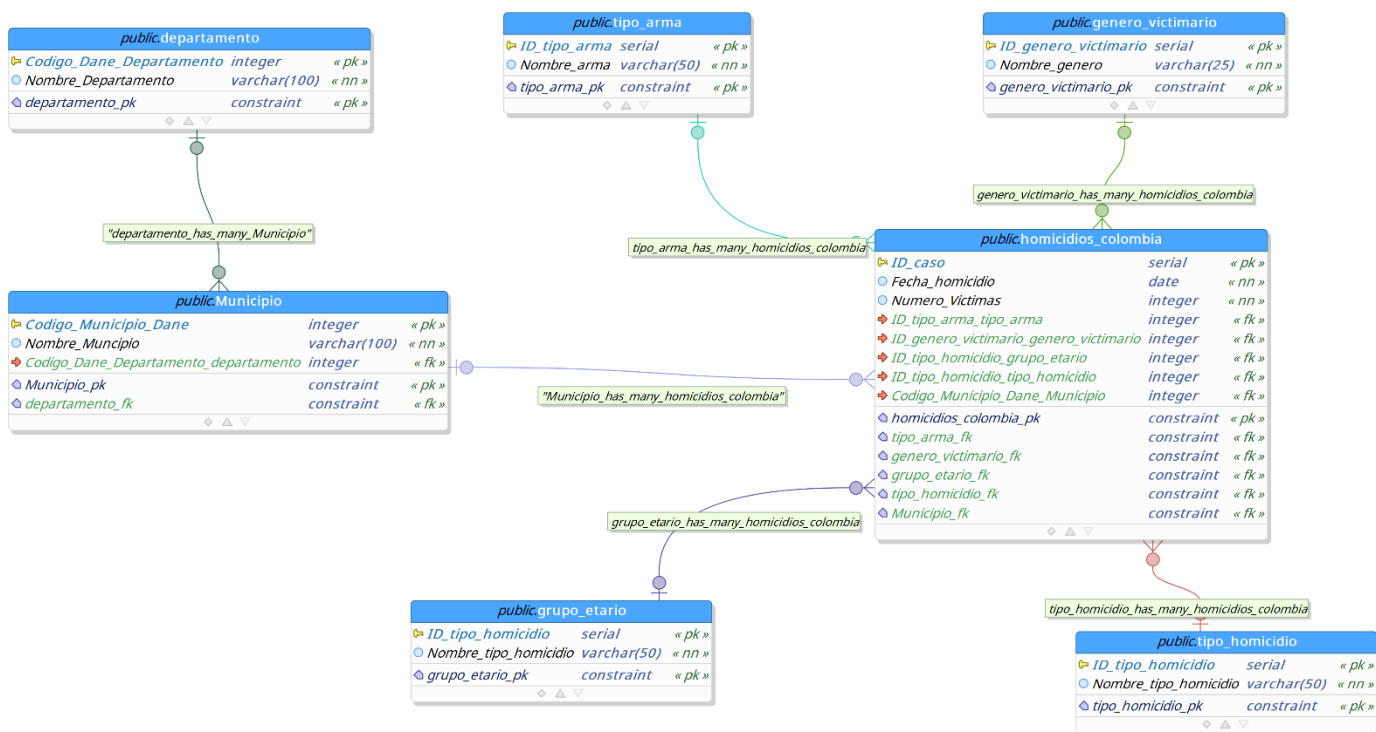


Ilustración 1 MODELO RELACIONAL 3RA FORMA NORMAL



Fuentes de Datos:

Para el siguiente proyecto se hizo uso de dos bases de datos sacadas directamente de la DATOS ABIERTOS, es decir que todos los datos utilizados son directamente sacados de la información pública que presenta el Gobierno Nacional en la siguiente página: www.datos.gov.co.

Reporte Homicidios y Homicidios accidente de tránsito Policía Nacional (2021):

La principal recolección de datos se realizó en esta base de datos publica a partir de los datos publicados por la Dirección General de la Policía Nacional, donde nos presenta información catalogada como seguridad y defensa sobre los delitos de homicidios desde el año 2010 hasta finales de octubre del presente año.

Para acceder a la base de datos nombrada esta es la dirección web

[<https://www.datos.gov.co/Seguridad-y-Defensa/Reporte-Homicidios-y-Homicidios-accidente-de-tr-n-s/ha6j-pa2r>].

Esta base de datos tiene un periodo de actualización trimestral, por tanto, el análisis presentado tiene datos hasta la penúltima actualización del año 2021, por tanto, al verificar la base de datos puede encontrar diferencias en los datos proporcionados.

La base de datos presenta un diseño simple organizado por 9 columnas entre ellas DEPARTAMENTO (Nombre del departamento), MUNICIPIO (Nombre del municipio), CODIGO DANE (Código numérico del municipio), ARMAS MEDIOS (Medio con el cual se realizó el Homicidio), FECHA HECHO (Fecha del hecho), GENERO (Genero del victimario), GRUPO ETARIO (Grupos catalogados por la edad del victimario). Los elementos de cada atributo o columna están inicialmente en formato texto con una cantidad aproximada un poco menor a los 190000 datos.

Departamentos y municipios de Colombia (2021):

Como fuente de información complementaria, esta base de datos es de carácter político-administrativo del país, publicado por el Departamento Administrativo Nacional de Estadística actualizada hasta el 5 de noviembre del presente año, se usó principalmente para obtener los códigos y nombres de las regiones para contrastar la información obtenida en la base de datos.

Para acceder a la base de datos nombrada esta es la dirección web

[<https://www.datos.gov.co/Mapas-Nacionales/Departamentos-y-municipios-de-Colombia/xdk5-pm3f>].

Esta base de datos podría ser actualizada con una frecuencia Anual por tanto se usó la base de datos del año 2021.

Repositorio GitHub:

Por último, todo el contenido del proyecto (tanto la base de datos y los códigos de desarrollo) estará en un repositorio de GitHub con nombre ProyectoIGDB en el siguiente enlace [<https://github.com/TheChieft/ProyectoIGDB>]. Recordemos que existió completa participación por juanis07 (Juanita Robles), Yucaloid (Johan Caro) y TheChieft (Andrés Yañez).

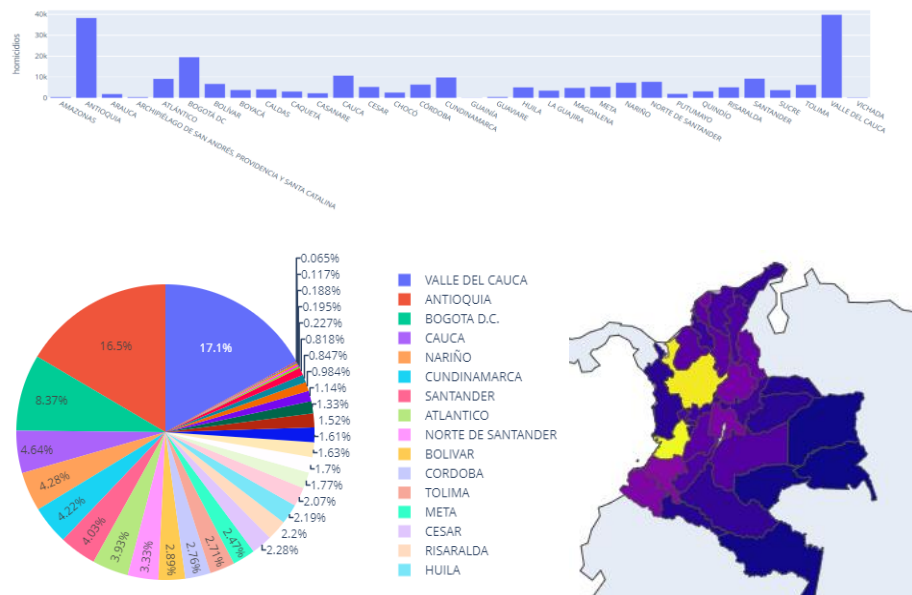
Para la completa funcionalidad de repositorio por favor siga las instrucciones del README dentro de GitHub

Implementación total del proyecto y análisis de los datos

Después de realizar todas las implementaciones antes presentadas en las anteriores entregas, se presentan los análisis a partir de las herramientas SQL en conjunto con DASH (Plotly), Para el manejo de la base de datos y la implementación visual de los datos.

Escenarios de análisis

1. Homicidios por departamentos: El primer análisis que realizamos fue el de conocer cómo se distribuyen los casos de homicidios por los departamentos del país en los últimos 10 años, en los cuales tuvimos los siguientes resultados



Como podemos observar los departamentos con más homicidios en Colombia son el departamento del Valle del Cauca y Antioquia con casi 40.000 homicidios cada uno y los departamentos con menos cantidad de homicidios son Vichada, Amazonas y el archipiélago de San Andrés, Providencia y Santa Catalina, no llegando a superar los 1000 homicidios en los últimos 10 años, siendo el menor de todos Guainía con menos de 160 homicidios.

La visualización de estos datos se realizó usando la función COUNT de SQL para que contara los casos de homicidio por departamento y luego con la función GROUP BY se adjuntaron a la tabla departamento (“todos los códigos de SQL están en el archivo proyectoSQL.py”)

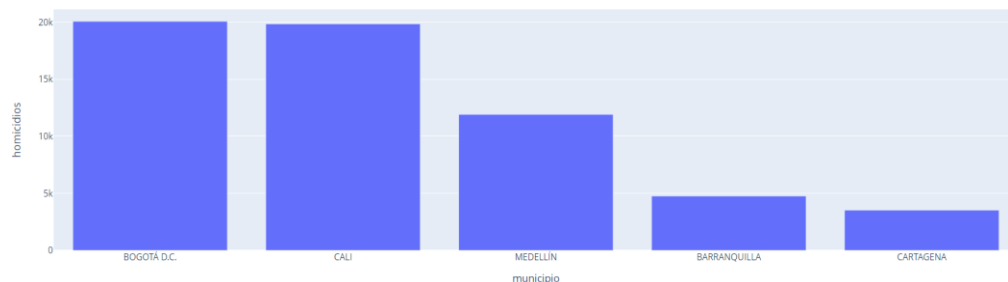
Luego de tener el código SQL por medio de Dash se generó el código para hacer la gráfica de barras usando en el eje x los departamentos y en el eje y la cantidad de casos de homicidios. También se realizó una gráfica de pastel aun que nos damos cuenta de que no es la mejor visualización por eso se realizó una última gráfica, un mapeado de Colombia con sus respectivas cantidades de homicidios (grafica de calor, a partir de un archivo JSON)

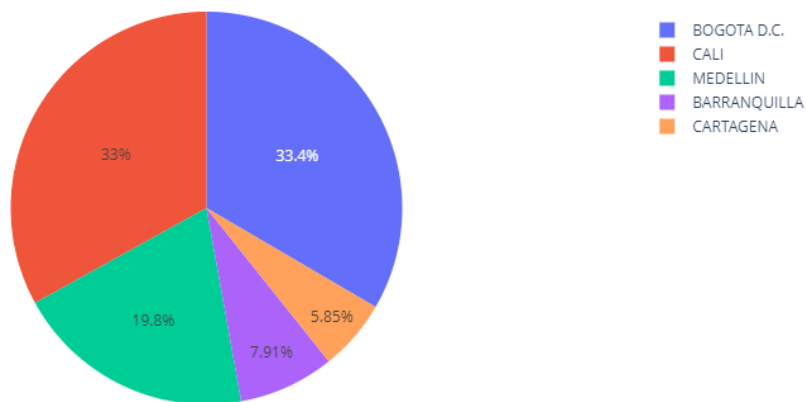
<https://gist.githubusercontent.com/john-guerra/43c7656821069d00dcbc/raw/be6a6e239cd5b5b803c6e7c2ec405b793a9064dd/Colombia.geo.json>].

Ventajas y desventajas de la proyección:

La proyección en barras nos da las ventajas de ver fácilmente como se distribuyen la cantidad de homicidios por los diferentes departamentos, pero discutiéndolo como grupo tiene la falla de que no tiene en cuenta las proporciones de habitantes por departamento.

2. Municipios con más homicidios: nuestro grupo quería conocer los 5 municipios en los cuales han ocurrido más homicidios en los últimos 10 años, en los cuales obtuvimos los siguientes resultados





Como podemos observar en los municipios de Colombia los que tienen más cantidad de homicidios son 5 de las ciudades más importantes del país Bogotá, Cali, Medellín, Barranquilla y Cartagena, las dos primeras superando la barrera de los 20.000 homicidios en los últimos 10 años.

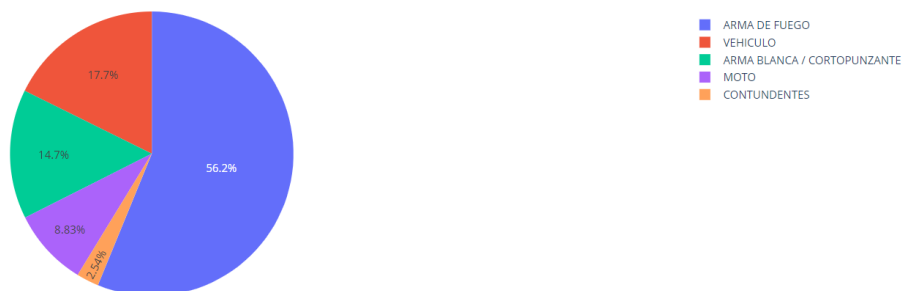
La visualización de estos datos fue muy parecida a la anterior solo que cambiamos los departamentos por la tabla de municipios y también se aplicó un ORDER BY en el COUNT del número de víctimas en forma descendente para que estuvieran ordenadas de mayor a menor.

Y luego en el Dash se imprimió en un código de barras las 5 primeras ciudades que al estar ordenadas en forma descendente serían las 5 con más homicidios

Ventajas y desventajas de la proyección:

Esta proyección tiene la misma ventaja y desventaja de la gráfica anterior que, aunque nos dice la cantidad de homicidios y nos la muestra de una manera fácil, tiene el problema de que no tiene en cuenta la proporción de población por departamento.

- Armas más utilizadas por homicidios: en esta proyección queríamos observar cuales eran las armas más comunes en un caso de homicidio, en nuestro análisis obtuvimos el siguiente grafico



Como podemos observar la mayoría de los homicidios en el país son cometidos por medio de armas de fuego con más de 120.000 casos, seguido por vehículos con aproximadamente 40.000 casos, es importante aclarar que los homicidios cometidos por vehículos solo cuentan aquellos accidentes de tránsito donde el conductor tuvo la culpa del incidente, esto se conoce como homicidio culposos. Y en el tercer puesto tenemos los asesinatos con armas cortopunzantes.

La visualización de esta tabla se realizó utilizando un COUNT en los casos de homicidio usando un GROUP BY para que los fuera contando por tipo de arma, luego se utilizó un ORDER BY para ordenarlos de forma descendente y luego por medio de Dash se graficaron los primeros 5 para graficar las armas más utilizadas en homicidios

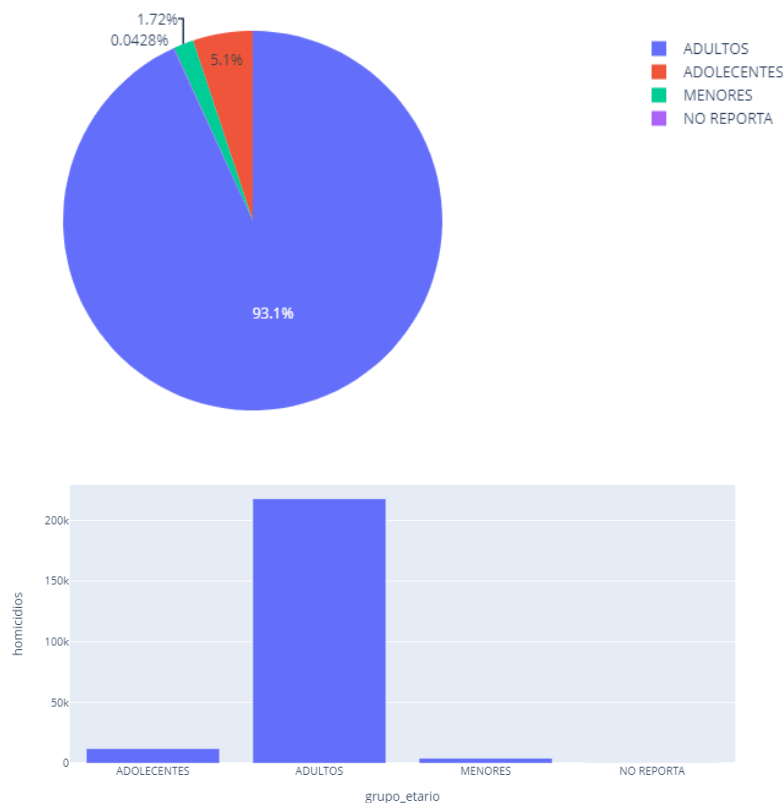
Ventajas y desventajas de la proyección:

Las gráficas de pie nos otorgan las ventajas de ver los porcentajes de una manera fácil y que la información sea más visual y estética, otra idea que tuvo nuestro grupo fue graficarla en una gráfica de barras, pero por temas de la realización de este documento creímos que la mejor forma era usar la gráfica de pie.

Otra ventaja de esta forma de grafica es que si quisiéramos ver el caso contrario de las armas menos utilizados solo tendríamos que poner el código SQL en forma ascendente y podríamos obtener ese dato muy fácilmente.

4. Grupo etario victimarios: antes de hablar de la visualización es importante aclarar que es un grupo etario, un grupo etario son grupos divididos por ciclos de vida, en el caso de nuestra base de datos tenemos menores (0 – 12) años, adolescentes [12-18) y adultos (> 18).

En esta visualización queríamos obtener la información de cuantos crímenes son cometidos por menores de edad y estos fueron los datos que obtuvimos



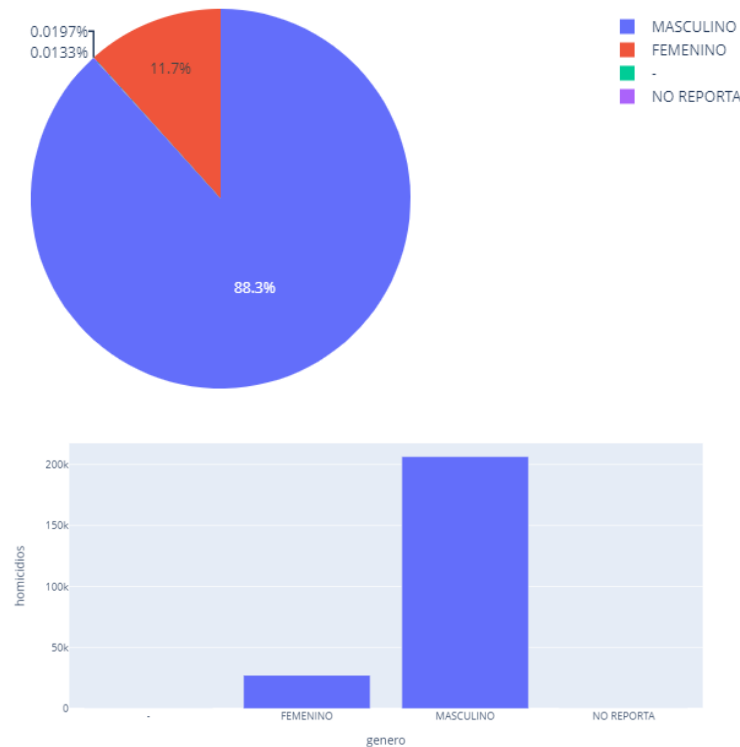
Como podemos observar el 6.82% de los homicidios en Colombia en los últimos 10 años son cometidos por menores de edad, el 93.1% por mayores de edad y solo el 0.0428% es desconocido el grupo etario del victimario, si pasamos esto a cantidades podemos decir que aproximadamente 15950 homicidios de los ocurridos en los últimos 10 años han sido cometidos por menores de edad, una cifra que aunque baja en comparación con los adultos es demasiado alta a nivel general estamos hablando de 1600 homicidios cometidos por menores al año.

La carga de y visualización de estos datos es muy parecida a la anterior un COUNT de casos agrupados por medio de GROUP BY en grupos etarios y luego pasada a Dash y graficada con una gráfica de pie.

Ventajas y desventajas de la proyección:

Nuestro grupo cree que la visualización de pie de estos datos es la mejor que podíamos hacer ya que directamente responde nuestra pregunta del porcentaje de crímenes cometidos por menores de edad y la verdad creemos que no aporta ninguna desventaja significativa. Mientras que para ver las cantidades el grafico de barras es la mejor opción.

5. Genero victimarios: en esta visualización queríamos observar la proporción de crímenes cometidos por hombres y mujeres, si tiende a una igualdad o si hay un sexo que cometa más homicidios que otros y esta fue la gráfica de nuestro análisis



Como podemos observar en la gráfica el 88.3% de los homicidios en el país son cometidos por hombres y solo un 11.7% por mujeres, podemos observar que hay un claro desbalance y que en caso de homicidios hay una desigualdad que muestra que los hombres cometen más asesinatos. Por otro lado, podemos observar que el 0.0133% de la gente que ha cometido un homicidio no se identifica como hombre o mujer y que un 0.0197% es desconocido (seguramente porque no se sepa quien cometió el homicidio).

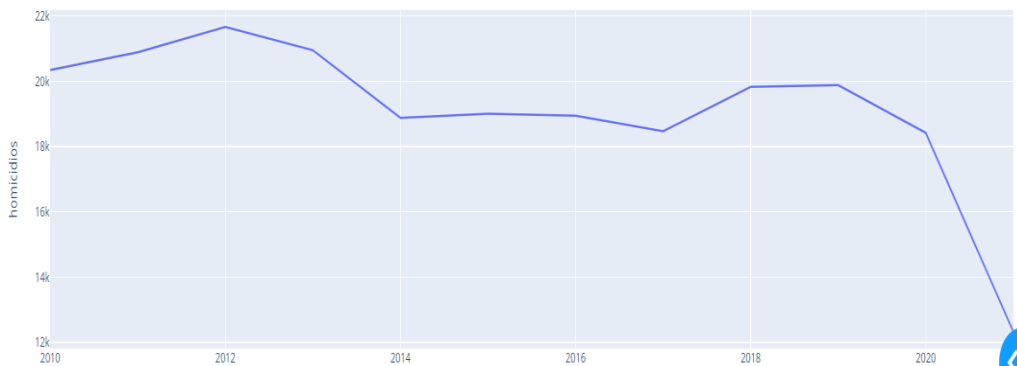
La carga de estos datos fue realizada prácticamente de la misma manera que la del análisis anterior solo que en vez de utilizar grupo etario se utilizó la información de la tabla genero del victimario.

Ventajas y desventajas de la proyección:

Elegimos esta visualización debido que, al ser solo 4 opciones de datos, las gráficas de pie quedan mejor al ser pocos datos, además que nos da

automáticamente los valores porcentuales que son los que necesitábamos para responder a nuestra pregunta inicial. Mientras que para ver las cantidades el gráfico de barras es la mejor opción.

6. Homicidios por años: en nuestro último análisis queríamos observar cómo estos homicidios se han distribuido a través de los años y si tienen una tendencia de aumento o de disminuir. Nuestros análisis lanzo los siguientes resultados



Como podemos observar los homicidios tienen una tendencia a la disminución pasando de 20.000 en 2010 a 12.000 en 2021 teniendo el pico más alto en el año 2012 con un aproximado de 22.000 homicidios.

La carga de datos se realizó por medio de la función `EXTRACT` para extraer el año de los casos de los homicidios y luego con un `COUNT` y un `GROUP BY` se fueron guardando la cantidad de homicidios por año.

Ventajas y desventajas de la proyección:

La ventaja por la cual escogimos la gráfica de línea fue para simular una línea de tiempo para que sea más fácil identificar los picos y disminuciones de los homicidios en el país, aunque la desventaja que tiene esto es que con la línea es más difícil ver el valor exacto donde esta

Conclusiones proyecto

- Se encontró una fuente de datos viable respaldada por la fiscalía nacional, con una gran cantidad de datos relacionados a los homicidios en el país los que cuales se pudieron subir satisfactoriamente a SQL y realizar una base de datos en tercera forma normal
- Se cargaron correctamente los datos a la base de datos, se realizó correctamente el estudio de los datos y se realizó correctamente la conexión

de dicha base de datos con Python, además con lo visto en clase sobre SQL con uso de Subconsultas, consultas, las funciones GROUP BY y ORDER BY se realizó la perfecta visualización y generación de los datos para pasarlos a Dash para la generación de los gráficos.

- Por medio de lo aprendido en clase de Dash y el módulo Psycogs2 se desarrollaron en perfecto estado 6 graficas utilizando Dash y Psycogs2 a las cuales se realizaron análisis con las preguntas iniciales del proyecto

Conclusiones de los datos

- Se concluyo que los departamentos con mayores casos de homicidios en el país son el valle del cauca y Antioquia rondando los 40.000 homicidios
- Se concluyo que los 5 municipios con más homicidios son 5 de las ciudades más importantes del país Bogotá, Medellín, Cali, Barranquilla y Cartagena
- Se concluyo que el medio más común para cometer un homicidio en el país es a través de las armas de fuego con el 56.2% de los homicidios del país
- Se concluyo que el 6.78% de los homicidios cometidos en Colombia son realizados por menores de edad y que la gran mayoría con un 93.2% fueron realizados por mayores de edad
- Se concluyo que la gran mayoría de homicidios son cometidos por los hombres con un porcentaje del 88.3% mientras que las mujeres tienen un 11.7%
- Se concluyo que la cantidad de homicidios tiene una tendencia de reducción con el pasar de los años y que en los últimos 10 años el año con más homicidios fue el 2012 llegando casi a los 22 homicidios.

Otras conclusiones

- Las bases de datos es una herramienta muy practica para manejar información de gran tamaño y poder analizar o llevarse una idea general del problema a tratar, al mirar las tablas ya nos pudimos dar una idea de las preguntas problemas que realizaríamos en el proyecto, lo cual significa que la estructura y orden es más fácil de interpretar.
- El manejo y uso de bases de datos lo hemos podido complementar en proyectos similares como el proyecto de la materia de probabilidad y estadística y otros proyectos personales que se han venido trabajando a la par de la universidad.