What is Data?

- A collection of facts, observations, measurements, or information,
    - Can be processed or analyzed.
- Numbers, text, images, audio, or video
- Data is fundamental as it serves as the basis for statistical analysis and inference.

Data is Relevant in Statistics for Reasons

1. Descriptive Statistics:
    - Describe and summarize characteristics of interest in a population or sample.
    - Provide measures,
        - Mean, median, mode, standard deviation, and percentiles,
    - Understand the central tendency, variability, and distribution of the data

2. Inferential Statistics
    - Making inferences or drawing conclusions about a population based on a sample.
    - Inferential statistics use probability theory and sampling techniques,
        - Generalize findings from a subset of data (sample) to a larger population.

3. Hypothesis Testing
    - Test hypotheses and determine the significance of observed differences or relationships.
    - Formulating a null hypothesis, and
    - An alternative hypothesis, collecting data, and assessing the evidence against the null hypothesis.

4. Statistical Modelling
    - Data is employed to build statistical models that represent relationships, patterns, or predictions in data.
    - Models can be used to analyze
        - Complex data
        - Identify trends
        - Make predictions or
        - Simulate scenarios

5. Decision Making
    - Data plays a vital role in decision-making.
    - Insights and evidence to support or guide decisions in various fields, including
        - Business
        - Economics
        - Medicine
        - Social Sciences etc

It is important to ensure that the data used in statistical analysis is accurate, reliable and representative of the population or phenomenon of interest

- Data Quality, including factors such as
    - Completeness

    - Accuracy
    - Consistency and relevance
- Influences the validity and reliability of statistical conclusions.

Different Types of Data

Numerical data:

- Continuous or discrete numerical values.

Two Subtypes:

Continuous Numerical Data:

- Can take any values within a specific range.
- Examples: Age, height, weight, temperature etc

Discrete Numerical Data:

- Consists of whole numbers or integers
- Examples: Number of siblings, number of pets, number of cars etc.
- Use statistical measures such as mean, median, mode, standard deviation, variance, percentiles etc.
- NumPy and Pandas libraries provide functions and methods to perform the calculations.

Categorical Data:

- Non-numerical or discrete values that belong to specific categories

Two Subtypes:

- Nominal Data: Represents categories that have no specific order or ranking.
- Examples: Gender, colour, occupation, marital status, etc.
- Ordinal Data: Represents categories that have a specific order or ranking.
- Examples: Education level, Customer satisfaction rating etc.
- Calculate frequencies, and proportions, and create data visualizations,
    - Bar plots
    - Pie Charts and Histograms
- Using libraries like pandas, matplotlib and seaborn.

Ordinal Data has Order in itself

Example : Performance (Poor, Average and Good)