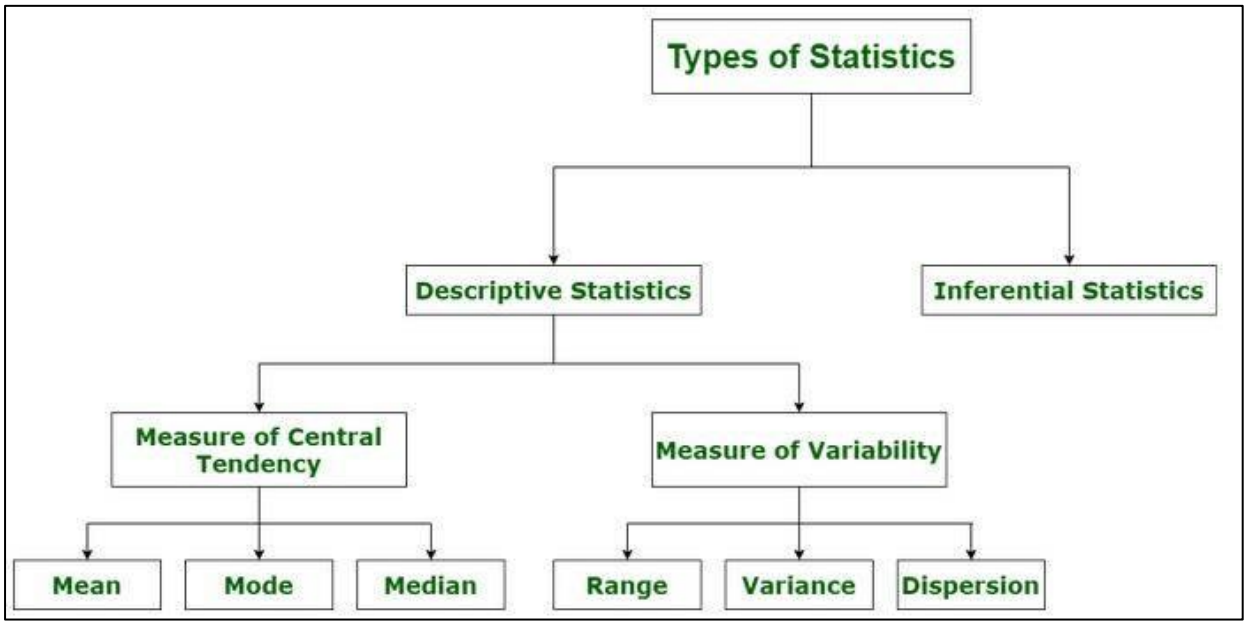
Statistics

Statistics is the science of analysing data.

Types of data in statistics:

- Descriptive Statistics
- Inferential Statistics



1. Descriptive Statistics

Descriptive statistics summarizes important features of a data set such as:

- Count
- Sum
- Standard Deviation
- Percentile
- Average
- Etc.. It is a good starting point to become familiar with the data.

We can use the describe() function in Python to summarize the data

```
import pandas as pd

full_data = pd.read_csv("Friends.csv", header=0, sep=",")

pd.set_option('display.max_columns',None)
pd.set_option('display.max_rows',None)

print (full_data.describe())
```

```
       Unnamed: 0      marks
count    4.000000   4.000000
mean     1.500000  58.000000
std      1.290994  27.580186
min      0.000000  37.000000
25%      0.750000  41.500000
50%      1.500000  48.500000
75%      2.250000  65.000000
max      3.000000  98.000000
```

Inferential Statistics

This branch of statistics takes a random data sample from a portion of the population to make predictions, draw conclusions based on that information, and generalize the results to represent the data on-hand.

Four different methodologies or types:

- Parameter Estimation
- Confidence Intervals
- Regression Analysis
- Hypothesis Test

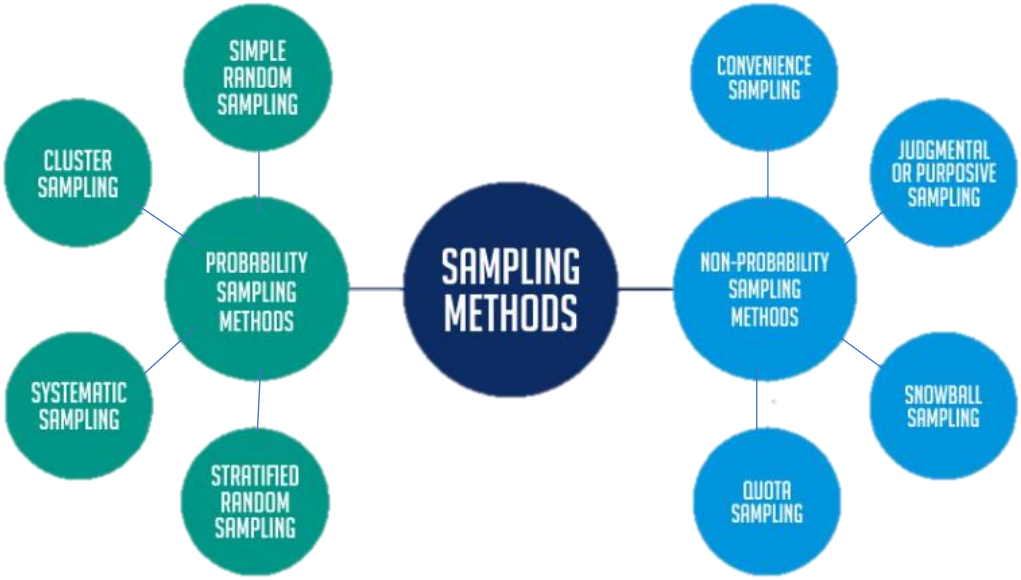
Population (N)

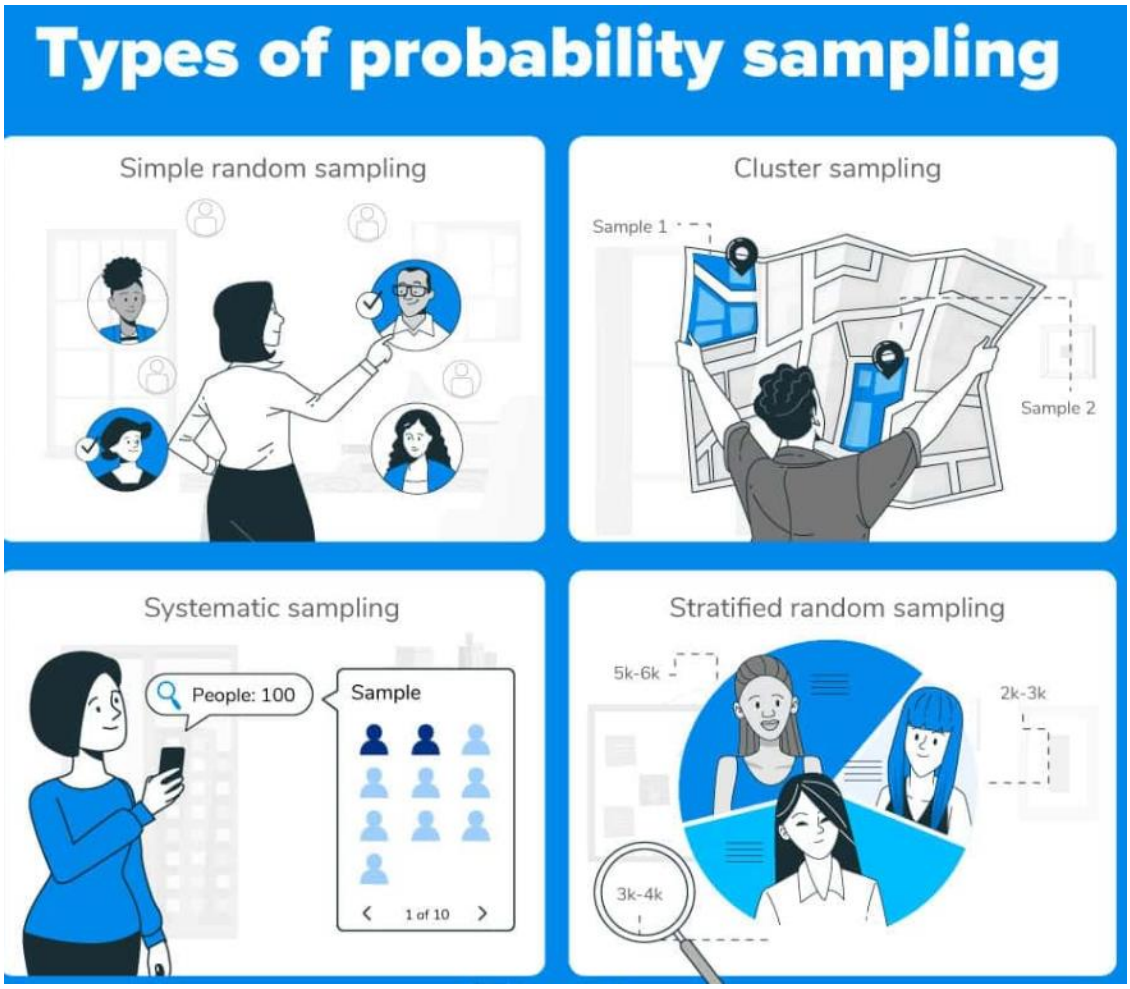
- The population includes all members of a specified group.
- Collecting data from an entire population can be time-consuming, expensive, and sometimes impractical or impossible.
- Includes all residents in the city

Sample (n)

- A sample is a subset of the population.
- Samples offer a more feasible approach to studying populations, allowing researchers to draw conclusions based on smaller, manageable datasets.
- Consists of 1000 households, a subset of the entire population

Sampling Techniques





	Probability Sampling Methods	Non-Probability Sampling Methods
Definition	Probability Sampling is a sampling technique in which samples from a larger population are chosen using a method based on the theory of probability.	Non-probability sampling is a sampling technique in which the researcher selects samples based on the researcher's subjective judgment rather than random selection.
Alternatively Known as	Random sampling method.	Non-random sampling method
Population selection	The population is selected randomly.	The population is selected arbitrarily.
Nature	The research is conclusive.	The research is exploratory.
Sample	Since there is a method for deciding the sample, the population demographics are conclusively represented.	Since the sampling method is arbitrary, the population demographics representation is almost always skewed.
Time Taken	Takes longer to conduct since the research design defines the selection parameters before the market research study begins.	This type of sampling method is quick since neither the sample nor the selection criteria of the sample are undefined.
Results	This type of sampling is entirely unbiased; hence, the results are also conclusive.	This type of sampling is entirely biased, and hence the results are biased, too, rendering the research speculative.
Hypothesis	In probability sampling, there is an underlying hypothesis before the study begins, and this method aims to prove the hypothesis.	In non-probability sampling, the hypothesis is derived after conducting the research study.