Desafío de Análisis de Datos sobre el Juzgado 10 | Escuela de Datos Vivos

Sobre el desafío:

Esta vez el challenge implica Datos Abiertos proporcionados por el [Juzgado Penal Contravencional y Faltas N° 10 Ciudad de Buenos Aires](https://twitter.com/jpcyf10) sobre denuncias y casos relacionados con violencia de género, con el apoyo del juez [Pablo Casas](https://www.linkedin.com/in/pablo-casas-09885a183/) (si, otro Pablo Casas..).

El objetivo es poder analizar los casos y encontrar patrones a través de un modelo de clustering para detectar los comportamientos más comunes denunciados.

Más allá de la posibilidad de practicar y crear un portfolio de trabajos, es una excelente oportunidad porque puede tener un impacto en la justicia concreto.

## Acerca de los datos

Aquí hay datos estructurados y no estructurados. **Estructurados** son los datos en la tabla, y **no estructurados** en las resoluciones en pdf (se puede tener acceso al texto de cada PDF[?](https://drive.google.com/drive/folders/0B9wNhp3GjjazUFBabDFpQzVXVEk?usp=sharing)), donde aparece texto. Esto se trata con técnicas de NLP (natural language processing).

Vamos a empezar con los estructurados.

Datos a utilizar:

[base\_de\_datos con perspectiva de género](https://drive.google.com/drive/folders/0B9wNhp3GjjazcmNTYzE4Rk1VQUU) (ahí mismo tienen el diccionario de datos). Esta carpeta de drive, pública, es parte del [repo general](https://drive.google.com/drive/folders/0B9wNhp3GjjazZ2VCQVZmM3MwTTQ), (hay varios documentos, pero probablemente no sirven para hacer análisis).

## Objetivos

**Proyecto principal**:

Realizar un modelo de clustering que use los datos estructurados para encontrar, por ejemplo, aquellos 3 ó 4 comportamientos más comunes. Cada registro pertenece a un cluster.

El punto 2), da origen a otro posible análisis que es encontrar los puntos que no se corresponden a un patrón general, estos son los outliers/detección de anomalías.

**Proyecto secundario:**

Plantear análisis exploratorio de los datos, con gráficos, y que los datos se lean dinámicamente del google drive (a medida que se agregan registros en ese excel gdoc).

Con lo cual a medida que salgan las resoluciones los gráficos serán actualizados. El Juez PC dijo que está super interesado en este punto, no se cuanto valor agregado puede agregarse considerando que el Juzgado ya tiene este sitio: <https://juzgado10pcyf.github.io/#/> (muy bien armado por cierto).

## Una vez que termine el proyecto

Parte del equipo del juzgado 10 están haciendo cursos en EDV, estimo que hay cierta probabilidad de que lo usen.

Como algunos de uds me sugirieron que querían presentar informes como en el trabajo, voy a hablar con el juez PC para que me diga cómo sería la presentación de un informe de tenor judicial.

## Fechas

Vamos a organizar uno o dos encuentros entre los alumnos/as (dependiendo de la convocatoria, con la misma mecánica del desafío anterior).

Estimamos que sean 2 semanas de desarrollo pensando en terminarlo antes del 18 de noviembre.

La presentación final la realizaremos el 25/11, el día internacional de la eliminación de la violencia contra la mujer.