

Description Report (Inventario)

- **Estructura del dataset**

- El archivo presenta un registro tipo serie temporal (time series), donde se recopilan múltiples mediciones a lo largo del tiempo para cada vaca.
- Cada registro representa una observación individual dentro de un periodo determinado.
 - Número total de columnas (features): 35
 - Número aproximado de registros: ~200
 - Periodo temporal: marzo a junio de 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).

- **Calidad de los datos**

- El dataset combina variables productivas, reproductivas y fisiológicas del ganado.
- Se observan valores faltantes (NaN) en varias columnas, especialmente en mediciones biológicas o de laboratorio.
- Algunas variables son irrelevantes o redundantes para el análisis principal (productividad, eficiencia y reproducción).
- Columnas relacionadas con parámetros de sangre (ppm) — grupo `DI` a `TD`.
 - Representan concentraciones químicas en sangre, pero con registros muy incompletos.
 - Se recomienda descartar este bloque si el enfoque es productivo/reproductivo.
- Variables como:
 - `Patada`
 - `Pezones encontrados`
 - `Usuario`
 - `RCS`presentan registros casi totalmente nulos o inconsistentes.

- **Recomendaciones**

- Limpieza de datos
 - Eliminar columnas con más del 80 % de valores nulos.
 - Sustituir valores faltantes con:

- 0 para indicadores binarios.
 - Promedio o mediana para variables continuas (ej. días, producción).
 - Convertir todas las variables numéricas a tipo `float64` para uniformidad.
- Gestión de fechas
 - Convertir todas las fechas (`fecha de nacimiento`, `fecha de parto esperada`, `fecha de inseminación`, etc.) al formato `datetime`.
 - Eliminar registros con eventos administrativos no útiles (“Rechazada”, “Aceptada”).
- Preparación para análisis estadístico
 - 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.
 - Si se buscan predictores de producción, concentrarse en: `Días en ordeño`, `Producción total en lactación`, `Nº de lactación`, `Edad`, `Número de inseminaciones`, `Días abiertos`.
- Documentación
 - Etiquetar las variables con nombres consistentes y significativos.
 - Añadir metadatos: unidad de medida, frecuencia, tipo de variable, descripción.