



# UNIVERSIDAD DEL ROSARIO

ESCUELA DE INGENIERÍA, CIENCIA Y TECNOLOGÍA, UNIVERSIDAD DEL ROSARIO

## Proyecto de Ingeniería de Datos: Segunda entrega

CLAUDIA LILIANA HERNANDEZ GARCIA [CLAUDIALI.HERNANDEZ@UROSARIO.EDU.CO](mailto:CLAUDIALI.HERNANDEZ@UROSARIO.EDU.CO)

Juan Sebastián Rodríguez [juansebasti.rodrig10@urosario.edu.co](mailto:juansebasti.rodrig10@urosario.edu.co)

Natalia Rojas Suarez [Nataliak.rojas@urosario.edu.co](mailto:Nataliak.rojas@urosario.edu.co)

Nicolas Peña [nicolasst.pena@urosario.edu.co](mailto:nicolasst.pena@urosario.edu.co)

Samuel Barbosa [samuel.barbosae@urosario.edu.co](mailto:samuel.barbosae@urosario.edu.co)

MACC

21 de Noviembre del 2022

### Primera Entrega

#### BASE DE DATOS

Reporte Delitos sexuales Policía Nacional  
Departamentos y municipios de Colombia

#### INTRODUCCIÓN

Los delitos sexuales son un gran grupo de delitos que afectan de distintas maneras la libertad sexual, dignidad y el desarrollo sexual. Estos delitos se abordan de diferentes maneras dependiendo el contexto sociocultural y legal de cada territorio. Actualmente podemos ver un poco más de sensibilización del tema hablándose más en medios, favoreciendo la difusión de conocimientos, información y testimonios.

Existen distintos tipos de delitos sexuales, por ejemplo la extorsión sexual. A continuación podemos ver un gráfico con los porcentajes de cada país de la región de Latinoamérica.

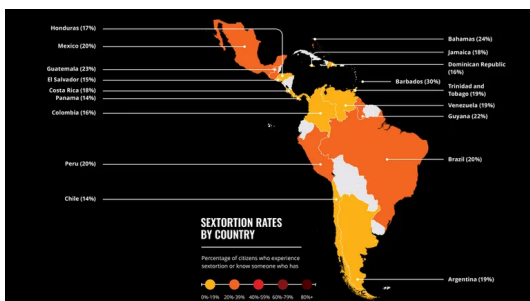


Figura 1: Extorsión sexual en Latino America.  
(Infobae, 2019)

Como podemos ver en la imagen anterior, la extorsión sexual existe en un 16 % en Colombia, lo que nos hace estar entre los porcentajes más bajos de América latina. En otro estudio donde se mide la violencia íntima a mujeres de entre 15 y 49 años, Colombia se encuentra con el porcentaje más alto, 18,4 %.

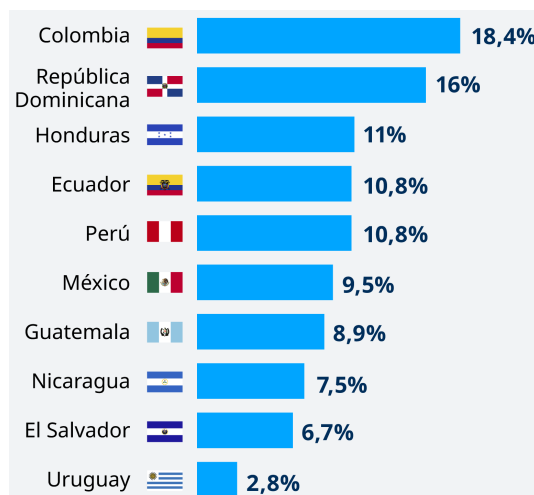


Figura 2: Prevalencia de violencia sexual.  
(OMS, 2019)

#### DESCRIPCION DEL PROBLEMA

Colombia es un país donde hay un sin número de delitos y violaciones contra la ley, muchas personas se sienten inseguras al estar en las calles de cualquier parte de la región. Nos enfocaremos principalmente en los delitos sexuales, intentando recrear una base

de datos de tal forma que sea clara de entender y sea lo más completa posible. Para esto usaremos dos bases de datos, una con la información de los delitos y otra con la información departamental. Esto lo haremos con el fin de poder obtener mejores resultados de búsqueda respecto a las regiones de Colombia y los delitos sexuales. Trataremos de encontrar las personas más vulnerables y las regiones para así poder tener un análisis claro para que la policía nacional o otras entidades puedan actuar en pro a la ciudadanía.

## REGLAS DE NEGOCIO

- Los delitos son realizados en varios departamentos de Colombia.
- Varios delitos son realizados en diferentes municipios de cada departamento.
- Cada departamento tiene un único código DANE.
- Cada municipio tiene un código DANE.
- Hay varios tipos de armas y/o medios para la realización del delito.
- Las fechas de cada delito están registradas en la base de datos.
- Se conoce el género y el grupo etario de la víctima.
- Cada delito está clasificado según su tipo.

## DIAGRAMA ENTIDAD RELACIONAL

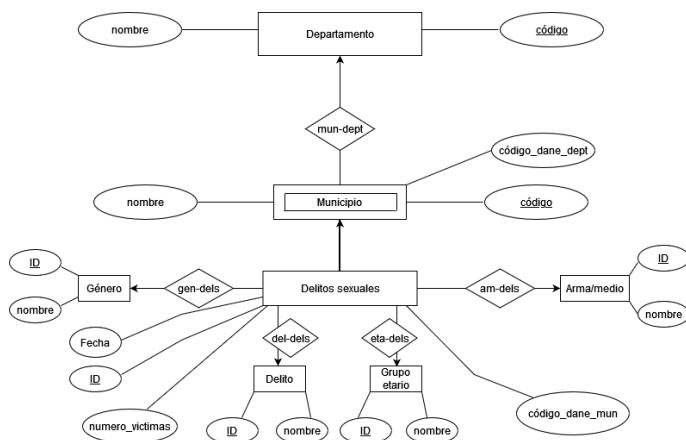


Figura 3: Diagrama entidad relacional.

## Segunda Entrega

## DIAGRAMA RELACIONAL NORMALIZADO EN TERCERA FORMA NORMAL

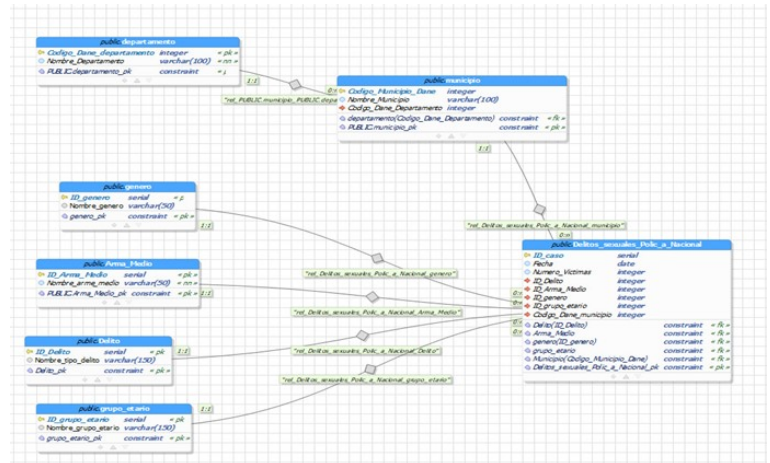


Figura 4: Diagrama entidad relacional.

## IMPLEMENTACION EN POSTGRESQL

De las dos conjuntos de datos que tomamos, la de delitos sexuales tenía una columna llamada "CODIGO DANE" referida al código DANE del municipio correspondiente. Esta columna se tuvo que retirar para pasar a la tercera forma normal, ya que contenía datos atómicos. Esta columna fue reemplazada por el otro conjunto de datos llamada "Departamentos y municipios de Colombia". Con la unión de conjuntos de datos se crearon siete tipos de tablas:

- Departamento.
- Municipio.
- Género.
- Arma\_Medio.
- Delito.
- Grupo\_Etario.
- Delitos\_sexuales\_Polic\_a\_Nacional
- Cada delito está clasificado según su tipo.

Y se creó una relación entre estas, como se puede observar en el diagrama relacional y así, transformar la base de datos en tercera forma normal.

## CARGA DE DATOS

Se descargaron los csv de la página del DANE, se tuvo que editar los archivos csv; se eliminaron las comillas, las tildes y se convirtió todo a mayúscula, utilizando la función buscar y reemplazar de excel. Luego se uso la función “COPY PUBLIC” de PostgreSQL. La ruta que se utilizó fue la del computador de Sebastián, por lo tanto esta ruta debe ser editada según donde se guarden los archivos csv. En el archivo tablas.txt (que se encuentra en el repositorio GIT en github) se puede apreciar la carga de datos y la creación de las tablas.

## POSIBLES ESCENARIOS DE ANÁLISIS

- Conocer el departamento con más delitos sexuales en la pandemia.
- Conocer el porcentaje de victimas sexuales que son menores de edad en Colombia.
- Identificar el genero más afectado por los delitos sexuales.
- Hallar el municipio con menos delitos sexuales en la pandemia.

## DESARROLLO EN PYTHON

Se creó el archivo de python y se hizo un import de la librería “psycopg2”, luego de realizar la conexión se usó la función “cursor.execute” para hacer las consultas de las tablas. En el repositorio GIT en github está adjunto el archivo conexión.py donde se puede apreciar el código.

### Tercera Entrega

## DASH

### I. Consultas

En primer lugar, se crean las consultas en postgresql que relacionan dos columnas para analizar las posibles situaciones que nos ayuden a solucionar el problema propuesto con base en nuestra base de datos. Antes de declarar cada consulta que representa cada relación establecida anteriormente, se importan las librerías psycopg2, que permite la conexión con Python, dash, que realiza las gráficas y pandas, que proporciona herramientas para leer los datos a partir del archivo .csv y crear el dataframe.

Luego, se declaran las funciones para realizar las graficas con los datos correspondientes. Por último, se ejecuta el módulo desde la consola retornando un enlace que nos lleva a una página donde se encuentran las gráficas creadas pudiendo así, interactuar con ellas para poder hacer el respectivo análisis. Lo dicho anteriormente se encuentra en el repositorio de GIT en su respectivo archivo.

## II. Resultados graficos del DASH

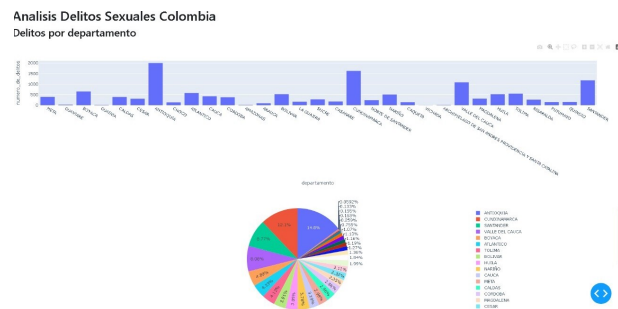


Figura 5: Histograma y diagrama de torta de delitos por departamento.



Figura 6: Mapa de calor de número de delitos por departamento.



Figura 7: Diagrama de torta de delitos según el grupo etario de la víctima.

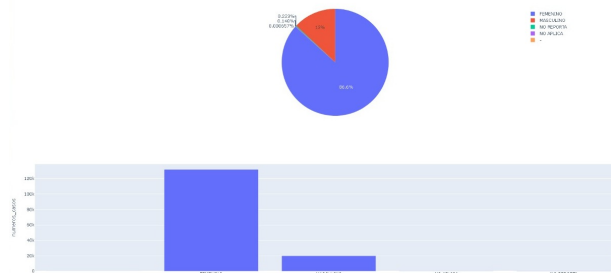


Figura 8: Histograma y diagrama de torta de delitos según el género de la víctima.



Figura 9: Histograma de número de delitos por municipio.

### III. Analisis de los graficos

- **Gráfica 1 y 2:** Como nuestro primer escenario de análisis es conocer los cinco departamentos con más delitos sexuales durante la pandemia, primero se filtro el tiempo en el cual se realizaron los delitos sexuales para poder analizar las graficas respecto a la pandemia. Se puede ver en el grafico que el departamento con más delitos de este carácter es Antioquia, aproximadamente con más de 2.000 delitos sexuales seguido de Cundinamarca, Santander, Valle del Cauca y Boyacá. Se puede pensar que la inseguridad ligada a Antioquia y a los otros cuatro departamentos podría relacionarse con la gran cantidad de abusos, también se puede decir que las redes son un medio importante para aumentar esta cifra, ya que durante la pandemia las restricciones de aislamiento no le permitian a las personas salir de sus casas pudiendo así, realizar ciber-abusos de tipo sexual.

**Resultados positivos:** Es posible identificar que Antioquia es el departamento con más delitos sexuales durante la pandemia.

**Resultados negativos:** Como hay una gran diferencia de estos cinco

departamentos con los demás, no es posible identificar una escala clara de los delitos para los departamentos con menor cantidad.

- **Gráfica 3:** En el gráfico de torta se puede evidenciar los porcentajes de delitos sexuales de cada grupo etario refiriendose a las víctimas. Se puede ver que el grupo etario más afectado por los delitos sexuales son los menores (refiriendose a menores de 14 años) con un 51 % y el grupo con menos víctimas son los adolescentes con 19.47 %.

**Resultados positivos:** Es claro el contraste del porcentaje de los niños con el los demás grupos etarios para identificarlo como el más afectado.

**Resultados negativos:** La base de datos sobre la que estamos trabajando tiene un error, al mirar el gráfico se puede ver que hay un grupo llamado “adolescentes” y otro llamado “adolescentes”, por lo cual para saber el porcentaje real de los adolescentes hay que sumar estas dos proporciones ya que hacen parte del mismo grupo etario. Además, hay un grupo nombrado “no aplica”, lo que nos dificulta saber si este porcentaje podría hacer que aumente algun grupo etario.

- **Gráfico 4:** Gracias al diagrama de torta podemos saber el porcentaje de delitos sexuales respecto al genero de la víctima, y con el histograma podemos ver la cantidad de víctimas también respecto al género. Podemos ver una diferencia contundente entre los hombres y las mujeres, siendo las mujeres más afectadas por los delitos sexuales. Más de 120.000 mujeres han sido víctimas, correspondiente al 86.6 %.

**Resultados positivos:** Las gráficas son claras y la diferencia está bastante marcada entre los géneros, así concluir es menos ambiguo.

**Resultados negativos:** Al haber un número tan grande de víctimas mujeres,

no es posible saber el número de no reportados y de no aplica de víctimas respecto al género.

- **Gráfica 5:** La gráfica muestra los cinco municipios del país con menos delitos sexuales, siendo Cumaribo, ubicado en el departamento del Vichada, el municipio menos afectado por los delitos sexuales teniendo solo 4 en tiempos de pandemia. Puerto Rondón (Arauca), La primavera (Vichada), Distracción (La Guajira) y Urumita (La Guajira) también hacen parte de los municipios con menos delitos sexuales durante los tiempos del apogeo del COVID-19.

**Resultados positivos:** El histograma es claro y se puede reconocer facilmente cuantos delitos sexuales se realizaron en cada uno de los cinco municipios en tiempo de pandemia.

**Resultados negativos:** La gráfica no especifica el intervalo de tiempo que estamos analizando, esto puede ocasionar confusión al momento de relacionar la gráfica y analizarla.

## CONCLUSIONES

Con los graficos anteriores podemos obtener que los lugares mas afectados por estos delitos son grandes ciudades, en las que tambien se encuentra el departamento de la capital. Los que menos obtuvieron estos delitos son los departamentos de La guajira, Arauca y Vichada, sabemos que estos departamentos han sido abandonados por el estado lo que nos puede llevar a la conclusion rapida de que la razon de que no tengan un gran numero de estos sea por la falta de presencia policial lo cual puede entorpecer el proceso de reporte de los mismos, pero es algo que con la informacion obtenida no podemos concluir. Ademas de esto el genero Femenino es el mas vulnerado por estos delitos, con los menores de 14.

Con esta informacion podremos sabemos cuales son los grupos mas vulnerables en nuestro pais y sus posibles ubicaciones. Esto no puede ayudar a la policia nacional a poder tener un mejor control sobre el tema y ayudar a las posibles victimas.

Tambien nos guia a nosotros como sociedad de ser mas consientes de que personas se encuentran en una probabilidad mas alta de poder sufrir alguno de estos delitos sexuales.

Agregado a esto, la base de datos que usamos nos proporciono muy buena informacion para poder llegar a estas soluciones, pero en algunos casos fue redundante o poco clara a quienes se referia, por esto recomendamos para proximas actualizaciones tener en cuenta estos detalles que pueden complicar el analisis. Asimismo, pedimos especificar, agregar, ya sea en otra base de datos o en la misma; estos factores de presencia policial en la zona, para asi poder comprender mas la situacion de seguridad en la region.

## REPOSITORIO GITHUB

Aquí se encuentra la carga de datos, la conexión y los archivos en formato csv modificados. **Repositorio.**

A continuacion agregaremos los perfiles de Github de cada uno de los integrantes.

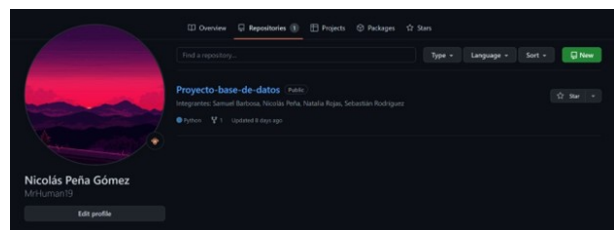


Figura 10: Perfil de github de Nicolás Peña Gómez.

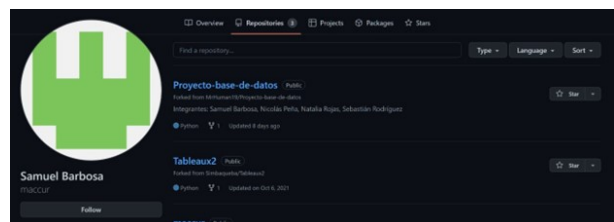


Figura 11: Perfil de github de Samuel Barbosa

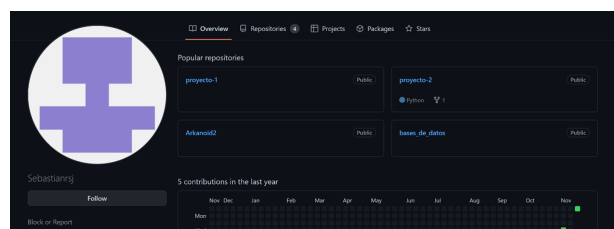


Figura 12: Perfil de github de Sebastián Rodríguez

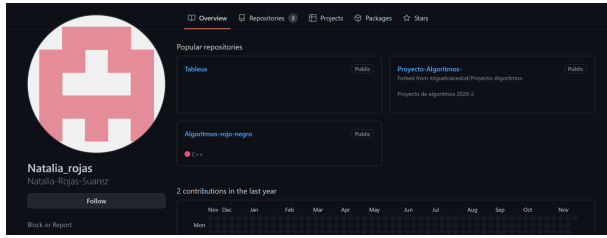


Figura 13: Perfil de github de Natalia Rojas

## REFERENCIAS

- 1 Infobae,(2019). *La extorsión sexual en América Latina y el Caribe (GLOBAL CORRUPTION BAROMETER – LATIN AMERICA THE CARIBBEAN 2019)*, Imagen. Infobae, [Link URL](#).
- 2 OMS, (2019). *Prevalencia a la violencia*

*intima*,Imagen. DW, [Link URL](#).

## BIBLIOGRAFIA

- 1 ENDS,(2015). *Encuesta Nacional de Demografía y Salud*, Texto. ENDS, [Link URL](#).
- 2 Perez Olmos, Isabel,(2001). *DICTAMENES SEXOLÓGICOS POR DELITO SEXUAL ANÁLISIS DE LA INFORMACIÓN FORENSE COLOMBIA 2001*, Texto. INML y CF, [Link URL](#).
- 3 Diaz Gomez,Andres; Pardo Lluch,Maria Jose, (2017). *Delitos sexuales y menores de edad: Una aproximación basada en las personas privadas de libertad en la isla de Gran Canaria*, Texto. Revista Electrónica de Ciencia Penal y Criminología, [Link URL](#).