ONE PAGE SUMMARY

IT & Software › Data Science › Data Wrangling

# Introduction to Data Wrangling

Master the art of transforming raw data into usable insights.

## DEFINITION

> Data wrangling is the process of cleaning and organizing raw data into a structured format. It prepares data for analysis by removing errors and inconsistencies.

## KEY CONCEPTS

### Data Cleaning

This involves correcting or removing inaccurate records from a dataset to improve quality.

### Data Transformation

This is the process of converting data into a suitable format for analysis, such as normalizing values.

### Data Integration

Combining data from different sources to provide a unified view for analysis.

### Data Enrichment

Enhancing existing data by adding relevant information from external sources.

### Data Validation

Checking data for accuracy and completeness to ensure it meets the required standards.

## EXAMPLES

→ Removing duplicate entries from a customer database.

→ Converting date formats from MM/DD/YYYY to YYYY-MM-DD.

→ Merging sales data from multiple regions into a single report.

## MEMORY TIPS

★ Remember 'CITE' for Data Wrangling: Clean, Integrate, Transform, Enrich.

★ Think of data wrangling as 'prepping' data like a chef preps ingredients.

★ Use the acronym 'VET' for validation: Verify, Examine, Test.

## COMMON MISTAKES

✗ Neglecting to document changes made during data wrangling.

✗ Overlooking the importance of data validation before analysis.

✗ Failing to back up original data before making modifications.

## QUICK RECAP

Data wrangling is essential for preparing raw data for analysis. It involves cleaning, transforming, and integrating data to ensure accuracy and usability. Mastering these techniques leads to better data insights.