

## Práctico 2- Limpieza y curación de datos

### Integrantes:

- 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

### 4. Análisis columna `violencia_tipo`:

- El `get_dummies` que les había propuesto era para obtener los tipos separados que ustedes obtienen antes de empezar con los `get_dummies`. Una vez que obtienen esto:

Psicológica	37748
Física	22822
No-Aplica	7658
Económica	1017
Sexual	661
Otras	177

ya no hace falta el `dumme`, pero quizás les sirva luego a la hora de los modelos.

### 8. Análisis columna llamado `derivacion`:

- Estos resultados son interesantes para mostrar y mencionar en la presentación.

## Variable Aleatoria Numéricas ¶

### 9. Análisis columna `victima_cantidad`

- La vez pasada les había sugerido que :  
“hipotetizaran el porque el número de víctimas podría ser 0 en 8161 casos“

Vi que trabajaron con este caso. Cuando eliminan porque *no existe un tipo de violencia* es relativo porque un conflicto familiar es la forma en que se tomó el dato, la violencia intrafamiliar

también es una violencia, pero sí se puede eliminar considerándolo un outlier.

- Con las asunciones, lo mejor siempre es eliminar(mientras no sean tantas filas) antes que cambiar los valores.
- Asumir que es error de tipeo y que 10 , 11, 12 debieron ser 1 es una asunción muy fuerte. Es preferible dejarlo y explicar que se lo considera un outlier pero no tienes información para decir que no es así o dropearlo antes que cambiarlo. Porque pueden estar cambiando algunas estadísticas como promedios por ejemplo.

#### 10. Análisis columna agresor\_cantidad

- Se está asumiendo para hacer los cambios que la columna tipo de violencia tiene más importancia o menos posibilidad de error que la columna agresor\_cantidad. No está mal, pero tengan en cuenta que hacer la deducción implicaría eso.
- Lo mismo de antes para:  
“Lo unificamos y le reasignamos como un caso a cada uno ya que la relación con la variable agresor\_relacion\_victima y llamada\_derivacion no nos aporta mucha información para que la cantidad de agresores sean de esa cantidad.”  
Asunción muy fuerte pasarlos a 1, es preferible dropear filas.

Están aprobadas, les sumo algunas sugerencias ya pensando en la presentación y mas información al respecto.

## Sugerencias para el video:

- No hay ningún análisis que tenga normalización de los datos. Es decir, para poder comparar llamante género debería tomarse cantidad igual a la cantidad mínima o comparar porcentajes. Si son menos hombres, se toma esa cantidad y se comparan en cuanto a llamado\_derivacion. Revisar con normalización, los análisis que deseen mostrar en el video.
- Analizar todos los meses del período de datos comprendido, los días del mes para todos los meses, y las horas. Y hacer tres boxplots: uno con los meses, otro con las estadísticas para los días y otro para las horas. Buscar un pico máximo y mínimo para cada uno.

## Presentación:

Grabar un video de no más de 10 minutos (sugerencia entre 5-7 minutos) esto es porque al final del ciclo de las mentorías van a asignar a los grupos a que vean videos de otros grupos y para entender los datos van a tener que ver el primer video y el que harán al final, es decir la idea es que ambos se complementen para que aquellos que tengan asignado ver sus videos puedan conocer sus datos, análisis, proceso y conclusiones. Entonces, para que no sea tan largo piden que los videos respeten ese límite **si o si**.

El 1 de julio deberán mandarme un enlace abierto o link para que pueda acceder tanto yo como cualquiera que desee ver el video. Estos videos los van a subir a una página para que estén disponibles para todos.

Realizar una presentación del problema y mostrar los datos con los que trabajan, solo desde el enfoque de estas dos materias que hemos puesto en práctica hasta ahora.

## Sugerencias para el armado del video

Contar brevemente qué datos les tocaron, qué tipos tenían y qué análisis les hicieron para conocerlos mejor.

Elaborar unas filminas con los gráficos que quieran compartir(aquellos que tengan resultados más importantes) con algunas breves explicaciones.

En el vídeo tratar de hablar las dos, pueden dividirse la presentación.

practiquen tomándose el tiempo.

Estructura sugerida:

- Introducción a los datos
- Principales hallazgos: lo más importante a contar
- Algún inconveniente que hayan tenido y como lo salvaron, o con que les gustaría continuar.
- Relaciones relevantes que hayan encontrado