Stats and Maths for Data Science

Topics/Points to Be Covered

1. Need of Statistics and Mathematics

2. Basic Terminology in Statistics

3. Types of Statistics

4. Introduction to probability

Need of Statistics and Mathematics

VisionNLP

In this emerging AI world data plays a major role.

Statistics is used to analyze the data and draw meaningful conclusions out of the data.

Statistics can be applied in various fields such as data science, research, social science, etc.

In the world of technology, numerous amount of data is being generated every second. e.g. online payment, post or comment on the social media sites, medical reports Etc.

Importance of Statistics

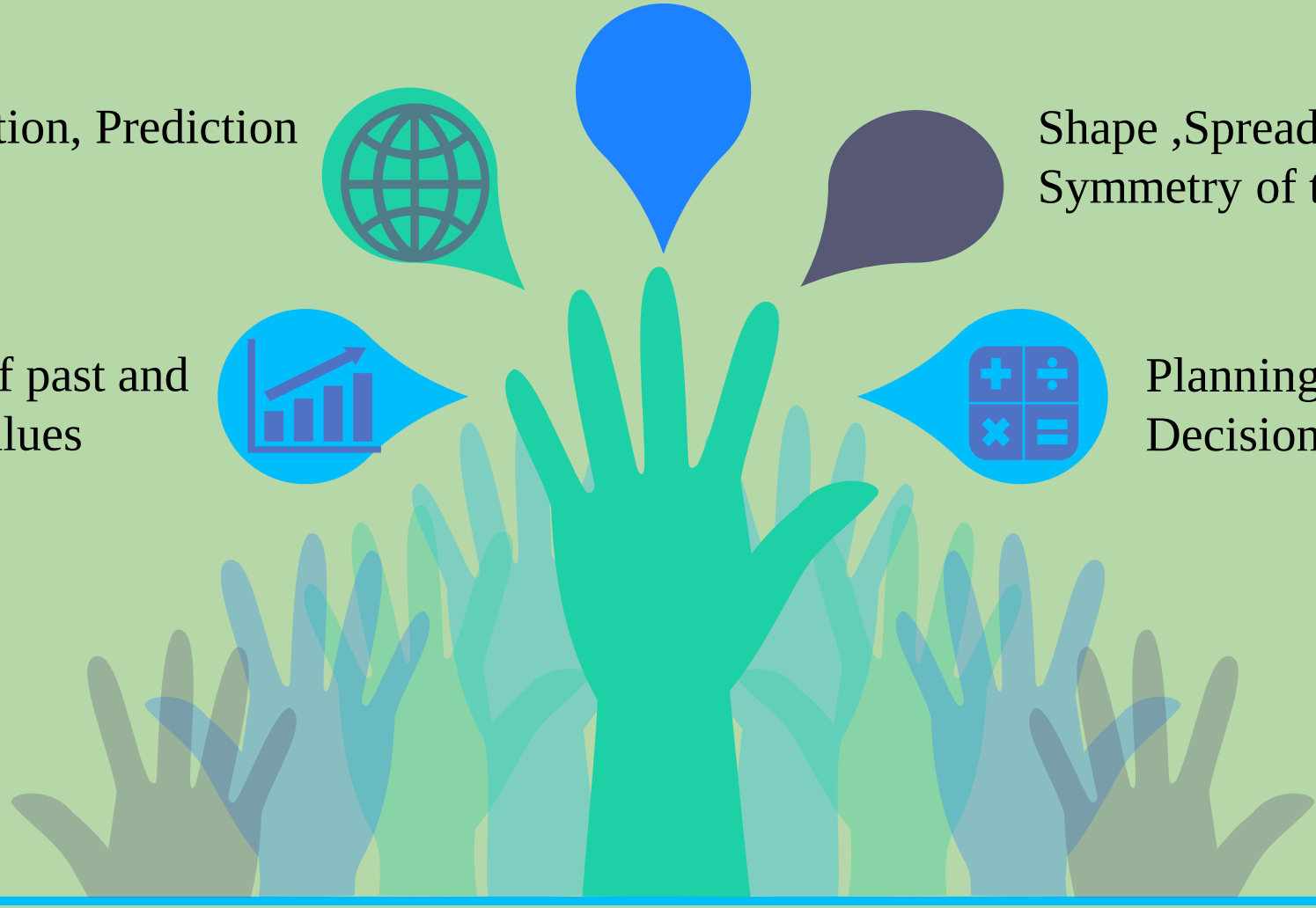
Collection of data

Estimation, Prediction

Shape ,Spread , and Symmetry of the Data

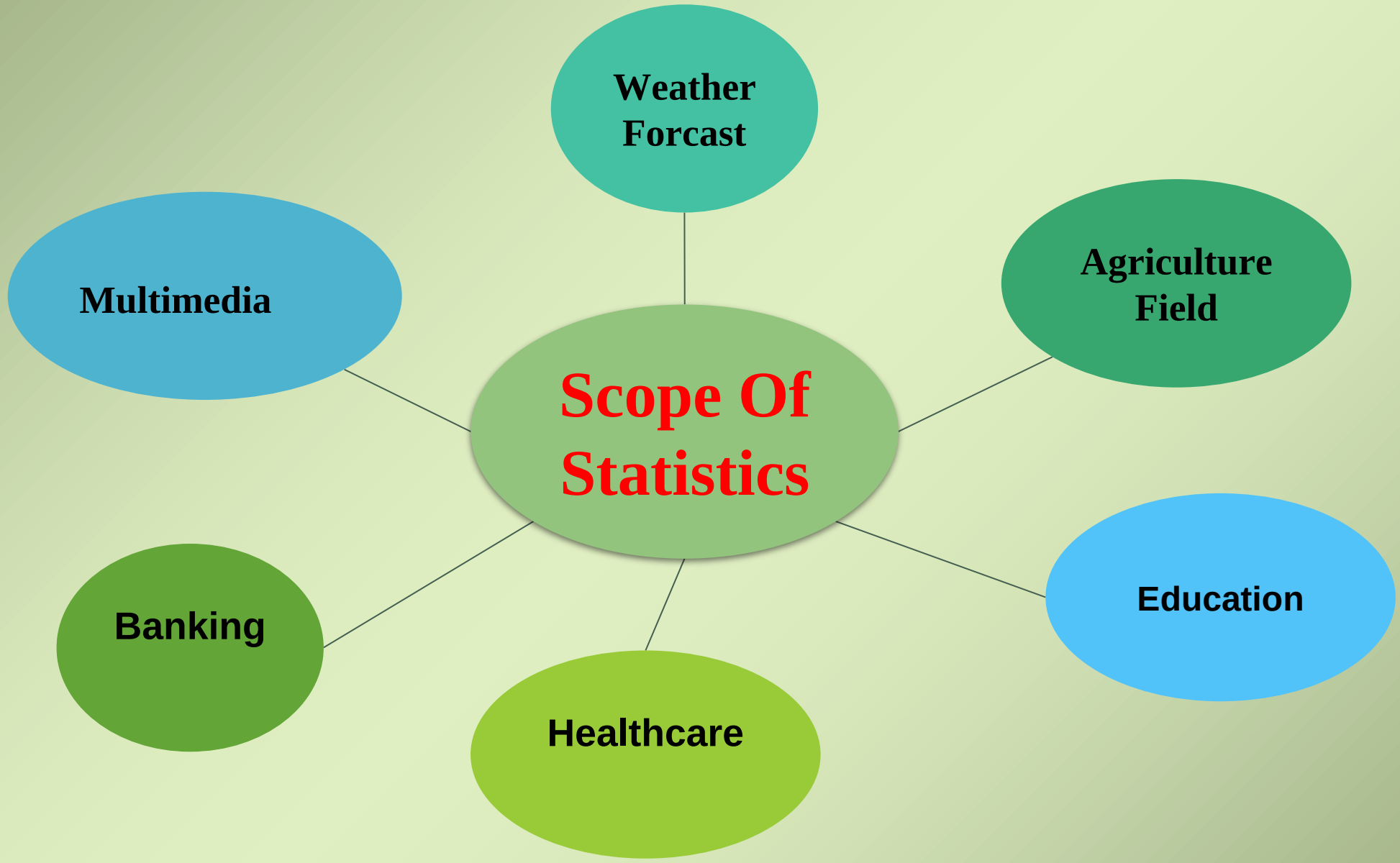
Comparison of past and future values

Planning, Controlling, Decision Making etc



Where we can use the Statistics

VisionNLP



Basic Terminology in Statistics

VisionNLP

Statistics: - Statistics is the branch of the Mathematical Science which deals with Collection, presentation, analysis and interpretation of the data.

Data:- The Information collected for statistical study is called the Data

Way of looking into problem by two person:

1. **Statistician**

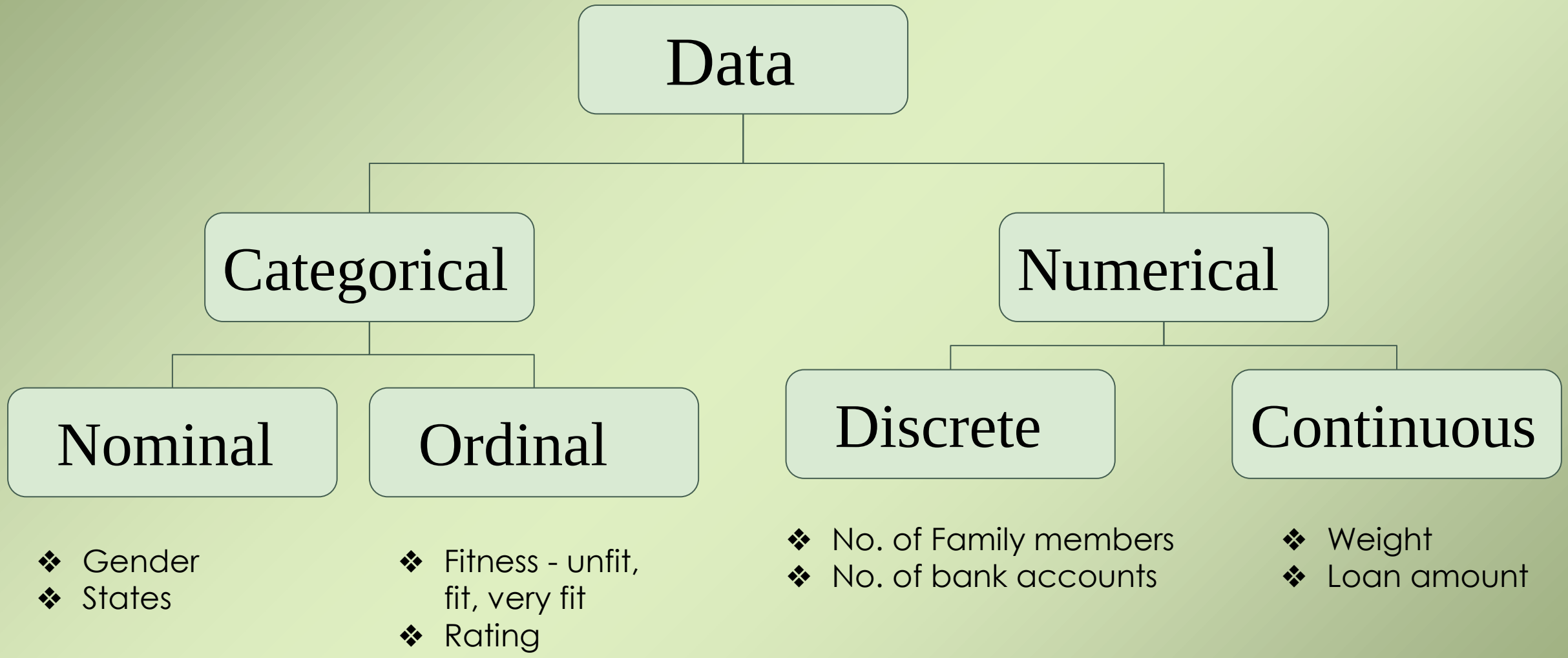
problem → we collect the data

1. **Machine learning Engineer**

We have a data → we will give a solution

Basic Terminology in Statistics

VisionNLP



Types Of Statistics

01

Descriptive Statistics

Collect, Organize, Summarize and present the data.
Tools:- Measures of Central Tendency and Dispersion.

For ex. Doctor suggests a patient to do a covid-19 test, based on what? symptoms… Here doc is not doing an inference, or building a predictive model or ai system.

02

Inferential Statistics

Making inference about the population from Sample or Drawing a conclusion ,Performing Estimation etc.
Tools:- Hypothesis Testing ,ANOVA.

For ex. How effective covishield vaccine is?

image credit: https://www.cuemath.com/data/terms-of-probability/

Introduction To Probability

What is Probability ?

There are a number of events in our day-to-day life and we're not sure whether it will occur or not. But we're eager to know which one will occur.

For instance,

1. it will rain today or not
2. chance of winning in a particular game eg. Tossing a coin

Every time we need to explain what is the chance of an event to occur, we talk in terms of Probability

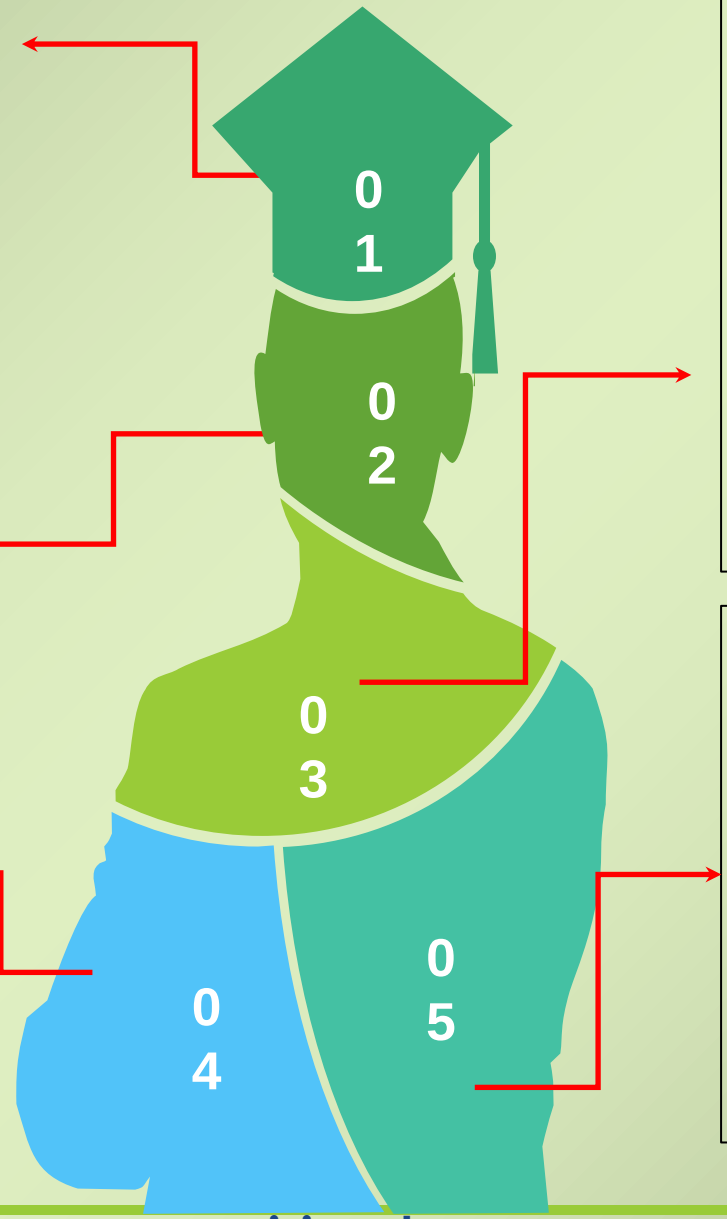
Basic Terminology in Probability

VisionNLP

Experiment:- Any process or experiment like rolling a dice, tossing a coin, drawing a card from a well-shuffled deck, etc.

Outcome:- Outcome means result of any event.
E.g Tossing of coin ,drawn out card, etc

Sample Point:- Element of the sample space is called the *Sample point.*
E.g. In the below sample space,
Ω = {HH, TT , HT, TH }
Here, H and T are the *Sample Point.*



Sample Space: - A set of all possible distinct outcomes in a random experiment is called sample ***space.*** Sample space is always denoted by the symbol omega (Ω)

E.g. Tossing of a two coin simultaneously

The sample space is,

Ω = {HH ,TT , HT, TH }

Event:-Any subset of the sample space of the random experiment is called an Event.
E.g.Tossing two coins simultaneously
The sample space is
Ω={HH,TT,HT, TH }
The events may be,
A ={HH} B ={HT} etc

Types of Events

VisionNLP

Simple Event

If the event contains only one sample point then it is called a **Simple Event.**

E.g. tossing of a single coin simultaneously
The sample space is,
Ω = {H, T}
A = {H}, B = {T}

Compound Event

If the event contains more than one sample point then it is called a Compound Event.

E.g. rolling of the single Die
The sample space is,
Ω = {1, 2, 3, 4, 5, 6}
The events may be,
A ={1,2} B ={1,2,3,4} ,C = {1,2,3,4,5,6} etc.

Sure Event

If the event contains all the sample points of the sample space then it is called a Sure Event

E.g. tossing of a single coin simultaneously
The sample space is,
Ω = {H, T}
Event A = getting either H or T..
A = {H,T} is called a Sure Event.

Impossible Event

An event that does not contain any sample point of the sample space is called an Impossible Event.

E.g. Rolling of the single Die
The sample space is,
Ω = {1, 2, 3, 4, 5, 6}
Event A = Getting a number multiple of 10.
A = {}

Mutually Exclusive Event :-

An event A and B said to be **Mutually Exclusive** if there is no common point between them.

If ball is red then it cannot be a black ,

E.g. rolling of the single Die

The sample space is,

$\Omega = \{1, 2, 3, 4, 5, 6\}$

Event A = Getting Odd no

$A = \{1,3,5\}$

Event A = Getting Even no

$B = \{2,4,6\}$

Complements Of Events:-

It's the collection of all sample points in the sample space which are not present in the Event A.

The complements of event A in sample space is denoted by A^c

E.g. $\Omega = \{1, 2, 3, 4, 5, 6\}$

$A = \{1,2\}$, $A^c = \{3,4,5,6\}$

$P(A^c) = 1 - P(A)$

Exhaustive Event:- The event A and B are said to be exhaustive when at least one of the events compulsory occurs.

$A \cup B = \Omega$

E.g. rolling of the single Die

The sample space is,

$\Omega = \{1, 2, 3, 4, 5, 6\}$

Event A = Getting Odd no

$A = \{1,3,5\}$

Event A = Getting Even no

$B = \{2,4,6\}$

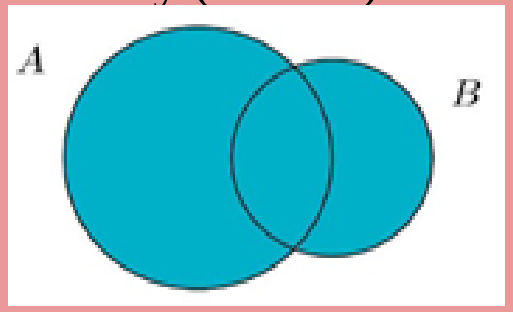
$A \cup B = \{1,2,3,4,5,6\}$

Thus ,

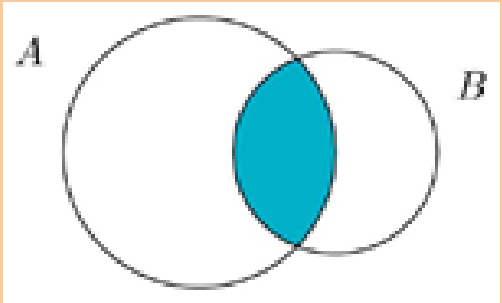
$A \cup B = \Omega$ i.e. A and B are **Exhaustive Events.**

Union of Events: - An event A and B is said to Union if at least one of two events A and B [either events A will happen or B or both A ,B will happen .

Its denoted by (A ∪ B)



Intersection of Two Events :-(A∩B) An event A and B is said to Intersection if happening of both A and B together.



Independence Of Event:- Independent Events are not affected by previous events.

The event has no effect on the probability of another event occurring.

I.e the probability that one event will occur has no impact on the probability of the other event that is taking place.

For instance, two independent events will be when:

Event A: you are rolling a dice

Event B: you are flipping a coin

By definition, Event A and B is said to be independent if

$$P(A \cap B) = P(A) * P(B)$$

Definition of Probability :- Probability is the likelihood or chance of an event occurring.

Formula:
P(A) = m /n where m = favourable outcome , n = total outcome.

For Example: Getting a Data Science job in first attempt. Here chances are 50% i.e 0.5

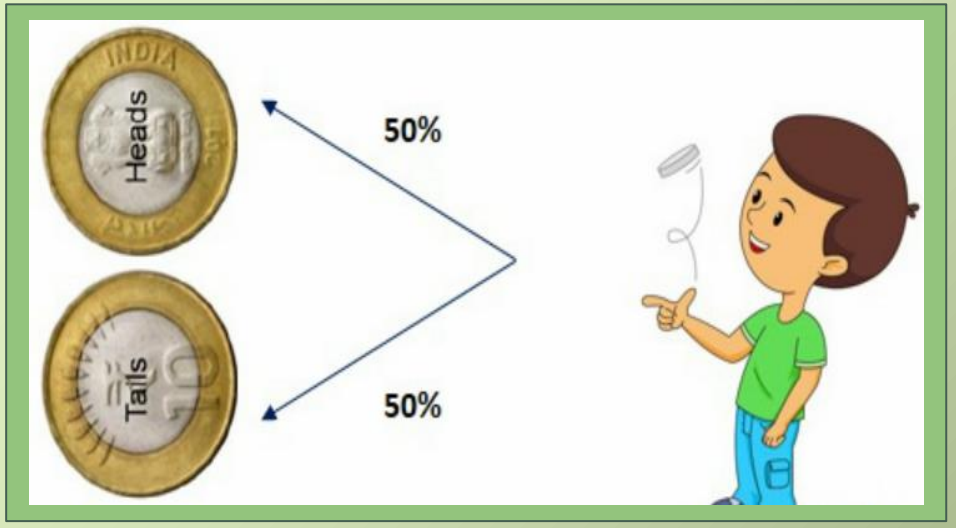
Remember this:

1. **Probability of any event always Lies between $0 < P \leq 1$.**
2. **Probability of all events always sum up to 1.**

Examples...

E.g. 1. Tossing of a two coin simultaneously
The sample space is,
Ω = {HH ,TT , HT, TH }
Find 1. Two heads
2. Prob.of getting One Head.
Solution:-

1. m = favourable outcome = 1
   n = total outcome = 4
   P(A) = m /n P(A) = ¼



1. Prob.of getting One Head.
   m = favourable outcome = 2
   i.e A = {HT ,TH}
   P(A) = m /n
   P(A) = 2 /4 P(A) = 1 /2 Thus , Prob.of getting One Head is ½ i.e. 0.5

E.g. 2.

A card is drawn at random from a pack of 52. Find the prob. Of

1. Drawing a red card.
2. Drawing a picture card.
3. Drawing a King card.

Solution:-

There is 52 cards in pack so,

n = 52

1. Let event A denote drawing a red card

Prob of drawing A red card is,

m = total no of favorable outcome is 26

P (A) = 26 /52 = 1/2

Prob of drawing A red card is = 0.5

2. Let event b is Drawing a picture card

there are total 12 picture card in a pack , so the total no of favorable outcome is 12

m = 12

P(B) = m/n = 12 / 52 = 0.23

The Prob of Drawing a King card is for Homework

E.g. 3 .

Rolling of the single Die

The sample space is,

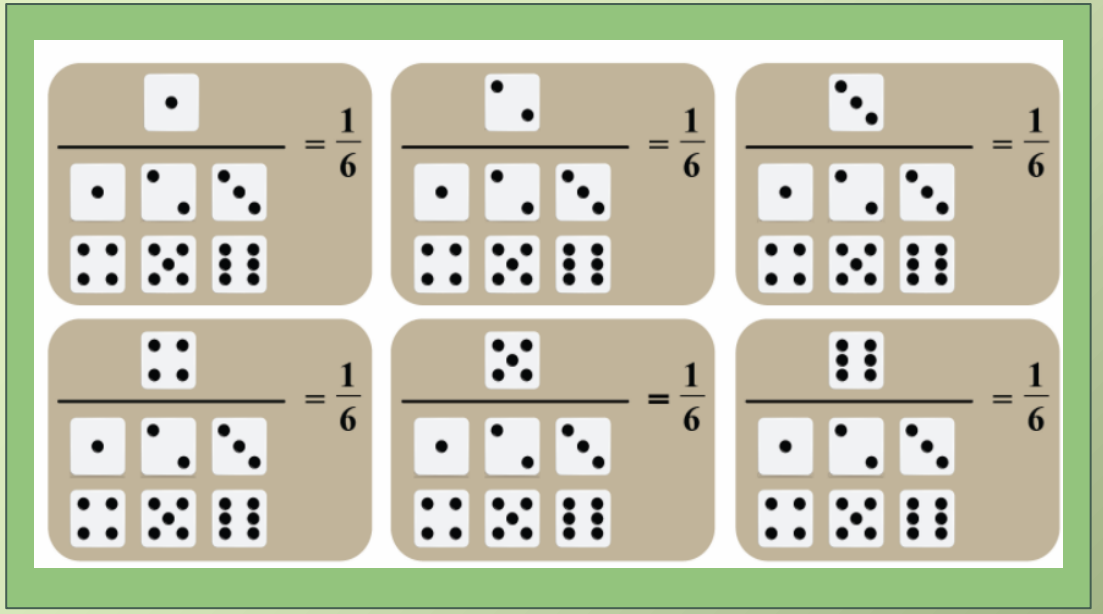
$\Omega = \{1, 2, 3, 4, 5, 6\}$

Find - 1. Prob of getting odd no is,

$A = \{1,3,5\}$

$P(A) = m/n = P(A) = 3/6 = 1/2$

P(A) = 0.5



2. Prob of getting no Greater than 4 is,

$B = \{5,6\}$

$m = 2$, $n = 6$

$P(B) = m/n$ $P(B) = 2/6$

P(B) = 0.33

Factorial:-

The Product of first n natural numbers is called a factorial.

Factorial is denoted by the symbol

“ ! “

n! = n*(n-1) *(n-2)*.....3*2*1.

Example:-

5! = 5*4*3*2*1

= 120

3! = 3*2*1

= 6.

Combination:-

A group of same or all of given objects without considering their order is called combination and is given by,

$$nCx = \frac{n!}{(n-x)! * x!}$$

$$nCx = \frac{n!}{(n-x)! * x!}$$

$$5C3 = \frac{5!}{(5-3)! * 3!}$$

$$5C3 = \frac{5 * 4 * 3 * 2 * 1}{2! * 3!}$$

$$5C3 = \frac{5 * 4 * 3 * 2 * 1}{(2 * 1) * (3 * 2 * 1)}$$

$$5C3 = \frac{60}{6}$$

$$5C3 = 10$$

Sampling

Population

Entire group Of People or Items Under Study

Sampling

Process of Selecting Sample From the Population

Sample

A portion OR Some part of the Population

Sampling

- Why not study everyone?? Why sampling
  - Lower cost
  - High speed of data collection
  - When it's impossible to study whole population

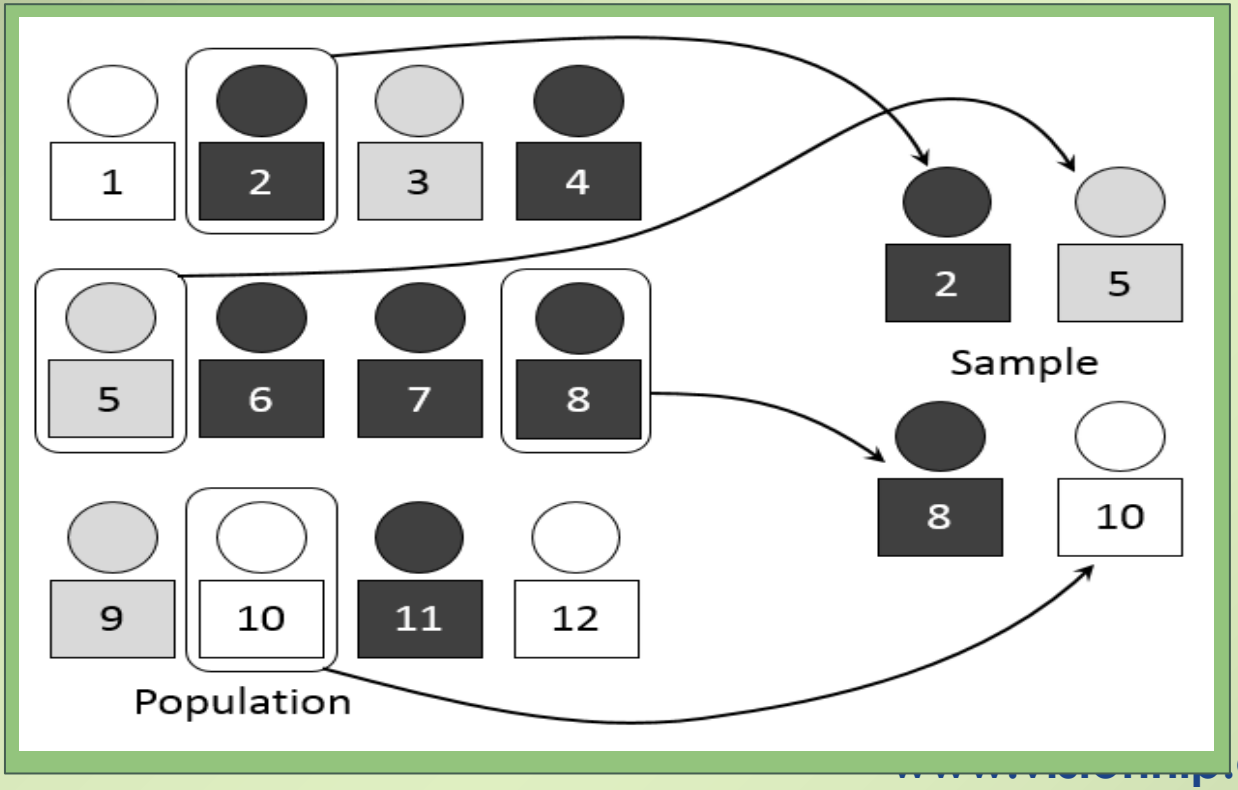
- Debt about census v/s sampling

What is Sampling:-

Sampling is the part of our day to day life which we use directly or indirectly.

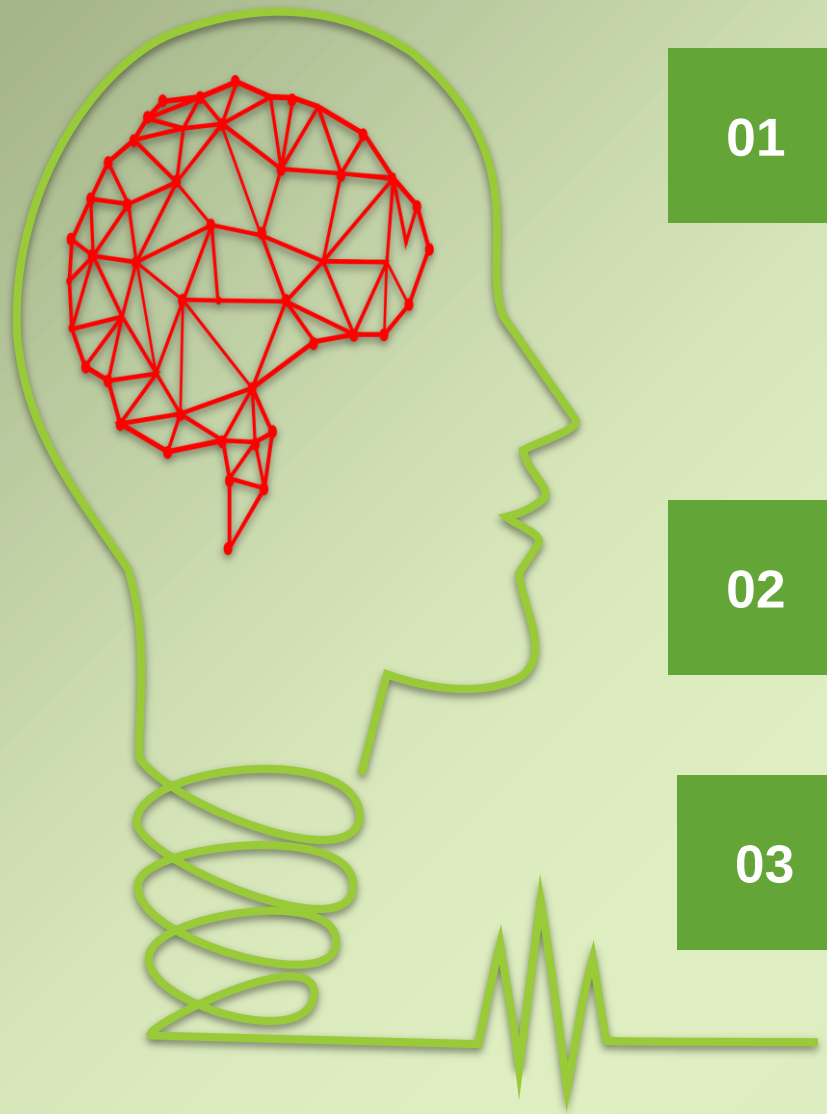
Examples

1. Inquiry about your trip plans with travel agency
2. Blood test
3. Soil testing



Basic Terminology in Sampling

VisionNLP



01

Population

A population is the total collection of elements about which we wish to make some inferences.

The number of elements belonging to the population known as Population size & its denoted by N.

02

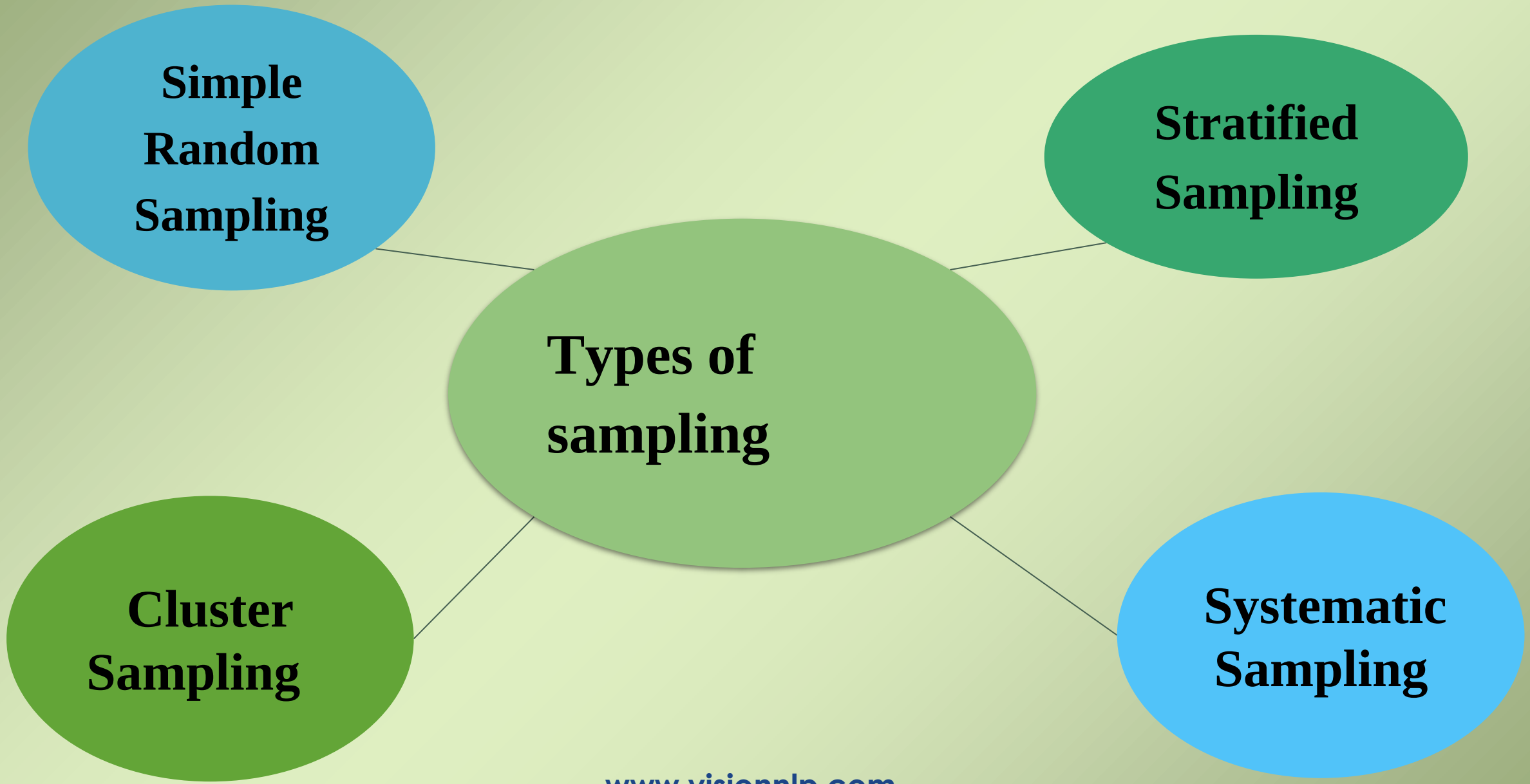
Sample:- Any part or unit selected from population.

03

Sampling:- The Process of drawing the Sample from the population called as Sampling.

Types of Sampling

VisionNLP



Simple Random Sampling

VisionNLP

As the name suggests it is a completely random method of selecting the sample. This is the easiest and most commonly used method of sampling. In this method each unit of the population has equal chance of being selected in the sample.