# Data is nothing but information.

```
-> Numerical Data

-> image Data

-> Video Data

-> Text Data

-> Audio Data

-> Speech
```

## Data -> information -> organisation info

Types of data

```
-> written format => stored in books => reusuability of data is hard

-> digital format => stored in servers => handling the data is going to be easy
```

## A. Structured Data:

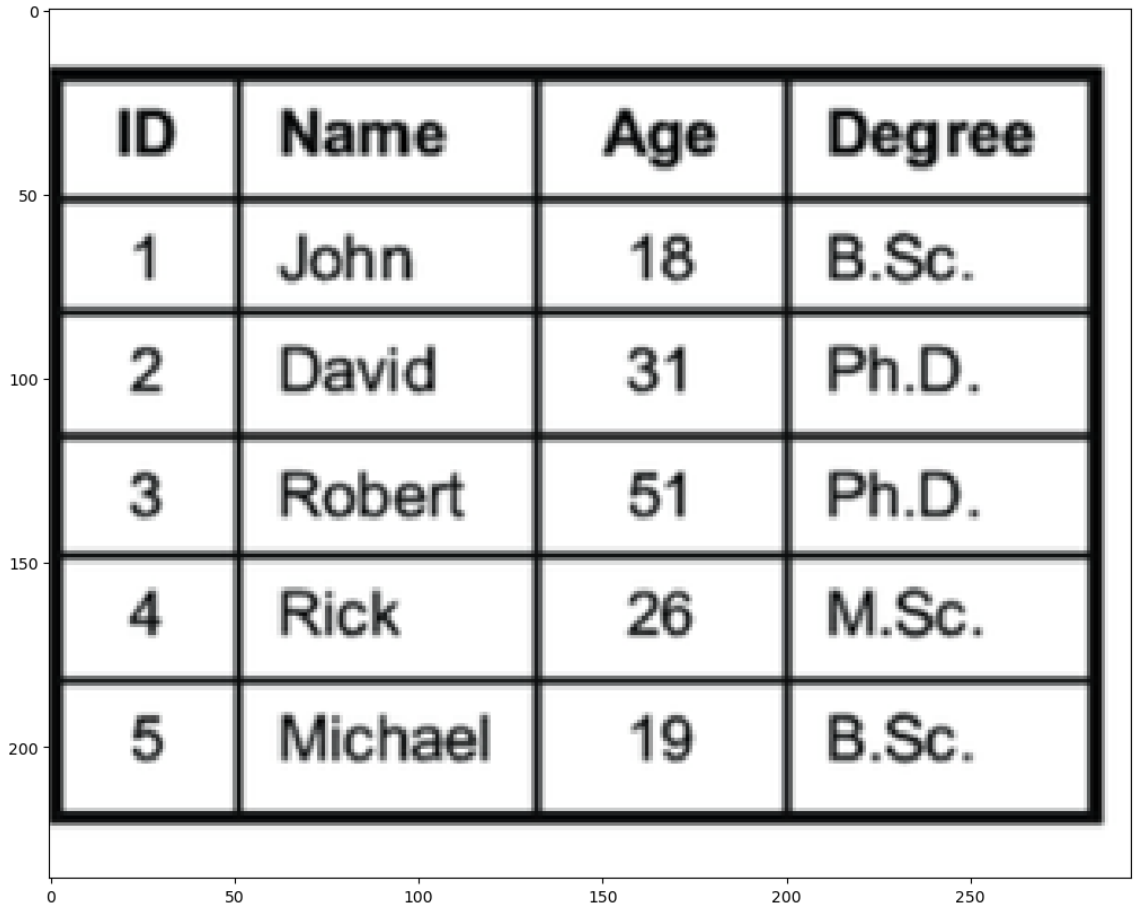
This Type Of Data Is Either Number Or Words. This Can Take Numerical Values But Mathematical Operations Cannot Be Performed On It. This Type Of Data Is Expressed In Tabular Format.

**E.G.) Sunny=1, Cloudy=2, Windy=3 Or Binary Form Data Like 0 Or1, Good Or Bad, Etc.**

In [1]:

```
import matplotlib.image as mpimg
import matplotlib.pyplot as plt
# Read Images
img = mpimg.imread(r"structure.png")
# Output Images
plt.figure(figsize=(20,10))
plt.imshow(img)

plt.show()
```



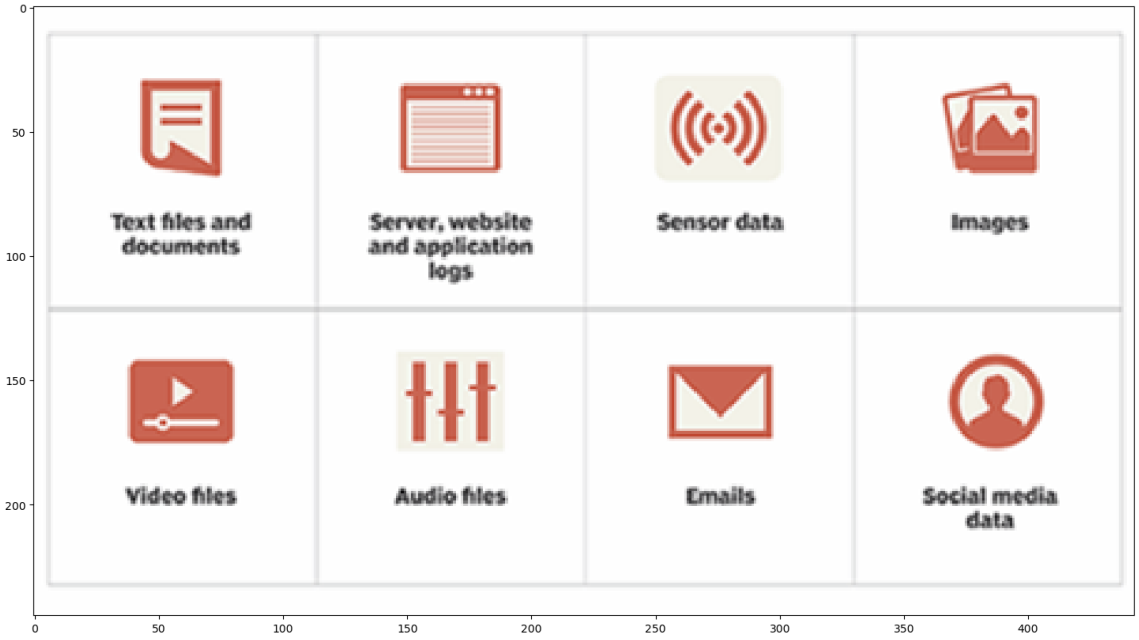

## B. Unstructured Data:

This Type Of Data Does Not Have The Proper Format And Therefore Known As Unstructured Data.This Comprises Textual Data, Sounds, Images, Videos, Etc.

In [2]:

```
import matplotlib.image as mpimg
import matplotlib.pyplot as plt
# Read Images
img = mpimg.imread(r"Unstructured.png")
# Output Images
plt.figure(figsize=(20,10))
plt.imshow(img)

plt.show()
```




# Population and Sample

Population

-> Population means whole data

=> Corana Vaccine => we are giving to all people

Sample

-> A small amount of data from the population

=> We are testing vaccine on samples

In [3]:

```
import matplotlib.image as mpimg
import matplotlib.pyplot as plt
# Read Images
img = mpimg.imread(r"ps.png")
# Output Images
plt.figure(figsize=(20,10))
plt.imshow(img)

plt.show()
```

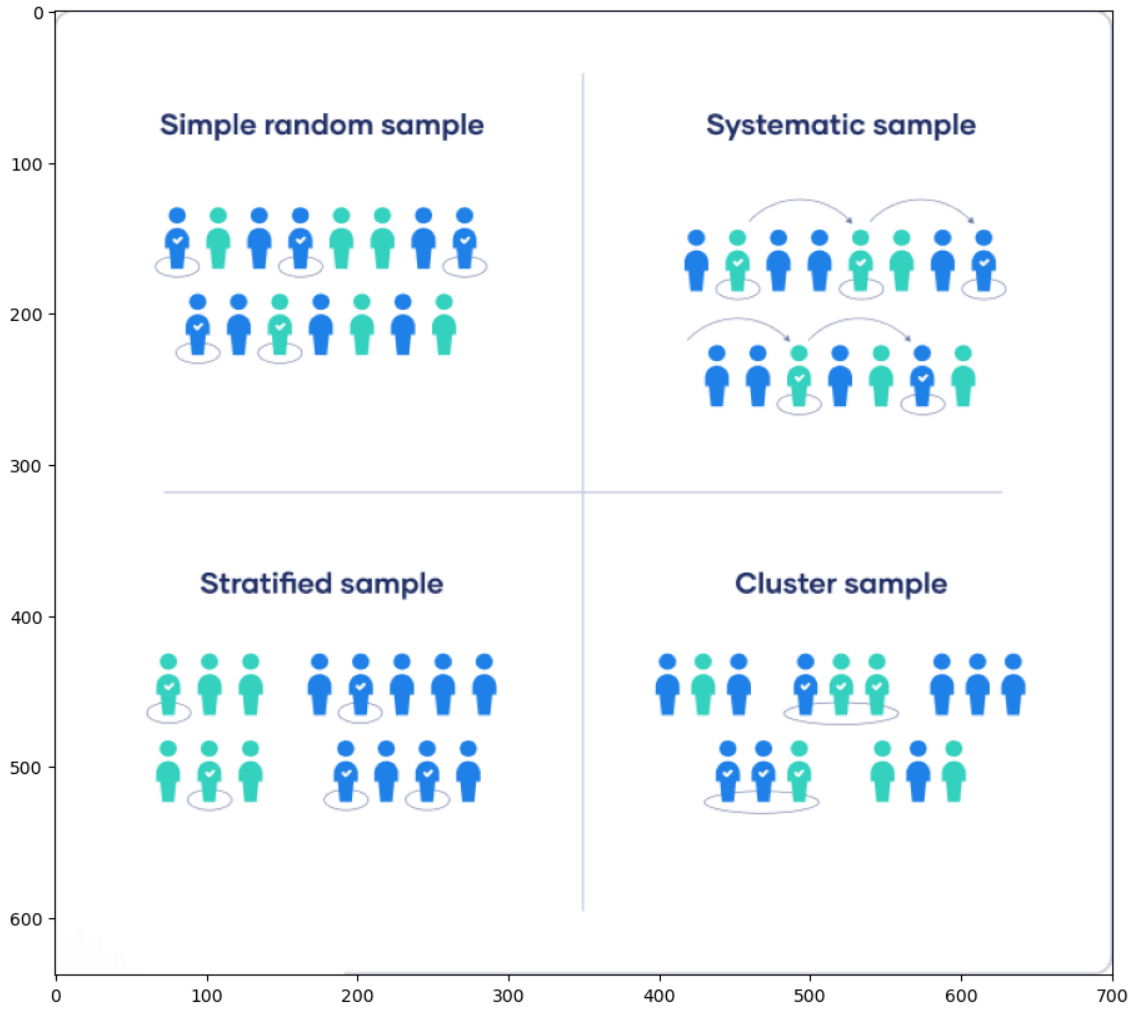

## Probability sampling methods

Probability sampling means that every member of the population has a chance of being selected. It is mainly used in quantitative research. If you want to produce results that are representative of the whole population, probability sampling techniques are the most valid choice.

In [4]:

```
img = mpimg.imread(r"probabilitysampling.png")
# Output Images
plt.figure(figsize=(20,10))
plt.imshow(img)

plt.show()
```




## 1. Simple random sampling

In a simple random sample, every member of the population has an equal chance of being selected. Your sampling frame should include the whole population.

To conduct this type of sampling, you can use tools like random number generators or other techniques that are based entirely on chance.

#### Example: Simple random sampling

You want to select a simple random sample of 100 employees of Company X. You assign a number to every employee in the company database from 1 to 1000, and use a random number generator to select 100 numbers.

## 2. Systematic sampling

Systematic sampling is similar to simple random sampling, but it is usually slightly easier to conduct. Every member of the population is listed with a number, but instead of randomly generating numbers, individuals are chosen at regular intervals.

**Example: Systematic sampling**

All employees of the company are listed in alphabetical order. From the first 10 numbers, you randomly select a starting point: number 6. From number 6 onwards, every 10th person on the list is selected (6, 16, 26, 36, and so on), and you end up with a sample of 100 people. If you use this technique, it is important to make sure that there is no hidden pattern in the list that might skew the sample. For example, if the HR database groups employees by team, and team members are listed in order of seniority, there is a risk that your interval might skip over people in junior roles, resulting in a sample that is skewed towards senior employees.

## 3. Stratified sampling

Stratified sampling involves dividing the population into subpopulations that may differ in important ways. It allows you draw more precise conclusions by ensuring that every subgroup is properly represented in the sample.

To use this sampling method, you divide the population into subgroups (called strata) based on the relevant characteristic (e.g. gender, age range, income bracket, job role).

Based on the overall proportions of the population, you calculate how many people should be sampled from each subgroup. Then you use random or systematic sampling to select a sample from each subgroup.

**Example: Stratified sampling**

The company has 800 female employees and 200 male employees. You want to ensure that the sample reflects the gender balance of the company, so you sort the population into two strata based on gender. Then you use random sampling on each group, selecting 80 women and 20 men, which gives you a representative sample of 100 people.

## 4. Cluster sampling

Cluster sampling also involves dividing the population into subgroups, but each subgroup should have similar characteristics to the whole sample. Instead of sampling individuals from each subgroup, you randomly select entire subgroups.

If it is practically possible, you might include every individual from each sampled cluster. If the clusters themselves are large, you can also sample individuals from within each cluster using one of the techniques above. This is called multistage sampling.

This method is good for dealing with large and dispersed populations, but there is more risk of error in the sample, as there could be substantial differences between clusters. It’s difficult to guarantee that the sampled clusters are really representative of the whole population.

**Example: Cluster sampling**

## Non-probability sampling methods

In a non-probability sample, individuals are selected based on non-random criteria, and not every individual has a chance of being included.

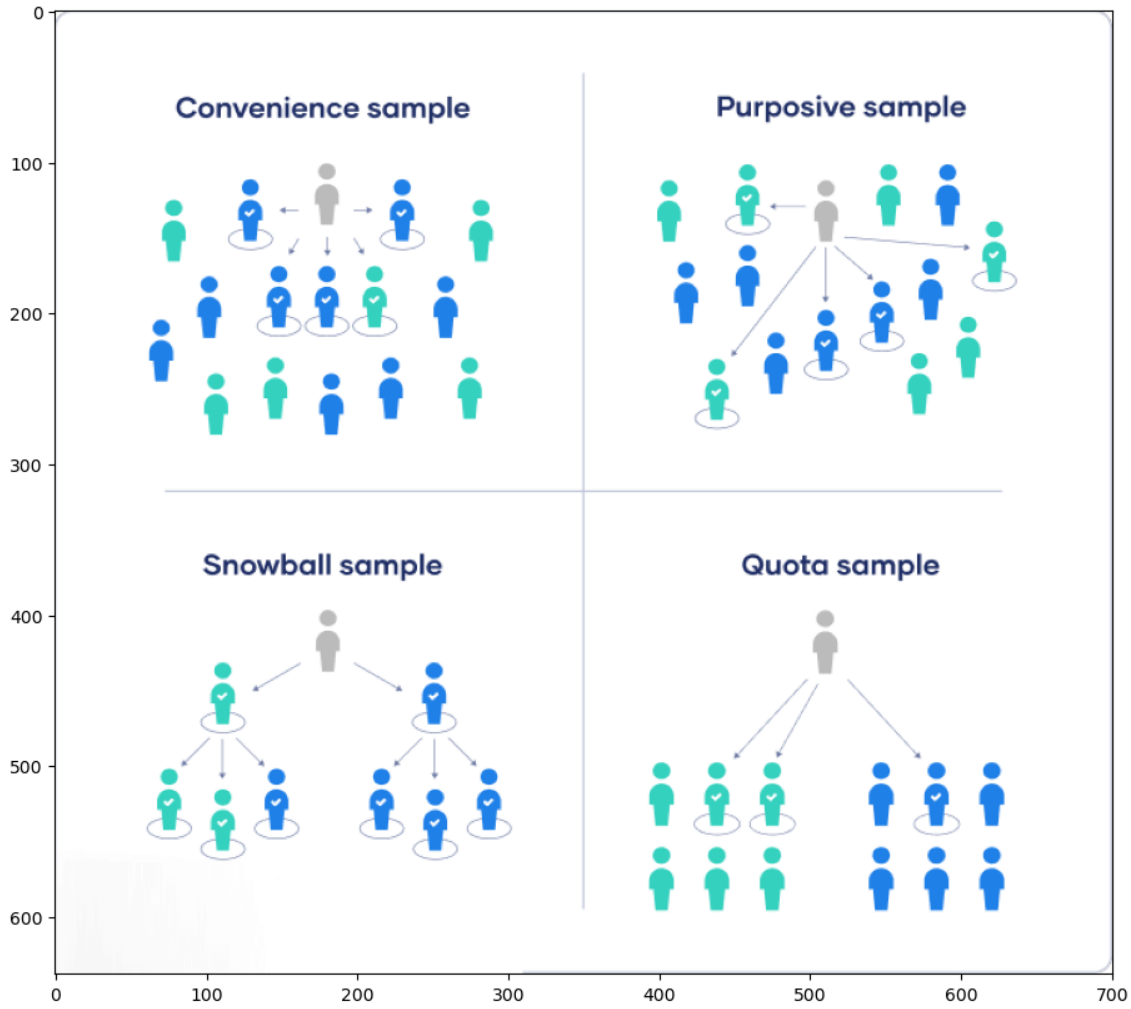
This type of sample is easier and cheaper to access, but it has a higher risk of sampling bias. That means the inferences you can make about the population are weaker than with probability samples, and your conclusions may be more limited. If you use a non-probability sample, you should still aim to make it as representative of the population as possible.

Non-probability sampling techniques are often used in exploratory and qualitative research. In these types of research, the aim is not to test a hypothesis about a broad population, but to develop an initial understanding of a small or under-researched population.

In [5]:

```
img = mpimg.imread(r"nonprobabilitysampling.png")
# Output Images
plt.figure(figsize=(20,10))
plt.imshow(img)

plt.show()
```




## 1. Convenience sampling

A convenience sample simply includes the individuals who happen to be most accessible to the researcher.

This is an easy and inexpensive way to gather initial data, but there is no way to tell if the sample is representative of the population, so it can't produce generalizable results.

**Example: Convenience sampling**

You are researching opinions about student support services in your university, so after each of your classes, you ask your fellow students to complete a survey on the topic. This is a convenient way to gather data, but as you only surveyed students taking the same classes as you at the same level, the sample is not representative of all the students at your university.

## 2. Voluntary response sampling

Similar to a convenience sample, a voluntary response sample is mainly based on ease of access. Instead of the researcher choosing participants and directly contacting them, people volunteer themselves (e.g. by responding to a public online survey).

Voluntary response samples are always at least somewhat biased, as some people will inherently be more likely to volunteer than others.

**Example: Voluntary response sampling**

You send out the survey to all students at your university and a lot of students decide to complete it. This can certainly give you some insight into the topic, but the people who responded are more likely to be those who have strong opinions about the student support services, so you can't be sure that their opinions are representative of all students.

## 3. Purposive sampling

This type of sampling, also known as judgement sampling, involves the researcher using their expertise to select a sample that is most useful to the purposes of the research.

It is often used in qualitative research, where the researcher wants to gain detailed knowledge about a specific phenomenon rather than make statistical inferences, or where the population is very small and specific. An effective purposive sample must have clear criteria and rationale for inclusion. Always make sure to describe your inclusion and exclusion criteria.

**Example: Purposive sampling**

You want to know more about the opinions and experiences of disabled students at your university, so you purposefully select a number of students with different support needs in order to gather a varied range of data on their experiences with student services.

## 4. Snowball sampling

If the population is hard to access, snowball sampling can be used to recruit participants via other participants. The number of people you have access to "snowballs" as you get in contact with more people.

**Example: Snowball sampling**

You are researching experiences of homelessness in your city. Since there is no list of all homeless people in the city, probability sampling isn't possible. You meet one person who agrees to participate in the research, and she puts you in contact with other homeless people that she knows in the area.

# Data Types based on Statistics

-> Quantitative Data

    => Data Consists of numbers

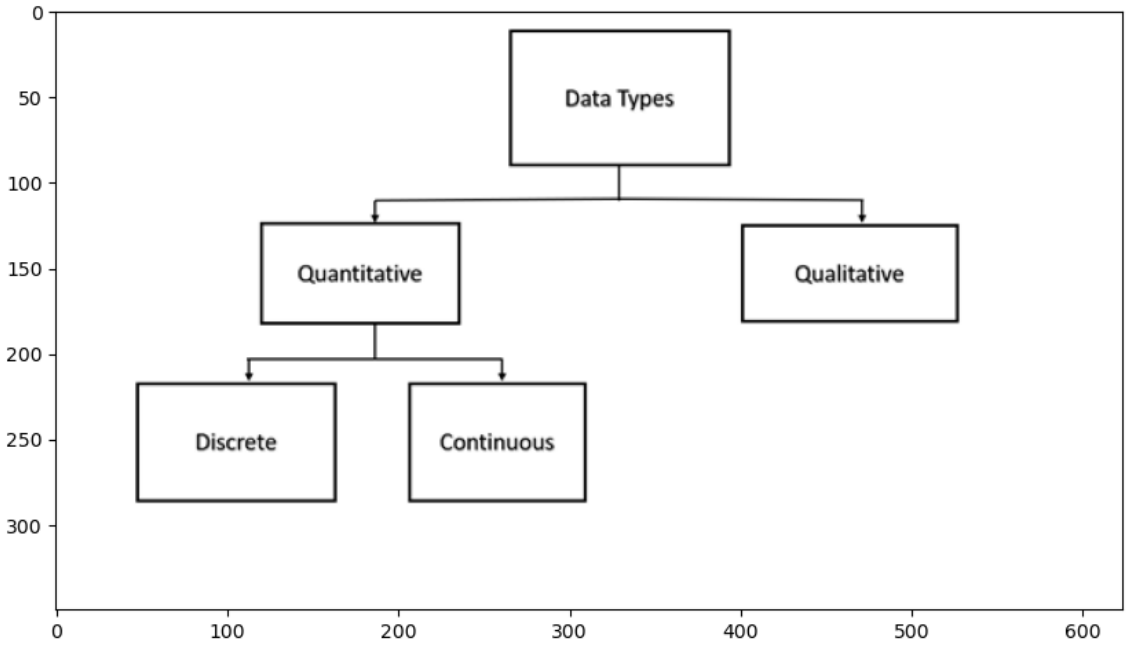
-> Qualitative Data

    => Data consists of Chars

In [6]:

```
img = mpimg.imread(r"dt.png")
# Output Images
plt.figure(figsize=(10,10))
plt.imshow(img)

plt.show()
```



## Quantitative Data

```
-> Discrete Data

    => Data is countable (Categories)

        -> Number of girls and boys information

-> Continuous Data

    => Data is measurable

    => Continous Data going to be two types

        -> Interval

            -> Values are going to have both negative and postive

                -> Temperature

        -> Ratio

            -> Starts at 0 and positive numbers

                -> weight
```

## 1. Quantitative Data Type: –

This Type Of Data Type Consists Of Numerical Values. Anything Which Is Measured By Numbers.

**E.G., Profit, Quantity Sold, Height, Weight, Temperature, Etc.**

This Is Again Of Two Types

## A.) Discrete Data Type: –

The Numeric Data Which Have Discrete Values Or Whole Numbers. This Type Of Variable Value If Expressed In Decimal Format Will Have No Proper Meaning. Their Values Can Be Counted.

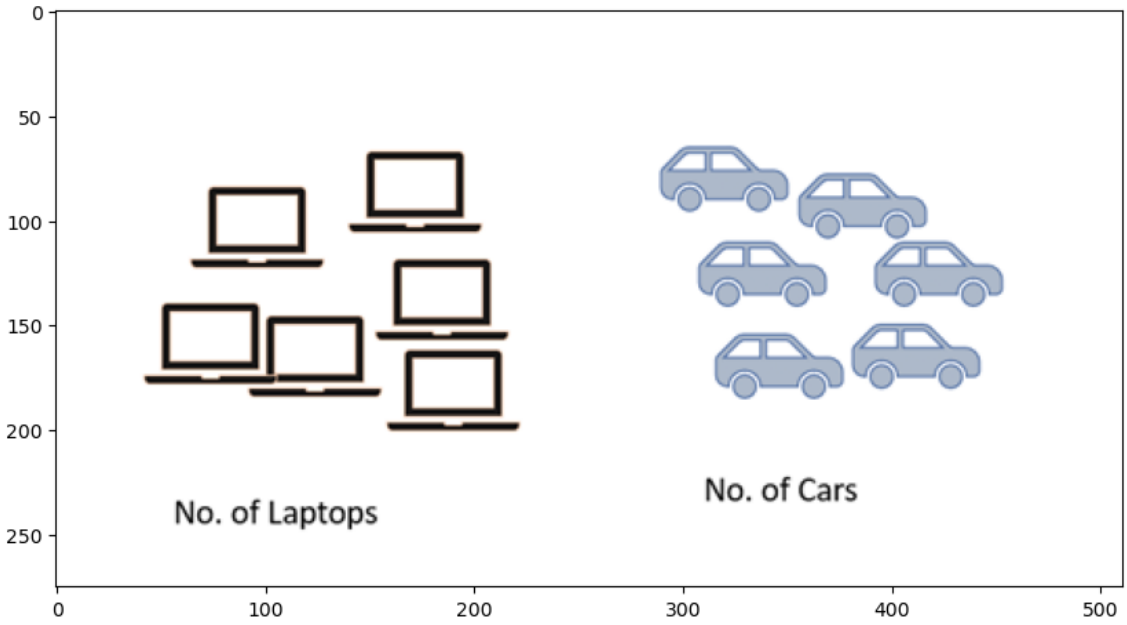
**E.G.: –**

No. Of Cars You Have, No. Of Marbles In Containers, Students In A Class, Etc.

In [7]:

```
img = mpimg.imread(r"ddt.png")
# Output Images
plt.figure(figsize=(10,10))
plt.imshow(img)

plt.show()
```




## B.) Continuous Data Type: –

The Numerical Measures Which Can Take The Value Within A Certain Range. This Type Of Variable Value If Expressed In Decimal Format Has True Meaning. Their Values Can Not Be Counted But Measured. The Value Can Be Infinite

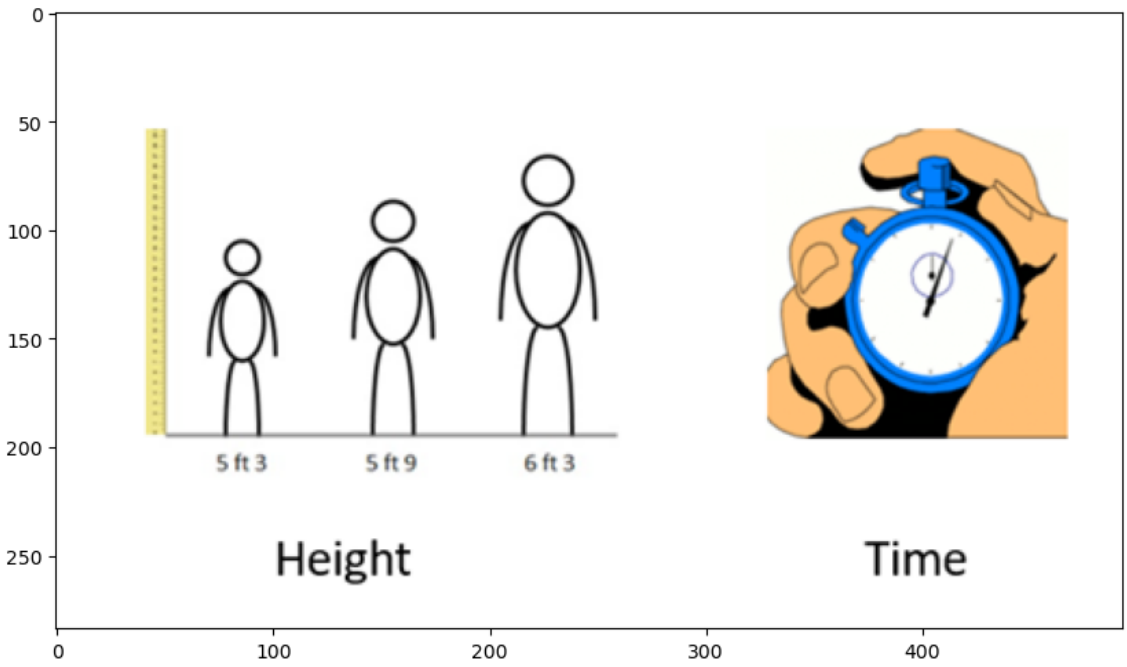
**E.G.: –**

Height, Weight, Time, Area, Distance, Measurement Of Rainfall, Etc.

In [8]:

```
img = mpimg.imread(r"cdt.png")
# Output Images
plt.figure(figsize=(10,10))
plt.imshow(img)

plt.show()
```




## I. Interval Data Type:

This Is Numeric Data Which Has Proper Order And The Exact Zero Means The True Absence Of A Value Attached. Here Zero Means Not A Complete Absence But Has Some Value. This Is The Local Scale.

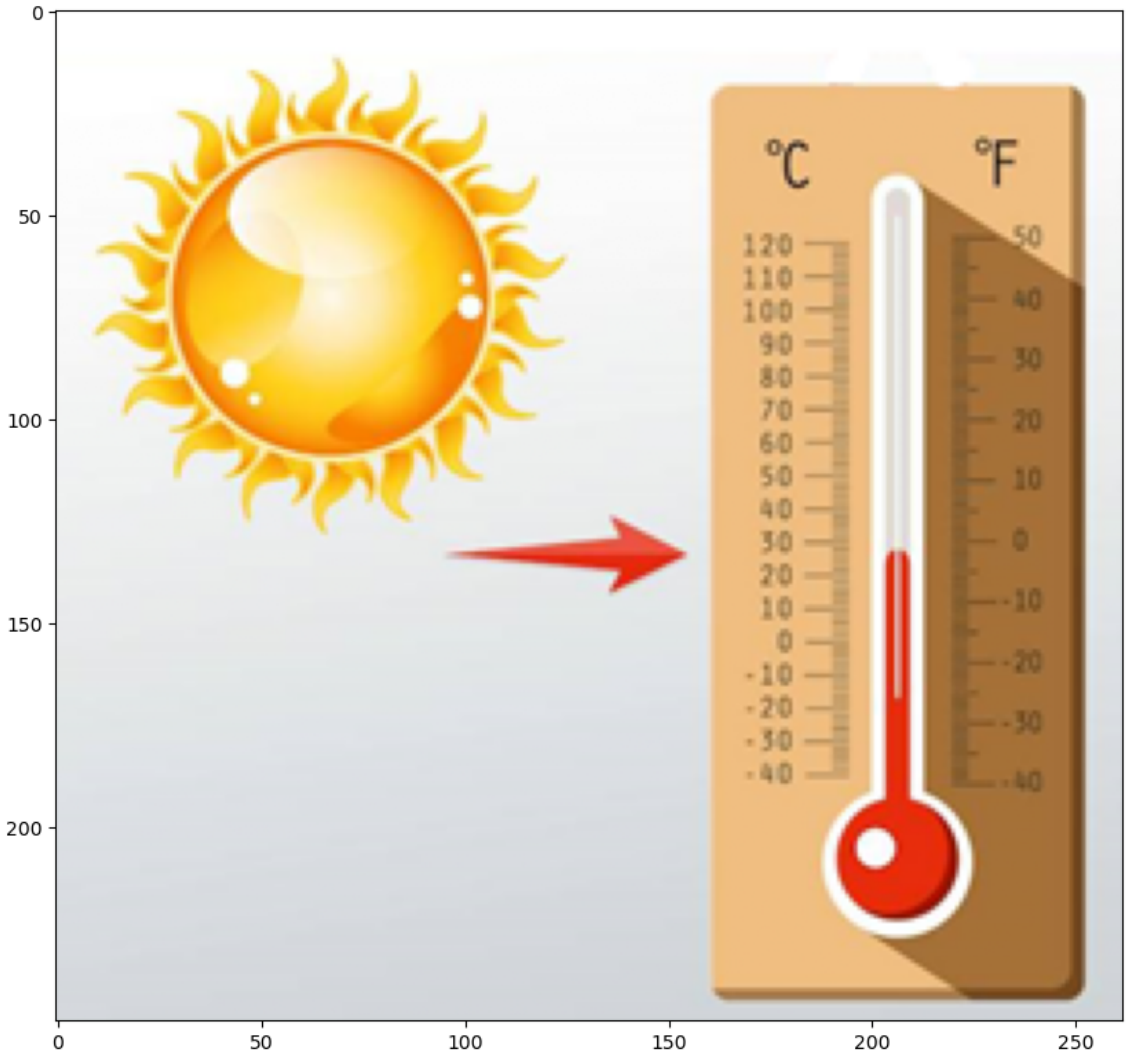
**E.G.,**

Temperature Measured In Degree Celsius, Time, Sat Score, Credit Score, PH, Etc. Difference Between Values Is Familiar. In This Case, There Is No Absolute Zero. Absolute

In [9]:

```
img = mpimg.imread(r"idt.png")
# Output Images
plt.figure(figsize=(10,10))
plt.imshow(img)

plt.show()
```




## IV. Ratio Data Type:

This Quantitative Data Type Is The Same As The Interval Data Type But Has The Absolute Zero. Here Zero Means Complete Absence And The Scale Starts From Zero. This Is The Global Scale.

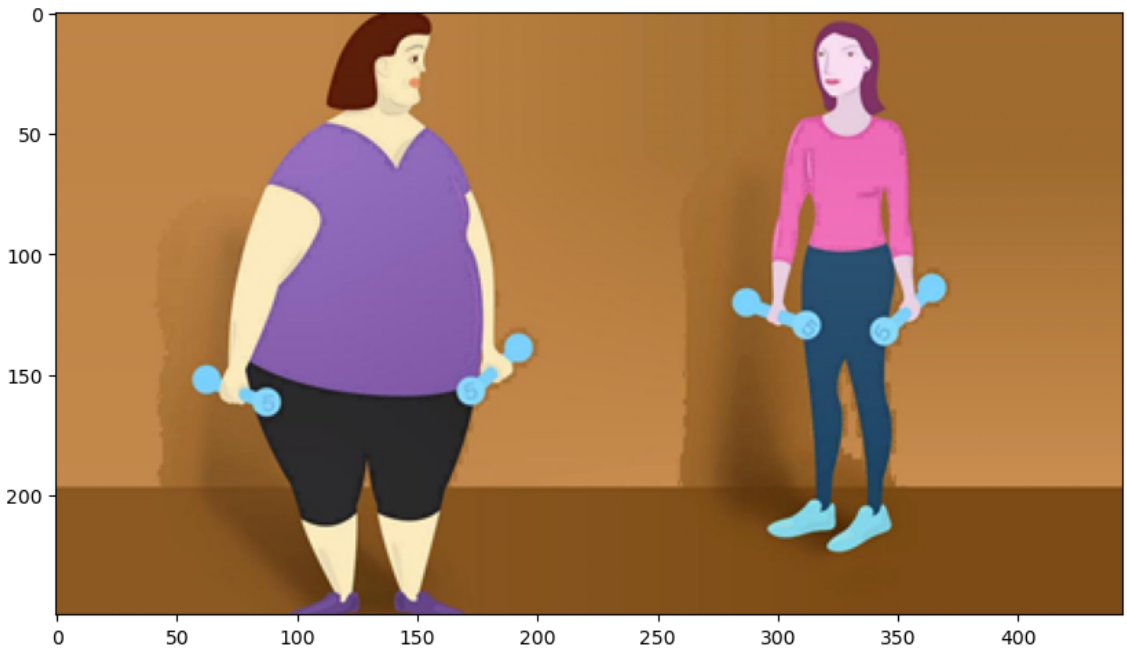
**E.G.,**

Temperature In Kelvin, Height, Weight, Etc.

In [10]:

```
img = mpimg.imread(r"rdt.png")
# Output Images
plt.figure(figsize=(10,10))
plt.imshow(img)

plt.show()
```




## Qualitative

```
-> Nominal (Names)

    -> Nominal data consists of nouns and we call that as unorder data

        => red, green,black, blue, yellow, white

-> Ordinal (Order)

    -> Ordinal data consists of action(Verbs and Adjectives). Data is in the fo
rm of order

        => Example: low, medium, high
```

## I. Nominal Data Type:

This Is In Use To Express Names Or Labels Which Are Not Order Or Measurable.

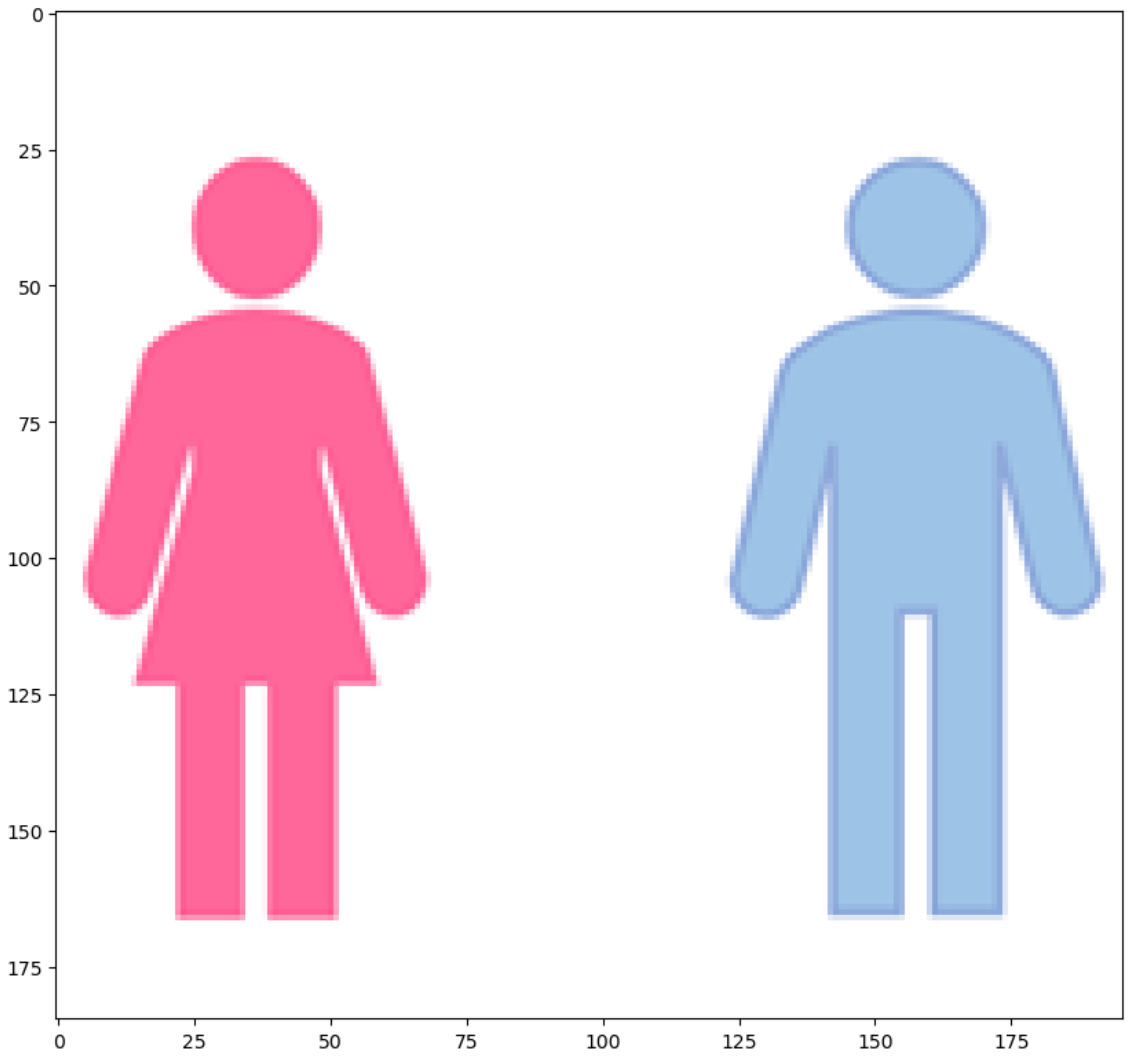
**E.G.,**

Male Or Female (Gender), Race, Country, Etc.

In [11]:

```
img = mpimg.imread(r"ndt.png")
# Output Images
plt.figure(figsize=(10,10))
plt.imshow(img)

plt.show()
```




## II. Ordinal Data Type:

This Is Also A Categorical Data Type Like Nominal Data But Has Some Natural Ordering Associated With It.

**E.G.,**

Likert Rating Scale, Shirt Sizes, Ranks, Grades, Etc.

In [12]:

```
img = mpimg.imread(r"odt.png")
# Output Images
plt.figure(figsize=(10,10))
plt.imshow(img)

plt.show()
```


## Programming and Stats Data Types

int, float .... => Quantitative Data

str, object ... => Qualitative Data

In [ ]:

```

```