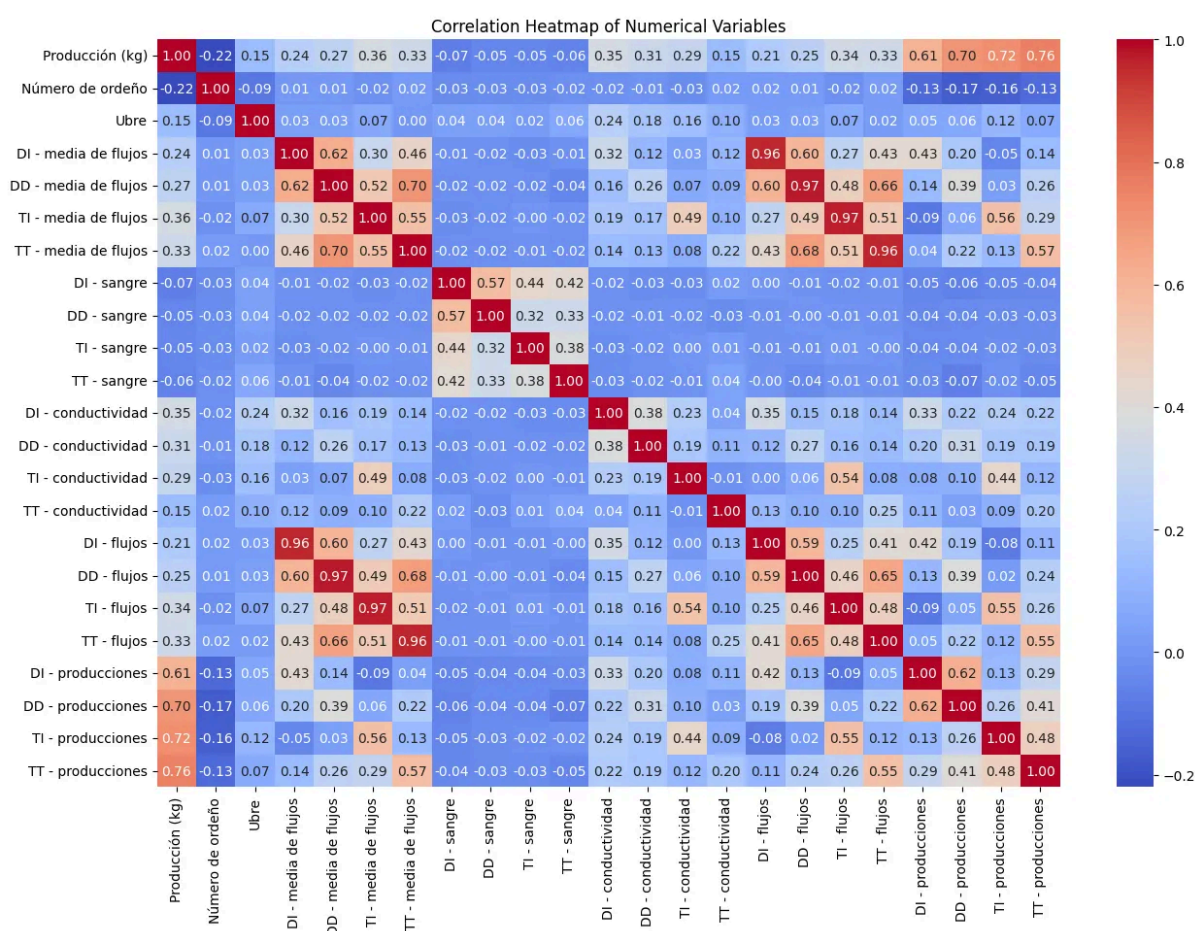


# Creación de dataset final (Primera iteración)

A partir del análisis exploratorio se identificó que el dataset contienen variables operativas y reproductivas que permiten reconstruir una tabla equivalente al registro productivo global. Esto permitió generar un dataset consolidado que resume el comportamiento diario de cada vaca, integrando producción, reproducción y características clave para el modelado y por eso mismo no realizamos más cambios transformaciones, ni imputaciones.

Nombre	Descripción
ID_Vaca	Identificador único de la vaca
Producción Total Diaria (kg)	Producción total de leche generada en el día
Producción Promedio por Ordeño (kg)	Promedio de producción por sesión de ordeño
Desviación Estándar Producción (kg)	Variabilidad en la producción por ordeño
Número de Ordeños	Cantidad de ordeños realizados durante el día
Duración Promedio (min)	Tiempo promedio por sesión de ordeño
Duración Total Diaria (min)	Tiempo total sumado de ordeños durante el día
Desviación Estándar Duración (min)	Variabilidad en la duración del ordeño
Intervalo Promedio entre Ordeños (min)	Tiempo promedio entre cada ordeño
Desviación Estándar Intervalo (min)	Variabilidad del intervalo entre ordeños
Producción por Minuto (kg/min)	Producción relativa por minuto de ordeño
Horas entre Ordeños	Intervalo de horas entre ordeños
Lactaciones	Número de lactaciones históricas
DEL	Días en leche de la vaca
Producción Media 7 Días (kg)	Media móvil de producción semanal
Inseminaciones Totales	Total de inseminaciones a lo largo de la vida

<b>Inseminación Actual</b>	Inseminaciones correspondientes a la lactación actual
<b>Producción del Día Anterior (kg)</b>	Producción registrada el día previo
<b>Producción Total (kg)</b>	Producción total acumulada en la lactación
<b>Días en Gestación</b>	Días transcurridos de la gestación
<b>Días para el Parto</b>	Días restantes para el parto
<b>Estado Reproductivo Codificado</b>	Estado reproductivo transformado a variable numérica



La matriz de correlación revela que las variables medidas en los cuatro cuartos de la ubre presentan patrones muy consistentes: los flujos, las producciones por cuarto, la conductividad y la presencia de sangre muestran correlaciones altas dentro de cada grupo, indicando que estas mediciones se comportan de manera similar entre DI, DD, TI y TT. Por otro lado, la producción total del ordeño muestra relaciones moderadas con algunos valores por cuarto, pero no se asocia fuertemente con sangre o conductividad. En general, las correlaciones fuera de estos grupos son débiles, lo que sugiere que cada conjunto de variables captura dimensiones fisiológicas distintas del proceso de ordeño.

Puntos claves:

- Las variables de flujos presentan correlaciones muy altas entre los cuatro cuartos.
- Las producciones por cuarto también muestran correlaciones moderadas–altas, indicando coherencia fisiológica.
- La conductividad tiene correlaciones moderadas dentro de su grupo, reflejando estabilidad entre cuartos.
- Las mediciones de sangre se correlacionan entre sí de forma moderada, pero poco con producción.
- La producción total (kg) solo muestra correlaciones moderadas con producciones por cuarto, y es independiente de sangre y conductividad.
- La mayoría de las relaciones entre grupos distintos de variables son débiles, lo que indica que cada grupo mide aspectos diferentes del ordeño.