

# Creación de dataset final (Tercera iteración)

Para crear el dataset final, se observó que los registros de los datos globales del Hato se pueden replicar con la información de los registros de ordeño y las fichas técnicas de las vacas. En el análisis exploratorio del dataset de los datos globales, se determinó que existen variables estadísticamente significativas y relevantes para cumplir con los objetivos, por lo que se tomó la decisión de transformar los datos combinados de ordeño y fichas en una tabla con información similar a la de los datos globales.

La tabla de los datos combinados contiene las siguientes columnas:

Nombre	Descripción
ID_Vaca	Identificador único de la Vaca
Producción Total Diaria (kg)	Producción total diaria de leche en kilogramos
Producción Promedio por Ordeño (kg)	Producción promedio de leche en kilogramos
Desviación Estándar Producción (kg)	Desviación estándar de producción de leche en kilogramos
Número de Ordeños	Número de ordeños realizados al día
Duración Promedio (min)	Duración promedio del tiempo de ordeño de una vaca
Duración Total Diaria (min)	Duración total del tiempo de ordeño diario de una vaca
Desviación Estándar Duración (min)	Desviación estándar del tiempo de ordeño de una vaca
Intervalo Promedio entre Ordeños (min)	Intervalo de tiempo entre ordeños
Desviación Estándar Intervalo (min)	Desviación estándar de intervalo de tiempo entre ordeños
Producción por Minuto (kg/min)	Producción de leche por minuto en kilogramos
Horas entre Ordeños	Horas de intervalo entre ordeños
Lactaciones	Número de lactaciones históricas por vaca
DEL	Días en leche de la vaca

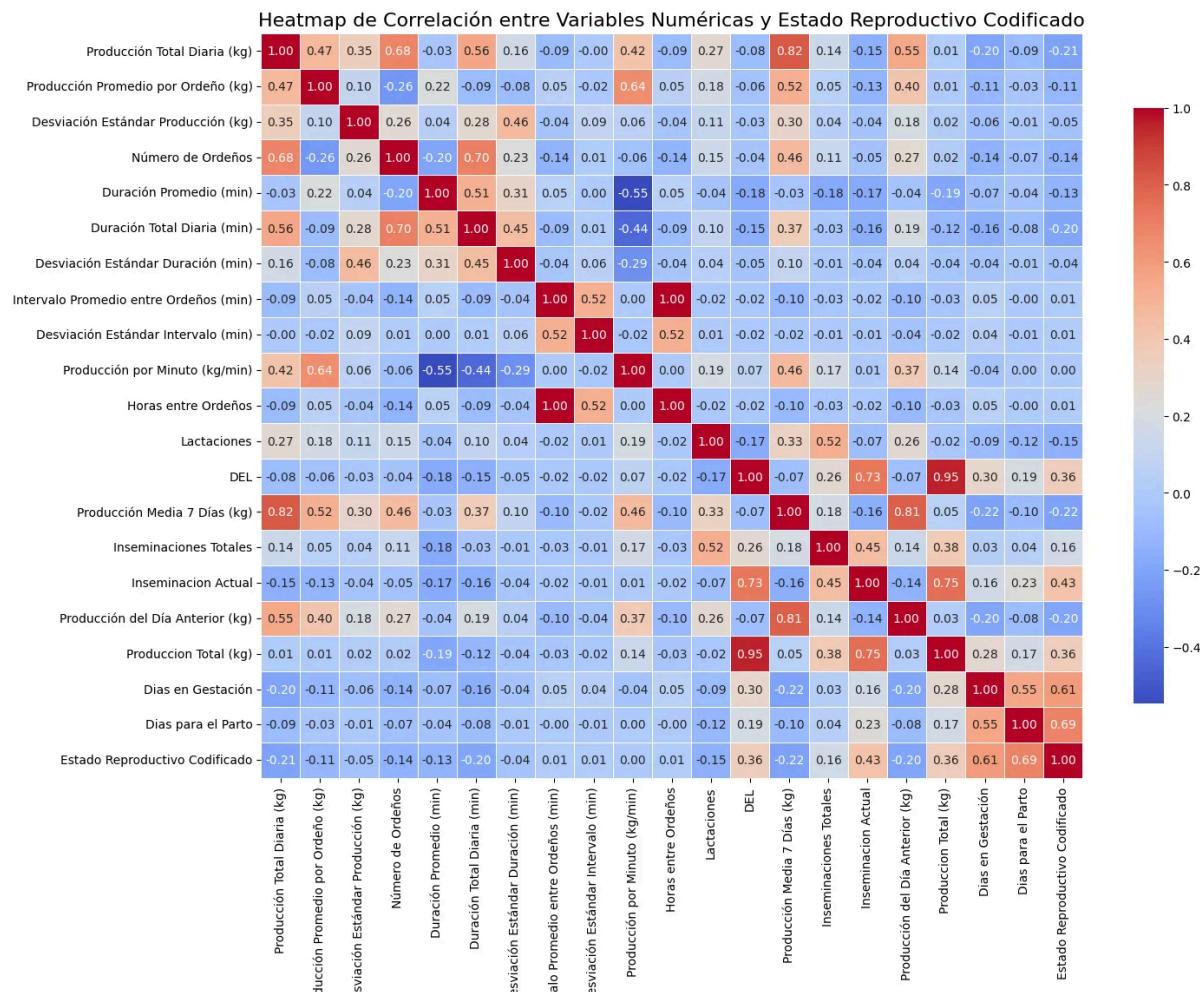
Producción Media 7 Días (kg)	Producción media de leche en los últimos 7 días
Tipo de Evento	Evento registrado de la vaca
Estado Reproductivo	Estado reproductivo de la vaca (fresca, inseminada, preñada, seca...)
Inseminaciones Totales	Número de inseminaciones totales a lo largo de la vida de la vaca
Inseminación Actual	Número de inseminaciones en la lactación actual
Producción del Día Anterior (kg)	Producción de leche del día anterior en kilogramos
Producción Total (kg)	Producción total de leche de la lactación actual en kilogramos
Días en Gestación	Número de días en que la vaca se encuentra preñada
Días para el Parto	Días restantes para el parto de la vaca

La tabla de los datos globales del hato contiene las siguientes columnas:

Nombre	Descripción
Número del animal	Identificador único de la Vaca
Nombre del grupo	Grupo o hato donde se encuentra la vaca actualmente
Número(s) de selección de animal	Sin Información
Estado de la reproducción	Estado reproductivo de la vaca (fresca, inseminada, preñada, seca...)
Días en ordeño	Días en leche de la vaca
Nº Lactación	Número de lactaciones históricas por vaca
Número de inseminaciones	Número de inseminaciones en la lactación actual
Días preñada	Número de días en que la vaca se encuentra preñada
Días para el parto	Días restantes para el parto de la vaca
Fecha de parto esperada	Fecha esperada para el parto de la vaca
Producción de leche ayer	Producción de leche del día anterior en kilogramos
Producción media diaria últimos 7 días	Producción media de leche en los últimos 7 días

Producción TOTAL en lactación	Producción total de leche de la lactación actual en kilogramos
-------------------------------	--

Los datos del dataset combinado (ordeños + fichas técnicas) son muy similares al dataset de los datos globales del hato y las variables más significativas para predecir las variables dependientes se encuentran dentro de los datos globales. Por lo que el dataset final será una transformación de los datos combinados en datos de registros globales de los hatos



Las variables numéricas que coinciden con los datos globales del Hato son:

1. Días para el Parto
2. Días en Gestación
3. Inseminación Actual
4. Producción Total (kg)
5. Días En Leche (DEL)
6. Producción del Día Anterior (kg)
7. Producción Total Diaria (kg)
8. Producción Media 7 Días (kg)

La correlación de estas variables presenta valores absolutos mayores a 0.2. Esto significa que existe una relación lineal apreciable entre ellas:

- Valores cercanos a 1 o -1 indican relaciones fuertes (positivas o negativas).
- Valores entre 0.2 y 0.5 indican una relación moderada, suficiente para sugerir que cuando una variable cambia, la otra también tiende a cambiar de manera consistente, aunque no perfectamente.
- En este caso, el hecho de que los valores absolutos superen 0.2 implica que no son relaciones débiles, y que estas variables sí comparten patrones comunes, lo cual puede ayudar a identificar comportamientos productivos y reproductivos del hato.