

# Informe de Diagnóstico y Transformación de Datos - Dataset Vivero

## 1. Introducción

Este informe presenta el proceso de diagnóstico y transformación del dataset del vivero. El objetivo es dejar los datos listos para el modelado en Machine Learning, abordando problemas de datos faltantes, codificación de variables categóricas y escalamiento de variables numéricas.

## 2. Diagnóstico Inicial

Se identificaron los siguientes problemas en el dataset original:

- Valores faltantes en variables numéricas como Altura\_cm, Ancho\_cm y Precio.
- Valores faltantes en la variable categórica Tipo.
- Posibles duplicados y registros incompletos.
- Diferentes escalas entre variables numéricas.

## 3. Transformaciones Realizadas

Para preparar el dataset se aplicaron las siguientes transformaciones:

1. Imputación de datos faltantes:
  - Media para variables numéricas.
  - Moda para la variable categórica Tipo.
2. Codificación de variables categóricas:
  - Se aplicó One-Hot Encoding a la variable Tipo.
3. Escalamiento de variables numéricas:
  - Se utilizó MinMaxScaler para normalizar las variables numéricas en un rango de 0 a 1.

## 4. Resultados

El dataset resultante no contiene valores faltantes y las variables categóricas han sido transformadas a formato numérico mediante One-Hot Encoding. Las variables numéricas han sido escaladas, reduciendo el impacto de las diferencias de magnitudes y facilitando el entrenamiento de modelos de Machine Learning.

## 5. Conclusiones

El dataset del vivero ahora se encuentra listo para ser utilizado en procesos de modelado y análisis predictivo. Las transformaciones aplicadas aseguran consistencia, completitud y compatibilidad con algoritmos de Machine Learning.