



**UNIVERSIDAD
CATÓLICA
DE CÓRDOBA**
JESUITAS

Final Base de Datos II

Docente: García Mattío, Mariano

Autores:

Murature, Lucio Antonio

(1701480@ucc.edu.ar)

Schröder, Matías Celestino

(1708485@ucc.edu.ar)

Bases de Datos II

Ciudad de Córdoba, 5 de Diciembre de 2019.

Introducción	4
Análisis de Requerimientos	5
Preguntas de negocio	5
Indicadores y perspectivas	5
Modelo conceptual	5
Análisis de Data Sources	7
Hechos e Indicadores	7
Mapeo	7
Granularidad	8
Modelo Conceptual Ampliado	9
Modelo Lógico del DW	10
Tipología	10
Tablas de dimensiones	10
Tabla de Hechos	11
Uniones	12
Integración de Datos	13
Carga inicial	13
Actualización	19
Cubo multidimensional	20
Algunas conclusiones...	23
Componentes BI	24
Respuestas encontradas	24
Reportes	28
Reporte 1	28
Reporte 2	29
Reporte 3	30
Reporte 4	32
Reporte 5	33
Tableros	35
Tablero 1	35
Tablero 2	38
Tablero 3	40

Introducción

Este trabajo es realizado para la materia Base de Datos II por Matías Celestino Schröder y Lucio Antonio Murature con el fin de utilizar la metodología Hefesto y las herramientas de Pentaho Server y Report Designer las cuales han sido inculcadas por el profesor Mariano García Mattío.

A la hora de especificar un dataset para trabajar estuvimos indagando en las distintas posibilidades que se encuentran en internet hasta llegar “Gun Violence Data” en el cual se especifica una gran cantidad de datos sobre incidentes de violencia armada alrededor de Estados Unidos.

La mayor cantidad de estos datos fueron obtenidos a partir de la técnica de scrapping a la página perteneciente a Gun Violence Archive la cual contiene datos desde Enero de 2013 a Marzo de 2018 inclusive.

Somos conscientes de que hay muchos incidentes no registrados en este dataset. Pero al mismo tiempo pensamos que hay cantidad suficiente para poder hacer una investigación de violencia armada y de poder hacer predicciones sobre tendencias futuras.

Para poder utilizar la metodología Hefesto vamos a suponer que nuestro cliente es el gobierno de Estados Unidos y quiere obtener información de distintos patrones dentro de la violencia armada.

Análisis de Requerimientos

Preguntas de negocio

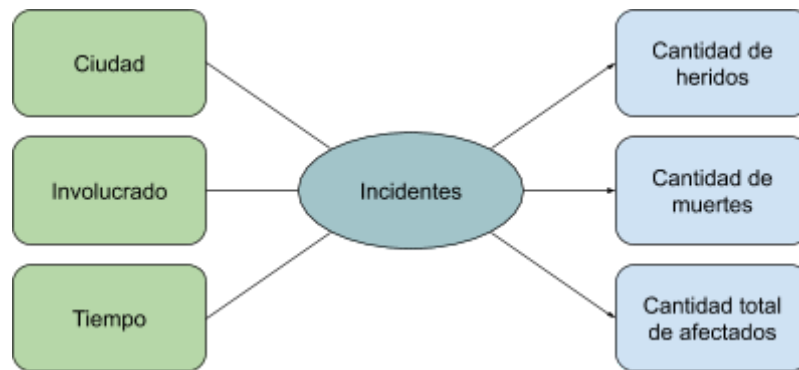
Como mencionamos anteriormente nuestro cliente supuestamente es el gobierno de Estados Unidos. Decidimos centrarnos en las distintas zonas donde ocurrieron los incidentes para ver en cuál de ellos debería aumentar el control, además de saber efectivamente si se están formando pandillas desde pequeña edad. Las preguntas que formulamos son los siguientes:

- ¿Qué ciudad posee la mayor cantidad de afectados por incidentes con armas? ¿Qué ciudad la menor? ¿Cuál es la cantidad de heridos y muertes?
- ¿Qué Estado posee la mayor cantidad de afectados por incidentes con armas? ¿Qué Estado la menor? ¿Cuál es la cantidad de heridos y muertes?
- ¿Qué sexo (Masculino o Femenino) tiene más carga de afectados en los Estados Unidos?
- ¿Qué edad es la predominante dentro de los casos de violencia con armas?
- ¿En qué año fueron más personas afectadas con violencia de armas?

Indicadores y perspectivas

Teniendo en cuenta estas preguntas, podemos afirmar que la cantidad de incidentes, la cantidad de heridos y la cantidad de muertes son los indicadores. Mientras que las perspectivas de análisis son respecto la ciudad (la que incluye estado), el tiempo e involucrado (incluyendo sexo y edad).

Modelo conceptual



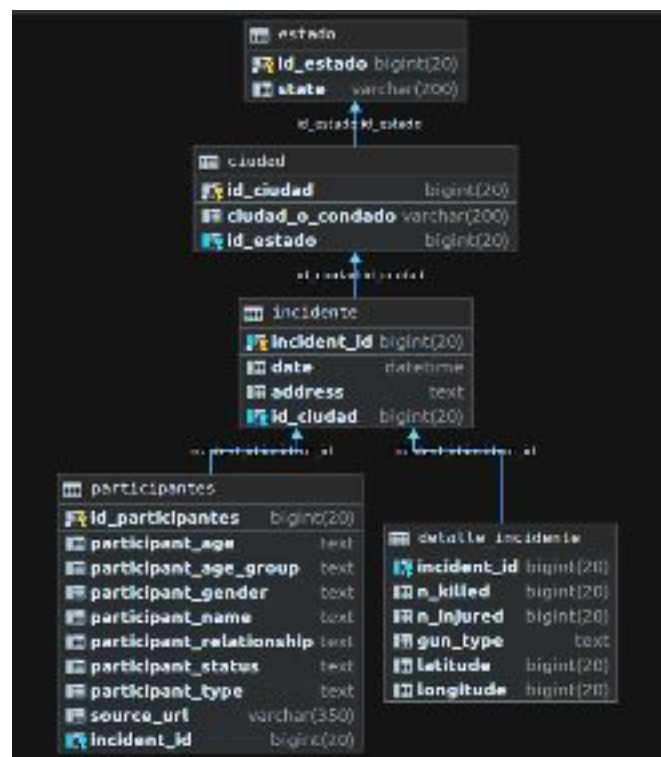
Análisis de Data Sources

Hechos e Indicadores

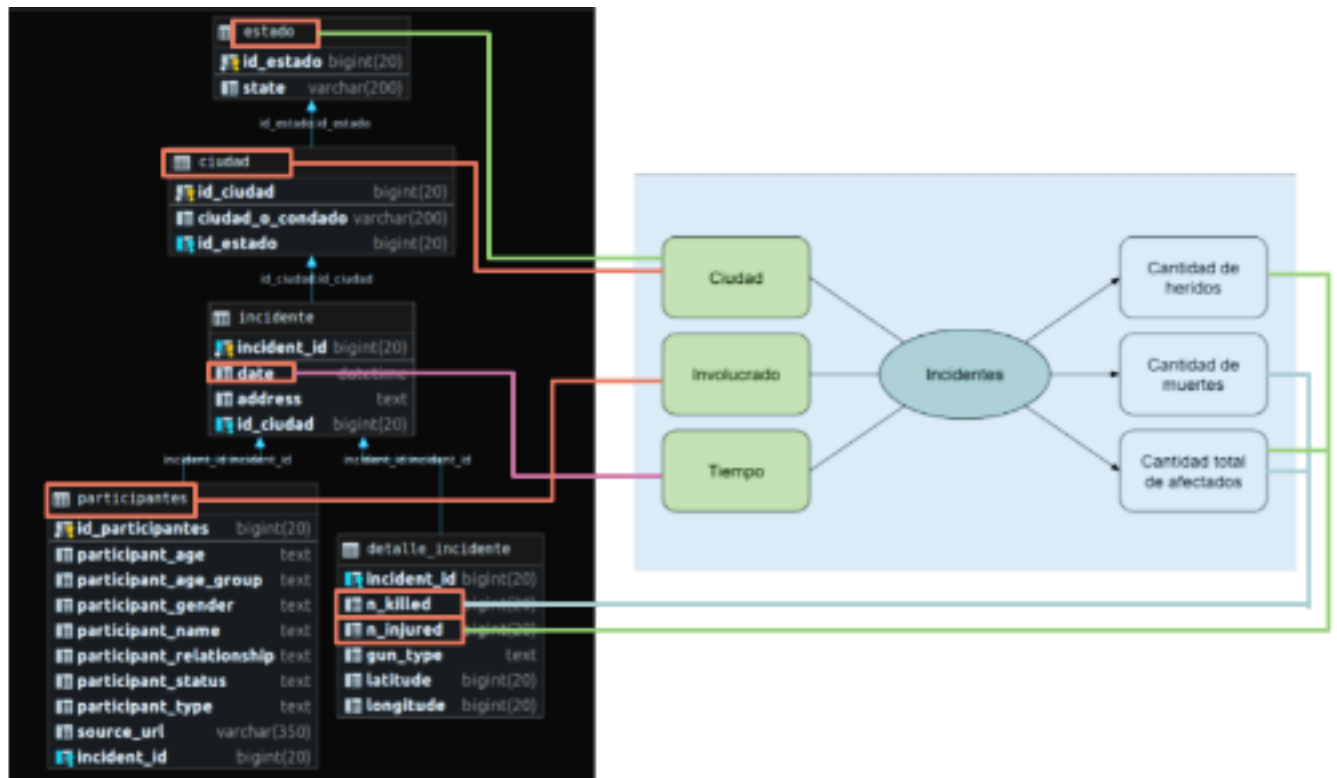
- Indicador: Cantidad de personas afectadas
 - Hechos: Cantidad de heridos + Cantidad de muertes
 - Función de agregación: SUM
- Indicador: Cantidad de personas heridas
 - Hechos: Cantidad de heridos por incidente
 - Función de agregación: SUM
- Indicador: Cantidad de personas muertas
 - Hechos: Cantidad de muertes por incidente
 - Función de agregación: SUM

Mapecto

Diagrama relacional:



Por lo tanto el mapeo entre los dos modelos sería el siguiente:



El mapeo realizado es el siguiente:

- La perspectiva Ciudad se relaciona con la tabla Ciudad y la tabla Estado.
- La perspectiva Tiempo se relaciona con el campo date de la tabla incidente.
- La perspectiva de Involucrado se relaciona con la tabla participantes.
- Los indicadores Cantidad de muertes, Cantidad de heridos y Cantidad de afectados se relacionan con los campos n_killed y n_injured de la tabla detalle_incidentes.

Granularidad

Con respecto a la Perspectiva Ciudad (la cual incluye la tabla ciudad y la tabla estado), los datos disponibles son los siguientes:

- `id_ciudad`: es la clave primaria de la tabla Ciudad, y representa unívocamente a una ciudad o condado en particular
- `Ciudad_o_condado`: posee el nombre de la ciudad o el condado
- `id_estado`: es la clave primaria del Estado, y representa unívocamente a un estado en particular
- `state`: posee el nombre del estado.

Con respecto a la Perspectiva de Involucrado, los datos disponibles son los siguientes:

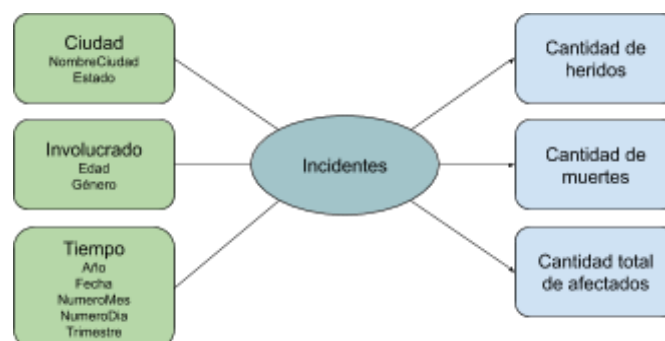
- `id_participantes`: es la clave primaria de participante, y representa unívocamente a un participante o grupo de participantes.
- `participant_age`: es un dato que representa la edad exacta uno de los involucrados en el incidente.
- `participant_age_group`: representación de la edades de los distintos involucrados.
- `participant_gender`: sexo del participante principal del incidente.
- `participant_name`: nombre del participante principal
- `participant_relationship`: relación que tenía el participante del incidente, respecto los heridos y muertos, por ejemplo, familiar, amigo o conocido.
- `participant_status`: estado después del incidente, refiriéndose a la salud.
- `participant_type`: culpable o sospechoso, según investigaciones.
- `source_url`: website de noticia sobre el incidente.

Con respecto a la perspectiva Tiempo, los datos disponibles son los siguientes:

- Fecha.
- Año.
- Trimestre.
- Número de mes.
- Número de día.

En este proceso no se tendrán en cuenta los datos de las horas, minutos y segundos, presentes en el dataset, debido a que los mismos no eran representativos. Se elige como mínima unidad de tiempo para no perder ningún dato relevante el “día”,

Modelo Conceptual Ampliado



Modelo Lógico del DW

Tipología

Se ha seleccionado el Esquema en Estrella ya que cumple con los requerimientos planteados y es simple de implementar y comprender.

Tablas de dimensiones

Perspectiva Ciudad:

- La nueva tabla de Dimensión tendrá como nombre dim_ciudad.
- Se agregará una clave principal con el nombre id_ciudad.
- Se modificará el nombre del campo NombreCiudad por nombre_ciudad.
- Se modificará el nombre del campo Estado por estado.

Perspectiva Involucrado:

- La nueva tabla de Dimensión tendrá como nombre dim_involucrado.
- Se agregará una clave principal con el nombre id_involucrado.
- Se modificará el nombre del campo Edad por edad.
- Se modificará el nombre del campo Género por genero.

Perspectiva Tiempo:

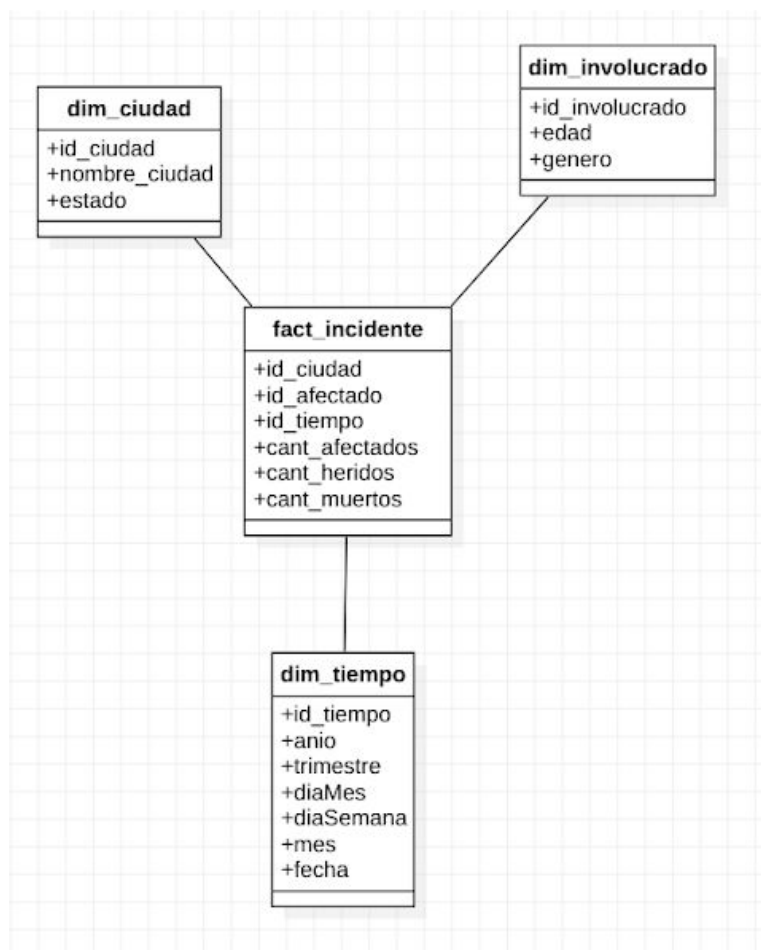
- La nueva tabla de Dimensión tendrá el nombre dim_tiempo.
- Se le agregará una clave principal con el nombre id_tiempo.
- Se modificará el nombre del campo Año por anio.
- Se modificará el nombre del campo Fecha por fecha.
- Se modificará el nombre del campo Trimestre por trimestre.
- Se modificará el nombre del campo NumeroMes por mes.
- Se modificará el nombre del campo NumeroDia por diaMes.

Tabla de Hechos

A continuación, se confeccionará la tabla de Hechos:

- La tabla de Hechos tendrá el nombre fact_incidentes.
- Su clave principal será la combinación de las claves principales de las tablas de Dimensiones antes definidas: id_ciudad, id_involucrado e id_tiempo.
- Se crearán dos Hechos, que se corresponden con los dos Indicadores:
 - Cantidad de heridos será indicado como cant_heroidos.
 - Cantidad de muertos será indicado como cant_muertos.
 - Cantidad total de personas afectadas como cant_afectados.

Uniones



Integración de Datos

Carga inicial

Proceso ETL Principal

El proceso ETL principal planteado para la Carga Inicial es el siguiente:



- Start: inicia la ejecución de los pasos en el momento en que se le indique.
- Las variables Fecha_Desde y Fecha_Hasta no son utilizadas en este DataWarehouse ya que se utiliza la totalidad de los datos a partir del dataset, que conforma desde el año 2013 hasta 2018.
- Carga de Dimensión dim_ciudad: ejecuta el contenedor de pasos que cargará la tabla de Dimensión dim_ciudad (más adelante se detallará).
- Carga de Dimensión dim_tiempo: ejecuta el contenedor de pasos que cargará la tabla de Dimensión dim_tiempo (más adelante se detallará).
- Carga de Dimensión dim_involucrado: ejecuta el contenedor de pasos que cargará la tabla de Dimensión dim_involucrado (más adelante se detallará).
- Carga de Dimensión fact_incidente: ejecuta el contenedor de pasos que cargará la tabla de Dimensión fact_incidente (más adelante se detallará).

Carga de Dimensión dim_ciudad

A continuación, se especificarán las tareas llevadas a cabo por Carga de Dimensión dim_ciudad. Este Paso es un Contenedor de Pasos, así que incluye las siguientes tareas:



Los datos serán obtenidos a partir de la siguiente consulta SQL a la base de datos:

```
SELECT c.id_ciudad, c.city_or_county, e.state  
FROM ciudad c inner join estado e on c.id_estado=e.id_estado;
```

Cargando se utiliza el paso “Insert/Update” de Pentaho Data Integration, bajo la siguiente configuración.

Insertar / actualizar

Nombre del paso: Insertar / Actualizar

Conexión: dw [Editar...] [Nuevo...] [Wizard...]

Esquema destino: [Examinar...]

Tabla destino: dim_ciudad [Examinar...]

Tamaño de transacción (commit): 100

No realizar actualizaciones: ☐

La(s) clave(s) para realizar la búsqueda de valor(es):

Campos de tabla	Comparador	Campos de Flujo1	Campos de Flujo2
1 id_ciudad	=	id_ciudad	
2 nombre_ciudad	=	city_or_county	
3 estado	=	state	

[Obtener campos]

Campos de actualización:

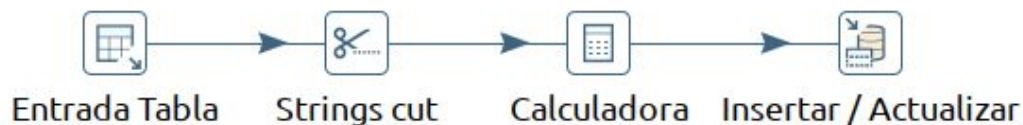
Campos de tabla	Campos de Flujo	Actualizar
1 id_ciudad	id_ciudad	Y
2 nombre_ciudad	city_or_county	Y
3 estado	state	Y

[Obtener campos de actualización] [Editar mapeo]

[Help] [OK] [Cancelar] [SQL]

Carga Dimensión dim_involucrado

Para la carga sobre la dimensión de Involucrado, se hará una consulta SQL sobre la base de datos, se hará un proceso de transformación de los datos para género y para la edad, y luego se insertarán en la tabla de dim_involucradodim_tiempo.



Para la obtención de los datos, se utilizara la siguiente consulta:

```

SELECT p.id_participantes, p.participant_age, p.participant_gender
FROM participantes p;
    
```

Para la carga se utiliza “Insert/Update” al igual que en la carga de dim_ciudad:

Nombre del paso: Insertar / Actualizar

Conexión: dw

Esquema destino: [vacío]

Tabla destino: dim_involucrado

Tamaño de transacción (commit): 100

No realizar actualizaciones: ☐

La(s) clave(s) para realizar la búsqueda de valor(es):

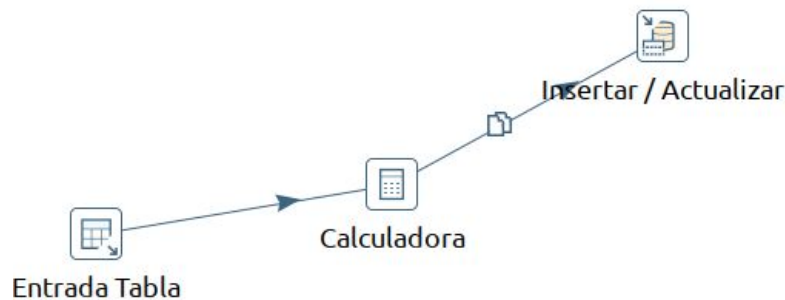
Campos de actualización:
1 id_involucrado = id_participantes

Campos de actualización:

Campos de actualización:
1 id_involucrado id_participantes N
2 edad edad_1 Y
3 genero genero Y

Carga Dimensión dim_tiempo

Al igual que en los demás pasos, para la carga de la siguiente dimensión se utilizara una consulta SQL sobre la base de datos, se hará una transformación para la fecha, y se insertará/actualizará en el modelo estrella, bajo una id generada a partir de la fecha.



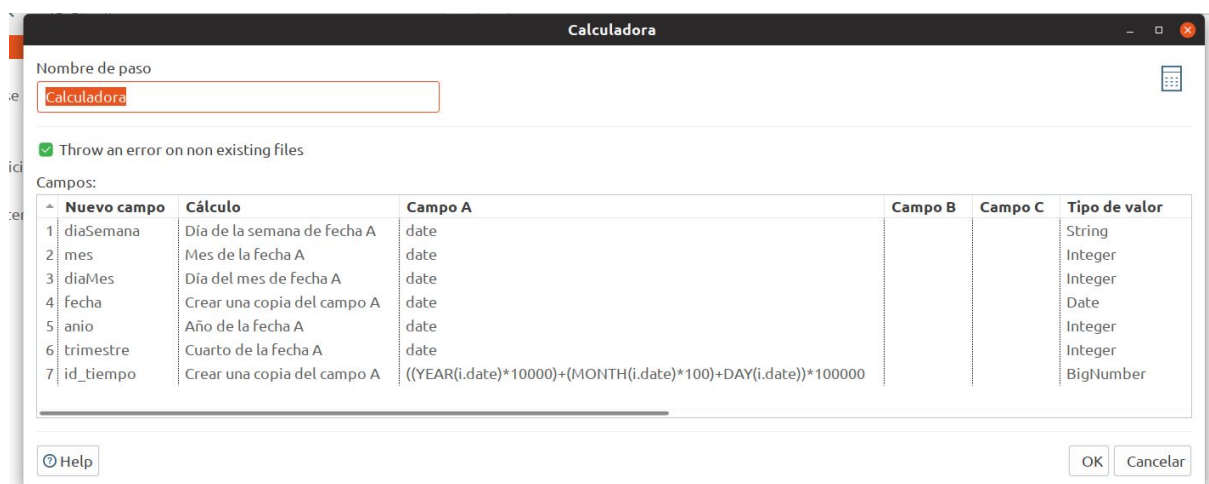
La obtención de datos se hará a partir de la siguiente consulta a la tabla incidente:

```

SELECT i.date,
((YEAR(i.date)*10000)+(MONTH(i.date)*100)+DAY(i.date))*100000
FROM incidente i;
    
```

Nótese el cálculo que se realiza a partir de la fecha del incidente, a partir de ésta y de la granularidad elegida, se conformará la idTiempo como clave primaria de la dimensión de tiempo, para identificar unívocamente cada día de cada año.

El step “Calculadora” nos va a ayudar a filtrar todos los datos que queremos de todas las fechas presentes en nuestra base de datos, con la siguiente configuración.



Para la inserción en la dimensión utilizamos, nuevamente, el paso “Insert/Update” bajo la siguiente configuración.

Insertar / actualizar

Nombre del paso: Insertar / Actualizar

Conexión: dw [Editar...] [Nuevo...] [Wizard...]

Esquema destino: [Examinar...]

Tabla destino: dim_tiempo [Examinar...]

Tamaño de transacción (commit): 100

No realizar actualizaciones: ☐

La(s) clave(s) para realizar la búsqueda de valor(es):

	Campo de tabla	Comparador	Campo de flujo1
1	id_tiempo	=	id_tiempo

[Obtener campos]

Campos de actualización:

	Campo de tabla	Campo de Flujo
1	fecha	fecha
2	diaSemana	diaSemana
3	mes	mes
4	diaMes	diaMes
5	anio	anio
6	trimestre	trimestre
7	id_tiempo	id_tiempo

[Obtener campos de actualización] [Editar mapeo]

[Help] [OK] [Cancelar] [SQL]

Carga Tabla de Hechos incidentes

Para ésta carga será imperativo utilizar varias uniones entre tablas, dada la necesidad de sacar las claves primarias que utilizamos en las dimensiones. Por más complicado que se vea, la transformación de Pentaho Data Integration es de lo más simple.



La obtención de datos está dada por la siguiente consulta SQL sobre la base de datos:

```
SELECT c.id_ciudad,  
((YEAR(i.date)*10000)+(MONTH(i.date)*100)+DAY(i.date))*100000 as id_tiempo,  
e.id_estado, sum(di.n_killed)+sum(di.n_injured) as cant_afectados, sum(di.n_killed)  
as cant_muertos, sum(di.n_injured) as cant_heridos  
, p.id_participantes as id_involucrado  
FROM estado e inner join ciudad c on e.id_estado = c.id_estado inner join incidente i  
on c.id_ciudad = i.id_ciudad inner join detalle_incidente di on i.incident_id =  
di.incident_id  
inner join participantes p on i.incident_id = p.incident_id  
GROUP BY c.id_ciudad,  
((YEAR(i.date)*10000)+(MONTH(i.date)*100)+DAY(i.date))*100000, e.id_estado,  
id_involucrado;
```

Si se lee con detenimiento, se puede dar cuenta cómo se relaciona el hecho de incidente con la dimensión tiempo, a partir del mismo cálculo para la idTiempo sobre la fecha del incidente.

A su vez, es imperativo también identificar los indicadores.

Para el step de “Insert/Update” se utiliza la siguiente configuración:

Nombre del paso: Insertar / Actualizar

Conexión: dw

Esquema destino:

Tabla destino: fact_incidente

Tamaño de transacción (commit): 100

No realizar actualizaciones: ☐

La(s) clave(s) para realizar la búsqueda de valor(es):

	Campo de tabla	Comparador	Campo de flujo1	Campo de flujo2
1	id_ciudad	=	id_ciudad	
2	id_tiempo	=	id_tiempo	
3	id_estado	=	id_estado	
4	id_involucrado	=	id_involucrado	

Obtener campos

Campos de actualización:

	Campo de tabla	Campo de Flujo	Actualizar
1	id_ciudad	id_ciudad	N
2	id_tiempo	id_tiempo	N
3	id_estado	id_estado	N
4	cant_afectados	cant_afectados	Y
5	latitud	latitud	Y
6	longitud	longitud	Y
7	id_involucrado	id_involucrado	N
8	cant_muertos	cant_muertos	Y
9	cant_heridos	cant_heridos	Y

Obtener campos de actualización

Editar mapeo

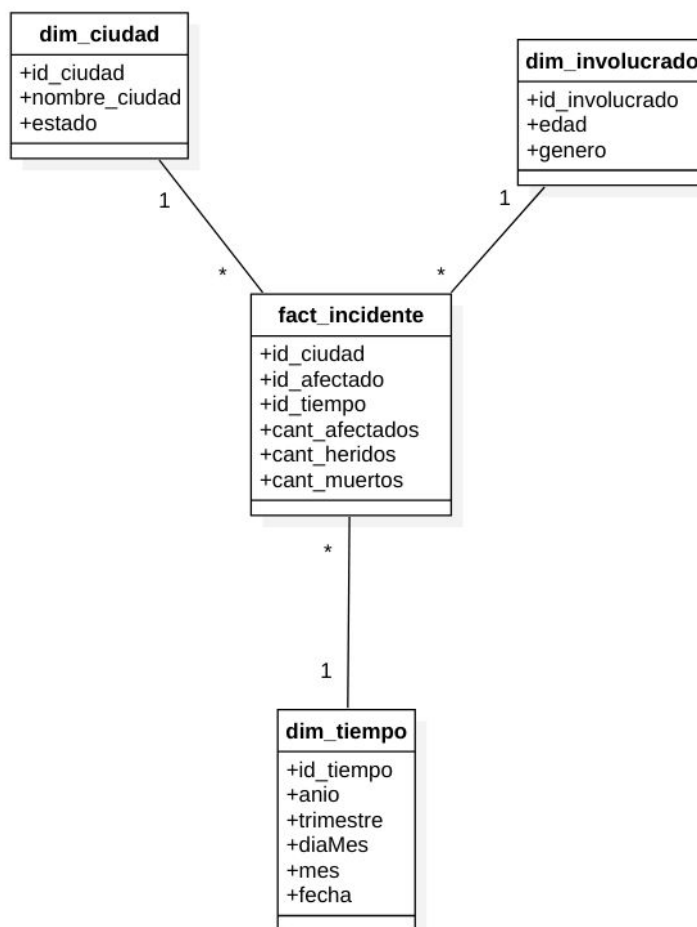
Help OK Cancelar SQL

Actualización

Los datos serán actualizados cada año, debido a que todos los años se publican actualizaciones en los datasets.

Para dicha actualización, se utilizará el mismo trabajo, que ejecuta todas las transformaciones previamente explicadas, sin colisionar datos, a partir del paso “Insert/Update” previamente mencionado.

Cubo multidimensional

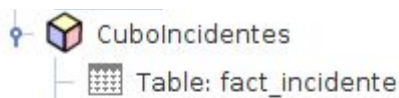


Pasos básicos

1. Se creará un Cubo Multidimensional llamado **CuboIncidentes**

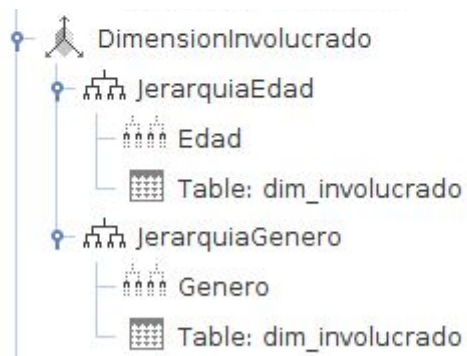


2. Se indicará que la tabla de Hechos utilizar es **fact_incidente**



3. Se creará un Dimensión para analizar a los involucrados. Su nombre será **DimensionInvolucrado**.
 - a. Se agregara una Jerarquía que se llamará **JerarquiaEdad**
 - i. Se indicará que JerarquiaEdad estará basada en los campos de la tabla de Dimensión **dim_involucrado**

- ii. Se añadirá a la JerarquiaEdad el atributo Edad que estará basado en el campo **edad**.
- b. Se añadirá una jerarquía llamada **JerarquiaGenero**
 - i. Se indicará que JerarquiaGenero estará basada en los campos de la tabla de Dimensión **dim_involucrado**
 - ii. Se añadirá a la JerarquiaGenero el atributo Genero que estará basado en el campo **genero**.



- 4. Se creará un Dimensión para analizar a las ubicaciones. Su nombre será **DimensionCiudad**.

- a. Se añadirá una jerarquía llamada JerarquiaCiudad
 - i. Se indicará que JerarquiaCiudad estará basada en los campos de la tabla de Dimensión **dim_ciudad**
 - ii. Se añadirá a la JerarquiaCiudad el atributo Estado que estará basado en el campo **estado**.
 - iii. Se añadirá a la JerarquiaCiudad el atributo Ciudad que estará basado en el campo **ciudad**. En donde existe una relación **padre-hijo** entre estado y ciudad respectivamente.

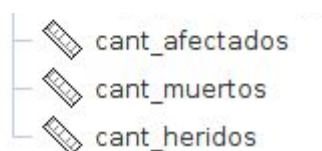


- 5. Se creará un Dimensión para analizar el tiempo. Su nombre será **DimensionTiempo**.

- a. Se añadirá una jerarquía llamada **JerarquiaTiempo**
- i. Se indicará que JerarquiaTiempo estará basada en los campos de la tabla de Dimensión **dim_tiempo**
 - ii. Se añadirá a la JerarquiaTiempo el atributo Anio que estará basado en el campo **anio**.
 - iii. Se añadirá a la JerarquiaTiempo el atributo Trimestre que estará basado en el campo **trimestre**. En donde existe una relación **padre-hijo** entre anio y trimestre respectivamente.
 - iv. Se añadirá a la JerarquiaTiempo el atributo Mes que estará basado en el campo **mes**. En donde existe una relación **padre-hijo** entre trimestre y mes respectivamente.
 - v. Se añadirá a la JerarquiaTiempo el atributo Dia que estará basado en el campo **diaNumero**. En donde existe una relación **padre-hijo** entre mes y dia respectivamente.



6. Se creará un Indicador, llamado **cant_afectados** que se calculará de la siguiente manera: $SUM(n_killed) + SUM(n_injured)$
7. Se creará un Indicador, llamado **cant_muertos** que se calculará de la siguiente manera: $SUM(n_killed)$
8. Se creará un Indicador, llamado **cant_heridos** que se calculará de la siguiente manera: $SUM(n_injured)$



Componentes BI

En esta sección del documento se plasmarán distintos componentes de Inteligencia de Negocios desarrolladas en la materia aplicadas a nuestro caso de estudio. En primer lugar, responderemos las preguntas planteadas en el desarrollo de la metodología, luego se mostrarán distintos tipos de reportes y tableros interactivos.

Respuestas encontradas

¿Qué ciudad posee la mayor cantidad de afectados por incidentes con armas? ¿Qué ciudad la menor? ¿Cuál es la cantidad de heridos y muertes?

Para responder esta pregunta se le realizará la siguiente consulta MDX al cubo multidimensional:

```
SELECT {[Measures].[cant_afectados], [Measures].[cant_muertos],  
[Measures].[cant_heridos]} ON COLUMNS,  
Order(Union(Union({[DimensionCiudad.JerarquiaCiudad].[All  
DimensionCiudad.JerarquiaCiudades]}, [DimensionCiudad.JerarquiaCiudad].[All  
DimensionCiudad.JerarquiaCiudades].Children),  
[DimensionCiudad.JerarquiaCiudad].[Illinois].Children),  
[Measures].[cant_afectados], DESC) ON ROWS FROM [CuboIncidentes]
```

Los resultados arrojan que la ciudad de **Chicago** en el estado de Illinois es la que posee mayor cantidad de incidentes con armas, sumando 2.104 muertos y 10.427 heridos haciendo un total de 12.531 afectados. Para la búsqueda de la menor ciudad se pudo dar cuenta que existen ciudades donde no hubo incidentes con armas registrados, por lo que poseen 0 heridos, muertos y afectados (un ejemplo es la ciudad de **Ala Moana** en el estado de Hawaii)


¿Qué Estado posee la mayor cantidad de afectados por incidentes con armas? ¿Qué Estado la menor? ¿Cuál es la cantidad de heridos y muertes?

Esta pregunta pudo ser respondida en la sección de conclusiones de este documento, el cubo multidimensional arrojó que el Estado de Illinois fue el que más incidentes poseyó, sumando una cantidad de 16.693 afectados, con 3.409 muertos y 13.514 afectados. Para más detalles sobre la consulta MDX realizada, ver la sección de “Algunas conclusiones...” del presente.

¿Qué sexo (Masculino o Femenino) tiene más carga de afectados en los Estados Unidos?

En order de responder esta pregunta, se realizará la siguiente consulta MDX a nuestro cubo multidimensional:

```
SELECT {[Measures].[cant_afectados]} ON COLUMNS,  
{[DimensionInvolucrado.JerarquiaGenero].[F],  
[DimensionInvolucrado.JerarquiaGenero].[M]} ON ROWS FROM  
[CuboIncidentes]
```

	Measures
JerarquiaGenero	 cant_afectados
F	25.449
M	141.305

Se puede dar cuenta que en los resultados se presenta una tremenda brecha cuantitativa referido al sexo (en el dataset estaba (mal)especificado como género) de los culpables de incidentes con armas en Estados Unidos. Para tener una mejor representación visual de la diferencia, decidimos hacer un gráfico tipo torta.



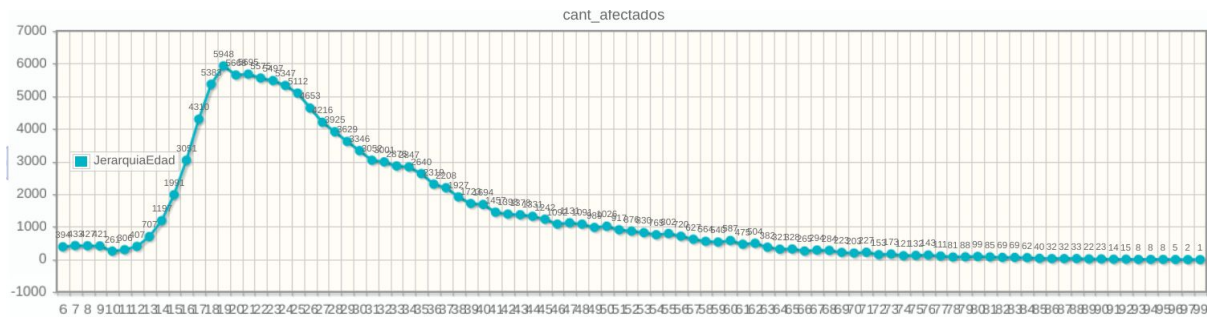
Con estos tipos de gráficos es muy fácil de ver las diferencias (en este caso muy amplias) en los números que los datos nos provee.

¿Qué edad es la predominante dentro de los casos de violencia con armas?

Para conocer este dato, realizaremos la consulta MDX especificada a continuación:

```
SELECT {[Measures].[cant_afectados]} ON COLUMNS,  
Hierarchize(Union({[DimensionInvolucrado.JerarquiaEdad].[All  
DimensionInvolucrado.JerarquiaEdads]},  
[DimensionInvolucrado.JerarquiaEdad].[All  
DimensionInvolucrado.JerarquiaEdads].Children)) ON ROWS FROM  
[CuboIncidentes]
```

Los resultados son impactantes, la edad predominante en incidentes con armas es de **19 años**, pero hay algo particular en lo que se podría llegar a analizar, para esto, veamos un gráfico que representa los resultados.



Se puede notar claramente como la gráfica parece representar una **distribución normal** en cuanto a las edades. Algo muy interesante para analizar. A su vez, también es interesante cómo crece la curva durante la adolescencia, y a partir de allí comienza a decrecer lentamente hasta hacerse cuasi-constante.

¿En qué año fueron más personas afectadas con violencia de armas?

Esta fue la primera pregunta que pudimos responder en el desarrollo del proyecto, es una consulta MDX muy simple dada por:

```
SELECT {[Measures].[cant_afectados]} ON COLUMNS,
Hierarchize(Union({[DimensionTiempo.JerarquiaTiempo].[All
DimensionTiempo.JerarquiaTiempos]}),
[DimensionTiempo.JerarquiaTiempo].[All
DimensionTiempo.JerarquiaTiempos].Children)) ON ROWS FROM
[CuboIncidentes]
```

Los resultados son claros:

	Measures
JerarquiaTiempo	cant_afectados
All DimensionTiempo.JerarquiaTiempos	178.870
2013	1.296
2014	35.559
2015	40.451
2016	45.646
2017	46.214
2018	9.704

Para el análisis no tomaremos en cuenta los años 2013 ni 2018 debido a que los datos no se presentan completamente en el dataset elegido. Más allá de esto, el año en el que más afectados se encontraron en Estados Unidos es en **2017** con una cantidad de 46.214, ¿tal vez se podría relacionar esto con la asunción del presidente Donald J. Trump?

Reportes

Reporte 1

Realizamos este reporte con el fin de representar un detalle por ciudad en un Estado seleccionado por el usuario final, especificando la cantidad de afectados, total por página, total de Estado, e indicando la máxima cantidad entre las ciudades.

El tablero busca responder ¿Qué ciudad posee la mayor cantidad de afectados por incidentes con armas?

Se realizará la siguiente consulta para obtener todos los Estados disponibles en el dataset en orden a tener una dropdown list.

```
SELECT DISTINCT dc.estado  
FROM fact_incidente inner join dim_ciudad dc on fact_incidente.id_ciudad =  
dc.id_ciudad;
```

Y, utilizando el parámetro **`\${estado}`** dado por la dropdown list, determinada por la anterior consulta, se hará una consulta a la base de datos para obtener los datos deseados.

```
SELECT dc.nombre_ciudad, SUM(cant_afectados) as afectados  
FROM fact_incidente inner join dim_ciudad dc on fact_incidente.id_ciudad =  
dc.id_ciudad WHERE dc.estado= `${estado}`  
GROUP BY dc.nombre_ciudad;
```

En el siguiente ejemplo se puede observar la cantidad de afectados en el estado de Connecticut:

Connecticut	
Reporte informativo de cantidad de muertos, heridos y afectados por Ciudad en el Estado de Connecticut	
Nombre de Ciudad	Cantidad de Afectados
Andover	2
Ansonia	11
Baltic (Sprague)	1
Berlin	3
Bethany	1
Bethel	3
Bolton	1
Bozrah	0
Branford	2
Bridgeport	414
Bridgewater	1
Bristol	22
Brooklyn	2
Burlington	2
Canaan (North Canaan)	3
Canton	0
Cheshire	3
Clinton	0

El ejemplo de reporte sobre el Estado de Connecticut se puede encontrar en el siguiente enlace: [Análisis detalle ciudades](#)

Reporte 2

La finalidad de éste reporte viene dada por analizar todos los Estados a la vez, dando como opción para especificar el año de los incidentes.

Se busca responder ¿En qué año fueron más personas afectadas con violencia de armas? y, al mismo tiempo, proveer de un análisis por Estado y dando un dato nuevo calculado con la herramienta de diseño de reportes especificando promedio de afectados por Estado en ese mismo año.


Para el parámetro **$\{anio\}$** utilizaremos una consulta para obtener todos los años disponibles en nuestro dataset.

```
SELECT DISTINCT dt.anio
FROM dim_tiempo dt inner join fact_incidente fi on dt.id_tiempo = fi.id_tiempo
```

Una vez obtenido el dato del año, para obtener el cuerpo del reporte, es decir, la totalidad de los datos que buscamos, realizamos la siguiente consulta.

```
SELECT dc.estado, SUM(cant_afectados) as afectados, SUM(cant_muertos) as
muertos
FROM dim_ciudad dc inner join fact_incidente fi on dc.id_ciudad = fi.id_ciudad
inner join dim_tiempo dt on fi.id_tiempo = dt.id_tiempo
WHERE dt.anio = ${anio}
GROUP BY dc.estado
```

Por ejemplo si seleccionamos el año 2017 podremos observar el siguiente reporte:



Estado	Afectados	Muertos
Alabama	1.400	544
Alaska	139	69
Arizona	587	289
Arkansas	653	216
California	3.376	1.423
Colorado	499	199
Connecticut	382	89
Delaware	297	56
District of Columbia	440	100
Florida	2.751	950
Georgia	1.715	658
Hawaii	28	8
Idaho	81	47
Illinois	4.588	959
Indiana	1.144	407
Iowa	300	71
Kansas	416	152
Kentucky	791	273

Para ver el reporte ejemplo completo, ingrese en el siguiente enlace: [Análisis año 2017](#).

Reporte 3

Para el siguiente reporte, buscamos responder la pregunta ¿Qué sexo (Masculino o Femenino) tiene más carga de afectados en los Estados Unidos? sumado a un promedio de la edad de los involucrados, utilizando la herramienta de **subreportes** del software de diseño de reportes. Cabe destacar que también se utilizó la técnica de **agrupación** para analizar por todos los Estados.

Éste reporte tiene un parámetro elegido por el usuario final que es el año, y se calcula igual que en el Reporte 2.

Para obtener los datos haremos la siguiente consulta a nuestra base de datos:

```
SELECT SUM(cant_afectados) as afectados, di.genero, dc.estado
FROM dim_tiempo dt inner join fact_incidente fd on dt.id_tiempo= fd.id_tiempo
inner join dim_involucrado di on fd.id_involucrado = di.id_involucrado
inner join dim_ciudad dc on fd.id_ciudad = dc.id_ciudad
WHERE (di.genero='M' or di.genero='F') and dt.anio=${anio}
GROUP BY dc.estado, di.genero;
```

Por otro lado, en el subreporte solamente se calcula el promedio de edad, tomando los parámetros que vienen del reporte padre. Se hace la siguiente consulta:

```
SELECT SUM(cant_afectados) as afectados, avg(di.edad), dc.estado
FROM dim_tiempo dt inner join fact_incidente fd on dt.id_tiempo= fd.id_tiempo
inner join dim_involucrado di on fd.id_involucrado = di.id_involucrado
inner join dim_ciudad dc on fd.id_ciudad = dc.id_ciudad
WHERE dt.anio=${anio} and edad is not null and dc.estado=${estado}
GROUP BY dc.estado;
```

Un ejemplo del reporte en el año 2016 es el siguiente:

Análisis de Sexos en Incidentes con Armas

Año: 2016

Alabama

		Promedio de Edad:	31,4349
F	173		
M	957		
Total: 1.130			

Alaska

		Promedio de Edad:	32,7551
--	--	-------------------	---------

Para tener una mejor visualización de este reporte recomendamos consultar este enlace: [Análisis sexos 2016](#)

Reporte 4

Para la cuarta entrega de reportes utilizamos la técnica de agrupamientos. Buscamos analizar todos los Estados, en todos los años, obteniendo también un total por año y un total histórico. En este reporte no se utilizan parámetros.

Obtenemos los datos a partir de:

```
SELECT dc.estado, SUM(cant_afectados) as afectados, dt.anio
FROM dim_tiempo dt inner join fact_incidente fi on dt.id_tiempo = fi.id_tiempo inner
join dim_ciudad dc on fi.id_ciudad = dc.id_ciudad
GROUP BY dt.anio, dc.estado
```

Indicamos también al software que las agrupaciones serán por año.

El reporte que se realizó es el siguiente:

Análisis de Estado por Año	
El presente representa un análisis por Estado sobre la cantidad de heridos provocados por incidentes con armas en los Estados Unidos	
2.013	
Alabama	12
Arizona	24
California	185
Colorado	13
Connecticut	15
Delaware	13
District of Columbia	44
Florida	58
Georgia	31

Se puede visualizar de forma completa en el siguiente enlace: [Análisis total histórico](#)

Reporte 5

Por último, tenemos el reporte que genera gráficos. Es un simple reporte el cual busca responder y visualizar la pregunta: ¿Qué ciudad posee la mayor cantidad de afectados por incidentes con armas? Aunque, se hace un análisis en el “TOP” 10 ciudades en cantidad de afectados histórico en un Estado determinado por un parámetro ***{estado}*** elegido por el usuario final.

La lógica de el parámetro es la misma que en Reporte 1, por lo que no se especificará para no redundar información.

Para obtener los datos usamos “LIMIT 10” combinado con un “ORDER BY DESC” para obtener el Top 10 ciudades. Entonces, la consulta SQL utilizada está dada por:

```
SELECT dc.nombre_ciudad, SUM(cant_muertos) as muertos, SUM(cant_heridos) as  
heridos  
FROM dim_ciudad dc inner join fact_incidente fi on dc.id_ciudad = fi.id_ciudad  
WHERE dc.estado={estado}  
GROUP BY dc.nombre_ciudad  
ORDER BY (sum(cant_muertos)+sum(cant_heridos)) DESC  
LIMIT 10;
```

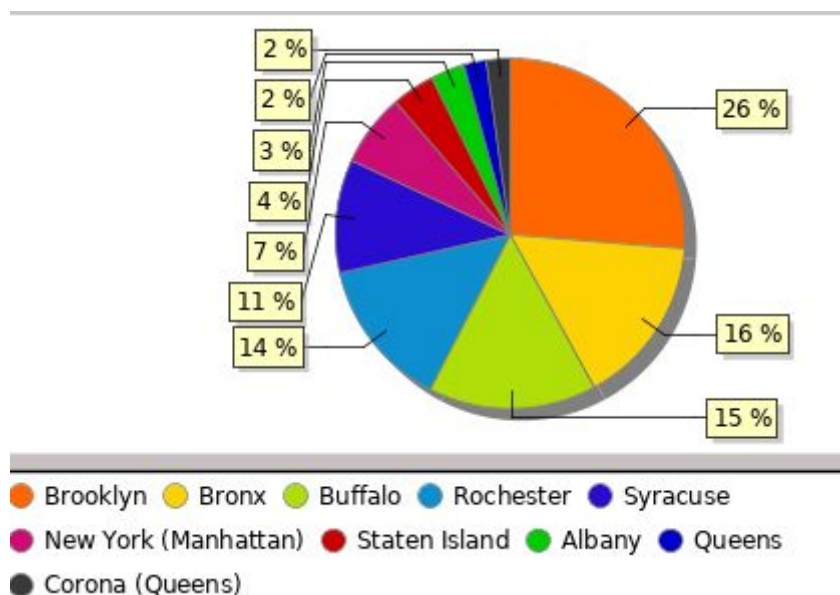
Ahora, como grupo, tomamos la decisión de hacer un display de los datos de forma plana (texto) y de una forma gráfica para tener una visualización, y por tanto, un mejor entendimiento de ¿qué me quieren decir estos datos?

Para ello usamos un gráfico de tipo torta donde se está configurado para analizar por cantidad de heridos en cada ciudad del “top” 10 del Estado elegido por parámetro.

Un ejemplo seleccionando el estado de Florida es el siguiente:

Florida		
Análisis gráfico de incidentes en las 10 ciudades del Estado con más impacto		
Ciudad	Muertos	Heridos
Jacksonville	421	1.126
Orlando	303	633
Miami	325	587
Tampa	151	302
West Palm Beach	125	259
Fort Myers	87	187
Tallahassee	67	126
Ocala	60	126
Miami Gardens	45	115
Bradenton	55	88

Respecto a estos datos, el gráfico en forma de torta que se generó es el siguiente:



Para tener una visualización completa de este tipo de reporte recomendamos verlo en el siguiente enlace: [Análisis TOP con gráfico](#)

Tableros

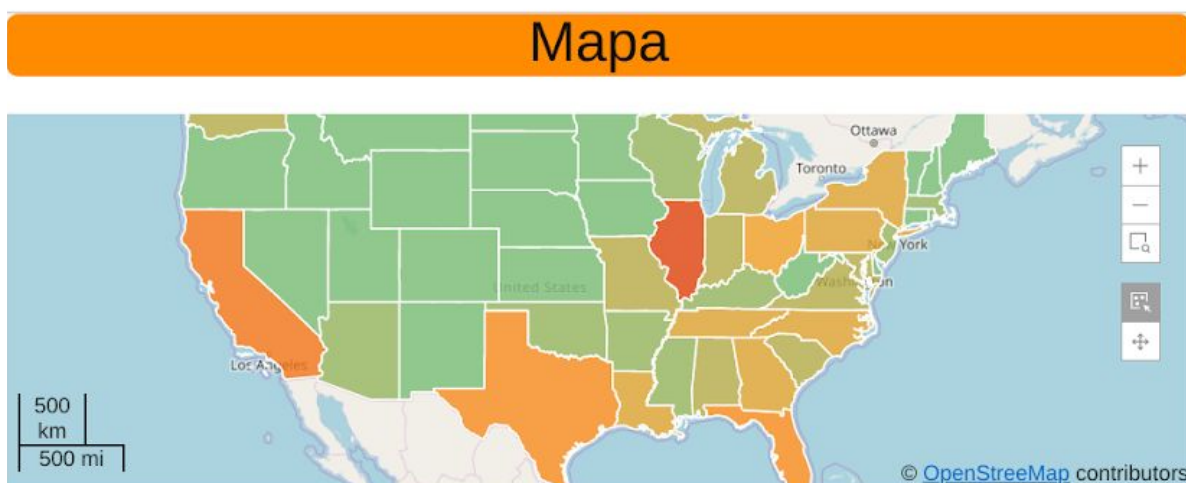
Tablero 1

El siguiente tablero que realizamos contiene el siguiente mapa, en el cual cada gama de color desde verde hacia lo más naranja representa desde menos afectados a más afectados respectivamente.

Esta tablero busca responder algunas de las preguntas realizadas al principio en la metodología hefesto pero algunas respecto a los distintos estados:

- Cuales son los estados con más afectados de forma visual.
- Cantidad de afectados por año de cada estado.
- División de la cantidad de muertes y heridos por estado.
- Cantidad de involucrados respecto el sexo por estado.

Todas estas se van a respuestas se van a aclarar a continuación.



Como se puede observar la división está hecha en los distintos estados. Esta división entre los distintos estados se realizó por las coordenadas de los mismos que están cargadas en un archivo JSON. Este JSON se tiene que cargar en el dashboard y la función antes de ejecutar el mapa es la siguiente:

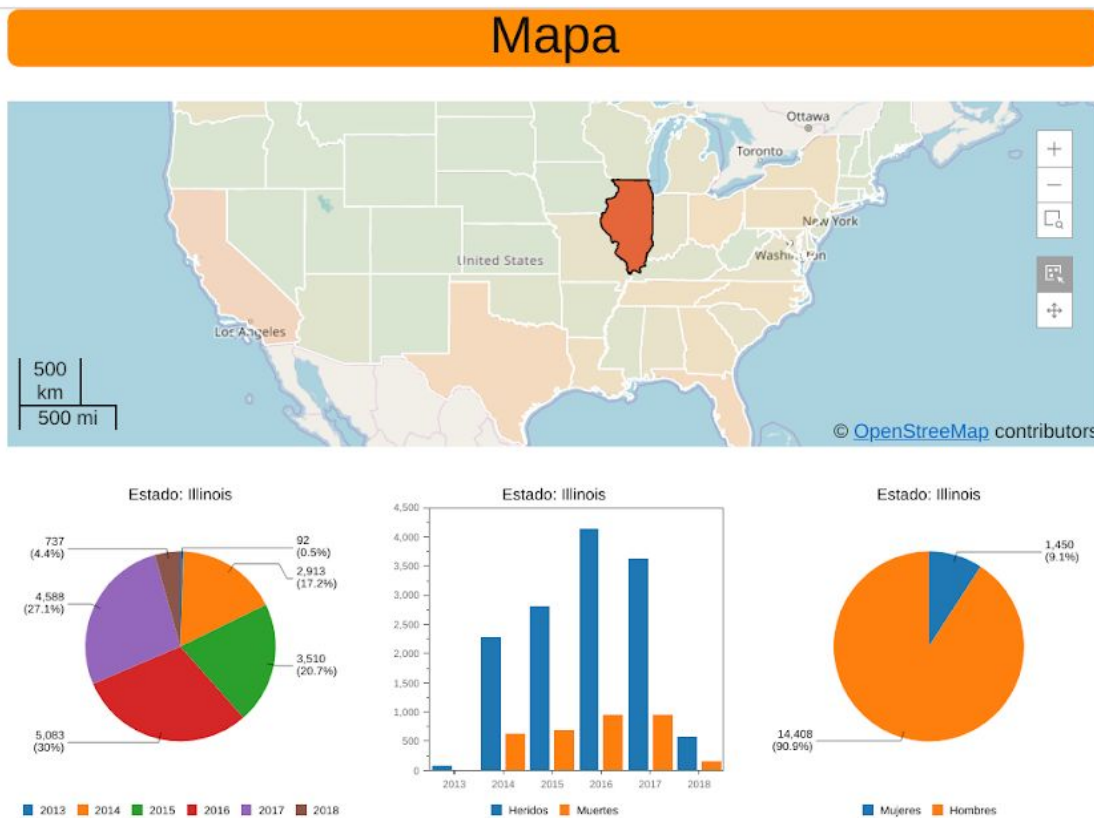
```
function f(scene){  
  var getResource = "";  
  this.shapeResolver = 'geoJSON';
```

```
this.setAddInOptions('ShapeResolver','geoJSON', {  
    url: getResource + '${solution:States.js}',  
    idPropertyName: 'NAME'  
});  
}
```

Respecto la gama de colores fue cargada a mano respecto la cantidad de afectados. Lo que se realizó fue agrupar en una tabla la cantidad total de afectados por estado. Se seleccionó al estado con más afectados y todos los totales de afectados se dividieron por esta cantidad. Para obtener mayores números, todos los resultados finales fueron multiplicado por 100 y como resultado quedó:

```
["Alabama", 28], ["Alaska", 3], ["Arizona", 12], ["Arkansas", 12],  
["California", 78], ["Colorado", 11], ["Connecticut", 9], ["Delaware", 6],  
["Florida", 64], ["Georgia", 38], ["Hawaii", 0], ["Idaho", 2], ["Illinois", 100],  
["Indiana", 26], ["Iowa", 6], ["Kansas", 8], ["Kentucky", 17], ["Louisiana", 38],  
["Maine", 1], ["Maryland", 28], ["Massachusetts", 12], ["Michigan", 27],  
["Minnesota", 8], ["Mississippi", 18], ["Missouri", 33], ["Montana", 1],  
["Nebraska", 5], ["Nevada", 10], ["New Hampshire", 1], ["New Jersey", 21],  
["New Mexico", 6], ["New York", 39], ["North Carolina", 40], ["North Dakota", 1],  
["Ohio", 48], ["Oklahoma", 14], ["Oregon", 7], ["Pennsylvania", 44], ["Rhode  
Island", 2], ["South Carolina", 27], ["South Dakota", 1], ["Tennessee", 37],  
["Texas", 65], ["Utah", 4], ["Vermont", 0], ["Virginia", 29], ["Washington", 29],  
["West Virginia", 5], ["Wisconsin", 17], ["Wyoming", 0],
```

Cuando seleccionamos un estado, en este caso Illinois, el tablero va a actualizar los siguientes gráficos. Además, se utiliza el parámetro Estado para el estado seleccionado. El parámetro se utiliza en las consultas de cada gráfico.



En primer lugar un gráfico en forma de torta en el que el total de afectados del estado se muestra separado por los distintos años en el que se obtuvieron estos acontecimientos. La consulta utilizada es la siguiente:

```
select anio, SUM(afectados) as afectados
from dimCiudad as c
inner join factIncidente as i on i.idCiudad=c.idCiudad
inner join dimTiempo as t on i.idTiempo = t.idTiempo
where c.estado = ${Estados}
group by anio;
```

En segundo lugar un gráfico de barras en el cual se muestra de todos los afectados del estado, cuáles fueron heridos y cuáles fueron muertes, también respecto a los años. La consulta utilizada es la siguiente:

```
select anio, SUM(heridos) as Heridos, SUM(muertes) as Muertes
from dimTiempo t inner join factIncidente fi on t.idTiempo = fi.idTiempo
inner join dimCiudad dC on fi.idCiudad = dC.idCiudad
where dC.estado = ${Estados}
```

group by anio

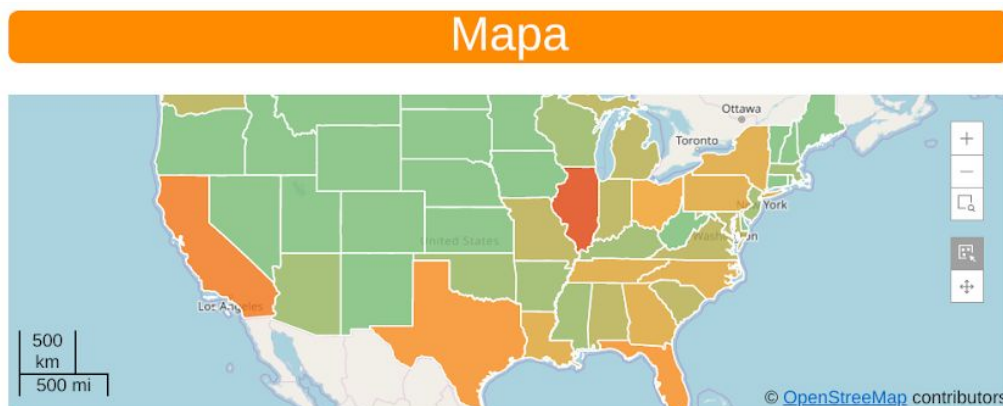
Por último tenemos un gráfico en forma de torta que muestra la división respecto el sexo de todos los afectados en el estado seleccionado. La consulta utilizada es la siguiente:

```
select 'Mujeres',COUNT(Sexo)
from dimInvolucrado di
inner join factIncidente fl on di.idInvolucrado = fl.idInvolucrado
inner join dimCiudad dC on fl.idCiudad = dC.idCiudad
where Sexo = 'F'
and estado = ${Estados}
union
select 'Hombres',COUNT(Sexo)
from dimInvolucrado di
inner join factIncidente fl on di.idInvolucrado = fl.idInvolucrado
inner join dimCiudad dC on fl.idCiudad = dC.idCiudad
where Sexo = 'M'
and estado = ${Estados}
```

Tablero 2

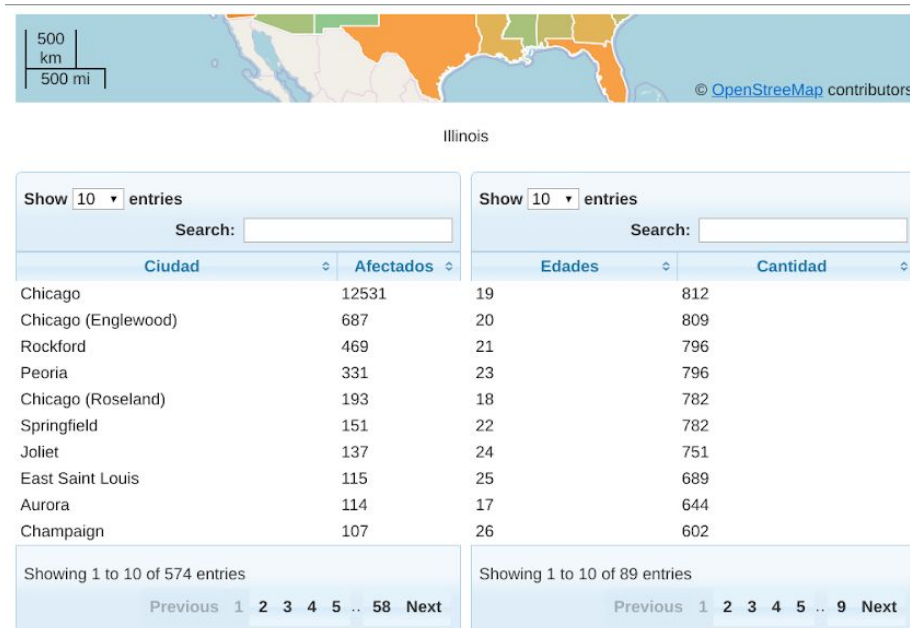
Este segundo tablero también busca responder algunas preguntas respecto únicamente de los estados:

- Ciudades con más afectados por estado.
- Edades predominantes de involucrados por estado.



Como podemos observar, este tablero contiene el mismo mapa que el anterior, por eso mismo, no vamos a volver a explicar cómo se realizó.

Al seleccionar un estado dentro del mapa las siguientes tablas se grafican.



La primer tabla contiene el nombre de las ciudades del estado seleccionado con su respectiva cantidad de afectados, acomodo de de manera decreciente. La consulta utilizada es la siguiente:

```
select ciudad as Ciudad, SUM(afectados) as Afectados
from dimCiudad inner join factIncidente fI on dimCiudad.idCiudad =
fI.idCiudad
where estado = ${Estados}
group by ciudad
order by SUM(afectados) desc
```

La segunda tabla contiene la edad de los distintos involucrados del estado seleccionado teniendo en cuenta la cantidad de involucrados por edad. Está ordenada de forma decreciente respecto la cantidad de involucrados por edad. La consulta realizada es la siguiente:

```
select Edad as Edades, COUNT(Edad) as Cantidad
```

from dimInvolucrado

inner join factIncidente fI on dimInvolucrado.idInvolucrado=fI.idInvolucrado

inner join dimCiudad dC on fI.idCiudad = dC.idCiudad

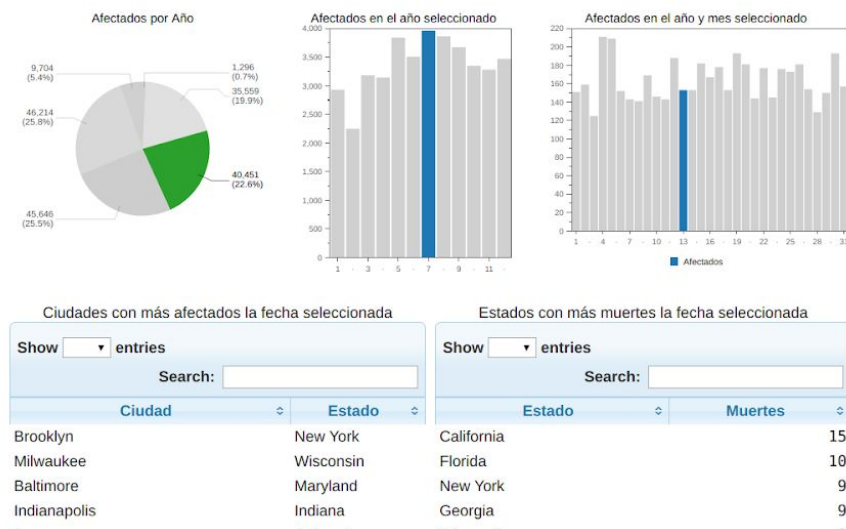
where estado = \${Estados}

group by Edad

order by COUNT(Edad) DESC

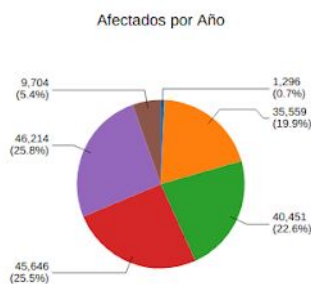
Tablero 3

Datos Respecto Fechas



El tablero ya presentado tiene un gráfico en forma de torta el cual contiene todos los años en el cual se realizó captura de datos.

Datos Respecto Fechas



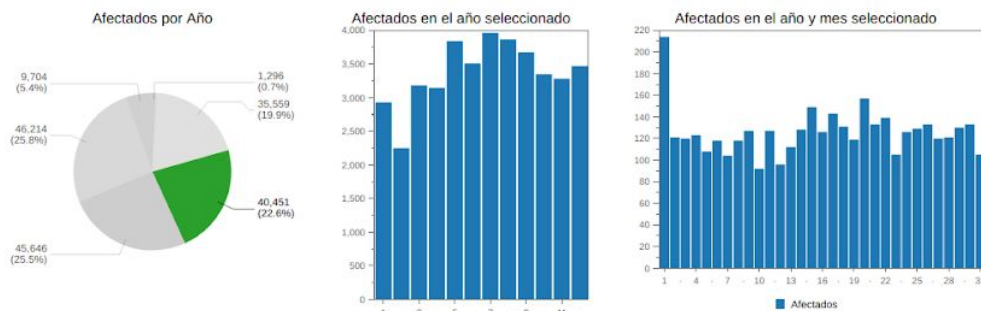
Este gráfico utilizó la siguiente consulta:

select anio, SUM(afectados) as Afectados

```
from factIncidente as fl
inner join dimTiempo as t on fl.idTiempo = t.idTiempo
group by anio;
```

Al seleccionar un año, se puede visualizar otro gráfico el cual contiene todos los meses y cuáles fueron los afectados que ocurrieron en el mismo. Además se utilizó el mes Enero como predeterminado por lo tanto se puede observar la cantidad de afectados por día en determinado año y mes.

Datos Respecto Fechas

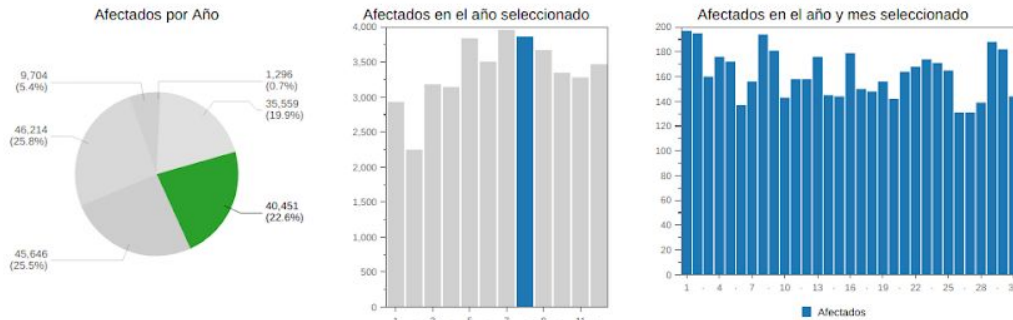


La consulta que se utilizó para este gráfico en forma de barra en el que se observan los meses es el siguiente:

```
select mes, SUM(afectados) as 'Afectados'
from factIncidente as fl
inner join dimTiempo dT on fl.idTiempo = dT.idTiempo
where anio = ${anio}
group by mes
```

Luego es posible seleccionar un mes en particular, cambiando así aquel que estaba como predeterminado.

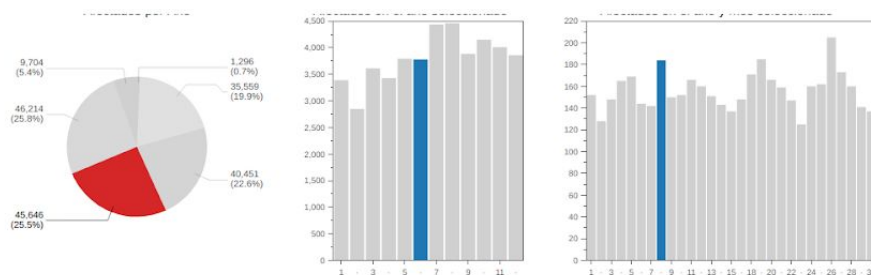
Datos Respecto Fechas



De esta forma podemos observar la diferencia por día respecto un año y un mes en particular. La consulta utilizada es la siguiente:

```
select dia, COUNT(dia) as Afectados
from factIncidente as fI
inner join dimTiempo dT on fI.idTiempo = dT.idTiempo
where mes = ${meses}
and anio = ${anio}
group by dia
```

Luego de esto es posible seleccionar un día en particular, de esta forma teniendo ya una fecha (día, mes y año en particular), para generar dos tablas. La primera indica la ciudad en la que hubo más afectados en esa fecha y su respectivo estado. La segunda indica cual es el estado en el que hubieron más muertes.



Ciudades con más afectados la fecha seleccionada

Show entries

Search:

Ciudad	Estado
Baltimore	Maryland
Chicago	Illinois
Charlotte	North Carolina
Washington	District of Columbia
Memphis	Tennessee

Showing 1 to 5 of 145 entries Previous Next

Estados con más muertes la fecha seleccionada

Show entries

Search:

Estado	Muertes
Florida	16
Illinois	15
North Carolina	14
South Carolina	11
Maryland	9

Showing 1 to 5 of 43 entries Previous Next

Las consulta utilizada para generar la primer tabla es la siguiente:

```
select ciudad as Ciudad, estado as Estado  
from dimCiudad  
inner join factIncidente fI on dimCiudad.idCiudad = fI.idCiudad  
inner join dimTiempo dT on fI.idTiempo = dT.idTiempo  
where mes = ${meses}  
and anio = ${anio}  
and dia = ${dia}  
group by ciudad, estado  
having count(muertes)  
order by count(muertes) desc
```

La consulta para generar la segunda tabla es la siguiente:

```
select estado, sum(muertes)  
from dimCiudad  
inner join factIncidente fI on dimCiudad.idCiudad = fI.idCiudad  
inner join dimTiempo dT on fI.idTiempo = dT.idTiempo  
where anio = ${anio}  
and mes = ${meses}  
and dia = ${dia}  
group by estado  
order by sum(muertes)desc
```

Conclusión

Hay que tener en cuenta que muchas de las conclusiones ya fueron planteadas en el análisis de los componentes gráficos de Inteligencia de Negocios utilizados.

Una pregunta muy interesante a analizar es la siguiente:

“¿Qué Estado posee la mayor cantidad de afectados por incidentes con armas? ¿Qué Estado la menor? ¿Cuál es la cantidad de heridos y muertes?”

Esta pregunta se puede resolver a través de la siguiente consulta:

```
SELECT {[Measures].[cant_afectados], [Measures].[cant_muertos],  
[Measures].[cant_heridos]} ON COLUMNS,  
Order(Union({[DimensionCiudad.JerarquiaCiudad].[All  
DimensionCiudad.JerarquiaCiudades]}), [DimensionCiudad.JerarquiaCiudad].[All  
DimensionCiudad.JerarquiaCiudades].Children), [Measures].[cant_afectados],  
DESC) ON ROWS FROM [CuboIncidentes]
```

Utilizamos un reporte básico para exportar los resultados de la consulta. Disponible en formato pdf: [Reporte Incidente-Estado](#)

Analizando la primera fila de Estado se puede dar cuenta que el Estado de **Illinois** es el que posee mayor cantidad de afectados en incidentes con armas, con una monstruosa cantidad de 16.693 personas.

Es interesante notar también que, por más que sea el que tiene menos afectados, no es el que tiene mayor cantidad de Muertes ni mayor cantidad de Afectados. Por lo tanto podríamos llegar a pensar que en Illinois hay inconvenientes pero ni los ciudadanos ni la policía responde con gatillo fácil como pasa en los estados de California, Texas y Florida en los cuales hay más muertes que en el estado de Illinois.