

## Práctico 2- Análisis exploratorio y curación de datos

**Entrega 17/06**

**Video presentacion 1: 24/06**

### 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>

## Objetivo

Analizar qué variables tienen datos suficientes como para obtener resultados significativos y con respecto a que (por ejemplo no hay cantidad suficiente de información para un análisis anual incluyendo el 2022) y ver que variables no y desecharlas según algún criterio comentándolo brevemente. Descripción de tipos de variables, limpieza de filas con NaNs. Explicar cuáles de los análisis de variables realizados, contienen información importante y cuáles bajo algún criterio se deciden descartar.

*Mucho de este trabajo ya lo hicieron para el práctico anterior, sugiero revisar las relaciones de las columnas con los datos temporales. Les comparto un notebook de ejemplo básico:*

<https://colab.research.google.com/drive/1H9AgoLFVdY5e0W4IRjBhMbG24VgsZWvU?usp=sharing>

## Consignas:

*Importante: trabajar sobre el mismo dataset que venían trabajando:*

1- Realizar limpieza de datos en caso de faltar o ser necesario, haciendo foco en los tipos de datos y búsquedas de criterios para limpiar filas (lo que espero es que expliquen con sus palabras la razón por la que deciden dropear algo).

La limpieza y curación total debe incluir:

- Análisis de tipos, NaNs, unificación de categorías equivalentes (este parte del trabajo ya está hecho).
- transformación de string a datos temporales (también hecho).
- Faltaría relacionar las demás columnas buscando ampliar el análisis ya hecho.
- Breve descripción de lo que se observa.

2- Visualizaciones para relaciones que se hayan quedado con ganas de experimentar o que tengan información que consideren de importancia para describir los datos y sus relaciones de manera más completa. Al pie de cada gráfico describir lo que se observa.

A modo de guía buscar responder las siguientes preguntas con el análisis:

- ¿De qué género son los/las denunciantes?
- ¿Qué relación existe con el/la agresor?
- ¿Qué vínculo existe entre la víctima con el llamante?
- ¿Cuál es la relación entre la derivación de la llamada y el llamante?
- ¿Cuál es la cantidad de víctimas/agresores?
- ¿Existieron días, o periodos sin llamadas? con llamadas mínimas?
- ¿Existen tendencias a la estacionariedad?
- ¿Cuáles son las fechas y horas que se dieron como pico?
- ¿Cuáles son las fechas y horas que se darán como pico a futuro?

Resumido:

- Mucho del trabajo que necesitaban hacer para este TP ya fue hecho, si se busca ampliar la limpieza y los análisis es para lograr un mayor conocimiento de aquello que van a tener que presentar en el Video presentación 1. La idea central del video es que conozcan los datos y muestren un poco el trabajo que hicieron para limpiarlos y conocerlos.
- Agregar lo que quedó como corrección del TP1. En particular desdoblar la información de tipos de violencia.
- Trabajar sobre una copia del notebook inicial
- Traten de trabajar en forma colaborativa en un github de presentación donde todas las integrantes puedan commitear sus desarrollos.