Subject: Computer	**Chapter:** Unit 04
Class: 9th	**Teacher: Javaid Iqbal**
Innovative Higher Secondary School	

MCQS

1. Which field combines computer science, statistics and business knowledge?
a) History
b) Data Science □
c) Geography
d) Psychology

2. Data Analytics mainly focuses on:
a) Music composition
b) Analyzing data □
c) Painting
d) Sports

3. The term 'dataset' refers to:
a) Single number
b) Collection of related data □
c) A drawing
d) A file only

4. What is a table in a dataset called?
a) Plot
b) Relation □
c) Graph
d) Chart

5. Rows in a table represent:
a) Attributes
b) Records □
c) Columns
d) Headings

6. Columns in a dataset are called:
a) Tuples
b) Attributes □
c) Shapes
d) Links

7. Categorical attributes are also known as:
a) Numerical
b) Qualitative □
c) Quantitative
d) Discrete

8. Numeric attributes include:
a) Nominal
b) Ordinal
c) Interval and Ratio □
d) Binary only

9. Which of the following is a binary attribute?
a) Red, Blue
b) Male, Female □
c) High, Medium, Low
d) Temperature

10. Which data type deals with continuous values?
a) Ordinal
b) Ratio-scaled □
c) Nominal
d) Binary

11. The process of collecting data from different sources is called:
a) Coding
b) Data Collection □
c) Formatting
d) Distribution

12. Summary statistics are used to:
a) Decorate reports
b) Describe data □
c) Delete data
d) Encrypt data

13. Which one represents data in rows and columns?
a) Tree
b) Table □
c) Pie chart
d) Paragraph

14. Analytical thinking is improved by:
a) Reading comics
b) Reading case studies □
c) Watching cartoons
d) Playing games

15. Big data is characterized by:
a) Small size
b) Low value
c) Huge volume □
d) Single file

16. Which technology connects devices and shares data automatically?
a) IoT □
b) HTML

c) WordPad
d) Paint

17. One of the main uses of big data in business is:
a) Cooking food
b) Predicting trends □
c) Printing books
d) Driving cars

18. Predictive modeling helps in:
a) Predicting future outcomes □
b) Singing songs
c) Drawing art
d) Making furniture

19. Which of these tools is used in big data?
a) MS Paint
b) Hadoop □
c) Notepad
d) PowerPoint

20. What does "structured data" mean?
a) Random text
b) Organized in rows and columns □
c) Sound files
d) Video clips

21. The term "unstructured data" refers to:
a) Table format
b) Image, audio, text □
c) CSV file
d) Excel sheet

22. Students learn about frequency tables to:
a) Play games
b) Interpret data □
c) Send emails
d) Write poetry

23. An ordinal attribute example:
a) A, B, C grades □
b) Height in cm
c) Gender
d) Country name

24. Continuous data can take:
a) Only whole numbers
b) Any value in a range □
c) Only text
d) Only zero

25. Ratio-scaled data has:
a) No true zero
b) A true zero point □
c) Only negative values
d) Only positive odd numbers

26. A dataset missing some entries is said to have:
a) Perfect records
b) Missing data □
c) Binary data
d) Nominal values

27. Business knowledge helps a data scientist to:
a) Understand business problems □
b) Dance better
c) Paint faster
d) Row a boat

28. The concept of IoT is related to:
a) Devices connected via internet □
b) Offline books
c) Cooking gas pipeline
d) Old age homes

29. Which one is a qualitative attribute?
a) Salary
b) Age
c) Color □
d) Speed

30. The cloud is used in big data for:
a) Storing and processing data □
b) Washing clothes
c) Cooking rice
d) Clicking selfies

31. In a dataset, each column is a:
a) Record
b) Attribute □
c) Row
d) Relation

32. Students read published stories to:
a) Ignore data
b) Critique visualizations □
c) Sleep
d) Dance

Exercise MCQs with Correct Answers

Q1. _____ is a structured or processed collection of data usually associated with a unique body of work.
a) Database
b) Dataset □
c) Data and Information
d) Information

Q2. _____ refers to the process of carefully examining and studying data to identify patterns, draw conclusions, or make the data meaningful.
a) Data analytics □
b) Data Predictions
c) Dataset
d) Database

Q3. _____ is the graphical representation of data through use of common charts, plots, infographics, and animations.
a) Data cleaning
b) Missing values
c) Data visualization □
d) Data hiding

Q4. _____ is subset of Machine Learning, with emphasis on the simulation or imitation of human brain's behavior by using artificial neural networks.
a) Data visualization
b) Computer vision
c) Deep learning □
d) Big Data

Q5. _____ is the use of data to predict future trends and events based on historical data.
a) Statistical analysis
b) Predictive analysis □
c) Graphical analysis
d) Deep learning

Q6. _____ is the fast rate at which data is received and acted on.
a) Volume
b) Velocity □
c) Variety
d) Vision

Q7. _____ includes the data which can only take certain values and cannot be further subdivided into smaller units.
a) Discrete data □
b) Continuous data
c) Ordinal data
d) Referral data

Q8. _____ is limitation of big data.
a) Statistical data
b) Unlimited growth of data □
c) Data visualization
d) Predictive maintenance

Q9. Customer satisfaction level such as satisfied, dissatisfied, and neutral are examples of _____ data type.
a) Ordinal data □
b) Continuous data
c) Numeric data
d) Discrete data

Q10. _____ is a method of collecting information from individuals.
a) Survey □
b) Data hiding
c) Data visualization
d) Data finding

Short Questions

Q1. Define Data Science.

Ans:

Data Science is a field that combines computer science, statistics, mathematics and business understanding to study data. Its purpose is to extract useful knowledge and make decisions based on data. It helps businesses and organizations make better decisions. Data Science uses various tools, software, and algorithms to discover patterns and predictions from data.

Q2. What is a Dataset?

Ans:

A dataset is a collection of related information or data which is arranged in a table form. It consists of rows and columns where each row is a record and each column is a property or attribute. A dataset is used to store observations. For example, a students' marks table is a dataset that stores marks of students.

Q3. Difference between Data and Information.

Ans:

Data refers to raw facts and figures that have no specific meaning. Information is data that has been processed or organized into a meaningful form. For example, marks written randomly are data while the average result prepared from those marks is information. Information helps in decision-making while data alone may not.

Q4. Write two sources of data.

Ans:

Data can be collected from many sources. Two common sources are sensors and online surveys. Sensors automatically collect temperature, speed, etc., while online surveys collect user opinions. These sources help collect real world data for analysis in data science.

Q5. What is Categorical data?

Ans:

Categorical data, also called qualitative data, refers to non-numeric data that describes qualities or

characteristics. Examples include gender, color, or blood group. This type of data is not measured with numbers but grouped in categories. Categorical data helps classify objects.

Q6. What is an Ordinal attribute?
Ans:
An ordinal attribute is a type of qualitative attribute where the values have a specific order or ranking. For example, size can be small, medium, or large. Similarly, grade A is higher than grade B. Although the order is known, exact difference between values is not measured.

Q7. Define Binary attribute.
Ans:
A binary attribute is a qualitative attribute that has only two possible values. Common examples are yes/no, male/female, true/false. Binary attributes are useful in classification because the answer is always one of two possible categories only.

Q8. Define Numeric data.
Ans:
Numeric data refers to data that is expressed in numbers and used for counting or measuring. Numeric data can be further divided into discrete and continuous data. Examples include height, age, weight, salary, etc. It is quantitative in nature and is used in calculations and statistics.

Q9. What is a Frequency table?
Ans:
A frequency table is a table that shows how many times each value appears in a dataset. It summarizes data by listing every distinct value and its count. It helps in understanding distribution of data and is used in data analysis to identify patterns and repetition.

Q10. What is Big Data?
Ans:
Big data refers to extremely large and complex datasets that cannot be handled by traditional database tools. It contains huge volume, high speed (velocity), and variety of data. It is generated from sources like social media, sensors, websites, etc. Big data helps companies make big decisions and predict future trends.

Q11. Mention one characteristic of Big Data and explain.
Ans:
One important characteristic is Volume, which means the large amount of data collected from various sources. This data can be in terabytes or even petabytes and is too large to store in regular computers. This characteristic makes big data valuable but also difficult to process.

Q12. What is Predictive Modeling?
Ans:
Predictive modeling is a process in which data is used to make predictions about future outcomes. It uses historical data and mathematical techniques to build a model that can forecast results. Businesses use predictive modeling to predict trends, sales, customer behavior, etc.

Q13. What is Internet of Things (IoT)?
Ans:
The Internet of Things (IoT) is a network of physical devices connected to the internet that collect and exchange data automatically. These devices include smart watches, sensors, smart home appliances, etc. They gather real-time data and send it for processing and monitoring.

Q14. Why is business knowledge important in Data Science?
Ans:
Business knowledge helps data scientists understand real-world problems faced by companies. It allows them to apply data analysis in a meaningful way that benefits the organization. Without business understanding, the analysis may not provide practical solutions. Therefore, bringing value to business is an important goal of Data Science.

Q15. What is an Attribute in a dataset?
Ans:
An attribute is a column in a table that represents a specific characteristic or property of data. For example, in a student table, attributes can be name, roll number, marks, etc. Each attribute gives information about the data stored in that specific column.

Q16. Define Ratio-scaled data.
Ans:
Ratio-scaled data is numeric data that has a true zero point and differences between values are meaningful. Examples include height, weight, age or income. If someone has zero weight or zero money it means nothing present. This type of data allows comparison like double, half, etc.

Q17. Define Interval-scaled data.
Ans:
Interval-scaled data is numeric data where the difference between values is meaningful but zero does not mean absence. An example is temperature in Celsius. The difference between 30°C and 20°C is 10 degrees, but 0°C does not mean no temperature. It allows addition and subtraction.

Q18. What is Data Visualization?
Ans:
Data visualization is the presentation of data in graphical format like charts, graphs and plots. It helps to understand data patterns and trends easily. Visual tools like bar charts, pie charts, and histograms make data easier to interpret than looking at raw numbers.

Q19. Explain Data Collection.
Ans:
Data Collection is the process of gathering data from different sources like surveys, experiments, or digital devices. It is the first step in data analysis. Good data collection ensures that information is accurate, complete and reliable for analysis.

Q20. What are Summary Statistics?
Ans:
Summary statistics are numbers like mean, median, mode, range used to describe and summarize a dataset. They give a quick overview of the data and help understand central tendency and spread. They are useful in making comparisons among different datasets.

Q21. What is Data Analytics?
Ans:
Data Analytics is the process of analyzing data sets to find patterns and draw conclusions. It helps in decision making by understanding trends and relations among data. Businesses use data analytics to improve performance and gain insights.

Q22. Difference between Structured and Unstructured data?
Ans:
Structured data is well organized in rows and columns such as tables in databases. Unstructured data has no fixed format such as text, images, audio, and videos. Structured data is easy to search and analyze while unstructured data requires special tools to interpret.

Q23. Role of Computer Science in Data Science.
Ans:
Computer Science provides the programming, algorithms, and storage technologies that are needed to process big data. It helps build software and applications that analyze and visualize data. Data scientists use tools like Python, R, databases, etc., which come from computer science.

Q24. Name one Big Data tool and explain its use.
Ans:
Hadoop is a popular big data tool used to store and process large datasets across many computers. It breaks big

data into small parts and processes them in parallel, making it faster and efficient. It helps companies handle big data that cannot fit into one machine.

Q25. What is Data Storage?
Ans:
Data storage means saving data in devices or cloud systems for future use. It can be stored in hard drives, servers or cloud platforms. Storing data properly ensures that it is safe and can be accessed whenever needed for analysis or reporting.

Q26. Define Discrete data.
Ans:
Discrete data is numeric data that can take only specific whole number values. For example, number of students in a class, number of cars, or number of books can only be whole numbers. Discrete data cannot have fractions or decimals.

Q27. What is Analytical Thinking?
Ans:
Analytical thinking is a skill where a person breaks down a problem into smaller parts, understands relationships, and finds solutions by logic. In data science, analytic thinking helps to study data carefully and make correct conclusions.

Q28. What does Business Knowledge help with in Data Science?
Ans:
Business knowledge helps a data scientist connect data analysis with real business needs. It guides them in making decisions that increase profit or improve services. Without business understanding, data results may not be useful for an organization.

Q29. Mention any one application of Big Data.
Ans:
One application of big data is in marketing where companies study customer behavior and preferences. By analyzing large amounts of customer data, companies can plan better advertisements and increase sales.

Q30. What is Cloud Computing?
Ans:
Cloud computing refers to the use of remote servers available on the internet to store, manage and process data. Users can access data anytime from anywhere without physical storage devices. It helps in handling large data and provides scalability and backup.

EXERCISE Short Question Answers – (2 Marks each)

1) Define data analytics and data science. Are they similar or different? Give reason.
Answer:
Data analytics is the process of examining data to find useful patterns, trends or conclusions. It helps answer specific questions using statistics and visual tools. Data science is a broader field that includes data analytics plus use of computer programming, machine learning and business knowledge to generate insights and predictions. They are related but data science is wider in scope than data analytics.

---

2) Can you relate how data science is helpful in solving business problems?
Answer:
Data science helps businesses by analyzing customer data, predicting sales and improving services. For example, by studying buying patterns, companies can launch targeted advertisements to increase profit. It also helps in detecting fraud, reducing cost and making better decisions. Thus, data science solves real-world business problems through data-based solutions.

---

3) Database is useful in the field of data science. Defend this statement.
Answer:
A database stores large amounts of data in an organized way so it can be easily accessed and managed. In data science, we need to process and analyze massive data smoothly. Database systems help in storing data securely, retrieving it fast, and updating it when needed. Without databases it would be almost impossible to handle big datasets for data science projects.

---

4) Compare machine learning and deep learning, in the context of formal & informal education.
Answer:
Machine learning teaches a computer using simple data and rules, like basic pattern recognition. Deep learning uses layered artificial neural networks to learn complex tasks just like the human brain. In formal education, machine learning examples include grade prediction or attendance monitoring. In informal settings, deep learning is used in facial recognition or personalized learning apps that adapt to student behavior.

---

5) What is meant by sources of data? Give three sources of data excluding those mentioned in the book.
Answer:
Sources of data are the places or methods from which data is collected for analysis. Besides book examples, three additional sources are:
• Mobile apps usage data
• Social media platforms (like Twitter, Facebook)
• Online transaction records (such as shopping websites).
These sources collect useful data for different industries.

---

6) Differentiate between database and dataset.
Answer:
A database is a structured collection of data stored in tables that can be managed and updated using special

software. It may contain many datasets and supports queries and transactions. A dataset is a single collection of related data usually organized in one table or sheet. Every dataset can be stored in a database but a database is a larger system overall.

---

7) Argue about the trends, outliers, and distribution of values in a data set? Describe.
Answer:
When analyzing a dataset, we often look at trends, which show how data changes over time, like sales increasing monthly. Outliers are unusual values far away from others, such as a very high or very low number, and can affect results. Distribution shows how data is spread out: whether values are crowded in one area or spread evenly. These features help understand the overall behavior of data.

---

8) Why are summary statistics needed?
Answer:
Summary statistics like mean, median, and mode help us quickly understand a dataset without checking each value. They show the central tendency and spread of data. For example, the average helps to compare performance of two classes. These statistics allow us to describe the data in a simple way to make decisions or identify issues.

---

9) Express big data in your own words. Explain three V's of big data with reference to email data.
Answer:
Big data refers to a huge amount of data that is difficult to process using normal tools. In an email inbox with hundreds of emails:
• Volume: A large number of emails store texts, pictures, files.
• Velocity: Emails arrive rapidly every minute.
• Variety: Emails have different types of data: text messages, attachments, spam etc. So, email inbox is a good example of big data.

---

10) Illustrate the purpose of data storage?
Answer:
Data storage means saving data in a safe place so it can be used in the future. It ensures that important information is not lost and can be accessed when needed for further analysis or reporting. Without proper storage, data could be lost or damaged. Data storage also helps in organizing and sharing data in big projects and businesses.

□ Long Questions Answers (Proper 6 Marks)

Q1: Explain the scope of data science as an interdisciplinary field.

Answer:
Data science is an interdisciplinary field because it combines multiple areas like computer science, mathematics, statistics, and business understanding. Computer science provides programming languages and algorithms for data processing. Mathematics and statistics help analyze data and find trends or patterns. Business knowledge guides the actual use of this data in solving real problems in companies. Data science is used in healthcare, finance, education, marketing, and many other areas to make better decisions. It helps organizations to make predictions, automate tasks, and improve performance. Thus, it covers a vast scope across both science and business sectors.

Q2: Differentiate between Data Science, Data Analytics, and Computer Science.

Answer:
Data Science is the broad field that deals with collecting, processing, analyzing, and interpreting large amounts of data to gain insights. It involves machine learning, artificial intelligence, statistics, and business knowledge.
Data Analytics is a part of data science that focuses mainly on analyzing data to find specific patterns or insights to solve immediate problems. It involves statistical techniques and visualization for decision-making.
Computer Science, on the other hand, is a field that deals with the theory and development of software and systems. It includes programming, algorithms, networking, database systems, etc.
In short, computer science provides the tools and technical foundation, data analytics performs analysis, and data science uses both to generate solutions and predictions.

Q3: Describe the various types of attributes with examples.

Answer:
Attributes are the characteristics or properties of data in a dataset. They can be divided into two main types: qualitative (categorical) and quantitative (numeric). Categorical attributes describe qualities and do not involve numbers. Examples include color (red, blue), gender (male, female) and city names.
Categorical attributes are further divided into: Nominal (no order, e.g., blood group), Binary (two values, e.g., Yes/No), and Ordinal (ordered, e.g., small, medium, large).
Numeric attributes involve numbers and can be counted or measured. They include discrete data (e.g., number of students) and continuous data (e.g., height, weight). Numeric attributes also include interval-scaled (e.g., temperature in Celsius) and ratio-scaled (e.g., age, salary). These attributes help classify different types of data in datasets for analysis.

Q4: Explain structured vs unstructured data with examples.

Answer:
Structured data is organized in a predefined format such as rows and columns in a table, making it easy to search, store, and analyze. It follows a fixed structure like spreadsheets or relational databases. Examples include student records, bank transaction tables, and inventory lists.
Unstructured data has no specific format or sequence. It can consist of text, images, audio, video or social media posts. This data is harder to organize and needs special software or tools for analysis. Examples include emails, WhatsApp messages, tweets, and photos.
Structured data provides quick analysis using simple tools, whereas unstructured data requires advanced tools like natural language processing or image recognition. In modern systems, both types of data are important to understand user behavior and make effective decisions.

Q5: Discuss the characteristics and importance of Big Data.

Answer:
Big Data refers to extremely large datasets that are difficult to handle with traditional data processing tools. Its main characteristics are Volume (huge amount of data), Velocity (high speed at which data is generated), and Variety (different types of data like text, images, audio). Some also include Veracity (accuracy of data) and Value (usefulness of data).
Big data is important because it helps organizations get deeper insights and make informed decisions. For example, companies use big data to analyze customer behavior, predict market trends, personalize advertisements, and improve services. It is also used in healthcare to monitor patient data, in education to track student performance, and in government planning. Thus, big data is valuable for improving efficiency, saving costs, and predicting future outcomes.

Q6: Explain data collection methods and importance of data storage.

Answer:
Data collection means gathering information from various sources such as surveys, questionnaires, sensors, experiments, and online forms. It is the first important step before any data analysis. Good data collection ensures that data is accurate, complete, and reliable. The collected data can be numeric or categorical depending on the requirement.
After data is collected, it must be stored safely for future use. Data storage involves saving data on computer systems, databases, USB drives or cloud storage. Proper storage ensures that data is secure from loss or damage and can be easily accessed when needed for analysis. Without proper storage, collected data may be useless or lost. Data storage also allows us to share and reuse collected data for other projects.

Q7: Describe predictive modeling and its use in businesses.

Answer:
Predictive modeling is the process of using historical data and statistical techniques to make predictions about future events. It uses machine learning algorithms to create a model that can identify patterns and trends in the data. Once the model is built, it can forecast future outcomes such as customer demand or sales growth.
Businesses use predictive modeling to improve their decisions and strategies. For example, it helps them understand customer behavior, predict which products will sell, detect fraud, and reduce risk. By predicting future trends, companies can plan better and gain a competitive advantage. Predictive modeling helps businesses save money, increase profit, and make smarter decisions.

Q8: Explain how summary statistics help in understanding data.

Answer:
Summary statistics are simple numerical measures that help provide a quick overview of a dataset. They include measures like mean (average), median, mode, range, variance and standard deviation. These values help us understand the central location of the data, its spread and variability.
For example, the average marks of a class give an idea of the performance of all students. Range tells us the difference between the highest and lowest marks. Using these statistics, we can identify if the data is consistent or has a lot of variation. Summary statistics are important because they allow quick comparison between different sets of data and help in making decisions without looking at all individual values.

Q9: Write a note on IoT and its connection with Big Data.

Answer:
The Internet of Things (IoT) is a system in which physical devices like sensors, home appliances, and machines are connected to the internet and can exchange data automatically. These devices collect data such as temperature, motion, heart rate, etc., and send it to servers for analysis.
IoT generates a large amount of data continuously, which becomes part of Big Data. This data can be analyzed to monitor systems in real-time, improve efficiency, reduce cost and increase safety. For example, smart homes, smart watches and automatic vehicles are all part of IoT that produce big data. Thus, IoT and Big Data work together: IoT collects the data and Big Data techniques analyze and store it for useful insights.

Q10: Explain the role of business knowledge in Data Science.

Answer:
Business knowledge is a very important component of data science because it helps understand the real-world problems that organizations face. A data scientist with business knowledge can analyze data in a way that provides valuable solutions that the company can use. Business understanding guides which questions to ask, what data to collect and how to interpret the results for practical use.
For example, if a company wants to increase sales, a data scientist must understand marketing strategies and customer behavior. This helps create models to predict what customers want and how to sell better. Without business knowledge, even accurate data analysis may not be useful for decision-making. Therefore, data science needs both technical skills and business understanding to create real value.

Q11: Discuss numeric data types: interval, ratio, discrete, continuous.

Answer:
Numeric data types are used for data that is expressed in numbers. There are four common types: interval, ratio, discrete, and continuous. Interval data includes values where the difference between them matters but zero has no true meaning, such as temperature in Celsius. Ratio data has a meaningful zero point and allows comparison like double or half, such as age, weight or salary.
Discrete data can only take whole number values, like number of books or students. Continuous data can take any value within a range, such as height or speed. These types help categorize numeric data and are important in data analysis and statistics for correct interpretation and calculations.

Q12: Explain how data is converted into information and knowledge.

Answer:
Data is the raw collection of facts and figures that has no proper meaning. When data is processed, organized or analyzed to make it meaningful, it becomes information. For example, a list of sales numbers is data, but calculating the total sales gives us information.
When information is linked with experience, context and understanding, it becomes knowledge. Knowledge allows us to make decisions and predictions. In data science, raw data is processed into useful information, and that information helps experts to make knowledgeable decisions for solving real-world problems.

Q13: Write advantages and applications of big data in real-world.

Answer:
Big Data has many advantages because it allows organizations to analyze huge amounts of information quickly. Businesses can identify customer behavior, improve marketing strategies, prevent fraud, and increase profits. Healthcare uses big data to monitor patient symptoms, improve treatments, and predict disease outbreaks.
In education, big data helps track student performance and learning patterns. In transportation, it analyzes traffic data to improve routes. Governments use big data for planning, safety and predicting natural disasters. Thus, big data provides better decision-making, cost saving and innovation in different real-world applications.

Q14: Define dataset and explain rows, columns, attributes with example.

Answer:
A dataset is a collection of related data organized in a table or file. It contains rows and columns. Each row in a dataset is called a record or tuple and represents one object or instance. Each column is known as an attribute or field which represents a specific property of the data.
For example, in a dataset of students, the rows may represent individual students, and the columns may represent attributes like name, age, class, and marks. The complete set of these rows and columns makes a dataset. This organized structure helps in easy storage, analysis and retrieval of information.

Q15: Compare categorical and numeric data with examples.

Answer:
Categorical data (qualitative data) consists of labels or names used to identify categories or groups. It is not measured in numbers. Examples include gender (male/female), colors (red/blue), or blood groups. Categorical data can be nominal (no order) or ordinal (with order, like small, medium, large).
Numeric data (quantitative data), on the other hand, includes numbers that are used to count or measure. It can be discrete (whole numbers like number of students) or continuous (any value in a range such as height or weight). Numeric data can be analyzed using mathematical operations. Thus, both types are used to represent different forms of information in datasets.

Exercise LONG Questions Answers (6 Marks Type)

---

1) Sketch the key concepts of data science in your own words.

Answer:
Data science is the study of data to extract meaningful insights for solving problems. It combines various disciplines such as computer science, statistics, mathematics, and business knowledge. The key concepts include data collection, where raw data is gathered; data cleaning, which removes errors; and data analysis, where useful patterns are found. It also includes data visualization to present results in graphs and charts. A major concept is predictive modeling which helps in forecasting future events using historical data. All these concepts are linked together to turn raw data into actionable knowledge for decision-making.

---

2) Develop your own thinking on the various data types used in data science.

Answer:
In data science, understanding data types is very important because it defines how data will be analyzed. There are two main types: qualitative (categorical) and quantitative (numeric). Categorical data tells about categories like gender, color, or yes/no. Nominal data has no order while ordinal data has a natural order like small, medium, large. Numeric data can be discrete, which is whole numbers like number of students, or continuous like weight or temperature. Interval data like Celsius has no true zero but difference is important, whereas ratio data has a true zero like age or price. Proper knowledge of these types helps us choose correct statistical methods in analysis.

---

3) Compare how big data is applicable to various fields of life. Illustrate your answer with suitable examples.

Answer:
Big data is used in many fields due to its ability to process large and complex data to make better decisions. In healthcare, big data helps doctors identify disease patterns and provide personalized treatment to patients. In business, companies study large customer data to predict demand and improve sales. In education, it helps to track student performance and design better teaching methods. Government uses big data for traffic analysis and weather forecasting. For example, online platforms like YouTube use big data to recommend videos based on previous watch history. Thus, big data plays a major role in daily life by improving services and decision-making in multiple sectors.

---

4) Relate the advantages and challenges of big data?

Answer:
Big data offers many advantages such as helping organizations make accurate decisions by analyzing massive datasets. It improves customer service, increases profit, reduces cost, and supports innovation. Big data also helps in predicting trends and detecting fraud. However, big data also has challenges. Handling its large volume and variety is difficult without advanced tools. There are privacy concerns as sensitive information about users is stored. It also needs skilled professionals for analysis and huge storage systems. So while big data brings many benefits, it also creates challenges such as complexity, data security and high cost of processing systems.

---

5) Design a case study about how data science and big data has revolutionized the field of healthcare.

Answer:
In healthcare, data science and big data have transformed the way hospitals and doctors treat patients. For example, a hospital gathers large amounts of data from patient records, lab tests and medical devices. Using big data tools, doctors can identify patterns to predict diseases early, such as risk of heart attack based on patient history. Through machine learning models, personalized treatment plans are made for each patient. In pandemic situations like COVID-19, big data helped track spread, monitor patients and manage hospital resources. Health apps also use data science to monitor heart rate and physical activities. Thus, big data has improved diagnosis, treatment and saved many lives by providing smarter healthcare systems.