MODULO 05 - EJERCICIO 04

ALEXIS YURI M.

Analiza una base de datos proporcionada por el docente y detecta problemas de calidad según al menos 5 dimensiones (por ejemplo: completitud, unicidad, validez, consistencia y actualidad).

1. Revisión de la Base de Datos – tickets\_soporte.json

La base contiene información de 1.000 tickets de soporte técnico con campos como:

ticket\_id, cliente\_id, descripcion, fecha\_apertura, fecha\_cierre, estado, agente, prioridad, canal, evaluacion\_cliente y tema.

2. Dimensiones de Calidad

Se eligen 5 dimensiones clave para evaluar la calidad de los datos según las siguientes definiciones:

- Completitud: Medir el porcentaje de datos faltantes.

- Unicidad: Identificar registros duplicados.

- Validez: Verificar que los datos cumplan con formatos y rangos válidos.

- Consistencia: Comprobar coherencia entre atributos relacionados.

- Actualización: Revisar si los datos están vigentes y actualizados.

3. Medición de Cada Dimensión

Dimension: Completitud.

Método de Medición: % de valores nulos por columna.

Resultado: descripcion: 3%, fecha\_cierre: 54.9%, evaluacion\_cliente: 50.9%, tema: 4.5%

Dimensión: Unicidad.

Método de Medición: Tickets duplicados por ticket\_id.

Resultado: 73 duplicados.

Dimensión: Validez.

Método de Medición: Fechas con formato correcto.

Resultado: fecha\_apertura: 90% de formato correcto, fecha\_cierre: 100% formato correcto.

Dimensión: Consistencia.

Método de Medición: fecha\_cierre ≥ fecha\_apertura

Resultado: 78.27% de casos consistentes.

Dimensión: Actualización.

Método de Medición: % de tickets abiertos sin fecha\_cierre.

Resultado: 47% de tickets abiertos sin fecha de cierre.

4. Problemas Detectados y Cuantificación de su impacto.

- Alta falta de completitud en fecha\_cierre (54.9%) y evaluacion\_cliente (50.9%).

- Duplicados en ticket\_id (73 casos), lo que impacta trazabilidad y métricas.

- Fechas inválidas en fecha\_apertura (10% de registros).

- Inconsistencias temporales en un 21.73% de los registros que contienen ambas fechas.

- Tickets abiertos sin cierre: casi la mitad (47%) no cuentan con fecha\_cierre.

5. Propuestas de Mejora

- Estandarizar formatos de fecha (AAAA-MM-DD) con validación automática al momento del registro.

- Eliminar duplicados en ticket\_id y asegurar que el identificador sea único y generado por el sistema.

- Definir campos obligatorios para evaluacion\_cliente y fecha\_cierre.

- Aplicar reglas de consistencia que impidan registrar fecha\_cierre anterior a fecha\_apertura.

- Monitorear la actualización de tickets abiertos, implementando alertas para resolverlos en tiempo oportuno.

Conclusión:

La base presenta problemas de calidad que afectan la confiabilidad del análisis, especialmente en completitud y unicidad. La implementación de validaciones automáticas, eliminación de duplicados y control de consistencia mejorará significativamente la calidad y utilidad de los datos para la toma de decisiones.