

Limpieza (Primera iteración)

Dataset de Fichas de vacas

Durante el análisis exploratorio se identificó que varias columnas no aportan información significativa para los objetivos del proyecto. Algunas variables, como ciertos identificadores internos, fechas generadas automáticamente o códigos del sistema, son elementos estáticos que no influyen en el comportamiento productivo de los animales. Por otro lado, columnas como Descripción o Comentarios pueden ser editadas manualmente por los usuarios, lo cual introduce variabilidad, inconsistencias y ruido semántico difícil de procesar de manera automática. Con base en ello, se determinó su eliminación para mantener únicamente información objetiva y estandarizada.

Asimismo, se encontraron atributos con un porcentaje considerable de valores nulos o que no aportan valor analítico, tales como registros de eventos aislados o campos administrativos sin impacto en el análisis productivo. Estas columnas fueron descartadas por no contribuir a la construcción de modelos consistentes ni a la interpretación de patrones relevantes dentro del conjunto de datos.

Dataset de Registros de Ordeño

En la etapa de transformación y limpieza se identificó que diversas columnas relacionadas con mediciones muy específicas o condiciones temporales de los animales no generaban información útil para el análisis. Variables como las características individuales de cada cuarto de la ubre mostraron alta variabilidad dependiente de factores externos por ejemplo, mastitis o alteraciones momentáneas lo que hacía que su inclusión introdujera ruido en lugar de aportar valor. Considerando estos aspectos, se optó por eliminar dichas columnas para trabajar únicamente con mediciones robustas y comparables.

Además, se aplicó un proceso de imputación de valores nulos siguiendo criterios técnicos definidos previamente. En columnas categóricas se sustituyeron los valores faltantes por etiquetas estándar que permitan mantener la integridad del análisis, mientras que en variables numéricas se llenaron los espacios vacíos con valores neutros (como ceros) que evitaran distorsionar las métricas generales del dataset. Este proceso asegura una base de datos más limpia, consistente y adecuada para las siguientes etapas del proyecto.