

UNIVERSIDAD DEL VALLE DE GUATEMALA  
Facultad de Ingeniería



**Propuesta de canalización de datos ETL de Violencia  
Intrafamiliar para almacenamiento en una base de datos  
relacional y análisis para extracción de información de  
indicadores de violencia contra la mujer**

Trabajo de graduación presentado por Andrea Carolina Argüello  
Barillas para optar al grado académico de Licenciado en Ingeniería en  
Ciencias de los Datos

Guatemala

2023







UNIVERSIDAD DEL VALLE DE GUATEMALA  
Facultad de Ingeniería



**Propuesta de canalización de datos ETL de Violencia  
Intrafamiliar para almacenamiento en una base de datos  
relacional y análisis para extracción de información de  
indicadores de violencia contra la mujer**

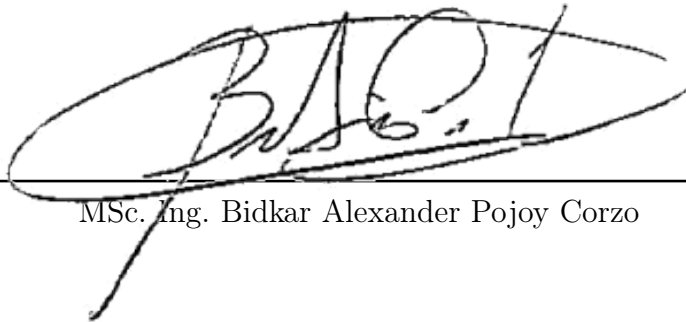
Trabajo de graduación presentado por Andrea Carolina Argüello  
Barillas para optar al grado académico de Licenciado en Ingeniería en  
Ciencias de los Datos

Guatemala,

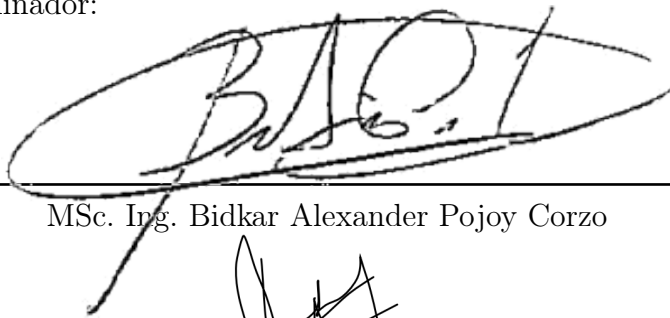
2023



Vo.Bo.:

(f)   
MSc. Ing. Bidkar Alexander Pojoy Corzo

Tribunal Examinador:

(f)   
MSc. Ing. Bidkar Alexander Pojoy Corzo

(f)   
MSc. Ing. José Antonio Medrano García

(f)   
MSc. Ing. José Vicente Herrera Juárez

Fecha de aprobación: Guatemala, 12 de enero de 2023.



Empiezo agradeciendo al MSc. Ing. Bidkar Alexander Pojoy Corzo, ya que la existencia del presente trabajo no hubiese sido posible sin su apoyo.

Agradezco al MSc. Ing. José Antonio Medrano García, por su ayuda a lo largo de la carrera, y a los catedráticos que me instruyeron, en la misma.

Agradezco a mis amistades y a mis padres, por acompañarme a navegar lo incierto, y aligerar la carga en los momentos de tormenta.

Finalmente, doy gracias a mi abuela, María Olimpia; quien a pesar de ya no estar, sigue siendo mi razón para seguir adelante.



<b>Prefacio</b>	<b>V</b>
<b>Lista de figuras</b>	<b>IX</b>
<b>Lista de cuadros</b>	<b>XI</b>
<b>Resumen</b>	<b>XIII</b>
<b>Abstract</b>	<b>XV</b>
<b>1. Introducción</b>	<b>1</b>
<b>2. Antecedentes</b>	<b>3</b>
2.1. Gobierno electrónico, <i>big data</i> y datos abiertos . . . . .	3
2.1.1. Comisión Presidencial de Gobierno Abierto y Electrónico . . . . .	4
2.1.2. Barómetro de Datos Abiertos . . . . .	4
2.2. Instituto Nacional de Estadística . . . . .	5
2.3. Violencia intrafamiliar y violencia contra la mujer . . . . .	6
<b>3. Justificación</b>	<b>9</b>
<b>4. Objetivos</b>	<b>11</b>
4.1. Objetivo general . . . . .	11
4.2. Objetivos específicos . . . . .	11
<b>5. Alcance</b>	<b>13</b>
<b>6. Marco teórico</b>	<b>15</b>
6.1. Gobierno electrónico . . . . .	15
6.2. Gobierno digital . . . . .	15
6.3. Ingeniería de datos . . . . .	15
6.3.1. Modelado de datos . . . . .	15
6.3.2. Datos abiertos . . . . .	16
6.3.3. Gobernanza de datos . . . . .	16

6.3.4.	Tipos de datos . . . . .	17
6.3.5.	Bases de datos . . . . .	17
6.4.	Análisis de datos . . . . .	18
6.4.1.	Visualización de datos . . . . .	19
6.5.	Canalización de datos . . . . .	20
6.5.1.	Procesamiento por lotes . . . . .	20
6.5.2.	ETL . . . . .	20
<b>7.</b>	<b>Marco metodológico</b>	<b>21</b>
7.1.	Cronograma de actividades . . . . .	21
7.2.	Modelo de base de datos preliminar . . . . .	21
7.3.	Canalización ETL . . . . .	22
7.3.1.	Limpieza de datos . . . . .	22
7.3.2.	Ingesta y transformación de datos . . . . .	22
7.4.	Análisis y visualización de datos . . . . .	22
7.4.1.	Definición de indicadores de rendimiento . . . . .	22
7.4.2.	Prototipo de tablero . . . . .	23
<b>8.</b>	<b>Resultados</b>	<b>25</b>
8.1.	Canalización ETL . . . . .	25
8.1.1.	Ingesta de datos . . . . .	25
8.1.2.	Limpieza de datos . . . . .	26
8.1.3.	Transformación de datos . . . . .	27
8.2.	Análisis y visualización de datos . . . . .	27
8.2.1.	Definición de indicadores de rendimiento . . . . .	27
8.2.2.	Prototipo de tablero . . . . .	33
8.2.3.	Presentación de resultados al INE . . . . .	37
<b>9.</b>	<b>Conclusiones</b>	<b>39</b>
<b>10.</b>	<b>Recomendaciones</b>	<b>41</b>
10.1.	Al INE . . . . .	41
10.2.	Para futuras iteraciones del proyecto . . . . .	41
<b>11.</b>	<b>Bibliografía</b>	<b>43</b>
<b>12.</b>	<b>Anexos</b>	<b>47</b>
12.1.	Anexo técnico . . . . .	82
12.2.	Códigos . . . . .	82
<b>13.</b>	<b>Glosario</b>	<b>85</b>

---

## Lista de figuras

---

1.	Cronograma de actividades . . . . .	21
2.	Diagrama entidad relación del modelo de datos preliminar. . . . .	22
3.	Diagrama de arquitectura y ETL preliminar. . . . .	23
4.	Captura de pantalla de aplicación en Shiny para carga de archivos. . . . .	25
5.	Diagrama entidad relación final de datos, generado con DBeaver. . . . .	28
6.	Captura de pestaña de indicadores de violencia intrafamiliar por datos de registro . . . . .	34
7.	Captura de pestaña de indicadores de violencia intrafamiliar por datos de la víctima . . . . .	34
8.	Captura de pestaña de indicadores de violencia intrafamiliar por datos del agresor . . . . .	35
9.	Captura de pestaña de indicadores de violencia intrafamiliar por datos de hecho	35
10.	Captura de pestaña de insights para el año 2021 . . . . .	36
11.	Captura de pestaña de indicadores por municipio . . . . .	36
12.	Captura de pestaña de indicadores por departamento . . . . .	37
13.	Reunión de aclaración de dudas con Karin Barrios, Patricia Zavaleta y Anabela de la Cruz . . . . .	47
14.	Minuta de primera reunión, información personal redactada. . . . .	48
15.	Modelo de datos preliminar. . . . .	81
16.	Tamaño de cada tabla en la base de datos . . . . .	82



---

## Lista de cuadros

---

1.	Comparación de características entre gestores de bases de datos. . . . .	19
2.	Comparación de características entre Power BI y Tableau. . . . .	20
3.	Indicadores de violencia contra la mujer presentados en el tablero . . . . .	29
4.	Indicadores de violencia intrafamiliar presentados en el tablero . . . . .	30
5.	Diccionario de variables de datos VIF . . . . .	48



El presente trabajo muestra el proceso de una canalización de datos ETL, sobre los datos públicos de Violencia Intrafamiliar provistos por el Instituto Nacional de Estadística (INE) de 2008 a 2021, con el objetivo de almacenar dicha información en una base de datos relacional para la extracción de información de indicadores de violencia contra la mujer. Esto, bajo el contexto de la “Política Nacional de Datos Abiertos 2018-2022” (2018), la cual menciona que el gobierno trabajará para “crear o respaldar programas e iniciativas que promuevan el desarrollo o creación conjunta de bases de datos, visualizaciones, aplicaciones y otras herramientas basadas en datos abiertos”.

Este proyecto fue realizado en un plazo menor a seis meses, utilizando tecnologías gratuitas y de fuente abierta, tales como el gestor de bases de datos relacionales PostgreSQL, el lenguaje de programación para análisis estadístico R y el software para análisis de datos Power BI. Entre los principales resultados cabe mencionar que es posible automatizar, en su mayoría, las reglas de limpieza, viéndose limitada en casos o errores que evidencian la ausencia de una gobernanza de datos o gestión de datos maestros constante. Por esto, se recomienda al INE la implementación de una arquitectura de datos para agilizar la producción de reportes y tableros periódicamente.



The present work shows the process of an ETL data pipeline, using the public data of Intrafamily Violence provided by the National Institute of Statistics (INE) from 2008 through 2021, with the objective of storing said information in a relational database for the extraction of information on indicators of violence against women. This, under the context of the “National Open Data Policy 2018-2022” (2018), which mentions that the government will work to “create or support programs and initiatives that promote the development or joint creation of databases, visualizations, applications and other tools based on open data”.

This project was carried out in less than six months, using free and open source technologies, such as the PostgreSQL relational database manager, the R statistical analysis programming language, and the Power BI data analysis software. Among the main results, it is worth mentioning that it is possible to automate, for the most part, the cleaning rules, being limited in cases or errors that show the absence of constant data governance or master data management. For this reason, the INE is recommended to implement a data architecture to speed up the production of periodic reports and dashboards.

En Guatemala, según lo indicado en el artículo 20 del Decreto Número 22-2008, Ley contra el femicidio y otras formas de violencia contra la mujer, (2008), el INE está obligado a generar indicadores e información sobre violencia contra la mujer. Por otro lado, el Decreto Número 97-96, Ley para prevenir, sancionar y erradicar la violencia intrafamiliar indica que la Procuraduría General de la Nación debe estimular “la investigación y recopilación de estadísticas e información pertinente sobre las causas, consecuencias y frecuencia de la violencia intrafamiliar, con el fin de evaluar las medidas estatales” (1996). Adicional a esto, el Decreto Número 5-2021 (2021), Ley para la simplificación de requisitos y trámites administrativos, destaca la necesidad de “modernizar la gestión administrativa por medio de la simplificación, agilización y digitalización de trámites administrativos, utilizando las tecnologías de la información y comunicación para facilitar la interacción entre personas individuales o jurídicas y dependencias del Estado”.

Bajo este contexto, el presente trabajo pretende definir una canalización de datos ETL utilizando el lenguaje de programación para análisis estadístico R sobre información histórica de violencia intrafamiliar, con destino a una base de datos relacional en el gestor PostgreSQL, para la entrada al *software* de análisis de datos Power BI para la visualización de los datos. Dichas tecnologías fueron seleccionadas debido a que son de fuente abierta y gratuitas, haciéndolas más accesibles que otras alternativas. Se concluyó que, si bien es posible realizar dicha canalización y automatizar la limpieza de datos, esta se ve limitada ya que actualmente existen problemas de gobernanza de datos o falta de gestión de datos maestros, debido a inconsistencias entre la información y el diccionario de variables de 2021.

Si bien el proyecto a presentar está planificado y contextualizado para bases de datos de violencia intrafamiliar, una metodología similar puede adaptarse a otras problemáticas de visualización similares, ajustando las necesidades específicas que puedan darse.



## 2.1. Gobierno electrónico, *big data* y datos abiertos

Para administrar el análisis en tiempo real de datos de transmisión de gran volumen, el gobierno de Estados Unidos e IBM colaboraron en 2002 para desarrollar una infraestructura en clúster escalable masivamente, cuyo resultado fueron IBM InfoSphere Stream e IBM Big Data, plataformas para el descubrimiento y la visualización de información de miles de fuentes en tiempo real. En 2012, la administración de Obama anunció la Iniciativa de Investigación y Desarrollo de Big Data, cuyos objetivos principales eran avanzar en tecnologías centrales de *big data* de última generación; acelerar el descubrimiento en ciencia e ingeniería; fortalecer la seguridad nacional; transformar la enseñanza y el aprendizaje y expandir la fuerza laboral necesaria para desarrollar y utilizar tecnologías de *big data*. El gobierno del Reino Unido fue uno de los primeros países de la UE en implementar programas de *big data*, estableciendo el Horizon Scanning Centre (HSC) en 2004 con el objetivo de mejorar la capacidad del gobierno para enfrentar desafíos interdepartamentales y multidisciplinarios. En Corea del Sur, la Iniciativa Big Data, lanzada en 2011 por el Consejo Presidencial sobre Estrategias Nacionales de Tecnologías de la información y la comunicación (TIC), tiene como objetivo establecer sistemas pangubernamentales de redes y análisis de *big data* y promover la convergencia de datos entre el gobierno y los sectores privados (Kim y col., 2014).

En la Declaración de Santo Domingo (2006), la Organización de los Estados Americanos (OEA) declaró que se debe solicitar a sus órganos, organismos y entidades “que continúen apoyando a los Estados Miembros en la incorporación de las TIC en sus planes nacionales de desarrollo, particularmente en los procesos de modernización de las instituciones públicas”. Adicionalmente, esta llama a “reconocer la importancia de la perspectiva de género y la necesidad de mejorar el acceso equitativo de la mujer a los beneficios de las TIC, y asegurar que éstas puedan convertirse en una herramienta fundamental para potenciar el papel de la mujer y promover la igualdad de género”.

En Guatemala, bajo el gobierno de Jimmy Morales, se formuló la “Política Nacional de Datos Abiertos 2018-2022” (2018) por el Ministerio de Gobernación y la Secretaría Nacional

de Ciencia y Tecnología. Este “reconoce que los datos deben ser presentados en formatos estructurados y estandarizados para apoyar la interoperabilidad, trazabilidad y reutilización efectiva”. Adicionalmente, el Decreto Número 5-2021 (2021), Ley para la simplificación de requisitos y trámites administrativos, “tiene por objeto modernizar la gestión administrativa por medio de la simplificación, agilización y digitalización de trámites administrativos, utilizando las tecnologías de la información y comunicación para facilitar la interacción entre personas individuales o jurídicas y dependencias del Estado”.

### **2.1.1. Comisión Presidencial de Gobierno Abierto y Electrónico**

A partir del Acuerdo Gubernativo número 346-2004 (2004), se creó la Comisión Presidencial para la Reforma, Modernización y Fortalecimiento del Estado y de sus Entidades Descentralizadas (COPRE), la cual fue suprimida por el gobierno de Álvaro Colom en 2008; uno de sus tres pilares fue el gobierno electrónico (Comisión Presidencial de Gestión Pública Abierta y Transparencia, 2019; *Gobierno Electrónico*, S.f.). El gobierno de Otto Pérez Molina acordó, según el Acuerdo Gubernativo 360-2012, la creación de forma temporal la Comisión Presidencial de Transparencia y Gobierno Electrónico (COPRET); esta tuvo dos reformas subsiguientes: la primera fue en el Acuerdo Gubernativo 41-2018, por el gobierno de Jimmy Morales, como la Comisión Presidencial de Gestión Pública Abierta y Transparencia, y la segunda fue en 2020 en el Acuerdo Gubernativo No. 45-2020, por Alejandro Giammattei, como la Comisión Presidencial de Gobierno Abierto y Electrónico. Entre sus actuales atribuciones se encuentra el “coadyuvar a la transformación de la gestión pública mediante la promoción de mecanismos y medidas que derivan de las políticas de gobierno en materia de innovación tecnológica y comunicación, participación ciudadana, rendición de cuentas y transparencia, en coordinación con los entes rectores según compete a cada materia” (Pérez Molina, 2020).

### **2.1.2. Barómetro de Datos Abiertos**

La Iniciativa Latinoamericana por los Datos Abiertos (ILDA) es parte de la iniciativa Open Data for Development (OD4D), como nodo regional para Latinoamérica. Surge en Montevideo, Uruguay en 2013, y su objetivo es promover las políticas y el uso de datos abiertos para favorecer el desarrollo inclusivo de la región. Tiene el apoyo del Centro de Investigaciones para Desarrollo de Canadá (International Development Research Centre-IDRC) (*ILDA*, S.f.).

El Barómetro de Datos Abiertos (Open Data Barometer-ODB) es un proyecto producido por la World Wide Web Foundation, con el apoyo de Omidyar Network, y su objetivo es “descubrir la verdadera prevalencia y el impacto de las iniciativas de datos abiertos en todo el mundo”. Este asigna una calificación basándose en tres subíndices, de tres componentes cada uno; estos son:

- Preparación: existencia de datos abiertos e intervenciones que respaldan el compromiso y la reutilización de datos abiertos. (1/3)
  - Gobierno

- Políticas (1/4)
  - Acción (1/4)
- Empresarios y negocios (1/4)
- Sociedad civil y ciudadana (1/4)
- Implementación: en la encuesta de expertos del Barómetro de datos abiertos, se le pide a los investigadores que completen una lista de verificación detallada para cada una de las 15 categorías de datos. Los tres subcomponentes se obtienen agrupando los conjuntos de datos en tres grupos, con base en un análisis cualitativo de las formas comunes en que se utilizan estas categorías de datos. (1/3)
  - Innovación (1/3)
  - Políticas sociales (1/3)
  - Responsabilidad (1/3)
- Impacto: se trataron publicaciones en línea, medios de comunicación y publicaciones académicas sobre los impactos de los datos abiertos como indicadores, y se pidió a los investigadores que califiquen el alcance del impacto en una escala de 0 a 10. (1/3)
  - Político (1/3)
  - Social (1/3)
  - Económico (1/3)

El resultado de cada subíndice es un puntaje de 0 a 100, y el resultado final del ODB es una calificación de 0 a 100, siendo esta el promedio de los tres subíndices (*Open Data Barometer - Leaders Edition: ODB Methodology - v1.0*, 2017).

La última evaluación global del ODB fue en 2016, mientras que en 2017-2018 se realizó una edición aplicada a países líderes, entre estos Guatemala (S.f.), quien obtuvo 24 y 26, respectivamente. En 2020, ILDA implementó una edición a nivel latinoamericano del Barómetro, el cual cubrió 24 países de América Latina y el Caribe en un período de 15 meses, de enero de 2019 a abril de 2020. Esta edición nuevamente incluyó a Guatemala, obteniendo un puntaje final de 31.06 de 100, con 44.17 en el subíndice de Preparación, 49 en el de Implementación y 0 en el de Impacto. Ninguno de los 15 conjuntos de datos proporcionan identificadores de datos para elementos clave en el conjunto de datos, mientras que solo 3 (presupuesto gubernamental detallado, datos detallados sobre el gasto público y contratos públicos) son legibles por computadora, reutilizables y están disponibles como un todo y están claramente identificados con una licencia abierta; por otro lado, no existen datos de propiedad de tierra ni de horarios de transporte público, mientras que si bien hay datos de registros de compañías, estas no se encuentran publicadas en línea por parte del gobierno en ninguna forma (ILDA, 2020; <https://barometrolac.org/>).

## 2.2. Instituto Nacional de Estadística

El Instituto Nacional de Estadística (INE) fue creado en 1985 por el Decreto Ley 3-85 (1985), y este tiene por objeto “formular y realizar la política estadística nacional, así como

planificar, dirigir, coordinar y supervisar las actividades del Sistema Estadístico Nacional”; además, entre sus funciones se encuentra el “establecer y mantener actualizado un inventario de las series estadísticas, producidas por las entidades y dependencias integrantes del Sistema Estadístico Nacional”. El Reglamento de Ley Orgánica del Instituto Nacional de Estadística además cita como objetivo del Sistema Estadístico Nacional “desarrollar, aplicar y mejorar los sistemas y técnicas adecuadas para el tratamiento de la información estadística”, y define actividad estadística como “el conjunto de procedimientos y métodos de recolección, procesamiento y divulgación de datos relativos a levantamiento de censos, estadísticas continuas, encuestas especiales, indicadores e índices en general, cuentas nacionales y demás programas macroestadísticos, así como el análisis e investigación estadística” (“Ley Orgánica del Instituto Nacional de Estadística y su Reglamento”, S.f.).

### 2.3. Violencia intrafamiliar y violencia contra la mujer

Uno de los Objetivos de Desarrollo Sostenible (ODS) propuestos por las Naciones Unidas (ONU) es “lograr la igualdad de género y empoderar a todas las mujeres y las niñas”, una de sus metas siendo el “eliminar todas las formas de violencia contra todas las mujeres y niñas en las esferas pública y privada, incluidas la trata y la explotación sexual y de otro tipo”. Cuentan con dos indicadores: la proporción de mujeres y niñas mayores de 15 años que “alguna vez han estado en pareja y que han sido objeto de violencia física, sexual o psicológica por parte de su pareja íntima actual o anterior en los últimos 12 meses, por forma de violencia y por edad”; y la proporción de mujeres y niñas mayores de 15 años “sometidas a violencia sexual por parte de personas que no sean su pareja íntima en los últimos 12 meses, por edad y lugar de ocurrencia” (*Goal 5: Achieve gender equality and empower all women and girls*, S.f.).

El artículo 11a de la Convención de Istanbul requiere la recolección de información para apoyar la prevención y lucha contra la violencia doméstica y violencia contra la mujer en Europa, indicando que los datos recopilados deben contener: el tipo de violencia; el sexo de la víctima y el sexo del agresor; la relación entre agresor y víctima; la edad de la víctima y la edad del victimario; y localización geográfica (Walby, 2016).

En Guatemala, se decretó en 1996 el Decreto Número 97-96, Ley para prevenir, sancionar y erradicar la violencia intrafamiliar. En su artículo 13, este nombra como ente asesor “encargado de las políticas públicas que impulsen la prevención, atención, sanción y erradicación de la violencia intrafamiliar” a la Procuraduría General de la Nación, el cual además debe estimular “la investigación y recopilación de estadísticas e información pertinente sobre las causas, consecuencias y frecuencia de la violencia intrafamiliar, con el fin de evaluar las medidas estatales”; esta además define en su primer artículo a la violencia intrafamiliar como “cualquier acción u omisión que de manera directa o indirecta causare daño o sufrimiento físico, sexual, psicológico o patrimonial, tanto en el ámbito público como en el privado, a persona integrante del grupo familiar, por parte de parientes o conviviente o exconviviente, cónyuge o excónyuge o con quien se haya procreado hijos o hijas” (“Decreto Número 97-96”, 1996). Bajo el gobierno de Alfonso Portillo, se emitió el Acuerdo Gubernativo No. 831-2000 (2000), Reglamento de la Ley para Prevenir, Sancionar y Erradicar la Violencia Intrafamiliar. A partir de este, se crea la Coordinadora Nacional para la Prevención de la Violencia

Intrafamiliar y en Contra de la Mujer (CONAPREVI), a la cual se le atribuye “estimular la investigación y recopilación de estadísticas e información pertinente sobre las causas, consecuencias, efectos y frecuencia de la violencia intrafamiliar y en contra de las mujeres, con el fin de evaluar e implementar las medidas estatales”. Además, entre sus integrantes del sector público se encuentra el Presidente de la Junta Directiva del Instituto Nacional de Estadística o su representante.

En 2007, la Comisión Económica para América Latina y el Caribe (ECLAC por sus siglas en inglés), en una reunión de expertos organizada por la ONU, propuso medir los siguientes aspectos de la violencia contra la mujer: tasa de violencia, tasa de violencia física, tasa de violencia psicológica, tasa de violencia sexual y tasa de violencia sin reportar. También recomendaron desagregar los datos para obtener indicadores más diferenciados por características como: área geográfica; condición de pobreza de los hogares; edad de la mujer; su estado de actividad; su nivel de educación; su estado de embarazo; su grupo étnico; y su relación con el agresor (AlmÉRas y col., 2001; *Indicators to measure violence against women*, 2007).

En 2008, se decreta el Decreto Número 22-2008, Ley contra el femicidio y otras formas de violencia contra la mujer. En su artículo 3, define la violencia contra la mujer como “toda acción u omisión basada en la pertenencia al sexo femenino que tenga como resultado el daño inmediato o ulterior, sufrimiento físico, sexual, económico o psicológico para la mujer, así como las amenazas de tales actos, la coacción o la privación arbitraria de la libertad, tanto si se produce en el ámbito público como en el ámbito privado.” En este mismo artículo además lista y define las siguientes como tipos de violencia contra la mujer:

1. económica: acciones/omisiones que repercuten en el uso, goce o acceso de una mujer a sus bienes materiales;
2. física: acciones de agresión que utiliza fuerza corporal, armas o sustancias para hacer daño, lastimar o enfermar a una mujer;
3. psicológica o emocional: acciones que producen daño o sufrimiento, psicológico o emocional, amenazas o violencia con objeto de intimidar, menoscabar el autoestima o controlar a la víctima; y
4. sexual: “acciones de violencia física o psicológica cuya finalidad es vulnerar la libertad e indemnidad sexual de la mujer” (“Decreto Número 22-2008”, 2008).

En su artículo 20, este indica que el INE está obligado a generar indicadores e información, y debe crear un Sistema Nacional de Información sobre Violencia contra la Mujer-SNIVM (“Decreto Número 22-2008”, 2008).



---

### Justificación

---

Con el nuevo milenio, varios gobiernos y países se comprometieron con los esfuerzos por adoptar nuevas tecnologías y con la transparencia de información. En el contexto de Guatemala, la “Política Nacional de Datos Abiertos 2018-2022” (2018) menciona que el gobierno trabajará para “crear o respaldar programas e iniciativas que promuevan el desarrollo o creación conjunta de bases de datos, visualizaciones, aplicaciones y otras herramientas basadas en datos abiertos”. Si bien el resultado del ODB fue mejor en 2020 que en años anteriores, Guatemala aún tiene espacios para mejora.

En los resultados publicados por ILDA, destaca el cero en el subíndice de Impacto, ya que varios Decretos refieren al INE u otra entidad a la recopilación y publicación de estadísticas e indicadores, parece que estos no se utilizan en la toma de decisiones. El INE publicó en el año 2022 un *dashboard* con indicadores para cada uno de los temas de los que recopila estadísticas. Sin embargo, la extracción de la información del mismo no está disponible de manera directa, y replicar el mismo se dificulta por el hecho de que no fue hasta 2019 que se publicó información en formato tipo Excel, siendo en años anteriores de tipo SPSS. Esto es concordante con lo presentado por ILDA, pues se indicó que el conjunto de datos de estadísticas de delitos no está en formatos legibles por computadora ni son reutilizables, y que este no se actualiza regularmente (algunos temas se actualizan de manera trimestral, mientras que otros se actualizan solo anualmente) y tampoco es fácil encontrar información sobre el mismo (<https://barometrolac.org/>).

IPAS, una ONG internacional, llevó a cabo una investigación primaria y un análisis de datos secundarios sobre violencia sexual en El Salvador, Guatemala, Honduras y Nicaragua entre 2009 y 2013; dicha iniciativa apoyada por UNFPA. Estudios realizados por el IPAS evidenciaron las “limitaciones en el reporte, análisis y difusión de datos sobre las diferentes formas de violencia sexual”, y abordaron la “necesidad de definir variables e indicadores consensuados a nivel nacional y regional para producir información y garantizar la comparabilidad de los datos provenientes de diferentes fuentes y sectores” (Fondo de Población de las Naciones Unidas (UNFPA), 2013). Dado que el INE publicó la Boleta de Violencia Intrafamiliar para el año 2021, puede asumirse que esta gestión aún se lleva en papel, por lo

que se puede asumir o esperar que el INE se vea sujeto al Decreto 5-2021; sin embargo, sería poco efectivo el digitalizar trámites sin tener una planificación de la estructura de datos a seguir, o sin tomar en cuenta que la nueva información se adecúe a información histórica, para no perder comparabilidad y dificultar el análisis. Citando a la Organización para la Cooperación y el Desarrollo Económicos (OECD), la falta de un modelo general de gobernanza de datos “puede conducir a la proliferación o duplicación de estándares de datos y soluciones técnicas para compartir datos, lo que dificulta la interoperabilidad de datos entre diferentes organizaciones y sectores, y afecta la posibilidad de integrar datos, procesos y organizaciones” (OECD, 2019)

Los gobiernos y entidades gubernamentales pueden reducir costos operativos haciendo uso de la nueva información que surge del *big data*, y esta puede ayudar a mejorar la toma de decisiones y aumentar la eficiencia y eficacia de la organización (Joseph y Johnson, 2013). La centralización de datos, además de permitir una automatización de los procesos de análisis y consecuentemente aumentar la eficiencia de organizaciones, permite que dicha información pueda utilizarse para modelos analíticos y algoritmos predictivos para evitar la reincidencia. Un estudio del uso de inteligencia artificial para la prevención de violencia contra la niñez cita que “el poder de *big data* para mejorar la precisión de la predicción del riesgo, combinando información sobre correlatos conocidos de violencia para producir índices de riesgo compuestos, es potencialmente mucho mayor que el de los humanos” (Daly y col., 2019; Hunt y col., 2020).

Las tecnologías a utilizar en el proyecto serán el gestor de bases de datos relacionales PostgreSQL, el lenguaje de programación para análisis estadístico R y el software para análisis de datos Power BI, dado que estas son tecnologías utilizadas actualmente por el INE y son aptas para la limpieza y transformación de datos, la gestión bases de datos relacionales y la obtención de indicadores estadísticos y visualizaciones de datos.

### 4.1. Objetivo general

Definir una metodología de extracción y transformación de datos utilizando el lenguaje de programación para análisis estadístico R sobre información histórica de violencia intrafamiliar con destino a una base de datos relacional en el gestor PostgreSQL, la cual servirá de entrada al *software* de análisis de datos Power BI para la toma de decisiones.

### 4.2. Objetivos específicos

- Definir un modelo de datos que contemple información necesaria relacionada con la violencia contra la mujer.
- Definir reglas de limpieza y criterios de calidad sobre el modelo de datos.
- Diseñar una canalización de datos para la migración de información pública de violencia intrafamiliar desde el origen de datos hasta una base de datos relacional en el gestor PostgreSQL utilizando el lenguaje de programación R.
- Definir y documentar indicadores clave de rendimiento (KPI por sus siglas en inglés) sobre violencia intrafamiliar elaborados sobre el modelo de datos implementado.
- Diseñar un prototipo de tablero en el *software* Power BI con indicadores de violencia contra la mujer relevantes a partir de la información transformada.



Este trabajo se enfocará en los datos recopilados anualmente por el INE en Guatemala sobre violencia intrafamiliar entre los años de 2008 a 2021, los disponibles al público, los cuales se encuentran almacenados en archivos de Excel y/o SPSS.

Se realizará un prototipo de integración completa de Inteligencia de Negocios (por sus siglas en inglés BI) que consta de modelo de datos y limpieza de los mismos. Se propondrá una estructura de datos relacional, con el objetivo de tener un control de calidad de la misma y para definir reglas de limpieza, la cual se realizará en R, al igual que la ingesta de datos; esto dado que la curva de aprendizaje del programa no representará una gran inversión en capacitaciones. Además, la implementación es más transparente y personalizable, y no requiere del uso de una interfaz gráfica.

El gestor de bases de datos a utilizar será PostgreSQL. Un comparativo de esta tecnología se encuentra en el marco teórico. Finalmente, el prototipo de tablero se realizará en Power BI, dado que esta es otra herramienta gratuita y soportada por Windows, y esta además puede conectarse directamente a diversas bases de datos.

El entregable final será el modelo de datos sugerido, los criterios y reglas de limpieza propuestas, la canalización programada en R y un tablero de Power BI accedendo a base de datos, presentando indicadores y visualizaciones a partir de la información transformada.



### 6.1. Gobierno electrónico

Según el Acuerdo Gubernativo número 346- 2004 (2004), el gobierno electrónico es la aplicación de las TIC “con el propósito de lograr eficacia y transparencia en la gestión del gobierno”, y sus objetivos son “la atención y participación ciudadana, buen gobierno en las gestiones internas institucionales, eficiencia, productividad y transparencia.”

### 6.2. Gobierno digital

El Plan de Gobierno Digital 2021-2026 (2021) cita que el gobierno digital se “refiere al uso de tecnologías digitales, como parte integral de las estrategias de modernización de los gobiernos para crear valor público”.

### 6.3. Ingeniería de datos

La ingeniería de datos comprende el recopilar, administrar y analizar grandes conjuntos de datos, con el fin de proporcionar los conjuntos de datos y las herramientas visuales correctos para el análisis por parte de científicos de datos. Involucra el desarrollo, mantenimiento, las pruebas y evaluaciones de soluciones de *big data* dentro de las organizaciones (*IBM Certified Data Engineer - Big Data*, S.f.).

#### 6.3.1. Modelado de datos

El modelado de datos es el proceso de crear una representación visual de un sistema de información completo o partes de él para comunicar conexiones entre puntos de datos y

estructuras. El objetivo es ilustrar los tipos de datos utilizados y almacenados dentro del sistema, las relaciones entre estos tipos de datos, las formas en que se pueden agrupar y organizar y sus formatos y atributos (IBM Cloud Education, 2020).

#### **6.3.1.1. *Data warehouse***

Un *data warehouse* es un tipo de sistema de gestión de datos diseñado para actividades de inteligencia de negocios (BI), especialmente análisis. Están destinados únicamente para realizar consultas y análisis. Los datos dentro del mismo generalmente se derivan de una amplia gama de fuentes, y a menudo son grandes cantidades de datos históricos (*What is a Data Warehouse?*, S.f.).

#### **6.3.1.2. *Star schema***

Un *star schema* o esquema de estrella es un tipo de esquema de bases de datos, un modelo dimensional implementado en sistemas de gestión de bases de datos relacionales, el cual consiste de una tabla central, llamada una tabla de “hechos”, que contiene las medidas numéricas del evento, rodeada por tablas de dimensiones con el contexto (*Esquema de estrella*, S.f. Kimball y Ross, 2013).

### **6.3.2. Datos abiertos**

Según la Política Nacional de Datos Abiertos (2018), estos son “datos digitales que son puestos a disposición del público con las características técnicas y jurídicas necesarias para que puedan ser usados, reutilizados y redistribuidos libremente por cualquier persona, en cualquier momento y en cualquier lugar” (Degenhart Asturias y col., 2018).

### **6.3.3. Gobernanza de datos**

El Comité de Sistemas de Seguridad Nacional (CNSS por sus siglas en inglés) de Estados Unidos define la gobernanza de datos como un conjunto de procesos que asegura el manejo formal de datos de una entidad o empresa, estableciendo parámetros de autoridad y toma de decisiones en relación a la información producida o manejada (CNSS, 2015).

En el sector público, la OECD destaca que dado el reciente uso indebido de datos por parte de empresas privadas, la mejora a la protección de los datos personales es una prioridad en agendas políticas. Definir un modelo general de gobernanza de datos puede evitar la proliferación o duplicación de estándares de datos y soluciones técnicas para compartir datos (OECD, 2019).

### 6.3.4. Tipos de datos

La CNSS (2015) define los datos como “información en una representación específica, generalmente como una secuencia de símbolos que tienen significado”.

#### 6.3.4.1. Estructurados

Los datos estructurados se encuentran altamente organizados, siguiendo un formato rígido, por lo que se dice que es fácil de descifrar y de organizar y buscar. Usualmente se maneja utilizando (Por sus siglas en inglés structured query language; lenguaje de consulta estructurada en español) lenguaje de programación desarrollado por IBM en 1974 (SQL) y utiliza datos cuantitativos (IBM Cloud Education, 2021a; Marr, 2019).

#### 6.3.4.2. No estructurados

Los datos no estructurados no pueden ser analizados por procesos y métodos convencionales, ya que no cuentan con un modelo de datos preestablecido o asignado. Usualmente son información cualitativa, como fotos, archivos de texto o audio, publicaciones en redes sociales, entre otros (IBM Cloud Education, 2021a; Marr, 2019).

#### 6.3.4.3. Semi estructurados

Los datos semi estructurados, como los datos no estructurados, tampoco tienen un modelo de datos predefinido y son más complejos que los datos estructurados, pero son más fáciles de almacenar que los datos no estructurados, ya que contienen características útiles para su clasificación (IBM Cloud Education, 2021a; Marr, 2019).

### 6.3.5. Bases de datos

Según el sitio de Oracle, una base de datos es “una colección organizada de información estructurada, o datos, normalmente almacenados electrónicamente en un sistema informático” (*What is a database?*, S.f.).

Es de particular interés las bases de datos relacionales, dado que estas son la manera más eficiente y flexible de acceder a datos estructurados. Esto es dado que elementos de una base de datos relacional se organizan como un conjunto de tablas con columnas y filas (*What is a database?*, S.f.).

#### 6.3.5.1. Sistema de gestión de bases de datos relacionales

Un sistema de gestión de bases de datos (por sus siglas en inglés DBMS) sirve como una interfaz entre la base de datos y los usuarios finales o programas. Permite recuperar, actualizar y administrar la organización y optimización de la información. Algunos de estos

son Microsoft SQL Server, Oracle Database, PostgreSQL y MySQL (*What is a database?*, S.f.).

### 6.3.5.2. Tipos de datos

La información en las columnas, en el caso de una base de datos relacional, representa información numérica o categórica. Algunos de los tipos de dato manejados son:

#### 1. Numéricos

- integer: números enteros
- float (punto flotante), double (doble precisión): números decimales

#### 2. Carácteres

- char: carácter

#### 3. Valores de fecha y hora

En particular, el lenguaje SQL maneja valores como char y varchar para definir cadenas de texto de tamaño fijo y variable con un máximo de caracteres, respectivamente, al igual que un tipo de dato de caracteres para identificar un Identificador único universal, un número de 16 bytes (UUID).

## 6.4. Análisis de datos

El análisis de datos puede definirse como la capacidad de derivar ciertas métricas o indicadores clave de rendimiento (KPI) de los datos (*What is Data Analytics?*, S.f.). Este análisis puede ser predictivo, prescriptivo (ayuda a determinar qué acción tomar), de diagnóstico (ayuda a explicar por qué ocurren ciertas cosas) o descriptivo (*What is Data Analytics?*, S.f.). Este último es en el que radica el reportaje, que es uno de los pilares de la inteligencia de negocios: responde al cómo, cuándo, cuántos y qué (*What is Data Analytics?*, S.f.). Un lenguaje utilizado para el análisis de datos es R, junto con la aplicación RStudio, dado que este es un lenguaje y entorno para computación estadística y gráficos.

Cuadro 1: Comparación de características entre gestores de bases de datos.

Característica	Oracle	PostgreSQL	MS SQL Server	MySQL
Permite filtros sobre funciones agregadas	No	Sí	No	No
Soporta PIVOT	Sí	No	Sí	No
Consultas en paralelo	Sí	Sí	Sí	No
Comparación utilizando expresiones regulares (RegEx)	Sí	Sí	No	Sí
Extraer parte de un valor <i>string</i> basado en RegEx ( <i>Substring</i> )	Sí	Sí	No	Sí
Reemplazar valores usando RegEx	Sí	Sí	No	Sí
Permite definir restricciones en llaves foráneas que se verifican solo en el momento de la confirmación	Sí	Sí	No	No
Permite definir verificaciones de restricciones ( <i>check constraints</i> )	Sí	Sí	Sí	Sí
Permite usar tipos de datos definidos por el usuario	Sí	Sí	No	No
Permite usar tipos de datos booleanos según el estándar SQL	No	Sí	No	No
Tipos de datos DATE (solo almacena fecha)	No	Sí	Sí	Sí
Tipos de datos para UUID	No	Sí	Sí	No
Permite restaurar enunciados	No	Sí	Sí	No
Lenguaje de definición de datos, subconjunto de SQL que describe datos y relaciones en una base de datos (DDL)				
Licencia gratuita para uso comercial	No	Sí	No	Sí

Fuente: Kellerer, S.f.

#### 6.4.1. Visualización de datos

La visualización de datos es la representación de datos mediante el uso de gráficos, tablas, diagramas, e infografías, con el objetivo de “comunicar relaciones de datos complejas y conocimientos basados en datos de una manera que es fácil de entender” (IBM Cloud Education, 2021b). El Harvard Business Review divide la visualización de datos en cuatro propósitos principales: generación de ideas, ilustración de ideas, descubrimiento visual y visualización de datos cotidianos.

Una herramienta para visualización de datos es Power BI Desktop es un programa desarrollado por Microsoft, su objetivo es ser “una solución de análisis empresarial basado en

la nube, que permite unir diferentes fuentes de datos, analizarlos y presentar un análisis de estos a través de informes y paneles” (Menendez, S.f.). Un comparativo de esta tecnología contra un *software* competidor, Tableau, se encuentra en el Cuadro 2.

Cuadro 2: Comparación de características entre Power BI y Tableau.

Característica	Power BI	Tableau
Sistemas operativos soportados	Windows	Windows, MacOS, Linux
Versión gratuita	Acceso a todas las funciones básicas de Power BI, conexión a múltiples fuentes de datos y pueden publicarse en la web	Permite conexión a Excel, CSV, y archivos JSON. todas las visualizaciones son públicas
Rango de precios de licencias individuales	Gratis o \$9.99 a \$20	\$35 a \$72
Licencias para empresas	Desde \$4995	\$12-\$15 por vista

Fuente: *Compare Tableau to Microsoft Power BI*, S.f. Edmond, 2021

## 6.5. Canalización de datos

IBM describe una canalización de datos o *data pipeline* como un método de ingesta de datos sin procesar para su análisis, luego de transferirlos a un almacén de datos (*What is a data pipeline?*, S.f.).

### 6.5.1. Procesamiento por lotes

El procesamiento por lotes, como el nombre lo implica, carga "lotes" de datos a un repositorio durante intervalos de tiempo establecidos; de acuerdo con IBM, “los trabajos de procesamiento por lotes forman un flujo de trabajo de comandos secuenciados, donde la salida de un comando se convierte en la entrada del siguiente comando” (*What is a data pipeline?*, S.f.).

### 6.5.2. ETL

Una canalización ETL (por sus siglas en inglés Extract, Transform, Load; extraer, transformar, cargar en español) sigue un orden específico, como su abreviatura lo indica. Las transformaciones son aplicadas hasta que los datos se han cargado al repositorio de datos (*What is a data pipeline?*, S.f.).

### 7.1. Cronograma de actividades

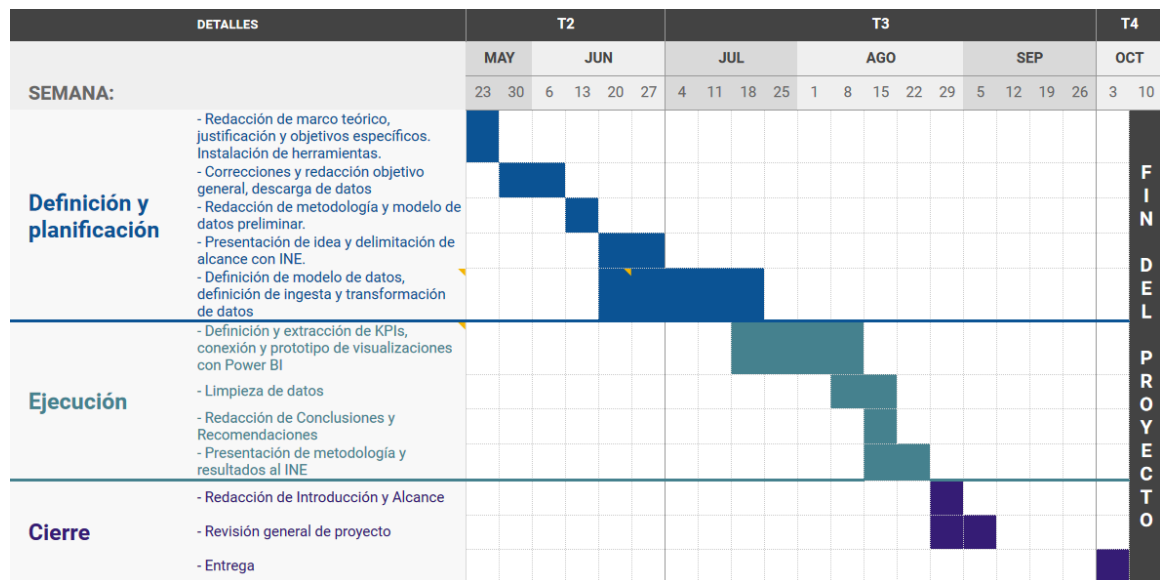


Figura 1: Cronograma de actividades

### 7.2. Modelo de base de datos preliminar

El diseño del modelo de bases de datos se realizará en el sitio [app.diagrams.net](http://app.diagrams.net). Se diseñará un modelo de base de datos en un esquema de estrella, tomando en consideración

las reglas de negocio del INE. El diagrama entidad relación preliminar se encuentra en la Figura 2.

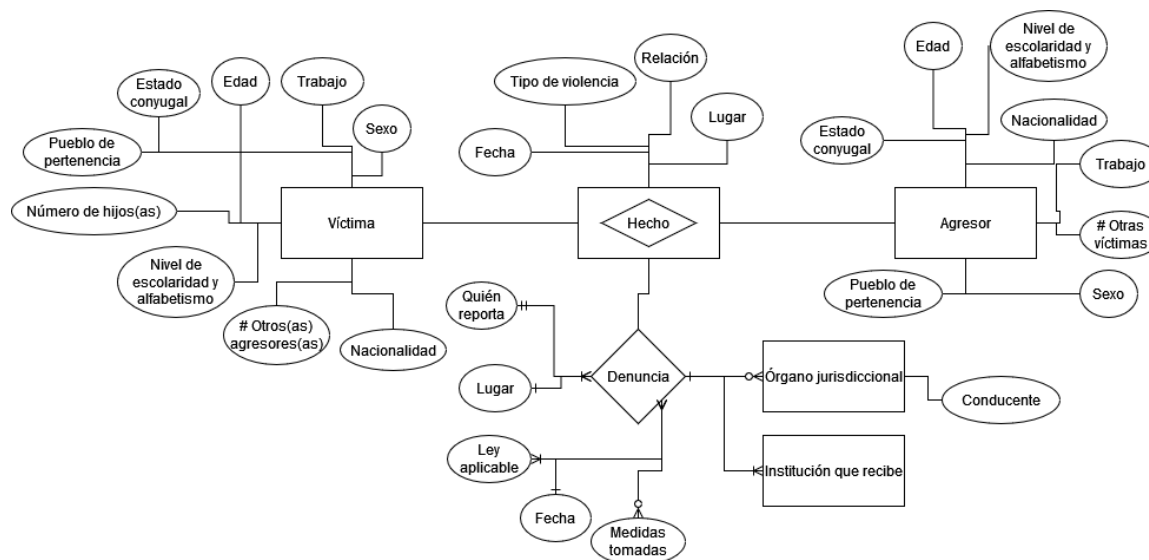


Figura 2: Diagrama entidad relación del modelo de datos preliminar.

## 7.3. Canalización ETL

### 7.3.1. Limpieza de datos

Se definirá una estrategia de limpieza de datos, al igual que procesos de control de calidad de los datos. Esta además deberá ser escalable y contemplar posibles cambios de variables a lo largo de los años.

### 7.3.2. Ingesta y transformación de datos

Se realizará un script en R para automatizar la ingesta y transformación de datos al modelo propuesto, para su carga a la base de datos relacional en PostgreSQL.

## 7.4. Análisis y visualización de datos

### 7.4.1. Definición de indicadores de rendimiento

A partir de información presentada por organizaciones tales como la ONU o ECLAC, se definirán indicadores para medir la violencia intrafamiliar y la violencia contra la mujer.

### 7.4.2. Prototipo de tablero

Se definirá un prototipo de tablero con visualizaciones relevantes y los KPIs definidos sobre información sobre violencia contra la mujer en Power BI, por medio de una conexión a la base de datos transformada.

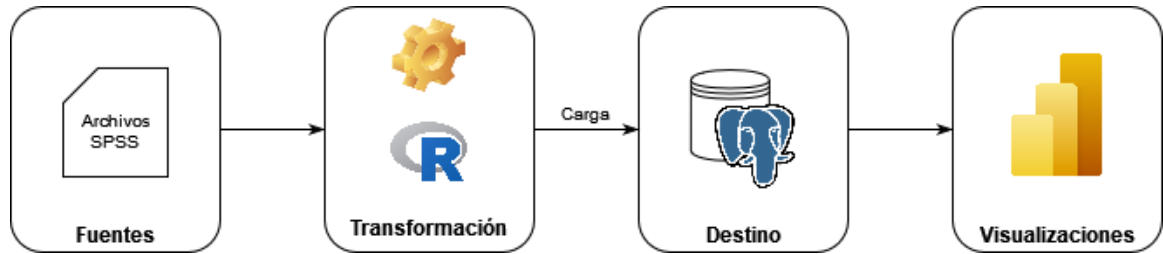


Figura 3: Diagrama de arquitectura y ETL preliminar.



## 8.1. Canalización ETL

### 8.1.1. Ingesta de datos

Se realizó, utilizando la librería para aplicaciones web Shiny, una aplicación en R para la compilación de información a partir de varios archivos de extensión .sav, y la publicación de la misma a una base de datos en PostgreSQL, la cual se utilizará como *staging*, es decir de almacenamiento temporal.

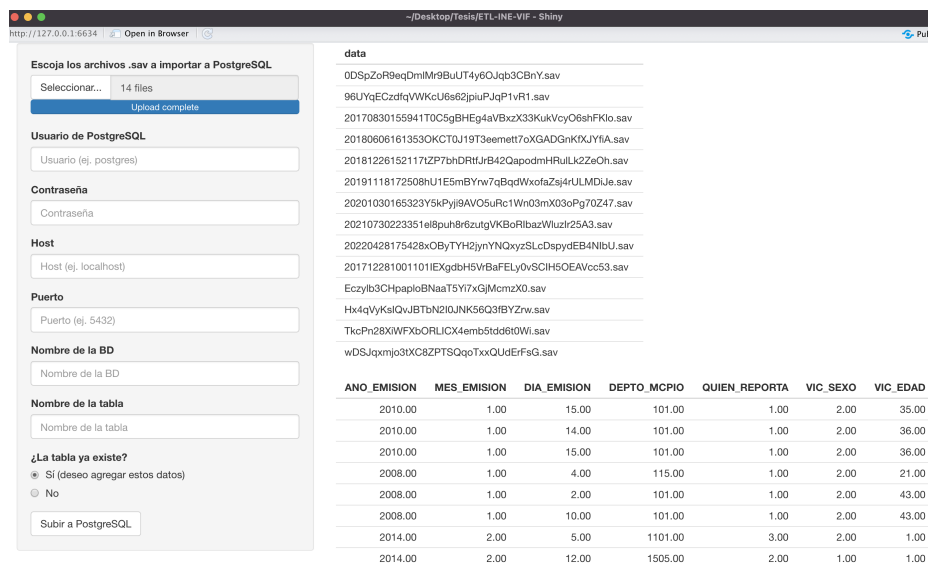


Figura 4: Captura de pantalla de aplicación en Shiny para carga de archivos.

### 8.1.2. Limpieza de datos

La estrategia de limpieza de datos se enfoca en diez de reglas, descritas a continuación:

- Se cambiaron todos los nombres de columnas a minúsculas, para mayor facilidad de trabajo en PostgreSQL.
- Varias de las bases de datos contenían el error ortográfico “agr\_gurpet”, el cual se corrigió a agr\_grupet.
- Se reemplazaron campos vacíos con el código equivalente a “Ignorado” para las columnas de vic\_alfab, agr\_alfab, vic\_escolaridad, agr\_escolaridad, vic\_est\_civ, agr\_est\_civ, vic\_trabaja, agr\_trabaja, organismo\_remite, organismo\_jurisdiccional, vic\_otras\_hom, vic\_otras\_muj, vic\_otras\_n\_as, vic\_otras\_n\_os, agr\_otros\_hom, agr\_otros\_muj, agr\_otros\_n\_as, agr\_otros\_n\_os, num\_hij\_hom, num\_hij\_muj, conducente, hec\_recur\_denun e inst\_denun\_hecho.
- Se reemplazaron campos vacíos de las columnas de agr\_ocup, agr\_dedica y vic\_ocup, vic\_dedica con el código equivalente a “Ignorado”, solo cuando agr\_trabaja y vic\_trabaja tenían el código de “Ignorado”, respectivamente.
- Se reemplazaron campos vacíos de la columna tipo\_discap con el código equivalente a “Ignorado” solo cuando vic\_disc tenía algún tipo de discapacidad o esta última tenía el código “Ignorado”.
- Se reemplazaron campos vacíos de la columna inst\_donde\_denuncio con el código equivalente a “Ignorado” solo cuando hec\_recur\_denun era equivalente a “Sí”; si hec\_recur\_denun era equivalente a “No”, se reemplazó inst\_donde\_denuncio a “No aplica”, y si era equivalente a “Ignorado”, se reemplazó este último al código equivalente a “Ignorado”.
- Se reemplazaron años menores a 1900 o valores vacíos en este campo por el equivalente a “Ignorado”. Si el año era equivalente “Ignorado”, los valores de mes y día cambiaron al código “Ignorado”.
- Se evaluó la validez de las fechas. Para fechas inválidas, si el día, mes y año eran distintos del código equivalente a “Ignorado”, se seleccionó la última fecha del mes.
- Se reemplazaron campos vacíos de la columna tipo\_medida con el código equivalente a “Ignorado” solo cuando se otorgaron medidas de seguridad, es decir, medidas\_seguridad era igual a 1; si medidas\_seguridad era igual a 2, es decir que no se otorgaron medidas de seguridad, este se reemplazó por un código equivalente a “Ninguno”, el cual se insertó para distinguir casos de valores vacíos en caso de si era ignorado si se otorgaron medidas.
- Si alguno de los datos se encontraba en la tabla de hechos/temporal, pero no en la de dimensión, se insertaron valores *dummy* o ficticios, los cuales deberían revisarse para respetar la gobernanza de datos.

### 8.1.3. Transformación de datos

Las tablas de dimensión se poblaron siguiendo la información publicada en los diccionarios del INE. Un ejemplo de estos, el del año 2021, puede encontrarse en los anexos. Cabe destacar las siguientes seis excepciones:

- Se agregó una descripción de “No aplica” para las dimensiones de `institucion_denuncia`, `organo_jurisdiccional`, `organo_remite`.
- Se agregó una descripción de “Ninguno” o “Ninguna” a las dimensiones de `dim_medida_seguridad` y `dim_discapacidad`.
- La dimensión tiempo agrupa los campos de año, mes y día, y además incluye datos útiles para inteligencia de negocio como el día de la semana.
- Las dimensiones `dim_escolaridad_vic` y `dim_escolaridad_agr` agrupan las columnas de `vic_alfab` y `vic_escolaridad` y `agr_alfab` y `agr_escolaridad`, respectivamente. Si la columna de alfabetismo es “Ignorado”, se coloca en la descripción de escolaridad como “Ignorado”. Por otro lado, si la columna de alfabetismo es “No”, se coloca en la descripción de escolaridad como “Ninguno”, esto para mantener congruencia ya que el INE nunca llena esta columna si la víctima o el agresor es analfabeta.
- Las dimensiones `dim_ocupacion_vic` y `dim_ocupacion_agr` agrupan las columnas de `vic_trabaja`, `vic_ocup` y `vic_dedica` y `agr_trabaja`, `agr_ocup` y `agr_dedica`.
- La dimensión `dim_ley_aplicable` no incluye descripciones que hacen referencia a más de una ley, como “Ambas leyes (VIF Y VCM)”, si no que estas se manejan como una relación muchos a muchos por medio de la tabla `bridge_ley`.

Se realizó un *script* en R para automatizar dichas transformaciones de datos al modelo propuesto, para su carga a la base de datos relacional en PostgreSQL.

El diagrama entidad relación de la base de datos puede observarse en la Figura 5.

## 8.2. Análisis y visualización de datos

### 8.2.1. Definición de indicadores de rendimiento

#### 8.2.1.1. Indicadores de violencia contra la mujer

De acuerdo con los Objetivos de Desarrollo Sostenible presentados por la ONU, se proponen como indicadores de rendimiento de violencia contra la mujer los descritos en el Cuadro 3.

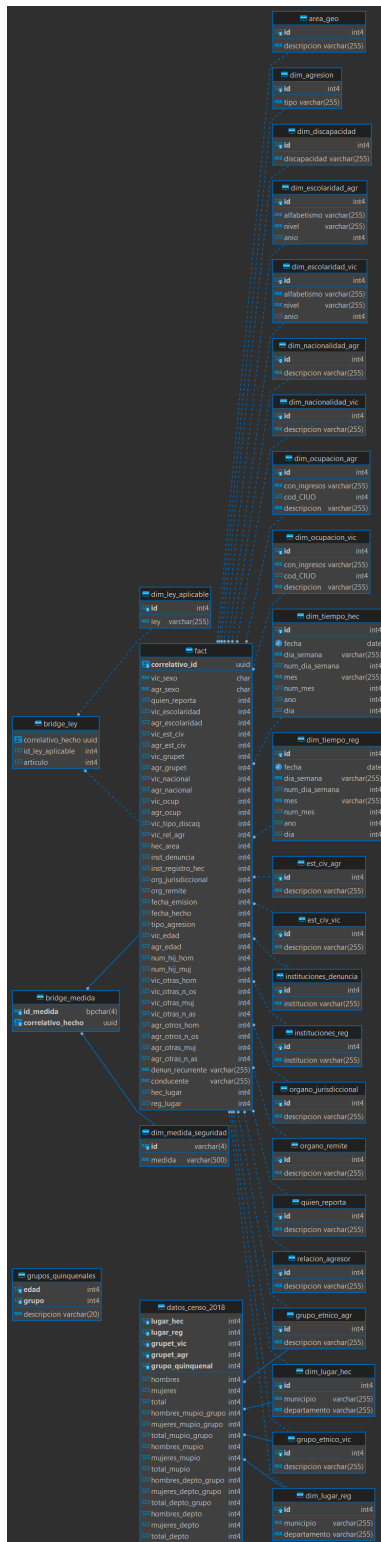


Figura 5: Diagrama entidad relación final de datos, generado con DBeaver.

Cuadro 3: Indicadores de violencia contra la mujer presentados en el tablero

KPI	Fórmula	Ejemplo
Tasa de cada 10 mil víctimas mujeres y niñas mayores de 15 años municipal, departamental y nacional	$\frac{M_Q}{T_Q} \times 10000 \quad (1)$ <p>Donde <math>M_Q</math> es el total de denuncias cuya víctima es una mujer mayor de 15 años y <math>T_Q</math> es el total de mujeres mayores de 15 años del municipio, departamento o a nivel nacional, según el censo nacional de 2018.</p>	<p>Para el año 2020, se registraron 1999 mujeres mayores de 15 años como víctimas de violencia contra la mujer en el departamento de Guatemala. Por lo tanto, la tasa es <math>\frac{1999 \times 10000}{1565878} = 12.77</math> denuncias por cada 10 mil mujeres mayores de 15 años en el departamento.</p>
Tasa de cada 10 mil mujeres y niñas mayores de 15 años que sujetas a algún tipo de violencia física, sexual o psicológica por parte de su conviviente, esposo o exconviviente, a nivel municipal y departamental, por forma de violencia y por grupo quinquenal de edad ( <i>Goal 5: Achieve gender equality and empower all women and girls</i> , S.f.)	$\frac{M_{QP}}{T_Q} \times 10000 \quad (2)$ <p>Donde <math>M_{QP}</math> es el total de mujeres mayores de 15 años sujetas a algún tipo de violencia física, sexual o psicológica por parte de su conviviente, esposo o exconviviente del municipio o departamento y <math>T_Q</math> es como se define en 1</p>	<p>Para el departamento de Guatemala, en el año 2021 388 mujeres entre 30 y 34 años de edad presentaron denuncias a las cuales aplica la Ley VCM. De estas, 100 reportaron que el tipo de agresión sufrida fue físico-psicológica. Por lo tanto, la tasa es <math>\frac{100 \times 10000}{1565878} = 0.59</math> denuncias por violencia físico-psicológica en el rango de 30 a 34 años por cada 10 mil mujeres mayores de 15 años en el departamento.</p>
Tasa de cada 10 mil mujeres y niñas mayores de 15 años sometidas a algún tipo de violencia sexual por parte de hombres que no sean su conviviente, esposo o exconviviente por edad y lugar de ocurrencia, a nivel municipal y departamental, por forma de violencia y grupo quinquenal de edad ( <i>Goal 5: Achieve gender equality and empower all women and girls</i> , S.f.)	$\frac{M_{Q\sim PS}}{T_Q} \times 10000 \quad (3)$ <p>Donde <math>M_{Q\sim PS}</math> es el total de mujeres mayores de 15 años sujetas a algún tipo de violencia sexual por parte de hombres que no sean su conviviente, esposo o exconviviente del municipio o departamento y <math>T_Q</math> es como se define en 1</p>	<p>Para el departamento de Chimaltenango, en el año 2021 4 mujeres entre 60 y 64 años de edad presentaron denuncias a las cuales aplica la Ley VCM, sufriendo el tipo de agresión sexual-patrimonial por alguien distinto a un esposo, conviviente o exconviviente hombre. Por lo tanto, la tasa es <math>\frac{4 \times 10000}{318622} = 0.12</math> denuncias por violencia sexual-patrimonial en el rango de 60 a 64 años por cada 10 mil mujeres mayores de 15 años en el departamento.</p>

(Continuación de Cuadro 3)

KPI	Fórmula	Ejemplo
Tasa de denuncia de cada 10 mil mujeres mayores a 15 años cuyo agresor es hombre, por grupo étnico, a nivel municipal y departamental	$\frac{M_{QG}}{T_{QG}} \times 10000 \quad (4)$ <p>Donde <math>M_{QG}</math> es el total de mujeres mayores de 15 años que denunciaron cualquier tipo de violencia ejercida por un hombre, por grupo étnico <math>G</math>, y <math>T_{QG}</math> es el total de mujeres mayores de 15 años del municipio o departamento del mismo grupo étnico <math>G</math>.</p>	Para el departamento de Chimaltenango, en el año 2021 2 mujeres garífunas de edad presentaron denuncias a las cuales aplica la Ley VCM. Por lo tanto, la tasa es $\frac{2 \times 10000}{289} = 69.02$ denuncias por cada 10 mil mujeres garífunas mayores de 15 años en el departamento.

Para 1, 2, 3 y 4, para poder tomarlos como indicadores de Violencia Contra la Mujer, la información fue filtrada para casos en los que la ley aplicable era el Decreto 22-2008.

El objetivo de 4 es mostrar si, a nivel municipal o departamental, existe alguna prevalencia de violencia por el factor de grupo étnico. Para esto, fue necesario incluir en el modelo de bases de datos información del censo agregado a departamental y municipal, por grupo étnico.

### 8.2.1.2. Indicadores de violencia intrafamiliar

Cuadro 4: Indicadores de violencia intrafamiliar presentados en el tablero

KPI	Fórmula	Ejemplo/interpretación
Distribución porcentual por departamento de registro/hecho	$\frac{D_d}{T_D} \times 100 \quad (5)$ <p>Donde <math>D_d</math> es el total de denuncias registradas en el departamento y <math>T_D</math> es el total de denuncias registradas/donde ocurrió el hecho.</p>	Es el porcentaje que representan las denuncias/hechos del departamento respecto al total de denuncias/hechos a nivel nacional.
Total de denuncias registradas anualmente/mensualmente	$T_D \quad (6)$ <p>Donde <math>T_D</math> es el total de denuncias registradas en el periodo dado.</p>	

(Continuación de Cuadro 4)

KPI	Fórmula	Ejemplo
Total de hechos anuales/- mensuales/por día de la se- mana	$T_H \quad (7)$ <p>Donde <math>T_H</math> es el total de hechos ocurridos en el periodo dado.</p>	
Distribución porcentual de institución que recibe la denuncia	$\frac{D_i}{T_D} \times 100 \quad (8)$ <p>Donde <math>D_i</math> es el total de denuncias registradas en una institución y <math>T_D</math> es el total de denuncias registradas.</p>	Es el porcentaje que representan las denuncias registradas en una institución respecto al total de denuncias a nivel nacional.
Distribución de quién reporta	$D_q \quad (9)$ <p>Donde <math>D_q</math> es el total de denuncias registradas, ya sea por una víctima, un familiar de la víctima, otro o ignorado.</p>	
Distribución porcentual de quién reporta	$\frac{D_q}{T_D} \times 100 \quad (10)$ <p>Donde <math>D_q</math> es el total de denuncias registradas, ya sea por una víctima, un familiar de la víctima, otro o ignorado, y <math>T_D</math> es el total de denuncias registradas.</p>	Es el porcentaje que representan las denuncias por quién reporta respecto al total de denuncias a nivel nacional.
Distribución de edad, según el sexo de la víctima/agresor	$D_{g,s} \quad (11)$ <p>Donde <math>D_g</math> es el total de denuncias registradas por grupo quinquenal de edad <math>g</math> y por sexo <math>s</math> de la víctima u agresor.</p>	
Distribución porcentual del grupo étnico de la víctima/agresor	$\frac{D_e}{T_D} \times 100 \quad (12)$ <p>Donde <math>D_e</math> es el total de denuncias registradas por grupo étnico de la víctima u agresor, y <math>T_D</math> es el total de denuncias registradas.</p>	Es el porcentaje que representan las víctimas/agresores de un grupo étnico respecto al total.

(Continuación de Cuadro 4)

KPI	Fórmula	Ejemplo
Distribución por sexo y nivel de escolaridad de la víctima/agresor	$D_{n,s} \quad (13)$ <p>Donde <math>D_{n,s}</math> es el total de denuncias registradas por nivel de escolaridad <math>n</math>, por sexo <math>s</math> de la víctima u agresor.</p>	
Distribución de relación con el agresor por sexo del agresor	$D_{r_a,s_a} \quad (14)$ <p>Donde <math>D_{r_a,s_a}</math> es el total de denuncias registradas por relación con el agresor <math>r_a</math> por sexo del agresor <math>s_a</math>.</p>	
Distribución de relación con la víctima por sexo de la víctima	$D_{r_v,s_v} \quad (15)$ <p>Donde <math>D_{r_v,s_v}</math> es el total de denuncias registradas por relación con la víctima <math>r_v</math> por sexo de la víctima <math>s_v</math>.</p>	
Distribución porcentual del sexo de la víctima/agresor según la ocupación de la víctima/agresor	$\frac{D_{t,s}}{T_{D_t}} \times 100 \quad (16)$ <p>Donde <math>D_{t,s}</math> es el total de denuncias registradas según la ocupación <math>t</math> (sí trabaja por un salario/ingreso, no o ignorado) de la víctima u agresor por el sexo de la víctima u agresor, <math>s</math>, y <math>T_{D_t}</math> es el total de denuncias registradas según la ocupación <math>t</math> de la víctima u agresor.</p>	<p>Para 2021, del total de 29487 víctimas que no trabajan por un salario o ingreso, 28865 eran mujeres, siendo entonces el <math>\frac{28865 \times 100}{29487} = 97.89</math> por ciento de víctimas que no trabajan por un salario o ingreso.</p>
Distribución porcentual del tipo de violencia del hecho	$\frac{H_v}{T_H} \times 100 \quad (17)$ <p>Donde <math>H_v</math> es el total de hechos por tipo de violencia <math>v</math> y <math>T_H</math> es el total de hechos ocurridos en el año.</p>	<p>Es el porcentaje que representan los hechos de un tipo de violencia dado respecto al total.</p>

(Continuación de Cuadro 4)

KPI	Fórmula	Ejemplo
Distribución del tipo de violencia del hecho	$H_v \quad (18)$ <p>Donde <math>H_v</math> es el total de denuncias por tipo de violencia <math>v</math>.</p>	
Distribución de recurrencia de denuncia por institución donde se denunció previamente	$D_r \quad (19)$ <p>Donde <math>D_r</math> es el total de denuncias por institución donde el hecho se denunció previamente. <math>r</math></p>	
Crecimiento de cantidad de denuncias recurrentes por institución donde se denunció previamente	$D_{r,i} - D_{r,i-1} \quad (20)$ <p>Donde <math>D_{r,i}</math> es el total de denuncias por institución donde el hecho se denunció previamente <math>r</math> en el año <math>i</math>.</p>	Es la diferencia de denuncias en una institución dada entre el año actual y el anterior.
Diferencia en denuncias respecto al año anterior, por quién denuncia	$D_{q,i} - D_{q,i-1} \quad (21)$ <p>Donde <math>D_{q,i}</math> es el total de denuncias por quién denuncia <math>q</math> en el año <math>i</math>.</p>	Es la diferencia de denuncias por quién reporta entre el año actual y el anterior.
Porcentaje de agresiones físicas por día de la semana	$\frac{H_{v \in \text{física}}}{T_H} \times 100 \quad (22)$ <p>Donde <math>H_{v \in \{\text{física}\}}</math> es el total de hechos donde el tipo de violencia <math>v</math> es algún tipo de violencia física (física, física-psicológica, física-patrimonial, etc.) y <math>T_H</math> es el total de hechos ocurridos en día de la semana.</p>	En 2021, 5718 hechos de violencia ocurrieron un día sábado, de los cuales 2949 fueron algún tipo de violencia física. Por lo tanto, el porcentaje de agresiones físicas para el día sábado es $\frac{2949 \times 100}{5718} = 51.57\%$ .

### 8.2.2. Prototipo de tablero

Se definió un prototipo de tablero con visualizaciones relevantes y los KPIs definidos sobre información sobre violencia contra la mujer y violencia intrafamiliar en Power BI, por medio de una conexión a la base de datos transformada. Se utilizó Power BI para realizar transformaciones adicionales para visualizaciones. Este tablero mantiene la misma estructura y varias de las visualizaciones que presenta el tablero oficial del INE, con algunas gráficas adicionales añadidas, o con modificaciones a las visualizaciones o presentación de los datos.

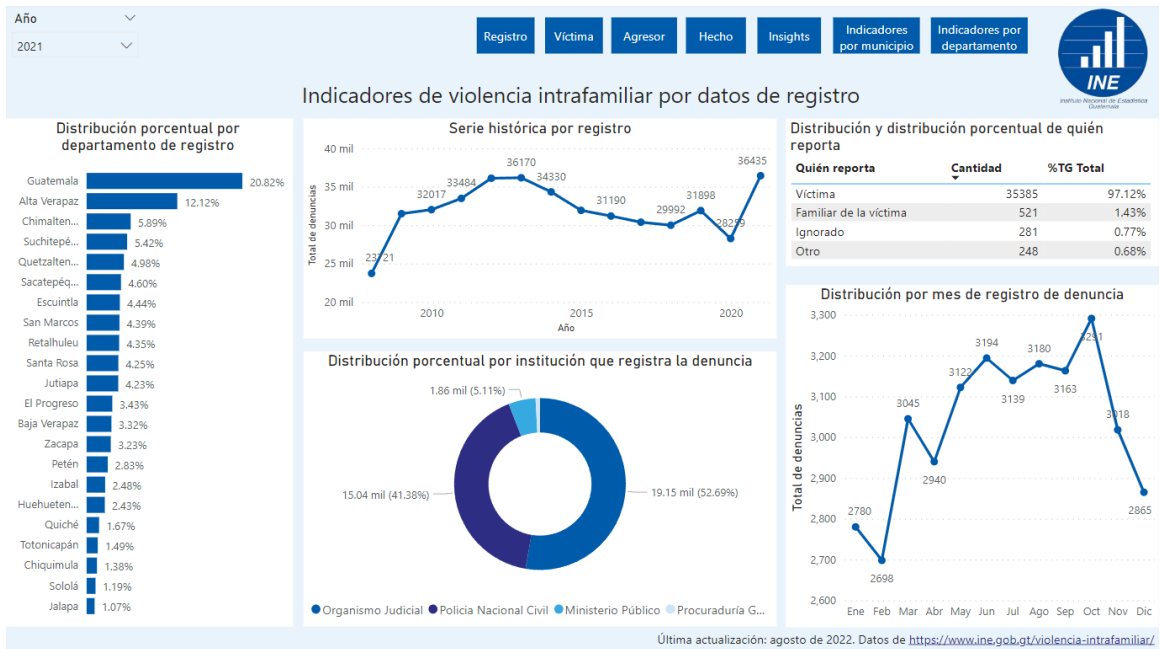


Figura 6: Captura de pestaña de indicadores de violencia intrafamiliar por datos de registro



Figura 7: Captura de pestaña de indicadores de violencia intrafamiliar por datos de la víctima

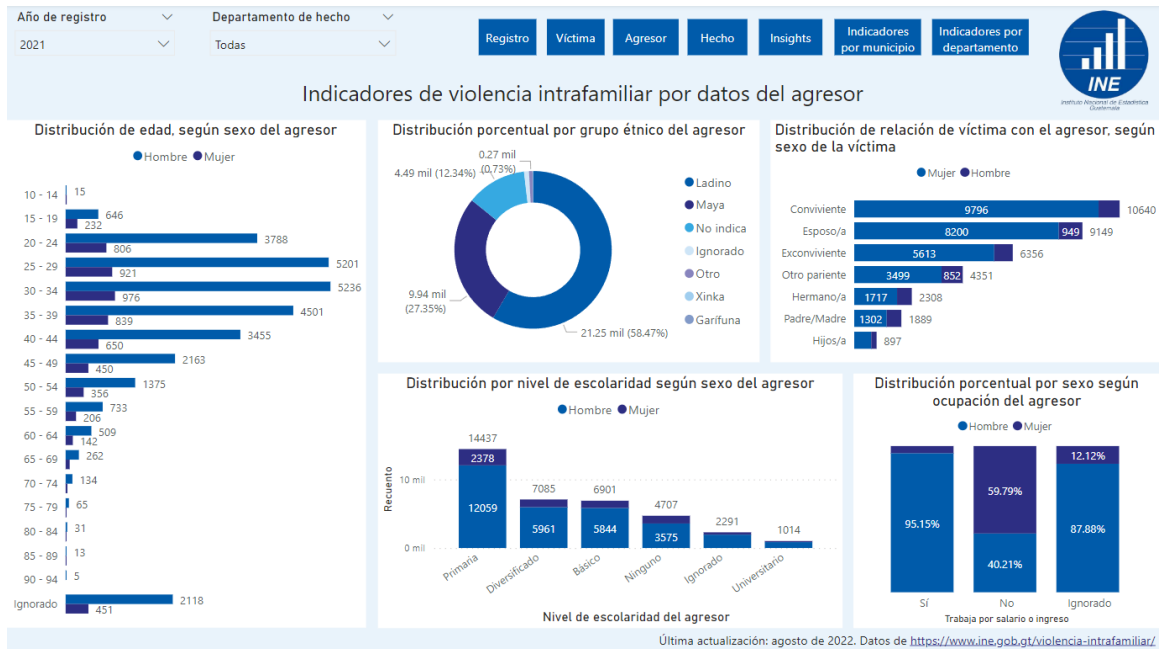


Figura 8: Captura de pestaña de indicadores de violencia intrafamiliar por datos del agresor

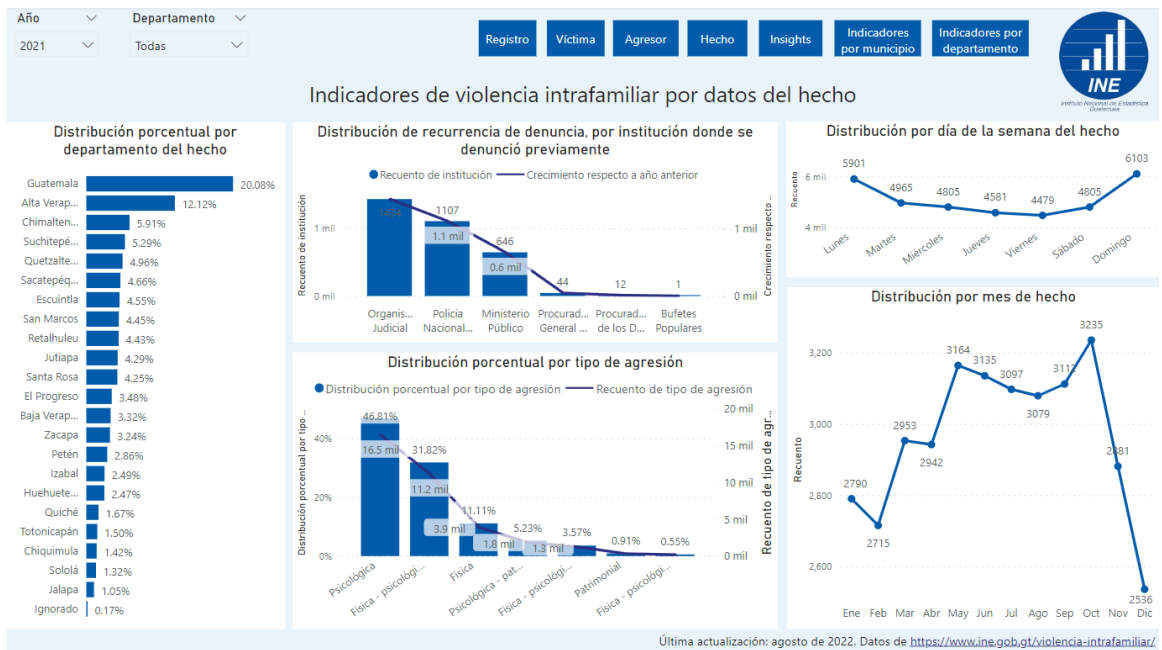


Figura 9: Captura de pestaña de indicadores de violencia intrafamiliar por datos de hecho



Figura 10: Captura de pestaña de insights para el año 2021

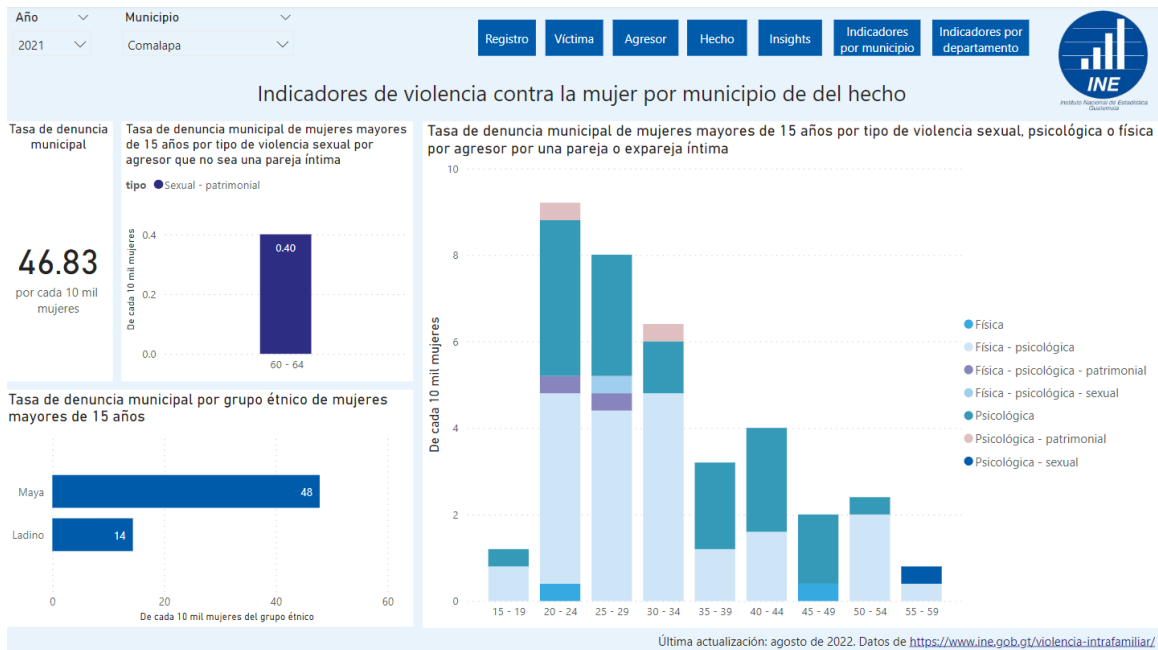


Figura 11: Captura de pestaña de indicadores por municipio

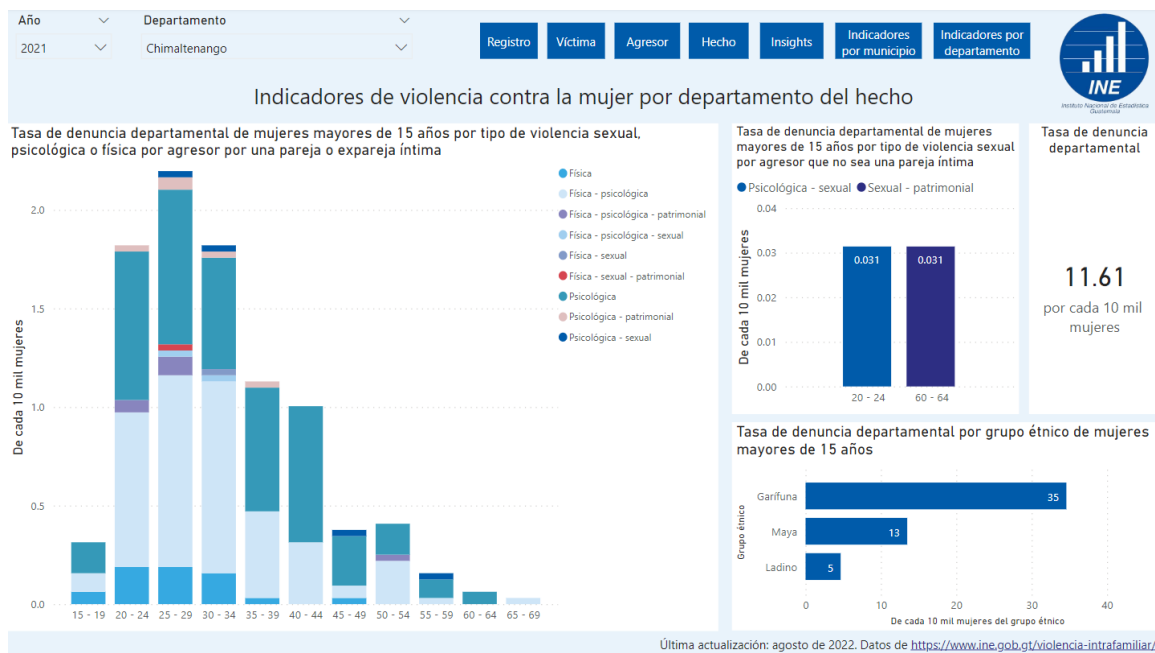


Figura 12: Captura de pestaña de indicadores por departamento

### 8.2.3. Presentación de resultados al INE

Como parte del trabajo, se realizó una sesión el viernes 24 de junio, en el cual estuvo presente el siguiente personal del INE: Karin Barrios, Jefe A.I. de la sección de Estadísticas Ambientales; Anabella de la Cruz, Jefe A.I. de la sección de Estadísticas Sociales; Patricia Zavaleta, Técnica Analista de la Unidad de Violencia Intrafamiliar. Las observaciones comentadas fueron las siguientes, y la minuta de la reunión puede encontrarse en los anexos:

- Se comentó que los *dashboards* de Power BI publicados a lo largo del sitio del INE son una parte de un plan institucional para la publicación de indicadores de cada tema. Este se realizó recopilando la información ya publicada en la página del INE.
- Se comentaron las tecnologías utilizadas actualmente, entrando en detalle que para la realización de los tableros fue necesaria la capacitación de personal.
- Se comentó a largos rasgos que de encontrar “mejoras a los tableros publicados actualmente, eran bienvenidas.”

Se tuvo una segunda reunión el día 7 de septiembre, con Karin Barrios, Anabela de la Cruz y Karina Escobar, Técnica Analista de la Unidad de Violencia Intrafamiliar. Esta duró 55 minutos, en el cual se destacan las siguientes observaciones: los indicadores calculados pueden beneficiarse de no utilizar la base estática de 2018, sino puede mejorarse utilizando las proyecciones del censo anuales. Sin embargo, el nivel más bajo al que esta llega es desagregada por sexo y edad a nivel municipal, por lo que se perderían los indicadores de prevalencia por grupo étnico. Adicionalmente, se comentó que las estadísticas publicadas anualmente de Violencia Contra la Mujer utilizan principalmente información del Ministerio Público, no de

reportes de Violencia Intrafamiliar; esto a pesar de que estos datos registran si una de las leyes aplicable a la denuncia, lo cual es determinado por un organismo jurisdiccional, es la ley de Violencia Contra la Mujer.

La retroalimentación del equipo del INE no mostró la recepción esperada, por lo que se propuso llegar a iteraciones del trabajo para encontrar un acuerdo común.

1. Se logró desarrollar automatización, identificando limitaciones debido a problemas de gobernanza de datos o falta de gestión de datos maestros. Entre estos, destaca la existencia de códigos de ocupaciones que se cambian o eliminan sin dejar registro o documentación de esto en los diccionarios, como .
2. Utilizando las herramientas propuestas, fue posible implementar una canalización ETL desde la fuente de datos del INE enfocados en el contexto violencia intrafamiliar. Además, es posible diseñar una base de datos relacional con modelo de estrella con la información pública.
3. La base de datos original contenía un total de 73 columnas (esto, después de excluir las columnas de filter\_\$ y de corregir la columna de agr\_gurpet). Fue posible reducir la dimensionalidad de la tabla de hechos a 40 columnas, utilizando 25 tablas de dimensión, agrupando la información de varias columnas, como lo fue el caso de la dimensión tiempo para el hecho, la cual agrupa las columnas originales de ANO\_HEC, MES\_HEC y DIA\_HEC, y la dimensión tiempo para el registro, la cual agrupa las columnas originales de ANO\_EMISION, MES\_EMISION y DIA\_EMISION.
4. El prototipo de *dashboard* presentado es dinámico, lo cual implica que puede ser filtrado por diversas dimensiones para su análisis, haciéndolo útil para permitir comparaciones entre años y/o departamentos.



### 10.1. AI INE

1. Se recomienda la implementación de un equipo o reglas de gobernanza de datos, al igual que una estandarización de procesos de limpieza de datos, todo esto con el objetivo de poder automatizar los procesos de limpieza de datos en la etapa previa a la publicación de información pública.
2. Como aspecto técnico, es posible implementar arquitectura de datos para agilizar la producción de reportes y tableros.
3. El agregar nuevas variables a las bases de datos públicas, como por ejemplo, una variable para identificar (anónimamente) a una víctima de otra, podría permitir evaluar la efectividad de las medidas otorgadas (si las hay), o sería útil para la realización de modelos de prevención de recurrencia.

### 10.2. Para futuras iteraciones del proyecto

1. Lo planteado en este trabajo puede reutilizarse para añadir modelos de ciencias de datos como lo es el modelado de series de tiempo. También, la metodología actual puede, a largos rasgos, adaptarse y reutilizarse para otras fuentes de datos.
2. Si se desea aportar a una entidad, se debe tener una alineación completa con necesidades de la misma. Se debe tener una documentación de los requisitos acordados, y definir un plan de comunicación con la entidad que se trabaje. Se recomienda la inclusión de metodologías de recopilación de requisitos del negocio o empresa para futuros proyectos de desarrollos de modelos analíticos.
3. Si bien el agrupar varias columnas en una misma tabla dimensión ayuda para la reducción de dimensionalidad, se debe procurar que la agrupación de columnas sea coherente y que estas estén relacionadas.

4. Se debe buscar que los indicadores presentados no solo informen o presenten datos estáticos, sino vale la pena incluir visualizaciones que puedan generar un impacto, por ejemplo, el evaluar por qué algunas instituciones tienen tantas denuncias recurrentes. Es valioso evaluar la creación de indicadores que puedan mostrar si existe alguna prevalencia por género, edad o grupo étnico.

- Alméras, D., Bravo, R., Milosavljevic, V., Montaña, S. & Rico, M. N. (2001). *Violence against women in couples: Latin America and the Caribbean : A proposal for measuring its incidence and trends*. ECLAC.
- Berger, O. (2004). *Acuerdo Gubernativo No. 346-2004*. [https://observatorioplanificacion.cepal.org/sites/default/files/instrument/files/G\\_ACUERDO\\_REFORMA\\_COPRE.pdf](https://observatorioplanificacion.cepal.org/sites/default/files/instrument/files/G_ACUERDO_REFORMA_COPRE.pdf)
- CNSS. (2015). *Committee on National Security Systems (CNSS) Glossary*. <https://rmf.org/wp-content/uploads/2017/10/CNSSI-4009.pdf>
- Comisión Presidencial de Gestión Pública Abierta y Transparencia. (2019). *Antecedentes del gobierno electrónico en Guatemala*. [https://transparencia.gob.gt/wp-content/uploads/2019/07/DGBOE\\_Compromiso8Hito2\\_Antecedentes-Gobierno-Electr%5C%C3%5C%B3nico.pdf](https://transparencia.gob.gt/wp-content/uploads/2019/07/DGBOE_Compromiso8Hito2_Antecedentes-Gobierno-Electr%5C%C3%5C%B3nico.pdf)
- Compare Tableau to Microsoft Power BI*. (S.f.). Tableau. <https://www.tableau.com/compare/tableau-power-bi>
- Daly, A., Devitt, K. & Mann, M. (Eds.). (2019). *Good data (Theory on Demand #29)*. Institute of Network Cultures.
- Decreto Ley 3-85. (1985). <https://www.ine.gob.gt/ine/wp-content/uploads/2017/10/Ley-Org%C3%A1nica-INE.pdf>
- Decreto Número 22-2008. (2008). *Diario de Centroamérica*.
- Decreto Número 5-2021. (2021). *Diario de Centroamérica*.
- Decreto Número 97-96. (1996). *Diario de Centroamérica*.
- Degenhart Asturias, E. A., Juárez Lucas, G. & Reyes Córdova, E. G. (2018). *Política Nacional de Datos Abiertos 2018-2022*. Ministerio de Gobernación. <https://uip.mingob.gob.gt/wp-content/uploads/2019/03/Pol%5C%C3%5C%ADtica-Nacional-de-Datos-Abiertos-2018-2022.pdf>
- Edmond, S. (2021). *Power BI vs Tableau: Which one should you choose?* Datacamp. [https://www.datacamp.com/blog/power-bi-vs-tableau-which-one-should-you-choose?gclid=EAIaIQobChMIwPu5t4qs3AIVAQAAAB0BAAAAEAAAYACAAEgJVzfD\\_BwE](https://www.datacamp.com/blog/power-bi-vs-tableau-which-one-should-you-choose?gclid=EAIaIQobChMIwPu5t4qs3AIVAQAAAB0BAAAAEAAAYACAAEgJVzfD_BwE)
- Esquema de estrella*. (S.f.). IBM Corporation. <https://www.ibm.com/docs/es/ida/9.1.2?topic=schemas-star>

- Fondo de Población de las Naciones Unidas (UNFPA). (2013). *The Role of Data in Addressing Violence Against Women and Girls*. [https://www.unfpa.org/sites/default/files/resource-pdf/finalUNFPA\\_CSW\\_Book\\_20130221\\_Data.pdf](https://www.unfpa.org/sites/default/files/resource-pdf/finalUNFPA_CSW_Book_20130221_Data.pdf)
- Goal 5: *Achieve gender equality and empower all women and girls*. (S.f.). United Nations Department of Economic y Social Affairs. <https://sdgs.un.org/goals/goal5>
- Gobierno Electrónico*. (S.f.). Gobierno de Guatemala. <https://transparencia.gob.gt/ejes-de-accion/gobierno-electronico/>
- Hunt, X., Tomlinson, M., Sikander, S., Skeen, S., Marlow, M., du Toit, S. & Eisner, M. (2020). Artificial Intelligence, Big Data, and mHealth: The Frontiers of the Prevention of Violence Against Children. *Frontiers in Artificial Intelligence*, 3. <https://doi.org/10.3389/frai.2020.543305>
- IBM Certified Data Engineer - Big Data*. (S.f.). IBM. <https://www.ibm.com/training/certification/50001501>
- IBM Cloud Education. (2020). *What is Data Modeling?* IBM. <https://www.ibm.com/cloud/learn/data-modeling>
- IBM Cloud Education. (2021a). *Structured vs. Unstructured Data: What's the difference?* IBM. <https://www.ibm.com/cloud/blog/structured-vs-unstructured-data>
- IBM Cloud Education. (2021b). *What is Data Visualization?* IBM. <https://www.ibm.com/cloud/learn/data-visualization>
- ILDA. (S.f.). LinkedIn. <https://gt.linkedin.com/company/ilda>
- ILDA. (2020). *Barómetro Regional de Datos Abiertos para América Latina y el Caribe 2020*. Indicators to measure violence against women. (2007). Naciones Unidas.
- Joseph, R. C. & Johnson, N. A. (2013). Big Data and Transformational Government. *IT Professional*, 15(6), 43-48. <https://doi.org/10.1109/MITP.2013.61>
- Kellerer, T. (S.f.). *SQL Feature Comparison*. SQL Workbench/J. [https://www.sql-workbench.eu/dbms\\_comparison.html](https://www.sql-workbench.eu/dbms_comparison.html)
- Kim, G.-H., Trimi, S. & Chung, J.-H. (2014). Big-Data Applications in the Government Sector. *Commun. ACM*, 57(3), 78-85. <https://doi.org/10.1145/2500873>
- Kimball, R. & Ross, M. (2013). *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling* (3.<sup>a</sup> ed.). Wiley.
- Ley Orgánica del Instituto Nacional de Estadística y su Reglamento. (S.f.). [https://www.ine.gob.gt/ine/wp-content/uploads/2022/01/Ley\\_Reglamento\\_INE\\_dic21.pdf](https://www.ine.gob.gt/ine/wp-content/uploads/2022/01/Ley_Reglamento_INE_dic21.pdf)
- Marr, B. (2019). *What's The Difference Between Structured, Semi-Structured And Unstructured Data?* Forbes. <https://www.forbes.com/sites/bernardmarr/2019/10/18/whats-the-difference-between-structured-semi-structured-and-unstructured-data/?sh=3cfbef262b4d>
- Menendez, J. C. (S.f.). *¿Qué es Power BI?* Microsoft. <https://www2.deloitte.com/es/es/pages/technology/articles/que-es-power-bi.html>
- OEA. (2006). *Declaración de Santo Domingo: Gobernabilidad y Desarrollo en la Sociedad del Conocimiento*. Organización de los Estados Americanos (OEA). <https://www.oas.org/docs/declarations/AG-DEC-46-Dec-de-Santo-Domingo-SPA.pdf>
- OECD. (2019). *The Path to Becoming a Data-Driven Public Sector*. <https://doi.org/https://doi.org/https://doi.org/10.1787/059814a7-en>
- The Open Data Barometer*. (S.f.). Open Data Barometer. <https://opendatabarometer.org/barometer/>
- Open Data Barometer - Leaders Edition: ODB Methodology - v1.0*. (2017). <http://opendatabarometer.org/doc/leadersEdition/ODB-leadersEdition-Methodology.pdf>

- Pérez Molina, O. (2020). Acuerdo Gubernativo No. 360-2012 y sus reformas. <https://transparencia.gob.gt/wp-content/uploads/Acuerdo-Gubernativo-360-2012-2020-1.pdf>
- Plan de Gobierno Digital 2021-2026*. (2021). Comisión Presidencial de Gobierno Abierto y Electrónico. [https://transparencia.gob.gt/wp-content/uploads/Plan-de-Gobierno-Digital\\_M-1.pdf](https://transparencia.gob.gt/wp-content/uploads/Plan-de-Gobierno-Digital_M-1.pdf)
- Portillo, A. (2000). *Acuerdo Gubernativo No. 831-2000*. <http://ww2.oj.gob.gt/justiciadegenero/wp-content/uploads/2014/07/Reglamento-de-la-ley-para-prev.-sancionar-y-erradicar-la-violencia-intrafam..pdf>
- Walby, S. (2016). *Ensuring data collection and research on violence against women and domestic violence: Article 11 of the Istanbul Convention*. Council of Europe. <https://edoc.coe.int/en/violence-against-women/7138-ensuring-data-collection-and-research-on-violence-against-women-and-domestic-violence-article-11-of-the-istanbul-convention.html>
- What is a data pipeline?* (S.f.). IBM. <https://www.ibm.com/topics/data-pipeline>
- What is a Data Warehouse?* (S.f.). Oracle. <https://www.oracle.com/database/what-is-a-data-warehouse/>
- What is a database?* (S.f.). Oracle. <https://www.oracle.com/database/what-is-database/>
- What is Data Analytics?* (S.f.). Oracle. <https://www.oracle.com/business-analytics/data-analytics/>



## CAPÍTULO 12

Anexos

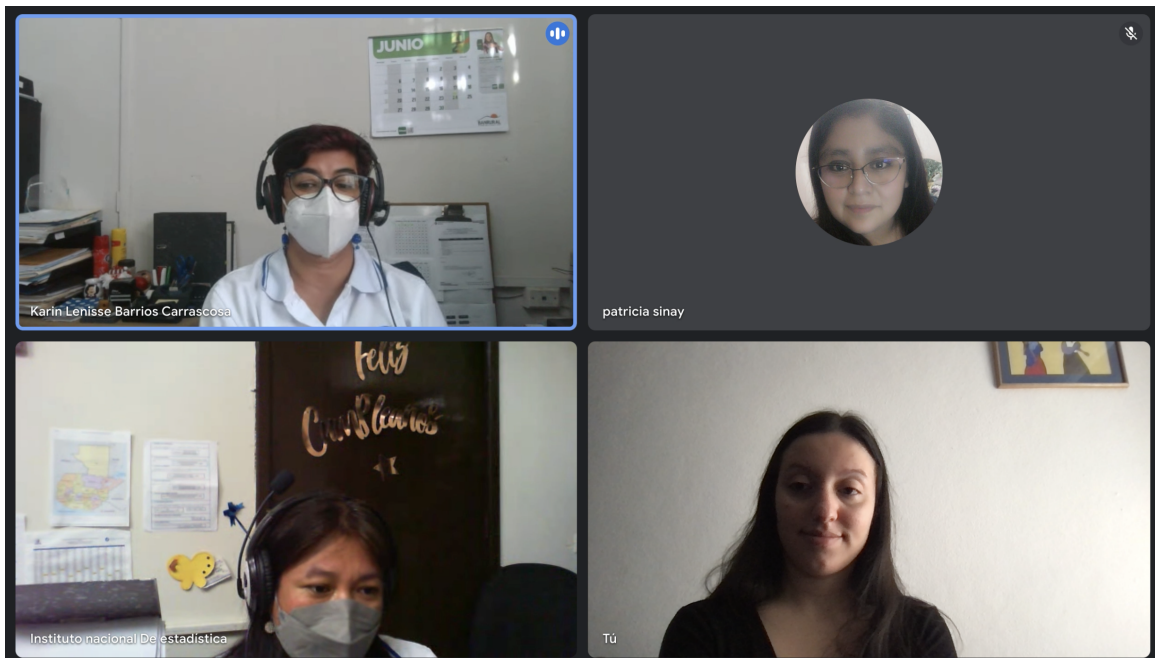


Figura 13: Reunión de aclaración de dudas con Karin Barrios, Patricia Zavaleta y Anabela de la Cruz



ANDREA CAROLINA ARGUELLO BARILLAS

para

mié, 29 jun, 8:00 ☆ ↶ ⋮

Buenos días a todos, espero que se encuentren bien.

Muchas gracias por su tiempo. Envío por esta vía los principales puntos de lo conversado el día viernes:

- El dashboard de Power BI de VIF publicado en el sitio del INE es una parte de un plan institucional para la publicación de indicadores de cada tema. Este se realizó recopilando la información ya publicada.
- Actualmente, se utilizan programas y software como SPSS, Excel, R y Power BI. El proyecto intentará apearse a tecnologías que ya se utilicen o que planeen utilizarse dentro del INE.
- La propuesta de proyecto pretende transformar dichos datos de VIF a un modelo de base de datos, culminando en un prototipo de dashboard conectado a los mismos. Este dashboard iterará sobre el ya publicado.

Cuando el prototipo de dashboard se encuentre más avanzado, agradecería poder contar con su apoyo para tener una reunión para recibir retroalimentación del mismo. Adicionalmente, quedo abierta a cualquier duda o comentario que pueda surgir.

Agradezco nuevamente su tiempo y apoyo. Sin más que agregar, quedo al pendiente de dudas sobre los puntos mencionados con anterioridad, o de qué puede agregarse, de no recibir respuesta, asumiré que todo está bien.

Saludos cordiales,  
Andrea Argüello

Figura 14: Minuta de primera reunión, información personal redactada.

Cuadro 5: Diccionario de variables base de datos Violencia Intrafamiliar 2021<sup>1</sup>

Código	Descripción del código	Valor	Etiqueta
DATOS DE REGISTRO			
ANO_EMISION	Año de emisión o registro de la denuncia en las fuentes de información	2020	2020
MES_EMISION	Mes de emisión o registro de la denuncia	1	Enero
		2	Febrero
		3	Marzo
		4	Abril
		5	Mayo
		6	Junio
		7	Julio
		8	Agosto
		9	Septiembre
		10	Octubre
		11	Noviembre
		12	Diciembre
DIA_EMISION	Día de emisión o registro de la denuncia	1 a 31	día de registro
		101	Guatemala
		102	Santa Catarina Pinula
		103	San José Pinula
		104	San José del Golfo
		105	Palencia
		106	Chinautla
		107	San Pedro Ayampuc
		108	Mixco
		109	San Pedro Sacatepéquez
		110	San Juan Sacatepéquez
		111	San Raymundo
		112	Chuarrancho
		113	Fraijanes
		114	Amatitlán
		115	Villa Nueva
		116	Villa Canales
		117	Petapa
		201	Guastatoya
		202	Morazán
		203	San Agustín Acasaguastlán
		204	San Cristóbal Acasaguastlán
205	El Jicaro		
206	Sansare		
207	Sanarate		
208	San Antonio La Paz		
301	Antigua Guatemala		
302	Jocotenango		
303	Pastores		
304	Sumpango		
305	Santo Domingo Xenacoj		
306	Santiago Sacatepéquez		
307	San Bartolomé Milpas Altas		
308	San Lucas Sacatepéquez		
309	Santa Lucía Milpas Altas		
310	Magdalena Milpas Altas		
311	Santa María de Jesús		

(Continuación de Cuadro 5)		
Código	Descripción del código	Etiqueta
		312 Ciudad Vieja
		313 San Miguel Dueñas
		314 Alotenango
		315 San Antonio Aguas Calientes
		316 Santa Catarina Barahona
		401 Chimaltenango
		402 San José Poaquil
		403 San Martín Jilotepeque
		404 Comalapa
		405 Santa Apolonia
		406 Tecpán Guatemala
		407 Patzún
		408 Pochuta
		409 Patzicía
		410 Santa Cruz Balanyá
		411 Acatenango
		412 Yepocapa
		413 San Andrés Itzapa
		414 Parramos
		415 Zaragoza
		416 El Tejar
		501 Escuintla
		502 Santa Lucía Cotzumalguapa
		503 La Democracia
		504 Siqinalá
		505 Masagua
		506 Tiquisate
		507 La Gomera
		508 Guanagazapa
		509 San José
		510 Iztapa
		511 Palín
		512 San Vicente Pacaya
		513 Nueva Concepción
		514 Sipacate
		601 Cuilapa
		602 Barberena
		603 Santa Rosa de Lima
		604 Casillas
		605 San Rafael Las Flores
		606 Oratorio
		607 San Juan Tecuaco
		608 Chiquimulilla
		609 Taxisco
		610 Santa María Ixhuitán
		611 Guazacapán
		612 Santa Cruz Naranjo
		613 Pueblo Nuevo Viñas
		614 Nueva Santa Rosa
		701 Sololá
		702 San José Chacayá
		703 Santa María Visitación
		704 Santa Lucía Utatlán
		705 Nahualá
		706 Santa Catarina Ixtahuacán
		707 Santa Clara La Laguna
		708 Concepción
		709 San Andrés Semetabaj
		710 Panajachel
		711 Santa Catarina Palopó
		712 San Antonio Palopó
		713 San Lucas Toliman
		714 Santa Cruz La Laguna
		715 San Pablo La Laguna
		716 San Marcos La Laguna
		717 San Juan La Laguna
		718 San Pedro La Laguna
		719 Santiago Atitlán
		801 Tonicapán
		802 San Cristobal Tonicapán
		803 San Francisco El Alto
		804 San Andrés Xecul
		805 Momostenango
		806 Santa María Chiquimula
		807 Santa Lucía La Reforma
		808 San Bartolo
		901 Quetzaltenango
		902 Salcajá
		903 Olinstepeque
		904 San Carlos Sija
		905 Sibilia
		906 Cabricán
		907 Cajolá
		908 San Miguel Siguilá
		909 Ostuncalco
		910 San Mateo
		911 Concepción Chiquirichapa
		912 San Martín Sacatepéquez
		913 Almolonga
		914 Cantel
		915 Huitán

(Continuación de Cuadro 5)		
Código	Descripción del código	Etiqueta
		916 Zunil
		917 Colomba
		918 San Francisco La Unión
		919 El Palmar
		920 Coatepeque
		921 Génova
		922 Flores Costa Cuca
		923 La Esperanza
		924 Palestina de Los Altos
		1001 Mazatenango
		1002 Cuyotenango
		1003 San Francisco Zapotitlán
		1004 San Bernardino
		1005 San José El Ídolo
		1006 Santo Domingo Suchitepéquez
		1007 San Lorenzo
		1008 Samayac
		1009 San Pablo Jocopilas
		1010 San Antonio Suchitepéquez
		1011 San Miguel Panán
		1012 San Gabriel
		1013 Chicacao
		1014 Patulul
		1015 Santa Bárbara
		1016 San Juan Bautista
		1017 Santo Tomás La Unión
		1018 Zumilito
		1019 Pueblo Nuevo
		1020 Río Bravo
		1021 San José La Máquina
		1101 Retalhuleu
		1102 San Sebastián
		1103 Santa Cruz Muluá
		1104 San Martín Zapotitlán
		1105 San Felipe
		1106 San Andrés Villa Seca
		1107 Champerico
		1108 Nuevo San Carlos
		1109 El Asintal
		1201 San Marcos
		1202 San Pedro Sacatepéquez
		1203 San Antonio Sacatepéquez
		1204 Comitancillo
		1205 San Miguel Ixtahuacán
		1206 Concepción Tutuapa
		1207 Tacaná
		1208 Sibinal
		1209 Tajumulco
		1210 Tejutla
		1211 San Rafael Pie de La Cuesta
		1212 Nuevo Progreso
		1213 El Tumbador
		1214 El Rodeo
		1215 Malacatán
		1216 Catarina
		1217 Ayutla
		1218 Ocós
		1219 San Pablo
		1220 El Quetzal
		1221 La Reforma
		1222 Pajapita
		1223 Ixchiguán
		1224 San José Ojetenam
		1225 San Cristóbal Cucho
		1226 Sipacapa
		1227 Esquipulas Palo Gordo
		1228 Río Blanco
		1229 San Lorenzo
		1230 La Blanca
		1301 Huehuetenango
		1302 Chiantla
		1303 Malacatancito
		1304 Cuilco
		1305 Nentón
		1306 San Pedro Necta
		1307 Jacaltenango
		1308 Soloma
		1309 Ixtahuacán
		1310 Santa Bárbara
		1311 La Libertad
		1312 La Democracia
		1313 San Miguel Acatán
		1314 San Rafael La Independencia
		1315 Todos Santos Cuchumatán
		1316 San Juan Atitán
		1317 Santa Eulalia
		1318 San Mateo Ixtatán
		1319 Colotenango
		1320 San Sebastián Huehuetenango

(Continuación de Cuadro 5)		
Código	Descripción del código	Etiqueta
		1321
		Tectitán
		1322
		Concepción Huista
		1323
		San Juan Ixcay
		1324
		San Antonio Huista
		1325
		San Sebastián Coatán
		1326
		Barillas
		1327
		Aguacatán
		1328
		San Rafael Petzal
		1329
		San Gaspar Ixchil
		1330
		Santiago Chimaltenango
		1331
		Santa Ana Huista
		1332
		Unión Cantinil
		1333
		Petatán
		1401
		Santa Cruz del Quiché
		1402
		Chiché
		1403
		Chinique
		1404
		Zacualpa
		1405
		Chajul
		1406
		Chichicastenango
		1407
		Patzité
		1408
		San Antonio Ilotenango
		1409
		San Pedro Jocopilas
		1410
		Cunén
		1411
		San Juan Cotzal
		1412
		Joyabaj
		1413
		Nebaj
		1414
		San Andrés Sajcabajá
		1415
		Uspantán
		1416
		Sacapulas
		1417
		San Bartolomé Jocotenango
		1418
		Canillá
		1419
		Chicamán
		1420
		Ixcán
		1421
		Pachalum
		1501
		Salamá
		1502
		San Miguel Chicaj
		1503
		Rabinal
		1504
		Cubulco
		1505
		Granados
		1506
		El Chol
		1507
		San Jerónimo
		1508
		Purulhá
		1601
		Cobán
		1602
		Santa Cruz Verapaz
		1603
		San Cristóbal Verapaz
		1604
		Tactic
		1605
		Tamahú
		1606
		Tucurú
		1607
		Panzós
		1608
		Senahú
		1609
		San Pedro Carchá
		1610
		San Juan Chamelco
		1611
		Lanquín
		1612
		Cahabón
		1613
		Chisec
		1614
		Chahal
		1615
		Fray Bartolomé de Las Casas
		1616
		Santa Catalina la Tinta
		1617
		Raxruhá
		1701
		Flores
		1702
		San José
		1703
		San Benito
		1704
		San Andrés
		1705
		La Libertad
		1706
		San Francisco
		1707
		Santa Ana
		1708
		Dolores
		1709
		San Luis
		1710
		Sayaxché
		1711
		Melchor de Mencos
		1712
		Poptún
		1713
		Las Cruces
		1714
		El Chal
		1801
		Puerto Barrios
		1802
		Livingston
		1803
		El Estor
		1804
		Morales
		1805
		Los Amates
		1901
		Zacapa
		1902
		Estanzuela
		1903
		Río Hondo
		1904
		Gualán
		1905
		Teculután
		1906
		Usumatlán
		1907
		Cabañas
		1908
		San Diego
		1909
		La Unión
		1910
		Huité
		1911
		San Jorge
		2001
		Chiquimula
		2002
		San José La Arada

(Continuación de Cuadro 5)			
Código	Descripción del código	Valor	Etiqueta
		2003	San Juan Erminta
		2004	Jocotán
		2005	Camotán
		2006	Olopa
		2007	Esquipulas
		2008	Concepción Las Minas
		2009	Quetzaltepeque
		2010	San Jacinto
		2011	Ipala
		2101	Jalapa
		2102	San Pedro Pinula
		2103	San Luis Jilotepeque
		2104	San Manuel Chaparrón
		2105	San Carlos Alzatate
		2106	Monjas
		2107	Mataquesuintla
		2201	Jutiapa
		2202	El Progreso
		2203	Santa Catarina Mita
		2204	Agua Blanca
		2205	Asunción Mita
		2206	Yupiltepeque
		2207	Atescatempa
		2208	Jerez
		2209	El Adelanto
		2210	Zapotitlán
		2211	Comapa
		2212	Jalpatagua
		2213	Conguaco
		2214	Moyuta
		2215	Pasaco
		2216	San José Acatempa
		2217	Quesada
QUIEN_REPORTA	Persona que reporta el hecho	1	Víctima
		2	Familiar de la víctima
		3	Otro
		9	Ignorado
DATOS GENERALES DE LA VÍCTIMA			
VIC_SEXO	Sexo de la víctima	1	Hombre
		2	Mujer
VIC_EDAD	Edad de la víctima	1 a 98	Edad de la víctima
		99	Edad ignorada de la víctima
TOTAL_HIJOS	Suma de hijos hombres e hijas mujeres	00 a 20	Total de hijos e Hijas
		99	Ignorado
NUM_HIJ_HOM	Número de hijos hombres de la víctima	00 a 20	Hijos hombres
		99	Ignorado
NUM_HIJ_MUJ	Número de hijas mujeres de la víctima	00 a 20	Hijas mujeres
		99	Ignorado
VIC_ALFAB	Sabe leer y escribir	1	Alfabeto
		2	Analfabeto
		9	Ignorado
		10	Ninguno
		21 a 26	Primero a sexto primaria
		29	Primaria (grado ignorado)
		31 a 33	Primero a tercero básico
		39	Básico (grado ignorado)
		44 a 46	Cuarto a sexto diversificado
		49	Diversificado (grado ignorado)
		51 a 57	Primer año a séptimo universitario (Incluye maestría)
		59	Universitario (año ignorado)
		99	Ignorado (tanto nivel como grado)
VIC_EST_CIV	Estado conyugal de la víctima	1	Soltero(a)
		2	Casado(a)
		3	Unido(a) (Unión declarada o no)
		4	Viudo(a)
		5	Otro
		9	Ignorado
VIC_GRUPET	Pueblo de pertenencia de la víctima	1	Ladino
		2	Maya
		3	Garífuna
		4	Xinka
		5	Otro
		6	No indica
		9	Ignorado
VIC_NACIONAL	Nacionalidad de la víctima	1	Guatemalteco

Código	(Continuación de Cuadro 5)		Etiqueta
	Descripción del código	Valor	
VIC_TRABAJA	Condición de empleo de la víctima	2	Extranjera
		9	Ignorado
		1	Si trabaja por un salario o ingreso
		2	No trabaja por un salario o ingreso
		9	Ignorado
		1111	Miembros del poder legislativo
		1112	Personal directivo de la administración pública
		1113	Jefes de pequeñas poblaciones
		1114	Dirigentes de organizaciones que presentan un interés especial
		1120	Directores generales y gerentes generales
		1211	Directores financieros
		1212	Directores de recursos humanos
		1213	Directores de políticas y planificación
		1219	Directores de administración y servicios no clasificados bajo otros epígrafes
		1221	Directores de ventas y comercialización
		1222	Directores de publicidad y relaciones públicas
		1223	Directores de investigación y desarrollo
		1311	Directores de producción agropecuaria y silvicultura
		1312	Directores de producción de piscicultura y pesca
		1321	Directores de industrias manufactureras
		1322	Directores de explotaciones de minería
		1323	Directores de empresas de construcción
		1324	Directores de empresas de abastecimiento, distribución y afines
		1330	Directores de servicios de tecnología de la información y las comunicaciones
		1341	Directores de servicios de cuidados infantiles
		1342	Directores de servicios de salud
		1343	Directores de servicios de cuidado de las personas de edad
		1344	Directores de servicios de bienestar social
		1345	Directores de servicios de educación
		1346	Gerentes de sucursales de bancos, de servicios financieros y de seguros
		1349	Directores y gerentes de servicios profesionales no clasificados bajo otros epígrafes
		1411	Gerentes de hoteles
		1412	Gerentes de restaurantes
		1420	Gerentes de comercios al por mayor y al por menor
		1431	Gerentes de centros deportivos, de esparcimiento y culturales
		1439	Gerentes de servicios no clasificados bajo otros epígrafes
		2111	Físicos y astrónomos
		2112	Meteorólogos
		2113	Químicos
		2114	Geólogos y geofísicos
		2120	Matemáticos, actuarios y estadísticos
2131	Biólogos, botánicos, zoólogos y afines		
2132	Agrónomos y afines		
2133	Profesionales de la protección medioambiental		
2141	Ingenieros industriales y de producción		
2142	Ingenieros civiles		
2143	Ingenieros medioambientales		
2144	Ingenieros mecánicos		

(Continuación de Cuadro 5)		
Código	Descripción del código	Etiqueta
		2145 Ingenieros químicos
		2146 Ingenieros de minas, metalúrgicos y afines
		2149 Ingenieros no clasificados bajo otros epígrafes
		2151 Ingenieros electricistas
		2152 Ingenieros electrónicos
		2153 Ingenieros en telecomunicaciones
		2161 Arquitectos
		2162 Arquitectos paisajistas
		2163 Diseñadores de productos y de prendas
		2164 Urbanistas e ingenieros de tránsito
		2165 Cartógrafos y agrimensores
		2166 Diseñadores gráficos y multimedia
		2211 Médicos generales
		2212 Médicos especialistas
		2221 Profesionales de enfermería
		2222 Profesionales de partería
		2230 Profesionales de medicina tradicional y alternativa
		2240 Practicantes paramédicos
		2250 Veterinarios
		2261 Dentistas
		2262 Farmacéuticos
		2263 Profesionales de la salud y la higiene laboral y ambiental
		2264 Fisioterapeutas
		2265 Dietistas y nutricionistas
		2266 Audiólogos y logopedas
		2267 Optometristas
		2269 Profesionales de la salud no clasificados bajo otros epígrafes
		2310 Profesores de universidades y de la enseñanza superior
		2320 Profesores de formación profesional
		2330 Profesores de enseñanza secundaria
		2341 Maestros de enseñanza primaria
		2342 Maestros preescolares
		2351 Especialistas en métodos pedagógicos
		2352 Educadores para necesidades especiales
		2353 Otros profesores de idiomas
		2354 Otros profesores de música
		2355 Otros profesores de artes
		2356 Instructores en tecnología de la información
		2359 Profesionales de la enseñanza no clasificados bajo otros epígrafes
		2411 Contables
		2412 Asesores financieros y en inversiones
		2413 Analistas financieros
		2421 Analistas de gestión y organización
		2422 Especialistas en políticas de administración
		2423 Especialistas en políticas y servicios de personal y afines
		2424 Especialistas en formación del personal
		2431 Profesionales de la publicidad y la comercialización
		2432 Profesionales de relaciones públicas
		2433 Profesionales de ventas técnicas y médicas (excluyendo la TIC)
		2434 Profesionales de ventas de tecnología de la información y las comunicaciones
		2511 Analistas de sistemas
		2512 Desarrolladores de software
		2513 Desarrolladores web y multimedia
		2514 Programadores de aplicaciones
		2519 Desarrolladores y analistas de software y multimedia y analistas no clasificados bajo otros epígrafes

(Continuación de Cuadro 5)		
Código	Descripción del código	Etiqueta
		2521 Diseñadores y administradores de bases de datos
		2522 Administradores de sistemas
		2523 Profesionales en redes de computadores
		2529 Especialistas en bases de datos y en redes de computadores no clasificados bajo otros epígrafes
		2611 Abogados
		2612 Jueces
		2619 Profesionales en derecho no clasificados bajo otros epígrafes
		2621 Archivistas y curadores de museos
		2622 Bibliotecarios, documentalistas y afines
		2631 Economistas
		2632 Sociólogos, antropólogos y afines
		2633 Filósofos, historiadores y especialistas en ciencias políticas
		2634 Psicólogos
		2635 Profesionales del trabajo social
		2636 Profesionales religiosos
		2641 Autores y otros escritores
		2642 Periodistas
		2643 Traductores, intérpretes y lingüistas
		2651 Artistas de artes plásticas
		2652 Músicos, cantantes y compositores
		2653 Bailarines y coreógrafos
		2654 Directores de cine, de teatro y afines
		2655 Actores
		2656 Locutores de radio, televisión y otros medios de comunicación
		2659 Artistas creativos e interpretativos no clasificados bajo otros epígrafes
		3111 Técnicos en ciencias físicas y químicas
		3112 Técnicos en ingeniería civil
		3113 Electrotécnicos
		3114 Técnicos en electrónica
		3115 Técnicos en ingeniería mecánica
		3116 Técnicos en química industrial
		3117 Técnicos en ingeniería de minas y metalurgia
		3118 Delineantes y dibujantes técnicos
		3119 Técnicos en ciencias físicas y en ingeniería no clasificados bajo otros epígrafes
		3121 Supervisores en ingeniería de minas
		3122 Supervisores de industrias manufactureras
		3123 Supervisores de la construcción
		3131 Operadores de instalaciones de producción de energía
		3132 Operadores de incineradores, instalaciones de tratamiento de agua y afines
		3133 Controladores de instalaciones de procesamiento de productos químicos
		3134 Operadores de instalaciones de refinación de petróleo y gas natural
		3135 Controladores de procesos de producción de metales
		3139 Técnicos en control de procesos no clasificados bajo otros epígrafes
		3141 Técnicos en ciencias biológicas (excluyendo la medicina)
		3142 Técnicos agropecuarios
		3143 Técnicos forestales
		3151 Oficiales maquinistas en navegación
		3152 Capitanes, oficiales de cubierta y prácticos

(Continuación de Cuadro 5)		
Código	Descripción del código	Etiqueta
	Valor	
	3153	Pilotos de aviación y afines
	3154	Controladores de tráfico aéreo
	3155	Técnicos en seguridad aeronáutica
	3211	Técnicos en aparatos de diagnóstico y tratamiento médico
	3212	Técnicos de laboratorios médicos
	3213	Técnicos y asistentes farmacéuticos
	3214	Técnicos de prótesis médicas y dentales
	3221	Profesionales de nivel medio de enfermería
	3222	Profesionales de nivel medio de partería
	3230	Profesionales de nivel medio de medicina tradicional y alternativa
	3240	Técnicos y asistentes veterinarios
	3251	Dentistas auxiliares y ayudantes de odontología
	3252	Técnicos en documentación sanitaria
	3253	Trabajadores comunitarios de la salud
	3254	Técnicos en optometría y ópticos
	3255	Técnicos y asistentes fisioterapeutas
	3256	Practicantes y asistentes médicos
	3257	Inspectores de la salud laboral, medioambiental y afines
	3258	Ayudantes de ambulancias
	3259	Profesionales de la salud de nivel medio no clasificados bajo otros epígrafes
	3311	Agentes de bolsa, cambio y otros servicios financieros
	3312	Oficiales de préstamos y créditos
	3313	Tenedores de libros
	3314	Profesionales de nivel medio de servicios estadísticos, matemáticos y afines
	3315	Tasadores
	3321	Agentes de seguros
	3322	Representantes comerciales
	3323	Agentes de compras
	3324	Agentes de compras y consignatarios
	3331	Declarantes o gestores de aduana
	3332	Organizadores de conferencias y eventos
	3333	Agentes de empleo y contratistas de mano de obra
	3334	Agentes inmobiliarios
	3339	Agentes de servicios comerciales no clasificados bajo otros epígrafes
	3341	Supervisores de secretaría
	3342	Secretarios jurídicos
	3343	Secretarios administrativos y ejecutivos
	3344	Secretarios médicos
	3351	Agentes de aduana e inspectores de fronteras
	3352	Agentes de administración tributaria
	3353	Agentes de servicios de seguridad social
	3354	Agentes de servicios de expedición de licencias y permisos
	3355	Inspectores de policía y detectives
	3359	Agentes de la administración pública para la aplicación de la ley y afines no clasificados bajo otros epígrafes
	3411	Profesionales de nivel medio del derecho y servicios legales y afines
	3412	Trabajadores y asistentes sociales de nivel medio
	3413	Auxiliares laicos de las religiones
	3421	Atletas y deportistas

(Continuación de Cuadro 5)		
Código	Descripción del código	Etiqueta
		3422 Entrenadores, instructores y árbitros de actividades deportivas
		3423 Instructores de educación física y actividades recreativas
		3431 Fotógrafos
		3432 Diseñadores y decoradores de interior
		3433 Técnicos en galerías de arte, museos y bibliotecas
		3434 Chefs
		3435 Otros profesionales de nivel medio en actividades culturales y artistas
		3511 Técnicos en operaciones de tecnología de la información y las comunicaciones
		3512 Técnicos en asistencia al usuario de tecnología de la información y las comunicaciones
		3513 Técnicos en redes y sistemas de computadores
		3514 Técnicos de la web
		3521 Técnicos de radiodifusión y grabación audio visual
		3522 Técnicos de ingeniería de las telecomunicaciones
		4110 Oficinistas generales
		4120 Secretarios (general)
		4131 Operadores de máquinas de procesamiento de texto y mecanógrafos
		4132 Grabadores de datos
		4211 Cajeros de bancos y afines
		4212 Receptores de apuestas y afines
		4213 Prestamistas
		4214 Cobradores y afines
		4221 Empleados de agencias de viajes
		4222 Empleados de centros de llamadas
		4223 Telefonistas
		4224 Recepcionistas de hoteles
		4225 Empleados de ventanillas de informaciones
		4226 Recepcionistas (general)
		4227 Entrevistadores de encuestas y de investigaciones de mercados
		4229 Empleados de servicios de información al cliente no clasificados bajo otros epígrafes
		4311 Empleados de contabilidad y cálculo de costos
		4312 Empleados de servicios estadísticos, financieros y de seguros
		4313 Empleados encargados de las nóminas
		4321 Empleados de control de abastecimientos e inventario
		4322 Empleados de servicios de apoyo a la producción
		4323 Empleados de servicios de transporte
		4411 Empleados de bibliotecas
		4412 Empleados de servicios de correos
		4413 Codificadores de datos, correctores de pruebas de imprenta y afines
		4414 Escribientes públicos y afines
		4415 Empleados de archivos
		4416 Empleados del servicio de personal
		4419 Personal de apoyo administrativo no clasificado bajo otros epígrafes
		5111 Auxiliares de servicio de abordó
		5112 Revisores y cobradores de los transportes públicos
		5113 Guías de turismo
		5120 Cocineros
		5131 Camareros de mesas
		5132 Camareros de barra
		5141 Peluqueros
		5142 Especialistas en tratamientos de belleza y afines

(Continuación de Cuadro 5)			
Código	Descripción del código	Valor	Etiqueta
		5151	Supervisores de mantenimiento y limpieza en oficinas, hoteles y otros establecimientos
		5152	Ecónomos y mayordomos domésticos
		5153	Conserjes
		5161	Astrólogos, adivinadores y afines
		5162	Acompañantes y ayudantes de cámara
		5163	Personal de pompas fúnebres y embalsamadores
		5164	Cuidadores de animales
		5165	Instructores de autoescuela
		5169	Trabajadores de servicios personales no clasificados bajo otros epígrafes
		5211	Vendedores de quioscos y de puestos de mercado
		5212	Vendedores ambulantes de productos comestibles
		5221	Comerciantes de tiendas
		5222	Supervisores de tiendas y almacenes
		5223	Asistentes de venta de tiendas y almacenes
		5230	Cajeros y expendedores de billetes
		5241	Modelos de moda, arte y publicidad
		5242	Demostradores de tiendas
		5243	Vendedores puerta a puerta
		5244	Vendedores por teléfono
		5245	Expedidores de gasolineras
		5246	Vendedores de comidas al mostrador
		5249	Vendedores no clasificados bajo otros epígrafes
		5311	Cuidadores de niños
		5312	Auxiliares de maestros
		5321	Trabajadores de los cuidados personales en instituciones
		5322	Trabajadores de los cuidados personales a domicilio
		5329	Trabajadores de los cuidados personales en servicios de salud no clasificados bajo otros epígrafes
		5411	Bomberos
		5412	Policías
		5413	Guardianes de prisión
		5414	Guardias de protección
		5419	Personal de los servicios de protección no clasificados bajo otros epígrafes
		6111	Agricultores y trabajadores calificados de cultivos extensivos
		6112	Agricultores y trabajadores calificados de plantaciones de árboles y arbustos
		6113	Agricultores y trabajadores calificados de huertas, invernaderos, viveros y jardines
		6114	Agricultores y trabajadores calificados de cultivos mixtos
		6121	Criadores de ganado
		6122	Avicultores y trabajadores calificados de la avicultura
		6123	Apicultores y sericultores y trabajadores calificados de la apicultura y la sericultura
		6129	Criadores y trabajadores pecuarios calificados de la cría de animales no clasificados bajo otros epígrafes
		6130	Productores y trabajadores calificados de explotaciones agropecuarias mixtas cuya producción se destina al mercado
		6210	Trabajadores forestales calificados y afines
		6221	Trabajadores de explotaciones de acuicultura
		6222	Pescadores de agua dulce y en aguas costeras
		6223	Pescadores de alta mar
		6224	Cazadores y tramperos

(Continuación de Cuadro 5)			
Código	Descripción del código	Valor	Etiqueta
		6310	Trabajadores agrícolas de subsistencia
		6320	Trabajadores pecuarios de subsistencia
		6330	Trabajadores agropecuarios de subsistencia
		6340	Pescadores, cazadores, tramperos y recolectores de subsistencia
		7111	Constructores de casas
		7112	Albañiles
		7113	Mamposteros, tronzadores, labrantes y grabadores de piedra
		7114	Operarios en cemento armado, enfoscadores y afines
		7115	Carpinteros de armar y de obra blanca
		7119	Oficiales y operarios de la construcción (obra gruesa) y afines no clasificados bajo otros epígrafes
		7121	Techadores
		7122	Parqueteros y colocadores de suelos
		7123	Revocadores
		7124	Instaladores de material aislante y de insonorización
		7125	Cristaleros
		7126	Fontaneros e instaladores de tuberías
		7127	Mecánicos-montadores de instalaciones de refrigeración y climatización
		7131	Pintores y empapeladores
		7132	Barnizadores y afines
		7133	Limpiadores de fachadas y deshollinadores
		7211	Moldeadores y macheros
		7212	Soldadores y oxicortadores
		7213	Chapistas y caldereros
		7214	Montadores de estructuras metálicas
		7215	Aparejadores y empalmadores de cables
		7221	Herrereros y forjadores
		7222	Herramientistas y afines
		7223	Reguladores y operadores de máquinas herramientas
		7224	Pulidores de metales y afiladores de herramientas
		7231	Mecánicos y reparadores de vehículos de motor
		7232	Mecánicos y reparadores de motores de avión
		7233	Mecánicos y reparadores de máquinas agrícolas e industriales
		7234	Reparadores de bicicletas y afines
		7311	Mecánicos y reparadores de instrumentos de precisión
		7312	Fabricantes y afinadores de instrumentos musicales
		7313	Joyereros, orfebres y plateros
		7314	Alfareros y afines (barro, arcilla y abrasivos)
		7315	Sopladores, modeladores, laminadores, cortadores y pulidores de vidrio
		7316	Redactores de carteles, pintores decorativos y grabadores
		7317	Artesanos en madera, cestería y materiales similares
		7318	Artesanos de los tejidos, el cuero y materiales similares
		7319	Artesanos no clasificados bajo otros epígrafes
		7321	Cajistas, tipógrafos y afines
		7322	Impresores
		7323	Encuadernadores y afines
		7411	Electricistas de obras y afines
		7412	Mecánicos y ajustadores electricistas
		7413	Instaladores y reparadores de líneas eléctricas
		7421	Mecánicos y reparadores en electrónica
		7422	Instaladores y reparadores en tecnología de la información y las comunicaciones

Código	(Continuación de Cuadro 5)		Etiqueta
	Descripción del código	Valor	
		7511	Carniceros, pescaderos y afines
		7512	Panaderos, pasteleros y confiteros
		7513	Operarios de la elaboración de productos lácteos
		7514	Operarios de la conservación de frutas, legumbres, verduras y afines
		7515	Catadores y clasificadores de alimentos y bebidas
		7516	Preparadores y elaboradores de tabaco y sus productos
		7521	Operarios del tratamiento de la madera
		7522	Ebanistas y afines
		7523	Reguladores y operadores de máquinas de labrar madera
		7531	Sastres, modistos, peleteros y sombrereros
		7532	Patronistas y cortadores de tela y afines
		7533	Costureros, bordadores y afines
		7534	Tapiceros, colchoneros y afines
		7535	Apelambradores, pellejeros y curtidores
		7536	Zapateros y afines
		7541	Buzos
		7542	Dinamiteros y pegadores
		7543	Clasificadores y probadores de productos (excluyendo alimentos y bebidas)
		7544	Fumigadores y otros controladores de plagas y malas hierbas
		7549	Oficiales, operarios y artesanos de artes mecánicas y de otros oficios no clasificados bajo otros epígrafes
		8111	Mineros y operadores de instalaciones mineras
		8112	Operadores de instalaciones de procesamiento de minerales y rocas
		8113	Perforadores y sondistas de pozos y afines
		8114	Operadores de máquinas para fabricar cemento y otros productos minerales
		8121	Operadores de instalaciones de procesamiento de metales
		8122	Operadores de máquinas pulidoras, galvanizadoras y recubridoras de metales
		8131	Operadores de plantas y máquinas de productos químicos
		8132	Operadores de máquinas para fabricar productos fotográficos
		8141	Operadores de máquinas para fabricar productos de caucho
		8142	Operadores de máquinas para fabricar productos de material plástico
		8143	Operadores de máquinas para fabricar productos de papel
		8151	Operadores de máquinas de preparación de fibras, hilado y devanado
		8152	Operadores de telares y otras máquinas tejedoras
		8153	Operadores de máquinas de coser
		8154	Operadores de máquinas de blanqueamiento, teñido y limpieza de tejidos
		8155	Operadores de máquinas de tratamiento de pieles y cueros
		8156	Operadores de máquinas para la fabricación de calzado y afines
		8157	Operadores de máquinas lavarropas
		8159	Operadores de máquinas para fabricar productos textiles y artículos de piel y cuero no clasificados bajo otros epígrafes

Código	(Continuación de Cuadro 5)		Etiqueta
	Descripción del código	Valor	
		8160	Operadores de máquinas para elaborar alimentos y productos afines
		8171	Operadores de instalaciones para la preparación de pasta para papel y papel
		8172	Operadores de instalaciones de procesamiento de la madera
		8181	Operadores de instalaciones de vidriería y cerámica
		8182	Operadores de máquinas de vapor y calderas
		8183	Operadores de máquinas de embalaje, embotellamiento y etiquetado
		8189	Operadores de máquinas y de instalaciones fijas no clasificados bajo otros epígrafes
		8211	Ensambladores de maquinaria mecánica
		8212	Ensambladores de equipos eléctricos y electrónicos
		8219	Ensambladores no clasificados bajo otros epígrafes
		8311	Maquinistas de locomotoras
		8312	Guardafrenos, guardagujas y agentes de maniobras
		8321	Conductores de motocicletas
		8322	Conductores de automóviles, taxis y camionetas
		8331	Conductores de autobuses y tranvías
		8332	Conductores de camiones pesados
		8341	Operadores de maquinaria agrícola y forestal móvil
		8342	Operadores de máquinas de movimiento de tierras y afines
		8343	Operadores de grúas, aparatos elevadores y afines
		8344	Operadores de autoelevadoras
		8350	Marineros de cubierta y afines
		9111	Limpiadores y asistentes domésticos
		9112	Limpiadores y asistentes de oficinas, hoteles y otros establecimientos
		9113	Personal doméstico
		9121	Lavaderos y planchadores manuales
		9122	Lavadores de vehículos
		9123	Lavadores de ventanas
		9129	Otro personal de limpieza
		9211	Peones de explotaciones agrícolas
		9212	Peones de explotaciones ganaderas
		9213	Peones de explotaciones de cultivos mixtos y ganaderos
		9214	Peones de jardinería y horticultura
		9215	Peones forestales
		9216	Peones de pesca y acuicultura
		9311	Peones de minas y canteras
		9312	Peones de obras públicas y mantenimiento
		9313	Peones de la construcción de edificios
		9321	Empacadores manuales
		9329	Peones de la industria manufacturera no clasificados bajo otros epígrafes
		9331	Conductores de vehículos accionados a pedal o a brazo
		9332	Conductores de vehículos y máquinas de tracción animal
		9333	Peones de carga
		9334	Reponedores de estanterías
		9411	Cocineros de comidas rápidas
		9412	Ayudantes de cocina
		9510	Trabajadores ambulantes de servicios y afines
		9520	Vendedores ambulantes (excluyendo de comida)
		9611	Recolectores de basura y material reciclable
		9612	Clasificadores de desechos

(Continuación de Cuadro 5)			
Código	Descripción del código	Valor	Etiqueta
		9613	Barrenderos y afines
		9621	Mensajeros, mandaderos, maleteros y repartidores
		9622	Personas que realizan trabajos varios
		9623	Recolectores de dinero en aparatos de venta automática y lectores de medidores
		9624	Acarreadores de agua y recolectores de leña
		9629	Ocupaciones elementales no clasificadas bajo otros epígrafes
		0110	Oficiales de las fuerzas armadas
		0210	Suboficiales de las fuerzas armadas
		0310	Otros miembros de las fuerzas armadas
		9998	Ocupaciones no especificadas en las anteriores
		9999	Ocupación ignorada
		1	Quehaceres del hogar
		2	Vive de su renta o jubilación
		3	Estudiar
		4	Desempleado
		6	Otro
		9	Ignorado
VIC_DEDICA	A que se dedica la víctima (sin ingresos)		
		1	Con discapacidad
		2	Sin discapacidad
		9	Ignorado
VIC_DISC	Discapacidad de la víctima		
		1	Ceguera
		2	Sordera
		3	Pérdida o discapacidad de las extremidades superiores
		4	Pérdida o discapacidad de las extremidades inferiores
		5	Deficiencia mental
		6	Otra discapacidad
		9	Ignorado
TIPO_DISCAQ	Tipo de discapacidad de la víctima		
		1	Esposo/a
		2	Conviviente
		3	Exconviviente
		4	Hijos/a
		5	Hijastros/a
		6	Padre/Madre
		7	Nieto/a
		8	Suegro/a
		9	Hermano/a
		10	Otro pariente
VIC_REL_AGR	Relación de la víctima con el agresor		
		00 a 20	Total otras víctimas
OTRAS_VICTIMAS	Total otras víctimas		
		99	Ignorado total otras víctimas
VIC_OTRAS_HOM	Otras víctimas hombres (de 13 años en adelante)		
		00 a 09	Otras víctimas hombres
		99	Ignorado otras víctimas hombres
VIC_OTRAS_MUJ	Otras víctimas mujeres (de 13 años en adelante)		
		00 a 09	Otras víctimas mujeres
		99	Ignorado otras víctimas víctimas
VIC_OTRAS_N_OS	Otras víctimas niños (de 00 a 12 años)		
		00 a 09	Otras víctimas niños
		99	Ignorado otras víctimas niños
VIC_OTRAS_N_AS	Otras víctimas niñas (de 00 a 12 años)		
		00 a 09	Otras víctimas niñas
		99	Ignorado otras víctimas niñas
DATOS DEL HECHO			
HEC_DIA	Día en que sucedió el hecho		
		1 a 31	día de ocurrencia
		99	Ignorado
		1	Enero
		2	Febrero
		3	Marzo
		4	Abril
		5	Mayo
		6	Junio
		7	Julio
		8	Agosto
		9	Septiembre
		10	Octubre
		11	Noviembre
		12	Diciembre
		99	Ignorado
HEC_MES	Mes en que ocurrió el hecho		
		2000	2000 en adelante (no puede ser superior al año de denuncia)
HEC_ANO	Año en que ocurrió el hecho		

(Continuación de Cuadro 5)		
Código	Descripción del código	Etiqueta
		9999 Ignorado año de ocurrencia del hecho
		101 Guatemala
		102 Santa Catarina Pinula
		103 San José Pinula
		104 San José del Golfo
		105 Palencia
		106 Chinautla
		107 San Pedro Ayampuc
		108 Mixco
		109 San Pedro Sacatepéquez
		110 San Juan Sacatepéquez
		111 San Raymundo
		112 Chuarrancho
		113 Fraijanes
		114 Amatitlán
		115 Villa Nueva
		116 Villa Canales
		117 Petapa
		201 Guastatoya
		202 Morazán
		203 San Agustín Acasaguastlán
		204 San Cristóbal Acasaguastlán
		205 El Jícara
		206 Sansare
		207 Sanarate
		208 San Antonio La Paz
		301 Antigua Guatemala
		302 Jocotenango
		303 Pastores
		304 Sumpango
		305 Santo Domingo Xenacoj
		306 Santiago Sacatepéquez
		307 San Bartolomé Milpas Altas
		308 San Lucas Sacatepéquez
		309 Santa Lucía Milpas Altas
		310 Magdalena Milpas Altas
		311 Santa María de Jesús
		312 Ciudad Vieja
		313 San Miguel Dueñas
		314 Alotenango
		315 San Antonio Aguas Calientes
		316 Santa Catarina Barahona
		401 Chimaltenango
		402 San José Poaquil
		403 San Martín Jilotepeque
		404 Comalapa
		405 Santa Apolonia
		406 Tecpán Guatemala
		407 Patzún
		408 Pochuta
		409 Patzicía
		410 Santa Cruz Balanyá
		411 Acatenango
		412 Yepocapa
		413 San Andrés Itzapa
		414 Parramos
		415 Zaragoza
		416 El Tejar
		501 Escuintla
		502 Santa Lucía Cotzumalguapa
		503 La Democracia
		504 Siquinalá
		505 Masagua
		506 Tiquisate
		507 La Gomera
		508 Guanagazapa
		509 San José
		510 Iztapa
		511 Palín
		512 San Vicente Pacaya
		513 Nueva Concepción
		514 Sipacate
		601 Cuilapa
		602 Barberena
		603 Santa Rosa de Lima
		604 Casillas
		605 San Rafael Las Flores
		606 Oratorio
		607 San Juan Tecuaco
		608 Chiquimulilla
		609 Taxisco
		610 Santa María Ixhuatán
		611 Guazacapán
		612 Santa Cruz Naranjo
		613 Pueblo Nuevo Viñas
		614 Nueva Santa Rosa
		701 Sololá
		702 San José Chacayá
		703 Santa María Visitación

(Continuación de Cuadro 5)		
Código	Descripción del código	Etiqueta
		704 Santa Lucía Utatlán
		705 Nahualá
		706 Santa Catarina Ixtahuacán
		707 Santa Clara La Laguna
		708 Concepción
		709 San Andrés Semetabaj
		710 Panajachel
		711 Santa Catarina Palopó
		712 San Antonio Palopó
		713 San Lucas Toliman
		714 Santa Cruz La Laguna
		715 San Pablo La Laguna
		716 San Marcos La Laguna
		717 San Juan La Laguna
		718 San Pedro La Laguna
		719 Santiago Atitlán
		801 Tonicapán
		802 San Cristobal Tonicapán
		803 San Francisco El Alto
		804 San Andrés Xecul
		805 Momostenango
		806 Santa María Chiquimula
		807 Santa Lucía La Reforma
		808 San Bartolo
		901 Quetzaltenango
		902 Salcajá
		903 Olintepeque
		904 San Carlos Sija
		905 Sibilía
		906 Cabricán
		907 Cajolá
		908 San Miguel Siguilá
		909 Ostuncalco
		910 San Mateo
		911 Concepción Chiquirichapa
		912 San Martín Sacatepéquez
		913 Almolonga
		914 Cantel
		915 Huitán
		916 Zunil
		917 Colomba
		918 San Francisco La Unión
		919 El Palmar
		920 Coatepeque
		921 Génova
		922 Flores Costa Cuca
		923 La Esperanza
		924 Palestina de Los Altos
		1001 Mazatenango
		1002 Cuyotenango
		1003 San Francisco Zapotitlán
		1004 San Bernardino
		1005 San José El Ídolo
		1006 Santo Domingo Suchitepé- quez
		1007 San Lorenzo
		1008 Samayac
		1009 San Pablo Jocopilas
		1010 San Antonio Suchitepéquez
		1011 San Miguel Panán
		1012 San Gabriel
		1013 Chicacao
		1014 Patulul
		1015 Santa Bárbara
		1016 San Juan Bautista
		1017 Santo Tomás La Unión
		1018 Zunilito
		1019 Pueblo Nuevo
		1020 Río Bravo
		1021 San José La Máquina
		1101 Retalhuleu
		1102 San Sebastián
		1103 Santa Cruz Muluá
		1104 San Martín Zapotitlán
		1105 San Felipe
		1106 San Andrés Villa Seca
		1107 Champerico
		1108 Nuevo San Carlos
		1109 El Asintal
		1201 San Marcos
		1202 San Pedro Sacatepéquez
		1203 San Antonio Sacatepéquez
		1204 Comitancillo
		1205 San Miguel Ixtahuacán
		1206 Concepción Tutuapa
		1207 Tacaná
		1208 Sibinal
		1209 Tajumulco
		1210 Tejutla
		1211 San Rafael Pié de La Cuesta
		1212 Nuevo Progreso

(Continuación de Cuadro 5)		
Código	Descripción del código	Etiqueta
		1213 El Tumbador
		1214 El Rodeo
		1215 Malacatán
		1216 Catarina
		1217 Ayutla
		1218 Ocós
		1219 San Pablo
		1220 El Quetzal
		1221 La Reforma
		1222 Pajapita
		1223 Ixchiguán
		1224 San José Ojetenam
		1225 San Cristóbal Cucho
		1226 Sipacapa
		1227 Esquipulas Palo Gordo
		1228 Río Blanco
		1229 San Lorenzo
		1230 La Blanca
		1301 Huehuetenango
		1302 Chiantla
		1303 Malacatancito
		1304 Cuilco
		1305 Nentón
		1306 San Pedro Necta
		1307 Jacaltenango
		1308 Soloma
		1309 Ixtahuacán
		1310 Santa Bárbara
		1311 La Libertad
		1312 La Democracia
		1313 San Miguel Acatán
		1314 San Rafael La Independencia
		1315 Todos Santos Cuchumatán
		1316 San Juan Atitán
		1317 Santa Eulalia
		1318 San Mateo Ixtatán
		1319 Colotenango
		1320 San Sebastián Huehuetenango
		1321 Tectitán
		1322 Concepción Huista
		1323 San Juan Ixcoy
		1324 San Antonio Huista
		1325 San Sebastián Coatán
		1326 Barillas
		1327 Aguacatán
		1328 San Rafael Petzal
		1329 San Gaspar Ixchil
		1330 Santiago Chimaltenango
		1331 Santa Ana Huista
		1332 Unión Cantinil
		1333 Petatán
		1401 Santa Cruz del Quiché
		1402 Chiché
		1403 Chinique
		1404 Zacualpa
		1405 Chajul
		1406 Chichicastenango
		1407 Patzité
		1408 San Antonio Iltotenango
		1409 San Pedro Jocopilas
		1410 Cunén
		1411 San Juan Cotzal
		1412 Joyabaj
		1413 Nebaj
		1414 San Andrés Sajcabajá
		1415 Uspantán
		1416 Sacapulas
		1417 San Bartolomé Jocotenango
		1418 Canillá
		1419 Chicamán
		1420 Ixcán
		1421 Pachalum
		1501 Salamá
		1502 San Miguel Chicaj
		1503 Rabinal
		1504 Cubulco
		1505 Granados
		1506 El Chol
		1507 San Jerónimo
		1508 Purulhá
		1601 Cobán
		1602 Santa Cruz Verapaz
		1603 San Cristóbal Verapaz
		1604 Tactic
		1605 Tamahú
		1606 Tukurú
		1607 Panzós
		1608 Senahú
		1609 San Pedro Carchá
		1610 San Juan Chamelco

(Continuación de Cuadro 5)			
Código	Descripción del código	Valor	Etiqueta
		1611	Lanquín
		1612	Cahabón
		1613	Chisec
		1614	Chahal
		1615	Fray Bartolomé de Las Casas
		1616	Santa Catalina la Tinta
		1617	Raxruhá
		1701	Flores
		1702	San José
		1703	San Benito
		1704	San Andrés
		1705	La Libertad
		1706	San Francisco
		1707	Santa Ana
		1708	Dolores
		1709	San Luis
		1710	Sayaxché
		1711	Melchor de Mencos
		1712	Poptún
		1713	Las Cruces
		1714	El Chal
		1801	Puerto Barrios
		1802	Livingston
		1803	El Estor
		1804	Morales
		1805	Los Amates
		1901	Zacapa
		1902	Estanzuela
		1903	Río Hondo
		1904	Gualán
		1905	Teculután
		1906	Usulután
		1907	Cabañas
		1908	San Diego
		1909	La Unión
		1910	Huité
		1911	San Jorge
		2001	Chiquimula
		2002	San José La Arada
		2003	San Juan Erminta
		2004	Jocotán
		2005	Camotán
		2006	Olopa
		2007	Esquipulas
		2008	Concepción Las Minas
		2009	Quetzaltepeque
		2010	San Jacinto
		2011	Ipala
		2101	Jalapa
		2102	San Pedro Pinula
		2103	San Luis Jilotepeque
		2104	San Manuel Chaparrón
		2105	San Carlos Alzatate
		2106	Monjas
		2107	Mataquescuintla
		2201	Jutiapa
		2202	El Progreso
		2203	Santa Catarina Mita
		2204	Agua Blanca
		2205	Asunción Mita
		2206	Yupiltepeque
		2207	Atescatempa
		2208	Jerez
		2209	El Adelanto
		2210	Zapotitlán
		2211	Comapa
		2212	Jalpatagua
		2213	Conguaco
		2214	Moyuta
		2215	Pasaco
		2216	San José Acatempa
		2217	Quesada
		9999	Ignorado
HEC_AREA	Area geográfica de ocurrencia	1	Urbana
		2	Rural
		9	Ignorado
HEC_TIPAGRE	Tipo de agresión sufrida por la víctima	1222	Física
		2122	Psicológica
		2212	Sexual
		2221	Patrimonial
		1122	Física - psicológica
		1212	Física - sexual
		1221	Física - patrimonial
		2112	Psicológica - sexual
		2121	Psicológica - patrimonial
		2211	Sexual - patrimonial
		1112	Física - psicológica - sexual
		1121	Física - psicológica - patrimonial

(Continuación de Cuadro 5)			
Código	Descripción del código	Valor	Etiqueta
		2111	Psicológica - sexual - patrimonial
		1211	Física - sexual - patrimonial
		1111	Física - psicológica - sexual - patrimonial
HEC_RECUR_DENUN	Reiteración de la denuncia (ha denunciado con anterioridad)	1	Si
		2	No
		9	Ignorado
INST_DONDE_DENUNCIO	En que institución denunció anteriormente	1	Ministerio Público
		2	Procuraduría General de la Nación
		3	Policía Nacional Civil
		4	Organismo Judicial
		5	Bufetes Populares
		6	Procuraduría de los Derechos Humanos
		9	Ignorado
DATOS DEL AGRESOR			
AGR_SEXO	Sexo del(a) agresor(a)	1	Hombre
		2	Mujer
AGR_EDAD	Edad del(a) agresor(a)	1 a 98	Edad del(a) agresor(a)
		99	Ignorado
AGR_ALFAB	Sabe leer y escribir	1	Alfabeta
		2	Analfabeta
		9	Ignorado
AGR_ESCOLARIDAD	Nivel de escolaridad del(a) agresor(a)	10	Ninguno
		21 a 26	Primero a sexto primaria
		29	Primaria (grado ignorado)
		31 a 33	Primero a tercero básico
		39	Básico (grado ignorado)
		44 a 46	Cuarto a sexto diversificado
		49	Diversificado (grado ignorado)
		51 a 57	Primer año a séptimo universitario (Incluye maestría)
		59	Universitario (año ignorado)
99	Ignorado (tanto nivel como grado)		
AGR_EST_CIV	Estado conyugal del (a) agresor(a)	1	Soltero(a)
		2	Casado(a)
		3	Unido(a)
		4	Viudo(a)
		5	Otro
		9	Ignorado
AGR_GRUPET	Pueblo de pertenencia del agresor(a)	1	Ladino
		2	Maya
		3	Garífuna
		4	Xinka
		5	Otro
		6	No indica
		9	Ignorado
AGR_NACIONAL	Nacionalidad del(a) agresor(a)	1	Guatemala
		2	Extranjera
		9	Ignorado
AGR_TRABAJA	Condición de empleo del(a) agresor(a)	1	Si trabaja por un salario o ingreso
		2	No trabaja por un salario o ingreso
		9	Ignorado
		1111	Miembros del poder legislativo
		1112	Personal directivo de la administración pública
		1113	Jefes de pequeñas poblaciones
		1114	Dirigentes de organizaciones que presentan un interés especial
		1120	Directores generales y gerentes generales
		1211	Directores financieros
		1212	Directores de recursos humanos
		1213	Directores de políticas y planificación
		1219	Directores de administración y servicios no clasificados bajo otros epígrafes

(Continuación de Cuadro 5)			
Código	Descripción del código	Valor	Etiqueta
		1221	Directores de ventas y comercialización
		1222	Directores de publicidad y relaciones públicas
		1223	Directores de investigación y desarrollo
		1311	Directores de producción agropecuaria y silvicultura
		1312	Directores de producción de piscicultura y pesca
		1321	Directores de industrias manufactureras
		1322	Directores de explotaciones de minería
		1323	Directores de empresas de construcción
		1324	Directores de empresas de abastecimiento, distribución y afines
		1330	Directores de servicios de tecnología de la información y las comunicaciones
		1341	Directores de servicios de cuidados infantiles
		1342	Directores de servicios de salud
		1343	Directores de servicios de cuidado de las personas de edad
		1344	Directores de servicios de bienestar social
		1345	Directores de servicios de educación
		1346	Gerentes de sucursales de bancos, de servicios financieros y de seguros
		1349	Directores y gerentes de servicios profesionales no clasificados bajo otros epígrafes
		1411	Gerentes de hoteles
		1412	Gerentes de restaurantes
		1420	Gerentes de comercios al por mayor y al por menor
		1431	Gerentes de centros deportivos, de esparcimiento y culturales
		1439	Gerentes de servicios no clasificados bajo otros epígrafes
		2111	Físicos y astrónomos
		2112	Meteorólogos
		2113	Químicos
		2114	Geólogos y geofísicos
		2120	Matemáticos, actuarios y estadísticos
		2131	Biólogos, botánicos, zoólogos y afines
		2132	Agrónomos y afines
		2133	Profesionales de la protección medioambiental
		2141	Ingenieros industriales y de producción
		2142	Ingenieros civiles
		2143	Ingenieros medioambientales
		2144	Ingenieros mecánicos
		2145	Ingenieros químicos
		2146	Ingenieros de minas, metalúrgicos y afines
		2149	Ingenieros no clasificados bajo otros epígrafes
		2151	Ingenieros electricistas
		2152	Ingenieros electrónicos
		2153	Ingenieros en telecomunicaciones
		2161	Arquitectos
		2162	Arquitectos paisajistas
		2163	Diseñadores de productos y de prendas
		2164	Urbanistas e ingenieros de tránsito
		2165	Cartógrafos y agrimensores
		2166	Diseñadores gráficos y multimedia
		2211	Médicos generales
		2212	Médicos especialistas
		2221	Profesionales de enfermería
		2222	Profesionales de partería
		2230	Profesionales de medicina tradicional y alternativa
		2240	Practicantes paramédicos
		2250	Veterinarios
		2261	Dentistas
		2262	Farmacéuticos

(Continuación de Cuadro 5)			
Código	Descripción del código	Valor	Etiqueta
		2263	Profesionales de la salud y la higiene laboral y ambiental
		2264	Fisioterapeutas
		2265	Dietistas y nutricionistas
		2266	Audiólogos y logopedas
		2267	Optometristas
		2269	Profesionales de la salud no clasificados bajo otros epígrafes
		2310	Profesores de universidades y de la enseñanza superior
		2320	Profesores de formación profesional
		2330	Profesores de enseñanza secundaria
		2341	Maestros de enseñanza primaria
		2342	Maestros preescolares
		2351	Especialistas en métodos pedagógicos
		2352	Educadores para necesidades especiales
		2353	Otros profesores de idiomas
		2354	Otros profesores de música
		2355	Otros profesores de artes
		2356	Instructores en tecnología de la información
		2359	Profesionales de la enseñanza no clasificados bajo otros epígrafes
		2411	Contables
		2412	Asesores financieros y en inversiones
		2413	Analistas financieros
		2421	Analistas de gestión y organización
		2422	Especialistas en políticas de administración
		2423	Especialistas en políticas y servicios de personal y afines
		2424	Especialistas en formación del personal
		2431	Profesionales de la publicidad y la comercialización
		2432	Profesionales de relaciones públicas
		2433	Profesionales de ventas técnicas y médicas (excluyendo la TIC)
		2434	Profesionales de ventas de tecnología de la información y las comunicaciones
		2511	Analistas de sistemas
		2512	Desarrolladores de software
		2513	Desarrolladores web y multimedia
		2514	Programadores de aplicaciones
		2519	Desarrolladores y analistas de software y multimedia y analistas no clasificados bajo otros epígrafes
		2521	Diseñadores y administradores de bases de datos
		2522	Administradores de sistemas
		2523	Profesionales en redes de computadores
		2529	Especialistas en bases de datos y en redes de computadores no clasificados bajo otros epígrafes
		2611	Abogados
		2612	Jueces
		2619	Profesionales en derecho no clasificados bajo otros epígrafes
		2621	Archivistas y curadores de museos
		2622	Bibliotecarios, documentalistas y afines
		2631	Economistas
		2632	Sociólogos, antropólogos y afines
		2633	Filósofos, historiadores y especialistas en ciencias políticas
		2634	Psicólogos
		2635	Profesionales del trabajo social
		2636	Profesionales religiosos
		2641	Autores y otros escritores

(Continuación de Cuadro 5)			
Código	Descripción del código	Valor	Etiqueta
		2642	Periodistas
		2643	Traductores, intérpretes y lingüistas
		2651	Artistas de artes plásticas
		2652	Músicos, cantantes y compositores
		2653	Bailarines y coreógrafos
		2654	Directores de cine, de teatro y afines
		2655	Actores
		2656	Locutores de radio, televisión y otros medios de comunicación
		2659	Artistas creativos e interpretativos no clasificados bajo otros epígrafes
		3111	Técnicos en ciencias físicas y químicas
		3112	Técnicos en ingeniería civil
		3113	Electrotécnicos
		3114	Técnicos en electrónica
		3115	Técnicos en ingeniería mecánica
		3116	Técnicos en química industrial
		3117	Técnicos en ingeniería de minas y metalurgia
		3118	Delineantes y dibujantes técnicos
		3119	Técnicos en ciencias físicas y en ingeniería no clasificados bajo otros epígrafes
		3121	Supervisores en ingeniería de minas
		3122	Supervisores de industrias manufactureras
		3123	Supervisores de la construcción
		3131	Operadores de instalaciones de producción de energía
		3132	Operadores de incineradores, instalaciones de tratamiento de agua y afines
		3133	Controladores de instalaciones de procesamiento de productos químicos
		3134	Operadores de instalaciones de refinación de petróleo y gas natural
		3135	Controladores de procesos de producción de metales
		3139	Técnicos en control de procesos no clasificados bajo otros epígrafes
		3141	Técnicos en ciencias biológicas (excluyendo la medicina)
		3142	Técnicos agropecuarios
		3143	Técnicos forestales
		3151	Oficiales maquinistas en navegación
		3152	Capitanes, oficiales de cubierta y prácticos
		3153	Pilotos de aviación y afines
		3154	Controladores de tráfico aéreo
		3155	Técnicos en seguridad aeronáutica
		3211	Técnicos en aparatos de diagnóstico y tratamiento médico
		3212	Técnicos de laboratorios médicos
		3213	Técnicos y asistentes farmacéuticos
		3214	Técnicos de prótesis médicas y dentales
		3221	Profesionales de nivel medio de enfermería
		3222	Profesionales de nivel medio de partería
		3230	Profesionales de nivel medio de medicina tradicional y alternativa
		3240	Técnicos y asistentes veterinarios
		3251	Dentistas auxiliares y ayudantes de odontología
		3252	Técnicos en documentación sanitaria
		3253	Trabajadores comunitarios de la salud

(Continuación de Cuadro 5)		
Código	Descripción del código	Etiqueta
		3254 Técnicos en optometría y ópticos
		3255 Técnicos y asistentes fisioterapeutas
		3256 Practicantes y asistentes médicos
		3257 Inspectores de la salud laboral, medioambiental y afines
		3258 Ayudantes de ambulancias
		3259 Profesionales de la salud de nivel medio no clasificados bajo otros epígrafes
		3311 Agentes de bolsa, cambio y otros servicios financieros
		3312 Oficiales de préstamos y créditos
		3313 Tenedores de libros
		3314 Profesionales de nivel medio de servicios estadísticos, matemáticos y afines
		3315 Tasadores
		3321 Agentes de seguros
		3322 Representantes comerciales
		3323 Agentes de compras
		3324 Agentes de compras y consignatarios
		3331 Declarantes o gestores de aduana
		3332 Organizadores de conferencias y eventos
		3333 Agentes de empleo y contratistas de mano de obra
		3334 Agentes inmobiliarios
		3339 Agentes de servicios comerciales no clasificados bajo otros epígrafes
		3341 Supervisores de secretaría
		3342 Secretarios jurídicos
		3343 Secretarios administrativos y ejecutivos
		3344 Secretarios médicos
		3351 Agentes de aduana e inspectores de fronteras
		3352 Agentes de administración tributaria
		3353 Agentes de servicios de seguridad social
		3354 Agentes de servicios de expedición de licencias y permisos
		3355 Inspectores de policía y detectives
		3359 Agentes de la administración pública para la aplicación de la ley y afines no clasificados bajo otros epígrafes
		3411 Profesionales de nivel medio del derecho y servicios legales y afines
		3412 Trabajadores y asistentes sociales de nivel medio
		3413 Auxiliares laicos de las religiones
		3421 Atletas y deportistas
		3422 Entrenadores, instructores y árbitros de actividades deportivas
		3423 Instructores de educación física y actividades recreativas
		3431 Fotógrafos
		3432 Diseñadores y decoradores de interior
		3433 Técnicos en galerías de arte, museos y bibliotecas
		3434 Chefs
		3435 Otros profesionales de nivel medio en actividades culturales y artistas
		3511 Técnicos en operaciones de tecnología de la información y las comunicaciones
		3512 Técnicos en asistencia al usuario de tecnología de la información y las comunicaciones
		3513 Técnicos en redes y sistemas de computadores
		3514 Técnicos de la web
		3521 Técnicos de radiodifusión y grabación audio visual
		3522 Técnicos de ingeniería de las telecomunicaciones

(Continuación de Cuadro 5)			
Código	Descripción del código	Valor	Etiqueta
		4110	Oficinistas generales
		4120	Secretarios (general)
		4131	Operadores de máquinas de procesamiento de texto y mecanógrafos
		4132	Grabadores de datos
		4211	Cajeros de bancos y afines
		4212	Receptores de apuestas y afines
		4213	Prestamistas
		4214	Cobradores y afines
		4221	Empleados de agencias de viajes
		4222	Empleados de centros de llamadas
		4223	Telefonistas
		4224	Recepcionistas de hoteles
		4225	Empleados de ventanillas de informaciones
		4226	Recepcionistas (general)
		4227	Entrevistadores de encuestas y de investigaciones de mercados
		4229	Empleados de servicios de información al cliente no clasificados bajo otros epígrafes
		4311	Empleados de contabilidad y cálculo de costos
		4312	Empleados de servicios estadísticos, financieros y de seguros
		4313	Empleados encargados de las nóminas
		4321	Empleados de control de abastecimientos e inventario
		4322	Empleados de servicios de apoyo a la producción
		4323	Empleados de servicios de transporte
		4411	Empleados de bibliotecas
		4412	Empleados de servicios de correos
		4413	Codificadores de datos, correctores de pruebas de imprenta y afines
		4414	Escribientes públicos y afines
		4415	Empleados de archivos
		4416	Empleados del servicio de personal
		4419	Personal de apoyo administrativo no clasificado bajo otros epígrafes
		5111	Auxiliares de servicio de abordó
		5112	Revisores y cobradores de los transportes públicos
		5113	Guías de turismo
		5120	Cocineros
		5131	Camareros de mesas
		5132	Camareros de barra
		5141	Peluqueros
		5142	Especialistas en tratamientos de belleza y afines
		5151	Supervisores de mantenimiento y limpieza en oficinas, hoteles y otros establecimientos
		5152	Ecónomos y mayordomos domésticos
		5153	Conserjes
		5161	Astrólogos, adivinadores y afines
		5162	Acompañantes y ayudantes de cámara
		5163	Personal de pompas fúnebres y embalsamadores
		5164	Cuidadores de animales
		5165	Instructores de autoescuela
		5169	Trabajadores de servicios personales no clasificados bajo otros epígrafes
		5211	Vendedores de quioscos y de puestos de mercado
		5212	Vendedores ambulantes de productos comestibles
		5221	Comerciantes de tiendas
		5222	Supervisores de tiendas y almacenes
		5223	Asistentes de venta de tiendas y almacenes

(Continuación de Cuadro 5)		
Código	Descripción del código	Etiqueta
		5230 Cajeros y expendedores de billetes
		5241 Modelos de moda, arte y publicidad
		5242 Demostradores de tiendas
		5243 Vendedores puerta a puerta
		5244 Vendedores por teléfono
		5245 Expendedores de gasolineras
		5246 Vendedores de comidas al mostrador
		5249 Vendedores no clasificados bajo otros epígrafes
		5311 Cuidadores de niños
		5312 Auxiliares de maestros
		5321 Trabajadores de los cuidados personales en instituciones
		5322 Trabajadores de los cuidados personales a domicilio
		5329 Trabajadores de los cuidados personales en servicios de salud no clasificados bajo otros epígrafes
		5411 Bomberos
		5412 Policías
		5413 Guardianes de prisión
		5414 Guardias de protección
		5419 Personal de los servicios de protección no clasificados bajo otros epígrafes
		6111 Agricultores y trabajadores calificados de cultivos extensivos
		6112 Agricultores y trabajadores calificados de plantaciones de árboles y arbustos
		6113 Agricultores y trabajadores calificados de huertas, invernaderos, viveros y jardines
		6114 Agricultores y trabajadores calificados de cultivos mixtos
		6121 Criadores de ganado
		6122 Avicultores y trabajadores calificados de la avicultura
		6123 Apicultores y sericultores y trabajadores calificados de la apicultura y la sericultura
		6129 Criadores y trabajadores pecuarios calificados de la cría de animales no clasificados bajo otros epígrafes
		6130 Productores y trabajadores calificados de explotaciones agropecuarias mixtas cuya producción se destina al mercado
		6210 Trabajadores forestales calificados y afines
		6221 Trabajadores de explotaciones de acuicultura
		6222 Pescadores de agua dulce y en aguas costeras
		6223 Pescadores de alta mar
		6224 Cazadores y tramperos
		6310 Trabajadores agrícolas de subsistencia
		6320 Trabajadores pecuarios de subsistencia
		6330 Trabajadores agropecuarios de subsistencia
		6340 Pescadores, cazadores, tramperos y recolectores de subsistencia
		7111 Constructores de casas
		7112 Albañiles
		7113 Mamposteros, tronzadores, labrantes y grabadores de piedra
		7114 Operarios en cemento armado, enfoscadores y afines
		7115 Carpinteros de armar y de obra blanca
		7119 Oficiales y operarios de la construcción (obra gruesa) y afines no clasificados bajo otros epígrafes
		7121 Techadores
		7122 Parqueteros y colocadores de suelos
		7123 Revocadores
		7124 Instaladores de material aislante y de insonorización

Código	(Continuación de Cuadro 5)		Etiqueta
	Descripción del código	Valor	
		7125	Cristaleros
		7126	Fontaneros e instaladores de tuberías
		7127	Mecánicos-montadores de instalaciones de refrigeración y climatización
		7131	Pintores y empapeladores
		7132	Barnizadores y afines
		7133	Limpiadores de fachadas y deshollinadores
		7211	Moldeadores y macheros
		7212	Soldadores y oxicortadores
		7213	Chapistas y caldereros
		7214	Montadores de estructuras metálicas
		7215	Aparejadores y empalmadores de cables
		7221	Herreros y forjadores
		7222	Herramientistas y afines
		7223	Reguladores y operadores de máquinas herramientas
		7224	Pulidores de metales y afiladores de herramientas
		7231	Mecánicos y reparadores de vehículos de motor
		7232	Mecánicos y reparadores de motores de avión
		7233	Mecánicos y reparadores de máquinas agrícolas e industriales
		7234	Reparadores de bicicletas y afines
		7311	Mecánicos y reparadores de instrumentos de precisión
		7312	Fabricantes y afinadores de instrumentos musicales
		7313	Joyereros, orfebres y plateros
		7314	Alfareros y afines (barro, arcilla y abrasivos)
		7315	Sopladores, modeladores, laminadores, cortadores y pulidores de vidrio
		7316	Redactores de carteles, pintores decorativos y grabadores
		7317	Artesanos en madera, cestería y materiales similares
		7318	Artesanos de los tejidos, el cuero y materiales similares
		7319	Artesanos no clasificados bajo otros epígrafes
		7321	Cajistas, tipógrafos y afines
		7322	Impresores
		7323	Encuadernadores y afines
		7411	Electricistas de obras y afines
		7412	Mecánicos y ajustadores electricistas
		7413	Instaladores y reparadores de líneas eléctricas
		7421	Mecánicos y reparadores en electrónica
		7422	Instaladores y reparadores en tecnología de la información y las comunicaciones
		7511	Carniceros, pescaderos y afines
		7512	Panaderos, pasteleros y confiteros
		7513	Operarios de la elaboración de productos lácteos
		7514	Operarios de la conservación de frutas, legumbres, verduras y afines
		7515	Catadores y clasificadores de alimentos y bebidas
		7516	Preparadores y elaboradores de tabaco y sus productos
		7521	Operarios del tratamiento de la madera
		7522	Ebanistas y afines
		7523	Reguladores y operadores de máquinas de labrar madera
		7531	Sastres, modistos, peleteros y sombrereros
		7532	Patronistas y cortadores de tela y afines
		7533	Costureros, bordadores y afines
		7534	Tapiceros, colchoneros y afines
		7535	Apelambradores, pellejeros y curtidores

(Continuación de Cuadro 5)		
Código	Descripción del código	Etiqueta
	Valor	
	7536	Zapateros y afines
	7541	Buzos
	7542	Dinamiteros y pegadores
	7543	Clasificadores y probadores de productos (excluyendo alimentos y bebidas)
	7544	Fumigadores y otros controladores de plagas y malas hierbas
	7549	Oficiales, operarios y artesanos de artes mecánicas y de otros oficios no clasificados bajo otros epígrafes
	8111	Mineros y operadores de instalaciones mineras
	8112	Operadores de instalaciones de procesamiento de minerales y rocas
	8113	Perforadores y sondistas de pozos y afines
	8114	Operadores de máquinas para fabricar cemento y otros productos minerales
	8121	Operadores de instalaciones de procesamiento de metales
	8122	Operadores de máquinas pulidoras, galvanizadoras y recubridoras de metales
	8131	Operadores de plantas y máquinas de productos químicos
	8132	Operadores de máquinas para fabricar productos fotográficos
	8141	Operadores de máquinas para fabricar productos de caucho
	8142	Operadores de máquinas para fabricar productos de material plástico
	8143	Operadores de máquinas para fabricar productos de papel
	8151	Operadores de máquinas de preparación de fibras, hilado y devanado
	8152	Operadores de telares y otras máquinas tejedoras
	8153	Operadores de máquinas de coser
	8154	Operadores de máquinas de blanqueamiento, teñido y limpieza de tejidos
	8155	Operadores de máquinas de tratamiento de pieles y cueros
	8156	Operadores de máquinas para la fabricación de calzado y afines
	8157	Operadores de máquinas lavarropas
	8159	Operadores de máquinas para fabricar productos textiles y artículos de piel y cuero no clasificados bajo otros epígrafes
	8160	Operadores de máquinas para elaborar alimentos y productos afines
	8171	Operadores de instalaciones para la preparación de pasta para papel y papel
	8172	Operadores de instalaciones de procesamiento de la madera
	8181	Operadores de instalaciones de vidriería y cerámica
	8182	Operadores de máquinas de vapor y calderas
	8183	Operadores de máquinas de embalaje, embotellamiento y etiquetado
	8189	Operadores de máquinas y de instalaciones fijas no clasificados bajo otros epígrafes
	8211	Ensambladores de maquinaria mecánica
	8212	Ensambladores de equipos eléctricos y electrónicos
	8219	Ensambladores no clasificados bajo otros epígrafes
	8311	Maquinistas de locomotoras
	8312	Guardafrenos, guardagujas y agentes de maniobras

Código	(Continuación de Cuadro 5)		Etiqueta
	Descripción del código	Valor	
		8321	Conductores de motocicletas
		8322	Conductores de automóviles, taxis y camionetas
		8331	Conductores de autobuses y tranvías
		8332	Conductores de camiones pesados
		8341	Operadores de maquinaria agrícola y forestal móvil
		8342	Operadores de máquinas de movimiento de tierras y afines
		8343	Operadores de grúas, aparatos elevadores y afines
		8344	Operadores de autoelevadoras
		8350	Marineros de cubierta y afines
		9111	Limpiadores y asistentes domésticos
		9112	Limpiadores y asistentes de oficinas, hoteles y otros establecimientos
		9113	Personal doméstico
		9121	Lavaderos y planchadores manuales
		9122	Lavadores de vehículos
		9123	Lavadores de ventanas
		9129	Otro personal de limpieza
		9211	Peones de explotaciones agrícolas
		9212	Peones de explotaciones ganaderas
		9213	Peones de explotaciones de cultivos mixtos y ganaderos
		9214	Peones de jardinería y horticultura
		9215	Peones forestales
		9216	Peones de pesca y acuicultura
		9311	Peones de minas y canteras
		9312	Peones de obras públicas y mantenimiento
		9313	Peones de la construcción de edificios
		9321	Empacadores manuales
		9329	Peones de la industria manufacturera no clasificados bajo otros epígrafes
		9331	Conductores de vehículos accionados a pedal o a brazo
		9332	Conductores de vehículos y máquinas de tracción animal
		9333	Peones de carga
		9334	Reponedores de estanterías
		9411	Cocineros de comidas rápidas
		9412	Ayudantes de cocina
		9510	Trabajadores ambulantes de servicios y afines
		9520	Vendedores ambulantes (excluyendo de comida)
		9611	Recolectores de basura y material reciclable
		9612	Clasificadores de desechos
		9613	Barrenderos y afines
		9621	Mensajeros, mandaderos, maleteros y repartidores
		9622	Personas que realizan trabajos varios
		9623	Recolectores de dinero en aparatos de venta automática y lectores de medidores
		9624	Acarreadores de agua y recolectores de leña
		9629	Ocupaciones elementales no clasificadas bajo otros epígrafes
		'0110	Oficiales de las fuerzas armadas
		'0210	Suboficiales de las fuerzas armadas
		'0310	Otros miembros de las fuerzas armadas
		9998	Ocupaciones no especificadas en las anteriores
		9999	Ocupación ignorada
		1	Quehaceres del hogar
		2	Vive de su renta o jubilación
		3	Estudiar
		4	Desempleado
		6	Otro

AGR\_DEDICA

A que se dedica el(a) agresor(a) (sin ingresos)

(Continuación de Cuadro 5)			
Código	Descripción del código	Valor	Etiqueta
		9	Ignorado
OTROS_AGRESORES	Total otros agresores	00 a 20	Total otros agresores
		99	Ignorado total otros agresores
AGR_OTROS_HOM	Otros agresores hombres (de 13 años en adelante)	00 a 09	Otros agresores hombres
		99	Ignorado otros agresores hombres
AGR_OTRAS_MUJ	Otras agresoras mujeres (de 13 años en adelante)	00 a 09	Otras agresoras mujeres
		99	Ignorado otras agresoras mujeres
AGR_OTROS_N_OS	Otros agresores niños (de 00 a 12 años)	00 a 09	Otros agresores niños
		99	Ignorado otros agresores niños
AGR_OTRAS_N_AS	Otras agresoras niñas (de 00 a 12 años)	00 a 09	Otras agresoras niñas
		99	Ignorado otras agresoras niñas
DATOS DE LA INSTITUCIÓN QUE REGISTRA EL HECHO			
INST_DENUN_HECHO	Nombre de la institución que registra el hecho	1	Ministerio Público
		2	Procuraduría General de la Nación
		3	Policía Nacional Civil
		4	Organismo Judicial
		5	Bufetes Populares
		6	Procuraduría de los Derechos Humanos
ORGANISMO - JURISDICCIONAL	Órgano jurisdiccional que recibe la denuncia	1	Juzgados de Paz
		2	Juzgados de Paz Comunitarios
		3	Juzgados de Paz Penal
		4	Juzgado de Paz Penal de Conocimiento a Prevención de Delitos de Narcotráfico, Defraudación y Contrabando Aduanero, Violencia Sexual, Explotación y Trata de Personas
		5	Juzgados de Paz Penal de Turno/24horas
		6	Juzgados de Paz Móvil
		7	Juzgado de Paz con Competencia Específica para la Protección en Materia de Violencia Intrafamiliar, y de Niñez y Adolescencia Amenazada o Violada en sus Derechos (de turno)
		8	Juzgados de Primera Instancia Penal, Narcoactividad y Delitos contra el Ambiente
		9	Juzgados de Primera Instancia Penal, Narcoactividad y Delitos contra el Ambiente de Turno/24 horas
		10	Juzgado de Turno de Primera Instancia Penal de Delitos de Femicidio y otras formas de de Violencia contra la contra la Mujer y Violencia Sexual, Explotación y Trata de Personas
		11	Juzgados de Primera Instancia Penal de Delitos de Femicidio y otras formas de Violencia contra la Mujer
		12	Tribunales de Sentencia Penal, Narcoactividad y Delitos contra el Ambiente
		13	Tribunales de Sentencia Penal de Delitos de Femicidio y otras formas de Violencia en contra de la Mujer
		14	Juzgados de Primera Instancia Ramo Mixto
		15	Juzgados de Primera Instancia de Familia
		16	Juzgados de Adolescentes en Conflicto con la Ley Penal
CONDUCTENTE	Se certificó lo conducente	1	Si
		2	No
		9	Ignorado
		1	Ley VIF ó Decreto 97-96
LEY_APLICABLE	Ley que se aplicó en el caso		

(Continuación de Cuadro 5)			
Código	Descripción del código	Valor	Etiqueta
		2	Ley VCM ó Decreto 22-2008
		3	Ambas leyes (VIF y VCM)
		4	Código Penal ó Decreto 17-73
		5	Otra
		6	Ley VIF y Código Penal
		9	Ignorado
ARTICULOVIF		001 a 014 999	Artículos de la ley VIF Artículos de la ley VIF igno- rado
ARTICULOVCM	Artículos aplicados (Se registraron hasta 4 artículos de una misma ley)	001 a 028 999	Artículos de la ley VCM Artículos de la ley VCM igno- rado
ARTICULOCODPEN		001 a 498 999	Artículos del Código Penal Artículos del Código Penal ignorado
ARTICULOTRAS		001 a 500 999	Artículos otras leyes Artículos otras leyes ignorado
MEDIDAS_SEGURIDAD	Medidas de seguridad otorgadas por el Organismo Judicial	1 2 9	Medidas otorgadas Medidas no otorgadas Ignorado
TIPO_MEDIDA	Tipo de medida de seguridad aplicada	a b c d e f g h i j k	Ordenar al presunto agresor que salga inmediatamente de la residencia común. Si se resiste, se utilizará la fuerza pública. Ordenar la asistencia obligatoria a instituciones con programas terapéutico-educativos, creados para ese fin. Ordenar el allanamiento de la morada cuando, por violencia intrafamiliar, se arriesgue gravemente la integridad física, sexual, patrimonial, o psicológica de cualquiera de sus habitantes. Prohibir que se introduzcan o se mantengan armas en la casa de habitación, cuando se utilicen para intimidar, amenazar o causar daño a alguna de las personas integrantes del grupo familiar. Decomisar las armas en posesión del presunto agresor, aún cuando tenga licencia de portación. Suspenderle provisionalmente al presunto agresor, la guarda y custodia de sus hijos e hijas menores de edad. Ordenar al presunto agresor abstenerse de interferir, en cualquier forma, en el ejercicio de la guarda, crianza y educación de sus hijos e hijas. Suspenderle al presunto agresor el derecho de visitar a sus hijos e hijas, en caso de agresión sexual contra menores de edad. Prohibir, al presunto agresor que perturbe o intimide a cualquier integrante del grupo familiar. Prohibir el acceso del presunto agresor al domicilio permanente o temporal de la persona agredida y a su lugar de trabajo o estudio. Fijar una obligación alimentaria provisional, de conformidad con lo establecido en el Código Civil.

(Continuación de Cuadro 5)			
Código	Descripción del código	Valor	Etiqueta
		l	Disponer el embargo preventivo de bienes, del presunto agresor. Para aplicar esta medida no será necesario ningún depósito de garantía. A juicio de la autoridad judicial competente, el embargo recaerá sobre la casa de habitación familiar y sobre los bienes necesarios para respaldar la obligación alimentaria en favor de la persona agredida y los dependientes que correspondan, conforme a la ley.
		m	Levantar un inventario de los bienes muebles existentes en el núcleo habitacional, en particular el menaje de casa u otros que le sirvan como medio de trabajo a la persona agredida.
		n	Otorgar el uso exclusivo, por un plazo determinado, del menaje de casa a la persona agredida. Deberán salvaguardarse especialmente la vivienda y el menaje amparado al régimen de patrimonio familiar.
		ñ	Ordenar, al presunto agresor, que se abstenga de interferir en el uso y disfrute de los instrumentos de trabajo de la persona agredida. Cuando ésta tenga sesenta (60) años o más o sea discapacitada, el presunto agresor no deberá interferir en el uso y disfrute de instrumentos indispensables para que la víctima pueda valerse por sí misma o integrarse a la sociedad.
		o	Ordenar al presunto agresor, la reparación en dinero efectivo de los daños ocasionados a la persona agredida, o a los bienes que le sean indispensables para continuar su vida normal. Se incluyen gastos de traslado, reparaciones a la propiedad, alojamiento y gastos médicos. El monto se hará efectivo en la forma y procedimiento que la autoridad judicial estime conveniente para garantizar que la misma sea cumplida.
		z	Se ignora la medida de seguridad
		1	Juzgados de Paz
		2	Juzgados de Paz Comunitarios
		3	Juzgados de Paz Penal
		4	Juzgado de Paz Penal de Conocimiento a Prevención de Delitos de Narcotráfico, Defraudación y Contrabando Aduanero, Violencia Sexual, Explotación y Trata de Personas
ORGANISMO_REMITE	Órgano jurisdiccional a donde se remitió	5	Juzgados de Paz Penal de Turno/24horas
		6	Juzgados de Paz Móvil
		7	Juzgado de Paz con Competencia Específica para la Protección en Materia de Violencia Intrafamiliar, y de Niñez y Adolescencia Amenazada o Violada en sus Derechos (de turno)
		8	Juzgados de Primera Instancia Penal, Narcoactividad y Delitos contra el Ambiente
		9	Juzgados de Primera Instancia Penal, Narcoactividad y Delitos contra el Ambiente de Turno/24 horas

(Continuación de Cuadro 5)			
Código	Descripción del código	Valor	Etiqueta
		10	Juzgado de Turno de Primera Instancia Penal de Delitos de Femicidio y otras formas de Violencia contra la Mujer y Violencia Sexual, Explotación y Trata de Personas
		11	Juzgados de Primera Instancia Penal de Delitos de Femicidio y otras formas de Violencia en contra de la Mujer
		12	Tribunales de Sentencia Penal, Narcoactividad y Delitos contra el Ambiente
		13	Tribunales de Sentencia Penal de Delitos de Femicidio y otras formas de Violencia en contra de la Mujer
		14	Juzgados de Primera Instancia Ramo Mixto
		15	Juzgados de Primera Instancia de Familia
		16	Juzgados de Adolescentes en Conflicto con la Ley Penal
		17	Ministerio Público (como órgano estatal)
		18	Este mismo órgano
		99	Ignorado

<sup>1</sup>ANO\_EMISION indica 2020, pero este es incorrecto

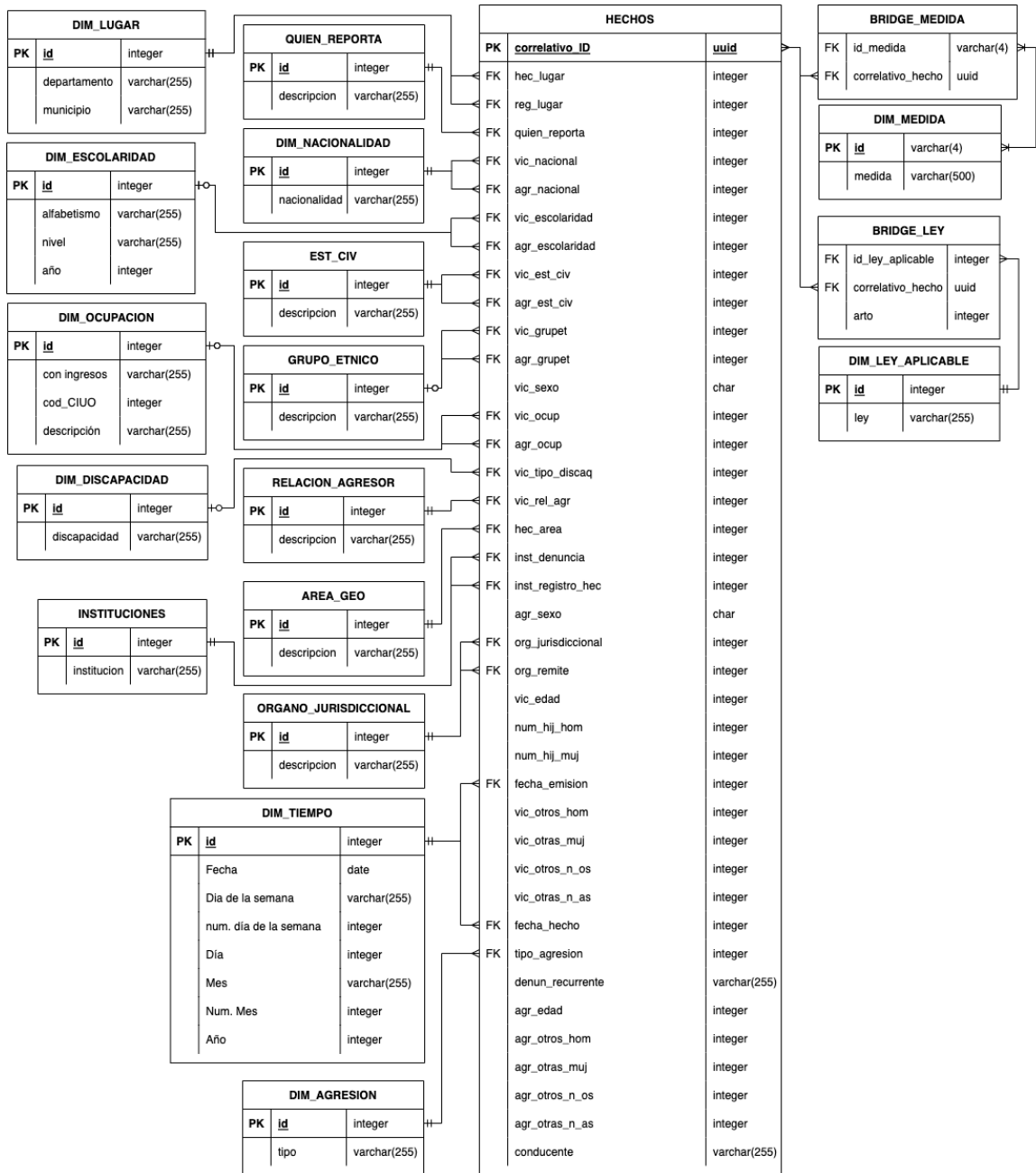


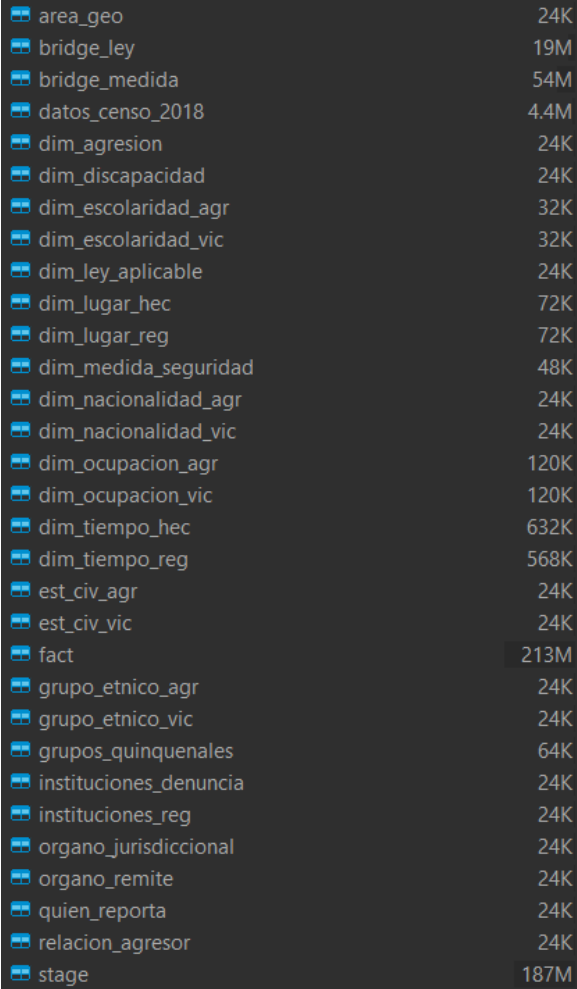
Figura 15: Modelo de datos preliminar.

## 12.1. Anexo técnico

Programas utilizados:

1. pgAdmin 4 versión 6.10
2. PostgreSQL 14.4, compilado por Visual C++ build 1914, 64-bit
3. Power BI versión 2.105.923.0 32-bit
4. R versión 4.2.1
5. RStudio 2022.07.0 Build 548

Tamaño total de la base de datos, incluyendo tablas utilizadas para visualizaciones: 491 MB



area_geo	24K
bridge_ley	19M
bridge_medida	54M
datos_censo_2018	4.4M
dim_agresion	24K
dim_discapacidad	24K
dim_escolaridad_agr	32K
dim_escolaridad_vic	32K
dim_ley_aplicable	24K
dim_lugar_hec	72K
dim_lugar_reg	72K
dim_medida_seguridad	48K
dim_nacionalidad_agr	24K
dim_nacionalidad_vic	24K
dim_ocupacion_agr	120K
dim_ocupacion_vic	120K
dim_tiempo_hec	632K
dim_tiempo_reg	568K
est_civ_agr	24K
est_civ_vic	24K
fact	213M
grupo_etnico_agr	24K
grupo_etnico_vic	24K
grupos_quinquenales	64K
instituciones_denuncia	24K
instituciones_reg	24K
organo_jurisdiccional	24K
organo_remite	24K
quien_reporta	24K
relacion_agresor	24K
stage	187M

Figura 16: Tamaño de cada tabla en la base de datos

## 12.2. Códigos

Enlace a repositorio del proyecto: <https://github.com/Andrea-Arguello/ETL-INE-VIF>

Listing 12.1: Aplicación de Shiny para *staging* de archivos .sav

```

## Descomentar si los paquetes no están instalados
# install.packages(
#   c("drat",
#     "tidyverse",
#     "RPostgres",
#     "DBI",
#     "shiny",
#     "shinyjs"
#   ))
# drat::addRepo("JanMarvin")
# install.packages("readspss")

library(readspss)
library(utils)
library(shiny)
library(shinycssloaders)
library(shinyjs)
library(DBI)
library(RPostgres)
library(tidyverse)

if(interactive()) {
  ui <- fluidPage(
    shinyjs::useShinyjs(),
    sidebarLayout(
      sidebarPanel(
        fileInput("files",
                  "Escoja los archivos .sav a importar a PostgreSQL",
                  multiple = T,
                  accept=".sav",
                  buttonLabel = "Seleccionar ...",
                  placeholder = "No se han seleccionado archivos"),
        textInput("user",
                  "Usuario de PostgreSQL",
                  placeholder="Usuario_(ej. postgres)"),
        passwordInput("password",
                      "Contraseña",
                      placeholder="Contraseña"),
        textInput("host",
                  "Host",
                  placeholder="Host_(ej. localhost)"),
        textInput("port",
                  "Puerto",
                  placeholder="Puerto_(ej. 5432)"),
        textInput("dbname",
                  "Nombre de la BD",
                  placeholder="Nombre de la BD"),
        textInput("table",
                  "Nombre de la tabla",
                  placeholder="Nombre de la tabla"),
        radioButtons("exists",
                     "¿La tabla ya existe?",
                     choices=list(
                       "Sí_(deseo agregar estos datos)" = 1,
                       "No" = 2)),
        actionButton("upload",
                     "Subir a PostgreSQL"),
      ),
      mainPanel(
        tableOutput("contents"),
        verbatimTextOutput(
          "Muestra de datos a subir (primeros 3 de cada archivo)"
        ),
        shinycssloaders::withSpinner(tableOutput("toupload")),
        shinycssloaders::withSpinner(textOutput("result"),
                                     type="7")
      )
    )
  )

server <- function(input, output) {
  extcheck <- reactive({
    for(i in 1:length(input$files[,1])) {
      file <- input$files[[i, 'datapath']]
      ext <- tools::file_ext(file)
      req(file)
      validate(
        need(ext == "sav",
             sprintf("%s no es .sav", (file))
        )
      )
    }
  })
  toUpload <- reactive({
    req(input$files)
    bind_rows(lapply(
      input$files$datapath,
      function(x){
        head(read.sav(x, convert.factors = F), 3)
      })
    ))
  })
}

```

```

upload <- eventReactive(input$upload, {
  validate(need(length(input$files[,1]) > 0,
    "No_se_han_seleccionado_archivos"))
  validate(need(input$user,
    "Ingresar_usuario_de_PostgreSQL"))
  validate(need(input$password, "Ingresar_Contraseña"))
  validate(need(input$host, "Ingresar_Host"))
  validate(need(input$port, "Ingresar_Port"))
  validate(need(input$dbname,
    "Ingresar_nombre_de_la_BD"))
  validate(need(input$table,
    "Ingresar_nombre_de_la_tabla"))
  validate(need(input$exists,
    "Seleccione_si_la_tabla_ya_existe_o_no"))
  disable("upload")
  out <- tryCatch({
    drv = RPostgres::Postgres()
    con = DBI::dbConnect(drv,
      user=input$user,
      password=input$password,
      host=input$host,
      port=input$port,
      dbname=input$dbname)
    if (input$exists == "1") {
      tableexists = DBI::dbExistsTable(con, input$table)
      if (!tableexists) {
        showModal(modalDialog(
          title = "Error",
          "No_existe_una_tabla_en_la_BD_con_el_nombre_ingresado"
        ))
        stop(
          "No_existe_una_tabla_en_la_BD_con_el_nombre_ingresado"
        )
      }
    }
    savcondense = bind_rows(lapply(
      input$files$datapath,
      function(x){read.sav(x, convert.factors = F)}))
    drv = RPostgres::Postgres()
    con = DBI::dbConnect(drv,
      user=input$user,
      password=input$password,
      host=input$host,
      port=input$port,
      dbname=input$dbname)
    if (input$exists == "1") {
      DBI::dbAppendTable(con,
        input$table,
        savcondense,
        row.names=NULL)
    } else {
      DBI::dbWriteTable(con,
        input$table,
        savcondense,
        row.names=F,
        overwrite=F)
    }
    showModal(modalDialog(
      title = "Éxito",
      "Se_han_subido_los_datos_correctamente"
    ))
    "Se_han_subido_los_datos_correctamente"
  },
  finally={enable("upload")}
)
})
output$contents <- renderTable({
  extcheck()
  input$files$name
})
output$stoupload <- renderTable({
  toUpload()
})
output$result <- renderText({upload()})
}
}
shinyApp(ui, server)

```

**big data:** Conjuntos de datos demasiado grandes y/o complejos para ser analizados y tratados por *software* de aplicación de procesamiento de datos tradicional. 3, 10, 15

**DDL:** Lenguaje de definición de datos, subconjunto de SQL que describe datos y relaciones en una base de datos. 19

**script:** Secuencia de instrucciones o comandos que automatiza una tarea al ejecutarse por una computadora. 27

**SQL:** (Por sus siglas en inglés structured query language; lenguaje de consulta estructurada en español) lenguaje de programación desarrollado por IBM en 1974. 17

**TIC:** Tecnologías de la información y la comunicación. 3, 15

**UUID:** Identificador único universal, un número de 16 bytes. 18, 19