Data Science

- Data science: multidisciplanary → uses - scientific methods, Algorithms, statistics, Visualisation, extract info from structured, Unstructured data.

- Big data: collection of data i.e, volume, Variety, Veracity, Velocity, Value (5Vs)
  - Veracity ↓ Uncertain & Inconsistancy.

→ AI > ML > NN > DL .

→ Datatypes and Scales -

Datatypes :- structured, Unstructured, semi structured, Quasi structured (Textual data + inconsistent)

:- Cross-sectional data - Time series data - Panel data

- Cross-sectional data ↓ Many variables collected same time or duration
- Time series data ↓ Single variable collected several times (week, month)
- Panel data ↓ Several variables collected several times. (Ex: Unemployee rate of countries)

→ Object ; Attributes :-

Qualitative

Attributes :- Nominal - Binary - ordinal - Continuous and discontinuous

- Numeric : (Interscaled, Ratio scaled). → Quantitative.

Nominal :- Name of things - Name of symbols: Ex: Lecturer, White, Professor.

ordinal :- Values with meaningful sequence (Ranking) Ex: Medium, High, Low.

Binary :- ⌈ Symmetric : Both values are equal Ex: Gender
(yes/No) ⌊ Asymmetric: Both values not equal Ex: Result.

Numeric :- Real/integer value, measure quantity.
↳ ⌈ Interval-scaled :- Diffirence of adjacent values consistent : Ex - Temparature,
- No predefined starting point (True zero val) - Dates.
↳ Ratio-scaled :
- Diffirence of adjacent values Ex: Height, weight, Length.
- Predefined True zero value No zero in diffirence of Height.

Continuous :- Infinite values, measured, Subdivisible, float Ex: Height, weight, time

Discontinuous: Specific values, counted, non subdivisible Ex: No of students, No of cars.

§→ Population :- set of possible observations

Sample :- Logical subset of the population.

§ Measures of centraltendency :-

1. Mean:
$\mu = \frac{\Sigma x}{N} = \frac{x_1 + x_2 \dots}{N}$

2, Weighted mean

3, Median
$= \frac{n+1}{2}$

4, Mode:
Most often occures.

Types of Data Analytics :-

1. Descriptive

- summarize historical data to identify patterns
- Used Aggregate functions on database.

2. Predictive

- Predicts future
- Predict probability of future occurrences
- Ex: Regression classification

3. Prescriptive

- Choose optimal actions to perform on isights from Descriptive and predidictive analysis

Ex: Linear programming
Meta-heuristics Alg.

Steps in Data Science:

Step 1: Setting Research goal:-
- Define research goal ⇒ Well defined, Deliverable
- Create project charter ⇒ objectives, Resources, Timeline.

Step 2 Retrieving Data :-
- Internal Data Ex: Servers. →
  - Data retrieval
  - Data ownership
- External data Ex: Facebook

Step 3: Data Preparation: (Preprocessing).

Step 4 : Data Exploration: -
- Simple graphs
- combined graphs
- Link and brush
- Non-graphic Techiques

1. Histogram
2. scatter plot
3. Box plot
4. pie chart
5. Bar.

Step 5 Build the model:-
- Model selection
- Model execution
- Diagnosis and model comparision.

$$MSE = \frac{1}{n}\Sigma(y-\hat{y})^2$$

Step 6 Presentation and Automation:-
- Presenting Data
- Automating Data Analysis