

# Description Report (Reporte)

- **Estructura del dataset**
  - El archivo presenta un registro tipo `Time Series`, donde se recopilan múltiples mediciones a lo largo del tiempo para cada vaca.
    - Número total de columnas (features): 22
    - Número aproximado de registros: 34
    - Periodo temporal: 22/03/2025 hasta 18/07/2025, con variaciones entre vacas (no todos los animales tienen la misma cantidad de observaciones).
    - Formato de variables:
      - Numéricas: `int64` y `float64`
      - Categóricas / de texto: `object`
      - Fechas y tiempos: `datetime` (en algunos casos deben convertirse correctamente)
    - La mayoría de los atributos son días transcurridos de un evento en específico.
- **Calidad de los datos**
  - El dataset combina variables productivas, reproductivas y salud del ganado.
  - Se observan valores faltantes (NaN) en varias columnas.
  - Algunas variables son irrelevantes o redundantes para el análisis principal (Días de ordeño, Ordeño desde 22/03/2025 hasta 18/07/2025, Días desde último evento de parto)
  - Variables como:
    - `#Pezones no encontrados en último ordeño`
    - `Días desde el último secado`
    - `Días desde el último tratamiento de vacunación`
    - `Días desde próximo celo`
    - `Días desde último Aborto`presentan registros casi totalmente nulos, algunos de ellos es debido a que al no presentar anomalías o alertas de riesgo, ningún valor es registrado.
- **Recomendaciones**
  - Eliminar columnas con más del 50 % de valores nulos.
  - Sustituir valores faltantes con:
    - `0` para indicadores binarios.
    - Promedio o mediana para variables continuas (ej. días, producción).

- Invertir los valores (ej: en vez de pezones NO encontrados cambiarlo a pezones encontrados para tener un atributo con valores que sirvan para el modelo y no valores NaN).
  - Convertir todas las variables numéricas a tipo `float64` para uniformidad.
  - Estandarizar variables numéricas antes de aplicar correlaciones o modelos (z-score o min–max).
  - Agrupar las vacas por ID individual y generar promedios o series de tendencia.