MODULO 05 - EJERCICIO 05

ALEXIS YURI M.

Analice la base de datos proporcionada por el docente (atenciones\_salud.csv) y detecte problemas de calidad según al menos 5 dimensiones (por ejemplo: completitud, unicidad, validez, consistencia y actualidad).

1. Revisión inicial de la base de datos con 5.000 registros.

Columnas analizadas:

- edad\_paciente

- sexo

- medico

- fecha\_visita

2. Dimensiones de calidad seleccionadas.

- Precisión: en columna edad\_paciente (entre 1 y 120 años).

- Completitud: en columna sexo (mínimo 98% de campos no nulos).

- Completitud: en columna medico (100% de campos no nulos).

- Validez: en columna fecha\_visita (formato debe ser DD/MM/AAAA).

- Consistencia: coherencia lógica de la edad del paciente.

3. Métodos de medición.

- Precisión: porcentaje de registros con edad entre 1 y 120 años.

- Completitud: porcentaje de registros no nulos o no vacíos.

- Validez: formato y existencia real de la fecha de visita.

- Consistencia: edades coherentes con rango lógico.

4. Resultados del diagnóstico.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |
|  | Dimensión | Columna | Umbral (%) | Resultado (%) | Cumple | Observaciones |  |
|  | Precisión | edad\_paciente | 100 | 87,3 | No | 635 registros con edad fuera de rango 0-121 (0, -10, -5, 130, 150, 999) |  |
|  | Completitud | sexo | 98 | 98,9 | Si | 55 registros con sexo vacío. |  |
|  | Completitud | medico | 100 | 89,18 | No | Se encontraron 541 valores nulos. |  |
|  | Validez | fecha\_visita | 100 | 100 | SI | Todos los registros con fechas con formato válido. |  |
|  | Consistencia | edad\_paciente | 99,9 | 89,24 | No | 538 registros con edad incoherente (-5, -10, 999, 150, 0) |  |
|  |  |  |  |  |  |  |  |

5. Impacto de los problemas.

- Edad fuera de rango: afecta el análisis demográfico y segmentación de los pacientes.

- Sexo incompleto: reduce precisión en informes por género.

- Fechas inválidas: dificulta análisis de tendencias en el tiempo.

- Inconsistencias: generan desconfianza y errores potenciales en la toma de decisiones.

6. Propuestas de mejora.

- Implementar validaciones en captura para edades y formatos de fecha.

- Usar listas desplegables (combobox) para el campo sexo.

- Agregar controles automáticos en procesos ETL.

- Monitorear la calidad de los datos usando tableros de control.

- Capacitar al personal encargado de ingresar los datos al sistema.