Day01 Statistics :
it is a branch mathematics where we can collect, organize , analyze and represent the data for decision making.

stats are 2 types

1.Descriptive stats:- is a summery that describes about data

2. Inferential stats: This is process of data analytics where we can make the conclusion about the data.

Q)what is data?
Data is a raw form of your information that can be measured.

Q)What is population?
population is nothing but overall data

Q)what is a sample?
sample is a small part of a population.

Types of sampling techniques :

1.simple random sampling : is a process of sampling where every member of population has equal chances to select.

2.Stratified sampling : is the process where population splits into non overlapping groups.

3. Systematic sampling: is probability sampling method where researchers select members from population at every Nth interval.

4.Convinence Sampling : is the process taking the sample data from those who has more knowledge on research data

Variable : is a container where we can store data and reuse it.

There 2 type of data:

Quantitative Variables : measured numerically → add, subtract,mul, it takes the numerical values
ex:- number of students in a class room

1. Discrete Variable: whole number
ex: number of childrens in a famiit → 1,2,3..etc
Total population in Hyderabad → 1cr
sqft of a house

2. Continuous Variable: A numeric value that has infinite number of values

ex:- Height of students in a class room
rent prices of houses in a area

#1. marks of students in a class room → continuous
#2. No of tigers in the zoo → descrete
#3. No of seats in a car → descrete

Qualitative Variable: it is a categorical data based on some characterstics we can derive some numerical values.

ex:- eye colour , breed of dog, level-of education, marital status...etc

ph.D → > 1
PG → 2
Bach →>3
secondary → 4
primary - 5

Variable measurment sclaes:

1. **Nominal data : categorical data**
2. **Ordinal data : order of the data matters**
3. **interval : order is matters and value also but 0 is not present → 60 - 70 f**
4. **ration : 2:3,4:5**

---X---

→ Descriptive Stats : is summery that describes or summerizes about the data

→ Q) what is the most common general value should we consider from data?
Ans) descriptive stats → Central limit theorem or central tendency theorem or central measure → 3 imp types of measures → Mean, median, Mode

a = [1,1,2,2,3,3,4,5,5,6,100] → outliers is extreme range of values
mean → 1+1+2+2+3+3+4+5+5+6+100 / 11 →12

mode → repeated elements → [1,2,3,5]

median → middle value before checking the median value ensure that your data is in ascending order →3

#when we need to use the mean, median and mode?
→ when there is outliers in the data we don't use the mean → mode or median
ex:- [dog, cat, dog, dog, dog] → mode (isused for categorical values)

2.measure of despersion → describes the spread of data and variation from centre value

1.range (maximum - minimum) → [7,4,3,2,3,5] → 5
2.Interquartile range(IQR) → it is defined as 75% percentile of data - 25%percentile of your data.
3.standard deviations →