Introduction of Data

BY: Rupesh Jawade

1.What is data ?

- Everything which is present around us ,and that thing arrange in meaning full manner is called “Data”.
- Data is a information or collection of facts (number ,words, measurement and observation , images,) that has been translate into a form for that computer can process.

2.What is information ?

- When data are processed, organized, structured or presented in a given context so as to make them useful, they are called “information”.
- Information is data that has been processed in such a way as to be meaningful to person who receives it .it is anything that is communicated

3.Differenciate between **DATA** And **INFORMATION**

DATA

- Data is unorganized and unrefined facts
- Data doesn’t depend on information.
- Data is an individual unit that contains raw materials which do not carry any specific meaning.
- Raw data alone is insufficient for decision making.
- An example of data is a student’s test score

INFORMATION

- Information is a group of data that collectively carries a logical
- Information comprises processed, organized data presented in a meaningful context meaning.
- Information depends on data.
- Information is sufficient for decision making.
- The average score of a class is the information derived from the given data.

4.How data is useful for us ?

- Data allows organizations to measure the effectiveness of a given strategy.
- They can be used to measure/record a wide range of business activities - both internal and external.
- Describing ,Diagnosing. Predicting. Prescribing. These are the four main uses of data.

- **Describing** is the use of data to set out **what has happened** – e.g. a website received 1,000 visitors in the first week of May this year
- **Diagnosing** is the use of data to explain **why something happened** – e.g. 1,000 visitors came to the website on that week because an email was sent to 10,000 customers
- **Predicting** is the use of data to define **what will happen** – e.g. if we send an email to 20,000 customers we will get 2,000 visitors to the website
- **Prescribing** is the use of data to define **what will be done** – e.g. we will send a weekly email

5.Differenciate Between Structured, semi-structured, unstructured data.

Structured Data

- Structured data is highly-organized and formatted so that it's easily searchable in relational databases.
- Structured data is quantitative data that consists of numbers and values.
- Structured data is stored in tabular formats like excel sheets or SQL databases.
- It is based on a relational database.
- Analysis It is easy to search.

Unstructured Data.

- Unstructured data has no predefined format or organization, making it much more difficult to collect, process, and analyze.
- Unstructured data is qualitative data that consists of audio, video, sensors, descriptions, and more.
- It is based on character and binary data.
- Analysis Searching for unstructured data is more difficult.

Semi- Structured data

- Semi structured data is a combination of structured and unstructured data because it contains elements of both.
- It is based on XML/RDF(Resource Description Framework).
- There are several types of semi-structured data.
- **JSON** files
- **HTML** web page.
- **CSV** files.

6.What is Big Data?

- The term Big data is used in the data definition to describe the data is in the petabyte range or higher.
- Big data refers to **data that is so large, fast or complex that it's difficult or impossible to process using traditional methods**. The act of accessing and storing large amounts of information for analytics has been around for a long time

7.What are the **Quantitative data & Qualitative data**

Quantitative data

- The data deal with number and things you can measure objectively is call quantitative data.
- Like :
  such as height ,width and length. Temperature and humidity, prices,area , volume

Qualitative data

- The data deal with characteristics and descriptors that can not be easily measured, but can be observed subjectively.
- Such as smells , tastes ,textures, attractiveness and colors

8.What are the different ‘v’s Big data

There are 5Vs in Big Data.

- 1.Variety
- 2.Volume.
- 3.value.
- 4.Veracity.
- 5.Velocity.

9.Name some popular tools that use in Big data?

1. Apache Spark
2. Apache Hadoop
3. Apache Flink
4. Google Cloud Platform
5. MongoDB
6. Sisense
7. RapidMiner

10 What are the different types of data? Explain?

- There are two types of data
- 1. Quantitative 2. Qualitative

- **1.Quantitative Data:**

The data deal with number and things that you can measure objectively such as duration or speed , is call quantitative data

There are two types in Quantitative data

1. Discrete :whole number that can't be broken down in smaller parts such as a number of item.

2. Continuous : Number that can be broken down in smaller parts such as height or weight.

Two types of continuous data

a) Interval : Number with known differences Between Variables Such as time

b) Ration: Number that have measurable intervals where difference can be determined such as height or weight.

2.Qualitative Data

- The data deal with characteristics and descriptors that can not be easily measured, but can be observed subjectively.
- Non- numerical data that is categorical such as yes/no responses or eye colour

- **There are two types**

a) Nominal : The data which is used for naming variables such as hair color

b) Ordinal : The data which is used to describe the order of value such as

1=Happy 2=neutral 3 = unhappy