

Reto 1: Análisis de un Conjunto de Datos de Ventas

Archivo: ventas.csv

Columnas clave: producto, categoría, precio, cantidad_vendida, fecha, región

Pasos a seguir:

1. **Inspección inicial**
 - o Revisa las primeras filas del archivo.
 - o Verifica que cada columna tenga el tipo de dato correcto (precio y cantidad deben ser numéricos, fecha en formato fecha).
 - o Cuenta cuántos valores nulos hay en cada columna.
 - o Observa si hay valores fuera de rango.
 2. **Limpieza de datos**
 - o Rellena valores nulos en precio con la mediana.
 - o Elimina duplicados si existen registros repetidos.
 - o Asegúrate de convertir la columna fecha al formato de fecha si no lo está.
 3. **Análisis univariado**
 - o Analiza la distribución de cantidad_vendida y precio.
 - o Usa **histogramas** con `sns.histplot()` para ambas columnas.
 - o Utiliza **boxplots** con `sns.boxplot()` para detectar valores atípicos (outliers).
 4. **Detección de outliers**
 - o Usa el rango intercuartílico (IQR) para calcular los outliers en precio y cantidad_vendida.
 - o Identifica cuántos valores extremos existen y decide si conservarlos o no.
 5. **Análisis bivariado**
 - o Usa `sns.scatterplot(x='precio', y='cantidad_vendida', hue='categoría')` para ver si existe relación entre ambas variables y cómo se comportan por categoría.
-

Reto 2: Exploración de Datos de Calidad del Vino

Archivo: winequality-red.csv

Columnas clave: variables químicas (pH, alcohol, sulphates, etc.), quality

Pasos a seguir:

1. **Inspección inicial**
 - o Revisa el número de filas y columnas.
 - o Verifica que todas las columnas sean numéricas.
 - o Observa los valores mínimos, máximos, medias y desviaciones estándar.
 2. **Limpieza de datos**
 - o Revisa si hay valores nulos en columnas como alcohol y pH.
 - o Rellena con la mediana o elimina si hay pocos registros nulos.
 3. **Análisis univariado**
 - o Usa `sns.histplot()` para observar la distribución de quality.
 - o Usa `sns.boxplot()` para detectar valores extremos en alcohol, pH, sulphates, etc.
 4. **Detección de outliers**
 - o Aplica IQR a las variables químicas clave.
 - o Justifica qué hacer con los valores extremos (corregir, eliminar o conservar).
 5. **Análisis bivariado**
 - o Genera un **mapa de calor de correlación** CON `sns.heatmap(corr_matrix, annot=True)` para detectar variables más ligadas a quality.
-

Reto 3: Análisis de Datos de Viajes en Taxi

Archivo: taxi_trips.csv

Columnas clave: distancia, duración, costo, hora_inicio, barrio

Pasos a seguir:

1. **Inspección inicial**
 - o Verifica si hora_inicio está en formato de fecha.
 - o Asegúrate de que distancia, duración y costo sean numéricos.
 - o Observa estadísticas básicas para cada variable.
 2. **Limpieza de datos**
 - o Elimina registros donde la distancia sea cero o negativa.
 - o Rellena valores faltantes en costo con el promedio del barrio.
 3. **Análisis univariado**
 - o Usa sns.histplot() para la distribución de distancia.
 - o Usa sns.boxplot() para encontrar viajes extremadamente largos.
 4. **Detección de outliers**
 - o Aplica IQR a distancia y costo.
 - o Evalúa si son errores de captura o situaciones reales.
 5. **Análisis bivariado**
 - o Crea un sns.scatterplot(x='distancia', y='costo', hue='barrio') para observar si existe proporcionalidad o outliers por zona.
-

Reto 4: Análisis de Datos de Clientes de un Banco

Archivo: clientes_banco.csv

Columnas clave: edad, saldo, ingresos, tipo_cuenta, default

Pasos a seguir:

1. **Inspección inicial**
 - o Observa la cantidad de nulos por columna.
 - o Verifica que las variables edad, saldo e ingresos sean numéricas.
 - o Comprueba los valores únicos de tipo_cuenta.
2. **Limpieza de datos**
 - o Rellena valores nulos en saldo con la mediana.
 - o Convierte tipo_cuenta en una categoría.
 - o Elimina duplicados si los hay.
3. **Análisis univariado**
 - o Usa sns.histplot() para visualizar edad, saldo e ingresos.
 - o Usa sns.boxplot() para detectar valores extremos en esas columnas.
4. **Detección de outliers**
 - o Identifica ingresos y saldos inusualmente altos.
 - o Clasifica clientes fuera del rango típico para investigar más.
5. **Análisis bivariado**
 - o Usa sns.boxplot(x='tipo_cuenta', y='saldo') para ver la dispersión.
 - o Usa sns.scatterplot(x='ingresos', y='saldo', hue='default') para ver patrones de impago.

☁️ Reto 5: Análisis de Datos de Clima

Archivo: clima.csv

Columnas clave: temperatura, humedad, precipitación, fecha, ciudad

Pasos a seguir:

1. Inspección inicial

- o Verifica que fecha esté en formato correcto.
- o Comprueba rangos de temperatura, humedad y precipitación.
- o Revisa si hay valores faltantes por ciudad o día.

2. Limpieza de datos

- o Rellena valores faltantes de precipitación con 0 si es apropiado.
- o Elimina registros con temperaturas fuera de lo razonable (< -50 o > 60 °C).

3. Análisis univariado

- o Usa `sns.histplot()` para observar la distribución de temperatura.
- o Usa `sns.boxplot()` para buscar valores extremos por ciudad.

4. Detección de outliers

- o Aplica IQR a temperatura y humedad.
- o Evalúa si los registros extremos corresponden a eventos climáticos reales.

5. Análisis bivariado

- o Usa `sns.scatterplot(x='temperatura', y='humedad', hue='ciudad')`.
- o Usa `sns.heatmap()` para visualizar correlaciones entre variables.