

**Carrera:** Tecnicatura Superior en Ciencia de Datos e Inteligencia Artificial.

**Institución:** Politécnico Malvinas Argentinas

**Materia:** Aprendizaje Automático

**Proyecto Aprendizaje Automático:** Mortalidad. Tasa general y distribución porcentual según principales causas de muerte y grupos de edad. Provincia de Tierra del Fuego AELAS. Años 2005-2019.

**Alumna:** Barrios, Yanina Soledad

**Profesor de la Catedra:** Mirabete, Martin

## Entrega 2: Descripción del Dataset y Origen

- En esta segunda entrega, debe tener acceso al dataset que utilizará en su proyecto de [Aprendizaje Automático](#).
- Proporcione una descripción completa del dataset, incluyendo la cantidad de instancias, características (columnas), tipos de datos, y cualquier información relevante.
- Informe sobre el origen del dataset, es decir, de dónde provienen los datos. Esto puede incluir la fuente, la fecha de adquisición y cualquier proceso de recopilación o preprocesamiento que haya realizado.
- Se recomienda utilizar dataset de dominio público (ipiec, datos.gob.ar, etc).
- Esta entrega debe ser realizada en el repositorio GIT, incluyendo archivos de dataset y documentos.

### Descripción completa del dataset

El dataset contiene un archivo de varias hojas:

Primero: Es un índice con la descripción del contenido.

Segundo: Tres hojas llamados cuadro 1, cuadro 2, cuadro 3, donde cada hoja contienen los datos específicos de mortalidad según las diferentes variables.

cuadro 1: Contiene los datos de las principales causas de muerte según la edad y para ambos sexo.

cuadro 2: Contiene los datos de las principales causas de muerte según la edad y para el sexo masculino.

Cuadro 3: Contiene los datos de las principales causas de muerte según la edad y para el sexo femenino.

Tercero: Es una hoja llamada Ficha Técnica, donde incluye información del dataset.

### Cantidad de Instancias (Filas)

Cada hoja presenta datos desde 2005 a 2019, agrupados por sexo y edad.

Hay en los tres dataset la misma cantidad de instancias 120 filas.

El dataset contiene 45 columnas en las hojas de cuadro 1, cuadro 2 y cuadro 3, incluyen información como:

**Las columnas contienen información como:**

- Año.
- Número de defunciones.

- **Tasa de mortalidad.**
- **Distribución porcentual.**

#### **Tipos de Datos:**

- **Categorías:** Causas de muerte, grupos de edad, sexo.
- **Númericos:** Tasas y cantidad de defunciones.

### **Informes Relevante**

Existen múltiples columnas que están divididas por año y puedes incluir valores nulos o encabezados repetitivos.

La estructura de los datos es un poco complicada, ya que en algunas filas contienen encabezados o están vacías.

### **Origen del Dataset**

- **Fuente:** La información es extraída del sitio Web <https://ipiec.tierradelfuego.gob.ar/>

**Dataset:** [https://ipiec.tierradelfuego.gob.ar/wp-content/uploads/2022/11/05\\_1\\_06-Mortalidad.-Tasa-general-y-distribucion-porcentual-segun-principales-causas-de-muerte-grupos-de-edad-y-sexo.-TDF.-Anos-2005-2019.xlsx](https://ipiec.tierradelfuego.gob.ar/wp-content/uploads/2022/11/05_1_06-Mortalidad.-Tasa-general-y-distribucion-porcentual-segun-principales-causas-de-muerte-grupos-de-edad-y-sexo.-TDF.-Anos-2005-2019.xlsx)

- **Período de Cobertura:** Los datos cubren el período entre los años 2005 y 2019.

**Fecha de Adquisición de datos:** Los datos fueron adquiridos mediado del mes de Octubre del 2024.

**Preprocesamiento:** Hasta el momento, no se ha realizado preprocesamiento específico sobre los datos. El dataset necesita limpieza, incluyendo:

- **Eliminar columnas vacías** (algunas columnas están completamente vacías).
- **Renombrar columnas** para una mejor interpretación.
- **Transformación de Datos:** Posiblemente se requiera transformar algunas columnas a tipos de datos más adecuados para su análisis, como convertir tasas a formato numérico adecuado.

Este análisis inicial nos permite identificar las transformaciones y los pasos necesarios para preparar el dataset para análisis más profundos, como la clasificación de causas de muerte o el análisis predictivo. El próximo paso sería realizar un preprocesamiento completo para garantizar que los datos estén en un formato adecuado para su análisis mediante aprendizaje automático.