

# Limpieza (Tercera iteración)

## **Dataset de Fichas de Vaca (HATO)**

Durante el análisis exploratorio se identificó que varias columnas no aportan información útil para los objetivos del proyecto. El tipo de evento y la fecha son datos estáticos generados automáticamente por el sistema, es decir, no están sujetos a redacción manual. En contraste, las columnas Descripción y Comentarios sí pueden ser editadas por un usuario, lo que introduce variabilidad y posibles inconsistencias; por ello, se decidió descartarlas.

Asimismo, las columnas Evento y Usuario presentan un alto porcentaje de valores nulos y no proporcionan información relevante para el análisis, por lo que también se determinó su eliminación.

## **Dataset de Registros de ordeño (HATO)**

Durante la transformación de los datos, y tras la validación con un experto en el área bovina, se determinó que todas las columnas relacionadas con los cuartos de las ubres no aportan valor al análisis. Esto se debe a que factores como la presencia de mastitis u otros eventos pueden afectar temporalmente la producción de un cuarto, generando variabilidad que no contribuye a los objetivos del estudio. Por esta razón, dichas columnas fueron eliminadas.

Además, se imputaron los valores nulos: en la columna MS se reemplazaron por “VMS NAN”, y en la columna MDI se sustituyeron por ceros, atendiendo al criterio técnico establecido.