

Quality Report (Global HATO)

- **Resultados de la exploración**

- ¿Están completos los datos? No
- ¿Faltan datos? Si

Atributo	¿Formato consistente?	# Missing	¿Valores suficientes ?	¿Aporta al proyecto? (Sí/No + por qué)
Número del animal	Sí (int)	0	Sí	Sí, identificador único esencial para agrupar historial por vaca.
Nombre del grupo	Sí (categoría/objeto)	0	Sí	Sí, permite segmentar entre VMS, secas, etc. Útil para filtrar.
Número(s) de selección de animal	No del todo (texto con variaciones)	373	No	No, muy incompleto y no relevante para la predicción productiva.
Estado de la reproducción	Parcialmente consistente	170	Sí	Sí, importante para modelos sobre preñez, producción, DEL, etc.
Días en ordeño (DEL)	Sí (float)	372	No	Sí, es una de las variables biológicamente más importantes, pero el dataset tiene demasiados faltantes. Se debe imputar o descartar.
Nº Lactación	Sí (float)	170	Sí	Sí, relevante: producción cambia por número de lactancia.
Número de inseminaciones	Sí (float)	385	No	Sí, pero extremadamente incompleta; poco útil sin imputación.
Días preñada	Sí (float)	475	No	Sí, pero los faltantes lo hacen casi inutilizable.
Días para el parto	Sí (float)	475	No	Sí, misma situación: útil pero muy incompleta.

Fecha de parto esperada	Parcialmente consistente	475	No	Sí, pero peor aún que días para el parto; demasiados NA.
Producción de leche ayer	Sí (float)	423	No	Sí, variable clave, pero con exceso de faltantes.
Producción media diaria últimos 7 días	Sí (float)	~420	No	Sí, muy útil para tendencias; pero muy incompleta.
Producción TOTAL en lactación	Sí (float)	~0	Sí	Sí, excelente atributo acumulativo sin NA. Muy útil para modelos de producción.

- **Problemas con los datos**

- Tiene demasiados valores nulos como producción, días en ordeño y reproducción.
- Algunas columnas no tienen formato consistente, por ejemplo fechas como texto.
- El dataset es pequeño y no permite modelos profundos.
- No contiene información temporal suficiente para análisis de series de tiempo.