UNIT 9

# Data Science and Data Gathering

## Student Learning Outcomes

**By the end of this chapter, you will be able to:**

- Understand Data Types: Identify and differentiate between qualitative and quantitative data, and explain their importance in data analysis.
- Organise and Analyse Data: Organise data effectively and apply basic data analysis techniques to draw meaningful conclusions from both structured and unstructured data.
- Utilize Data Collection Methods: Gain knowledge of various data collection methods and tools, including surveys, questionnaires, and online sources, and understand how to use these methods to gather relevant data.
- Describe Data Storage Techniques: Describe different data storage techniques such as spreadsheets, databases, data warehouses, and NoSQL, and understand their applications in managing data.
- Apply Data Visualization Techniques: Organize the importance of data visualization and be able to use various tools to create visual representations of data, such as charts and graphs, to effectively communicate information.
- Pre-process and Analyse Data: Learn data pre-processing techniques and apply quantitative and qualitative analysis methods to interpret data and generate insights.
- Utilize Collaborative Tools and Cloud Computing: Understand the role of cloud storage and collaborative tools in data management, and apply best practices for data protection and remote access.
- Address Ethical Data Considerations: Be aware of ethical practices related to data privacy, confidentiality, and legal frameworks, and understand the importance of secure data handling.
- Explore Data Science Fundamentals: Gain an introduction to data science, including its definition, scope, and fundamental principles, and understand the impact of data science on various applications.
- Understand Big Data and Its Applications: Explain the concept of big data, its characteristics, and its practical applications in domains, such as retail, healthcare, finance, and transportation.
- Recognize Data Science Tools and Techniques: Learn about various data science tools and techniques, including predictive modelling and graph analytics, and understand their applications in solving real-world problems.
- Recognize Future Trends in Data Handling: Identify emerging trends in data analytics, artificial intelligence, and machine learning, and make informed predictions about future developments in digital data management and analysis.

## Introduction

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 organising and analysing data, using both quantitative and qualitative techniques. We'll also look into data visualisation, which helps us understand complex information through charts and graphs. Next, we will discuss collaborative tools and cloud computing, along with ethical issues involved in working with data. Finally, we introduce data science, big data, and future trends in digital data handling, showing how technology is advancing data analysis to make it more powerful and insightful.

## 9.1 Data

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.

**Examples:**

1. **Weather Data**: Information on temperature, humidity, and wind speed collected from weather stations.
2. **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.
3. **Survey Responses**: Answers from a survey about customer satisfaction. For example, 80% of customers rating their experience as "satisfactory" or better.
4. **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.
5. **Social Media Data**: Likes, comments, and shares on social media posts. For example, a post receiving 200 likes and 50 comments.

Understanding data is essential in today's world, as it allows us to comprehend situations, make informed decisions, solve problems, and drive innovation.

**Tidbits**

**Weather Data Impact:** Weather data helps meteorologists predict storms and extreme weather conditions. By analyzing weather data, scientists can issue early warnings for hurricanes, potentially saving lives.

## 9.2 Data Types

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

### 9.2.1 Qualitative Data

Qualitative data refer to categories or labels used to describe the qualities or characteristics of something rather than its quantity. This type of data offer a way to categorize and provide insights into opinions, behaviours, and experiences through descriptions rather than numbers. 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 colours of cars in a parking lot (e.g., red, blue, green), where "name" and "colour"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).

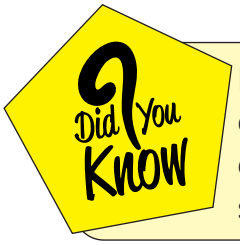
**Types of Qualitative data**

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 colours (red, blue, green). Nominal data is helpful for grouping and categorizing information. For instance, a survey asking about favourite colours might list responses such as 'red', 'blue', or 'green', which help indicate preferences without ranking them. On nominal data, you can check for equality, group items into categories (e.g., all red fruits together), count items in each category, and find the most frequent category (mode).
- **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.

In addition to the operations applicable to nominal data, ordinal data allow for

comparisons (e.g., bachelor's level is higher than light school), ranking items (e.g., from least to most satisfied), finding the median value (e.g., median satisfaction level), and analysing frequency distribution (e.g., the number of people at each education level).




Did you know that every day, people generate over 2.5 quintillion bytes of data? That's a 2 followed by 18 zeros! This data comes from various sources like social media, online shopping, and sensors in smart devices.

### 9.2.2 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 by the number of months.

**Types of Quantitative Data**

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

- **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?"

**Operations Performed on Discrete Data**

All operations that can be performed on nominal and ordinal data can also be performed on discrete data. In addition to logical and grouping operations discussed in qualitative data section, arithmetic operations (addition, subtraction) and statistical operations (average, range) can be performed on discrete data. **For example**, the difference between marks of two students in a class can be calculated by subtracting their marks. Statistical operations like calculating the average marks of students in a class (75, 63, 92, 55, 70 their average is 71), and determining the range of marks in class (minimum value 55 to maximum value 92) can also performed.

- **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

Continuous data are used when measuring items and can help answer questions about the extent or duration of something. 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.

**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.

Table 9.1 compares four types of data highlighting their nature, examples, and application in data analysis.

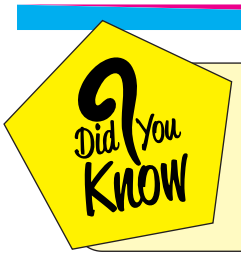
**Tidbits**

**Social Media Data Trends:** Social media platforms collect vast amounts of data about user interactions. For example, Facebook uses this data to recommend friends and tailor ads based on users' likes and interests.

| 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 rank. | Ranks items in a meaningful order. | Counts specific items or occurrences. | Measures the extent or duration of something. |
| **Representation** | Words, labels, symbols. | Words, labels, symbols with order. | Numbers, whole or discrete values. | Numbers, including fractions and decimals. |
| **Characteristics** | Non-numeric, categorical. | Ordered, but differences between categories are not uniform. | Countable, often whole numbers. | Measurable, can include fractions or decimals. |
| **Examples** | Gender (male, female); Types of fruits (apple, banana, orange); Colours (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). | Height of students (150.5 cm, 160.2 cm); Weight of fruits (1.5 kg, 2.2 kg); Temperature (22.5°C, 23.7°C). |
| **Operations** | Equality, grouping, counting, mode. | Equality, comparison, ranking, median, frequency distribution. | Equality, grouping, counting, arithmetic operations without division | Equality, grouping, arithmetic operations including division |
| **Statistical Analysis** | Mode, frequency counts. | Median, mode, frequency distribution. | Mean, median, range, standard deviation. | Mean, median, range, standard deviation. |
| **Usage** | Categorizing, understanding preferences or attributes. | Understanding order or levels of a characteristic. | Counting specific occurrences or items. | Measuring precise amounts, durations, or extents. |



**Table 9.1 Comparison of Data Types**




**Digital Data Explosion:** The amount of digital data created and consumed globally is so vast that it's predicted to reach 175 Zettabytes by 2025. That's a 1 followed by 21 zeros!

**Class activity**

**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:

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



## 9.3 Organising and Analysing Data

Organizing data systematically is very important for clear analysis and interpretation. When data is well-organised, it helps reduce errors. For example, imagine you have a list of students and their test scores, a messy, list might lead to accidentally recording a score under the wrong student's name. Organizing the data neatly in a table reduces such mistakes.

### 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 9.2.

| Student | Math | Science | English |
|---|---|---|---|
| Ali | 85 | 78 | 90 |
| Sara | 78 | 88 | 85 |
| Ahmed | 92 | 82 | 87 |
| Fatima | 90 | 80 | 89 |
| Bilal | 67 | 75 | 70 |

Table 9.2

Tables like this help present data clearly, making it easy to compare and analyse the scores of different students across various subjects.

**Charts:** Charts are visuals 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

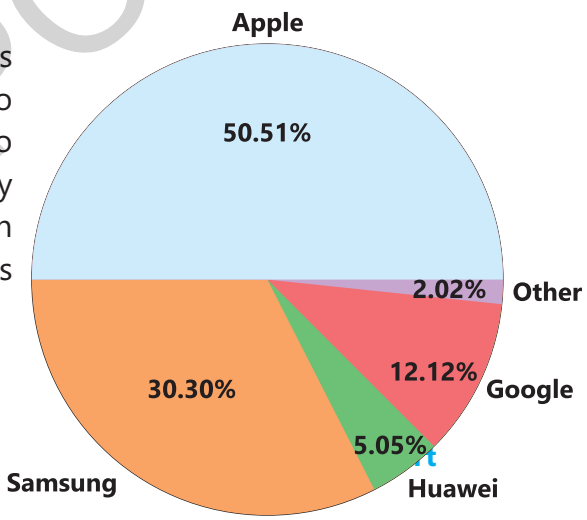



Fig 9.9 chart

**Graphs:** Graphs are visuals 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.


Fig 9.2 Graphs

### 9.3.1 Data Collection Methods and Tools

**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:

• **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

**Tidbits**

- A free tool from Google that helps you create surveys and collect responses online. Link for Google Forms is https://forms.google.com/
- Microsoft Office tools allows you to create surveys and quizzes Link for Microsoft Forms is https://forms.office.com/
- SurveyMonkey is also a popular tool for creating detailed surveys. Link for SurveyMonkey is https://www.surveymonkey.com/

- **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? (e.g., 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.

### 9.3.1.1 Gathering Data from Online Sources

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. Use Google search engine to search for articles and check if the information is accurate by looking at who wrote it and where it's from.

**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.

**Class activity**

**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., favourite 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.

## 9.4 Data Types

With respect to storage and processing, data has two types: structured and unstructured data.

### 9.4.1 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 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 9.3.

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

Table 9.3 Structured data

### 9.4.2 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."

**Tidbits**

**Structured vs. Unstructured Data:** Structured data, like spreadsheets, is neatly organised into rows and columns, making it easy to search and analyse. Unstructured data, like emails or social media posts, is more chaotic and requires special tools to process.

## 9.5 Data Storage Techniques

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.

### 9.5.1 Spreadsheets

Spreadsheets are tools that help us organise data in rows and columns, much like a simpler 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. You can perform calculations, create charts, and sort data easily. These are commonly used for personal budgeting, simple data analysis, and small business inventory management.

### 9.5.2 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. These are used in many everyday applications like banking systems, online shopping websites, and school records.

### 9.5.3 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 BigQuery,** a fast and scalable data warehouse offered by Google Cloud.

### 9.5.4 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. Unlike traditional databases that use tables, NoSQL databases can use various structures like documents, key-value pairs, graphs, or columns to store data. NoSQL databases are often used in big data applications, real-time web applications, and content management systems.

**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.

## 9.6 Data Visualization

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.

### 9.6.1 Importance and Benefits of Data Visualization

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.

### 9.6.2 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:

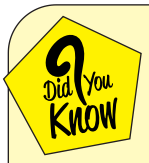
**Nominal Data:** Nominal data represents categories without any specific order. Bar charts and pie charts are great for showing nominal data.

**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.

Visualization technique of line graphs, scatter plot and box plot are commonly used for continuous data.




**Human Brain and Visualization:** Did you know that the human brain processes visuals 60,000 times faster than text? This is why charts and graphs can make complex information easier to understand quickly!

**Visualizing Ancient Data:** The earliest known example of data visualization dates back to 1786 when William Playfair created the first line graph. He used it to show trade and economic data!

## 9.7 Data Pre-Processing and Analysis

### 9.7.1 Data Pre-processing

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.

#### 9.7.1.1 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.

**Identifying Errors, Outliers, and Biases:**

**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 opinions 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.

### 9.7.2 Implementing Data Validation and Cleaning Processes

Once we identify issues, we need to fix them by validating and cleaning the data.

**Data Validation** involves checking data completeness and accuracy.

**Validating data completeness** means to ensure that no data is missing.

**Example**

For Example making sure every student has a test score recorded.

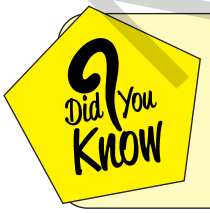
**Validating data accuracy** ensures the data is correct. **For example** verifying that all test scores are between 0 and 100.

**Data Cleaning** involves error removing, handling missing data and dealing with outliers. Errors are either corrected or the incorrect data is deleted.

**Example**

For example changing a score of 105 to a valid number like 95, or if the valid number is not known, deleting this record from data.

**To handle missing data**, decide how to deal with gaps for example if a student's test score is missing, you might use the average score of the class to fill in the gap. To **deal with Outliers**, decide whether to keep, change, or remove unusual values. **For example** investigating why a student scored 5 and determining if it's an error or a true value.




**Data Cleaning as Cooking:** Think of data cleaning like preparing ingredients for a recipe. Just like you wash and chop vegetables before cooking, you need to clean data to ensure it's ready for analysis.

### 9.7.3 Data Analysis Techniques

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.

#### 9.7.3.1 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.

The mean is useful when you want to find a single value that represents a typical data point in a dataset. It can only be applied on numeric data and works best when data lacks extreme values or outliers.

- **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.

The median is useful when the dataset has outliers (extremely high or low values) that could distort the mean. It gives a better representation of the typical value in such cases.

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

**Example**

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

The mode is useful for understanding the most common value in a dataset. It is particularly helpful for qualitative data where we want to know the most frequent category.

- **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. They provide degree to which data points differ from the average value (mean) or median. Here are the key measures of spread:
- **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 the one observation
- $\bar{x}$= the mean value of all observations
- $n$ = the number of observations
- $\sum$ = 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:

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

Standard deviation is the square root of the variance. It provides a measure of the average distance of each data point from the mean. For example, for data set [2,4,6], the variance is 4, So the standard deviation is $\sqrt{4} = 2$.

### 9.7.3.2 Qualitative Analysis

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. For instance, by reading through student essays, researchers can find common themes like "difficulty with math" or "enjoyment of science." This method provides a deeper understanding of the underlying themes and patterns in the data.

Class activity

**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.

## 9.8 Collaborative Tools and Cloud Storage

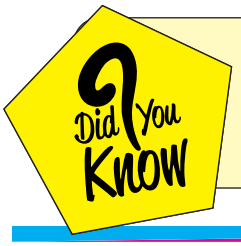
In today's world, working together and sharing information online is essential. Collaborative tools make this possible, allowing people to work together, share files, and access their work from anywhere.

### 9.8.1 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.

### 9.8.2 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




**Cloud Storage Evolution:** Storing files in the cloud is like having an infinite bookshelf. Instead of cluttering your physical space, your documents, photos, and videos can be stored online, freeing up room for more important things.

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.

### 9.8.3 Data Backups

**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. For example, if you are working on a school project and save it on your computer. To ensure you don't lose your work, you create a backup by saving a copy on Google Drive or a USB. This way, if your computer crashes or the file is accidentally deleted, you can easily recover it from Google Drive or the USB. Following is the activity flow for data backup on a cloud service.

- **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.

### 9.8.4 Collaborative Authoring

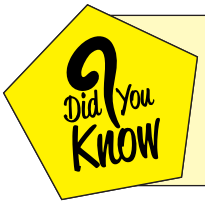
**Collaborative authoring** is the process of multiple people working together to create, edit, and improve a document or project in real-time, often using online tools and platforms that allow everyone to contribute and see each other's changes instantly. For example, if your teacher assigns a group project to design presentation on climate change, your group can use Google Slides to design the presentation. All students do their respective task such as to create the outline, design various slides including introduction, significance, conclusion etc., and everyone can comment and suggest changes. Here everyone can add their slides and notes to the same file remotely at their convenience. 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.

### 9.8.5 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.




**Global Collaboration:** Cloud-based tools have enabled people to work together on projects from different parts of the world. For instance, a student in Pakistan can collaborate on a science project with peers in the USA and Australia simultaneously.

## 9.9 Introduction to Data Science

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. For example, you might discover that students who study in groups tend to score higher. By using data science, you can make better decisions and find solutions to everyday problems. It combines computer skills, math, and business knowledge to turn raw data into useful insights, just like turning ingredients into a delicious recipe.

### 9.9.1 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. By learning data science, you can solve problems more effectively and make smarter choices, whether at school, work, or in everyday life.

### 9.9.2 Interdisciplinary Nature of Data Science:

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.

### 9.9.3 Data Science Workflow

Data Science Workflow, 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. This is like drawing a map to help someone find their way. For example, you might create a bar chart showing the number of students who like each sport, making it clear which sports are most popular.

**Example**

Imagine a school wants to understand why some students arrive late. They start by collecting data on arrival times (data collection). Next, they clean the data by fixing any errors (data cleaning). They then analyse the data to find patterns, like whether bad weather or traffic is causing the delays (data analysis). They interpret these patterns to understand why students are late (data interpretation). Finally, they create a chart to show the most common reasons for being late (data visualization). This entire process of collecting, cleaning, analysing, interpreting, and visualizing data is what data science is all about!

**Class activity**

**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:**
  1. 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.

## 9.10 Big Data and its Applications

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. This pile is so big and varied that special tools are needed to make sense of it all. Big Data involves using advanced technology to analyse and find useful information in this massive amount of data.

### 9.10.1 Introduction to Big Data

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."

**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.

### 9.10.2 Practical Applications of Big Data

**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.

In addition to improving product offerings, Big Data helps businesses understand their customers better. By examining customer feedback and behaviour, companies can make their products and services better.

**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.

### 9.10.3 Tools and Techniques in Data Science

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

#### 9.10.3.1 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 a particular author from a huge collection.

These tools are important for data scientists as they help in make sense of large amounts of data, provide valuable insights, and support informed decisions.

### 9.10.3.2 Data Science Techniques

Data science uses a variety of techniques to solve problems and gain insights from data. These techniques help in making predictions and understanding complex relationships within data.

**Predictive Modelling:** Predictive modelling is a technique used to forecast future events based on historical data. For example, a school wants to identify students who might need extra help with their studies. By using predictive modelling, the school can analyse past performance data to predict which students may struggle in the future. This allows them to provide additional support to those students before problems arise.

**Graph Analytics:** Graph analytics is a method for analysing relationships between different data points. It helps to visualize and understand connections

**Class activity**

**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.

and interactions. For instance, in social media, graph analytics can show how users are connected through their friendships or interactions. This helps companies understand social networks better and identify influential people within those networks, for marketing and information dissemination.

### 9.10.3.3 Applications of Data Science Techniques

Data science techniques are widely used across various industries to solve problems and make better decisions. Here are some real-world applications:

**Retail Industry:**

In retail, data science tools help stores understand shopping patterns and improve sales. For example, a store might analyze data on customer purchases and discover that people who buy winter coats often buy gloves as well. This information allows the store to place gloves near the coats, making it easier for customers to find them and increasing the chances of additional sales. By using data to organise their stores better, retailers can enhance the shopping experience and boost revenue.

**Healthcare:**

Hospitals use data science techniques to monitor patient health and predict trends. For instance, doctors may use predictive modelling to analyse patients' medical histories and identify those at risk of developing certain diseases, such as diabetes or heart disease. This helps doctors take preventative measures or offer early treatment, improving patient care and potentially saving lives.

**Finance:**

Banks and financial institutions use data science to detect and prevent fraud. For example, if a bank notices unusual spending patterns, such as a sudden large withdrawal from an account, it can use predictive modelling to flag these transactions as potentially fraudulent. This helps the bank act quickly to prevent theft and protect customers' money.

**Sports:**

In sports, data science techniques are used to enhance player performance and team strategies. For example, coaches analyse player statistics and game data to identify strengths and weaknesses. By understanding these patterns, they can develop strategies to improve team performance and make more informed decisions during games.

These applications show how data science techniques help solve problems and improve outcomes across different fields, making them invaluable tools for informed decision making and achieving better results.

### 9.10.3.4 Predictions for the Future of Digital Tools in Data Management and Analysis:

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.

**Class activity**

**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.

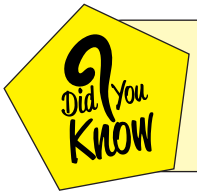
**Enhanced Automation:** In the future, data tools will be able to handle more tasks automatically, making data analysis faster and more accurate. For example, imagine a program that automatically cleans up data by fixing errors and organizing information without human intervention. This means you won't have to spend time on these repetitive tasks and can focus on more complex analysis, like finding trends or making predictions based on the data.

**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. For instance, new technologies might use advanced encryption methods to ensure that only authorized people can see sensitive data, such as your personal health records or financial information. This will help keep data safe from hackers and other security threats.

**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. This will help businesses and researchers quickly understand important trends and make decisions without manually processing the data.

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

For instance, interactive dashboards may ebnable users to explore data from different angles, such as clicking on different chart sections for more detailed information. This makes it easier to detect patterns and insights in the data, such as discovering which products are most popular in different regions.
These advancements in digital data handling will change how we analyse and use data, making it an exciting field with many possibilities for the future.




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

## Class activity

**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

## Summary

- Data refers to raw facts we gather about things around us, which can then process to extract useful information.
- 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.
- Ordinal data represents categories with a meaningful order, though the differences between categories are not uniform.
- Quantitative data are numbers used to measure the quantity or amount of something.
- Continuous data consists of values that can take any number within a given range, including fractions or decimals.
- 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.
- Spreadsheets organise data in rows and columns, much like a simpler version of a database.
- Databases are like digital filing cabinets where information is stored in an organised way.
- Data Warehouses are specialized databases designed for storing and analysing large amounts of data collected from various sources.
- 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.
- Data visualization turns numbers and information into visual representations, making data easier to understand.
- Data pre-processing is the initial and important step in working with data. It involves cleaning and organizing it for analysis.
- Data Cleaning involves error removing, handling missing data and dealing with outliers.
- Data analysis involves examining data to discover useful information, draw conclusions, and support decision-making.
- Qualitative analysis deals with non-numeric data such as text, images, and sounds. It helps us understand meanings, concepts, and experiences.
- Remote access refers to the ability to connect to and use a computer or network from a distant location.
- A backup is a copy of important data or files stored separately from the original, used to protect against data loss.

- Collaborative authoring is the process where multiple people work together to create, edit, and improve a document or project.
- Data science is like being a detective, solving problems using data instead of solving crimes.
- Data science involves gathering, analysing, and interpreting large amounts of data to find patterns and useful information.
- Data Analytics is a branch of data science focussed on analysing data to understand it better and support decisions making.
- Big Data is a term used to describe extremely large data sets that are so complex they can't be managed with traditional tools.


**Multiple Choice Questions (MCQs)**

1. What is data?

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?

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?

a) Nominal Data b) Ordinal Data
c) Discrete Data d) Continuous Data

4. What is an example of continuous data?

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?

a) Ordinal Data b) Discrete Data
c) Nominal Data d) Continuous Data

6. How can you organise data to make it easier to analyse?

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?
   a) Microsoft Word  b) Google Forms
   c) Excel Spreadsheets  d) Adobe Photoshop
8. What is the main purpose of data collection?
   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?
   a) To generate random numbers  b) To convert text into data
   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?
   a) Microsoft Excel  b) Google Sheets
   c) Tableau  d) PowerPoint
11. What is the first step in the data science process?
   a) Data Cleaning  b) Data Analysis
   c) Data Collection  d) Understanding the problem
12. What does the 'Volume' characteristic of Big Data refer to?
   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?
   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?
   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.

## Short Questions

1. What is the difference between qualitative and quantitative data?
2. Give an example of continuous data and explain why it is considered continuous.
3. Which method would you use to collect opinions from a large group of people about a new school policy?
4. What type of data is the number of students in your class?
5. Why is it important to organise data into tables or charts before analyzing it?

6. What is one advantage of using online tools like Google Forms for collecting survey data?
7. Why might you need to integrate data from different sources when working on a project?
8. Describe a scenario where discrete data might be more useful than continuous data.
9. Explain why data visualization is important. How does it help in understanding complex information?
10. Describe what a line graph is used for and provide an example of data that could be displayed using a line graph.
11. Explain the use of scatter plots in visualizing continuous data. Provide an example of a situation where a scatter plot would be useful.

## Long Questions

1. Explain the differences between qualitative and quantitative data. Provide examples of each type.
2. Describe the process of conducting a survey to gather data about students' favourite extracurricular activities.
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.
4. Analyse the benefits and challenges of using digital tools like Google Forms for data collection.
5. Imagine you are tasked with organizing a school event and need to collect data on students' preferences for activities and refreshments.
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.
7. Describe a situation where non-numeric data is essential. How would you collect, store, and analyse this type of data? Discuss the tools and techniques you would use to ensure the data is accurately interpreted.
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.
9. Discuss the importance and benefits of data visualization. Why is it essential for businesses and decision-makers to use charts, graphs, and dashboards?
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.