



INGENIERÍA DE BASE DE DATOS

Ejercicios de Repaso (Examen Final)

INDICACIONES PARA LOS ESTUDIANTES

1. Descargue las bases de datos desde el siguiente repositorio:
<https://github.com/oespinozah/Examen-Final-Repaso>
2. Importe las bases de datos en R y realice un análisis inicial para identificar problemas de calidad.
3. Aplique los criterios de calidad de la norma ISO/IEC 25012 y cualquier problema encontrado.
4. Corrija los problemas identificados y elabore un reporte en RMarkdown para cada caso.
5. Suba los reportes a un repositorio de GitHub, incluyendo un README con una descripción general de su repositorio.

CASO 1

Eres un analista de datos en una empresa que gestiona ventas en diferentes regiones del país. La empresa necesita que la base de datos de ventas cumpla con la norma ISO/IEC 25012, que establece criterios de calidad como precisión, completitud, consistencia, actualidad y trazabilidad. Tu tarea es cargar esta base de datos en R (base_ventas.csv), analizarla y verificar su calidad según estos estándares.

La base de datos con la que van a trabajar contiene las siguientes variables:

- id_vendedor: Identificador único del vendedor (valores posibles: enteros positivos).
- nombre_vendedor: Nombre del vendedor (valores posibles: cadenas de texto).
- zona: Zona asignada al vendedor (valores posibles: 'Norte', 'Sur', 'Este', 'Oeste').
- fecha_venta: Fecha en que se realizó la venta (valores posibles: fechas en formato 'YYYY-MM-DD').
- monto_venta: Monto de la venta en soles (valores posibles: números decimales)



CASO 2

Eres un analista de datos en una empresa de logística que gestiona envíos a nivel nacional. La empresa necesita que la base de datos de envíos cumpla con la norma ISO/IEC 25012, que establece criterios de calidad como precisión, completitud, consistencia, actualidad y trazabilidad. Además de cargar y analizar la base de datos (base_envíos.csv), también deberás crear una variable para la duración del envío y agrupar los envíos por región, calculando la duración promedio del envío.

La base de datos con la que van a trabajar contiene las siguientes variables:

- id_envío: Identificador único del envío (valores posibles: enteros positivos).
- origen: Ciudad de origen del envío (valores posibles: cadenas de texto).
- destino: Ciudad de destino del envío (valores posibles: cadenas de texto).
- fecha_envío: Fecha en que se realizó el envío (valores posibles: formato 'YYYY-MM-DD').
- fecha_entrega: Fecha en que se entregó el envío (valores posibles: formato 'YYYY-MM-DD').
- monto_envío: Monto del envío en soles (valores posibles: números decimales).

CASO 3

Eres un analista de datos en una empresa de marketing digital que gestiona campañas publicitarias en diferentes plataformas. La empresa necesita que la base de datos de campañas cumpla con la norma ISO/IEC 25012, que establece criterios de calidad como precisión, completitud, consistencia, actualidad y trazabilidad. Tu tarea es cargar esta base de datos en R (base_campañas.csv), analizarla y verificar su calidad según estos estándares.

La base de datos con la que van a trabajar contiene las siguientes variables:

- id_campaña: Identificador único de la campaña (valores posibles: enteros positivos).
- nombre_campaña: Nombre de la campaña (valores posibles: cadenas de texto).
- plataforma: Plataforma en la que se ejecuta la campaña (valores posibles: 'Facebook', 'Google', 'Instagram', 'Twitter').
- fecha_inicio: Fecha de inicio de la campaña (valores posibles: formato 'YYYY-MM-DD').
- presupuesto: Presupuesto asignado a la campaña en dólares (valores posibles: números decimales).



CASO 4

Eres un analista de datos en una empresa de seguros que gestiona pólizas de seguros de vida. La empresa necesita que la base de datos de pólizas cumpla con la norma ISO/IEC 25012, que establece criterios de calidad como precisión, completitud, consistencia, actualidad y trazabilidad. Además de cargar y analizar la base de datos (base_polizas.csv), debe crear una variable para la antigüedad de la póliza en años y filtrar las pólizas activas.

La base de datos con la que van a trabajar contiene las siguientes variables:

- id_poliza: Identificador único de la póliza (valores posibles: enteros positivos).
- nombre_cliente: Nombre del cliente (valores posibles: cadenas de texto).
- fecha_inicio: Fecha de inicio de la póliza (valores posibles: formato 'YYYY-MM-DD').
- fecha_fin: Fecha de fin de la póliza (valores posibles: formato 'YYYY-MM-DD' o NA para pólizas activas).
- monto_cobertura: Monto de cobertura en dólares (valores posibles: números decimales).
- estado: Estado de la póliza (valores posibles: 'Activa', 'Cancelada', 'Expirada').

CASO 5

Eres un analista de datos en una clínica que gestiona las citas médicas de sus pacientes. La clínica necesita que la base de datos de citas médicas cumpla con la norma ISO/IEC 25012, que establece criterios de calidad como precisión, completitud, consistencia, actualidad y trazabilidad. Además de cargar y analizar la base de datos, también deberás crear una variable para indicar si la cita fue cumplida o cancelada y filtrar las citas cumplidas.

La base de datos con la que van a trabajar contiene las siguientes variables:

- id_cita: Identificador único de la cita (valores posibles: enteros positivos).
- id_paciente: Identificador único del paciente (valores posibles: enteros positivos).
- nombre_paciente: Nombre del paciente (valores posibles: cadenas de texto).
- fecha_cita: Fecha de la cita (valores posibles: formato 'YYYY-MM-DD').
- hora_cita: Hora de la cita (valores posibles: formato 'HH:MM').
- estado_cita: Estado de la cita (valores posibles: 'Cumplida', 'Cancelada').



CASO 6

Eres un analista de datos en una empresa de telecomunicaciones que gestiona las suscripciones de sus clientes. La empresa necesita que la base de datos de suscripciones cumpla con la norma ISO/IEC 25012, que establece criterios de calidad como precisión, completitud, consistencia, actualidad y trazabilidad. Además de cargar y analizar la base de datos, deberás crear una variable para indicar la duración de la suscripción en meses y filtrar las suscripciones activas.

La base de datos con la que van a trabajar contiene las siguientes variables:

- id_suscripcion: Identificador único de la suscripción (valores posibles: enteros positivos).
- id_cliente: Identificador único del cliente (valores posibles: enteros positivos).
- nombre_cliente: Nombre del cliente (valores posibles: cadenas de texto).
- fecha_inicio: Fecha de inicio de la suscripción (valores posibles: formato 'YYYY-MM-DD').
- fecha_fin: Fecha de fin de la suscripción (valores posibles: formato 'YYYY-MM-DD' o NA para suscripciones activas).
- estado: Estado de la suscripción (valores posibles: 'Activa', 'Cancelada', 'Expirada').

CASO 7

Eres un analista de datos en una empresa de comercio electrónico que gestiona pedidos y envíos a clientes en todo el país. La empresa necesita que la base de datos de pedidos cumpla con la norma ISO/IEC 25012, que establece criterios de calidad como precisión, completitud, consistencia, actualidad y trazabilidad. Además de cargar y analizar la base de datos (base_pedidos.csv), también deberás crear variables para la duración del envío y el mes del pedido y filtrar los pedidos completados. Finalmente, deberás identificar los productos más vendidos por mes, calcular el valor total de ventas por cliente y generar una lista de los 10 mejores clientes.

La base de datos con la que van a trabajar contiene las siguientes variables:

- id_pedido: Identificador único del pedido (valores posibles: enteros positivos).
- id_cliente: Identificador único del cliente (valores posibles: enteros positivos).
- nombre_producto: Nombre del producto (valores posibles: cadenas de texto).
- cantidad: Cantidad del producto pedido (valores posibles: enteros positivos).
- fecha_pedido: Fecha en que se realizó el pedido (valores posibles: formato 'YYYY-MM-DD').
- fecha_envio: Fecha en que se envió el pedido (valores posibles: formato 'YYYY-MM-DD').
- estado_pedido: Estado del pedido (valores posibles: 'Completado', 'Cancelado', 'Pendiente').
- precio_unitario: Precio unitario del producto (valores posibles: números decimales)



CASO 8

Eres un analista de datos en una empresa de servicios financieros que gestiona transacciones bancarias. La empresa necesita que la base de datos de transacciones cumpla con la norma ISO/IEC 25012, que establece criterios de calidad como precisión, completitud, consistencia, actualidad y trazabilidad. Además de cargar y analizar la base de datos, también deberás crear una variable para indicar la cantidad de días desde la última transacción, Filtrar las transacciones completadas, Identificar transacciones con montos anormalmente altos.

La base de datos con la que van a trabajar contiene las siguientes variables:

- id_transaccion: Identificador único de la transacción (valores posibles: enteros positivos).
- id_cliente: Identificador único del cliente (valores posibles: enteros positivos).
- fecha_transaccion: Fecha en que se realizó la transacción (valores posibles: formato 'YYYY-MM-DD').
- monto: Monto de la transacción en dólares (valores posibles: números decimales).
- tipo_transaccion: Tipo de transacción (valores posibles: 'Depósito', 'Retiro', 'Transferencia').
- estado_transaccion: Estado de la transacción (valores posibles: 'Completada', 'Pendiente', 'Fallida').

CASO 9

Eres un analista de datos en una empresa de seguros que gestiona reclamaciones de seguros. La empresa necesita que la base de datos de reclamaciones cumpla con la norma ISO/IEC 25012, que establece criterios de calidad como precisión, completitud, consistencia, actualidad y trazabilidad. Además de cargar y analizar la base de datos, también deberás crear nuevas variables que indiquen la duración de la resolución de la reclamación en días y el mes de la reclamación, filtrar las reclamaciones aprobadas, identificar reclamaciones con montos anormalmente altos, detectar clientes con un número inusualmente alto de reclamaciones en un corto período de tiempo, calcular el valor total de reclamaciones por cliente y generar una lista de los clientes con reclamaciones sospechosas.

La base de datos con la que van a trabajar contiene las siguientes variables:

- id_reclamacion: Identificador único de la reclamación (valores posibles: enteros positivos).
- id_cliente: Identificador único del cliente (valores posibles: enteros positivos).
- fecha_reclamacion: Fecha en que se realizó la reclamación (valores posibles: formato 'YYYY-MM-DD').
- fecha_resolucion: Fecha en que se resolvió la reclamación (valores posibles: formato 'YYYY-MM-DD').
- monto_reclamacion: Monto de la reclamación en dólares (valores posibles: números decimales).



- tipo_reclamacion: Tipo de reclamación (valores posibles: 'Accidente', 'Robo', 'Incendio', 'Otro').
- estado_reclamacion: Estado de la reclamación (valores posibles: 'Aprobada', 'Rechazada', 'Pendiente').

CASO 10

Eres un analista de datos en una empresa de salud que gestiona registros médicos de pacientes. La empresa necesita que la base de datos de registros médicos cumpla con la norma ISO/IEC 25012, que establece criterios de calidad como precisión, completitud, consistencia, actualidad y trazabilidad. Además de cargar y analizar la base de datos (base_registros_medicos.csv), también deberás crear variables para la duración de la estadía y el mes de la admisión, filtrar los registros de pacientes dados de alta. Identificar registros con duraciones de estadía anormalmente largas o cortas, detectar pacientes con un número inusualmente alto de visitas en un corto período de tiempo, calcular el valor total de los costos por paciente y generar una lista de los pacientes con costos sospechosos.

La base de datos con la que van a trabajar contiene las siguientes variables:

- id_registro: Identificador único del registro médico (valores posibles: enteros positivos).
- id_paciente: Identificador único del paciente (valores posibles: enteros positivos).
- fecha_admision: Fecha en que se admitió al paciente (valores posibles: formato 'YYYY-MM-DD').
- fecha_alta: Fecha en que se dio de alta al paciente (valores posibles: formato 'YYYY-MM-DD' o NA para pacientes aún hospitalizados).
- costo: Costo total de la estadía en el hospital en dólares (valores posibles: números decimales).
- diagnostico: Diagnóstico principal del paciente (valores posibles: cadenas de texto).
- estado: Estado del paciente (valores posibles: 'Alta', 'Hospitalizado', 'Fallecido').

Lima, sábado 06 de julio del 2024