

Práctico 1- Análisis y Visualización

Integrantes:

- Kunzmann, Natalia
- Rinaudo, Eleana
- Sulca, Anahí

Descripción

Los datos corresponden a información recopilada correspondiente a los llamados de todo el país atendidos por los y las profesionales del 137 sobre casos de Violencia Familiar desde Enero de 2017 hasta Febrero del 2022 organizados de manera unificada.

Pueden consultar en:

<http://datos.jus.gob.ar/dataset/linea-137-victimas-de-violencia-familiar>

El dataset contiene información de id, descripción del llamante, género del llamante, tipo de violencia, género de la víctima, cantidad de víctimas, cantidad de agresores, género del agresor, relación con la víctima, fecha y hora de la llamada.

repositorio con los datos listos para procesar:

<https://github.com/IsabelRivadero/ir-mentoría2022>

Variables Aleatorias Categóricas

1. Análisis columna llamante_descripción:

- Está bien lo que hicieron de llevar al valor otro/a los valores que aparecían representados de distintas formas, pero se podían considerar representadas por la misma.
- En todos los gráficos mencionan que van a realizar un histograma pero lo que realizan es en realidad un Countplot es decir, se trata de un bar graph ya que están comparando variables discretas. Por eso les quería compartir un enlace y una comparativa para aclarar lo que es una duda frecuente: <https://www.kaggle.com/general/303648>

Comparison Chart

BASIS FOR COMPARISON	HISTOGRAM	BAR GRAPH
Meaning	Histogram refers to a graphical representation, that displays data by way of bars to show the frequency of numerical data.	Bar graph is a pictorial representation of data that uses bars to compare different categories of data.
Indicates	Distribution of non-discrete variables	Comparison of discrete variables
Presents	Quantitative data	Categorical data
Spaces	Bars touch each other, hence there are no spaces between bars	Bars do not touch each other, hence there are spaces between bars.
Elements	Elements are grouped together, so that they are considered as ranges.	Elements are taken as individual entities.
Can bars be reordered?	No	Yes
Width of bars	Need not to be same	Same

- Sobre la observación:
Si bien el 30,59% de las llamadas fueron hechas por las víctimas, también se podría observar que más del 60% de las llamadas es la suma de los porcentajes restantes indicaría que la mayor parte de las llamadas las realiza un externo a la víctima.

3. Análisis columna llamante_vinculo_ninios_presentes

- Está bien mencionar el porcentaje solo de las tres primeras variables, además se puede mencionar porque no se tuvieron en cuenta los demás porcentajes ya sea por considerarlos secundarios o poco significativos por ejemplo.
- Está muy bien el proceso de ver que valores toma la variable para luego unificar en una misma variable en caso de ser apropiado, con esto bastante de la limpieza ya va a estar hecho para el próximo TP.

4. Análisis columna violencia_tipo

- En el segundo agrupamiento, donde se llevan varias columnas a "Otras", quedaría para el próximo TP quedará como pendiente corregir ya que se pierde información de otros tipos de violencia. Esto podría ser llevando los siguientes valores posibles:

Física, sexual y psicológica

Económica y psicológica

Sexual y psicológica

Sexual, económica y psicológica

Física, sexual, psicológica y económica

por ejemplo, para el valor representado por "Física, sexual y psicológica" llevar a 3 columnas una para Física, una para Sexual y otra para psicológica y hacer que cada una de estas columnas tenga un 1 en la fila cada vez que se midió este tipo de violencia es decir, que SI hay una información correspondiente a el valor medido en Física, sexual y psicológica; y cero en caso contrario. Esto es para poder desdoblar sin perder información y poder de alguna manera

separar para contar cuántas veces hubo violencia Física, sexual, psicológica o Económica.

- Este cambio va a afectar a los resultados de otros tipos de violencias, en particular a la psicológica.
- Además podrían plantear una hipótesis antes de comenzar el experimento, para comprobar o no luego. Como por ejemplo: la violencia psicológica es en porcentaje mayor.

6. Análisis columna agresor_genero

El género es del agresor, se confunde en el cálculo de los porcentajes ya que se menciona que es el género de las víctimas.

7. Análisis columna agresor_relacion_victima ¶

Se podría mencionar alguna hipótesis al respecto de porque tantos de los datos que han sido tomados indican desconocimiento del vínculo víctima agresor

Variable Aleatoria Numéricas ¶

9. Análisis columna victima_cantidad

- el 75 % de los casos se mantienen por debajo de 2 la cantidad de víctimas.
- hipotetizar de porque la victima podría ser 0 en 8161 casos, no se cuentan como víctimas si no hubo desenlace? si no se solicitó ayuda presencial?. Esto podría servir para sugerir que relaciones serían interesantes de observar en más detalle en alguna instancia.

10. Análisis columna agresor_cantidad

- Lo mismo que en el caso anterior pero para agresor_cantidad= 0. Esto además modifica el promedio. Fijense que no debería poder quedar por debajo de 1 individuo el número de agresores

Relaciones relevantes

b. Llamante y derivación de la llamada

- NS/NC y transgenero podría graficarse y luego hecho un gráfico igual pero excluyéndose por no ser suficientes datos para poder concluir algo aclarando este criterio.

c. Cantidad de víctimas vs Cantidad de agresores

- llamado_fecha_anio:
Calcularon el promedio de víctimas y agresores por año
- Algo importante de decir de ambos gráficos es que el promedio de víctimas siempre es mayor al promedio de agresores lo que quiere decir que es más común un agresor con más de una víctima.
- En el primer gráfico pueden hipotetizar con respecto a porque el pico de agresores en 2021. O buscar algún tipo de información del contexto, como por ejemplo: el contexto pandémico podría explicar el incremento en el número máximo de víctimas que comenzó en 2020, tuvo su pico en 2021 y luego comenzó un descenso de 13 a 10 mientras que el descenso de victimarios es de 7 a 4.
- En el segundo gráfico se podría observar algo con respecto a las tendencias a la baja en víctimas, aunque siempre superando en 3 a los victimarios.

En líneas generales:

- Muy buen trabajo, están aprobadas quedense tranquilas.
- Traten de para el próximo tp agregar lo mencionado en:

4. Análisis columna violencia_tipo: que será parte del TP2 ya que forma parte de la limpieza y curación de los datos.

- Observar: victima_cantidad, y agresor_cantidad en los casos que es cero. Es llamativo que existan llamadas sin víctimas o sin agresor. Definir un criterio, tratar de realizar un análisis que pueda servir para explicar, o si piensan que son casos que se deberían quitar y explicar porque.

Todo esto lo pueden completar junto con el próximo TP.