Class 1
Machine Learning With Python

WHAT IS STATISTICS.?

Statistics can be defined as a branch of mathematics dealing with the collection , analyzing, interpretation and presentation of the data.

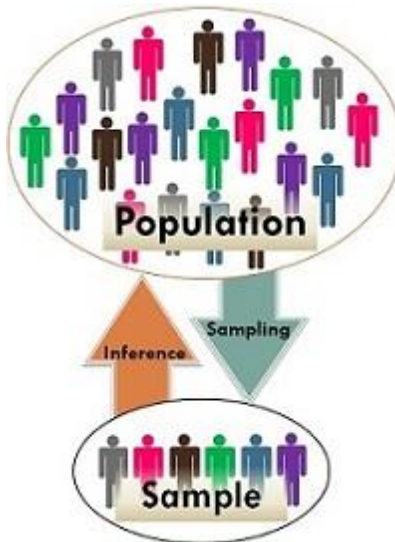**Another definition of statistics :-**
Study of sample attributes is called statistics .

Population :-

the data that is collected overall from a studying experiments.

Sample :-

A representative subset of population is called as sample.

for example :-
Suppose we have a Cheese Cake and we cut a slice from it, so Cheese Cake is **Population** and slice will be representative **Sample.**

SAMPLING:-

Sampling is used to draw a sample from population

for example:-
Suppose we cut the Cake vertically that will be useful and if we cut the cake horizontally then that is useless, so **technique** used to draw samples from population is called sampling.

SYMBOLS FOR REPRESENTATION

Population		Sample	
Total Population	**(N)**	Total Sample	**(n)**
Mean	(μ) mu	Mean	$(\bar{x}$ "x-bar")
Standard Deviation	σ "sigma"	Standard Deviation	s
Capital + Greek		Low Case	

TYPES OF STATISTICS

1.Descriptive

2.Infrential

Descriptive:- This Statistics is covered with describing and graphing the data **(Generally done on population data).**

Inferential:- This Statistics is used to draw Inferences/Conclusions from the data **(Generally done on Sample data).**

Types Of Statistics

- **Descriptive**
  - **Measures of Central Tendency**
    - **Mean**
      - **Harmonic**
      - **Geometric**
    - **Median**
    - **Mode**
  - **Measures of dispersion/Spread/Variance**
    - **Variance**
    - **Standard Daviation**
    - **Range**
    - **Frac Files**
      - **Quartiles**
      - **Decile**
      - **Percentile**
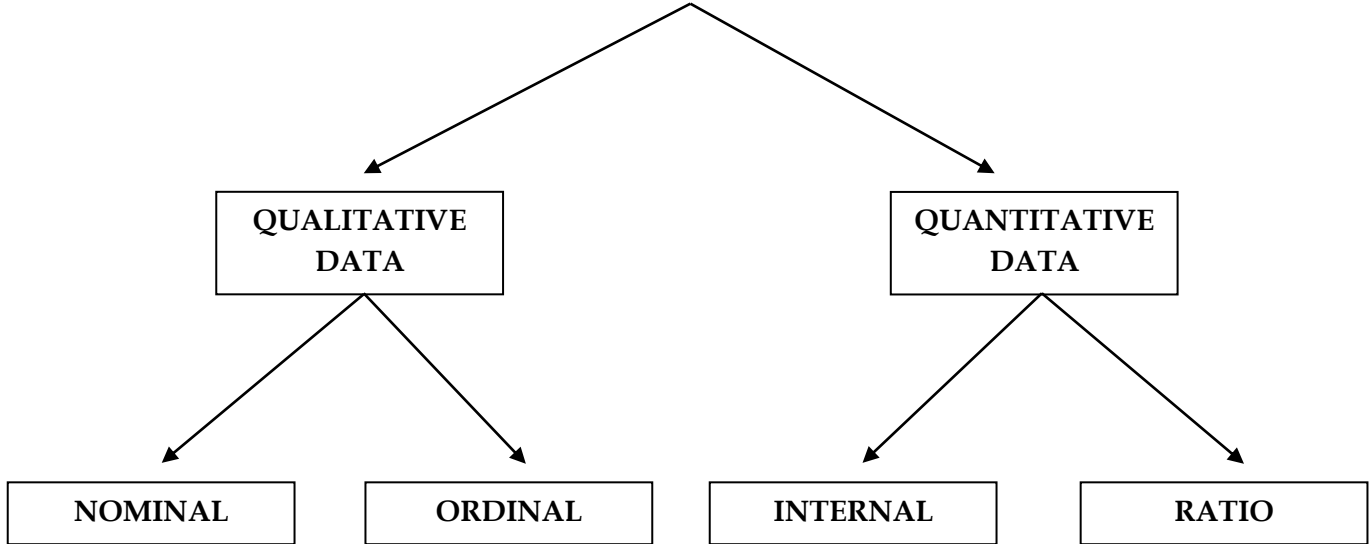   - **Z-score**
- **Inferential**
  - **Hypothesis Testing**
  - **Probabilities**
  - **Regression**
  - **Correlation, etc.**

COLLECTION OF THE DATA

| Primary Data Collection. | Secondary Data Collection. |
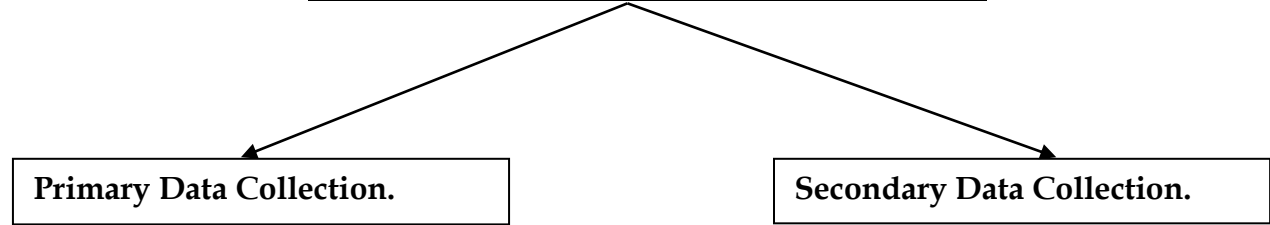
S.No.	Primary Data Collection	Secondary Data Collection
1.	Data collected from Source is called primary	Data collected/ Inferred/ process from primary is called as secondary
2.	for example :- Survey, questioners, online poles, Interviews, feedbacks, etc	Databases, Apps data, Social Media, etc
3.	Raw	Processed
4.	Expensive (as it includes more man power)	Cheap
5.	Collection with the aim of study and experiments	That is not specific, statistician needs to check which variables are useful, **That is why predictive modeling is done on Secondary data**.

TYPES OF DATA

```
                        TYPES OF DATA
                              |
            +-----------------+-----------------+
            |                                   |
    +---------------+                   +----------------+
    | QUALITATIVE   |                   | QUANTITATIVE   |
    |    DATA       |                   |     DATA       |
    +---------------+                   +----------------+
        |     |                             |      |
     +------+ +--------+              +----------+ +--------+
     |NOMINAL| |ORDINAL|              | INTERNAL | | RATIO  |
     +------+ +--------+              +----------+ +--------+
```

Note:- Qualitative Data is also Known as **Categorical Data,** and Quantitative Data is also known as **Numerical Data.**

Ques:-Why should we know the type of data.?

Ans:- If you don't know that your data falls on which class then you cannot be able to recognize which statistical measure to be used.

Different data types would revise different analysis, that is why it is important to know the data types.