

# **Data Science**



**By:**  
**Dr. Shikha Deep**

# EDA

## Contents :

1. Data Summarization
2. Data Distribution
3. Measuring Asymmetry

# Data Summarization

Data summarization refers to techniques for producing compact representations of data sets by using descriptive measures such as mean, median, variance, frequency distributions, and graphical plots.

- It means getting a quick overview of data using summary statistics (mean, median, std, min, max, count) and tables (frequency counts).

# Data Distribution

Data distribution describes how data values are spread across the possible range of values.

- It shows **shape of data** (normal, skewed, uniform).

# Measuring Asymmetry

Asymmetry of data distribution is measured by:

- **Skewness** → Direction & degree of asymmetry.
  - Skewness  $> 0$  → Right skewed (tail on right).
  - Skewness  $< 0$  → Left skewed (tail on left).
  - Skewness  $\approx 0$  → Symmetric.
- **Kurtosis** → Peakedness of distribution.
  - High kurtosis → heavy tails (outliers).
  - Low kurtosis → flat distribution.

