

Proyecto 1

Luis Pedro Lira 23669, Fernando Rocha 23501, Juan Francisco Martínez 23617

2026-02-02

Introducción

La violencia intrafamiliar constituye un problema social de alta relevancia debido a su impacto en la estabilidad emocional, social y económica de los hogares. Por otro lado, el divorcio representa una consecuencia legal de conflictos persistentes dentro del núcleo familiar, los cuales pueden estar relacionados con dinámicas de violencia y deterioro de la convivencia.

En este proyecto se analiza la posible relación entre los divorcios y la violencia intrafamiliar utilizando datos oficiales correspondientes al año 2022. A través de técnicas de análisis exploratorio de datos y minería de datos, se busca identificar patrones, comportamientos y relaciones entre ambas problemáticas, con el fin de obtener una comprensión más profunda de su interacción a nivel agregado.

El análisis se centra en una exploración descriptiva inicial de los conjuntos de datos, la caracterización de sus variables y la evaluación de posibles relaciones estadísticas entre los indicadores de divorcio y violencia intrafamiliar.

Objetivos

Objetivo General Analizar la relación entre los divorcios y la violencia intrafamiliar mediante técnicas de análisis exploratorio y minería de datos, utilizando información correspondiente al año 2022.

Objetivos específicos - Describir las características principales de los conjuntos de datos de divorcios y violencia intrafamiliar. - Identificar y clasificar las variables numéricas y categóricas presentes en cada conjunto de datos. - Realizar un análisis exploratorio de las variables relevantes mediante medidas descriptivas y representaciones gráficas. - Evaluar posibles relaciones entre indicadores de divorcio y violencia intrafamiliar.

Descripción general de los datos

Para el desarrollo del presente proyecto se utilizaron dos conjuntos de datos correspondientes al año 2022. El primero contiene información relacionada con los procesos de divorcio, mientras que el segundo recopila registros asociados a casos de violencia intrafamiliar. Ambos conjuntos de datos fueron obtenidos a partir de fuentes oficiales y procesados previamente para su uso en el análisis.

Descripción de las variables: Violencia intrafamiliar

El conjunto de datos de violencia intrafamiliar contiene información detallada sobre las características sociodemográficas de las víctimas y agresores, así como aspectos relacionados con el hecho de violencia reportado. Las variables incluidas permiten analizar dimensiones como edad, sexo, escolaridad, ocupación, estado civil y relación entre víctima y agresor, entre otras.

La mayoría de las variables en el conjunto de datos corresponden a codificaciones numéricas que representan categorías, mientras que otras variables como la edad y las fechas corresponden a valores numéricos continuos o discretos.

El conjunto de datos de violencia intrafamiliar presenta valores faltantes (NA) en diversas variables, los cuales corresponden principalmente a información no aplicable a todos los registros. Adicionalmente, algunas variables contienen códigos numéricos utilizados para representar valores no especificados, los cuales serán tratados como valores faltantes durante el análisis exploratorio.

Descripción de las variables: Divorcios

El conjunto de datos de divorcios contiene información asociada a los procesos de disolución del vínculo matrimonial, incluyendo características temporales, geográficas y sociodemográficas de las personas involucradas. Las variables disponibles permiten analizar aspectos como la edad, escolaridad, ocupación y nacionalidad de los cónyuges, así como el lugar y fecha de ocurrencia del divorcio.

Es importante destacar que el conjunto de datos contiene valores faltantes (NA) en algunas variables, los cuales corresponden principalmente a información no aplicable o no disponible para ciertos registros. Adicionalmente, algunas variables presentan valores codificados para representar información no especificada, los cuales serán tratados adecuadamente durante el análisis exploratorio.

Exploración de variables numéricas

En esta sección se analizan las principales variables numéricas de los conjuntos de datos de violencia intrafamiliar y divorcios, utilizando medidas de tendencia central, medidas de orden y representaciones gráficas.

```
##      Min. 1st Qu. Median     Mean 3rd Qu.    Max.    NA's
##      1.00   24.00  31.00   33.28   39.00   98.00  5046
```

La edad de la víctima presenta un rango amplio, desde 1 hasta 98 años, lo cual evidencia que la violencia intrafamiliar afecta a personas de distintas etapas de la vida. La mediana de edad es de 31 años, mientras que la media es de aproximadamente 33 años, lo que indica una ligera asimetría positiva en la distribución. El 50% de las víctimas se concentra entre los 24 y 39 años de edad, sugiriendo que los adultos jóvenes representan el grupo más afectado. Se identificaron 5,046 registros sin información de edad, los cuales serán considerados como valores faltantes en el análisis.

```
##      Min. 1st Qu. Median     Mean 3rd Qu.    Max.    NA's
##      0.00   1.00   2.00    2.13    3.00   19.00  70366
```

El número total de hijos de la víctima presenta una mediana de 2 hijos, lo que indica que la mayoría de los casos corresponde a personas con familias pequeñas. La media es cercana a la mediana, lo cual sugiere una distribución relativamente equilibrada. El 75% de las víctimas tiene tres hijos o menos, mientras que los valores máximos observados corresponden a casos aislados de familias numerosas. Se identificó una proporción considerable de valores faltantes, los cuales fueron tratados adecuadamente durante el análisis exploratorio.

```
##      Min. 1st Qu. Median     Mean 3rd Qu.    Max.    NA's
##     15.0    26.0   31.0    32.8    38.0   81.0  34721
```

La edad de la mujer en los procesos de divorcio presenta una mediana de 31 años y una media de aproximadamente 33 años, lo que indica que los divorcios se concentran principalmente en edades adultas jóvenes. El 50% de los casos se encuentra entre los 26 y 38 años, mientras que los valores máximos corresponden a casos aislados de edades avanzadas. La distribución muestra una ligera asimetría positiva y se identifican valores faltantes que fueron considerados adecuadamente durante el análisis.

```
##      Min. 1st Qu. Median   Mean 3rd Qu.   Max.   NA's
##    15.00   29.00   34.00  36.02   41.00  98.00 34795
```

La edad del hombre en los procesos de divorcio presenta una mediana de 34 años y una media de aproximadamente 36 años, lo que sugiere que los hombres tienden a ser ligeramente mayores que las mujeres al momento del divorcio. El rango intercuartílico se ubica entre los 29 y 41 años, concentrando la mayoría de los casos en edades adultas. La distribución presenta una asimetría positiva y algunos valores extremos correspondientes a edades avanzadas.