



Primer proyecto BDnR: Estrategia jurídica contra la violencia de género

Equipo 1

Integrantes:

Ana Paola Carmona Méndez

César André San Luis Martínez

Karol Josafat Cisneros Suárez

Natalia Wilson Robles

Clase:

Base de datos no relaciones

Profesor:

Felipe López Gamino

Contenido

Introducción.....	3
Problemas encontrados.....	3
Solución	4
Características de la solución	5
Obtención y almacenamiento de los datos	6
Estructura de las "tablas" de la Base de Datos.....	6
Resultados	8
 Señales de alerta.....	9
 Impacto de la propuesta.....	10
 Concientización de la violencia.....	11
Conclusión.....	12
Bibliografía	13

Introducción

La violencia contra las mujeres es una problemática social urgente que requiere ser entendida. En este proyecto buscamos analizar información pública sobre atenciones a mujeres víctimas de violencia para identificar tendencias, tipos de agresión, distribución geográfica, entre otras cosas para comprender la magnitud del problema.

Para manejar y analizar un volumen amplio y diverso de información, se utilizó una base de datos NoSQL (MongoDB), es así como mediante consultas y gráficas, los datos se transformaron en evidencia clara y accesible, útil para la prevención, la protección y la urgencia de reformar las políticas públicas de nuestro país.

Problemas encontrados

Durante el desarrollo del proyecto se presentaron varios retos técnicos y de gestión de datos:

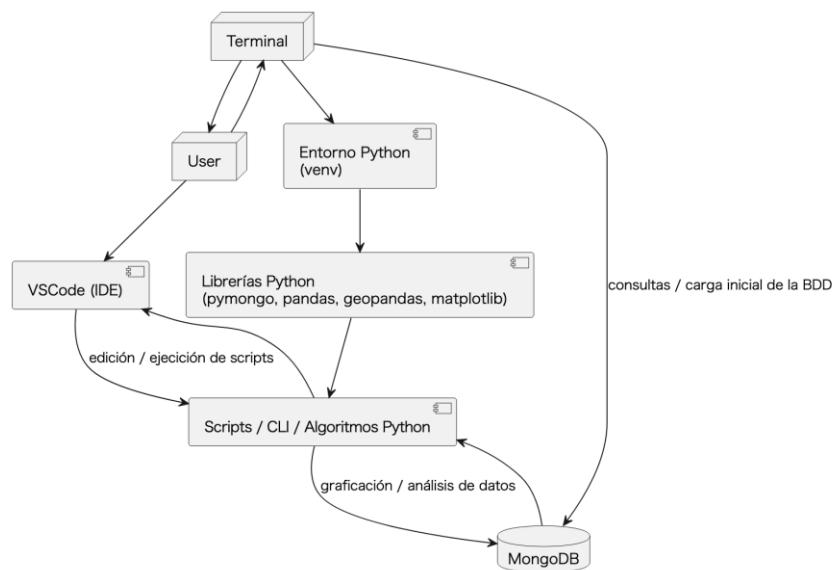
1. **Disponibilidad y formato de las bases de Datos:** Uno de los principales desafíos fue encontrar bases de datos públicas. Debido a que muchos portales de datos abiertos presentan información desactualizada, falsa o en formatos difíciles de procesar. Para resolverlo, se seleccionó el conjunto de datos de *atenciones a mujeres víctimas de violencia* por su nivel de detalle, confiabilidad y estructura compatible con MongoDB.
2. **Conexión entre MongoDB y Python:** Configurar la conexión entre ambas herramientas requirió instalar controladores específicos (pymongo) y ajustar las rutas locales para que la base se comunicara correctamente con el entorno de desarrollo, en este caso trabajamos con VScode.
3. **Tipos de datos y consistencia de la información:** Los datos contenían inconsistencias en los tipos de variables (fechas en formato texto, campos vacíos o mal registrados), al igual que se presentaban bastantes NA. Esto se solucionó mediante transformaciones y validaciones en Python antes y después de insertarlos en MongoDB, garantizando una lectura uniforme para las consultas y gráficas. Comprobando que todo se ejecutara a la perfección.
4. **Cambio de enfoque técnico:** En cursos anteriores habíamos trabajado con bases de datos relacionales (SQL), pero en este proyecto fue necesario adaptarnos al enfoque

NoSQL, ya que el curso se centra en este tipo de tecnología. Se utilizó MongoDB por su flexibilidad para manejar información no estructurada y de gran volumen. Aunque implicó un cambio importante, permitió realizar consultas más dinámicas y adecuadas al análisis social planteado.

Solución

El flujo de trabajo consistió en 4 pasos: **a)** Mediante una lluvia de ideas, se discutió entre varias bases de datos y se llegó a un acuerdo. **b)** Una vez seleccionada la BDD, esta se cargó a MongoDB mediante la terminal. **c)** Luego, se comenzó un análisis exploratorio de los datos, aquí surgieron varios debates internos respecto a qué hacer con ciertos tipos de datos (N/A y None) y cómo tomarlos en cuenta, fue durante esta etapa que surgieron las primeras ideas para consultas y gráficas. **d)** Con las consultas elegidas, cada miembro del equipo se dedicó a su implementación en Python, así como la generación de gráficas y tablas mediante librerías como Pandas, Matplotlib y Geopandas.

Se necesitaba almacenar datos en una base de datos no relacional (MongoDB), ejecutar varios algoritmos en Python para procesar esos datos y generar gráficas desde un entorno de desarrollo (VSCode) y/o la terminal. La solución implementada se puede representar mediante el diagrama presentado a continuación.



Los componentes de la solución son:

- Terminal: desde donde se cargó la BDD a MongoDB y se crearon los ambientes virtuales (pyEnv) que nos permitieron profundizar el análisis de los datos mediante gráficas.
- VSCode: IDE que nos permitió editar y hacer prueba y error de los scripts de Python que utilizamos de una forma más rápida y eficiente.
- PyEnv y Librerías: Se instalaron librerías como pymongo, geoPandas, Pandas, Matplotlib y Polars dentro del ambiente virtual como buena práctica para evitar interferencia con otros proyectos / sistemas de otros miembros de equipo. Estas librerías nos permitieron hacer scripts en Python que posteriormente fueron utilizados para la generación de gráficas y el análisis de datos.

- MongoDB: el manejador de BBDD que utilizamos para la elaboración del proyecto, lo elegimos por su rapidez, facilidad de uso, flexibilidad y familiaridad entre los miembros del equipo.

Características de la solución

Las principales funcionalidades de la solución propuesta incluyen:

- La visualización geográfica (por entidad federativa) de la frecuencia de reportes de violencia de género.
- La clasificación por tipo de relación con el agresor y rangos de edad de las llamadas registradas en la base de datos.
- El registro en una línea temporal de la frecuencia de las llamadas.
- La visualización por tipo de violencia manifestada en los reportes.

El equipo consideró que estas funcionalidades son claves para el propósito del proyecto, pues es gracias a ellas que podemos utilizar los datos presentados para entender y dimensionar el impacto y frecuencia de estas problemáticas en la sociedad mexicana.

Para hacer uso de estas funcionalidades, todos los miembros del equipo tuvieron que seguir un proceso de configuración (setup), de forma que los scripts, librerías y archivos funcionasen de la forma deseada en cualquier ambiente y sistema operativo. Este proceso consiste en:

1. Una correcta instalación y funcionamiento de MongoDB y Python 3.
2. Una correcta carga de la BDD utilizando la terminal o Compass.
3. La creación de un espacio de trabajo en una carpeta específica en la computadora de cada miembro.
4. La creación de un ambiente virtual usando ‘venv’ dentro de la carpeta del espacio de trabajo. (`python3 -m venv venv_name`)
5. La instalación de las librerías necesarias para la ejecución de todos los scripts creados por los demás miembros del equipo. (`pip install ‘librería’`)

El seguimiento de estas instrucciones fue crucial para el trabajo colaborativo y el correcto funcionamiento de los scripts creados, minimizando el tiempo perdido en la solución de problemas relacionados con diferencias en sistema operativo, versión de librerías, etc.

Obtención y almacenamiento de los datos

La obtención de los datos constituyó una etapa fundamental del proyecto, ya que era necesario contar con información en un formato adecuado para la realización de las consultas y análisis requeridos. Dado que el tema que tratamos, la violencia de género en nuestro país, implica el manejo de información sensible, resultó esencial garantizar la fiabilidad, integridad y tratamiento ético de los datos.

Por ello, se optó por utilizar la base de datos pública proporcionada por la Secretaría de las Mujeres de la Ciudad de México, correspondiente al programa *Abogadas de las Mujeres*. Esta fuente oficial asegura la validez de la información y el cumplimiento de los estándares éticos necesarios para su uso académico.

Para su procesamiento, los datos fueron descargados en formato CSV desde la fuente oficial.

El almacenamiento en MongoDB permitió trabajar con una estructura flexible y no relacional, facilitando la ejecución de consultas complejas (por ejemplo, filtrado por tipo de violencia o alcaldía) sin necesidad de esquemas rígidos.

Esta estructura también permitió agregar información contextual durante el análisis, como indicadores por zona o clasificaciones adicionales, sin modificar la base principal.

Estructura de las "tablas" de la Base de Datos

La base de datos se almacenó en MongoDB, donde la información se organiza en colecciones en lugar de tablas.

Cada documento dentro de la colección representa un caso individual de atención brindado por el programa *Abogadas de las Mujeres* de la Secretaría de las Mujeres de la Ciudad de México.

- Colección principal:

Cada documento almacena información sobre la usuaria, el tipo de violencia reportada, la relación con el agresor y las características del evento.

A continuación, se describen los principales campos y su tipo de dato:

Campo	Tipo de dato	Descripción
_id	ObjectId	Identificador único generado automáticamente por MongoDB.
fecha	Date	Fecha en la que se registró la atención o denuncia.
agencia	String	Unidad o coordinación territorial donde se brindó el servicio.
usuaria_sexo	String	Sexo de la persona atendida (en todos los casos “MUJER”).
usuaria_nacionalidad	String	Nacionalidad de la usuaria.
usuaria_ent_fed_nacimiento	String	Entidad federativa de nacimiento.
usuaria_estado_civil	String	Estado civil reportado.
usuaria_edad	Entero	Edad de la usuaria al momento del registro.
usuaria_hijas, usuaria_hijos, usuaria_hijes	Entero	Número de hijas, hijos o hijes declarados.
relacion_con_agresora	String	Vínculo con la persona agresora (por ejemplo, “CÓNYUGE”).
agresora_ocupacion	String	Ocupación registrada del agresor/a.
ci_iniciada	String	Indica si se inició o no una carpeta de investigación.
delito	String	Tipo de delito identificado, en caso de existir.

representacion	String	Indica si se proporcionó representación jurídica.
violencia_modalidad	String	Modalidad del contexto de violencia (familiar, comunitaria, institucional, etc.).
solicitud_medidas	String	Si se solicitaron medidas de protección.
alto_riesgo	String	Indica si el caso fue clasificado como de alto riesgo.
riesgo_feminicida	String	Indica si se identificó riesgo de feminicidio.
vio_psicoemocional, vio_fisica, vio_economica, vio_patrimonial, vio_sexual, vio_simbolica, vio_obstetrica, vio_reproductivos, vio_feminicida	String	Variables binarias ("Sí" / "No") que especifican el tipo de violencia reportada.

Esta estructura de base de datos permitió almacenar información diversa y sensible de manera flexible, sin la necesidad de un esquema rígido. El uso de MongoDB facilitó la gestión de campos opcionales o con valores faltantes, garantizando al mismo tiempo la anonimización de los datos personales y el cumplimiento de principios éticos en el manejo de información sensible.

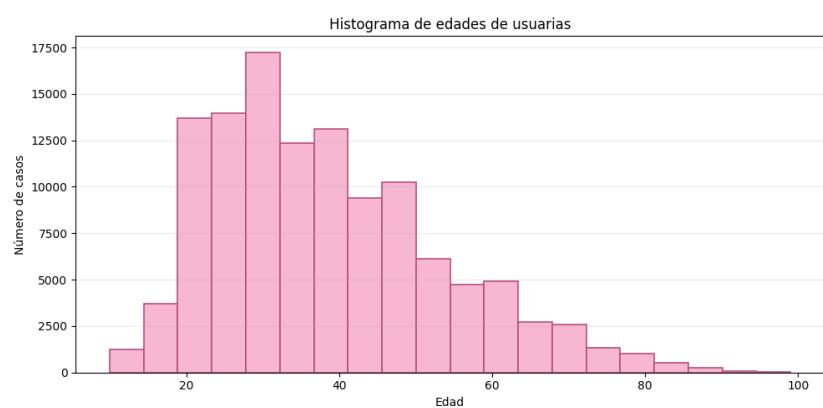
Resultados

Como se ha mencionado anteriormente, la elección de una base de datos de índole social y con información valiosa con respecto a la realidad de la mujer mexicana, nos permite llegar a resultados interesantes sobre los focos de violencia, la respuesta que ha tenido esta propuesta gubernamental por la sociedad y la importancia de la concientización a través del análisis de los datos. Asimismo, reconocemos la responsabilidad que recae en nuestras manos no sólo en cuanto a ser fieles a la información, sino también a visualizar de la forma más clara y accesible los resultados.

A continuación, planteamos una interpretación de los resultados obtenidos, dejando en claro que la realidad de las mujeres en nuestro país es tan diversa y compleja que no se puede encerrar en simples gráficas, sin embargo, nos invitan a reflexionar y a formular soluciones.

Señales de alerta

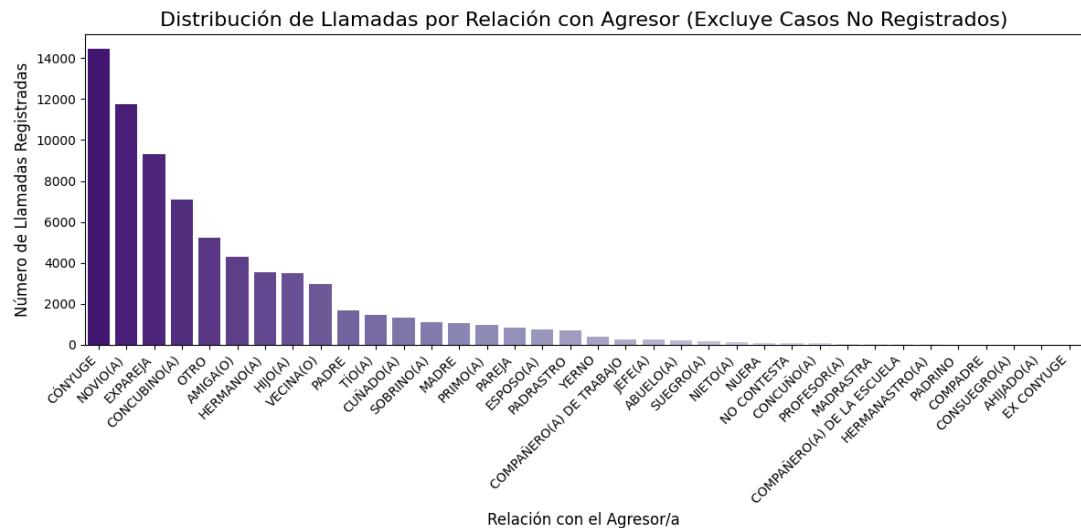
Primeramente, un aspecto importante que orientó nuestros objetivos fue el buscar aquellas señales de alerta que rodean el contexto de la mujer en nuestro país. Por lo tanto, nos pareció significativo enfocarnos tanto en la edad de la víctima como en la relación de ésta con la persona agresora.



La gráfica “Histograma de edades de usuarias” nos muestra la distribución de llamadas por rango de edades. Podemos ver que el grupo con mayor número de llamadas son las mujeres de entre 20 a 35 años de edad. Esto se

puede deber a varios aspectos, tales como ser el periodo de vida con primer acercamiento a relaciones formales y convivencia social (lo que se traduce en posibles situaciones de riesgo), ser aún dependientes económicamente, inicio de vida sexual o ser el grupo con mayor acceso a la tecnología y a la información, lo que facilita la denuncia de abusos. A pesar de que las llamadas se concentren en estos apartados, es de gran importancia resaltar que en todos los grupos de edad sigue habiendo registros, desde la niñez hasta la vejez, donde no necesariamente las mujeres en estos grupos sufren de menos violencia, sino que tienen menos acceso a los canales de demanda por diferentes razones.

Ahora, la gráfica “Distribución de llamadas por relación con agresor” muestra otra señal de alerta: desde donde proviene el abuso. Para facilitar la visualización de la información decidimos quitar el apartado de relaciones no registrada, que era el grupo con mayor número de llamadas (lo que se puede deber a la privacidad de la usuaria). Al encabezar la lista las relaciones matrimoniales, de noviazgo y expareja, podríamos intuir que gran parte de la violencia surge en entornos donde hay una relación íntima con el agresor (que involucra tanto



intimidad sexual como emocional). De igual forma, analizando otros apartados notables, podemos ver cómo la familia es una exponencial cuna de abuso.

Impacto de la propuesta

La propuesta “Atención de abogadas de las mujeres”, surgida en 2019 para orientar jurídicamente a mujeres víctimas de violencia de género, ha tenido alcance a nivel nacional, lo que permite un análisis más complejo en cuanto a los diferentes contextos de nuestro país. El siguiente mapa refleja cómo ha sido el uso de este recurso por entidad.

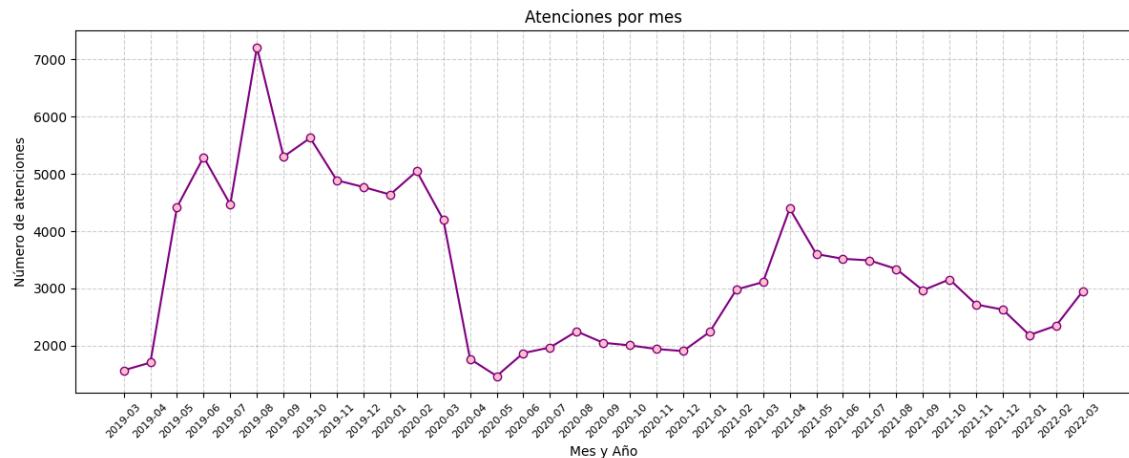


Es visible la concentración de llamadas en el centro-sur del país y la marginalización de este recurso en el norte. Este resultado puede ser reflejo de la notable centralización de recursos en nuestro país con respecto a la capital, la Ciudad de México. De igual forma, aquí pueden contemplarse las distintas concepciones culturales en cuanto

a la violencia de género en las diferentes regiones del país.

Otro aspecto interesante que comunica bastante sobre el impacto de esta propuesta en nuestro país es la distribución de las llamadas a lo largo del periodo de 2019 a 2022. Se nota un alza en las llamadas al inicio de su implementación (lo que puede reflejar la necesidad que había de

una herramienta así); sin embargo, durante el segundo semestre de 2020 se nota una caída en los registros, lo que coincide con el confinamiento por la pandemia de COVID-19; por este motivo, la baja podría deberse a una constante vigilancia de la víctima por parte del agresor (juntos en casa), interrupción de los servicios públicos y falta de espacios seguros.



Concientización de la violencia

Por último, el formato de las llamadas logró registrar un aspecto valioso como lo es el tipo de violencia denunciada, así como el nivel de riesgo al que estuvo expuesta la víctima. Más allá de la subjetividad que estas categorías pueden representar para el espectador, nos resulta de

gran importancia la concientización sobre las distintas formas en las que se puede manifestar la violencia (todas dignas de atención) y qué tan peligrosas pueden llegar a ser.

--- CONSULTA: Conteo de Tipos de Violencia por Nivel de Riesgo ---			
	Bajo/Medio Riesgo	Alto Riesgo	Riesgo Feminicida
Tipo_Violencia			
vio_psicoemocional	51267	3618	1102
vio_fisica	35501	3771	1424
vio_patrimonial	4573	389	126
vio_sexual	14300	1131	179
vio_simbolica	132	16	5
vio_obstetrica	22	3	2
vio_reproductivos	80	15	5
vio_feminicida	143	93	199

La consulta realizada muestra cómo la violencia psicoemocional, física y sexual encabezan el número de llamadas en todos los niveles de riesgo, debido probablemente al mayor entendimiento que la sociedad en general tiene de cómo se representan estos abusos; no obstante, este tipo de análisis nos invita a concientizar sobre aquellos abusos “invisibles” al entendimiento del público en general, tal como la violencia obstétrica, reproductiva y patrimonial, que están al fondo de los registros.

Conclusión

El desarrollo de nuestro proyecto constituye un ejercicio fundamental en la aplicación de la ciencia de datos con un propósito social. Si bien hoy en día estamos acostumbrados a lucrar con los datos o usarlos con fines puramente económicos o de eficiencia, este trabajo demuestra que su valor trasciende en un potencial impactante, al dotar de contexto y voz a una problemática humana que a menudo se vive en silencio.

La riqueza de esta información nos ofrece un esbozo íntegro de una realidad sumamente compleja. Sin embargo, nuestro objetivo central, al abordar un tema tan sensible como lo es la realidad de la mujer mexicana, no es la obtención de respuestas definitivas, pues reducir un fenómeno multifacético a simples cifras y gráficos sería una grave e irresponsable simplificación. Por el contrario, buscamos operar como un promotor para la acción social, al generar preguntas y cuestionamientos que incomoden e inviten a reflexionar sobre cómo actuamos ante una problemática de la que todos somos parte. Los datos nos obligan a preguntarnos, por ejemplo, por qué la violencia usualmente se encuentra dentro de un núcleo tan íntimo como lo es la familia. De igual manera, cuestionan las reglas y protocolos enfocados en prevenir la violencia en los diferentes grupos etarios. Asimismo, nos hacen pensar en lo que les quedamos debiendo a las mujeres en un período de gravedad como lo fue la pandemia.

En esencia, este análisis se convierte en una confrontación a la sociedad con la realidad de las mujeres en nuestro país y las múltiples formas en que la violencia se puede manifestar. Su mayor utilidad es iluminar estas grietas sociales, recordándonos la responsabilidad ética que nuestra profesión nos da en la búsqueda de una sociedad más justa.

Bibliografía

- Matplotlib Development Team. (2024). Pyplot summary. Matplotlib Documentation. Recuperado de https://matplotlib.org/stable/api/pyplot_summary.html
- MongoDB Inc. (2024). Connect to MongoDB using MongoClient. MongoDB Documentation. Recuperado de <https://www.mongodb.com/docs/drivers/node/current/connect/mongoclient/>
- ONU Mujeres. (2021, noviembre). La pandemia de COVID-19 y la violencia contra la mujer: qué nos revelan los datos. Recuperado de <https://www.unwomen.org/es/noticias/reportaje/2021/11/la-pandemia-de-covid-19-y-la-violencia-contra-la-mujer-que-nos-revelan-los-datos>
- Pandas Development Team. (2024). 10 minutes to pandas. Pandas Documentation. Recuperado de https://pandas.pydata.org/docs/user_guide/10min.html#min
- Secretariado Ejecutivo del Sistema Nacional de Seguridad Pública - Centro Nacional de Información (SESNP-CNI). (2020). Información sobre violencia contra las mujeres: incidencia delictiva y llamadas de emergencia 9-1-1. México.
- Secretaría de las Mujeres de la Ciudad de México. (2019). Atención de abogadas de las mujeres. Portal de Datos Abiertos de la CDMX. Recuperado de <https://datos.cdmx.gob.mx/dataset/atenciones-abogadas-mujeres>
- Strotgen. (n.d.). mexico-leaflet/states.geojson at master. Recuperado de strotgen/mexico-leaflet. GitHub. <https://github.com/strotgen/mexico-leaflet/blob/master/states.geojson>