

Description Report (ID Vaca(1204-8794))

- **Estructura del dataset**

- El dataset parece seguir una estructura de tipo **Time Series**. Donde se miden muchas cosas en un gran lapso de tiempo para la misma vaca.
 - El dataset contiene 35 features
 - El número de instancias es de ~200 Instancias
 - El número de instancias parece variar entre cada vaca
 - El periodo de los registros varía en días pero todos están en el rango de Marzo-Junio del 2025

- **Calidad de los datos**

- El dataset tiene diversos tipos de variables:
 - Datos de tipo Int y Float64
 - Objects como fechas
 - Features donde todos sus registros parecen ser nulos
- 11/35 features cuentan con todos (o casi) todos sus registros nulos o sin relevancia
 - De estas 11, en la categoría de Sangre (ppm) tenemos un grupo de 4 columnas que van desde DI-TD. Estas columnas nos indican la cantidad de sangre por partes de millón en la sangre. Como son pocos registros y en general el proyecto no se enfoca en estos rubros, se recomienda descartar este grupo de columnas.

- **Recomendaciones**

- ID vacas tiene un rango de tiempo de solo Marzo-Junio. Por el rango de tiempo y que en esencia, tiene la misma información que REGISTRO ORDEÑO, se recomienda descartar todos los csv de este grupo. Si se usa, se recomienda trabajar siguiendo lo siguiente:
 - Imputar datos (dependiendo del contexto) con 0's, promedio o eliminar completamente los registros.
 - Varias columnas, como Patada, Pezones Encontrados, Usuario y RCS no tienen registros útiles o la mayoría son nulos, debido a esto convendría eliminarlos.
 - Modificar el tipo de datos numéricos a Float64 para evitar conflictos entre Int y Float.
 - En la feature **Acción** puede ser interesante tener en cuenta las fechas donde la vaca fue RECHAZADA, ya que no brindan información adicional al modelo sabiendo que en ese momento ninguna otra columna tiene un registro. Esto último por que puede ser

una variable de estudio en relación a los días de lactancia y periodo de secado.