

ASSIGNMENT 1

1. Differentiate between data and information

Ans:

DATA	INFORMATION
Data is a collection of individual statistics, facts or items of information	Information is a data that is processed, organized, and structured
In raw form and unprocessed and unstructured	Processed and structured
Might be meaningless on its own	Always meaningful
Relies on data	Does not rely on information
Comes in the form like numbers, figures, and statistics	Comes as words, thoughts, and ideas
May be difficult to understand	Relatively easy to understand

2. How is data useful to us?

Ans:

- For informed decision making
- For problem solving
- For greater understanding
- For improving processes
- For understanding behaviour

3. What is big data?

Ans:

Big data is a collection of data that is huge in volume and growing with time. It can be structured, unstructured, and semi structured that are being collected from different sources. Big data refers to a process that is used when

traditional data mining and handling technique cannot uncover the insights and meaning of the underlying data.

4. Differentiate between structured, semi structured, and unstructured data?

Ans:

Structured data –

Structured data is data whose elements are addressable for effective analysis. It has been organized into a formatted repository that is typically a database. It concerns all data which can be stored in database SQL in a table with rows and columns. They have relational keys and can easily be mapped into pre-designed fields. Today, those data are most processed in the development and simplest way to manage information. Example: Relational data.

Semi-Structured data –

Semi-structured data is information that does not reside in a relational database but that has some organizational properties that make it easier to analyse. With some processes, you can store them in the relation database (it could be very hard for some kind of semi-structured data), but Semi-structured exist to ease space. Example: XML data.

Unstructured data –

Unstructured data is a data which is not organized in a predefined manner or does not have a predefined data model thus it is not a good fit for a mainstream relational database. So, for Unstructured data, there are alternative platforms for storing and managing, it is increasingly prevalent in IT systems and is used by organizations in a variety of business intelligence and analytics applications. Example: Word, PDF, Text, Media logs.

5. What are quantitative and qualitative data?

Ans:

- Quantitative data refers numerical or measurable data.
Example: weight, distance, length.

- Qualitative data studies imply on personal accounts or documents that illustrate in detail how people think or respond with the society.

Example: what language you speak, your opinion on something.

6.What are different V's in big data?

Ans:

- Velocity
- Volume
- Variety
- Values
- Veracity
- Variability

7.Name some popular tools used in big data?

Ans:

- Apache Spark
- Apache Hadoop
- Apache Flink
- Google cloud platform
- MongoDB
- Sisense
- RapidMiner

8.What are different types of data ? Explain

Ans:

- Qualitative data or categorical data

Qualitative data, also known as the categorical data, describes the data that fits into the categories. Qualitative data are not numerical. The categorical information involves categorical variables that describe the features such as a person's gender, home town etc. Categorical measures are defined in terms of natural language specifications, but not in terms of numbers.

Normal data

the variables without providing the numerical value. Nominal data is also called the nominal scale. It cannot be ordered and measured. But sometimes, the data can be Nominal data is one of the types of qualitative information which helps to label qualitative and quantitative. Examples of nominal data are letters, symbols, words, gender etc.

- Ordinal Data

Ordinal data/variable is a type of data that follows a natural order. The significant feature of the nominal data is that the difference between the data values is not determined. This variable is mostly found in surveys, finance, economics, questionnaires, and so on.

The ordinal data is commonly represented using a bar chart. These data are investigated and interpreted through many visualisation tools. The information may be expressed using tables in which each row in the table shows the distinct category.

- Quantitative or Numerical Data

Quantitative data is also known as numerical data which represents the numerical value (i.e., how much, how often, how many). Numerical data gives information about the quantities of a specific thing. Some examples of numerical data are height, length, size, weight, and so on. The quantitative data can be classified into two different types based on the [data sets](#). The two different classifications of numerical data are discrete data and continuous data.

- Discrete Data

Discrete data can take only discrete values. Discrete information contains only a finite number of possible values. Those values cannot be subdivided meaningfully. Here, things can be counted in whole numbers.

Example: Number of students in the class

- Continuous Data

Continuous data is data that can be calculated. It has an infinite number of probable values that can be selected within a given specific range.

Example: Temperature range

