



# Introduction to Data Science

---



# Table of Content

- **What is Data Science**
- **Why Data Science important**
- **Data Science Workflow**
- **Components of Data Science**
- **Descriptive vs Inferential Statistics**
- **Regression Analysis example**
- **Types of Data**
- **Data Formats**



# What is Data Science?

## **Definition:**

Data Science is the field that uses scientific methods, algorithms, and systems to extract knowledge and insights from structured and unstructured data.

## **Purpose:**

Turn raw data into useful information for decision-making.





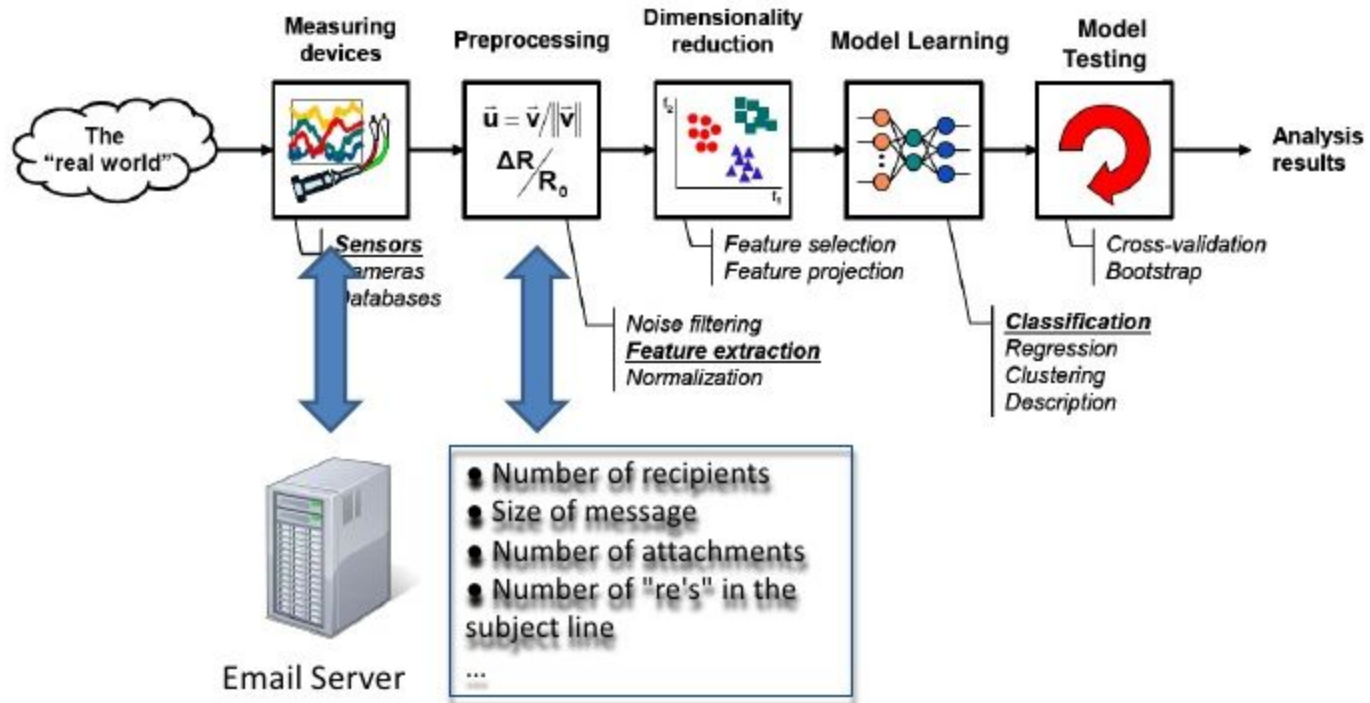
# Why is Data Science Important?

- **Data is growing exponentially**
- **Organizations rely on data to make decisions**
- **Drives innovation in healthcare, finance, e-commerce, and more**





# Data Science Workflow





# Components of Data Science

- **Statistics & Mathematics**
- **Programming (Python, R, etc.)**
- **Data Visualization**
- **Machine Learning**
- **Domain Knowledge**



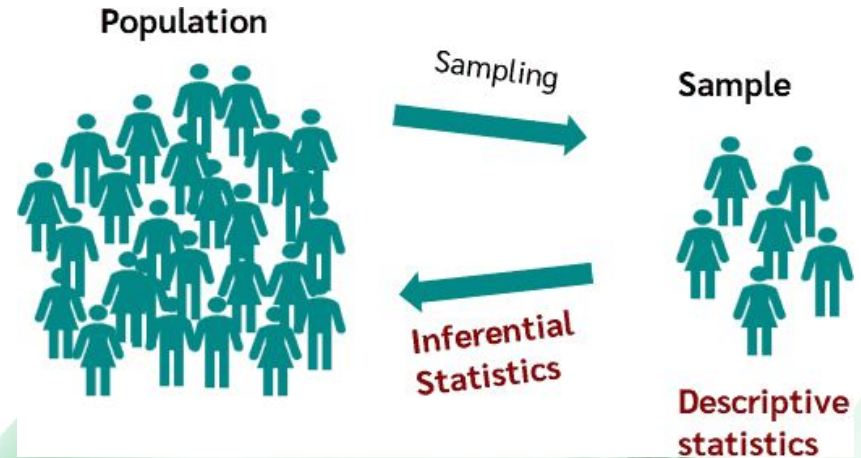
# Introduction to Statistics in Data Science

Statistics is the foundation of data analysis.

Helps in summarizing, understanding, and interpreting data.

## Two main types:

- Descriptive Statistics
- Inferential Statistics





# What is Descriptive Statistics?

## Definition:

Descriptive statistics involves methods for organizing and summarizing data.

## Purpose:

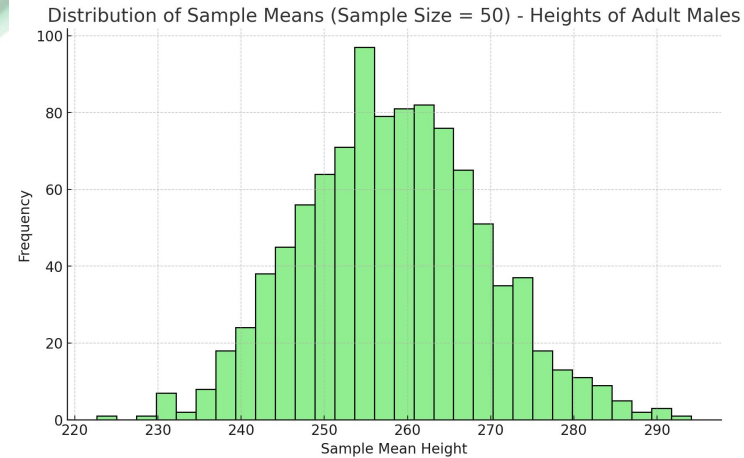
To describe patterns and characteristics of a dataset without making conclusions beyond the data.

## Includes:

Measures of central tendency: mean, median, mode

Measures of dispersion: range, variance, standard deviation

Data visualization: histograms, boxplots, bar charts







# What is Inferential Statistics?

## Definition:

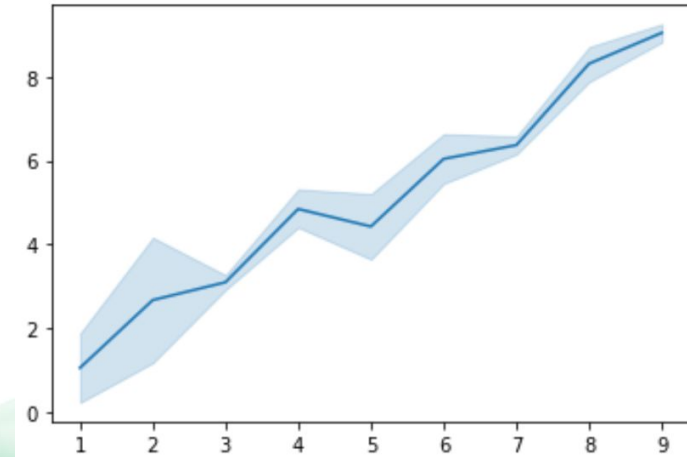
Inferential statistics uses sample data to make generalizations or predictions about a larger population.

## Purpose:

To make conclusions, test hypotheses, and determine relationships.

## Includes:

- Hypothesis testing
- Confidence intervals
- Regression analysis





# Linear Regression Analysis with Example:

To understand the relationship between variables.

**Goal:** Predict student exam score based on hours studied.

**Formula:**

$$\text{Score} = a + b \times (\text{Hours Studied})$$

Where:

a = intercept (score when study hours = 0)

b = slope (how much score increases with each hour studied)

Ex: If a = 50 and b = 5, then studying for 4 hours gives:

$$\text{Score} = 50 + 5 \times 4 = 70$$



# What is Data?

## Definition:

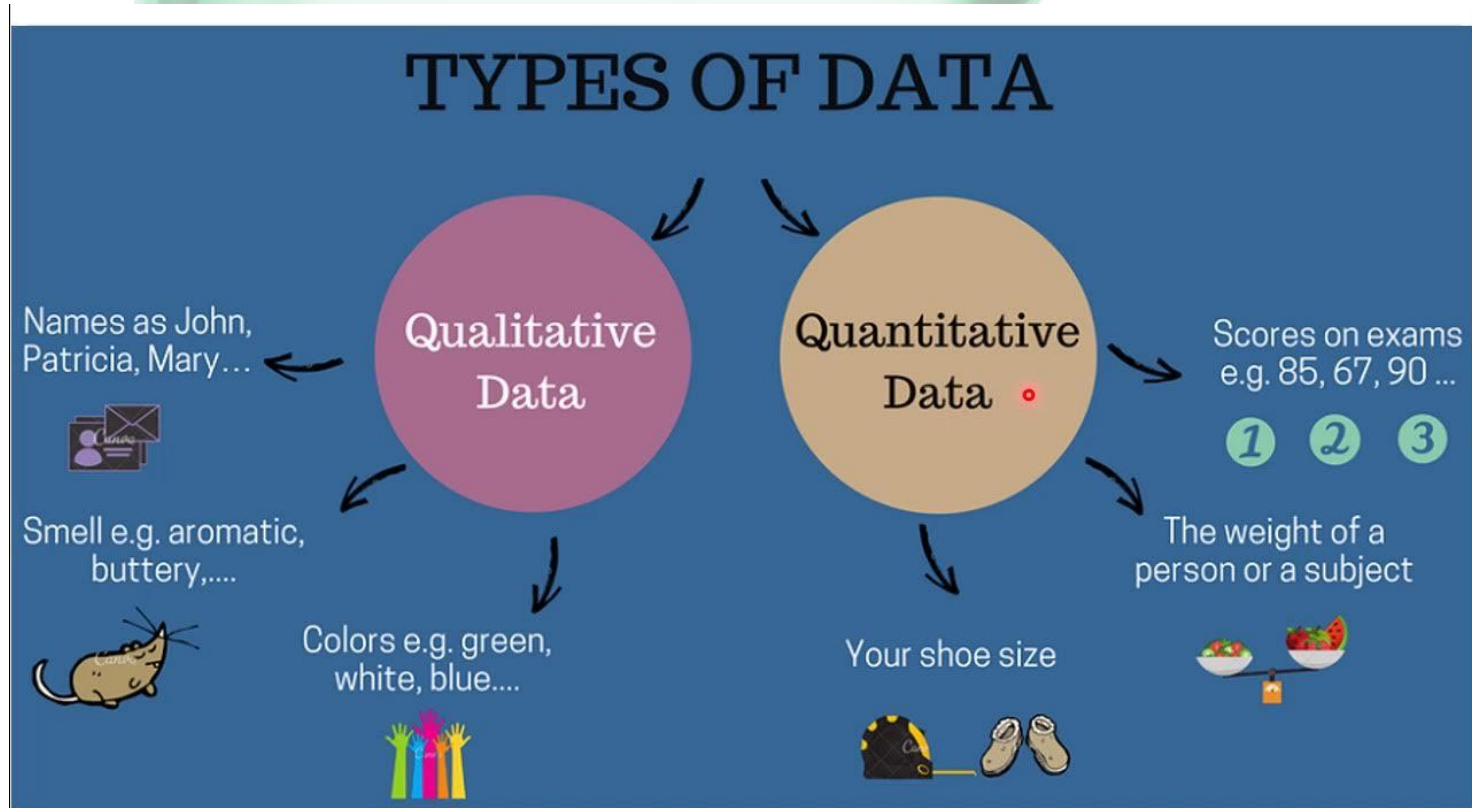
Data is a collection of facts, figures, or information that can be processed and analyzed.

## Types of Data:

- Qualitative vs Quantitative
- Structured vs Unstructured



# Qualitative vs Quantitative Data



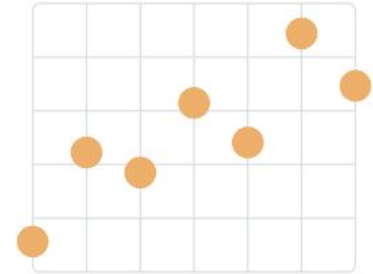


# Quantitative (Numeric) Data

These variables represent measurable quantities and can be either:

- **Discrete:** Countable, finite values.
  - **Examples** include the number of students in a class, the number of cars in a parking lot. Can not be decimal.
- **Continuous:** Infinite possibilities within a range.
  - **Examples** include height, weight, temperature.

● ● ● DISCRETE DATA



● ● ● CONTINUOUS DATA



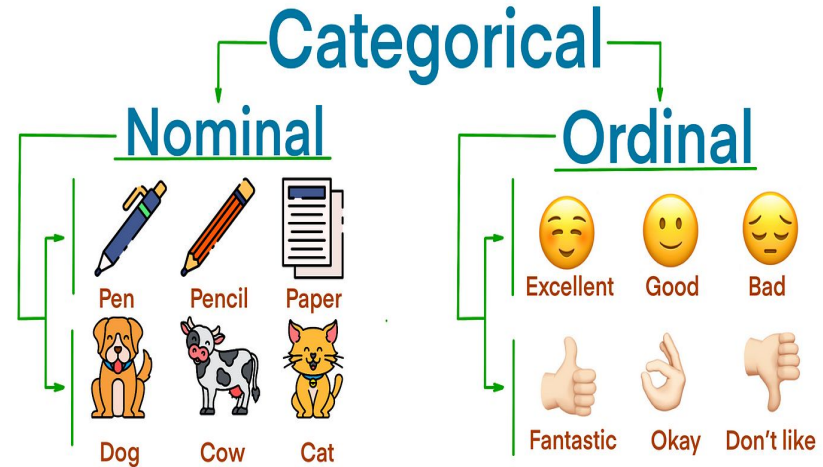




# Qualitative(Categorical) Variables

These variables represent categories or groups and can be either:

- **Nominal:** No inherent order among categories.
  - **Examples** include colors (red, blue, green), gender (male, female).
- **Ordinal:** There is an inherent order, but the difference between levels is not quantifiable.
  - **Examples** include customer satisfaction ratings (poor, fair, good, excellent), education levels (high school, bachelor's, master's).





# Structured vs Unstructured Data

Type	Description	Example
Structured	Organized in rows/columns	Excel sheets, SQL tables
Unstructured	Free-form, lacks structure	Images, emails, audio, text



UNSTRUCTURED DATA



STRUCTURED DATA



# Data Formats

**Structured:** CSV, Excel, SQL

**Semi-structured:** JSON, XML

**Unstructured:** Images, Videos, Emails, Social Media

Book1.csv - Notepad

File Edit Format View Help

```
Product,Price[$],Color,Material,Weight [lbs],Original Design,Warranty [years]
Door stopper,25,Grey,Metal,0.5,Yes,1
Hanger,60,Brown,Wood,12,Yes,2
Chair,80,Green,Wood,10,No,2
Table,120,Green,Wood,20,Yes,3
Vase,20,White,Glass,2,Yes,1
```

Viewer Text

Paste Copy Format Remove white space Clear Load JSON data

```
{
  "Students": [
    {
      "Name": "Amit Goenka",
      "Major": "Physics"
    },
    {
      "Name": "Smita Pallod",
      "Major": "Chemistry"
    },
    {
      "Name": "Rajeev Sen",
      "Major": "Mathematics"
    }
  ]
}
```



# Thank you

---