

## Short Introduction of Unit

This chapter explores various aspects of data management and its impact on our understanding of information. We begin by examining different types of data and effective ways to collect and store it. You will learn methods for organizing and analyzing data, using both quantitative and qualitative techniques.

### Q.1 Explain Data with suitable examples.

09509001

**Ans.** Data consists of raw facts collected about things around us that we can process to generate useful information. It can take many forms, such as numbers, words, measurements, observations, or even images and sounds, and may originate from various sources.

#### Example

- Weather Data:** Information on temperature, humidity, and wind speed collected from weather stations.
- Sales Data:** Records of product sales, quantities, and prices at a store. For example, selling 50 units of a product at 100 Rupees each in a day.
- Survey Responses:** Answers from a survey about customer satisfaction. For Example, 80% of customers rating their experience as "satisfactory" or better.
- Website Data:** Information on website visitors, such as number of visitors per day and the pages they visit. For example, 500 visitors viewing the homepage on a Monday.
- Social Media Data:** Likes, comments, and shares on social media posts. For example, a post receiving 200 likes and 50 comments.

#### Did You Know?

The amount of digital data created and consumed globally is so vast that it's predicted to reach 175 Zettabytes by 2025.

### Q.2 Describe Qualitative Data type categorically with its types.

09509002

**Ans:** Data can be divided into two broad categories namely qualitative and quantitative.

**Qualitative Data:** Qualitative data refer to categories or labels used to describe the qualities or characteristics of something rather than its quantity. Key characteristics of qualitative data include non-numeric, descriptive, and categorical attributes.

- Non-Numeric:** Qualitative data is represented by words, labels, or symbols instead of numbers. It describes attributes rather than quantities. Examples include the names of students in a class (e.g. Ali, Badar, Qasim) and the colors of cars in a parking lot (e.g. red, blue, green), where "name" and "color" are attributes.
- Categorical:** Qualitative data can be into categories or classes based on their characteristics. Examples include types of fruit (e.g. apple, banana, orange), job titles (e.g., manager, engineer, accountant), and book genres (e.g., fiction, non-fiction, mystery).

#### Types of Qualitative data

Qualitative data is further classified into two types namely Nominal and Ordinal data.

- Nominal Data:** Nominal data is used to label or categorize items without implying any order. Each item is unique and separate. Examples include gender (male, female), types of fruits (apple, banana, orange), and colors (red, blue, green). Nominal data is helpful for grouping and categorizing information.

- **Ordinal Data:** Ordinal data represents categories with a meaningful order, though the differences between categories are not uniform. Examples include customer satisfaction ratings (satisfied, neutral, unsatisfied), education levels (high school, bachelor's, master's), and shirt sizes (small, medium, large, extra-large). Ordinal data helps in ranking items meaningfully.

### Q.3 What do you know about Quantitative Data?

09509003

**Ans.** Quantitative data consists of numbers used to measure the quantity or amount of something. These data types answer questions like "How much?" or "How long?" and can be useful for mathematical calculations and statistical analyses. Key characteristics of quantitative data include being numerical, measurable, countable, and arithmetical).

- **Numerical:** Quantitative data is expressed in numbers, representing a measurable quantity. Examples include heights in centimeters, weights in kilograms, and test scores.
- **Measurable:** Quantitative data can be measured using instruments or tools. Examples include using a ruler for length, a scale for weight, and a thermometer for temperature.
- **Countable:** Quantitative data can be counted or enumerated, particularly for discrete data. Examples include the number of students in a class or the number of cars in a parking lot.
- **Arithmetical:** Quantitative data can be used in arithmetic operations. For instance, multiplying the unit price of a fruit by its weight to find the total price, or calculating annual school fees by multiplying monthly fees by number of months.

### Types of Quantitative Data

Quantitative data can be further classified into two main types: discrete and continuous.

1. **Discrete Data:** Discrete data consist of distinct, separate values that are countable, often in whole numbers. Discrete data answer questions like "How many?" or "How often?"

For example, counting oranges in a basket might show there are 16, answering "How many?" similarly, tracking the number of oranges falling from a tree each week, with counts like 5 on Monday, 7 on Wednesday, and 4 on Friday, helps answer "How often?"

2. **Continuous Data:** Continuous data consists of values that can take any number within a given range, including fractions or decimals. Examples include student heights (150.5 cm, 160.2 cm), fruit weights (1.5 kg, 2.2 kg), and room temperatures (22.5°C, 23.7°C).

### Operations Performed on Continuous Data

All operations that can be performed on discrete data can also be performed on continuous data. In addition to logical, grouping and arithmetic operations division can also be performed on continuous data. For example, you can divide a 2.5 kg of meat (continuous data) among ten persons (yielding 0.25 kg each) but cannot divide 3 cars (discrete data) among five persons.

### Q.4 Make a Detailed Comparison on different data types with the help of table.

09509004

**Ans:**

Aspect	Nominal Data	Ordinal Data	Discrete Data	Continuous Data
Definition	Labels or categories without a specific order.	Categories with a meaningful order, but uneven differences.	Distinct, separate values that are countable.	Values that can take any number within a range.
Purpose	Categorizes items without implying	Ranks items in a	Counts specific	Measure the extent or duration

	rank.	meaningful order.	items or occurrences	of something.
<b>Representation</b>	Words, labels, symbols.	Words, labels, symbols with order.	Numbers, whole or discrete values.	Numbers, including fractions and decimals.
<b>Characteristics</b>	Non-numeric, categorical	Ordered, but differences between categories are not uniform.	Countable, often whole numbers	Measurable, can include fractions or decimals.
<b>Examples</b>	Gender (male, female); Types of fruits (apple, banana, orange); Colors (red, blue, green).	Satisfaction ratings (satisfied, neutral, unsatisfied); Education levels (high school, bachelor's, master's); Shirt sizes (small, medium, large).	Number of students in a class (30); Number of cars in a parking lot (15); Number of books on a shelf (50);	(150.5 cm, 160.2 cm); Weight of fruits (1.5 kg, 2.2 kg); Temperature (24.5° C, 23.7° C).
<b>Operations</b>	Equality, grouping, counting, mode.	Equality, comparison, ranking, median, frequency distribution,	Equality, grouping, counting, arithmetic operations without division	Equality, grouping, arithmetic operations including division
<b>Statistical Analysis</b>	Mode, frequency counts.	Median, mode, frequency distribution.	Mean, median, range, standard deviation.	Mean, median, range, standard deviation.
<b>Usage</b>	Categorizing, understanding preferences or attributes.	Understanding order or levels of a characteristic	Counting specific occurrences or items,	Measuring precise amounts, durations, or extents.

### Q. 5 Describe the importance of organizing data.

**Ans.** Organizing data systematically is very important for clear analysis and interpretation. When data is well-organised, it helps reduce errors.

#### Importance of Organising Data

Proper organisation saves time. Think about searching for a book in a messy room versus a neatly arranged bookshelf. Similarly, when data is organised, it's easier to find and analyse. For instance, if you have sales data arranged by date in a spreadsheet, you can quickly see how sales have changed over time without having to search through random numbers.

- Moreover, organised data improves clarity. When data is presented clearly, it's easier to understand and interpret. For example, a chart showing monthly sales figures is much easier to understand than a long list of numbers.
  - By organising data into tables, charts, and graphs, you can quickly grasp what the data is saying, making it simpler to draw conclusions and make decisions.
- To make data easy to understand, it can be organised into tables, charts, and graphs. Here's how:

**Data Tables:** Imagine you have data about students' scores in different subjects. An example of this data is shown in table.

Student	Math	Science	English
Sajid	85	78	90
Hammad	78	88	85
Rizwan	92	82	87
Sufiyan	90	80	89
Khuram	67	75	70

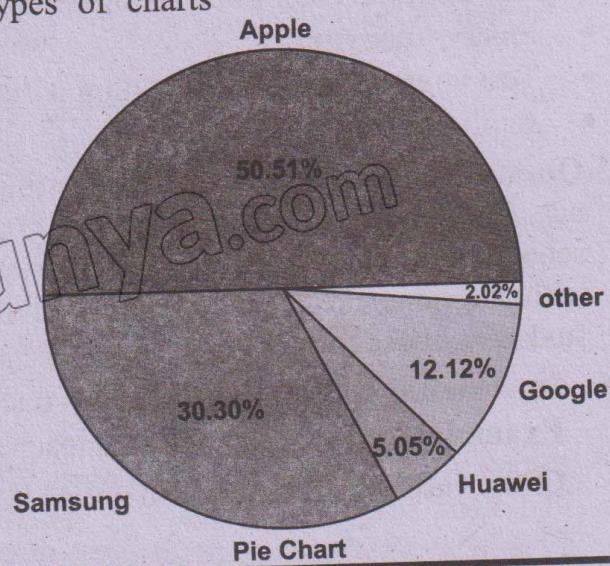
### Q.6 What do you know about Charts and Graphs?

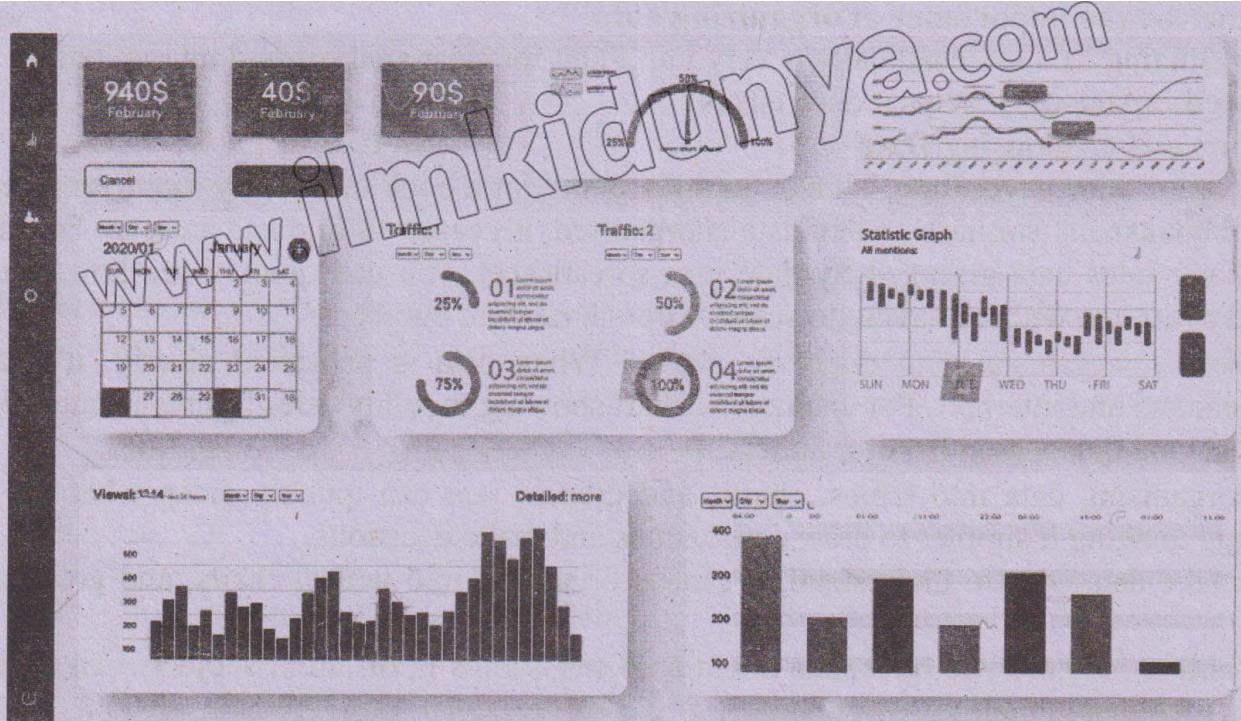
**Ans. Charts:** Charts are visual representation of data designed to make complex information easier to understand. Charts help identify patterns, trends and outliers in datasets. Common types of charts include:

- Bar charts
- Line charts
- Pie charts

**Graphs:** Graphs are visual tools used to represent data and show relationship between different data points. Common types of graphs include:

- Line graphs
- Bar graphs
- Scatter plots
- Histograms etc.





## Graphs

### Q.7 Describe the role of Surveys and Questionnaires in Collection of Data.

09509007

**Ans. Data collection** is the process of gathering information to answer questions, make decisions or understand something better. There are different methods and tools for collecting data, each with its own way of gathering and recording information.

#### Methods of Data Collection

Here's a simple explanation of the main methods:

**1. Surveys:** Surveys collect information from people by asking them questions. This can be done on paper, over the phone, or online. For example, to find out your classmates' favourite ice cream flavours, you might create a survey with questions like "What is your favourite ice cream flavour?" and give it to your classmates to fill out.

#### Best Practices in Survey Design and Administration

- Be clear and specific
- Keep it short and simple
- Use multiple choice and rating scales
- Ensure anonymity
- Test your survey before sending it out
- Analyse the results

**2. Questionnaires:** Questionnaires are similar to surveys but are often written forms that people fill out. They usually have a set of questions that need to be answered. For instance, if your school wants to know which activities students enjoy most, they might give out a questionnaire with questions like, "Which school activity do you enjoy the most? (eg, sports, art, music)", where students choose their answers from the options provided.

- **Interviews:** Talking to individuals one-on-one to gather detailed information.

**Example:** Interviewing a school teacher to understand their experience and challenges.

- **Observations:** Watching and noting what happens in a particular situation.

**Example:** Observing how students behave during a group project to understand how they work in a team.

**Online Data Sources:** These include websites, databases, and digital tools where you can find and collect information. For example, if you're researching the most popular pets, you might use the internet to find statistics or articles about pet ownership.

#### **Q. 8 How data can be collected from Online Sources? Explain in details.**

09509008

**Ans.** In today's world, we can find lots of information online to help with school projects. We can gather, extract and integrate this information to answer almost any question. For example, if you want to know how social media affects teenagers, you would collect information from different sources, extract useful data from it and integrate it into a coherent and comprehensible report. Here's how you can do it.

**Data Gathering:** Data gathering means collecting information from various places to help with your research or school projects.

**1. Online Databases:** These are special websites where you can find a lot of information, such as articles and reports. For example, you can use Google Scholar or IEEE Xplore databases. Search for keywords like "how social media affects teenagers" to find useful articles and studies.

**2. Online Resources:** These include educational websites, government portals, and news outlets. For example, Wikipedia, the National Institute of Health (USA), and BBC.com.

**Data Extraction:** Data extraction is about finding and saving the most relevant information from a larger set of data.

**1. Identify Relevant Data:** Look for specific details about how social media impacts teenagers. Search for keywords like "teenage social media usage statistics" and find sections in articles with important statistics or quotes.

**2. Copy and Save:** Highlight the key information you find. For example, if an article says "70% of teenagers feel stressed because of social media, copy this information and paste it into a document. Keep your data organised by topic or source.

**3. Use Tools:** Some websites have tools to help you gather and organise data. Look for options like "Export" or "Download" to get articles in formats like PDFs.

**Data Integration:** Data integration involves combining information from different sources to get a complete perspective.

**1. Organise Data:** Put all your saved information into one place, such as a document or spreadsheet. For example, create a Google Sheets file titled "Social Media Impact Research" where you list all key statistics and quotes.

**2. Compare and Merge:** Review data from different sources and find common themes. For instance, if many sources say that social media increases anxiety among teenagers, note this. Combine similar findings into one summary.

**3. Create Summaries:** Write a brief overview of what you found. Highlight the main points, such as how many teenagers report feeling anxious due to social media. Make charts or graphs to show trends, such as bar graph of anxiety levels among teens.

#### **Q. 9 Differentiate between Structured and Unstructured data.**

09509009

**Ans.**

##### **1. Structured Data:**

Structured data is organised and formatted to be easily searchable and analysable. Examples include data in spread sheets and traditional databases. Think of a spread sheet with rows and

**Did you know?**  
Every day, people generate over 2.5 quintillion bytes of data.

columns where each row is a record (such as student) and each column is an attribute (such as Student ID, Student Name, Class, Date of Birth, Fee Status, and Height) as shown in table.

Student ID	Student Name	Class	Date of Birth	Fee Status	Height
001	Ali Akbar	9 <sup>th</sup>	3/25/2009	Paid	4.7
002	Faheem Aslam	9 <sup>th</sup>	5/7/2008	Paid	4.9
003	Munir Ahmad	9 <sup>th</sup>	6/11/2009	Unpaid	5.2
004	Khalid Mahmood	9 <sup>th</sup>	9/13/2009	Paid	5.6
005	Kamran Malik	9 <sup>th</sup>	7/21/2009	Paid	5.3

## 2. Unstructured Data

Unstructured data is more free-form and doesn't fit into a specific format. Examples include text from emails, social media posts, videos, and images. This type of data is harder to organise but can be very valuable. Develop a sound sense of discipline, character, initiative and a solid academic background. You must devote yourself whole-heartedly to your studies, as that is your first obligation to yourselves, your parents and the State, you must learn to obey for only then you can learn to command.

### Q.10 Briefly discuss different data storage techniques.

09509010

**Ans.** When we talk about data storage methods, we're referring to the different ways we can save and organise information so that we can easily access and use it later. Here we look at four important data storage technologies: **Spreadsheets, Databases, Data Warehouses, and NoSQL**.

#### 1. Spreadsheets

Spreadsheets are tools that help us organise data in rows and columns, much like a simple version of a database. They are often used for tasks that don't require complex data management. Spreadsheets like Microsoft Excel or Google Sheets allow users to input data into cells organised in rows and columns.

#### 2. Databases

Databases are like digital filing cabinets where information is stored in an organised way. They are designed to handle large amounts of structured data and allow users to easily find and manipulate this data. Information in databases is stored in tables, similar to spreadsheets. Each table has rows and columns, where rows represent individual records, and columns represent the attributes of those records.

#### 3. Data Warehouses

Data Warehouses are specialized types of databases designed for storing and analysing large amounts of data collected from various sources. They are used to help organizations make decisions based on data analysis. Data warehouses are used by companies to analyze business performance, track sales, and understand customer behaviour. Prominent examples of data warehouse include **Amazon Redshift**, a data warehouse service offered by Amazon Web Services, and **Google Big-Query**, a fast and scalable data warehouse offered by Google Cloud.

#### 4. NoSQL

NoSQL stands for "Not Only SQL" and refers to a variety of database technologies that are designed to handle different types of data that may not fit well into traditional databases. NoSQL

databases are flexible and can store unstructured data, which is data that doesn't follow a specific format. **Example:** Examples of popular NoSQL databases include **MongoDB** which stores data in JSON-like documents and **Cassandra**, which is designed to handle large amounts of data across many servers without a single point of failure.

**Q.11 Explain the concept of data visualization. How does it help in understanding complex data? Provide examples of different types of visualizations and their applications in real-life scenarios.** 09509011

**Ans:** Data visualization is the process of turning numbers and information into pictures. These pictures make it easier for us to understand what the data is telling us. When we look at data in the form of charts or graphs, it becomes simpler to see patterns, trends, and relationships.

### **Visualizing Different Data Types**

Different types of data can be visualized in different ways. Here are some common types of data and how we can visualize them:

**Nominal Data:** Nominal data represents categories without any specific order. Bar charts and pie charts are great for showing nominal data.

**Ordinal Data:** Ordinal data represents categories with a specific order but without a uniform scale. Bar charts and stacked bar charts are effective for visualization for ordinal data.

**Discrete Data:** Discrete data consists of distinct, separate values. Visualization technique of histograms and dot plots are useful.

**Continuous Data:** Continuous data can take any value within a range. Line graphs, scatterplot and box plot are commonly used for continuous data.

**Q.12 Discuss the importance and benefits of data visualization. Why is it essential for businesses and decision-makers to use charts, graphs, and dashboards?** 09509012

**Ans:** Data visualization is important because it makes understanding data much easier. Imagine trying to understand a list of hundreds of numbers. It's tough, right? But if those numbers are displayed as a bar chart, it's much easier to see what's going on. Visualizations provide quick insights, allowing you to grasp what the data is saying without having to read through all the details. This is particularly useful for making better and faster decisions. For example, a business can quickly see which products are selling well and which are not. There are several techniques and tools for visualizing data including charts, graphs and dashboards. Some popular tools for visualizing data include Microsoft Excel, Google Sheets and Tableau.

**Microsoft Excel:** A spreadsheet tool that can create various charts and graphs.

**Google Sheets:** Similar to Excel, it allows you to create and share visualizations online.

**Tableau:** A powerful tool specifically designed for creating detailed and interactive visualizations.

**Microsoft Power BI:** Power BI allows users to create a wide variety of visualizations, including charts, graphs, maps, and more, to represent data in an intuitive and understandable format.

### **Did you know?**

AI helps everyday gadgets like virtual assistants (**Siri and Alexa**) understand your questions and perform tasks, making technology feel more initiative.

**Q.13 Explain the process of data Pre-Processing and Evaluate Data Techniques.** 09509013

**Ans.** Data pre-processing is the first and most important step in working with data. It involves getting the data ready for analysis by cleaning and organizing it. Think of it like preparing ingredients before cooking a meal; you need to wash, chop, and measure everything so that the cooking process goes smoothly.

## Data Pre-processing Techniques

In data pre-processing, we use various techniques to ensure the data is accurate and ready for analysis. Here are some important steps and techniques:

### Evaluating Data Quality

Before using data, we need to check its quality, ensuring it is accurate, complete, and reliable. We ask questions like: Is any data missing? Are there errors or incorrect entries? Is the data consistent and up-to-date?

**Example:** Imagine you have a list of students' names and their test scores. You need to check if all students have scores recorded, if the scores are correct, and if they are from the current session.

**Errors** are mistakes in the data.

**Example:** Check, if the data is within a valid range. For instance, if maximum marks of a subject are 100 and a student's score is recorded as 105, it's clearly an error because scores can't exceed 100.

**Outliers** are unusual or extreme values that don't fit the pattern of the rest of the data.

**Example:** In a list of test scores, if most students scored between 50 and 80, but one student scored 5, the score of 5 is an outlier.

**Biases** are distortions that affect the accuracy of the data.

**Example:** If a survey only includes answers from students in one school, it may not represent the opinion of all students in the city. Hence if you want to analyse opinion of entire city students then this data would be biased due to the limited sample.

## Q.14 Describe data analysis techniques with the help of quantitative & statistical analysis.

09509014

**Ans.** Data analysis involves examining data to discover useful information, draw conclusions, and support decision-making. There are two main types of data analysis: quantitative and qualitative.

### Quantitative Analysis

Quantitative analysis deals with numbers and measurable data. It helps us understand patterns, relationships, and trends in numeric data.

**Statistical Analysis.** Statistical analysis uses math to make sense of data. Here are some key concepts:

- **Measures of Centre:** Measures of centre are statistical tools that help us understand the central point or typical value in a set of data. They provide a summary of a dataset by identifying the middle point. The three most common measures of centre are the mean, median, and mode.
- **Mean (Average):** The sum of all values divided by the number of values representing the average.

### Example

For example, if five students have scores of 70, 80, 90, 100, and 110, the mean score is  $(70+80+90+100+110)/5=90$ .

- **Median:** The middle value when all values are arranged in order. If there is an even number of values, the median is the average of the two middle values.

**Example:** For example, for the scores 70, 80, 90, 100, and 110, the median is 90. For 70, 80, 90, 100, 110, and 120, the median is 95.

- **Mode:** The value that appears most frequently. There can be more than one modes if multiple values appear with the same frequency.

**Example:** For example, in the scores 70, 80, 80, 90, and 100, the mode is 80 and the scores 70, 80, 80, 90, 100 has two modes 70 and 80.

- **Measures of Spread:** Measures of spread (also known as measures of dispersion) are statistical tools used to describe the amount of variation or diversity in a dataset.
- **Range:** The range is the simplest measure of spread. It is the difference between the highest and lowest values. For example, for the test scores 60, 75, 83, 91, and 95, the range is  $95 - 60 = 35$ .
- **Variance:** Variance measures how spread out the values are from the mean. It gives a sense of how much the values in a dataset vary from the mean. Formula for the variance is:

$$S^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}$$

Where:

- $S^2$  Sample Variance
- $x_i$  = the value of one observation
- $\bar{x}$  = the mean value of all observations
- $n$  = the number of observations
- $\Sigma$  = Sigma symbol; used to represent sum of values

For example, for the dataset [2, 4, 6], the mean is 4. The variance is calculated as  $[(2 - 4)^2 + (4 - 4)^2 + (6 - 4)^2] / 2 = (4 + 0 + 4) / 2 = 8/2 = 4$ .

**Standard Deviation:** This measures the amount of variation or spread in a set of data. A low standard deviation means that the data points are close to the mean, while a high standard deviation means that the data points are spread out over a wide range. Formula for standard deviation is:

$$S = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$$

### Q.15 What are the methods of Qualitative Analysis?

09509015

**Ans.** Qualitative analysis deals with non-numeric data such as text, images, and sounds. It helps us understand meanings, concepts, and experiences.

#### Methods for Analysing Qualitative Data

One common method is **Content Analysis**, which involves counting the number of times specific words or themes appear in a text. For example, researchers might analyse a set of interviews to count how many times students mention "homework stress." This method helps to quantify the presence of specific themes or ideas within qualitative data.

Another method is **Thematic Analysis**, which involves identifying and Interpreting themes or patterns within qualitative data.

### Q.16 Describe Cloud Storage and Remote Access.

09509016

**Ans.** In today's world, working together and sharing information online is essential. Collaborative tools make this possible, allowing people to work together, share files, and access their work from anywhere.

#### 1. Cloud Storage for Data Management

Using cloud storage for managing data has become an essential part of how we store, access, and share information. It allows us to save files on the Internet for access from any device, make backups to protect our data, and work on documents with others in real-time.

#### 2. Remote Access

Remote access refers to the ability to connect to and use a computer or network from a distant location. This means you can work on files, use software, or access resources on a

computer or network that is not physically near you. For example, if you save a school project on a cloud storage service like Google Drive while at home, you can later access and edit that same project from a computer at school. Remote access is made possible through the Internet, allowing you to connect to your files and applications from anywhere, as long as you have an Internet connection. Following is the activity flow for remote access.

- **Save Your Work:** Open Google Drive and upload your project file.
- **Access from Anywhere:** Log into Google Drive from any device (computer, tablet, phone) and open your project.
- **Continue Working:** Make changes, add new information, or review your work from wherever you are.

#### **Q.17 How do data backups support collaborative authoring, and what best practices ensure both data security and seamless collaboration among multiple users?**

09509017

**Ans.** Data Backups are copies of important data or files stored separately from the original to protect against data loss. Backups are essential to ensure that you can recover your data if something goes wrong, such as accidental deletion, hardware failure, or a computer virus.

- **Automatic Backups:** Set up your device to automatically back up important files to a cloud service like OneDrive.
- **Restore Files:** If you lose a file, log into your cloud account and download the backup version.
- **Resume Work:** Continue working on your restored file without worrying about losing data.

#### **Collaborative Authoring**

Collaborative authoring is the process of multiple people working together to create, edit, and improve a document or project in real-time, often using online tools and platforms that allow everyone to contribute and see each other's changes instantly. Following is the activity flow for collaborative authoring.

- **Create a Shared Presentation:** In Google Slides, create a presentation called "Climate Change Presentation" and share it with your friends.
- **Upload Files:** Each group member work on their slides and notes.
- **Edit Together:** Open the files and make edits together, ensuring everyone's input is included.

#### **Benefits of Collaborative Tools**

1. **Enhanced Productivity:** When multiple students work on different sections of a project simultaneously the project gets done faster. This saves time and allows everyone to contribute their best work.
2. **Version Control:** Google Docs automatically saves every change, so you can go back to previous versions if needed. You don't lose work, and you can see who made which changes in the document.

#### **Q.18 What is Data Science, and how does its workflow function? Could you provide an explanation?**

OR

#### **What do you know about Data Science and its Workflow? Explain.**

09509018

**Ans.** Data science is like being a detective, but instead of solving crimes, you solve problems using data. Imagine you want to figure out why some students do better in exams than others. Data science helps you gather information about students' study habits, look for patterns, and find out what makes a difference.

#### **Understanding Data Science**

Knowing about data science is important because it helps us make better decisions in our daily lives. For example, if you understand how to analyse data, you can figure out the best time to study when you're most focused. Businesses use data science to learn what their customers like and improve their products. Sports teams use data to create better strategies and enhance their performance. Even doctors use data science to predict health trends and provide better care. Data science is unique because it combines different fields to work with data effectively. It includes:

- **Computer Science:** For handling and organizing data.
- **Mathematics and Statistics:** For analysing data and finding patterns.
- **Business Knowledge:** For applying these insights to real-life problems and making informed decisions.

### Data Science Workflow

- **Data Science Workflow:** It refers to the systematic process used by data scientists to extract insights and knowledge from data. Here are the steps of Data Science Workflow process:
- **Problem Identification:** Understanding and clearly defining the problem you are trying to solve.
- **Data Collection:** Gather information from various sources. For example, if you want to know how many students like different sports, you might conduct a survey. This is like collecting ingredients for a recipe.
- **Data Cleaning:** After collecting data, it's important to clean it by removing errors and organizing it. Imagine your room is messy, and you need to put everything in its right place. Data cleaning involves fixing mistakes and making the data useful.
- **Data Analysis:** This step involves looking closely at the data to find patterns or answers. For example, if you have weather data from the past month, you can analyse it to predict if it will rain tomorrow. It's like solving a puzzle by examining the pieces.
- **Data Interpretation:** Once you have analysed the data, you need to understand it and draw conclusions. This is like reading a story and understanding its meaning. For example, after analysing survey results, you might conclude that most students prefer outdoor sports over indoor ones.
- **Data Visualization:** To make data easier to understand, you can create charts and graphs.

Q.19 Provide a comprehensive analysis on the following terms:

09509019

**Big Data**

**Volume**

**Velocity**

**Variety**

Ans. Big Data is a term used to describe extremely large and complex data sets that can't be managed with traditional tools. To understand Big Data better, we can look at its defining characteristics, often referred to as the "**Three Vs.**"

Big Data refers to extremely large and complex sets of data that are difficult to process using traditional methods. Imagine trying to organise a huge pile of documents, photos, and videos from thousands of people. Big Data involves using advanced technology to analyse and find useful information in this massive amount of data.

**Volume:** This refers to the sheer amount of data collected. For example, consider the countless posts, likes, and comments shared on social media platforms every day. This enormous amount of information is an example of "volume" in Big Data. It's like trying to fit a huge pile of puzzle pieces into one box—they just keep coming!

**Velocity:** This is about the speed at which data is generated and processed. For instance, think about how quickly posts are sent and received. The fast flow of this information means that data is constantly being updated and processed in real-time. Imagine a busy highway with cars zooming by, this is similar to the rapid pace at which data moves in Big Data.

**Variety:** This refers to the different forms data can take. Data is not just numbers; it can also include text, images, videos, and more. For example, a company might collect customer reviews as text, product photos as images, and sales figures as numbers. Just like having a mixed bag of different types of candy, Big Data includes a wide range of data types and formats.

### Q.20 Describe Applications of Big Data.

09509028

**Ans:** Following are the applications of big data that plays an active role in data science:

**Businesses:** Big Data is valuable because it helps businesses make better decisions. By analysing vast amounts of data, companies can uncover trends and patterns that might not be noticeable with smaller data sets. For instance, a retailer can use Big Data to determine which products are most popular at different times of the year. This helps them stock the right items and improve sales.

**Retail:** Stores use Big Data to understand what customers like and to boost sales. For example, an online shop might analyse your browsing history and past purchases to suggest products you might be interested in. It's like how a friend might recommend books based on the ones you've enjoyed before.

**Healthcare:** Hospitals and doctors use Big Data to monitor patient health and predict disease outbreaks. For instance, by analysing patient records, they can anticipate flu seasons and prepare vaccines in advance. It's similar to preparing for a storm by checking weather patterns to ensure you're ready.

**Finance:** Banks use Big Data to detect fraud and manage risks. For example, analysing transactions helps them spot unusual activities that might indicate fraud. This is like having a security system that alerts you if someone tries to break into your house.

**Transportation:** Companies like Uber use Big Data to optimize routes and manage rides. They analyse data from previous trips to find the quickest routes and predict where rides will be needed most. It's like a GPS system that learns the best routes based on past traffic patterns.

### Q.21 Which type of tools are used in data science modeling? Explain All.

09509028

**Ans:** Data science relies on various tools and technologies to help manage, analyse, and visualise data effectively. These tools and techniques are essential for turning raw data into useful insights.

#### Data Science Tools

**Excel:** Excel is a widely used spreadsheet program that helps you organise and analyse data. For example, if you have a list of your test scores, you can use Excel to create graphs that show your progress over time. This makes it easy to see how you're improving and identify any patterns in your performance.

**Python:** Python is a popular programming language used in data science for analysing data and building data models. It comes with powerful libraries like Pandas, for data manipulation and Matplotlib for creating graphs. For instance, you might use Pandas to analyse data from a survey and then use Matplotlib to create a chart that shows the results.

**R:** R is another programming language designed for statistical analysis and data visualization. It is particularly good at handling complex data and presenting it in a clear way. For example, you might use R to analyse a dataset from a scientific experiment and create detailed plots that make the results easier to understand.

**SQL:** SQL (Structured Query Language) is used to manage and query databases. For instance, if you need to find all students who scored above 90 on a test from a large database, SQL allows you to extract this specific information quickly. It's like asking a librarian to find all books by

particular author from a huge collection. These tools are important for data scientists as they help in make sense of large of data, provide valuable insights, and support informed decisions.

## Q.22 What are the predictions for the future of digital tools in data management and analysis?

09509022

**Ans:** As technology continues to advance, several exciting changes are expected in the field of data management and analysis some of these are described as follows:

**Enhanced Automation:** In the future, data tools will be able to handle more tasks automatically, making data analysis faster and more accurate. For example, imagine a program that automatically cleans up data by fixing errors and organizing information without human intervention.

**Improved Data Privacy:** As we collect and use more personal data, keeping it safe will become even more important. Future tools will have better security features to protect our information from unauthorized access.

**Integration of AI and Data Science:** Artificial Intelligence (AI) will be increasingly integrated with data science tools to make analysing large amounts of data easier and faster. For example, AI-powered tools could automatically generate reports and create charts using the latest data.

**Advances in Visualization:** Future data tools will offer innovative ways to visualize data, helping us understand complex information more easily.

## Q.23 Explain the differences between qualitative and quantitative data. Provide examples of each type.

09509023

**Ans.** Data can be divided into two broad categories namely qualitative and quantitative.

1. **Qualitative Data:** Qualitative data refer to categories or labels used to describe the qualities or characteristics of something rather than its quantity. Key characteristics of qualitative data include non-numeric, descriptive, and categorical attributes.
  - **Non-Numeric:** Qualitative data is represented by words, labels, or symbols Instead of numbers. It describes attributes rather than quantities. Examples include the names of students in a class (e.g. Ali, Badar, Qasim) and the colors of cars in a parking lot (e.g, red, blue, green), where "name" and "color" are attributes.
  - **Categorical:** Qualitative data can be into categories or classes based on their characteristics. Examples include types of fruit (e.g, apple, banana, orange). job titles (e.g., manager, engineer, accountant), and book genres (e.g., fiction, non-fiction, mystery).
2. **Quantitative Data:** Quantitative data consists of numbers used to measure the quantity or amount of something. These data types answer questions like "How much?" or "How long?" and can be useful for mathematical calculations and statistical analyses. Key characteristics of quantitative data include being numerical, measurable, countable, and arithmetical.
  - **Numerical:** Quantitative data is expressed in numbers, representing a measurable quantity. Examples include heights in centimetres, weights in kilograms, and test scores.
  - **Measurable:** Quantitative data can be measured using instruments or tools. Examples include using a ruler for length, a scale for weight, and a thermometer for temperature.
  - **Countable:** Quantitative data can be counted or enumerated, particularly for discrete data. Examples include the number of students in a class or the number of cars in a parking lot.
  - **Arithmetical:** Quantitative data can be used in arithmetic operations. For instance, multiplying the unit price of a fruit by its weight to find the total price, or calculating annual school fees by multiplying monthly fees number of months.

**Q.24 Compare and contrast continuous and discrete data. Use examples to show how each type of data might be used in a school setting, such as in measuring student performance or tracking attendance.**

09509024

**Ans. Continuous Data:** Continuous data consists of values that can take any number within a given range, including fractions or decimals. Examples include student heights (150.5 cm, 160.2 cm), fruit weights (1.5 kg, 2.2 kg), and room temperatures (22.5°C, 23.7°C).

**Example 1:** For example, measuring the height of a plant, might reveal that it is 32.5 centimetres tall. If you measure how long it takes for the plant to grow from 20 to 30 centimetres, might show that it took 15 days. These measurements provide detailed information about the amount (height) or duration (growth time), helping us understand how much or how long something occurs.

• **Discrete Data:** Discrete data consist of distinct, separate values that are countable, often in whole numbers. Discrete data answer questions like "How many?" or "How often?"

**For example:** counting oranges in a basket might show there are 16, answering "How many?" similarly, tracking the number of oranges falling from a tree each week, with counts like 5 on Monday, 7 on Wednesday, and 4 on Friday, helps answer "How often?"

### **Example 2: Measuring student performance**

**Continuous Data:** Monitoring a student's GPA throughout time to identify trends in academic achievement.

**Discrete Data:** Counting the number of assignments completed by each student during the semester.

**Tracking Attendance:** Monitoring the percentage of attendance (e.g., 92.5%) throughout the year.

Discrete data refers to the total number of days a student was present or absent.

**Q.25 Analyze the benefits and challenges of using digital tools like Google Forms for data collection.**

09509025

**Ans. Google forms** are often used for creating surveys since they allow you to arrange events, ask questions of your staff or clients, and collect a variety of information in an easy and effective manner. Google forms support a variety of question formats, including short responses, paragraphs, multiple selection, verification boxes, pull-down, linear scale, and grids with numerous possibilities, among others.

#### **Benefits**

1. Advanced users can utilize regular expressions to customize the types of data that can be entered into a field. This further customizes the form.
2. Google Forms allows us to preview how the survey will appear before sending it to the recipients.
3. We can send the form via email, integrate it into our website, or share the link via social media or other channels.

#### **Challenges**

1. Google Forms has basic customization possibilities, but it may not have advanced tools for more complex survey designs. Users seeking highly customizable layouts or intricate question branching may find the platform too restricted.
2. While Google Forms offers a variety of question kinds, such as multiple-choice, short answer, and dropdown, it may not cover the entire range of question types required for specific research or survey needs. Users seeking sophisticated question kinds or experimental designs may need to look into alternate survey platforms.

**Q.26** Imagine you are tasked with organizing a school event and need to collect data on students' preferences for activities and refreshments.

09509026

**Ans.** Collecting information about students' preferences for activities and refreshments is critical when organizing a school event. Input can be gathered using a digital tool such as Google Forms or printed surveys. The survey should include areas for activities (e.g., sports, arts, games) and refreshments (checkboxes for popular items such as pizza and beverages). Once the data has been spread and collected, evaluate it to detect trends. For example, if the majority of students prefer outside games, emphasize them, and base the menu on popular beverages. Sharing the final plan with students demonstrates that their input was respected, resulting in a tailored and interesting event.

**Q.27** Explain the role of tables and charts in data analysis. Provide an example of how you could use a table or chart to present data about students' grades in different subjects.

09509027

**Ans. Tables:** Tables are rows and columns of numbers and language (mostly numbers). They enable quick access to and relatively simple comparison of data. If the data is organized chronologically (for example, sales figures over a ten-year period), the table can reveal trends—patterns of increasing or decreasing activity. Tables, on the other hand, are not always the most vivid or dramatic way to depict such trends or correlations between data; for that, a line graph.

**Charts:** Charts and graphs are simply another way to convey data that is also presented in tables. At the same time, charts and graphs provide less detail or precision than tables. Consider the contrast between a table of sales figures over a ten-year period and a line graph of the same data. The graph provides a clearer sense of the overall trend, but not the exact monetary amount. Other forms of charts and graphs include horizontal bar charts, vertical bar charts, and pie charts.

#### Example

Student Name	Computer	Math	Urdu
Jawad	85	75	58
Sajid	75	65	72
Hammad	76	66	78

**Q.28** Describe a situation where non-numeric data is essential. How would you collect, store, and analyze this type of data? Discuss the tools and techniques you would use to ensure the data is accurately interpreted.

09509028

**Ans. Non-Numeric:** Qualitative data is represented by words, labels, or symbols instead of numbers. It describes attributes rather than quantities. Examples include the names of students in a class (e.g. Ali, Badar, Qasim) and the colors of cars in a parking lot (e.g. red, blue, green), where "name" and "color" are attributes. More advanced tools, such as NVivo or ATLAS. They specialize in storing and maintaining qualitative data.

**NVivo** is a strong qualitative data analysis software that organizes, analyses, and visualizes non-numerical data.

**ATLAS** like NVivo, is a powerful qualitative data analysis program that enables researchers to manage, analyze, and visualize massive amounts of unstructured or qualitative data

## Topic Wise Short Questions (Additional)

### Data & Its Components

**Q.1 Differentiate between Qualitative and Nominal Data:** 09509029

**Ans.** Qualitative data refers to categories or labels that describe qualities or characteristics rather than quantities. **Nominal data** is used to label or categorize items without implying any order.

**Q.2 Distinguishes between Structured and Unstructured data.** 09509030

**Ans.** **Structured Data** is organized and formatted for easy searching and analysis: examples include data in spread-sheets and traditional databases. **Unstructured Data** is more free-form and doesn't fit into a specific format.

**Q.3 What is Spreadsheet?** 09509031

**Ans.** A spreadsheet is a digital application or software tool that organizes, analyzes, and stores data in tabular format. It is composed of a grid of rows and columns, with each intersection forming a cell.

**Q.4 Define Data Backup.** 09509032

**Ans.** A backup is a copy of important data or files stored separately from the original, used to protect against data loss.

**Q.5 What do you know about NoSQL?** 09509033

**Ans.** NoSQL stands for "Not Only SQL" and refers to a variety of database technologies designed to handle different types of data that may not fit well into traditional databases.

**Q.6 Define Data Pre-Processing.** 09509034

**Ans.** Data pre-processing is the initial and important step in working with data. It involves cleaning and organizing it for analysis.

**Q.7 What does represent Ordinal data?** 09509035

**Ans.** Ordinal data represents categories with a meaningful order, though the differences between categories are not uniform.

**Q.8 Define data science.** 09509036

**Ans.** Data science is the branch of knowledge, in which computer programming skills along with mathematics and statistics are used to extract meaningful information from a collection of data.

**Q.9 What is a computer system?** 09509037

**Ans.** A computer system is a fundamental and important part of modern life. It has revolutionized the way we work, communicate, learn, and entertain ourselves. In today's world, we are surrounded by a lot of data, which may be on our computer system or otherwise.

**Q.10 Describe the role of big data in the Internet of Things (IoT).** 09509038

**Ans.** Big data plays a significant role in the Internet of Things (IoT) ecosystem by enabling the collection, storage, processing, and analysis of massive volumes of data generated by IoT devices. IoT devices generate continuous data.

**Q.11 Describe data visualization.** 09509039

**Ans.** Data visualization is the graphical representation of data to get meaningful insights, trends, and patterns from data. The visual elements which help in data visualization are charts, graphs, maps, figures, and dashboards, etc.

**Q.12 What are the online data sources? Give example.** 09509040

**Ans.** These include websites, databases, and digital tools where you can find and collect information. For example, if you're researching the most popular pets, you might use the internet to find statistics or articles about pet ownership.

**Q.13 What is a Survey?** 09509041

**Ans.** It is a method of collecting information from individuals. The basic purpose of a survey is to collect data to describe different characteristics such as usefulness, quality, price, kindness, etc.

#### **Q.14 Distinguishes between Primary data collection and Secondary data collection.**

**Ans.**

09509042

- **Primary Data Collection:** Involves the collection of original data directly from the data source or via direct interaction with the respondent.

- **Secondary Data Collection:** Involves data collection using existing data collected by someone else for some purpose. Such data is usually available in the form of published material like research papers, books, websites, etc.

#### **Q.15 How the process of data collecting can be measured?**

09509043

**Ans.** Data collection is the systematic process of gathering, documenting, and organizing data from a variety of sources to extract significant insights, facilitate informed decision-making, or allow for rigorous analysis.

#### **Q.16 Describe the role of data science.**

09509044

**Ans.** Data Science plays a crucial role in determining the place of various products in various shelves of the supermarket.

Data science involves gathering, analysing, and interpreting large amounts of data to find patterns and useful information.

### **Statistics**

#### **Q.17 What do you know about summary statistics?**

09509045

**Ans.** It is information about the data in a sample. It can help understand the values better. It may include the total number of values, minimum value, maximum value, along with the mean value and the standard deviation corresponding to a data collection.

#### **Q.18 Define the term Statistics and Probability?**

09509046

**Ans.** Statistics is the analysis of the frequency of past events and probability is to predict the likelihood of future events. Statistics is used in various fields such as

science, business, economic, social sciences, and more, to understand patterns, trends, and relationships within data.

#### **Q.19 Define E-Commerce.**

09509047

**Ans.** E-commerce, short for electronic commerce, refers to the buying and selling of goods and services over the internet. It encompasses a wide range of online transactions, including retail sales, business-to-business (B2B) transactions, online auctions, and digital downloads of products like software, music, and books.

### **Big Data**

#### **Q.20 What are the three Vs of big data?**

09509048

**Ans.** The three Vs of big data are:

- **Volume:** Refers to the amount of data. Big data deals with huge volumes of low-density, unstructured data.
- **Velocity:** Refers to the speed of data, which is the fast rate at which data is received.
- **Variety:** Refers to the various formats and types of data that are available. Traditional data types were structured and fit neatly in a relational database.

#### **Q.21 Describe any two benefits of big data.**

09509049

**Ans.**

- **Predictive Maintenance:** It is a maintenance strategy that uses the analysis of existing data to predict when equipment, machinery, or a product is likely to fail.
- **Fraud and Compliance:** Big data analytics can identify and detect unusual suspicious patterns and anomalies.

### **Quantitative analysis**

#### **Q.22 Define Range.**

09509050

**Ans.** The range is the simplest measure of spread. It is the difference between the highest and lowest values. **For example**, for the test scores 60, 75, 83, 91, and 95, the range is  $95 - 60 = 35$ .

### **Q.23 Define Variance.**

09509051

**Ans.** Variance measures how spread out the values are from the mean. It gives a sense of how much the values in a dataset vary from the mean. Formula for the variance is:

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

**Q.24 Write the formula of Standard Deviation.** 09509052

**09509052**

**Ans.** Formula for standard deviation is:

$$S = \sqrt{\frac{\sum (xi - \bar{x})^2}{n - 1}}$$

## Topic Multiple Choice Questions (Additional)

## Choose the correct option

## Data & Its Analysis



7. Which of the following is an example of interval scaled attribute? 09509059

- (a) Age
- (b) Temperature in Celsius
- (c) Height
- (d) Weight

## **Big Data**

8. Which of the following is not one of the characteristics of big data? 09509060

(a) Veracity      (b) Velocity  
(c) Validity      (d) Volume

9. Which is one of the key applications of big data in business? 09509061

  - (a) Social media management
  - (b) Customer behavior analysis
  - (c) Weather forecasting
  - (d) Artistic creativity

10. How does data science add value to big data? 09509062

  - (a) By creating more data
  - (b) By making data storage more efficient
  - (c) By extracting insights and predictions from data
  - (d) By ensuring data quality and consistency

11. In the context of big data, what does the 3Vs refer to? 09509063

- (a) Veracity, Velocity, Veracity
- (b) Volume, Value, Velocity
- (c) Validity, Variety, Volume
- (d) Variety, Velocity, Volume

# Data Science

12. Which of the following key concepts involves transforming raw data into a structured format that is suitable for analysis? 09509064

- (a) Data modeling  
 (b) Data exploration  
 (c) Data collection  
 (d) Data ethics
- 13. What is the role of domain knowledge in data science?** 09509065
- (a) Writing code for data analysis  
 (b) Conducting statistical tests  
 (c) Understanding the specific context of data  
 (d) Visualizing data
- 14. Which big data technology is commonly used to store and process large datasets in a distributed manner?** 09509067
- (a) SQL databases  
 (b) Hadoop  
 (c) Relational databases  
 (d) Data warehouses
- 15. Which of the following best describes the concept of big data?** 09509068
- (a) Data that is stored in traditional databases  
 (b) Data that is too small to be analyzed effectively  
 (c) Data that is characterized by high volume, velocity, and variety  
 (d) Data that is only generated by social media platforms
- 16. What is the primary challenge associated with traditional data processing techniques when dealing with big data?** 09509069
- (a) Lack of storage capacity  
 (b) Slow processing speed  
 (c) Inability to handle diverse data types and large volumes  
 (d) Limited availability of data visualization tools
- 17. Which programming language is commonly used for big data analytics and machine learning tasks?** 09509070
- (a) Java  
 (b) Python  
 (c) C++  
 (d) Ruby
- 18. What is the primary purpose of data preprocessing in big data analytics?** 09509071
- (a) To increase data storage costs  
 (b) To reduce the volume of data  
 (c) To prepare raw data for analysis  
 (d) To eliminate the need for data visualization
- 19. \_\_\_\_\_ is a structured or processed collection of data usually associated with a unique body of work.** 09509072
- (a) Data base  
 (b) Dataset  
 (c) Data and Information  
 (d) Information
- 20. \_\_\_\_\_ refers to the process of carefully examining and studying data to identify patterns, draw conclusions, or make the data meaningful.** 09509073
- (a) Data analytics  
 (b) Data predictions  
 (c) Dataset  
 (d) Database
- 21. \_\_\_\_\_ is the graphical representation of data through use of common charts, plots, infographics and animations** 09509074
- (a) Data cleaning  
 (b) Missing values  
 (c) Data visualization  
 (d) Data hiding
- 22. \_\_\_\_\_ is the fast rate at which data is received and acted on.** 09509075
- (a) Volume  
 (b) Velocity  
 (c) Variety  
 (d) Vision
- 23. \_\_\_\_\_ includes the data which can only take certain values and cannot be further subdivided into smaller units.** 09509076
- (a) Statistical data  
 (b) Continuous data  
 (c) Ordinal data  
 (d) Referral data
- 24. \_\_\_\_\_ is limitation of big data.** 09509077
- (a) Statistical data  
 (b) Unlimited growth of data  
 (c) Data visualization  
 (d) Predictive maintenance

- 25.** Customer satisfaction level such as satisfied, dissatisfied, and neutral are examples of \_\_\_\_\_ data type.  
 (a) Ordinal data  
 (b) Continuous data  
 (c) Numerical data  
 (d) Discrete data

09509078

- 26.** \_\_\_\_\_ is a method of collecting information from individuals. 09509079  
 (a) Survey  
 (b) Data hiding  
 (c) Data visualization  
 (d) Data finding

### Answer Key

1	c	2	c	3	c	4	d	5	c	6	c	7	b	8	c	9	b	10	c
11	d	12	a	13	c	14	b	15	c	16	c	17	b	18	c	19	b	20	a
21	c	22	b	23	c	24	b	25	a	26	a								

### Solved Exercise

**Choose the correction option.**

- 1. What is data?** 09509080  
 (a) Processed information  
 (b) Raw facts gathered about things  
 (c) A collection of numbers only  
 (d) A list of observed events
- 2. Which of the following is an example of qualitative data?** 09509081  
 (a) Temperature readings in degrees Celsius  
 (b) Number of students in a class  
 (c) Favourite ice cream flavours  
 (d) Test scores out of 100
- 3. What type of data involves distinct, separate values that are countable?** 09509082  
 (a) Nominal Data  
 (b) Ordinal Data  
 (c) Discrete Data  
 (d) Continuous Data
- 4. What is an example of continuous data?** 09509083  
 (a) Number of cars in a parking lot  
 (b) Height of students in centimetres  
 (c) Types of fruits  
 (d) Shirt sizes (small, medium, large)
- 5. What type of data is used to categorize items without implying any order?** 09509084  
 (a) Ordinal Data (b) Discrete Data  
 (c) Nominal Data (d) Continuous Data

- 6. How can you organise data to make it easier to analyse?** 09509085  
 (a) By writing it in long paragraphs  
 (b) By creating tables, charts, and graphs  
 (c) By storing it in random files  
 (d) By keeping it in a messy notebook
- 7. Which tool can be used to create surveys online?** 09509086  
 (a) Microsoft Word  
 (b) Google Forms  
 (c) Excel Spreadsheets  
 (d) Adobe Photoshop
- 8. What is the main purpose of data collection?** 09509087  
 (a) To create random numbers  
 (b) To gather information to answer questions or make decisions  
 (c) To delete old data  
 (d) To format text documents
- 9. What is the primary purpose of data visualization?** 09509088  
 (a) To convert text into data  
 (b) To generate random numbers  
 (c) To make data easier to understand by turning it into pictures  
 (d) To hide complex data

**10. Which tool is specifically designed for creating detailed and interactive visualizations?**

09509089

- (a) Microsoft Excel
- (b) Google Sheets
- (c) Tableau
- (d) PowerPoint

**11. What is the first step in the data science process?**

09509090

- (a) Data Cleaning
- (b) Data Analysis
- (c) Data Collection
- (d) Understanding the problem

**12. What does the 'Volume' characteristic of Big Data refer to?**

09509091

- (a) The speed at which data is generated
- (b) The different forms data can take

- (c) The sheer amount of data being collected
- (d) The way data is processed

**13. What is an outlier in a dataset?** 09509092

- (a) The most frequent value
- (b) The average of all values
- (c) An unusual or extreme value that doesn't fit the pattern.

- (d) The middle value when all values are arranged in order

**14. What does data encryption do?** 09509093

- (a) It converts data into a code to prevent unauthorized access.
- (b) It makes data available to everyone online.
- (c) It automatically deletes old data.
- (d) It speeds up internet connection.

### Answer Key

1	b	2	c	3	c	4	b	5	c	6	b	7	b	8	b	9	c	10	c
11	d	12	c	13	c	14	a												

### Short Questions

**Q.1 What is the difference between qualitative and quantitative data?**

09509094

**Ans.** Qualitative data refer to categories or labels used to describe the qualities or characteristics of something rather than its quantity.

Quantitative data consists of numbers used to measure the quantity or amount of something. These data types answer questions like "How much?" or "How long?".

**Q.2 Give an example of continuous data and explain why it is considered continuous.**

09509095

**Ans.** Continuous data consists of values that can take any number within a given range, including fractions or decimals. Examples include student heights (150.5 cm, 160.2

cm), fruit weights (1.5 kg, 2.2 kg), and room temperatures (22.5°C, 23.7°C).

**Q.3 Which method would you use to collect opinions from a large group of people about a new school policy?** 09509096

**Ans.** Questionnaires are similar to surveys but are often written forms that people fill out. They usually have a set of questions that need to be answered.

**Q.4 What type of data is the number of students in your class?** 09509097

**Ans.** The number of students in your class is quantitative (numerical) data, specifically discrete data, because it represents a countable quantity.

**Q.5 Why is it important to organize data into tables or charts before analyzing it?** 09509098

**Ans.** Proper organization saves time. Think about searching for a book in a messy room versus a neatly arranged bookshelf.

Similarly, when data is organized, it's easier to find and analysis.

**Q.6 What is one advantage of using online tools like Google Form for collecting survey data?**

09509099

**Ans.** A free tool from Google that helps you create surveys and collect responses online.

Link for Google Forms is  
<https://forms.google.com>

**Q.7 Why might you need to integrate data from different sources when working on a project?**

09509100

**Ans.** Data integration involves combining information from different sources to get a complete perspective.

**1. Organize Data:** Put all your saved information into one place, such as a document or spreadsheet.

**2. Compare and Merge:** Review data from different sources and find common themes. For instance, if many sources say that social

**Q.8 Describe a scenario where discrete data might be more useful than continuous data.**

09509101

**Ans:** When a school organizes many activities and wishes to track the number of students who attend each one, discrete data is more relevant than continuous data.

**Q.9 Explain why data visualization is important. How does it help in understanding complex information?**

09509102

**Ans:** Data visualization is important because it makes understanding data much easier. Imagine trying to understand a list of hundreds of numbers. It's tough, right? But if those numbers are displayed as a bar chart, it's much easier to see what's going on.

**Q.10 Describe what a line graph is used for and provide an example of data that could be displayed using a line graph.**

09509103

**Ans:** Line graphs, often known as line charts or line plots, are data visualization tools. It is a type of graph that depicts data in graphical format, making the raw data more intelligible.

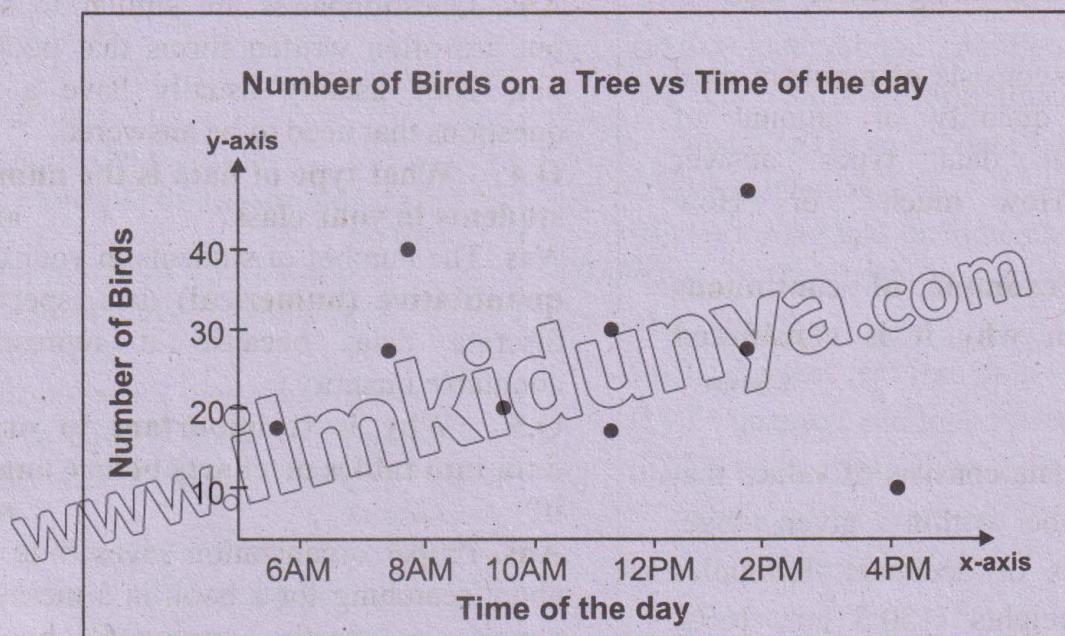
**Example:** A chart displaying the number of students in a class that attend school every day for a week.

**Q.11 Explain the use of scatter plots in visualizing continuous data. Provide an example of a situation where a scatter plot would be useful.**

09509104

**Ans:** A scatter plot is a useful tool for displaying the relationship between two variables and recognizing patterns like positive, negative, or no association.

**Example**



## Long Questions

1. Explain the differences between qualitative and quantitative data. Provide examples of each type. 09509105

Ans. Long Question No. 23

2. Describe the process of conducting a survey to gather data about students' favorite extra curricular activities. 09509106

Ans. Long Question No. 7 (Point Number 1)

3. Compare and contrast continuous and discrete data. Use examples to show how each type of data might be used in a school setting, such as in measuring student performance or tracking attendance. 09509107

Ans. Long Question No. 24

4. Analyze the benefits and challenges of using digital tools like Google Forms for data collection. 09509108

Ans. Long Question No. 25

5. Imagine you are tasked with organizing a school event and need to collect data on students' preferences for activities and refreshments. 09509109

Ans. Long Question No. 26

6. Explain the role of tables and charts in data analysis. Provide an example of how you could use a table or chart to present data about students' grades in different subjects. 09509110

Ans. Long Question No. 27

7. Describe a situation where non-numeric data is essential. How would you collect, store, and analyze this type of data? Discuss the tools and techniques you would use to ensure the data is accurately interpreted. 09509111

Ans. Long Question No. 28

8. Explain the concept of data visualization. How does it help in understanding complex data? Provide examples of different types of visualizations and their applications in real-life scenarios. 09509112

Ans. Long Question No. 11

9. Discuss the importance and benefits of data visualization. Why is it essential for businesses and decision-makers to use charts, graphs, and dashboards?

Ans. Long Question No. 12

10. Differentiate between nominal, ordinal, discrete, and continuous data. For each type, describe a suitable visualization technique and provide a specific example of how this technique can be used to represent that type of data effectively.

Ans. Long Question No. 4

## Activities

### Activity 1

#### Exploring Data Types

Activity Type: Group

#### Task Details:

Group Formation: Divide the class into small groups. Assign each group a type of data (Qualitative, Quantitative, Nominal, Ordinal, Discrete, Continuous).

Research and Design: Each group researches their assigned data type and creates a poster or presentation. They should include:

09509113

1. Definition: Clear explanation of the data type.
2. Characteristics: Key features of the data type.
3. Examples: Real-life examples or scenarios where this data type is used.
4. Operations: Possible operations or analyses that can be performed with this data type.

Presentation: Each group presents their poster to the class explaining their data type and answering any questions from classmates.

Class Discussion: Facilitate a class discussion on how different data types are used in various real-life situations and data analysis.

For more activities, you can visit, <https://mentossacademia.com>

#### **Ans. Class Work/ Lab Work/ Practical Work**

### **Activity 2**

09509114

#### **Data Collection and Organization**

**Activity Type:** Individual

##### **Activity Task Details:**

- **Create Survey:**

1. Design a short survey with clear, specific questions on a topic of choice (e.g., favorite school subject).

2. Ensure questions are designed to gather meaningful data.

- **Collect Responses:**

1. Share the survey link with classmates or a small group to gather responses.
2. Collect data over a specified period.

- **Organise Data:**

1. Enter the collected data into a spreadsheet.
2. Organise the data into tables and use basic spreadsheet functions to clean and prepare it.

- **Create Visuals:**

1. Use spreadsheet tools to create at least one chart or graph (e.g., bar chart, pie chart).
2. Label charts clearly and ensure they represent the data accurately.

- **Share Findings:**

1. Present the findings to the class.
2. Explain the visuals and what they reveal about the data.

#### **Ans. Class Work/ Lab Work/ Practical Work**

### **Activity 3**

09509115

#### **Analysing and Interpreting Data**

**Activity Type:** Group

##### **Activity Task Details**

- **Data Organization**

1. Each group receives a data set (e.g., student test scores).
2. Organise the data into a structured table.

- **Create Graphs**

1. Use graphing tools or paper to create graphs (e.g., bar chart, line graph) to represent the data visually.
2. Ensure graphs are appropriately labelled and scaled.

- **Analyse Results**

1. Analyse the graphs to identify trends, patterns, or outliers in the data.

2. Discuss possible interpretations of the results.

- **Report Findings**

1. Prepare a brief report summarizing the analysis.

2. Share the report with the class, highlighting key findings and insights.

**Ans. Class Work/ Lab Work/ Practical Work**

09509116

#### **Activity 4**

##### **Visualizing Data Types**

**Activity Type:** Group

**Activity Task Details:**

##### **Assign Data Types**

1. Divide the class into groups, each assigned a specific data type (nominal, ordinal, discrete, continuous).

##### **Create Visualizations**

2. Each group creates a poster using appropriate charts or graphs for their assigned data type:

- **Nominal Data:** Bar charts or pie charts.

- **Ordinal Data:** Bar charts or stacked bar charts.

- **Discrete Data:** Histograms or dot plots.

- **Continuous Data:** Line graphs, scatter plots, or box plots.

- **Present Findings:**

1. Each group presents their poster and explains why the chosen visualization method is appropriate.

2. Discuss the effectiveness of different visualizations for different data types.

**Ans. Class Work/ Lab Work/ Practical Work**

09509117

#### **Activity 5**

##### **Data Cleaning and Validation**

**Activity Type:** Individual

**Activity Task Details:**

##### **Evaluate Data:**

1. Review a data table with intentional errors (e.g., incorrect values, missing entries).

2. Identify errors, missing values, and outliers.

##### **Clean Data:**

1. Apply techniques to correct errors (e.g., correcting incorrect values).

2. Fill in missing values using reasonable methods.

3. Address outliers by investigating and deciding on appropriate actions.

##### **Explain Changes:**

1. Write a brief explanation of the changes made.

2. Justify why each correction was necessary for accurate analysis.

**Ans. Class Work/ Lab Work/ Practical Work**

09509118

#### **Activity 6**

##### **Exploring Online Data Sources**

**Activity Type:** Individual or Pair

**Activity Task Details:**

- **Data Gathering:**

1. Choose a topic (e.g., impact of social media on teenagers).

2. Use online databases and resources to find relevant and reliable information.
- **Data Extraction:**
    1. Extract key pieces of information from sources
    2. Save the data in a document or spreadsheet.
  - **Data Integration:**
    1. Organise the collected data into a coherent report or summary.
    2. Include any charts or graphs that represent the data.
  - **Presentation:**
    1. Present findings to the class.
    2. Explain the sources of information and how the data supports the conclusions.

**Ans. Class Work/ Lab Work/ Practical Work**

09509119

## **Activity 7**

### **Design Your Own Survey**

**Objective:** Learn how to design and analyse a survey for data collection.

**Required Material:** Paper and pencils, Access to survey creation tools

(optional: Google Forms or similar tools)

#### **Activity Task Details:**

- **Survey Design:**

1. Students design a survey on a topic of interest (e.g., favourite books, study habits).
2. Include a variety of question types (e.g., multiple-choice, short answer).

- **Collect Responses:**

1. If possible, use an online tool to distribute the survey to classmates or friends. Alternatively, gather responses on paper.

- **Analyse Data:**

1. Compile the survey results into a data table.
2. Create visual representations (e.g., bar charts, pie charts) to display the survey results.

- **Report:**

1. Write a brief report summarizing the survey results.
2. Include insights or trends observed from the data.

For more activities visit, <https://mentorsacademia.com>

**Ans. Class Work/ Lab Work/ Practical Work**