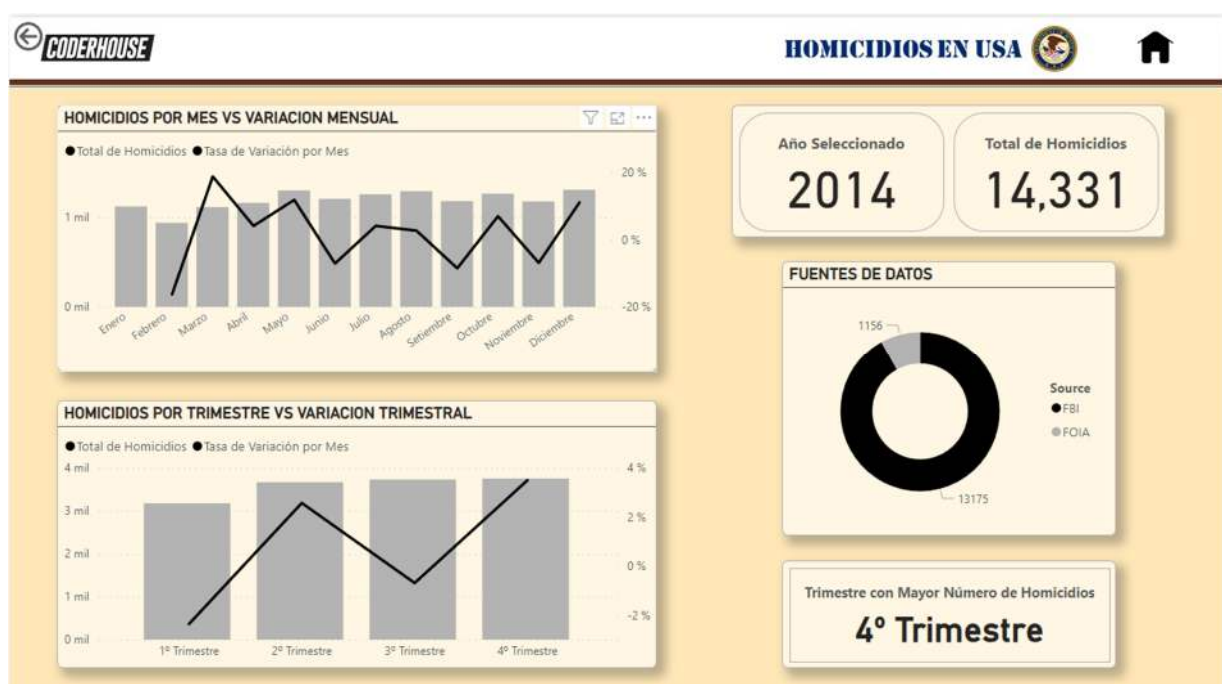
- Gráfico de anillos que muestra de donde se ha obtenido la información de los casos.



- Tarjeta que muestra el trimestre con mayor índice de homicidios en el año seleccionado.

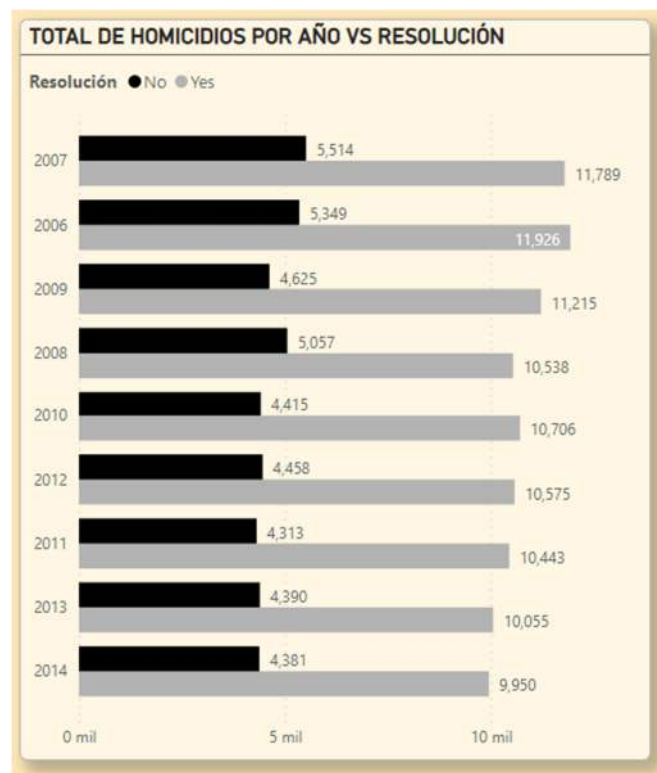


La página “Detalle de Años” quedaría de esta manera:



En la página “General 2” se genera lo siguiente:

- Gráfico de barras agrupadas que muestra el total de homicidios resueltos y no resueltos por año.



- Medidores que muestran el total de homicidios resueltos y no resueltos junto con un marcador para cambiar los medidores a porcentaje.



- Medidores que muestran el porcentaje de homicidios resueltos y no resueltos con respecto al total junto con un marcador para cambiar los medidores a cantidades.



- Top 3 de Estados con mayor y menor ratio de homicidios resueltos.

TOP 3 - ESTADOS CON MAYOR RATIO DE HOMICIDIOS RESUELTOS	
North Dakota	97.89 %
Montana	94.58 %
Idaho	93.73 %

TOP 3 - ESTADOS CON MENOR RATIO DE HOMICIDIOS RESUELTOS	
Illinois	38.07 %
District of Columbia	50.61 %
New Jersey	54.98 %

- Segmentaciones para filtrar por Estado y por mes y un marcador para eliminar los filtros.

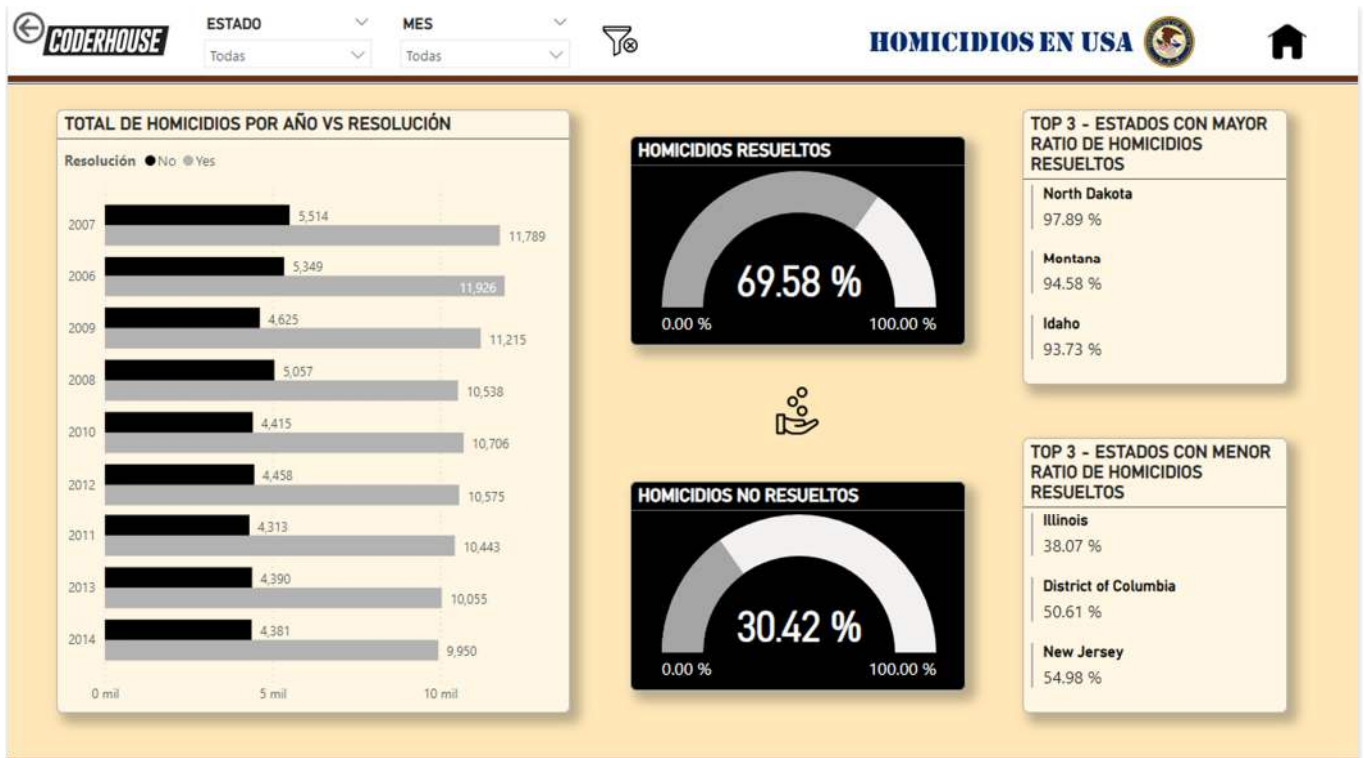
ESTADO

Todas

MES

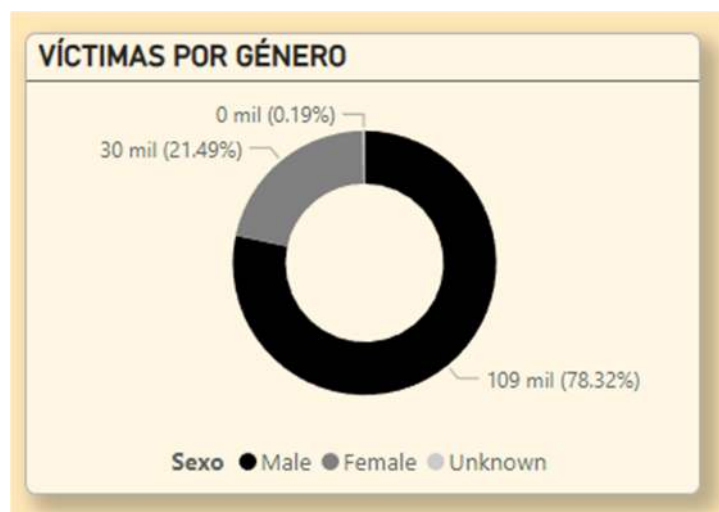
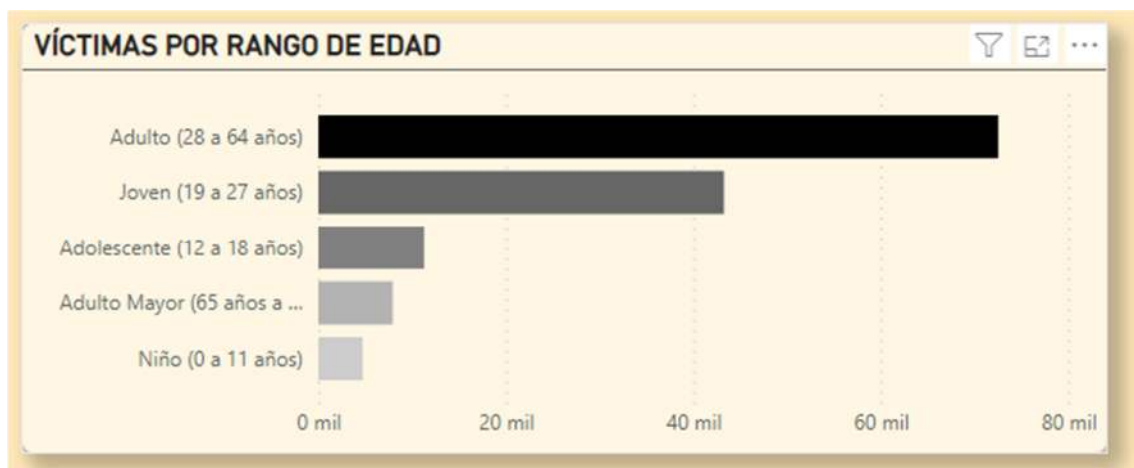
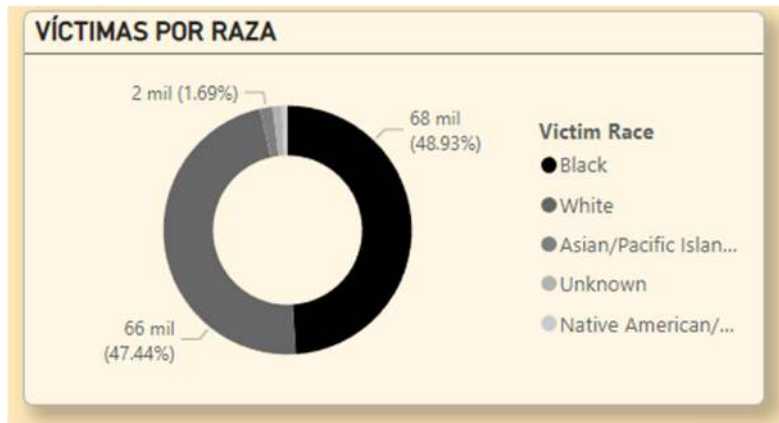
Todas

De esta manera quedaría la página del informe:

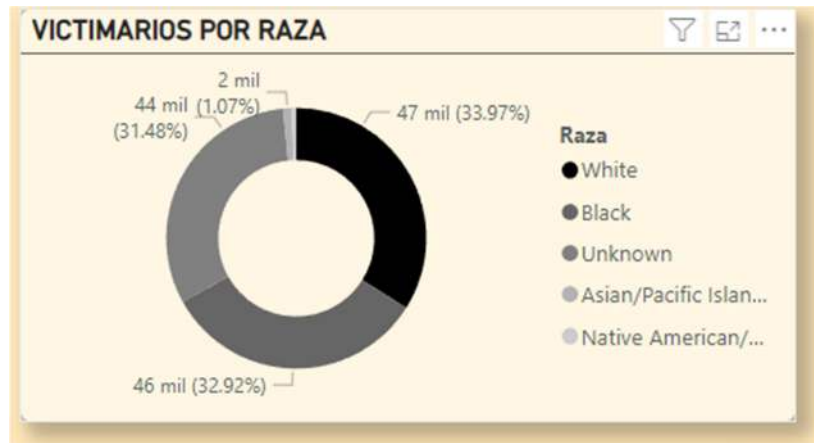


En la página “Víctimas y Victimarios” se genera lo siguiente:

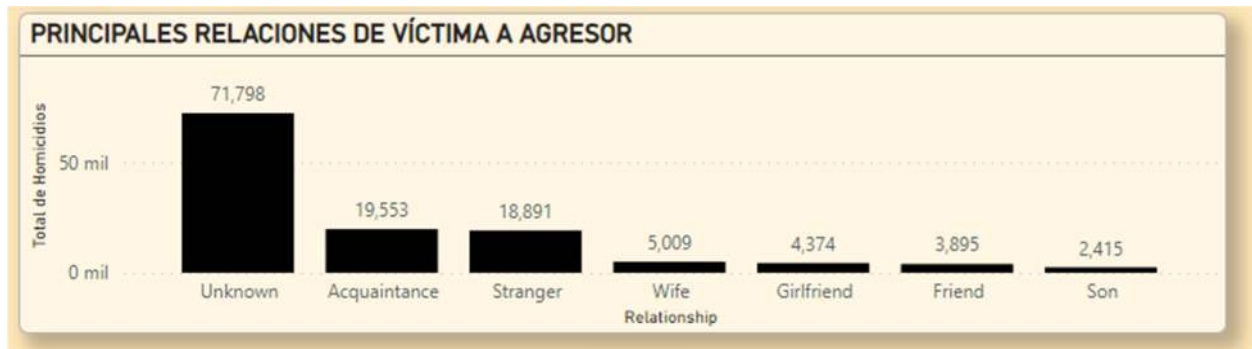
- Gráfico de anillos que muestra los homicidios de las víctimas por raza, gráfico de barras que muestra homicidios de víctimas por rango de edad y gráfico de anillos que muestra los homicidios de víctima por género.



- Mediante el uso de un marcador, todos los gráficos anteriores pasan a mostrar información sobre los victimarios.



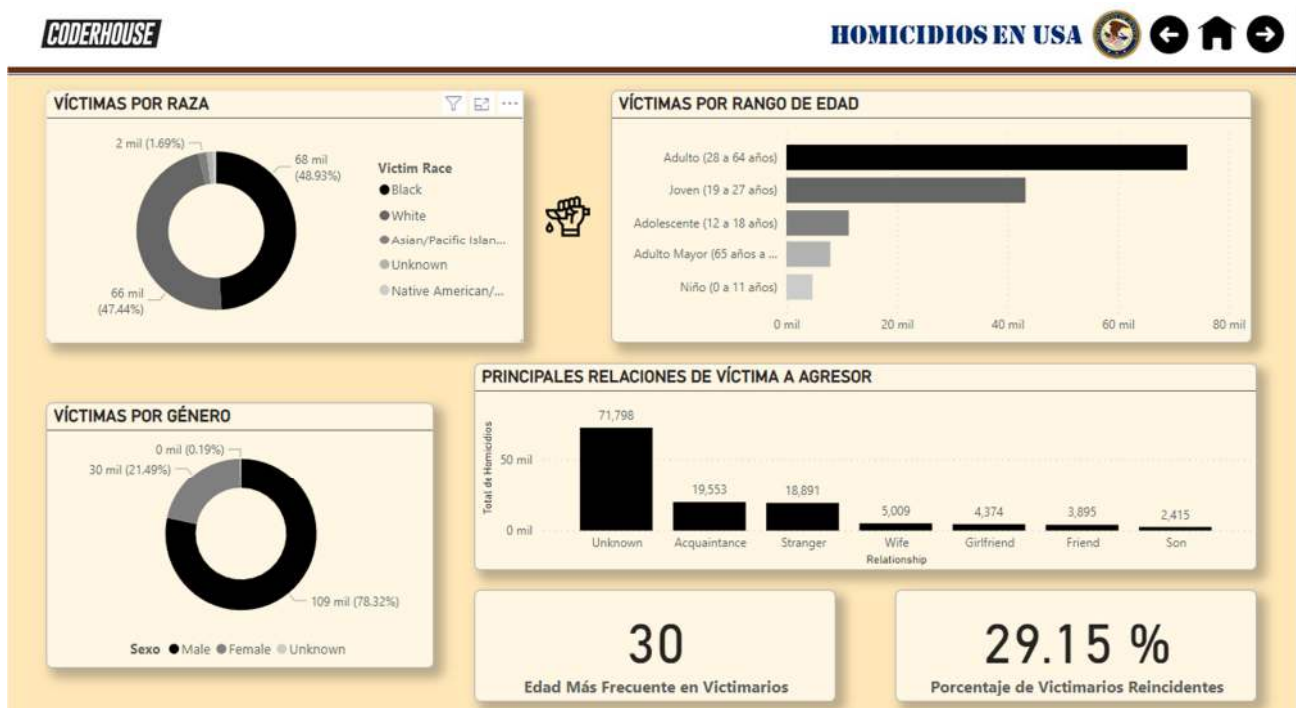
- Gráfico de columnas con las principales relaciones entre la víctima y el agresor.



- Tarjetas que muestran la edad más frecuente en victimarios y el porcentaje de victimarios reincidentes.



De esta manera quedaría la página:



Finalmente, en la página conclusiones va lo siguiente:

- Tarjetas que muestran las 3 conclusiones generadas en una medida:

El promedio
anual es de
15,522 casos.

La resolución promedio
de homicidios al año es
del 69.61%

El 3° Trimestre de cada
año es donde se suele
registrar la mayor cantidad
de homicidios.

De esta manera quedaría la página “Conclusiones”:



C
O
N
C
L
U
S
I
O
N
E
S

El promedio
anual es de
15,522 casos.

La resolución promedio
de homicidios al año es
del 69.61%

El 3° Trimestre de cada
año es donde se suele
registrar la mayor cantidad
de homicidios.



CONCLUSIÓN.

Considerando el objetivo inicial de identificar en qué época o estados es dónde se produce la mayor cantidad de homicidios en el año, se puede observar que generalmente, en cada 3er trimestre del año hay un alza considerada de homicidios, por lo que se debería tener en cuenta para reforzar seguridad o idear estrategias de seguridad en esas fechas. Asimismo, se puede observar que en los estados de California, Texas y Florida es donde hay mayor índice de homicidios, por lo que habría que incrementar la seguridad de esas zonas.