# UNIT-1

## *DATA EVOLUTION*

FIRST LET US KNOW ABOUT

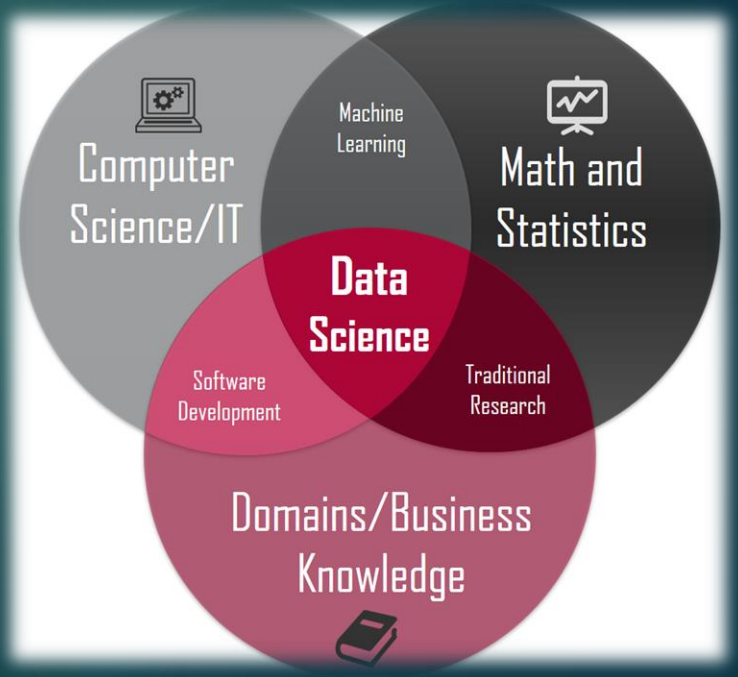
# WHAT IS DATA?

- Data is information such as facts and numbers used to analyze something or make decisions.
- The information that has been translated into a form that is efficient for movement or processing.
- A data is a collection of discrete values that convey information, describing the quantity, fact, statistics, other basic units of meaning, or simply sequences of symbols that may be further interpreted.

- NOW LET'S SEE DATA SCIENCE

# DATA SCIENCE

- Data science is the field of study that combines domain expertise, programming skills, and knowledge of mathematics and statistics to extract meaningful insights from data.
- The domain of the study that deals with vast volume, and uses algorithms.
- Data science includes developing strategies for analyzing data, exploring, and visualizing the data



# Data to Data science

- The vast amount of collected data is converted into a file, data file, or in .xlsx, CSV, etc.

Data
A group of persons data is collected
Age, hobbies, height, weight

Converted data in tabular form

| s.no | age | Hobbies | Height | Weight |
|---|---|---|---|---|
| 1 | | | | |
| 2 | | | | |
| 3 | | | | |

- This tabular form is known as a dataset.
- These datasets are used in computer programming languages like python, etc

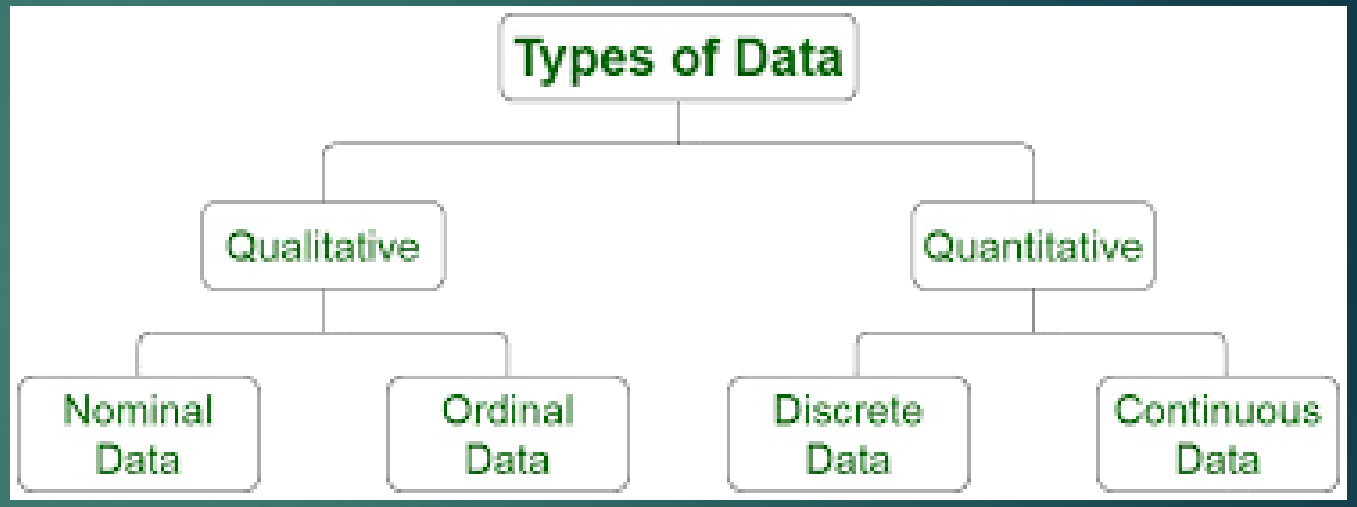
# Data Scientist


- Data scientists are analytical experts who have technical skills.
- As we all know, today the world is paying a huge amount for a data scientist.
- Data scientist salary in India ranges between 4.5 lakh rupees to 25.9 lakhs

# Types of data

- There are two types of data.

1. Qualitative
    a. Nominal data
    b. Ordinal data
2. Quantitative
    a. Discrete data
    b. Continuous data



# Lets briefly look at the types of data

## Nominal vs Ordinal Data

### Nominal Data

Nominal data is used just for labeling variables, without any type of quantitative value. The name 'Nominal' comes from the Latin word "nomen" which means 'name'.

The nominal data just name a thing without applying it to an order. Actually, the nominal data could just be called "labels."

### Ordinal data

Ordinal data is data which is placed into some kind of order by their position on a scale.

Ordinal data may indicate superiority. However, you cannot do arithmetic with ordinal numbers because they only show sequence.

### Examples

- Gender (Women, Men)
- Hair color (Blonde, Brown, Brunette, Red, etc.)
- Marital status (Married, Single, Widowed)
- Ethnicity (Hispanic, Asian)

### Examples

- The first, second and third person in a competition.
- Letter grades: A, B, C, and etc.
- When a company asks a customer to rate the sales experience on a scale of 1-10.
- Economic status: low, medium and high.

intellspot.com

## Discrete vs Continuous Data

| DISCRETE | EXAMPLES | PICS |
|---|---|---|
| Discrete data is a count that involves only integers. The discrete values cannot be subdivided into parts. For example, the number of children in a class is discrete data. You can't count 1.5 kids. | • The number of students in a class.<br>• The number of workers in a company.<br>• The number of home runs in a baseball game.<br>• The number of test questions you answered correctly |  |

| CONTINUOUS | EXAMPLES | PICS |
|---|---|---|
| Continuous data could be meaningfully divided into finer levels. It can be measured on a scale or continuum and can have any numeric value. For example, you can measure your height at very precise scales — meters, centimeters, millimeters, etc. | • The amount of time required to complete a project.<br>• The height of children.<br>• The square footage of a two-bedroom house.<br>• The speed of cars. |  |

intellspot.com

# Data Evaluation

- By the name itself we can see that the data is evaluated here.
- Data Evaluation is a scientific and statistical evaluation to determine what type of data, quantity, and quality is needed either during a project or at the end of the project.

| | Indicator [How will we know it?] | Methods [How will we gather the information? What tools shall we use?] | Timing [When do we collect the data?] | Responsibility [Who will collect the data?] | Notes [What should we keep in mind?] |
|---|---|---|---|---|---|
| Evaluation Question 1: [To what extent ...] | | | | | |
| 1.1 | | | | | |
| 1.2 | | | | | |
| 1.3 | | | | | |
| Evaluation Question 2: [In what ways ...] | | | | | |
| 2.1 | | | | | |
| 2.2 | | | | | |
| 2.3 | | | | | |
| Evaluation Question 3: [To what extent ...] | | | | | |
| 3.1 | | | | | |
| 3.2 | | | | | |
| 3.3 | | | | | |
| Evaluation Question 4: [How appropriate ...] | | | | | |
| 4.1 | | | | | |
| 4.2 | | | | | |
| 4.3 | | | | | |

# *Data source*

- Data source is nothing but a location where data is being used to originate from.
- The source can be a flat file, XML file, or any other format that a system can read.
- There are three types of data sources:

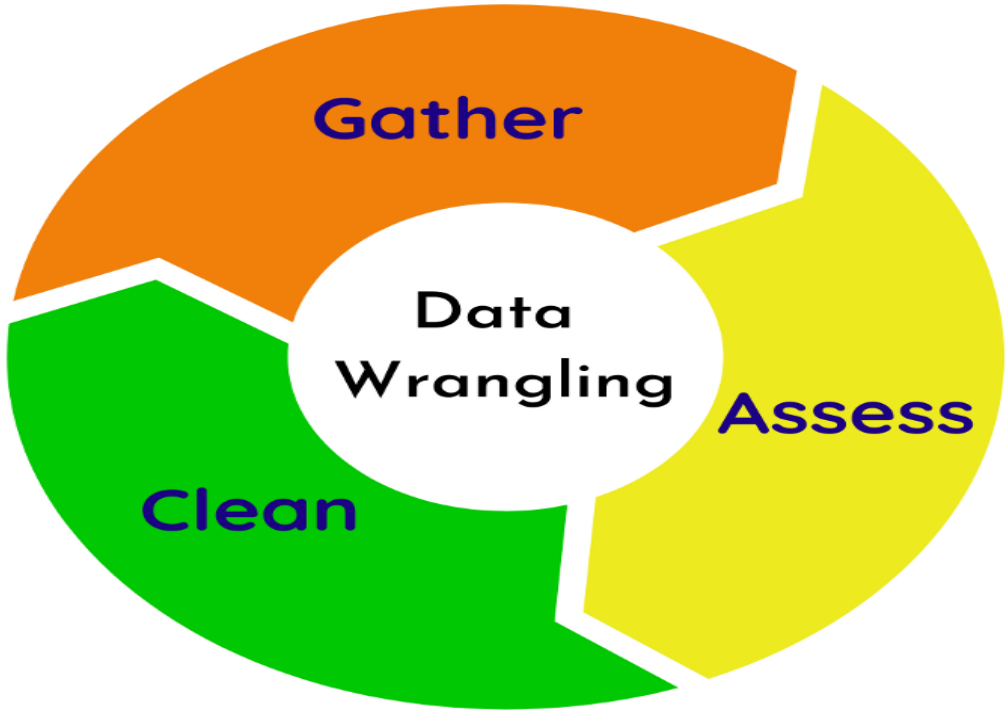
a) Relational

b) Multidimensional

c) Dimensional modelled relational

# Data wrangling

- Data wrangling is the process of removing errors and combing complex data sets to make them more accessible and easier to analyze.
- Data wrangling includes cleaning data, converting one form of data into another form of data, mapping data, and storing the data.



# Data all around us

- ✓ Data is available everywhere in the world in our day-to-day life.
- ✓ Data can be presented in the form of numbers, such as time, date, temperature, etc.
- Data is collected by online surveying, many other ways.

# Main concept of data science

Collection of data, Design, and Analysis

- Collection of Data from various sources.
- Design the data means how the data should be gathered and programming
- Analysis: Output of the data and what you observed or noticed from the above data.

# Big data

- Big data refers to significant volumes of data that cannot be processed effectively with the traditional applications that are used.
- The processing of this big data begins with raw data that isn't aggregated and is most often impossible to store in the memory of a single computer.

- In simple words: The data that contains greater variety, arriving in increasing volumes.
- Big data is larger, more complex data sets, especially from new data sources.

# Thank you