

## SEGUNDA PARTE LABORATORIO AUTO - 1

### Etapa 1: Procesamiento y limpieza de datos

#### Objetivos específicos

- Cargar y explorar datos en formato **JSON**.
- Transformar una estructura jerárquica a un **DataFrame de pandas**.
- Limpiar, convertir y preparar los datos para su análisis.

#### Actividades

##### 1. Carga del archivo JSON

- Cargue el archivo `pokemonDB_dataset.json` usando la librería `json` o `pandas`.
- Explore su estructura inicial (`keys()`, `items()`) para entender su jerarquía.

##### 2. Revisión de estructura y tipos

- Use `df.info()` y `df.describe(include='all')` para identificar tipos de datos.
- Determine cuántas variables son numéricas y cuántas categóricas.

##### 3. Limpieza y conversión de campos

Convierta columnas como "Height", "Weight", "HP Base", "Attack Base", "Defense Base", etc., a valores numéricos.

Elimine unidades ("m", "kg", "lbs") y convierta los campos a tipo float.

Separe las columnas compuestas, por ejemplo:

"Type" → Type1, Type2

"Gender" → Male (%), Female (%)

##### 4. Normalización y verificación final

- Verifique valores nulos y duros:
- Cree un nuevo DataFrame `df_clean` con los campos limpios listos para análisis:

```
["Type1", "Type2", "HP Base", "Attack Base", "Defense Base", "Speed Base", "Height_m", "Weight_kg", "Base Exp"]
```

## **Etapla 2: Exploratory Data Analysis (EDA)**

### **Objetivos específicos**

- Aplicar técnicas de análisis exploratorio de datos.
- Extraer estadísticas descriptivas y patrones visuales.
- Interpretar relaciones entre variables.

### **Preguntas y ejercicios de análisis**

En cada pregunta se indica si debe realizar **análisis estadístico** o **visualización gráfica**.

#### **1. Análisis estadístico**

¿Cuál es el promedio, mínimo y máximo de los atributos base (HP, Attack, Defense, Speed) de todos los Pokémon?

#### **2. Análisis gráfico**

Cree un histograma para visualizar la distribución de los valores de Base Exp.

Interprete si la distribución es simétrica o sesgada.

#### **3. Análisis gráfico**

Realice un **boxplot** comparando los valores de Attack Base entre los tipos principales (Type1).

Identifique qué tipo tiene Pokémon con ataques más altos en promedio.

#### **4. Análisis estadístico**

¿Cuál es el top 5 de especies (Species) más frecuentes en el dataset?

## 5. Análisis gráfico

Genere un **gráfico de barras** que muestre la cantidad de Pokémon por tipo principal (Type1).

¿Qué tipo es el más común?

## 6. Análisis estadístico

Calcule la **correlación** entre los atributos HP Base, Attack Base, Defense Base y Speed Base.

¿Qué atributos están más correlacionados entre sí?

## 7. Análisis gráfico

Cree un **heatmap** (mapa de calor) con la matriz de correlaciones obtenida en el punto anterior.

Interprete los resultados.

## 8. Análisis gráfico

Realice un **diagrama de dispersión** (scatter plot) entre Weight\_kg y Attack Base.

¿Existe relación entre el peso de un Pokémon y su capacidad de ataque?

## 9. Análisis estadístico

Determine el **promedio de altura y peso** por tipo principal (Type1).

Interprete cuál tipo tiende a tener Pokémon más grandes.

## 10. Análisis gráfico

Construya un **gráfico de violín o boxplot múltiple** comparando el atributo Speed Base entre los tipos Flying, Electric y Ground.

¿Qué tipo de Pokémon tiende a ser más rápido?

## Entregables

Cada grupo debe entregar:

1. Un proyecto en github titulado “Laboratorio 2: Data processing-EDA”, donde estarán los códigos (preferiblemente notebooks) de la parte 1 y la parte 2 del laboratorio y sus bases de datos
2. Un breve reporte en markdown como archivo Readme del laboratorio, que contenga:
  - Las respuestas interpretativas a las 10 preguntas.
  - Los gráficos generados y comentarios sobre ellos.
  - Conclusión general sobre qué factores parecen influir más en las estadísticas base de los Pokémon.