

Assignment 1

1. Differentiate between data and information

Ans:

Data	Information
Raw form and unstructured	Processed and structured
Data is independent of information	Information is dependent on data
May or may not be meaningful	Always meaningful
Difficult to understand	Easy to understand
It may not be in order	It should be in order
Data is an individual unit	Information is a group of data

2. How data is useful to us

- Helps in identifying relationships exist between 2 variables
- Helps in predicting the future based on previous trend of data
- Helps in detecting fraud
- Helps in determining pattern that exist between data

3. What is big data?

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 techniques cannot uncover the insights and meaning of the underlying data. The main aim is to solve new problems or old problems in a better way.

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

Structured data:

It is well defined structure and follows a consistent order. It is easily accessible and usable by people. It is stored in well defined schemas like database. It tends to be tabular and has rows and columns defines its attribute in a very clear manner.

Unstructured data:

Data which is not structured is called unstructured data. It has an internal structure but is not structured via predefined data models or schema. It may or may not be textual. And can also be stored with a non relational database like Nosql.

Semi structured data:

This type of data has the elements of both structured and unstructured data. One good example of semi-structured data is JSON.

5. What are quantitative data and qualitative data?

Quantitative studies refers numerical or measurable data. In contrast, qualitative studies imply on personal accounts or documents that illustrate in detail how people think or respond within society. The major difference between the qualitative and quantitative data is that quantitative data is about the numbers and the qualitative data is descriptive.

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

Big data is best described with the six - volume, variety, velocity, value, veracity and variability.

7. Name some popular tools used in big data

Apache Spark, Apache Hadoop, Apache Flink, Google Cloud Platform, MongoDB, Sisense, RapidMiner.

8. What are different types of data? Explain

Basically we have 2 type of data- Qualitative data and Quantitative data.

Qualitative data, also known as the categorical data. The categorical information involves categorical variables that describe the features such as a person's gender, home town etc.

Again this type can be divide into nominal and ordinal. Nominal data is one of the types of qualitative information which helps to label the variables without providing the numerical value. It cannot be ordered and measured. Whereas Ordinal data is a type of data that follows a natural order. The ordinal data is commonly represented using a bar chart. These data are investigated and interpreted through many visualisation tools.

Quantitative data is also known as numerical data which represents the numerical value. 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 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. And 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.