

Proyecto Final de BI - Informe

Diego Hiriart, Luis Corales

Facultad de Ingeniería y Ciencias Aplicadas, Universidad de las Américas
Quito, Ecuador

diego.hiriart@udla.edu.ec

luis.corales.martinez@udla.edu.ec

Abstract—El presente informe contiene el proceso para obtener y analizar datos mediante una herramienta OLAP. Los datos en cuestión corresponden a los crímenes de homicidios, asesinatos, femicidios, y sicarios en el Ecuador desde el 2016. Mediante la elaboración del análisis, se identificaron tendencias de aumento para el crimen en el país, sobre todo para asesinatos. Además, se identifican ciertos picos en las incidencias de femicidios. Los resultados obtenidos demuestran alzas en el crimen en Ecuador, lo cual indica que se requiere mayor análisis a nivel social para encontrar explicaciones a esta tendencia.

I. INTRODUCCIÓN

Recientemente, el Ecuador se encuentra atravesando una oleada de crimen muy importante. A simple vista, a partir de noticias y conmoción social, los delitos que más parecen presentar crecimiento en sus incidencias son los homicidios. Al mismo tiempo, durante la pandemia el número de femicidios vio un aumento, es importante saber si esta tendencia se ha mantenido, aumentado, o disminuido.

Considerando esta premisa, es de gran importancia analizar los datos existentes sobre estos delitos, y determinar qué relaciones se pueden encontrar con las últimas épocas de crisis que se han vivido en el Ecuador debido a la pandemia. Con el análisis de estos datos, se podría confirmar que exista una relación entre la crisis de pandemia y el aumento de ciertos delitos.

II. PREPARACIÓN DEL PROYECTO

A. Alcance

La delimitación de análisis a delitos de homicidios (sin una circunstancia que lo excuse, quitar la vida a una persona sin premeditación), asesinatos (de manera premeditada, matar a una persona buscando los medios necesarios), sicarios, y femicidios es necesaria para un enfoque más centrado, con menos ruido de variables relacionadas a más categorías. Respecto al marco de tiempo a considerar, se utilizarán los datos desde enero de 2019 hasta la fecha más reciente disponible al tiempo de elaboración (marzo de 2022). Esto permite cubrir las estadísticas antes durante, y después de la pandemia para los delitos mencionados.

A partir del análisis de estos datos se puede encontrar si existe una correlación entre el confinamiento por pandemia y el incremento o decremento de homicidios, asesinatos, sicarios, y femicidios; sin embargo, no se puede determinar si es una relación casual o causal. Se podría identificar si una relación existe, pero un análisis más detallado de las instancias de los crímenes y del contexto social de los mismos sería necesario para determinar con certeza si la relación es causal.

B. Fuentes de datos

Para la obtención de datos, se hizo una búsqueda exhaustiva de fuentes que proporcionaran los datos requeridos, donde se encontró que el Ministerio de Gobierno [1] recopila anualmente los registros de reportes de crímenes de alta incidencia, tales como homicidio, femicidio, asesinato y sicariato. Estos datos están disponibles al público almacenados como un Excel con información desde el 2014 hasta marzo del 2022, detallando los siguientes datos por crimen: tipo de crimen, provincia, cantón, mes, año, tipo de arma, rango de edad de la víctima, sexo del perpetuador y número de homicidios en el reporte.

Adicionalmente, se obtuvieron datos de las localidades (provincia y cantón) en las que se reportan cada crimen. Específicamente, se asociaron a cada cantón los datos de población [2] (proyectadas para el 2020 del último censo) y la superficie de cada cantón [3] (para obtener densidad poblacional), ambos datos obtenidos del INEC. Usando estos datos, se podrían ver más relaciones entre la fluctuación de incidencias de crímenes y regiones del Ecuador, lo que ayudaría a análisis más completo.

C. Limitaciones

Aunque los datos prepandemia se extienden hasta el 2014, los datos post pandemia solo llegan (actualmente) a marzo de 2022. Esta limitación se debe al corto tiempo que ha pasado desde el inicio del regreso a la normalidad. Se puede realizar un análisis de la incidencia de crímenes para varios años antes de la pandemia, pero esto no es posible para el tiempo después de la misma. Debido a esto, no se podrá ver claramente cómo se comportan las tendencias de cada crimen después de la pandemia de igual manera que antes de esta.

III. DESARROLLO

Como primer paso, se necesita agrupar todos los datos obtenidos en una base de datos. Para ello, se usó SQL Server con el fin de almacenar los datos de las fuentes en tablas organizadas. Este proceso de poblar una base de datos con la información de las fuentes fue necesario para tener una mejor manera de luego elaborar el proceso de ETL. A partir de los datos en hojas de cálculo de Excel obtenidas de entidades públicas, se llena esta base de datos para filtrar y organizar la información de las fuentes.



Ilustración 1 Diagrama de la base de datos fuente

Una vez obtenidos y organizados los dato fuente, se debe generar el Data Mart. Para esto, se llevó a cabo un proceso de ETL (Extract Transform and Load; extraer, transformar y cargar), que consiste en extraer datos de las fuentes con el fin de agruparlos y cargarlos en una base de datos relacional. La extracción de datos para el proceso se vale de la base de datos fuente, su transformación consiste en obtener datos sin repeticiones innecesarias y ajustarlos para obtener dimensiones y la tabla de hechos para el Data Mart.



Ilustración 2 Diagrama del Data Mart

Para poder analizar la información del Data Mart, se requiere utilizar OLAP. Mediante el uso de Analysis Services de SQL

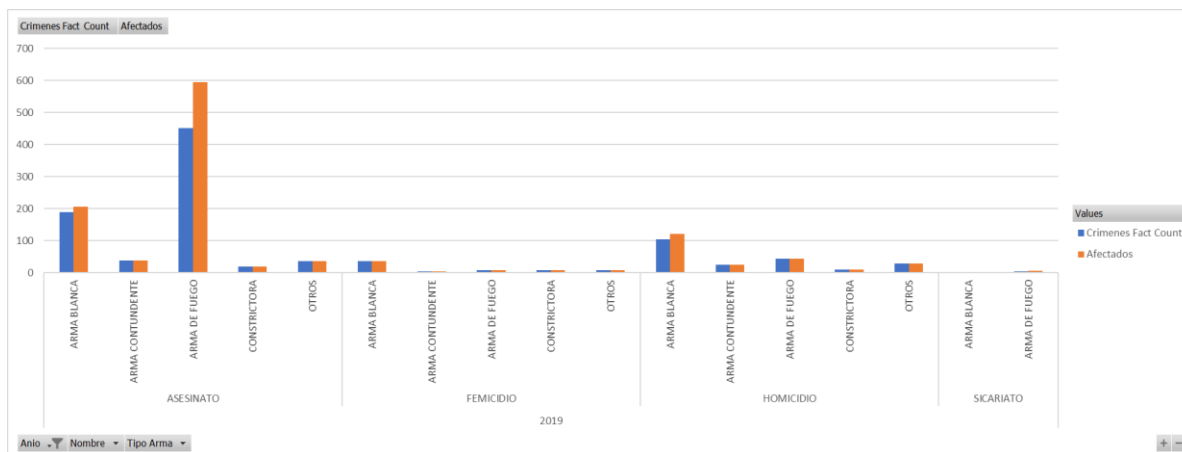
Server, se puede utilizar el Data Mart para generar un cubo OLAP. Este cubo es multidimensional, y a partir de este se pueden crear reporterías para visualizar los datos y usarlos como conocimiento.

Con la obtención del cubo multidimensional, se podrán generar reportes en los cuales sea posible observar más a detalle las estadísticas de los crímenes. Para un análisis adecuado de tendencias y picos o bajas en las ocurrencias de los delitos, se visualizará el rango del 2016 a 2022. En caso de estar visualizándose la cantidad anual de algún delito, se tomar en cuenta que el año 2022 aún no finaliza, y los datos estarán incompletos para el mismo.

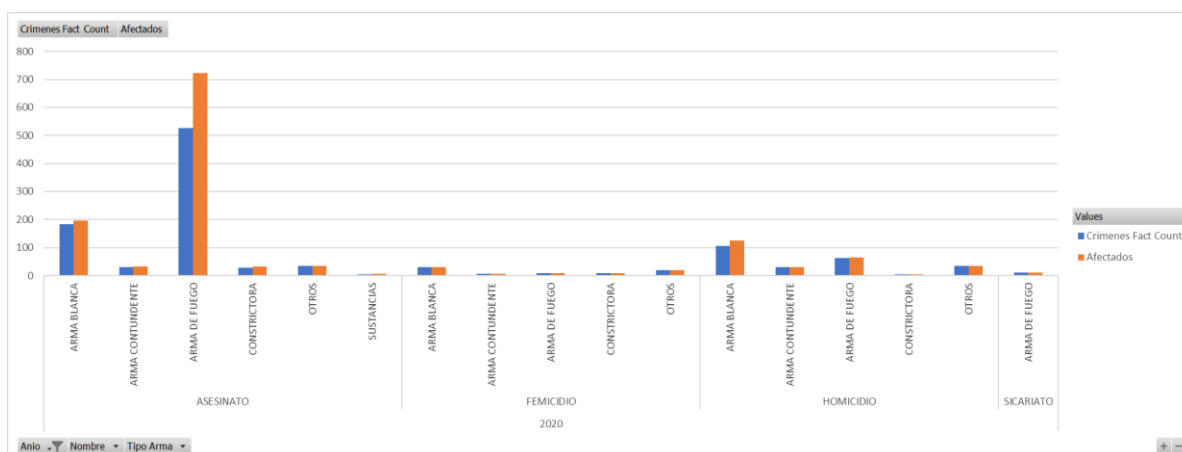
IV. RESULTADOS

Se analizaron los datos obtenidos del cubo OLAP realizando un filtrado en tablas dinámicas de Excel. Usando este proceso, se pudieron obtener estadísticas específicas de cada tipo de crimen y sus incidencias en los años registrados. Es de especial utilidad observar las gráficas de totales de crímenes y afectados en cada año, así como los cuadros de tendencias para cada delito mes por mes.

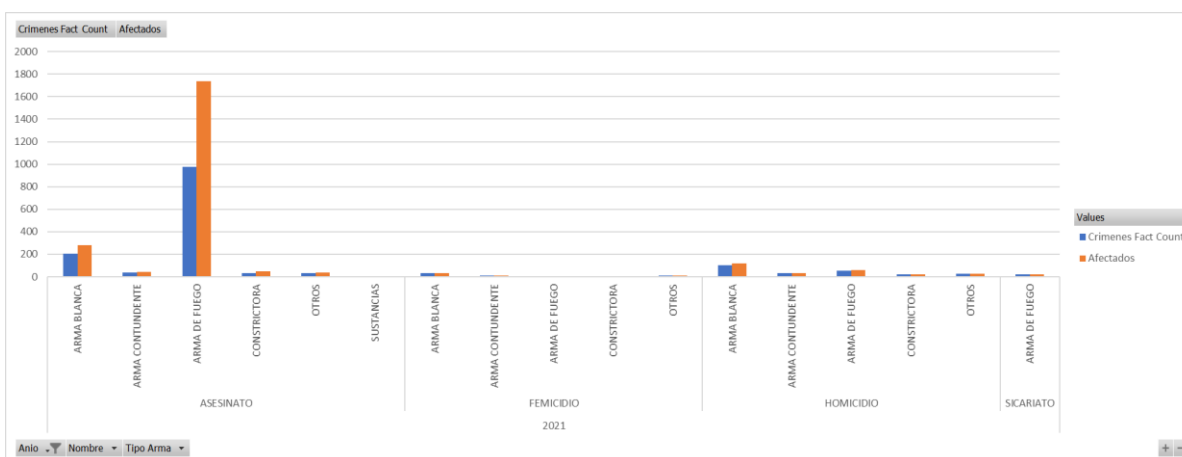
Primero, se deberían observar los totales para todos los crímenes año por año. Las siguientes gráficas muestran la cantidad de ocurrencias y los afectados (víctimas) por cada crimen desde el 2019 hasta el 2021.



Gráfica 1 Cantidad de delitos y afectados en 2019, con detalle de arma utilizada



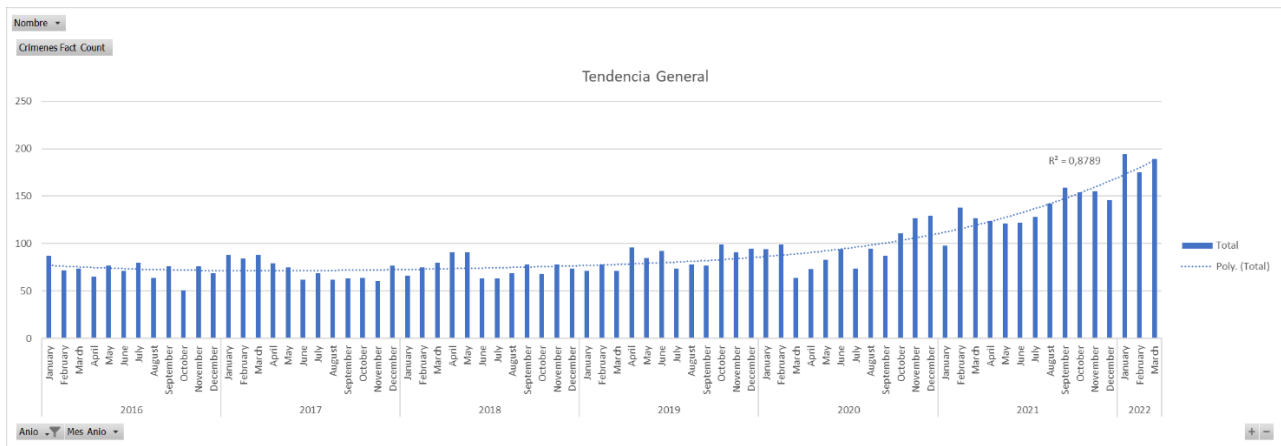
Gráfica 2 Delitos y afectados en el 2020



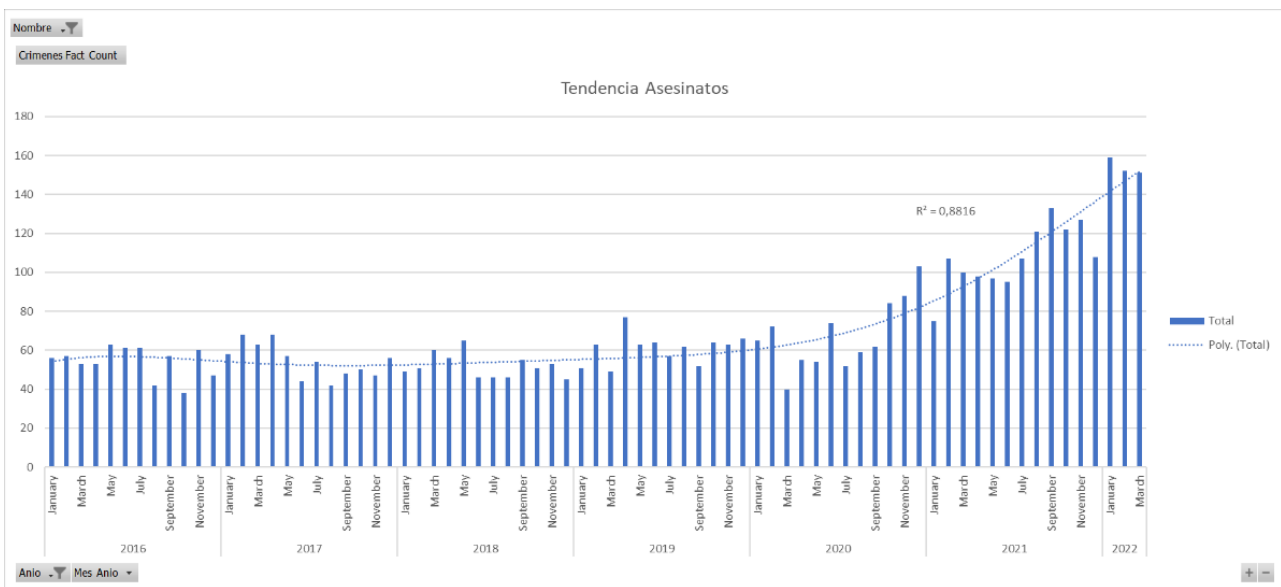
Gráfica 3 Incidencias de crímenes y cantidad de afectados en el 2021

También se obtuvieron gráficas con las tendencias de las cantidades de crímenes mes por mes desde el 2016 hasta marzo de 2022. La tendencia general para el crimen es de aumento, evidenciado por la línea de tendencia polinomial de grado 6. Usando el mismo tipo de línea de tendencia, se observa que los

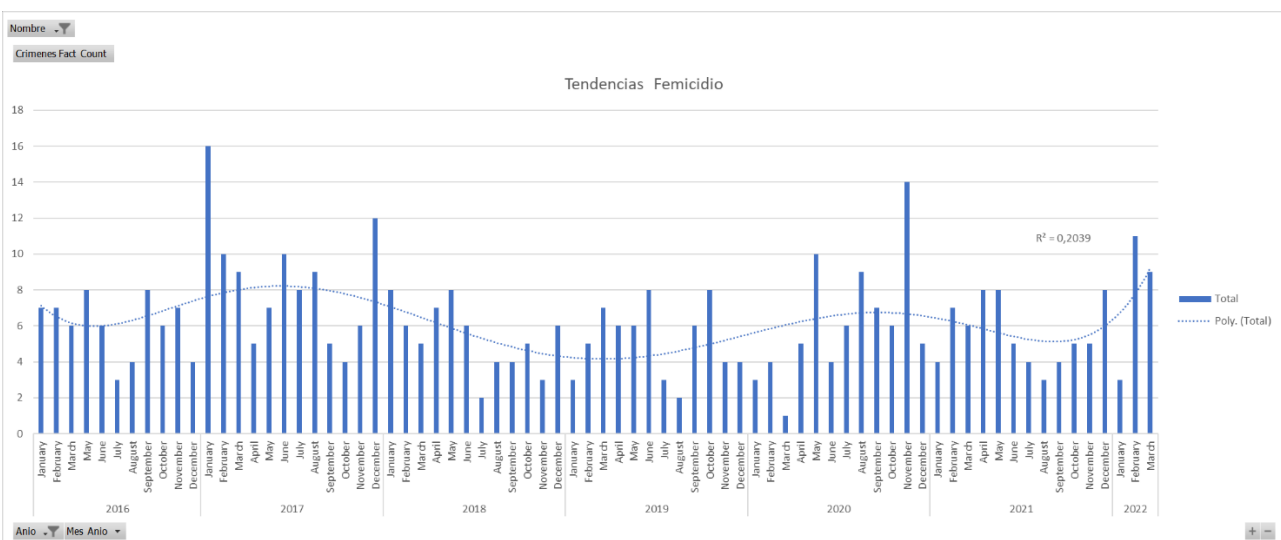
crímenes de homicidio, asesinato y sicariato no tienen una tendencia observable (a juzgar por el valor R^2). Sin embargo, los asesinatos tienden a aumentar y se observan picos de femicidios en el 2017 y 2020. Adicional, se ve que el número de homicidios se ha mantenido elevado desde 2017.



Gráfica 4 Tendencia mensual de crímenes desde 2016 a marzo de 2022



Gráfica 5 Tendencias de asesinatos mensuales

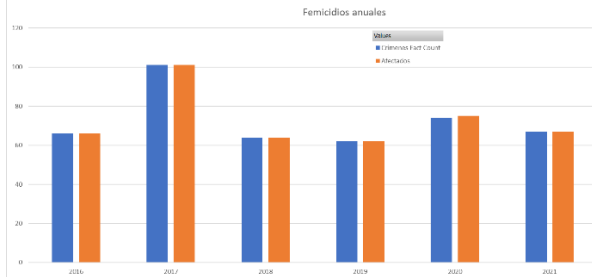


Gráfica 6 Tendencias de femicidios mensuale

Para observar con más detalle los picos en incidencias de femicidios es necesario verlos de manera anual desde 2019 al 2021. Filtrando solo este crimen y observando los valores totales de incidencias y afectados, se pueden ver dos particularidades.

Primero, en el 2017 existe un pico de más de 100 femicidios y personas afectadas; y segundo que, en el 2020 (año inicial de la pandemia), también existió un alza de ocurrencias con 74 femicidios y 75 personas afectadas. En comparación, en los años

2019 y 2021 la cantidad de ocurrencias fueron 62 y 67, respectivamente (las personas afectadas son la misma cantidad).



Gráfica 7 Femicidios anuales y personas afectadas

V. DISCUSIÓN

Ese tipo de análisis cuantitativo de fenómenos sociales es útil para identificar tendencias en los mismos, y a partir de esto profundizar el conocimiento que se tenga e identificar que acciones tomar para su mejora. Al tratarse de crímenes de alto impacto social en el presente caso, su análisis es de especial importancia para determinar situaciones que afectan a la de seguridad pública del país.

Con el análisis de Business Intelligence realizado en este proyecto, fue posible determinar rápidamente características de las ocurrencias de crímenes en el Ecuador. Poder identificar tendencias y picos de crímenes en un proyecto pequeño en relación con lo que podría hacer una entidad estatal, significa que se podrían obtener resultados de seguridad pública relevantes para afrontar el crimen. Herramientas de BI como la del presente proyecto o la que dispone en línea el Ministerio de Gobierno [1], permiten a la población tener una idea de la situación social del país sin sumergirse en informes complejos o ser parte de los organismos que los manejan.

Sería de utilidad combinar la analítica de datos para obtener información detallada de la actividad criminal en el país junto con medidas de mejora social y cumplimiento de la ley. La razón de esto es que su combinación permitiría atender el problema de la oleada reciente de crimen desde más enfoques. No solo se haría mediante la respuesta de servicios de seguridad pública, sino también mediante la intervención adecuada a nivel social para identificar y atender causas raíz de problemas que afectan a toda la sociedad. Adicionalmente, se podrían identificar épocas y regiones del país en las que las ocurrencias de crímenes son altas y comprarlas con situaciones específicas que se estén desarrollando para identificar más causas de crímenes específicos.

VI. CONCLUSIONES

- Es clara la existencia de una tendencia de aumento en la cantidad de crímenes cometidos en Ecuador. Aunque este no es el caso para cada tipo de crimen de manera individual, de manera generalizada la tendencia de aumento está soportada por un valor R^2 de 0.87, esto indica que lo más seguro es que las incidencias de crímenes sigan en aumento.
- Aunque los datos de femicidios del 2020 tienen un pico significativo en comparación al 2019, no puede determinarse que el aumento este directamente relacionado con la pandemia. Puede inferirse que el confinamiento o conmoción social de alguna manera afectaron las

incidencias de femicidios, pero se necesita un análisis cualitativo más profundo para determinarlo.

- Tanto el ETL como el cubo OLAP permitieron realizar el filtrado y minado de datos para analizar la información requerida. El análisis mediante Business Intelligence de datos relacionados al ámbito social permiten obtener información importante sobre fenómenos en la sociedad.

VI. RECOMENDACIONES

- Es necesario realizar una nueva investigación en próximos años donde se puedan recolectar más datos en los años que representan la nueva normalidad (post pandemia). Esto permitiría encontrar tendencias que grafiquen acertadamente como han variado los crímenes estipulados durante los años.
- Investigar a profundidad los hechos (aparte de la pandemia) que hayan ocurrido en cada mes que presente anomalías (números muy altos o bajos dentro de una categoría con respecto a otros meses). Esta investigación podría aportar más factores al análisis para determinar si realmente la pandemia fue el principal hecho para que los crímenes se vieran en aumento.
- Configurar adecuadamente el ordenado de los datos generados en el OLAP. Esto es importante para evitar inconvenientes en la representación de los datos en una gráfica. Se debe realizar la configuración para que los datos estén ordenados alfabéticamente, numéricamente, o cumpliendo algún otro parámetro.

REFERENCIAS

- [1] Ministerio de Gobierno, «Homicidios Intencionales – Porcentaje de Variación Interanual.» Indicadores de Seguridad Ciudadana, 2022. [En línea]. Disponible en: <http://cifras.ministeriodegobierno.gob.ec/comisioncifras/inicio.php>.
- [2] INEC, «Proyección de la Población Ecuatoriana, por años calendario, según cantones 2010-2020.» INEC, 2020. [En línea]. Disponible en: <https://www.ecuadorencifras.gob.ec/proyecciones-poblacionales/>.
- [3] INEC, «Población, superficie (km2), densidad poblacional a nivel parroquial.» INEC, 2022. [En línea]. Disponible en: <https://www.ecuadorencifras.gob.ec/informacion-censal-cantonal/>.

ANEXOS

- 1) Scripts SQL utilizados para la creación de la base de datos fuente y Data Mart.
- 2) Proyectos de Visual Studio 2019 tipo SQL Server Integration Services (SSIS) y SQL Server Analysis Services (SSAS) para la carga de datos fuente, creación del Data Mart, y procesamiento OLAP.
- 3) Datos fuente en Excel de crímenes, población y superficie cantonal.
- 4) Archivos de respaldo de las bases de datos fuente, Data Mart y OLAP.