



Maestría en Explotación de Datos y Descubrimiento del Conocimiento

Universidad de Buenos Aires

TÍTULO DEL TRABAJO

---

Victoria Colombo

fecha

## Resumen

# Introducción

La violencia sexual comprende una multiplicidad de conductas o intentos de conductas, que van desde actos hasta comentarios sexuales, dirigidos contra la sexualidad de otra persona de manera coercitiva. El trabajo con datos sobre violencia sexual presenta complicaciones porque los datos suelen ser escasos o presentar gran cantidad de faltantes (Ferris, 2002, p. 150). Uno de los motivos es que las víctimas o su entorno a menudo se rehúsan a denunciar o participar en encuestas sobre este tipo de agresiones, o proveen información incompleta. Esto puede deberse a la vergüenza y el estigma social frecuentemente asociado no solo con la violencia sexual sino con la sexualidad en general, pero también a la falta de acceso a la justicia, al temor a las represalias por parte de los agresores, o el temor a que la denuncia no sea creída (Murphy-Oikonen, McQueen, Miller, Chambers, and Hiebert, 2022). Otros posibles motivos para la escasez y/o mala calidad de los datos pueden ser la falta de vías adecuadas para recabar esta información, o la negligencia o desconocimiento de procedimientos adecuados por parte de oficiales de policía encargados de recibir denuncias. A pesar de las dificultades en la recolección de datos, diversos estudios a nivel mundial logran identificar patrones frecuentes en la violencia sexual. Para este trabajo, resultan relevantes dos de ellos: la mayoría de las víctimas son mujeres, mientras que los perpetradores suelen ser hombres (Ferris, 2002, p. 149; Contreras et al., 2016, p. 15); y en la mayoría de los casos, los agresores son personas conocidas por las víctimas, como parejas, exparejas u otros conocidos (García-Moreno et al., 2005, p. 9, Unicef et al., 2018, p. 22, Ferris, 2002, p. 151).

La clasificación de las identidades de género de víctimas y perpetradores es compleja. Por un lado, muchos estudios clasifican a las personas únicamente como hombres o mujeres, omitiendo las identidades de género disidentes.<sup>1</sup> Por otro lado, aunque se reportan pocos casos de violencia sexual contra hombres cisgénero, es probable que estén subrepresentados debido a los prejuicios y estigmas sociales sobre la masculinidad que dificultan las denuncias y el acceso a la justicia para estas víctimas (Ferris, 2002, p. 149). Analizar esas complejidades excede a este trabajo de especialización. En mi análisis las categorías de género de víctimas, victimarios y llamantes se limitan a las registradas en el *dataset*: hombre, mujer, y transgénero, sin especificar si es un hombre o una mujer transgénero. Reconozco esto como una limitación no solo de mi trabajo sino también de los datos disponibles.

La recopilación, sistematización, y análisis de datos sobre violencia sexual por parte de los Estados es crucial para planificar y llevar adelante políticas efectivas de prevención, asistencia, y erradicación de la violencia sexual. En Argentina, si bien no hay un sistema estatal único y centralizado de este tipo de información, existen entidades judiciales y programas estatales que, además de ofrecer auxilio, asistencia y/o acceso a la justicia, recaban datos sobre violencia sexual, y mantienen un registro público de ellos. Unos de esos programas es Las Víctimas contra las Violencias.

Desde el año 2016, en el marco del programa Las Víctimas contra las Violencias, dependiente del Ministerio de Justicia de la Nación, la línea 137 funciona las 24 horas del día para solicitar asistencia en casos de violencia sexual o familiar<sup>2</sup>. El programa cuenta con equipos de intervención de abogadas, psicólogas, y trabajadoras sociales. Al

---

<sup>1</sup>Entre los estudios e informes consultados para este trabajo, solamente el *Relevamiento de fuentes secundarias de datos sobre violencia sexual* de la Unidad Fiscal Especializada en Violencia contra las Mujeres (UFEM) (2023) menciona identidades de género cuando especifica que la violencia sexual “afecta particularmente a las mujeres cis y personas LGBTI+” (p.7).

<sup>2</sup>Además, desde 2020 cuenta también con el canal de *Whatsapp* (54911) 3133-1000.

recibir una llamada solicitando asistencia se coordina el envío de equipos móviles para proveer a la víctima, en base a las necesidades del caso, de contención emocional, acompañamiento a un hospital y/o a radicar una denuncia, y/o a un lugar seguro donde pueda alojarse (Ministerio de Justicia de la República Argentina, 2022).

Los registros de las llamadas a la línea 137 se encuentran digitalizados desde 2017 y están disponibles en el Portal de Datos Abiertos de la Justicia Argentina. Allí se encuentran publicados cuatro tipos de *datasets* por año: llamados e intervenciones domiciliarias por situaciones de violencia familiar, y llamados e intervenciones domiciliarias por situaciones de violencia sexual. Los registros no están exentos de los problemas frecuentes antes mencionados en los datos sobre violencia sexual, ya que presentan información faltante de dos maneras: celdas vacías en el caso de las variables numéricas de edad, y respuestas NS/NC (no sabe-no contesta) en lugar de SI o NO en el resto de las variables categóricas. Teniendo en cuenta la clasificación de datos faltantes que se origina en Rubin (1976), los datos faltantes en el *dataset* de llamados son posiblemente del tipo *missing at random* (MAR) y *missing not at random/ non-ignorable missing data* (MNAR). Es decir, o bien los datos faltan por motivos que tienen que ver con otras variables (MAR), o bien el valor de los datos que faltan está relacionado con el motivo mismo por el que faltan (MNAR).

En este trabajo analizo llamados para reportar violencia sexual a la línea 137 entre 2017 y 2021, e intento predecir valores faltantes de la variable “víctima convive con el agresor”.

## Datos

Para este trabajo descargué del Portal de Datos Abiertos mencionado arriba 5 *datasets* en formato *csv* de llamados a la línea 137 para solicitar asistencia por violencia sexual. Los archivos pertenecen, a razón de uno de por año, al período entre enero de 2017 y julio de 2021. En dos variables los nombres (no así el contenido) diferían en los distintos años, por lo tanto realicé las siguientes modificaciones para poder unificar los 5 archivos en un solo *dataset*:

- Eliminé la variable *caso\_id*, que solo existe a partir de 2020.
- El nombre de la variable *llamado-provincia\_id*, llamada hasta el año 2019: *llamado-provincia\_indec\_id*, fue normalizado en 2017, 2018 y 2019 a *llamado-provincia\_id*.

El *dataset* final unificado consta de 19143 observaciones y 54 variables, en su mayoría categóricas, que aportan información sobre la víctima, la persona denunciante, el contexto del hecho y el tipo de violencia sufrida. En el cuadro 1 se puede ver un detalle de las variables y su tipo.

Acá nota al pie sobre python<sup>3</sup>.

El siguiente paso fue limpiar los datos de errores de tipeo, valores cargados con sinónimos o abreviaturas, y errores ortográficos:

- Normalicé en la variable *victima\_vinculo\_agresor* el valor “Ex pareja de la víctima” que aparecía también como “Ex pareja”, “Ex-pareja de la víctima” y “Expareja de la víctima”, otro tanto hice con “Pareja de la víctima”.
- Normalicé para todas las variables pertinentes los valores SI, NO, y NS/NC dejándolos en mayúscula, ya que aparecían en distintos formatos: minúscula, mayúscula inicial, con tildes en SI, etc.

<sup>3</sup>El análisis exploratorio y el resto del trabajo con datos fue y será realizado en Python

Cuadro 1: Resumen de las variables.

Descriptor	Tipo variable	Variable(s)
Víctima	Cuantitativa	victima_edad
	Cualitativa	victima_genero, victima_nacionalidad, victima_discapacidad, victima_vinculo_agresor, victima_convive_agresor, victima_a_resguardo
Llamante	Cuantitativa	llamante_edad
	Cualitativa	llamante_genero, llamante_vinculo
Llamado	Ordinal	llamado_fecha_hora
	Cualitativa	caso_id, llamado_provincia, llamado_provincia_id, caso_judicializado, hecho_lugar
Violencia sexual	Cualitativa	vs_violacion_via_vaginal, vs_violacion_via_anal, vs_violacion_via_oral, vs_tentativa_violacion, vs_tocamiento_sexual, vs_intento_tocamiento, vs_intento_violacion_tercera_persona, vs_grooming, vs_exhibicionismo, vs_amenazas_verbales_contenido_sexual, vs_explotacion_sexual, vs_explotacion_sexual_comercial, vs_explotacion_sexual_viajes_turismo, vs_sospecha_trata_personas_fines_sexuales, vs_existencia_facilitador_corrupcion_nnya, vs_obligacion_sacarse_fotos_pornograficas, vs_eyaculacion_partes_cuerpo, vs_acoso_sexual, vs_iniciacion_sexual_forzada_inducida, vs_otra_forma_violencia_sexual, vs_no_sabe_no_contesta
Otras violencias	Cualitativa	ofv_sentimiento_amenaza, ofv_amenazas_explicitas, ofv_violencia_fisica, ofv_intento_ahorcar, ofv_intento_quemar, ofv_intento_ahogar, ofv_amenaza_muerte, ofv_uso_sustancias_psicoactivas, ofv_intento_privacion_libertad, ofv_privacion_libertad, ofv_uso_arma_blanca, ofv_uso_arma_fuego, ofv_enganio_seducion, ofv_intento_matar, ofv_uso_animal_victimizar, ofv_grooming, ofv_otra_forma_violencia, ofv_no_sabe_no_contesta

- Normalicé todos los valores “Sin datos” a NS/NC por cosniderarlos equivalentes.
- Qité espacios de comienzo y final de *strings* para solucionar problemas del tipo “Madre”  $\neq$  “ Madre”
- Convertí en la variable llamante\_vinculo el valor “Vecino” a “Vecina/o” ya que no necesariamente se refiere unívocamente a personas de género masculino.
- Normalicé en llamado\_provincia “Ciudad Autónoma de Buenos Aires” a “CABA” ya que ambos valores aparecían.
- Corregí en llamado\_provincia las instancias de “Santa Fé” a “Santa Fe”.

Luego, las variables numéricas *victima\_edad* y *llamante\_edad* tenían errores de carga evidentes, aparecían edades como: 125, 221, 324. Eliminé esos valores y dejé las celdas vacías como datos faltantes de edad.

A modo de análisis exploratorio, realicé histogramas univariados para ver la frecuencia de las categorías de las variables: *victima\_genero*, *victima\_edad*, *victima\_convive\_agresor*, *victima\_vinculo\_agresor*, y *hecho\_lugar*. Además, construí la variable agresor conocido no conocido para explorar la dicotomía extraños-conocidos en relación a las víctimas y los agresores. La variable estación del año, fin de semana y momento del día para explorar frecuencia de los llamados. Exploré la variable provincia para ver la cantidad de llamados por región.

relación entre convive y otras variables como edad de la víctima y de quien llama y tmb quién llama (instituciones etc)

histogramas y gráficos exploratorios de esas variables

Datos faltantes

datos faltantes en forma de NS/NC y en forma de datos vacíos. Dónde están y cuántos son. Cuántos en la variable de interés victima convive agresor. posible relación de datos faltantes de edad con quien llama tanto en edad de la victima como en edad de quien llama relación entre los faltante de edad y NSNC en convive

Las variables de vs y ofv toman los valores SI NO.

En general en todas las variables de violencia hay muchos más NO que sí.

## Metodología

El objetivo es imputar los NS/NC de convive como si o no. Primero quiero ver si utilizando un método de visualización que permita plasmar interacciones entre variables reduciendo las dimensiones a 2 para poder ser representado gráficamente se pueden detectar asociaciones lineales entre los valores de convive y el resto de las variables. Para esto usé NMDS. Luego, para intentar predecir los NSNC como si o no usé SVM.

A modo de un segundo preprocesamiento para poder llevar a cabo estas tareas, decidí reducir las dimensiones del dataset a mano primero agrupando variables y reduciendo la cantidad de categorías en algunas otras variables. LA REDUCCIÓN DE LA CARDINALIDAD DE algunas variables es especialmente útil para la aplicación de SVM ya que esta conlleva encodear las features y para uno de los encoders elegidos, one-hot, la alta cardinalidad de features puede resultar problemática.

Habiendo hecho los gráficos de más arriba para explorar posibles interacciones entre dos o tres variables, me pareció valioso explorar más dimensiones y plasmarlo en dos dimensiones. Para eso apliqué NMDS

Reducción manual de dimensiones Algunas resultan muy poco informativas como vs explotacion sexual viajes turismo (0,02) y ofv intento matar (0,01) resultan muy poco informativas (ocurrencia de 0,02 y 0,01).

Para intentar

Explicar agrupamiento de variables.

quizás  
ner cuántos  
tenían  
problem

## Resultados

### Discusión y conclusiones

cruzamiento de datos ovd líneas de asistencia, observatorio de género. acceso y análisis de datos extensivo a provincias, no solo benos aires

en *La guerra contra las mujeres*, (2016), Rita Segato habla de la violencia sexual como algo siempre dirigido hacia cuerpos femeninos y *feminizados* (resaltado propio). Con esto último quiere decir cuerpos percibidos o contruidos por los abusadores como femeninos con respecto a posiciones de poder: menores, débiles, racializados, pertenecientes a disidencias sexuales. Esto se condice con datos sobre la mayor incidencia de la violencia sexual contra identidades masculinas durante la niñez y la adolescencia, es decir, en períodos en que los cuerpos y los sujetos son más vulnerables, y por lo tanto, también percibidos como feminizados (Contreras, Both, Guedes, and Dartnall, 2016; Unidad Fiscal Especializada en Violencia contra las Mujeres (UFEM), 2023; Ferris, 2002).

Si las denuncias de violencia sexual contra disidencias de género representan una minoría en los datos, ¿quiere decir esto que esas personas sufren menos violencia sexual?, ¿O quiere decir que, como minoría social, están subrepresentados en general y que tienen menos acceso a la justicia?

## Referencias

- Juan Manuel Contreras, S Both, A Guedes, and E Dartnall. Violencia sexual en latinoamérica y el caribe: análisis de datos secundarios. iniciativa de investigación sobre la violencia sexual., 2016.
- Lorraine E Ferris. World report on violence and health: Edited by etienne g. krug, linda l. dahlberg, james a. mercy, anthony zwi and rafael lozano. geneva: World health organization, 2002. *Canadian Journal of Public Health= Revue Canadienne de Santé Publique*, 93(6):451, 2002.
- Claudia García-Moreno, Henrica AFM Jansen, Mary Ellsberg, Lori Heise, Charlotte Watts, et al. *WHO multi-country study on women's health and domestic violence against women*. World Health Organization, 2005.
- Ministerio de Justicia de la República Argentina. Nueva Línea 137: ampliación de servicios de atención contra las violencias y para el acceso a derechos. <https://www.argentina.gob.ar/noticias/nueva-linea-137-ampliacion-de-servicios-de-atencion-contras-las-violencias-y-para-el-acceso>, 2022.
- Jodie Murphy-Oikonen, Karen McQueen, Ainsley Miller, Lori Chambers, and Alexa Hiebert. Unfounded sexual assault: Women's experiences of not being believed by the police. *Journal of interpersonal violence*, 37(11-12): NP8916–NP8940, 2022.
- Donald B Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 1976.
- Rita Laura Segato. *La guerra contra las mujeres*. Traficantes de sueños, 2016.
- Unicef et al. Un análisis de los datos del programa «las víctimas contra las violencias», 2018.
- Unidad Fiscal Especializada en Violencia contra las Mujeres (UFEM). Relevamiento de fuentes secundarias de datos sobre violencia sexual Información a nivel nacional y de la Ciudad Autónoma de Buenos Aires. <https://www.mpf.gob.ar/ufem/violencia-sexual/>, 2023.



## Anexo