

Quality Report (ID Vaca(1204-8794))

- **Resultados de la exploración**

- En este caso, como los registros de cada vaca cambian en consideración a otra vaca, para fines prácticos estaremos analizando solamente el reporte de la vaca **1204.csv**

NOTA: El número total de instancias de cada registro de vaca puede tener más o menos datos pero en general, las conclusiones a las que se llegan en cada columna aplican para todos los registros ya que normalmente todos siguen esta tendencia de tener las mismas columnas sin datos. Se recomienda que se considere un merge de todos los csv como se hace en REPORTE DE ORDEÑO.

- ¿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 fue ordeñada 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	7	Si	Si, la variable como pivote puede ser muy útil para identificar correlaciones.

Número de ordeño	Si	7	No	Si, dependiendo del número de lactancias podríamos detectar mayores productoras y por ende más o menos lactancias.
RCS (* 1000 células / ml)	Si	259	No	No, no se tiene registro en ningún ID.csv.
Patada	Si	242	No	Si, se puede usar como variable Dummy.
Incompleto	Si	231	No	Si, se puede usar como variable Dummy.
Pezones no encontrados	Si	241	No	Si, se puede usar como variable Dummy.
Ubre (Estado AMD)	Si	7	Si	No, ya que al ser algo más empírico para el modelo no representa importancia alguna.
Pezón (Estado AMD)	Si	236	No	No, no se tiene suficiente información y no brinda información relevante para el modelo.
DI (Media de los flujos (kg/min))	Si	7	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.
DD (Media de los flujos (kg/min))	Si	7	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.
TI (Media de los flujos (kg/min))	Si	7	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.
TD (Media de los flujos (kg/min))	Si	7	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.

DI (Sangre (ppm))	Si	7	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 (ppm))	Si	7	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 (ppm))	Si	7	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.
TD (Sangre (ppm))	Si	7	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 (mS / cm))	Si	7	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.
DD (Conductividad (mS / cm))	Si	7	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.
TI (Conductividad (mS / cm))	Si	7	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.
TD (Conductividad (mS / cm))	Si	7	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.

EO/PD (Misc)	Si	0	Si	Si, a partir del cambio en el robot (corral) podemos saber cuando entro a otro robot de ordeña.
Usuario	No	259	No	No, no hay datos suficientes para evaluar.
Destino Leche	Si	7	Si	Si, para la parte de producción puede ser un buen índice para saber si se desecho o no la leche.
Razón de la desviación	Si	259	No	No, no hay datos suficientes para evaluar.
DI (Flujos máximos (kg/min))	Si	7	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.
DD (Flujos máximos (kg/min))	Si	7	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.
TI (Flujos máximos (kg/min))	Si	7	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.
TD (Flujos máximos (kg/min))	Si	7	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.
DI (Producciones (kg))	Si	7	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.
DD (Producciones (kg))	Si	7	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.

TI (Producciones (kg))	SI	7	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.
TD (Producciones (kg))	Si	7	Si	Si, nos puede ayudar para ver alguna relación entre el tiempo de secado y ordeña.

- **Problemas con los datos**

- ID vacas tiene un rango de tiempo de solo Marzo a Junio. Por el rango de tiempo y que en esencia, tiene la misma información que REGISTRO ORDEÑO, se recomienda descartar todos los csv.
- La primera fila del dataset corresponde a diferentes grupos para las columnas. Así que, a la hora de la programación resultará útil usar como header la fila 2 para el nombre de las demás features, es decir, las demás filas.
- Varias columnas, como Patada, Pezones Encontrados, Usuario y RCS no tienen registros útiles o la mayoría son nulos.
- En la categoría de Sangre (ppm) tenemos un grupo de columnas que van desde DI-TD. Estas columnas son pocos registros y en general el proyecto no se enfoca en estos rubros.
- Las variables no se encuentran estandarizadas por completo ya que en algunos registros se ven mezclas de tipos de datos numéricos.
- Varios registros tienen datos nulos.