# Plan de Trabajo

**Tïtulo Investigación**

Sistema Unificado de Registros de Incidentes relativos a la violencia de género

**Tema Investigación**

Creación de un Sistema Unificado de Registros para que las víctimas de violencia de género puedan administrar los datos y la información de sus casos, y así evitar la revictimización ocurrida cuando una recurre a las autoridades y éstas le requieran información relacionada con las situaciones de violencia.

Este sistema brindará una interfaz para historizar y dar acceso inmediato y digital, a documentos, incidencias y acciones que son relevantes para la asistencia de una persona en situación de violencia de género, creando un registro perdurable, unificado y accesible de estos incidentes. La incorporación de un módulo de análisis estadístico y visualización de los datos, el objeto de esta beca, permitirá que se utilicen estos registros para la creación de estadísticas de alto impacto que sirvan de herramienta en la generación de políticas públicas. Teniendo en cuenta durante todo el desarrollo el buen tratamiento de los datos sensibles presentes.

**Preguntas disparadoras que hayan llevado a elegir el tema de la investigación**

¿Qué problemáticas reales afectan hoy en día en Argentina a las mujeres que sufren de violencia de género? ¿Cuáles son las causas de estas problemáticas y qué actores intervienen? ¿Cómo se puede contribuir desde la tecnología, a enfrentar estas problemáticas?

**Pregunta de investigación**: ¿Qué herramientas puede brindar la tecnología para afrontar las situaciones de violencia de género, tanto para las víctimas como para el estado?

**Literatura existente u otros trabajos que se relacionan y constituyen antecedentes**

* Nicolás Abbate, “*Sesgos de datos con perspectiva de género: Cómo incorporar a la población TTNB+ a las encuestas en América Latina*”.
* Sistema Único de Registro de Denuncias por Violencia de Género (URGE).
* Stephen A. Thomas (2015), *Data Visualization with JavaScript*.
* Jake VanderPlas (2016), *Python Data Science Handbook*.
* Recurso online: API Georef de la Secretaría de Innovación Tecnológica del Sector Público (Jefatura de Gabinete de Ministros) en datos.gob.ar. URL: <https://datos.gob.ar/dataset/jgm_8/archivo/jgm_8.1>.
* Recurso online: Dataset de Provincias, Departamentos y Municipios en formato GeoJSON URL: <https://datos.gob.ar/ar/dataset/jgm-servicio-normalizacion-datos-geograficos/archivo/jgm_8.16>.
* Recurso online: Datos Abiertos del Juzgado Penal Contravencional y Faltas Nro. 10 de la Ciudad Autónoma de Buenos Aires: <https://drive.google.com/drive/folders/0B9wNhp3GjjazZ2VCQVZmM3MwTTQ?resourcekey=0-s10MrFRR0gmOd1TgrDZs7g>.

**Objetivos del proyecto**

*Objetivos Generales*

* Desarrollo de un Sistema Unificado de Registros para víctimas de violencia de género, que sea capaz de generar estadísticas significativas y ayude a evitar la revictimización.
* Formación académica y profesional de la becaria en las temáticas abordadas.

*Objetivos Específicos*

* Relevamiento del estado actual de la problemática.
  + Descripción del estado actual.
  + Desarrollo de procesos e instituciones intervinientes.
* Exploración y aprendizaje (metodologías, tecnologías y librerías).
  + Librerías de visualización de datos.
  + Tecnologías de representación de datos geográficos.
  + Buenas prácticas en el resguardo y tratamiento de datos sensibles.
  + Herramientas de machine learning aplicables.
* Estudio de las librerías seleccionadas:
  + Estudio de la utilización de las librerías.
  + Mejores prácticas en la implantación e integración en proyectos.
* Recopilación de datos y creación de dataset de prueba.
* Desarrollo de Prototipo SUR-Gen:
  + Puesta en marcha de prototipo con datos de prueba.
  + Refinamiento iterativo (símil metodologías ágiles, y/o PDCA) sobre prototipo, requerimientos y modelos de datos.
  + Implementación del módulo de análisis estadístico y visualización de datos.
* Testeo de casos de uso sobre prototipo y módulo de visualización.
* Documentación del desarrollo.
* Presentación en congresos, revistas y eventos científico-tecnológicos.

**Metodologías y herramientas a explorar y utilizar**

* Python, IPython, Jupyter Notebooks:
  + Junto con el stack de *data science*, como entorno para exploración y análisis de datos. Luego se buscará refinar los análisis exploratorios que se realicen y convertirlos en funcionalidades específicas del sistema.
* Django:
  + Como tecnología de base para el sistema, brinda simplicidad para representar el modelo de datos y lograr un desarrollo rápido, correcto y mantenible del backend.
* Librerías de visualización de datos:
  + Incluyendo librerías de python como Matplotlib, Seaborn, Plotly y Altair.
  + Matplotlib y Seaborn permiten la creación de gráficos estáticos.
  + Plotly permite también gráficos interactivos en Jupyter.
  + Altair emite gráficos interactivos en formato VegaLite que pueden resultar interesantes para explorar los datos en más detalle.
* Tecnologías de representación de datos geográficos:
  + GeoPandas, geoplotlib, y folium como librerías.
  + GeoJSON y shapefile como formatos de representación de datos.
  + Interacción e integración con otras herramientas y tecnologías utilizadas.
* Tecnologías y herramientas de *machine learning*:
  + Numpy como librería de base para representación de datos en forma de tensores.
  + Pandas como librería de alto nivel para el manejo de datos. Pandas se integra muy bien con muchas de las librerías de visualización de datos, y permite hacer procesamiento de alto nivel, además de poder leer y escribir en múltiples formatos de archivo y bases de datos.
  + Scikit-learn provee múltiples algoritmos de preprocesamiento y análisis de datos de *machine learning* complementarios a las herramientas que brindan numpy y pandas.

**Cronograma de actividades con estimación de tiempos**

| **Objetivos/ Mes** | *1°* | *2°* | *3°* | *4°* | *5°* | *6°* | *7°* | *8°* | *9°* | *10°* | *11°* | *12°* |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Relevamiento del estado actual de la problemática. |  |  |  |  |  |  |  |  |  |  |  |  |
| Exploración y aprendizaje (metodologías, tecnologías y librerías). |  |  |  |  |  |  |  |  |  |  |  |  |
| Estudio de las librerías seleccionadas |  |  |  |  |  |  |  |  |  |  |  |  |
| Recopilación de datos de estudio. |  |  |  |  |  |  |  |  |  |  |  |  |
| Desarrollo de Prototipo SUR-Gen |  |  |  |  |  |  |  |  |  |  |  |  |
| Testeo de casos de uso sobre prototipo y módulo de visualización. |  |  |  |  |  |  |  |  |  |  |  |  |
| Documentación del desarrollo. |  |  |  |  |  |  |  |  |  |  |  |  |
| Presentación en congresos, revistas y eventos científico-tecnológicos. |  |  |  |  |  |  |  |  |  |  |  |  |

El sombreado intenso indica las tareas prioritarias en un momento dado, y el sombreado claro aquellas que se tienen en cuenta pero no concentran la atención atención.

**Resultados esperados de los análisis a realizar y entregables**

Desarrollo de SUR-Gen, incorporando el Módulo de Análisis Estadístico y Visualización de Datos que ayude a las autoridades a analizar, procesar y visualizar la información volcada en el sistema, ayudándolos a entender la problemática de violencia de género en un municipio o región donde se implante.

Serán entregables del proyecto los videos mensuales de informe de avance, y las memorias e informe final del proyecto. También los trabajos académicos y presentaciones realizadas en congresos que se lleven a cabo durante la duración de la beca relacionados con el tema de proyecto.