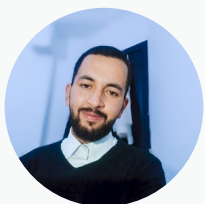




swipe  
to learn

# DATA CLEANING



**BERZIGOU Hamza**

# INTRODUCTION :

**Data Cleaning: A crucial step in data analysis involving refining, correcting, and preparing raw data for meaningful insights and accurate decision-making.**



# IDENTIFY DUPLICATES:

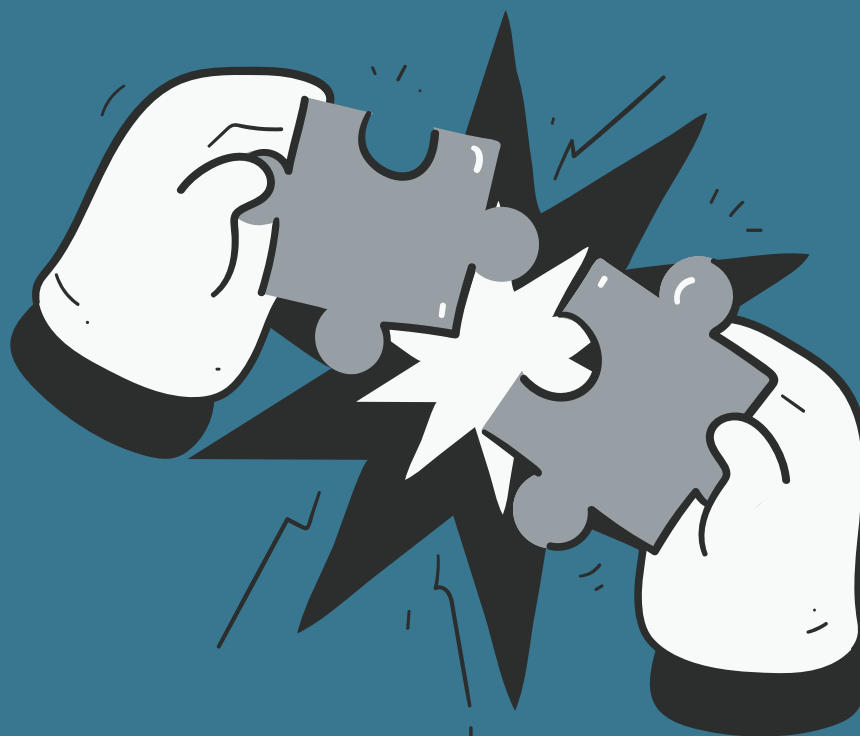
Duplicate detection is vital. It involves finding and removing repetitions to prevent skewed analysis and maintain data integrity.



**BERZIGOU Hamza**

# HANDLE MISSING :

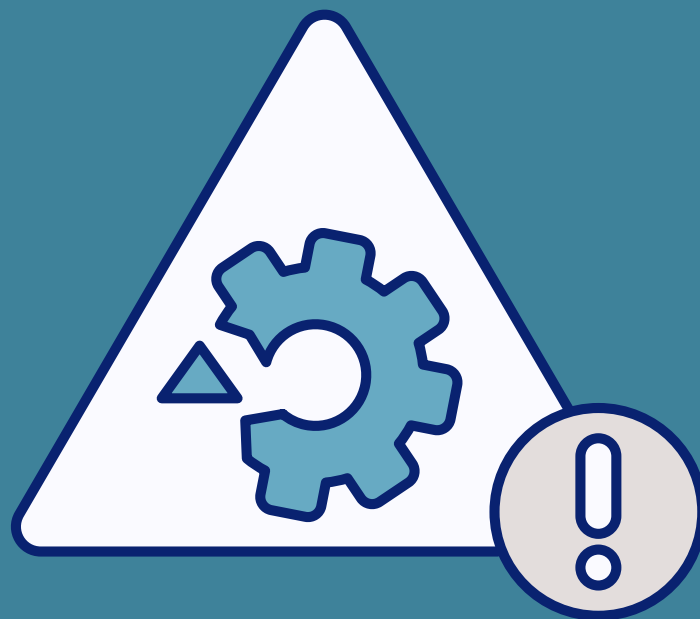
Addressing missing data is key. Techniques include imputation, deletion, or analysis modifications, depending on the context and significance.



**BERZIGOU Hamza**

# CORRECT ERRORS:

Spotting and fixing errors in data sets, like outliers or incorrect entries, is essential for maintaining the accuracy and reliability of your data.



**BERZIGOU Hamza**

# NORMALIZE DATA :

Data normalization involves adjusting values to a common scale, essential for comparing data accurately and effectively in analysis.



**BERZIGOU Hamza**

# DATA FORMATTING :

Consistent data formatting ensures uniformity. This includes standardizing dates, categoricals, and numerical formats for seamless integration and analysis.



**BERZIGOU Hamza**

# VALIDATE QUALITY :

Regular quality checks guarantee data reliability. Validation rules help ensure data accuracy and relevance over time.



**BERZIGOU Hamza**





# FILTER & SORT :

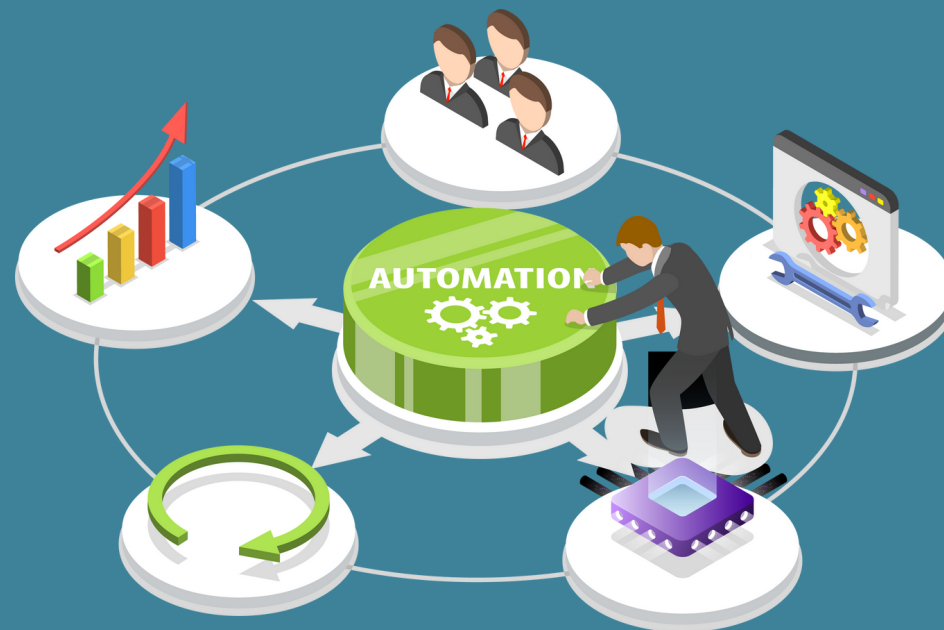
Filtering and sorting data helps in focusing on relevant information, making the analysis process more efficient and manageable.



**BERZIGOU Hamza**

# AUTOMATION TOOLS:

Employ data cleaning tools and automation to streamline the process, reducing manual errors and saving valuable time.



**BERZIGOU Hamza**



# USEFUL



Data Cleaning is an ongoing process, not a one-time task. Regular maintenance ensures continued accuracy and relevance of data.



**BERZIGOU Hamza**



**LIKE IT?  
REPOST IT**

