

## Clase 7: Ingeniería de características.

FM849 - Programación Científica para Proyectos de Inteligencia Artificial (IA)

## Contenidos de hoy

- ▶ Limpieza y Preparacion de Datos.
- ▶ Agregacion de Datos.
- ▶ Agrupamiento de Datos.

# Motivación

En general, los datos reales vienen con problemas:

- ▶ Datos faltantes.
- ▶ Datos desordenados.

Es importante entregar datos de calidad a modelos. Esto permite generar modelos robustos y confiables.

# Limpieza y Preparación de Datos.

## ► Tratamiento de valores faltantes

- Detectar valores NA → `isna()`, `notna()`
- Eliminar valores NA → `dropna()`, `fillna()`

NA se refiere a valores nulos/no disponibles.

## ► Corrección de tipos de datos

- Cambiar el tipo → `astype()`
- Crear fechas → `to_datetime()`

## ► Limpieza básica

- Reemplazar valores: → `replace()`
- Remover duplicados: → `duplicated()`, `drop_duplicates()`

# Agregación de Datos

- ▶ **Cálculo de estadísticas resumen**

- ▶ Calcular estadísticas → `mean()`, `sum()`, `count()`
- ▶ Calcular estadísticas → `min()`, `max()`, `std()`

- ▶ **Agregación por grupos**

- ▶ Agregar datos → `agg()`

- ▶ **Transformaciones agregadas**

- ▶ Transformar datos → `transform()`

# Agrupamiento de Datos

- ▶ **Agrupamiento por variables categóricas**
  - ▶ `groupby()`
- ▶ **Agrupamiento por intervalos**
  - ▶ `cut()`
  - ▶ `qcut()`
- ▶ **Reestructuración de datos**
  - ▶ `pivot(), pivot_table()`
  - ▶ `melt()`

# Vamos a continuar con el ejemplo de la clase anterior

Vamos a explorar un dataset de Pokemons: <https://colab.research.google.com/drive/1X7xoTisfFrGZFfrQ9R0ckTaYzv8Xw8TB?usp=sharing>.



Figura 1: Pokemones de primera generación.

## Referencias:

- ▶ Wes McKinney. (2022). Python for Data Analysis. Third Edition.