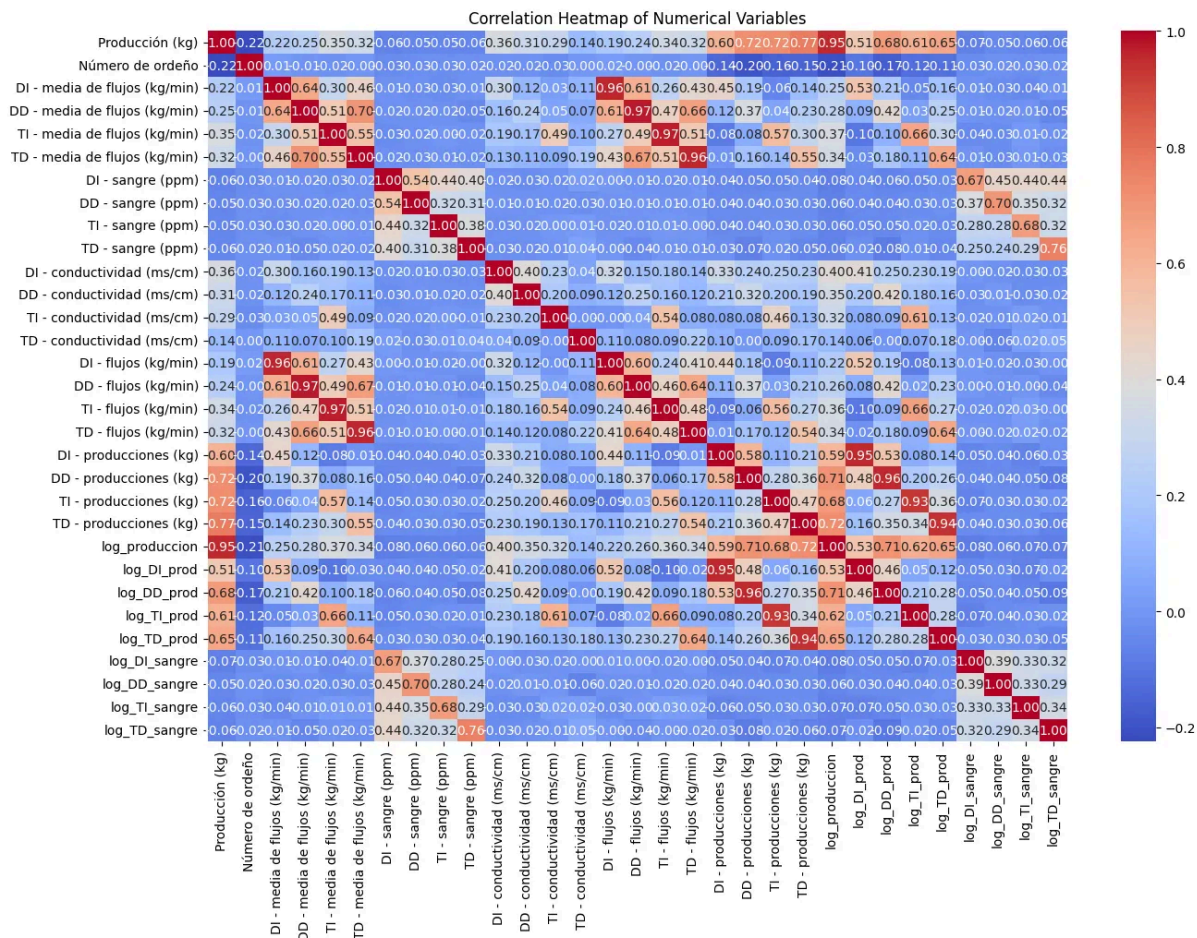


# Creación de dataset final (Segunda iteración)

El resultado de nuestro pipeline es un dataset intermedio llamado aquí `df_cleaned`, que representa la versión depurada y enriquecida de los registros de ordeño resultado de la limpieza, feature engineering y las transformaciones.

Nombre	Descripción
Acción	Tipo de acción registrada durante el ordeño.
Duración (mm:ss)	Duración total del ordeño expresada en minutos y segundos.
Producción (kg)	Producción total de leche generada durante el ordeño.
Número de ordeño	Identificador o número correspondiente al ordeño del día.
DI - media de flujos (kg/min)	Flujo promedio de leche del cuarto Delantero Izquierdo.
DD - media de flujos (kg/min)	Flujo promedio de leche del cuarto Delantero Derecho.
TI - media de flujos (kg/min)	Flujo promedio de leche del cuarto Trasero Izquierdo.
TD - media de flujos (kg/min)	Flujo promedio de leche del cuarto Trasero Derecho.
DI - sangre (ppm)	Nivel de sangre detectada en el cuarto Delantero Izquierdo (ppm).
DD - sangre (ppm)	Nivel de sangre detectada en el cuarto Delantero Derecho (ppm).
TI - sangre (ppm)	Nivel de sangre detectada en el cuarto Trasero Izquierdo (ppm).
TD - sangre (ppm)	Nivel de sangre detectada en el cuarto Trasero Derecho (ppm).
DI - conductividad (ms/cm)	Conductividad eléctrica de la leche del cuarto Delantero Izquierdo.
DD - conductividad (ms/cm)	Conductividad eléctrica de la leche del cuarto Delantero Derecho.

TI - conductividad (ms/cm)	Conductividad eléctrica de la leche del cuarto Trasero Izquierdo.
TD - conductividad (ms/cm)	Conductividad eléctrica de la leche del cuarto Trasero Derecho.
EO/PO	Tipo de extracción: Ordeño Estándar (EO) o Post Ordeño (PO).
Destino Leche	Clasificación del destino final de la leche (aprobada, rechazo, descarte).
DI - flujos (kg/min)	Flujo instantáneo de leche del cuarto Delantero Izquierdo.
DD - flujos (kg/min)	Flujo instantáneo de leche del cuarto Delantero Derecho.
TI - flujos (kg/min)	Flujo instantáneo de leche del cuarto Trasero Izquierdo.
TD - flujos (kg/min)	Flujo instantáneo de leche del cuarto Trasero Derecho.
DI - producciones (kg)	Producción específica del cuarto Delantero Izquierdo.
DD - producciones (kg)	Producción específica del cuarto Delantero Derecho.
TI - producciones (kg)	Producción específica del cuarto Trasero Izquierdo.
TD - producciones (kg)	Producción específica del cuarto Trasero Derecho.
datetime	Marca temporal exacta del ordeño (fecha y hora).
log_produccion	Transformación logarítmica de la producción total.
log_DI_prod	Transformación logarítmica de la producción del cuarto DI.
log_DD_prod	Transformación logarítmica de la producción del cuarto DD.
log_TI_prod	Transformación logarítmica de la producción del cuarto TI.
log_TD_prod	Transformación logarítmica de la producción del cuarto TD.
log_DI_sangre	Transformación logarítmica del nivel de sangre en DI.
log_DD_sangre	Transformación logarítmica del nivel de sangre en DD.
log_TI_sangre	Transformación logarítmica del nivel de sangre en TI.
log_TD_sangre	Transformación logarítmica del nivel de sangre en TD.



La matriz de correlación del dataset procesado evidencia patrones fisiológicos consistentes entre los cuatro cuartos de la ubre, así como relaciones claras entre las características productivas del ordeño y las variables derivadas. El conjunto de datos, resultante de la limpieza y del feature engineering aplicado durante el pipeline, mantiene coherencia estadística y no presenta correlaciones anómalas, lo cual reafirma que no fue necesario aplicar imputaciones o transformaciones adicionales.

En términos generales, se observan cuatro grandes grupos de variables con comportamientos internos muy definidos:

1. Flujos (medios e instantáneos)
2. Producciones por cuarto
3. Sangre (ppm)
4. Conductividad eléctrica (ms/cm)

Adicionalmente, la producción total y sus correspondientes transformaciones logarítmicas presentan relaciones esperadas con los indicadores de flujo y volumen, confirmando consistencia fisiológica y operativa.