

MANEJO DE VALORES

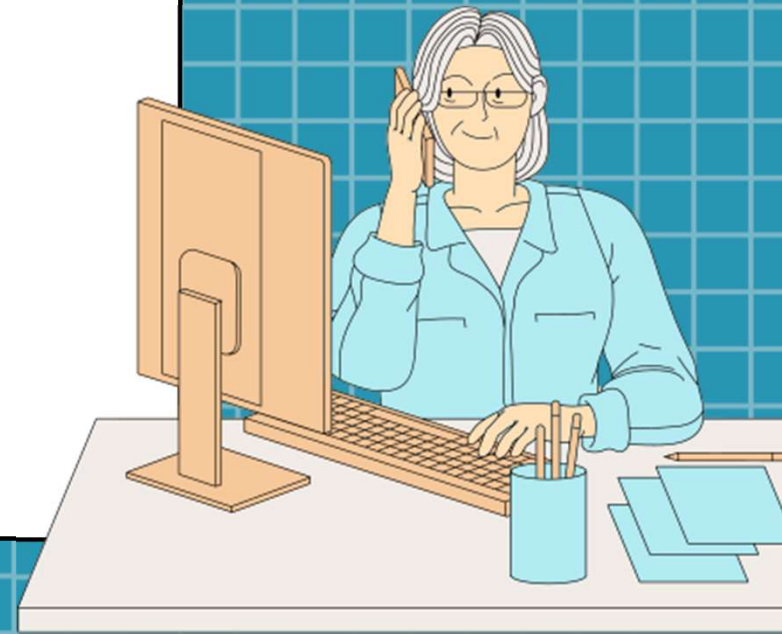
FALTANTES Y

DATOS ATÍPICOS.



¿COMO ASI?

En el análisis de datos, es común encontrarse con valores faltantes y datos atípicos. Estos pueden afectar los resultados y la calidad del análisis, por lo que es crucial manejarlos adecuadamente.



¿QUÉ SON LOS VALORES FALTANTES?

Son datos que están ausentes o no disponibles. Pueden ocurrir por diversas razones, como errores en la recolección de datos o pérdida de información.



IDENTIFICACIÓN DE VALORES FALTANTES

Podemos identificar valores faltantes utilizando:

- **Visualización:** Gráficos de dispersión o matrices de calor.
- **Funciones:** En Python, funciones como `isnull()` y `sum()` de pandas son útiles.



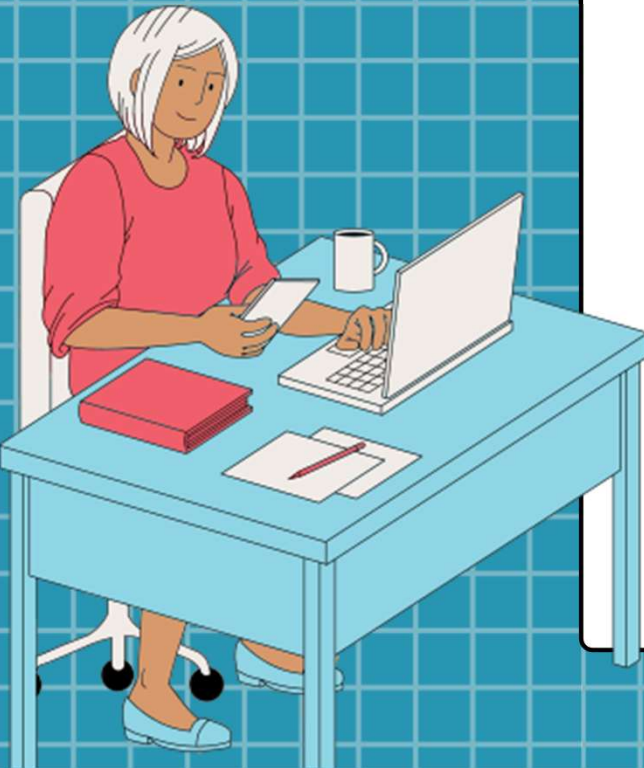
MÉTODOS PARA MANEJAR VALORES FALTANTES

1. Eliminación:

- Eliminar filas o columnas con valores faltantes.
- Útil cuando la cantidad de valores faltantes es pequeña.

2. Imputación:

- Sustituir valores faltantes con la media, mediana, moda o un valor predeterminado.
- Técnicas más avanzadas incluyen el uso de algoritmos de machine learning.



EJEMPLO DE IMPUTACIÓN

Este código reemplaza los valores faltantes en la columna especificada con la media de la misma.

```
import pandas as pd  
df['columna'] = df['columna'].fillna(df['columna'].mean())
```

¿QUÉ SON LOS DATOS ATÍPICOS?

Son valores que se desvían significativamente de otros valores en un conjunto de datos. Pueden indicar errores o variabilidad natural.

Identificación de Datos Atípicos

- **Visualización:** Diagramas de caja (box plots) y gráficos de dispersión.
- **Estadísticas:** Usar la desviación estándar o el rango intercuartílico (IQR).



MÉTODOS PARA MANEJAR DATOS ATÍPICOS

1. Eliminación:

- Eliminar datos que se encuentren fuera de un rango establecido.
- Puede ser necesario si los datos atípicos son errores de medición.

2. Transformación:

- Aplicar transformaciones logarítmicas o cuadráticas para reducir el impacto de los atípicos.

3. Imputación:

- Reemplazar datos atípicos con valores menos extremos, como la media o mediana.



EJEMPLO DE MANEJO DE DATOS ATÍPICOS

Este código reemplaza valores atípicos en la columna especificada con NaN.

```
import pandas as pd  
df['columna'] = np.where(df['columna'] > umbral_superior, np.nan, df['columna'])
```