

UNIT-1

Introduction to Data Wrangling: What Is Data Wrangling?- Importance of Data Wrangling
-How is Data Wrangling performed?- Tasks of Data Wrangling-Data Wrangling Tools-
Introduction to Python-Python Basics-Data Meant to Be Read by Machines-CSV Data-JSON
Data-XML Data.

Introduction to Data Wrangling

What is Data Wrangling?

Data Wrangling, also known as data cleaning, is the process of organizing, and transforming raw data into a structured and usable format for analysis. This critical step ensures that data is accurate, consistent, and ready for decision-making, predictive modeling, or visualization.

Key Steps in Data Wrangling



Data Collection: Gathering data from various sources such as databases, APIs and raw files.

Data Structuring: Organizing data into a structured format (e.g., rows and columns) that aligns with the requirements models.

Data Cleaning: Handling missing values, removing duplicates, and fixing data.

Data Enrichment: Enhancing the data by integrating additional information from other sources to provide context

Data Validation: Verifying the accurate and consistency of the data to ensure it meets quality standards.

Data Transformation: Converting data into a uniform format, including parsing dates, standardizing text, and normalization techniques.

Why is Data Wrangling Important?

Improves Data Quality: Ensures that insights are based on accurate and reliable information.

Saves Time: Streamlines the analytical process by preparing the data in advance.

Supports Better Decision-Making: High-quality data leads to more accurate analysis and predictions.

Scalability: Prepares data for integration with advanced systems like AI and machine learning.