

PRÁCTICA 1: BUSINESS INTELLIGENCE

Trabajo de análisis sobre la utilización de programas
destinados al estudio de datos a gran escala.

Realización de la primera
práctica de la asignatura
Introducción a los
Sistemas de Información.

índice

Introducción	2
Tabla de datos	2
Indicadores utilizados	3
Gráficos	4
Conclusiones	11
Miembros y funciones	12
Bibliografía	13

Introducción

En este trabajo de análisis estudiaremos las gráficas deducidas de un conjunto de datos (*dataset*) a gran escala sobre homicidios cometidos en diferentes ciudades del mundo entre los años 1980 y 2014, **centrándonos principalmente en los crímenes realizados en EEUU**, con el fin de mejorar la comprensión del funcionamiento de los sistemas web para obtener conjuntos de datos y de los sistemas software de procesamiento analítico. Por ello, hemos escogido la plataforma **Kaggle** como la fuente de nuestro conjunto de datos, disponible gratuitamente en la red; mientras que como sistema de procesamiento analítico hemos elegido **Microstrategy**, debido a que ofrecía mejores servicios que otros sistemas software gratuitos del mismo tipo, como es el acceso completo a todas las funcionalidades del software y cursos de formación gratuitos en su página web oficial.

Tabla de datos

La tabla de datos que se ha utilizado consta de campos como la cantidad, edad, raza y sexo, tanto de víctimas, como de homicidas, siendo estos los campos más utilizados. Sin embargo, también hemos contado con otros atributos como la ciudad y el estado donde se produjo el crimen, el tipo de dicho crimen, la agencia que lo registró, el arma usada y la fuente que proporcionó la información, así como el mes y el año de estos crímenes, entre otros más.

En total, el *dataset* seleccionado consta de 24 columnas, 2 de ellas como atributo geográfico (ciudad y estado).

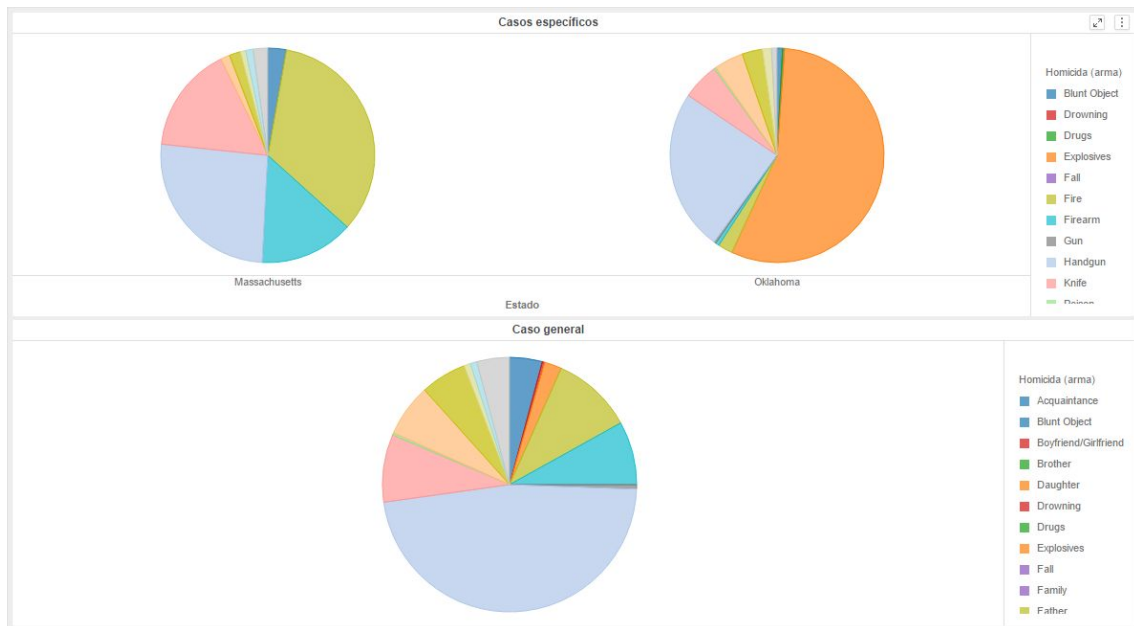
Indicadores utilizados

Los indicadores empleados han sido la **cantidad de víctimas** y la **cantidad de homicidas**, ya que ofrecen valores métricos y con los que poder realizar operaciones de cálculo diversas. Todas las gráficas se basan en uno u otro indicador.

En este caso, no se han creado nuevos indicadores puesto que el conjunto de datos ofrecía análisis más que variados y suficientes.

Gráficos

Armas más usadas en los casos de homicidio



Este gráfico representa el porcentaje de armas usadas en los homicidios estudiados. En este caso, hemos representado ese porcentaje en general y 2 casos particulares correspondientes a 2 estados de EEUU en **3 gráficos circulares**.

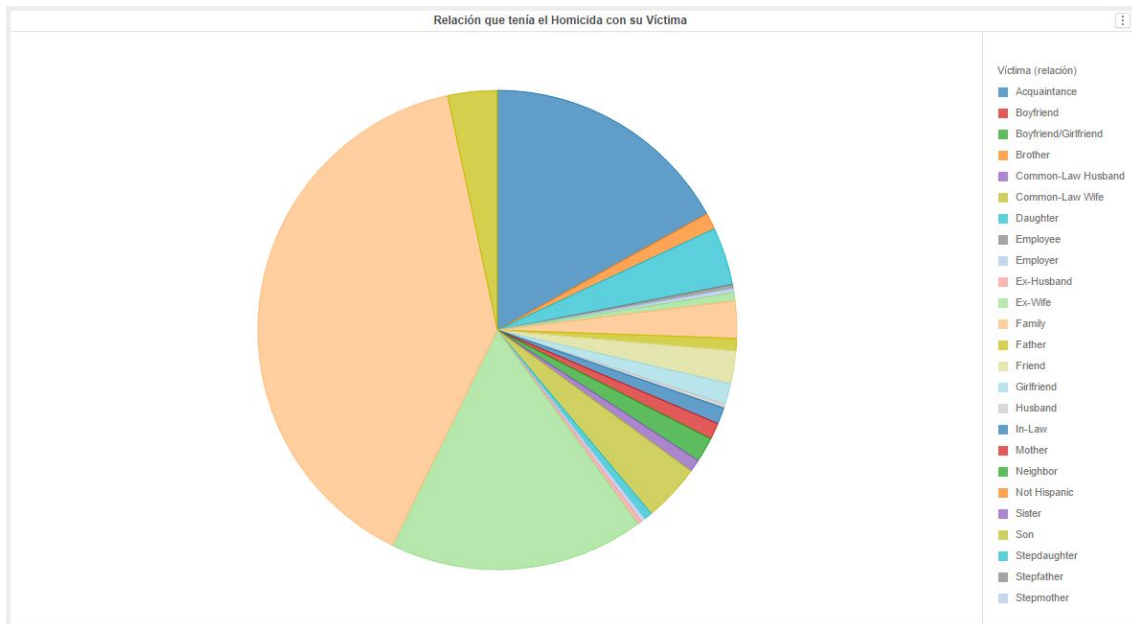
Hemos decidido usar este tipo de gráfico puesto que es la forma más simple de ver los datos representados en forma de porcentaje, así como podemos ver la diferencia que hay entre el uso de distintas armas. Se han elegido Massachusetts y Oklahoma como casos particulares a señalar, debido a que el arma más usada en esos estados (explosivos y fuego respectivamente) no corresponde con el arma más usada en general.

Recursos:

Atributos: Arma y Estado.

Indicadores: Cantidad de víctimas.

Relación de la víctima con el homicida



En este gráfico se muestran los diferentes tipos de relación que tenía la víctima con el homicida, hemos decidido utilizar un **gráfico circular** de nuevo, ya que facilita mucho la comparación entre ellos.

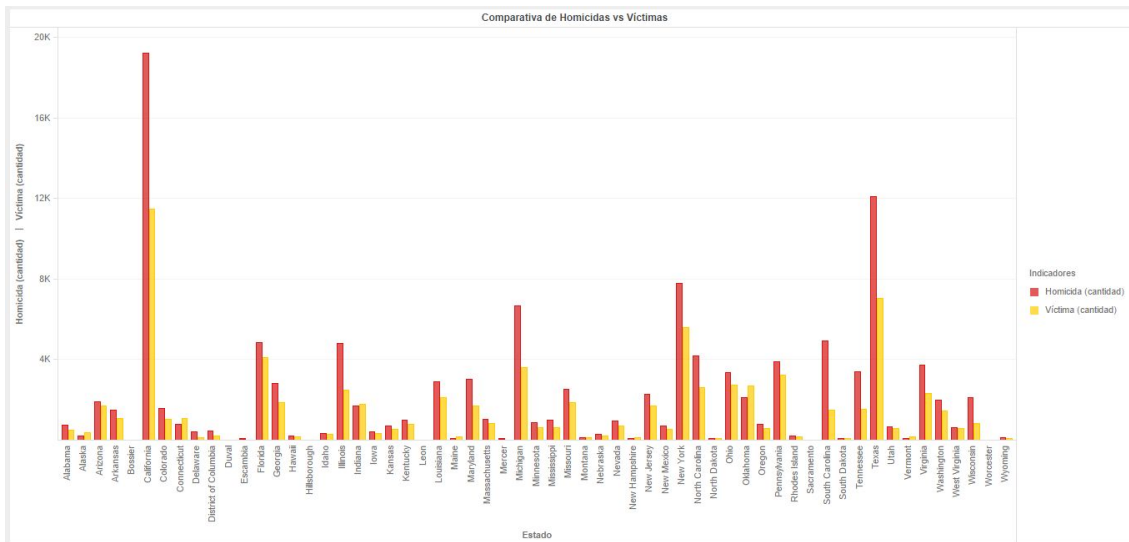
Este gráfico es el más adecuado para mostrar este tipo de información, donde existen una variedad en los datos, ya que puede observarse de una forma sencilla no solo el valor más repetido, sino también la variabilidad del resto de ellos.

Recursos:

Atributos: Relación de la víctima.

Indicadores: Cantidad de víctimas.

Comparativa de homicidas y víctimas por Estado de EEUU



Este **gráfico de barras** muestra una comparativa de la cantidad de homicidas y víctimas naturales de cada Estado de los EEUU.

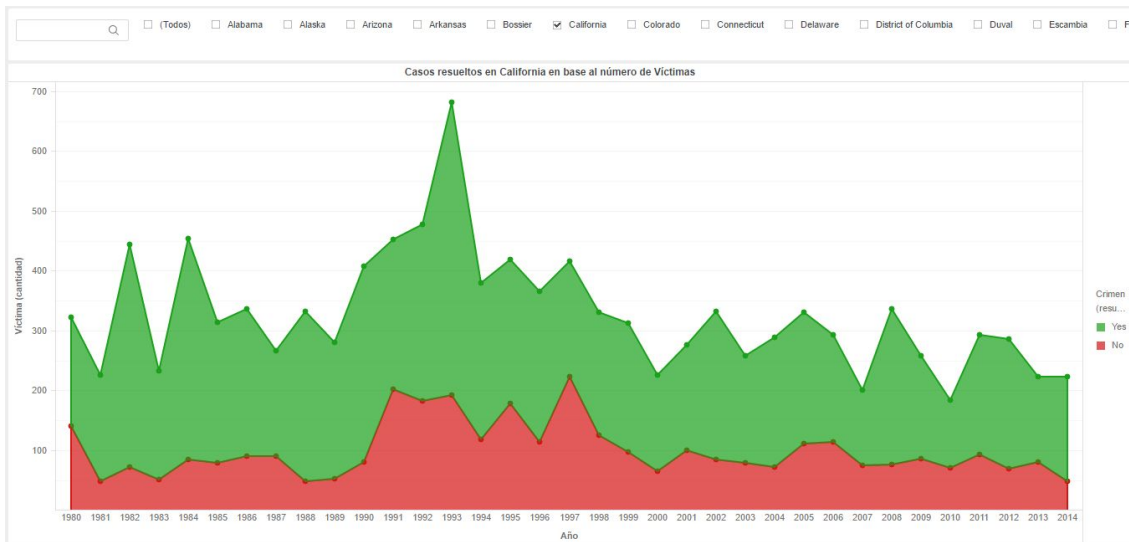
Hemos utilizado un gráfico de barras puesto que es la forma más cómoda y simple de mostrar esta información, pudiendo ver datos como los estados con más homicidas y víctimas, la diferencia que existe en número de homicidas y víctimas para cada estado... etc. Básicamente, proporciona de forma clara y visual una comparación rápida entre datos registrados en los diferentes estados.

Recursos:

Atributos: Estado.

Indicadores: Cantidad de víctimas y Cantidad de homicidas.

Evolución de la resolución de crímenes en California



Aquí se muestra la evolución de la cantidad de homicidios resueltos (verde) y no resueltos (rojo) en base al número de víctimas totales en California, entre los años 1980 y 2014. Hemos elegido este lugar debido a que en el estudio previo, descubrimos que ese estado tenía el índice más alto de homicidios y queríamos ver su evolución.

Para mostrar estos datos, hemos usado un **gráfico de área**, una de las alternativas para mostrar el desarrollo de unos valores a lo largo del tiempo como el gráfico de barras, solo que de esta forma podemos ver la tendencia que toman estos valores de una forma muy intuitiva.

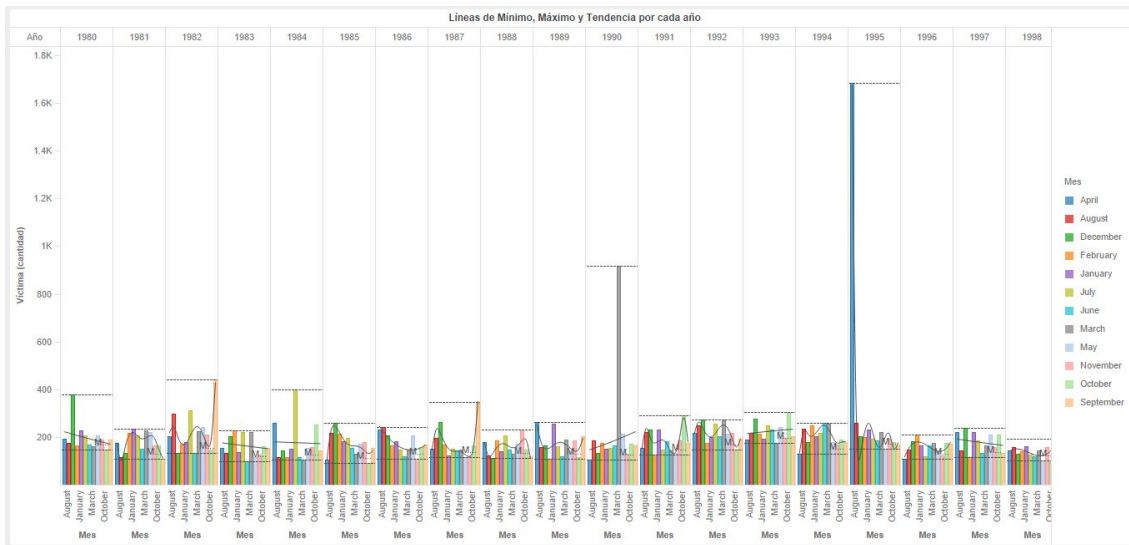
Recursos:

Atributos: Crimen resuelto y Año.

Indicadores: Cantidad de víctimas.

Visualización: Marcadores activados.

Víctimas en EEUU al mes desde el 1980 al 1998



Aquí tenemos otro **gráfico de barras** donde mostramos la cantidad de víctimas mensual en los EEUU durante los años 1980-1998.

Hemos utilizado este gráfico porque además de ser la forma más visual para ver la evolución de los datos a lo largo del tiempo, queríamos añadir valores como el mínimo y el máximo de víctimas en función del año, así como la curva de tendencia anual para reflejar más información.

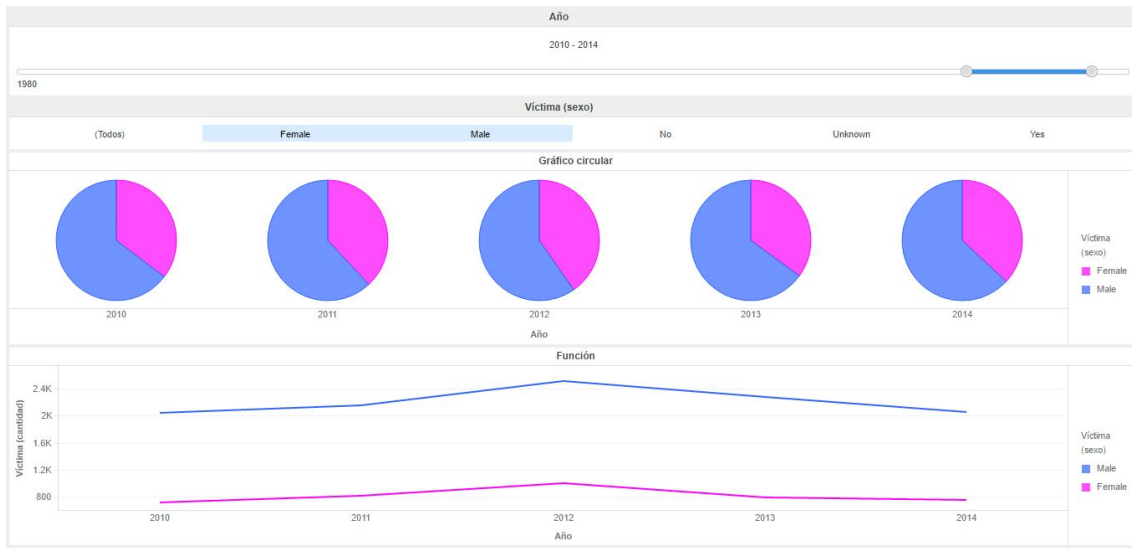
Recursos:

Atributos: Mes y Año.

Indicadores: Cantidad de víctimas.

Visualización: Máximo, Mínimo y Línea de tendencia (todo anual).

Víctimas en EEUU en base al sexo entre 2010 y 2014



Para estos datos hemos decidido unir dos gráficos distintos, el **gráfico circular** y el **gráfico de línea** mostrando la diferencia por sexo de las víctimas de homicidio en los EEUU entre los años 2010 y 2014.

Ambos gráficos se basan en el estudio de las víctimas en función de su sexo: en el primero, puede observarse de claramente los porcentajes de ambos sexos durante el tiempo; mientras que en el segundo, destaca la facilidad con la que puede visualizarse el número de víctimas en ambos sexos, en lugar de un tanto por ciento.

Recursos:

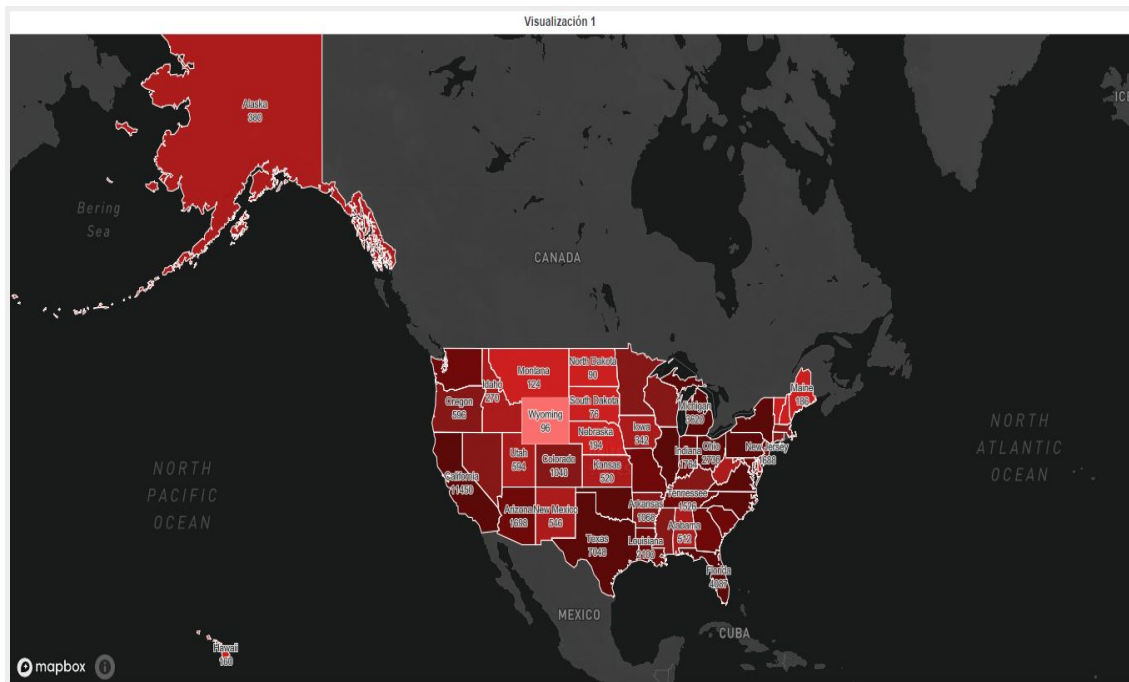
Atributos: Sexo de la víctima y Año.

Indicadores: Cantidad de víctimas.

Filtros: Sexo de la víctima (hombre y mujer) y Año (2010 - 2014).

Visualización: Colores personalizados y dependientes entre vistas.

Víctimas en EEUU clasificadas por estado



Aquí podemos ver la cantidad de víctimas de homicidio en cada estado de los EEUU representadas en un **mapa**.

Hemos decidido usar este tipo de representación porque es una forma muy visual de ver el número de muertos en cada estado, ya que están coloreados de menor (claro) a mayor (oscuro) número de víctimas. Se trata de una visualización muy práctica que ofrece ver rápidamente en qué zonas del país han habido más víctimas.

Recursos:

Atributos: Estado.

Indicadores: Cantidad de víctimas.

Visualización: Marcador de área, Etiquetas de nombre y valores.

Conclusiones

Microstrategy es una herramienta muy potente como sistema de información para el soporte de decisiones. Hemos podido construir gráficos sobre un conjunto de datos de forma intuitiva y muy sencilla, permitiéndonos analizarlos de diferentes formas y aportándonos información muy útil. Aplicando esto a la toma de decisiones de una empresa, consideramos que resulta una acción muy beneficiosa.

Por otro lado, uno de los factores más importantes ha sido descubrir que este software ofrece todas sus funcionalidades aún siendo gratuito, lo que creemos que permite un inicio a la *Data Science* con una herramienta muy completa.

En cuanto al **conjunto de datos** que hemos utilizado, era bastante amplio, pero hemos podido extraer la información que mostramos en las gráficas casi sin dificultad, lo que nos hace pensar qué pasaría si el conjunto de datos fuese aún más extenso: poniéndonos en situación de una empresa, pensamos que podrían obtenerse aún más información útil y relevante para la misma, lo que supondría que la toma de decisiones fuera más eficiente, produciendo así más beneficios.

Miembros y funciones

José Ignacio Martín Codes.

1. *Búsqueda de datasets.*
2. *Realización de la presentación.*
3. *Desarrollo del informe.*
 - 3.1. *Introducción.*
 - 3.2. *Conclusión.*

Iván Gómez Lanzas.

1. *Búsqueda de datasets.*
2. *Realización de las gráficas.*
3. *Desarrollo del informe.*
 - 3.1. *Descripción de las gráficas.*
 - 3.2. *Conclusión.*
 - 3.3. *Bibliografía.*

Antonio J. Galán Herrera.

1. *Búsqueda de datasets.*
2. *Realización de las gráficas.*
3. *Desarrollo del informe.*
 - 3.1. *Descripción de las gráficas.*
 - 3.2. *Conclusión.*
 - 3.3. *Bibliografía.*
 - 3.4. *Modificaciones y correcciones finales.*

Luis Pardo Pascual.

1. *Búsqueda de datasets.*
2. *Realización de las gráficas.*
3. *Desarrollo del informe.*
 - 3.1. *Descripción de la tabla de datos.*
 - 3.2. *Razonamiento de los indicadores.*
 - 3.3. *Conclusión.*

Bibliografía

Conjunto de datos.

<https://www.kaggle.com/jyzaguirre/us-homicide-reports>

Análisis de las gráficas.

<https://datavizcatalogue.com/ES/>