



Es hora de que pongas en práctica todo lo aprendido. 

En este apartado trabajaremos en un **caso práctico de evaluación de calidad de datos**, donde el equipo de datos de una empresa te ha convocado para realizar un **análisis conceptual de la calidad de los datos** y proponer un plan de mejora.

Más adelante conseguirás las resoluciones para que valides tus respuestas y puedas monitorear tu progreso. 

¡Manos a la obra!

1. **Desafío** : Una empresa de logística ha detectado **errores frecuentes** en la entrega de paquetes. Al revisar su base de datos de pedidos, encuentran que muchos registros tienen *direcciones incompletas, códigos postales mal formateados, duplicados, nombres de cliente inconsistentes y fechas poco confiables*.

↪ Tu desafío es:

- a. Identificar qué **dimensiones de calidad** se están viendo afectadas.
- b. Explicar **cómo esos problemas pueden impactar** en los procesos del negocio.
- c. Proponer al menos **3 métricas concretas** que permitan monitorear la calidad de los datos.
- d. Diseñar una estrategia general de **mejora y gobernanza** de los datos.
- e. Reflexionar sobre cómo aplicarías un **assessment de calidad** en este contexto.

2. **¿Dónde se lleva a cabo?** 

En un documento de texto

3. **Tiempo de dedicación** 

1 Hora.

4. Recursos

→ <https://www.talend.com/resources/data-quality/>

→ <https://atlan.com/data-quality-dimensions/>

5. Plus

- Si querés profundizar aún más, podés comparar tu análisis con los principios de la **norma ISO 8000** y evaluar si la empresa los está cumpliendo.
- También podés bosquejar una **plantilla de checklist** que usarías en una auditoría de datos.

6. Condición

Esta práctica o ejercitación **no requiere ser entregada y/o evaluada** por el mentor. No obstante puedes compartir tus resultados con el resto de los bootcampers y construir conocimiento en conjunto.

Resolución del ejercicio:

Ejemplo de análisis:

1. Dimensiones afectadas:

- Compleitud: direcciones incompletas.
- Validez: códigos postales con formato incorrecto.
- Consistencia: nombres distintos para el mismo cliente.
- Exactitud: fechas erróneas o poco confiables.
- Unicidad: registros duplicados.

2. Impacto en el negocio:

- Retrasos en entregas.
- Aumento de reclamos.
- Costos extra por envíos fallidos.
- Baja confiabilidad de los reportes de desempeño.
- Dificultades para hacer segmentación de clientes o análisis predictivos.

3. Métricas propuestas:

- Porcentaje de registros con dirección completa.
- Cantidad de duplicados por mes.
- Índice de registros válidos según formato postal.

4. Estrategia de mejora:

- Establecer reglas de validación en los formularios.
- Usar listas de referencia (catálogo de códigos postales).
- Implementar un flujo de revisión manual para datos críticos.
- Consolidar duplicados usando algoritmos de similitud textual.

5. Assessment de calidad:

- Aplicaría un cuestionario basado en dimensiones clave.
- Revisaría una muestra aleatoria de registros.
- Analizaría la trazabilidad del dato: ¿de dónde viene? ¿quién lo modifica?
- Incluiría entrevistas con usuarios para entender puntos de dolor.