Statistics:

Statistics is a science which deals with methodologies to gather, review, analyze, visualize and draw conclusions from data.

Basically there are 2 types of statistics;

- **Descriptive Statistics:**
  - Analyze, visualize, summarize the data
  - Histogram, Bar Plot, Pie Chart, Probability distribution Function, etc.
  - Measure of central tendency (mean, median, mode)
  - Measure of dispersion (variance, standard deviation)
- **Inferential Statistics:**
  - Infer properties of the under distribution of the data
  - Inferential Statistics infers properties of population of data (sample)
  - Eg. from Indian election, what do you infer ? avg height of male and female population in indian → from a sample we map and infer
  - Z-test, T-test, Anova Test, F-test, Chisquare test.

Sampling Methods:

1. Simple Random Sampling → picking sample at random
2. Stratified Sampling → picking data that is a representation of all the stratas of the population. Stratified meaning layer wise;
   - Population → [10-20,20-40,40-60,60+] years of age → the sample (Eg.8) should contain people from all the stratas;
     i.e. 10-20 - 2, 20-40 - 2, 40-60 - 2, 60+ - 2 → total 8.
3. Systematic Sampling → picking data systematically from every 5th row (or) every 10th entry should be considered in the sample. (picking at periodic intervals)
4. Convenience Sampling
5. Cluster Sampling

Types of Data:

- Categorical Data
- Numerical Data
  - Discrete Data [Eg. no. of ppl in a family 2, 3, 8, etc.]
  - Continuous Data [Eg. age, height → 68.9 kgs, 181.2 cms, etc.]

Note: PINCODE is categorical data

- Nominal Data : No order [Eg. color, gender]
- Ordinal Data :
  In ordinal data rank is very important [Eg. ranks of students, degree of education]