

# Quality Report (Registro Ordeño)

- **Resultados de la exploración**

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

Atributo	¿Tiene formato consistente?	# valores faltantes	¿Tiene valores suficientes para usarlos?	¿Aporta información útil para el proyecto? ¿Porque Si/No?
Hora de inicio	Si	0	Si	Si, nos sirve como punto de partida para conocer los tiempos de ordeña y días de cada vaca.
Acción	Si	0	Si	Si, como variable categórica nos ayuda a saber si es que la vaca se ordeña o no en ese momento.
Duración (mm:ss)	Si	0	Si	No, el número de horas en la producción no es necesaria si es que ya se tiene la fecha de la lactancia.
Producción (kg)	Si	1,150	Si	Si, la variable como pivote puede ser muy útil para identificar correlaciones.
Número de ordeño	Si	1,150	No	Si, dependiendo del número de ordeños podríamos detectar mayores productoras y por ende más o menos lactancias.
RCS (* 1000 células / ml)	Si	49,193	No	No, no se tiene ningún registro en cualquier ID.csv.
Patada	Si	42,240	No	Si, se puede usar como variable Dummy
Incompleto	Si	43,815	No	Si, se puede usar como variable Dummy

<b>Pezones no encontrados</b>	Si	43,472	No	Si, se puede usar como variable Dummy
<b>Ubre</b>	Si	1,165	Si	No, ya que al ser algo más empírico para el modelo no representa importancia alguna.
<b>Pezón</b>	Si	44,208	No	No, no se tiene suficiente información y no brinda información relevante para el modelo.
<b>DI - media de flujos</b>	Si	2,588	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.
<b>DD - media de flujos</b>	Si	3,734	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.
<b>TI - media de flujos</b>	Si	4,988	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.
<b>TD - media de flujos</b>	Si	4,179	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.
<b>DI - sangre</b>	Si	2,588	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña y como es que afecta al desechado de la leche.
<b>DD - sangre</b>	Si	3,734	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña y como es que afecta al desechado de la leche.
<b>TI - sangre</b>	Si	4,988	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña y como es que afecta al desechado de la leche.

<b>TT - sangre</b>	Si	4,179	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña y como es que afecta al desechado de la leche.
<b>DI - conductividad</b>	Si	2,588	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.
<b>DD - conductividad</b>	Si	3,734	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.
<b>TI - conductividad</b>	Si	4,988	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.
<b>TT - conductividad</b>	Si	4,179	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.
<b>EO/PO</b>	Si	0	Si	Si, a partir del cambio en el robot (corral) podemos saber cuando entro a otro robot de ordeña.
<b>Usuario</b>	No	49,602	No	No, no hay datos suficientes para evaluar.
<b>Destino Leche</b>	Si	1,165	Si	Si, para la parte de producción puede ser un buen índice para saber si se desechar o no la leche.
<b>Razón de la desviación</b>	Si	48,486	No	No, no hay datos suficientes para evaluar.
<b>DI - flujos</b>	Si	2,588	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.
<b>DD - flujos</b>	Si	3,734	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.
<b>TI - flujos</b>	Si	4,988	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.

<b>TT - flujos</b>	Si	4,179	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.
<b>DI - producciones</b>	Si	2,588	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.
<b>DD - producciones</b>	Si	3,734	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.
<b>TI - producciones</b>	SI	4,988	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.
<b>TT - producciones</b>	Si	4,179	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.

- **Problemas con los datos**

- Manejar el registro de ordeño de uno por uno llega a ser poco práctico. Para la implementación se recomienda mergear todos los registros en uno solo.
- Varios registros tienen datos nulos. Por el contexto de la situación, para features con registros nulos se recomienda imputar con 0's o borrar con 0's. Por otro lado, promedios, medias o alguna otra técnica de estadística descriptiva puede sesgar los registros individuales de cada vaca.
- Se tuvieron que hacer modificaciones a nombre de algunas features, ya que aparecieran features con el mismo nombre y esto lo vuelve poco intuitivo. Las features renombradas se modificaron en función al nombre del header de la primera columna del .csv.
- Para el proceso ETL se recomienda tener un estándar en el tipo de dato numérico que se va a manejar. Ya sea **Int** o **Float64**.