

Quality Report (Registro Ordeño HATO)

- 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	0	Si	Si, la variable como pivote puede ser muy útil para identificar correlaciones.
Número de ordeño	Si	0	No	Si, dependiendo del número de ordeños podríamos detectar mayores productoras y por ende más o menos lactancias
RCS (* 1000 células / ml)	Si	2,444	No	No, no se tiene ningún registro.
Patada	Si	1,819	No	Si, se puede usar como variable Dummy.

Incompleto	Si	2,298	No	Si, se puede usar como variable Dummy.
Pezones no encontrados	Si	2,245	No	Si, se puede usar como variable Dummy.
Ubre	Si	0	Si	No, ya que al ser algo más empírico para el modelo no representa importancia alguna.
Pezón	Si	2,118	No	No, no se tiene suficiente información y no brinda información relevante para el modelo.
DI - media de flujos	Si	0	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.
DD - media de flujos	Si	0	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.
TI - media de flujos	Si	0	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.
TD - media de flujos	Si	0	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.
DI - sangre	Si	0	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.
DD - sangre	Si	0	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.
TI - sangre	Si	0	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.

TT - sangre	Si	0	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.
DI - conductividad	Si	0	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.
DD - conductividad	Si	0	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.
TI - conductividad	Si	0	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.
TT - conductividad	Si	0	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.
Usuario	No	2,444	No	No, no se tiene ningún registro.
Destino Leche	Si	0	Si	Si, para la parte de producción puede ser un buen indice para saber si se desecho o no la leche.
Razón de la desviación	Si	2,441	No	No, no se tiene ningún registro.
DI - flujos	Si	0	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.
DD - flujos	Si	0	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.
TI - flujos	Si	0	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.
TT - flujos	Si	0	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.

DI - producciones	Si	0	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.
DD - producciones	Si	0	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.
TI - producciones	Si	0	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.
TT - producciones	Si	0	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.

- **Problemas con los datos**

- Los identificadores se encuentran en el nombre del archivo. Además, algunas vacas no tienen muchos registros.
 - 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**.