

Collection Report (Consistencia entre datasets)

- Comparando inventario_total_180725.csv y reporte_180725.csv se logró encontrar una coincidencia parcial ya que la mayoría de las vacas están presentes en ambas fuentes, aunque algunas aparecen solo en una.

Aspecto validado	Evidencia	Resultado	Acción sugerida
IDs de vaca	Cruce de claves primarias	Coincidencia parcial	Revisar vacas sin correspondencia
Fechas	Campos Ultima día de Inseminación y Días de Leche	Formatos mixtos (fecha / número)	Estandarizar formato de fecha
Variables repetidas	Columnas duplicadas entre reportes	Baja redundancia	Fusionar solo variables relevantes

- A partir de los nuevos datos obtenidos durante la segunda y tercera visita al CAETEC, comparamos ficha_vacas.csv y ficha_vacas_HATO donde la mayoría de las observaciones son datos únicos debido a que son comentarios que el experto o encargado en turno documentó durante el evento que se registró.
- De igual forma se comparó registro_ordeño.csv y registro_ordeño_HATO donde se logró encontrar una coincidencia en los atributos que se despliegan.
- Para estos datos la mayoría de las vacas no se repiten debido a que en ficha_vacas.csv y registro_ordeño.csv son archivos que contienen solamente la información de las vacas del corral 6 (corral inicial que nos proporcionaron), por otro lado los archivos de ficha_vacas_HATO y registro_ordeño_HATO contienen los datos de las vacas de todo el CAETEC (que fueron solicitados debido a que el objetivo del proyecto requería más vacas fuera del estado de secado).