

# Description Report (Registro Ordeño)

- **Estructura del dataset**
  - El dataset parece seguir una estructura de tipo **Panel Data**. Donde se miden muchas variables de todas las vacas en diferentes horas.
    - El dataset contiene 35 features
    - El número de instancias, juntando todos los registros de las 44 vacas, tenemos una cantidad de 49602
    - El periodo de los registros van desde los años de 2021 a 2025
      - Donde, en los datos se observan vacas de 1 hasta 4 lactancias.
      - De manera manual, se descargaron los 44 archivos de las vacas. Esto último en referencia al corral de donde estamos sacando los datos (el corral 6)
- **Calidad de los datos**
  - El dataset tiene diversos tipos de variables:
    - Datos de tipo Datetime
    - Objects como variables categóricas
    - Algunas de las variables son anotaciones, por lo que no podrán ser variables categóricas.
    - Las únicas variables de tipo categóricas corresponden a objetos siendo:
      - Hora de Inicio (fecha en formato DD-MM-YYYY HH:MM)
      - Duración (en formato MM:SS)
    - Muchas de las variables columnas solo tienen datos nulos.
- **Recomendaciones**
  - Unir todos los dataframes individuales de cada vaca para poder analizar en volumen todos los registros, que tienen las mismas instancias, en un solo dataframe final.
  - En la feature **Acción** puede ser interesante tener en cuenta fechas donde la vaca fue RECHAZADA, ya que no brindan información adicional al modelo sabiendo que en ese ordeño 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.
  - Eliminar las columnas donde la mayoría de los datos son nulos.
  - Categorizar las variables de objeto, para tener los tipos de eventos y de esto tomar como pivote algunas de las importantes, para tener las variables que son importantes para cuando se clasifiquen.

- Crear variables dummies para las columnas de patada, incompleto y pezones encontrados, ya que estos registros pueden llegar a ser útiles para el modelo.
- Eliminar la columna de Ubre, ya que esta columna llega a ser más de interpretación.