

UNIVERSIDAD DEL VALLE DE GUATEMALA

Minería de datos

Sección 30

Ing. Leonel Guillen del Valle



Análisis exploratorio

Bryan Carlos Roberto España Machorro - 21550

Javier Alejandro Prado Ramirez - 21486

Diego Jose Valdez Soto - 21328

GUATEMALA, 05 de febrero 2024

Situación problemática

La violencia intrafamiliar es un problema creciente que plantea serias preocupaciones en nuestra comunidad. A pesar de los esfuerzos de concientización y prevención, los informes de casos de violencia doméstica continúan en aumento, lo que indica una urgente necesidad de acción.

Este fenómeno afecta a individuos de todas las edades, géneros y clases sociales, generando un impacto devastador en la seguridad y bienestar de los miembros de las familias afectadas. Desde daños físicos y emocionales hasta secuelas psicológicas a largo plazo, la violencia intrafamiliar deja cicatrices profundas en quienes la padecen.

Problema científico

El problema científico que surge de la situación planteada es comprender en profundidad los factores subyacentes que contribuyen al aumento de la incidencia de violencia intrafamiliar en nuestra comunidad. Es fundamental investigar y analizar estos factores para identificar patrones, tendencias y correlaciones significativas que nos ayuden a comprender mejor las dinámicas de la violencia doméstica. Esto implica no solo identificar los factores de riesgo que predisponen a las personas a ser víctimas o perpetradores de violencia intrafamiliar, sino también comprender cómo estos factores interactúan entre sí y cómo pueden influir en la dinámica de las relaciones familiares. Solo a través de una comprensión profunda de estos aspectos podremos desarrollar estrategias preventivas y de intervención efectivas para abordar el problema de la violencia intrafamiliar en nuestra comunidad.

Objetivos

Objetivo general:

Identificar y comprender los factores de riesgo asociados con la violencia intrafamiliar utilizando análisis de datos, con el fin de desarrollar estrategias efectivas de prevención e intervención.

Objetivo específico:

- Investigar la relación entre las características sociodemográficas de las víctimas y los agresores y la incidencia de violencia intrafamiliar. Esto incluirá analizar variables como la edad, género, estado civil, nivel educativo, ocupación, número de hijos, entre otros, para identificar patrones y tendencias significativas.
 - Evaluar la asociación entre los diferentes tipos de violencia reportados (física, verbal, económica, psicológica, sexual) y las características contextuales de los incidentes, como la hora del día, el día de la semana y el lugar de ocurrencia. Se llevará a cabo un análisis detallado de la frecuencia y distribución de estos tipos de violencia para comprender mejor sus dinámicas y patrones.
- 1) Explore los datos para encontrar preguntas interesantes y guías de investigación.
Para esto:
 - a) Comience describiendo cuantas variables y observaciones tiene disponibles, el tipo de cada una de las variables.

```

Número de observaciones: 1000
Número de variables: 36

Tipos de variables:
fecha_de_apertura          object
dia_de_apertura            object
fecha_ocurrencia_hechos    object
dia_ocurrencia             object
hora_militar_ocurrencia_hechos object
conforman_unidad_domestica object
fisica                     object
verbal                     object
economica                  object
psicologica                object
sexual                     object
genero_m_f_victima         object
edad_victima               float64
victima_conflicto_armado    object
etnia_victima              object
estado_civil_victima        object
escolaridad_victima         object
corregimiento_victima       object
comuna_de_la_victima        object
barrio_victima              object
ocupacion_victima           object
...
comuna_agresor              object
barrio                      object
tipo_vivienda               object
dtype: object

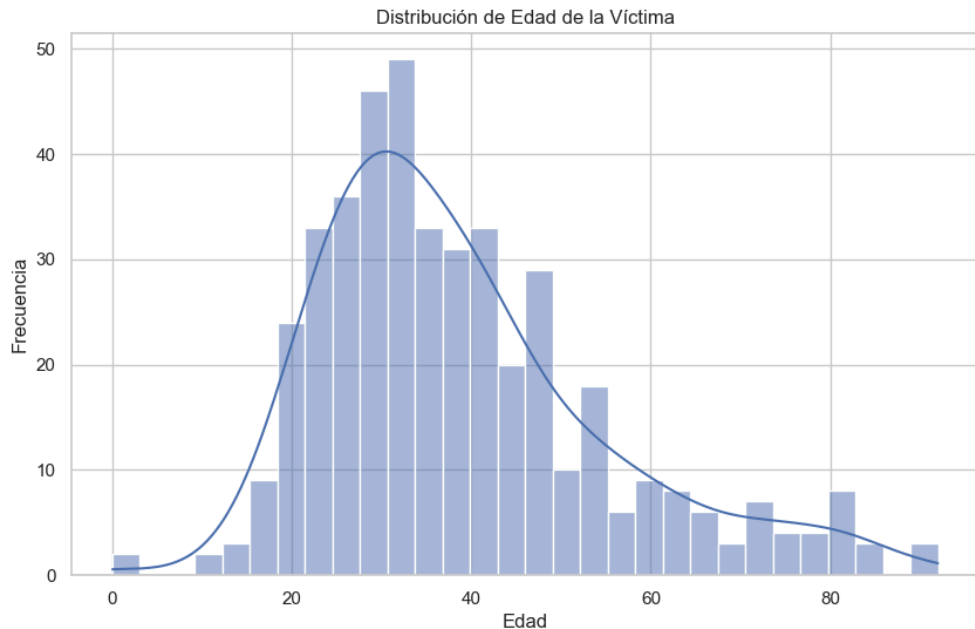
```

- b) Haga un resumen de las variables numéricas e investigue si siguen una distribución normal y tablas de frecuencia para las variables categóricas, escriba lo que vaya encontrando.

```
Variables numéricas:  
edad_victima          float64  
no_hijos_victima      float64  
numero_hermanos_victima float64  
no_de_hijos_agresor   float64  
edad_agresor          float64  
dtype: object
```

Variable 'edad_victima'

```
Resumen estadístico de edad_victima:  
count    439.000000  
mean      38.772210  
std       16.165611  
min        0.000000  
25%       28.000000  
50%       35.000000  
75%       46.000000  
max       92.000000  
Name: edad_victima, dtype: float64
```



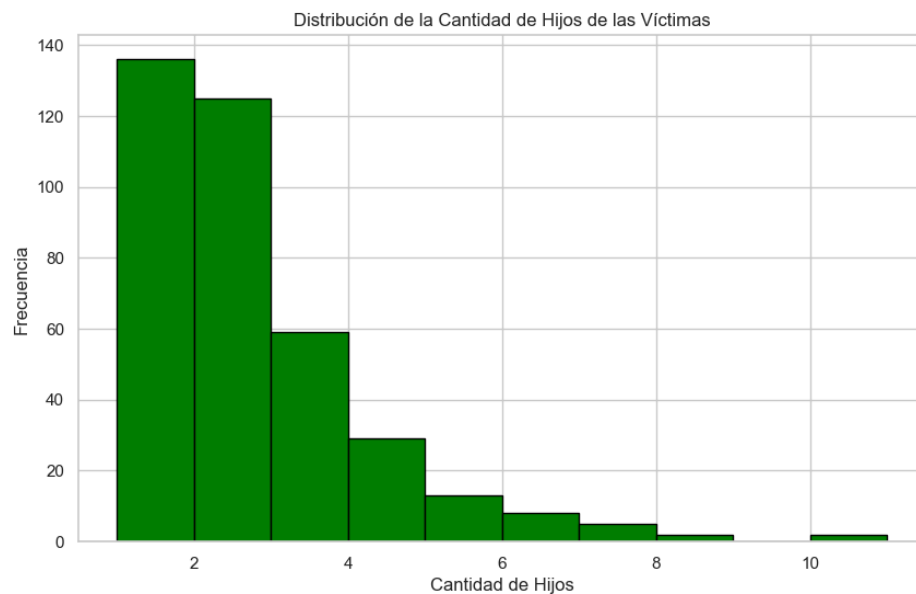
Analisis

Proporcionamos una vista rápida de la centralidad y la dispersión. La media y la mediana pueden darte una idea de la ubicación central, mientras que la desviación estándar te informa sobre la variabilidad. Los valores mínimo y máximo te indican el rango de edad de las víctimas.

Histograma: Ayuda a visualizar la distribución general de las edades. Por ejemplo, si la mayoría de las edades se agrupan en un rango específico o si hay una distribución uniforme a lo largo de varios rangos de edad. También puedes identificar la presencia de sesgos en la distribución.

Variable 'no_hijos_victima':

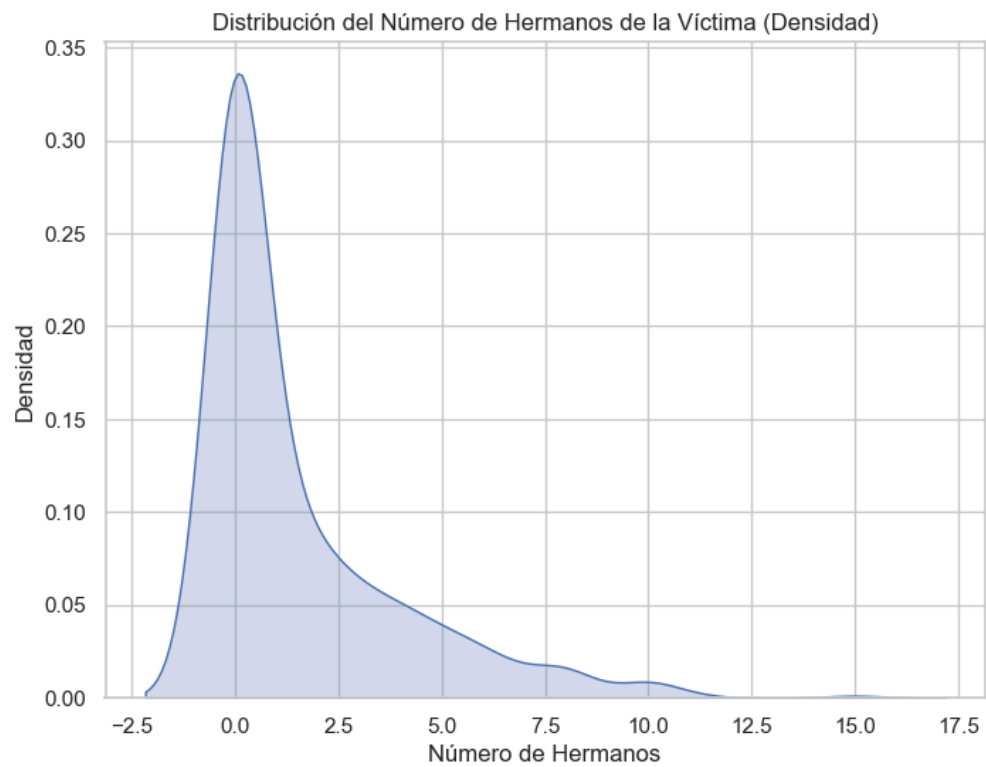
```
Resumen estadístico de no_hijos_victima:  
count      439.000000  
mean        1.965831  
std         1.599434  
min         0.000000  
25%         1.000000  
50%         2.000000  
75%         3.000000  
max         10.000000  
Name: no_hijos_victima, dtype: float64
```



Vemos que la mayoría de las personas que tienden a ser víctimas tienen entre 2-3 hijos.

variable 'numero_hermanos_victima':

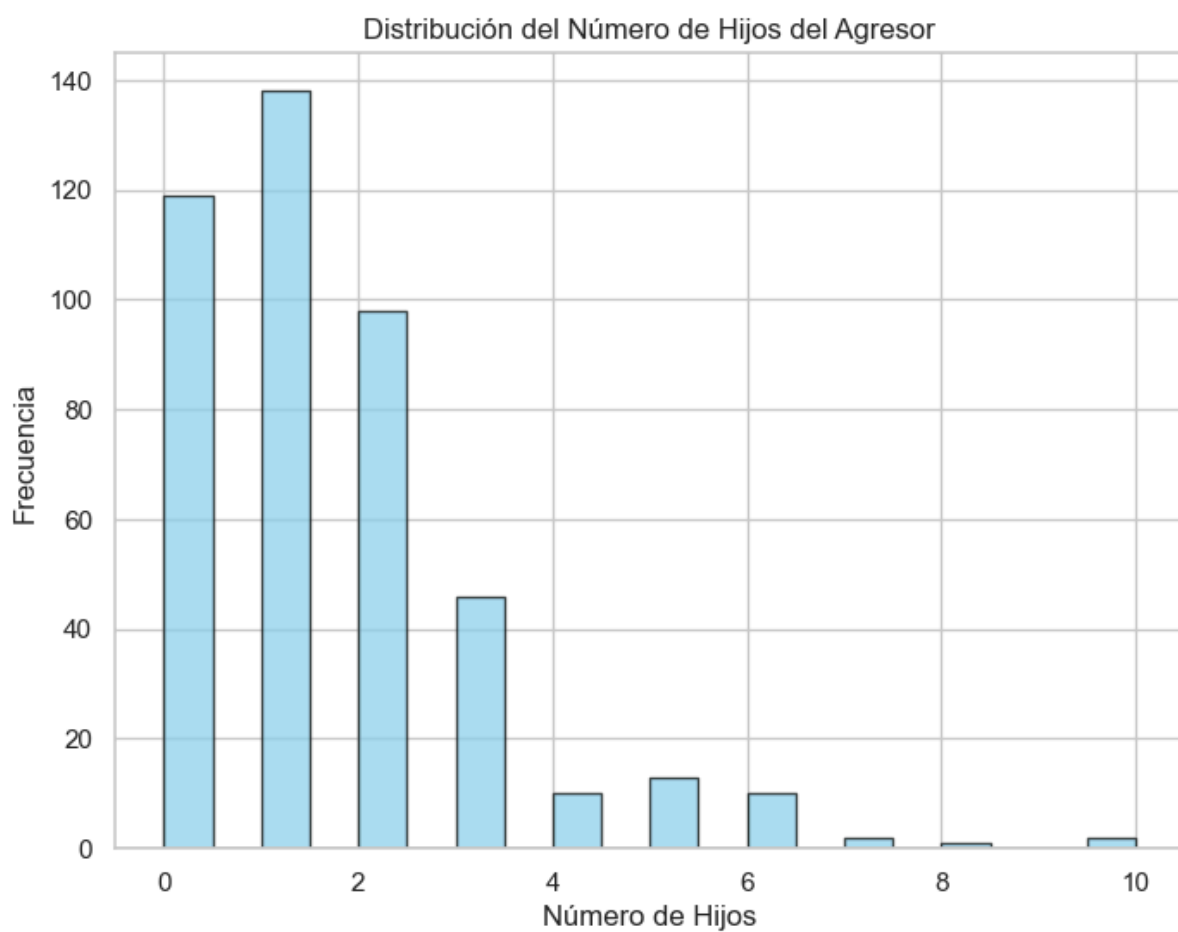
```
Resumen estadístico de numero_hermanos_victima:  
count      439.000000  
mean        1.546697  
std         2.444261  
min         0.000000  
25%         0.000000  
50%         0.000000  
75%         2.000000  
max         15.000000  
Name: numero_hermanos_victima, dtype: float64
```



Podemos ver que muchas de las víctimas tienen entre 0 a 2 hermanos

variable 'no_de_hijos_agresor'

```
Resumen estadístico de no_de_hijos_agresor:  
count      439.000000  
mean        1.546697  
std         1.585316  
min         0.000000  
25%         0.000000  
50%         1.000000  
75%         2.000000  
max         10.000000  
Name: no_de_hijos_agresor, dtype: float64
```



Nos damos cuenta que el agresor pormedio tiene de 0 a 2 hijos, lo que hace que podemos quitar la idea que por el estrés de tener muchos hijos hace que sea por el estrés que exista violencia intrafamiliar