

Violencia intrafamiliar y comunitaria en Nuevo León

Gabriela Sánchez Yepéz

Posgrado en Ingeniería de Sistemas
Facultad de Ingeniería Mecánica y Eléctrica
Universidad Autónoma de Nuevo León

Abstract

Se presenta un análisis de incidentes de violencia intrafamiliar y comunitaria en el estado de Nuevo León reportados durante el año 2018. Es importante identificar los grupos vulnerables y posibles causas al problema que permitan planear estrategias de prevención. Para realizar dicho análisis, se utilizan distintas herramientas que proporciona el lenguaje de programación Python.

Keywords:

1. Introducción

Este trabajo busca aplicar herramientas de ciencia de datos que permitan analizar el efecto de distintos factores presentes en incidentes de violencia intrafamiliar y comunitaria en el estado de Nuevo León, con el objetivo de identificar vulnerables y posibles causas al problema que permitan planear estrategias de prevención.

La estructura del artículo es la siguiente: *

2. Metodología

En esta sección se especifican las características de los datos con los que se trabaja así como las herramientas utilizadas para el análisis de dichos datos.

2.1. Datos

Los datos fueron proporcionados por la doctora Patricia L. Cerda Pérez, en formato `xlsx`.

Los datos se presentan en dos archivos, el primero contiene los incidentes de violencia intrafamiliar y comunitaria presentados durante los meses de enero a noviembre en el año 2018 y el segundo, los incidentes del resto del año, teniendo un total de 16 410 reportes.

Los registros contienen 47 columnas de las cuales únicamente 17 proporcionan información relevante, sin datos personales, que serán usados en el análisis. Estos registros proveen datos sobre lugar, fecha y hora de los incidentes, así como información sobre la(s) víctima(s) y agresor(es).

2.2. Herramientas

En esta subsección se explica la selección de herramientas para los distintos análisis que fueron realizados.

Preprocesamiento de los datos

Para realizar un preprocesamiento de los datos primero se utilizó la herramienta `BASH` que permitió convertir el archivo de los datos a un formato manejable, esto es, `csv`, utilizando la instrucción

```
ssconvert datos.xlsx vf.csv
```

La siguiente fase de preprocesamiento se realizó con la librería `PANDAS`. El primer paso es unir los datos en un solo archivo, eliminar la información que no se utiliza y cambiar los tipos de los datos para no tener problemas una vez que se empiece a procesar la información.

Estadística descriptiva y visualización de la información

El análisis estadístico básico así como una visualización de la información se realizó con distintas herramientas proporcionadas por la librería `PANDAS`.

Para poder trabajar con esta librería es necesario cargar los datos para que se lean en el formato que se requiere, esto se logra con

```
import pandas as pd
vf = pd.read_csv("vf.csv")
```

En el análisis descriptivo se revisan los datos que pueden estudiarse cuantitativamente tales como la cantidad de incidentes reportados por mes, hora y fecha, entre otros. Esto se puede realizar utilizando la siguiente instrucción

```
vf.mes.value_counts(sort=True, normalize=False,
                    dropna=True)
```

los resultados nos permiten observar el impacto de dichos factores en el total de los incidentes.

Estos y otros resultados, tales como las edades de las víctimas y agresores, pueden visualizarse gráficamente utilizando la librería `PLOTLY`.

Para poder graficar estos datos, la información correspondiente debe ser procesada ya que hay reportes para los cuales no se especifica la edad del agresor o la víctima; en algunos casos el reporte fue llenado con un *NE* y en otros casos se reportó un rango por ejemplo *40 a 45 años*. La forma en que se procede a limpiar estos datos es la siguiente

```
edad_v = vf.edad_v # edades de las victimas
edad_v = []
for dato in edad_v:
    s = str(dato).replace(",", " ")
    pedazos = s.split()
    while "a" in pedazos:
        pos = pedazos.index("a")
        desde = int(pedazos[pos - 1])
        hasta = int(pedazos[pos + 1])
        prom = (desde + hasta) // 2
        edad_v.append(prom)
        pedazos = pedazos[:pos] + pedazos[pos + 2:]
    edad_v += pedazos
edad_v = list(filter(lambda dato: dato != "NE",
    edad_v))
edad_v = list(filter(lambda dato: dato != "nan",
    edad_v))
```

Una vez que los datos estén en el formato adecuado se procede a graficar. Por ejemplo, teniendo la información de las edades de las víctimas y agresores en los diccionarios *di_edadA* y *di_edadV*, se puede graficar un diagrama de barras:

```
import plotly
import plotly.plotly as py
import plotly.graph_objs as go

trace1 = go.Bar(
    x = list(di_edadA.keys()),
    y = list(di_edadA.values()),
    name = 'Edad agresores'
)
trace2 = go.Bar(
    x = list(di_edadV.keys()),
    y = list(di_edadV.values()),
    name = 'Edad victimas'
)

data = [trace1, trace2]
layout = go.Layout(
    barmode = 'group'
)
```

Modelos lineales

Modelos lineales con *scipy.stats*

Modelos lineales son para representar una variable de inters como una funcin de uno o ms factores conocidos de tal forma que se maximice la proporcin de variacin explicada por el modelo.

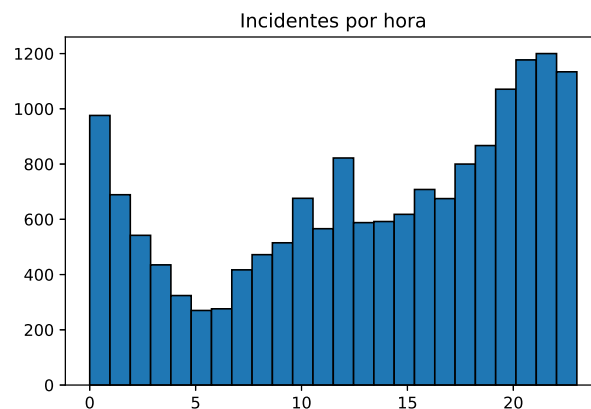


Figura 1: Cantidad de incidentes reportados por hora durante el año 2018.

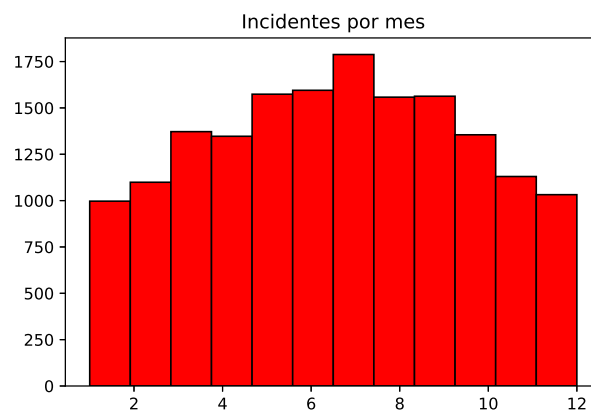


Figura 2: Cantidad de incidentes reportados por mes durante el año 2018.

3. Resultados

En esta sección se presentan los resultados del análisis y experimentos que se realizan.

Primero se muestran gráficos que permiten una visualización de la información con la que se trabaja.

Las Figuras 1 y 2 muestran un histograma de la cantidad de incidentes reportados de acuerdo a las horas y de acuerdo a los meses respectivamente. De ellas podemos observar que el rango de horarios de las 20:00 - 01:00 horas es el que reporta una mayor cantidad de incidentes mientras que en el caso de los meses, julio fue el mes más violento de ese año.

Para visualizar el rango de edades de víctimas y agresores, primero se agrupan en décadas para evitar obtener *peines* por la falta de información reportada. La Figura 3 muestra la gráfica que corresponde a esta información. De esta figura se resalta la cantidad de incidentes con víctimas entre 0-10 años pues se reportan 922 casos, además se reportan 111 casos de víctimas mayores de 80 años.

También es notorio que la mayoría de los incidentes ocurre

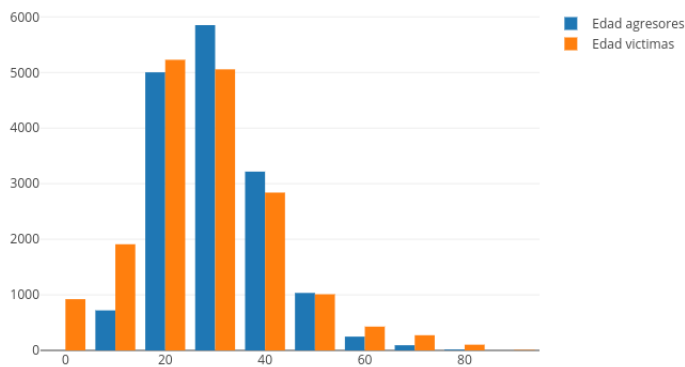


Figura 3: Rango de edades de víctimas y agresores.

con agresores y víctimas de edades de 30-50 años.

4. Conclusiones

Referencias

Agradecimientos

Agradezco al Consejo Nacional de Ciencia y Tecnología (CONACyT) por la beca otorgada. A la doctora Patricia L. Cerda Pérez por proveer los datos utilizados en el estudio y a la doctora Elisa Schaeffer por la guía proporcionada para la realización de este trabajo.