

Collection Report (Problemas para obtener datos)

1. **Dependencia total de terceros:** El acceso a los datos fue entregado completamente por el equipo docente en función a CAETEC. No teníamos acceso directo a las bases de datos fuente, lo que generó demoras en la recepción y limitó la autonomía del proyecto.
 1. **Disponibilidad de terceros:** Debido a que los datos que solicitamos fueron adicionales, la disponibilidad de tanto la doctora como del representante del SF era muy limitada, hubo varios días de espera sin respuesta. Posteriormente se decidió ir directamente al CAETEC para recopilar los datos, sin embargo de igual forma la disponibilidad de la doctora para la aprobación y envío de los datos terminó en más de una visita muy prolongada de aprox. 6 horas.
2. **Falta de Documentación y Contexto:** Los datasets recibidos carecían de un glosario o documentación. Campos como **Última OCC (x 1000)** o **Éxito IO (h)** no tenían una definición clara, por lo que tuvimos que realizar consultas para interpretarlos, con el riesgo asociado de malas interpretaciones.
3. **Falta de datos históricos:** Los archivos proporcionados solo contienen el estado actual de las vacas, sin conservar versiones anteriores. Esto impide el análisis de series temporales para variables clave como cambios de precios o segmentación de clientes.
 1. Con los archivos nuevos adquiridos esta limitante puede estar solucionada ya que los datos que nos proporcionaron cuentan con registros de hasta 4 lactancias de las vacas lo que significa que estos registros se mantienen guardados durante un periodo muy largo.
4. **Calidad no verificada previamente:** La calidad de los datos era completamente desconocida al momento de la recepción. No existían métricas preliminares de completitud, unicidad o consistencia, trasladando toda la carga de validación a la fase de preparación.