

taleemzone.com

CLASS 9TH • COMPUTER SCIENCE

6

Data Science and Data Gathering

Complete Notes

TOPICS COVERED:

1. Data and Data Types
2. Organising and Analysing Data
3. Data Tables, Charts, and Graphs
4. Data Collection Methods
5. Structured and Unstructured Data
6. Data Visualization
7. Data Pre-Processing
8. Data Validation and Cleaning
9. Cloud Storage and Data Backups
10. Introduction to Data Science
11. Tools in Data Science

INTRODUCTION

This unit 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. We shall learn methods for organising and analysing data, using both quantitative and qualitative techniques. We'll also look into data visualization, 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.

SECTION 1: DATA AND DATA TYPES**Q.1 What is Data? Explain its types.****Answer:****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.

Data Types

Data can be divided into two broad categories:

1. Qualitative Data

Qualitative data refers to categories or labels used to describe the qualities or characteristics of something rather than its quantity. This type of data offers a way to categorize and provide insights into opinions, behaviours, and experiences through descriptions rather than numbers.

Key Characteristics:

- **Non-Numeric:** Represented by words, labels, or symbols instead of numbers
- **Categorical:** Can be divided into categories or classes based on characteristics

EXAMPLE

- Names of students in a class (e.g., Arshad, Hina, Fatima)
- Types of fruit
- Colors, gender, nationality

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:

- **Numerical:** Expressed in numbers, representing a measurable quantity
- **Measurable:** Can be measured using instruments or tools
- **Countable:** Can be counted or enumerated
- **Arithmetical:** Can be used in arithmetic operations

EXAMPLE

- Heights, weights, and test scores
- Number of students, number of cars
- Temperature readings

SECTION 2: ORGANISING AND ANALYSING DATA

Q.2 Define Organising and analysing data. Also write the importance of Organising data.

Answer:**Organising and Analysing Data**

Organising 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

Reason	Explanation
Saves Time	Just like searching for a book in a messy room versus a neatly arranged bookshelf, organised data is easier to find and analyse.
Improves Clarity	When data is presented clearly, it's easier to understand and interpret.
Better Decision Making	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.

Q.3 What are data tables, charts, and graphs, and how do they help in organising and understanding data?**Answer:****Data Tables**

Data tables present information in rows and columns, making it easy to compare and analyse data.

Example: Students Data

Student	Math	Science	English
Arshad	85	78	90
Hina	78	88	85
Fatima	92	82	87
Faisal	90	80	89
Ayaan	67	75	70

Charts

Charts are visual representations of data designed to make complex information easier to understand. Charts help identify patterns, trends, and outliers in datasets.

Chart Type	Description	Use
Bar Chart	Uses rectangular bars to represent and compare values of different categories	Compare quantities (e.g., number of students in different classes)
Line Chart	Shows data points connected by straight lines	Show changes or trends over time
Pie Chart	Circular graph divided into slices representing portions of a whole	Show percentages or proportions of a total

Graphs

Graphs are visual tools used to represent data and show relationships between different data points.

Graph Type	Description	Example
Line Graph	Series of data points connected by straight lines	Temperature changes during the day, sales growth over months
Bar Graph	Uses rectangular bars to represent values	Number of students in different classes
Scatter Plot	Displays data as points representing relationship between two variables	Identifying patterns or correlations
Histogram	Type of bar graph showing distribution of numerical data	Data divided into intervals (bins)

SECTION 3: DATA COLLECTION

Q.4 What is data collection, and what are some common methods used to collect data?

Answer:**Data Collection**

Data collection is the process of gathering information to answer questions, make decisions, or understand something better.

Methods of Data Collection

Method	Description	Example
Surveys	Collect information from people by asking questions (paper, phone, online)	Survey about favorite ice cream flavours
Questionnaires	Written forms with a set of questions to be answered	School activity preference questionnaire
Interviews	One-on-one conversations to gather detailed information	Interviewing a teacher about their experience
Observations	Watching and noting what happens in a particular situation	Observing student behavior during group projects
Online Data Sources	Websites, databases, and digital tools	Researching statistics about pet ownership

Best Practices in Survey Design

Practice	Explanation
Be clear and specific	Use simple, unambiguous language
Keep it short and simple	Avoid unnecessary questions
Use multiple choice and rating scales	Makes responses easier to analyze
Ensure anonymity	Encourages honest responses
Test your survey	Ensure questions work properly and are easy to understand
Analyse the results	Draw useful conclusions from responses

TIDBITS

Question	Answer
Q. Why should you test your survey before sending it out?	To make sure the questions work properly and are easy to understand.
Q. Why should you analyse the survey results?	To understand the responses and draw useful conclusions.
Q. What are Google Forms?	A free tool from Google used to create surveys and collect responses online.
Q. What is Survey Monkey?	A popular tool used for creating detailed surveys.

SECTION 4: STRUCTURED AND UNSTRUCTURED DATA

Q.5 What are the two main types of data (with respect to storage and processing point of view), and how do structured and unstructured data differ?

Answer:

1. Structured Data

Structured data is organised and formatted to be easily searchable and analyzable.

Characteristics:

- Organized in a clear format (rows and columns)
- Easy to store, search, and analyse

Examples: Data in spreadsheets, traditional databases, names, dates, phone numbers

Example Table:

Student ID	Student Name	Class	Date of Birth	Fee Status	Height
001	Hina	9th	3/25/2009	Paid	4.7
002	Fatima	9th	5/7/2008	Paid	4.9
003	Aun Izhar	9th	6/11/2009	Unpaid	5.2
004	Ayaan	9th	9/13/2009	Paid	5.6
005	Areeda	9th	7/21/2009	Paid	5.3

2. Unstructured Data

Unstructured data is more free-form and doesn't fit into a specific format.

Characteristics:

- No predefined format
- Harder to organize
- Requires processing to extract meaning

Examples: Emails, social media posts, videos, images, text documents

Difference between Structured and Unstructured Data

Feature	Structured Data	Unstructured Data
Format	Organized in clear format (rows and columns)	No predefined format
Searchability	Easy to store, search, and analyse	Requires processing to extract meaning
Examples	Names, dates, phone numbers	Emails, videos, social media posts

SECTION 5: DATA VISUALIZATION

Q.6 What is data visualization, and which tools can be used to create charts and graphs for understanding data?

Answer:**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.

Data Visualization Tools

Tool	Description
Microsoft Excel	A spreadsheet tool that can create various charts and graphs
Google Sheets	Similar to Excel, allows you to create and share visualizations online
Tableau	A powerful tool specifically designed for creating detailed and interactive visualizations
Microsoft Power BI	Allows users to create a wide variety of visualizations including charts, graphs, maps, and more

SECTION 6: DATA PRE-PROCESSING

Q.7 What is data pre-processing, and why is it important to check for errors, outliers, and biases before analysing data?

Answer:**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.

1. Evaluating Data Quality

Before using data, we need to check its quality, ensuring it is accurate, complete, and reliable.

Questions to Ask:

- 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.

2. Identifying Errors, Outliers, and Biases

Term	Definition	Example
Errors	Mistakes in the data	A score recorded as 105 when maximum marks are 100
Outliers	Unusual or extreme values that don't fit the pattern	A score of 5 when most students scored between 50 and 80
Biases	Distortions that affect the accuracy of the data	Survey only includes answers from one school, not representing all students in the city

SECTION 7: DATA VALIDATION AND CLEANING

Q.8 What are data validation and cleaning, and how do quantitative and qualitative analysis help in understanding data?

Answer:

1. Data Validation and Cleaning

Data Validation involves checking data completeness and accuracy:

- **Completeness:** Ensuring no data is missing (e.g., every student has a test score recorded)
- **Accuracy:** Ensuring data is correct (e.g., verifying all test scores are between 0 and 100)

Data Cleaning involves:

- **Removing Errors:** Correcting incorrect data or deleting it
- **Handling Missing Data:** Deciding how to deal with gaps (e.g., using the average score to fill in missing values)

- **Dealing with Outliers:** Deciding whether to keep, change, or remove unusual values

2. Data Analysis Techniques

Data analysis involves examining data to discover useful information, draw conclusions, and support decision-making.

Types of Data Analysis

Analysis Type	Description	Focus
Quantitative Analysis	Deals with numbers and measurable data	Understanding patterns, relationships, and trends in numeric data
Qualitative Analysis	Deals with non-numeric data (text, images, sounds)	Understanding meanings, concepts, and experiences

SECTION 8: CLOUD STORAGE AND DATA BACKUPS

Q.9 What is cloud storage, and how do data backups and collaborative tools help in managing and protecting information?

Answer:

1. Cloud Storage

Cloud storage is a service model that stores data on remote servers accessed over the Internet, rather than on local disks.

Benefits:

- **Remote Access:** Ability to connect to and use data from any location
- **File Sharing:** Easy to share files with others
- **Accessibility:** Access files from any device with internet connection

DID YOU KNOW?**Q. What is cloud storage and why is it useful?**

Answer: Cloud storage allows you to store files online, like having an infinite bookshelf. It frees up physical space and keeps your documents, photos, and videos accessible from anywhere.

2. Data Backups

Data Backups are copies of important data or files stored separately from the original to protect against data loss.

Key Features:

- **Automatic Backups:** Automatically saves your files to prevent data loss
- **Restore Files:** Recovers deleted or previous versions of your files
- **Resume Work:** Accesses your files from anywhere to continue working seamlessly

Example: If you are working on a school project and save it on your computer, you create a backup by saving a copy on Google Drive or a USB to ensure you don't lose your work.

3. Benefits of Collaborative Tools

Benefit	Explanation
Enhanced Productivity	Multiple students can work on different sections simultaneously, saving time
Version Control	Google Docs automatically saves every change; you can go back to previous versions and see who made which changes

SECTION 9: INTRODUCTION TO DATA SCIENCE

Q.10 Define Data Science, and explain its interdisciplinary nature and importance in daily life.

Answer:**1. Data Science**

Data science is like being a detective, but instead of solving crimes, you solve problems using data. It combines computer skills, math, and business knowledge to turn raw data into useful insights, just like turning ingredients into a delicious recipe.

Example: You might 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. You might discover that students who study in groups tend to score higher.

2. Importance of Data Science

Knowing about data science is important because it helps us make better decisions in our daily lives:

Application	Example
Personal Decisions	Figure out the best time to study when you're most focused
Business	Learn what customers like and improve products
Sports	Create better strategies and enhance performance
Healthcare	Predict health trends and provide better care

3. Interdisciplinary Nature of Data Science

Data science combines different fields:

Field	Contribution
Computer Science	Handling and organizing data
Mathematics and Statistics	Analysing data and finding patterns
Business Knowledge	Applying insights to real-life problems and making informed decisions

SECTION 10: TOOLS IN DATA SCIENCE

Q.11 What are some common tools used in data science, and how do they help in managing, analysing, and visualizing data?

Answer:

Tools 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.

Tool	Description	Example
Microsoft Excel	Widely-used spreadsheet program for organizing and analyzing data	Create graphs showing test score progress over time

QUICK REFERENCE: DATA TYPES

Data Classification

text

DATA

|

+-----+-----+

||

QUALITATIVE DATA QUANTITATIVE DATA

(Categorical/ (Numerical/

Descriptive) Measurable)

||

Examples: Examples:

- Names - Heights
- Colors - Weights
- Fruit types - Test scores
- Gender - Number of students

Comparison Table

Feature	Qualitative Data	Quantitative Data
Nature	Non-numeric, descriptive	Numeric, measurable
Values	Categories or labels	Numbers
Questions Answered	"What type?" "Which category?"	"How much?" "How many?"
Examples	Names, colors, gender	Height, weight, scores
Analysis	Categorization	Mathematical calculations

SUMMARY

Topic	Key Points
Data	Raw facts collected; can be qualitative or quantitative
Qualitative Data	Non-numeric, descriptive, categorical
Quantitative Data	Numeric, measurable, countable, arithmetical
Data Collection	Surveys, questionnaires, interviews, observations, online sources
Structured Data	Organized in rows and columns; easy to search and analyse
Unstructured Data	Free-form; harder to organize; requires processing
Data Visualization	Turning data into pictures (charts, graphs) to understand patterns
Data Pre-Processing	Cleaning and organizing data before analysis
Cloud Storage	Storing data on remote servers accessed over the Internet
Data Science	Interdisciplinary field combining computer science, math, and business
Data Science Tools	Microsoft Excel, Google Sheets, Tableau, Power BI

End of Unit 6 Notes