Class 11

Answer key

Part-A: Employability Skills

DATA SCIENCE PART-A

1. Communication Skills-III

Unsolved Questions

Section A (Objective Type Questions)

A. 1. a. 2. b. 3. c.
4. c. 5. d.

B. 1. assertive 2. paragraph 3. Visual
4. open-ended, closed-ended 5. Possessive

C. 1. False 2. True 3. False 4. False 5. True

Section B (Subjective Type Questions)

A. 1. The four types of sentences are:

i. **Declarative Sentence:** Used to state facts or opinions. Example: We own a parrot.

ii. **Imperative Sentence:** Used to give commands or instructions. Example: Wash your dirty clothes.

iii. **Interrogative Sentence:** Used to ask questions. Example: Have you read this book?

iv. **Exclamatory Sentence:** Used to express strong emotions. Example: Alas! We lost the game.

2. Assertive people have more effective communication because they express their opinions clearly without ignoring others, give due respect to others, accept different opinions, and maintain a balance between their needs and others' needs. They use positive language, seek positive results, and foster an atmosphere of mutual respect, which leads to better and more meaningful interactions.

3. **Formal Greeting**

a. Formal greeting is used if you do not know the person. It is used to greet an elder, elderly person, or people with whom we have a formal relationship such as a teacher or a client.

b. This is used more often in schools, colleges, offices, meetings, and at work.

c. Example: Good morning. How are you?

Informal Greeting

a. Informal greetings are used when you are talking to friends, family or when you know the person very well.

b. Used at home, in parties.

c. Example: Hi! What's up?

4. **Verbal Communication**

   Verbal communication is the exchange of information using words. This is what most people use as a communication method. Verbal communication is important. If you don't use the right words, you'll be confused and you won't be able to convey your message.

   Non-Verbal Communication

   Nonverbal communication is the transmission of a message without using words. We send signals and messages to others through our expressions, gestures, and body posture.

5. Three factors that affect communication are:

   i. Emotions and Feelings

   ii. Educational Qualification

   iii. Working Environment

6. Phonetics is the study of human sounds and phonology is the systematic classification of sounds in one or more specific languages. Sounds can be divided into consonants and vowels. The first can be characterized by 1) place, 2) mode of pronunciation and 3) voice. The English alphabet has 26 letters from A to Z. However, each of these letters can be pronounced differently in different words. We write words in English language using 26 letters. However, while speaking English, more than 26 sounds are used. For this reason, the spelling of a word does not always match its pronunciation.

   Let us understand the sounds we use. All English words are made of three basic types of sounds—vowels, Diphthongs (combination sound of two vowels) and consonants.

7. Passive communication involves not expressing one's opinions openly to avoid conflict. Passive communicators often stay quiet even when criticised, fail to assert their needs, and exhibit uncomfortable body language. This style can lead to frustration, anxiety, and lack of confidence over time.

B. Kanchan's communication style seems to involve passive-aggressive tendencies. Passive-aggressive communication is characterized by indirect expression of hostility, resentment, and frustration. In Kanchan's case, her sweet exterior to everyone may be a passive way of interacting, but internally resenting and muttering comments about classmates, telling lies, and not taking responsibility indicate a passive-aggressive approach. This style often involves avoiding direct confrontation but expressing negative feelings in an indirect or subtle manner.

Lab Activity

Do it yourself.

2. Self-Management Skills-III

Unsolved Questions

Section A (Objective Type Questions)

A. 1. a. 2. d. 3. c. 4. b. 5. c.

B. 1. Personal Hygiene 2. Time-bound 3. Listening
4. belief 5. external motivation

C. 1. True 2. False 3. True 4. True 5. False

Section B (Subjective Type Questions)

A. 1. Specific, Measurable, Achievable, Relevant, and Time-Bound are the acronyms for SMART goals. Defining these factors in relation to your goal helps guarantee that your objectives are met within a reasonable time frame. This method avoids generalisations and guessing, establishes a defined timeframe, and makes tracking progress and identifying missed milestones easier.

2. Communicate regularly, Show interest in getting to know your contacts. Listen to them patiently and share information if required.

3. Four qualities of self-motivated people are:
   - Know what they want from life
   - Focused
   - Understand what is important
   - want to fulfil their dreams

4. **Opinions** - It means the outlooks and judgments formed about something or someone, and this may not necessarily be based on actual facts. For example, opinion about a person's pet choices.

   Values - Values are the principles or standards of action; your own judgment about what is important in your life. They have a great influence on a person's behaviour and attitude. They also serve as a general guideline in all situations. Some good values are honesty, equality, community service, etc.

5. Interests are the things we love to do. Interests could be
   - activities that you enjoy doing in your spare time that make you happy.
   - activities that interest you or you would even do if no one asked you.
   - activities you want to learn or want to do in the future.

Sometimes interests may not match abilities. In these cases, you can improve your skills or go a different route. For example, you may like badminton (interest), but you may not be good at the game. In this case, you can play just for fun, not aim for a badminton career.

6. A team is a group of people who work together to achieve a common goal. Each team has goals to achieve. The process of working together in a group is teamwork. Everyone on the team plays an important role. Each individual on the team is called a team member. Each team member needs to understand his/her role in the team.

   The benefits of teamwork are:

   - There is no individual pressure to succeed because everyone has a role to play in the team
   - All members have a support system to help correct mistakes made by any team member
   - You feel good when the team succeeds and this helps to build your confidence

7. Knowing what you are good at and what doesn't work for you, can help you turn your weaknesses into strengths and your strengths into extraordinary talents. Analysing your strengths and weaknesses is helpful. An analysis of strengths and weaknesses begins with knowing and understanding yourself first.

   To identify your strengths and weaknesses, think about the activities you enjoy/dislike doing.

   Find your strengths

   - Spend time thinking about what you do well
   - Think about what you've always done well
   - Think about what others appreciate about you

   Find your weaknesses

   - Identify areas where you have difficulty and things you find difficult to do
   - Critically analyse the comments you receive about yourself from other people
   - Be open to feedback and accept your weakness without feeling bad. Think of it as an area for improvement.

B. This is called Self-reflection or Self-Inquiry.

Lab Activity

Do it yourself.

3. ICT Skills-III

Unsolved Questions

Section A (Objective Type Questions)

A. 1. b. 2. c. 3. d. 4. a. 5. b.

B. 1. Print preview 2. Edit Menu 3. tooltip 4. status 5. Ctrl+U

C. 1. True 2. True 3. False 4. False 5. False

Section B (Subjective Type Questions)

A. 1. Tracking changes in an OpenOffice document means keeping records of the changes made to the document. Example: A student creates a report and gives it to his teacher for review. Just as a teacher modifies a printed copy with a red pen, the teacher can make changes to a document using the Track Changes mode. This highlights the changes and makes it easier for students to identify and correct mistakes later.

2. Text formatting means changing the appearance of text. Different parts of the text may appear in different styles and sizes. You can bold, underline, or highlight the text to make it visually appealing.

3. A word processor is a software application that helps you create documents by entering, modifying (editing text), arranging (formatting), and printing text.

4. The Print Preview feature allows the user to see how the document will look after printing.

5. The Standard toolbar can be found directly behind the menu bar. It has buttons that symbolise operations like New, Open, Save, and Print.

   By default, the Formatting toolbar is situated next to the Standard toolbar. It has buttons that symbolise text-editing functions including font, text size, bold, numbering, and bullets.

6. The menu bar is at the top of a document window. When you click on any menu option, a submenu with commands will open. Let us understand the functions of the options in the menu bar.

   - File – contains commands like New, Open, Save, and Print which apply to the entire document.
   - Edit – contains document editing tools like Cut, Copy, Paste, Find & Replace etc.
   - View - contains commands like Zoom in/out, Web, Status Bar, Ruler, etc. that are used to change the display of a document.
   - Insert: Contains commands to add and insert various objects into the document. For example images, shapes, diagrams, headers, footers, page numbers, and so on.
   - Format: Contains commands to change the appearance of text.
   - Style: Contains commands that determine the general style of the document. You can choose a new style for the document from here.
   - Table: Contains commands to insert or delete tables in your document and format them.

7. Text, like files and directories, can be moved in a variety of ways once it has been selected: Ctrl+X to cut, Ctrl+V to transfer your cursor to the new location of text, and Ctrl+V to paste. Copy with Ctrl+C, move your mouse to the new location of text, and then paste with Ctrl+V. Right-click the selected text and choose Copy or Cut, then right-click the new text place and choose Paste.

You may shift text between pages, paragraphs, and even applications or windows.

B. Shruti can use the Find and Replace feature to make her work easier and faster

Lab Activity

Do it yourself.

4. Entrepreneurial Skills-III

Unsolved Questions

Section A (Objective Type Questions)

A.	1.	c.	2.	a.	3.	b.	4.	d.	5. b.
B.	1.	Services	2.	amount	3.	Perseverance	4.	feedback	5. Scale-up
C.	1.	True	2.	True	3.	False	4.	True	5. False

Section B (Subjective Type Questions)

A. 1.
- Concept: What should my business concept be? How do I know this concept will work?
- Capital: How much capital need be raised to establish a business?
- Obtaining materials: Where should the materials be obtained and at what cost?
- Manufacturing: How do you manufacture (if you are in the product business)?
- Pricing: At what price should I sell the goods in order to generate a sufficient profit? How can I find out how much the consumer is willing to pay?
- Marketing and advertising: How would I inform my customers about my business in order to get them to come to me?

2. Entrepreneurship is the process of coming up with new methods to combine resources. When the market value generated by this novel combination of resources exceeds the market value generated by these resources individually or in other combinations, the entrepreneur makes a profit.

   Four conditions must be met in order to be considered for the possibility to become an entrepreneur. First, there must be chances or situations in which resources can be recombined to generate value. Second, entrepreneurship necessitates individual distinctions, such as preferred access to specific persons or the capacity to identify information about opportunities. Third, taking risks is essential. Fourth, the entrepreneurial process necessitates the management of people and resources.

3. Countless entrepreneurs have taken risks to develop their enterprises to this point. Taking risks, on the other hand, does not imply going into company blindly and expecting huge

outcomes. Instead, handling risks successfully necessitates meticulous planning and an underlying strategy.

The true reasons why entrepreneurs opt to take risks may be boiled down to the five factors listed below.

- You will never know unless you try.
- Taking risks educates you.
- Risk taking is linked to innovation and opportunity.
- Those that are willing to take risks already have a competitive advantage.
- Risk-takers may be happy and satisfied with their lives.

4. Planning is fundamental to business growth. It helps to create a vision for a well-established business. A successful business requires strategy to please its customers and excel despite cut throat competition. Let us now learn what exactly is business plan and why it is necessary. After starting a business, entrepreneurs need to think about how to continue to improve and grow by adding value to their business for attracting customers. The following principles need to be considered:
   - Quality
   - Scale-up
   - Provide some offers to your customers.
5. Business planning is important to entrepreneurs. Estimating the amount of material required - Knowing the amount of material required lets the entrepreneur determine the cost of manufacturing a product or the cost of providing a service.
6. **Fulfilled Needs** – These are the needs that customers understand. The needs are met by various companies or governments. For example, the need for travelling modes is covered by both private bus services and state bus or metro services.

   Partially Fulfilled Needs – These are needs that are met through different products or services, but customers are not completely satisfied and still have problems using them. For example, people always use taxi services like ola, uber and auto-rickshaws to get to different places, but they struggle to find one on time and end up paying an extra amount for it.

   Unfulfilled and Known Needs – These needs are known to customers, but are not met by anyone in the market. For example, residents of small towns had no electricity at all times, and therefore their children could not study after sunset. A company recognized this problem and installed solar lamps, which are charged during the day and can be used at night.

   Unknown Needs – These are needs that people have but are not aware of or expect to be addressed by a business. For example, 10 years ago people didn't know that online education is possible. However, through innovation, entrepreneurs have found the technology to make it happen.

7. a. Open-mindedness b. Perseverance

8. A successful entrepreneur must possess several key values. Three important values are:
   - **Confidence:** Confidence means believing in yourself and your ideas.

     Example: An entrepreneur takes the first step to start a new business and continues to try new things to grow the business, even if they face failures.
   - **Independence:** Independence means the ability to work alone and make decisions without needing constant supervision.

     Example: An entrepreneur sets their goals independently and works towards achieving them, taking full responsibility for success.
   - **Perseverance:** Perseverance means not giving up even when facing difficulties.

     Example: An entrepreneur continues to find creative ways to overcome problems and always focuses on achieving the goal, despite any failures or setbacks

B.
- Lack of Sufficient and High-Quality Data
- Difficulty in Accounting for Multiple Variables

Lab Activity

Do it yourself.

5. Green Skills-III

Unsolved Questions

Section A (Objective Type Questions)

A. 1. c. 2. a. 3. b. 4. d. 5. c.

B. 1. ODF Plus 2. Diu 3. agriculture and transport
4. stakeholder 5. Forests

C. 1. False 2. True 3. False 4. True 5. True

Section B (Subjective Type Questions)

A. 1. Liter of Light is a global grassroots initiative dedicated to providing those with limited or no access to electricity with affordable, sustainable solar light.

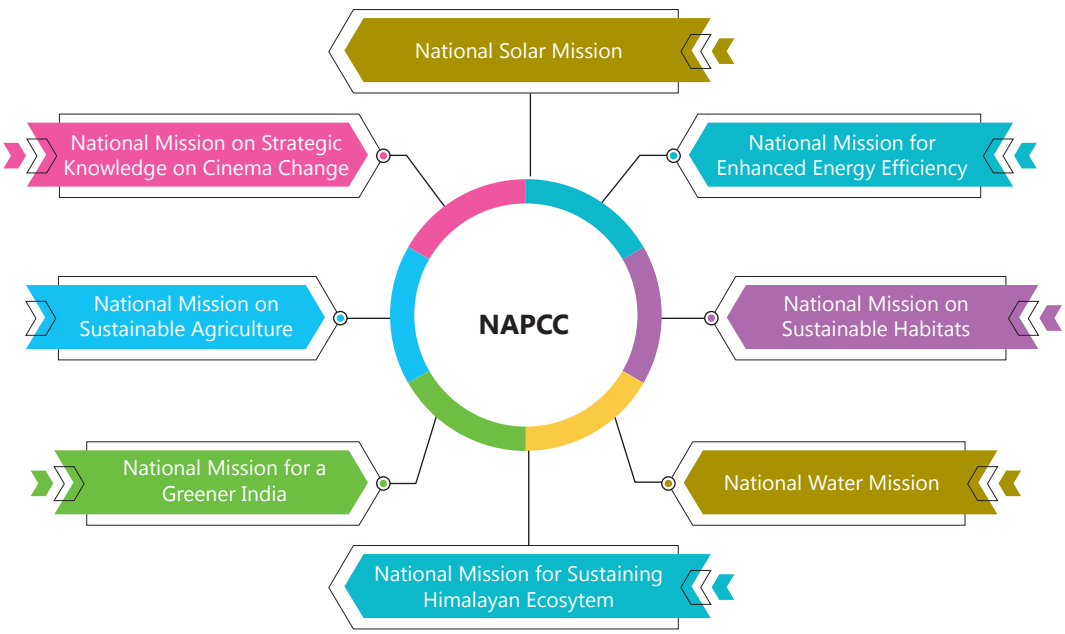
Liter of Light volunteers teach impoverished communities throughout the world how to use recycled plastic bottles and locally sourced materials to illuminate their homes, businesses, and streets through a global network of partnerships.

Liter of Light has placed over 350,000 bottle lights in over 15 countries while also teaching green skills to empower grassroots entrepreneurs at each stop.

2. • Private agencies in a green economy:
   - Enforcing laws and policies enacted by the government
   - Ensuring that government regulations are followed by engaging in official missions
   - Reporting individuals or groups that violate the law
   - Educating people and raising awareness about green economy initiatives
   - Informing the government and assisting it in implementing green policies
3. Agriculture - Agriculture is the process of raising foods for human consumption. It is the most significant component of our economy since it provides us with food. Agriculture has the potential to harm the environment by overusing land resources, cutting down forests, and polluting air, water, and soil with hazardous chemical fertilizers and pesticides.

   Transport - The transportation sector consumes the most gasoline and diesel and emits the most pollution.

   A green economy's transportation sector strives
   - to use cleaner fuels such as electricity for cars.
   - encouraging the use of public transportation, cycling, and other forms of alternative transportation to discourage the use of automobiles
4. Science and Technology - Science and technology are required in all sectors of sustainable development. Scientists can explain technical issues to governments and policymakers. They can also offer long-term answers to improve people's lives. Building Capacity for Sustainable Development is one of them.

   Other roles include - investing in the Next Generation's Education, providing Information Access, increasing the scientific basis for decision-making and educating the public
5. 

6. A stakeholder is defined as somebody who has an interest or concern in something (company, policy, or activity). A green economy has many stakeholders since it impacts everyone and requires the assistance of many individuals. Example of a stakeholder - Native Tribes - Native or indigenous people are the original inhabitants of a territory who continue to live by their old way of life. Many such tribes in India's hills and forests can benefit from the green economy. For example, tribals in Nangur village in Bastar district, Chattisgarh have greatly benefited from using various parts of the 'Mahua' plant to make eco-friendly plates, cones, and bowls. The tribals have turned this into a lucrative local business.
7. Water Management
    - One of our most valuable resources is water. Billions of people throughout the world lack access to safe drinking water and improved sanitation facilities. Population explosion is also making the situation worse. Conserve water by
    - turning off the faucet when not in use
    - Use wastewater treatment to prevent pollution of water sources.

B. She should file her complaint with the National Green Tribunal (NGT).

Lab Activity

Do it yourself.

Class 11

Answer key

Part-B: Subject Specific Skills

DATA SCIENCE PART-B

1. Ethics in Data Science

Unsolved Exercise

Section A
(Objective Type Questions)

A. 1. d 2. d 3. a 4. d
5. d 6. b 7. c

B. 1. True 2. True 3. False 4. True 5. False

Section B
(Subjective Type Questions)

A. 1. Ethical norms are necessary as they promote knowledge, truth, and avoidance of error. For example, they prohibit fabricating, falsifying, or misrepresenting research data to promote the truth, and reduce error or flaws.

2. Insider threats are threatening the organisation for given reasons:
   - Insiders act maliciously most of the times. That's why it's harder to detect their harmful activities than it is to detect external attacks.
   - Insiders know flaws in an organisation's cybersecurity.
   - Insiders know the location and nature of sensitive data they can misuse.

3. Espionage means to spy on a business to obtain confidential information about its operations. Trespassing onto a competitor's property or accessing their files without consent, impersonating an outside worker to know company trade secrets and other sensitive information, etc. are the acts of espionage.

4. Data Governance is the process or procedure that organisations uses to manage, utilise, and protect their data. In this context, data can mean either all or a subset of a company's digital and/or hard copy assets. In fact, defining what data means to an organisation is one of the data governance best practices.

5. Insider attacks are threatening as they can lead to a variety of repercussions, from fines for non-compliance with cybersecurity necessities to the loss of customer trust.

Three reasons:

1. Insiders act maliciously most of the time.
2. Insiders know flaws in an organisation's cybersecurity.
3. Insiders know the location and nature of sensitive data.

B. 1. Over time, data has gone from being small, structured, and easy to handle with tools like spreadsheets, to becoming massive, dynamic, and complex, which we call "big data." This shift has led to the need for new techniques and tools to manage and make sense of this flood of information. Google has been a big player in this evolution, not just with its own platforms like Search and YouTube, but also by sharing key technologies like Kubernetes and TensorFlow with the world through open-source projects. These innovations have made it easier for everyone to work with big data and use it to solve problems.

2. Differences between insider and outsider threats in terms of access, motivation, detection difficulty, collaboration, and preventive measures.

Aspect	**Insider Threat**	**Outsider Threat**
Nature of Access	Individuals with authorized access to organization's systems and data	Individuals without authorized access attempting to breach systems
Motivation	Personal gain, vendetta, or negligence	Malicious intent such as financial gain or disruption
Examples	Disgruntled employee leaking data to a competitor, Employee falling for a phishing scam	Hacker infiltrating network for ransomware, Competitor hiring hacker for DDoS attack
Detection Difficulty	Harder to detect due to authorized access	May be easier to detect through network monitoring
Collaboration	Insiders may collaborate with outsiders for nefarious purposes	Outsiders may exploit insider knowledge or vulnerabilities
Preventive Measures	Employee training, Access control policies, Monitoring employee behavior	Cybersecurity defenses like firewalls, antivirus, IDS

3. Sensitive data may include:
   - Trade secrets and intellectual property
   - Regulated data
   - Sensitive data
   - Information about products and internal research
   - Financial and personal staff information
   - Source code

- Business plans and other sensitive corporate information.

4. • **Password attacks:** Users' passwords are cracked either by guessing a possible password or through a dictionary attack by cyber attackers. To prevent such attacks, employees must ensure strong passwords and account lockout policy, that locks out a user after a certain number of wrong attempts.
   - **IP spoofing:** In Internet Protocol (IP) spoofing, an attacker tries to convince a system that it is communicating with a trusted entity like a trusted website. Therefore, the attacker provides an IP address known to the system instead of its real IP address to gain access by breaking system's firewall.
5. You must ensure your organisation has policies and training in place to prevent data breaches before hackers get through. This includes:
   - Training employees on security hygiene.
   - Having a BYOD (bring your own device) policy.
   - Segment your Wi-Fi network, so that employees who want to use it for their devices can do so without.
   - Use multilevel access authorisations, only allowing those who need access to the most sensitive information in your organization.
   - Monitor your network for anything out of the ordinary.
   - Use encryption to protect the most sensitive data.

Higher Order Thinking Skills

1. Ethical principles for conducting a research that involves dealing with other people's data are as follows.

   Ethical Principles for Researchers:

   a. **Informed Consent:** Researchers should obtain consent from potential participants after providing them with comprehensive information about the study.

   b. **Minimize Harm:** Efforts should be made to minimize any potential harm or risks to the participants involved in the research.

   c. **Anonymity and Confidentiality:** Researchers must ensure the anonymity and confidentiality of the participants' information to protect their privacy.

   d. **Avoid Deception:** Researchers should refrain from using deceptive practices in their studies to maintain transparency and integrity.

   e. **Right to Withdraw:** Participants should have the right to withdraw from the research at any time without facing any consequences.

   Researcher Responsibilities:

   a. **Obtaining Consent:** Researchers have the responsibility to obtain explicit consent from individuals who agree to participate in their studies.

   b. **Mitigating Risks:** Researchers should take measures to identify and minimize any potential risks or harm to the participants throughout the research process.

c. **Protecting Privacy:** It is the duty of researchers to ensure that participants' identities and personal information are kept confidential and anonymized.

d. **Ensuring Transparency:** Researchers should conduct their studies with honesty and transparency, avoiding any form of deception or manipulation.

e. **Respecting Participants' Rights:** Researchers must respect the autonomy of participants by giving them the freedom to withdraw from the study if they choose to do so.

2. Yes, there can be differences in expectations regarding what is considered ethical when handling data online compared to offline.

a. **Accessibility and Reach:** Online data can be accessed by a larger audience and replicated easily, potentially impacting privacy expectations compared to offline data, which may have more limited reach.

b. **Permanence and Replicability:** Online data can be stored indefinitely and easily replicated, raising concerns about privacy and security, whereas offline data may have more limited storage and distribution.

c. **Anonymity and Identity Protection:** Online interactions often involve pseudonyms or anonymous profiles, leading to different perceptions of anonymity compared to face-to-face interactions offline.

d. **Regulatory Environment:** Different regulations and standards govern online versus offline data handling, which can shape expectations regarding ethical behavior in each context.

e. **User Awareness and Education:** Users may have varying levels of awareness and understanding of data privacy risks and best practices online versus offline, influencing their expectations of ethical data handling.

3. Here are some concise pointers outlining different motives behind cyber attacks:

a. **Espionage:** Gathering sensitive information for political, economic, or strategic purposes.

b. **Hacktivism:** Promoting a cause, protesting against organizations or governments, or raising awareness about social issues.

c. **Cyber Warfare:** Disrupting or undermining the infrastructure, economy, or security of rival nations.

d. **Personal Vendettas:** Seeking revenge or retaliation against specific individuals or entities.

e. **Disruption of Services:** Targeting critical infrastructure or businesses to disrupt services or cause chaos.

f. **Data Manipulation or Destruction:** Intentionally manipulating or destroying data to cause harm or financial loss.

g. **Extortion:** Demanding payment through tactics like ransomware to release encrypted data.

h. **Intelligence Gathering:** Collecting information on potential targets, vulnerabilities, or security measures.

Applied Project

Do it yourself.

2. Assessing Data

Unsolved Exercise

Section A
(Objective Type Questions)

A. 1. b 2. c 3. c 4. d
5. a 6. a 7. a

B. 1. True 2. False 3. False 4. False 5. True

Section B
(Subjective Type Questions)

A. 1. Cause here is "playing a lot in Casino" and effect is "losing all his money".

2. Trial assessments, like experiments, test things by examining treatment (independent) and response (dependent) variables. In statistics, experiments aim to understand cause and effect relationships. Researchers conclude that the treatment caused an effect on the response variable.

3. Agent causation is the concept that only 'things' have the power to change the world. Agent causation requires the specification of a 'thing' that caused the effect. For example, the brick broke the window. Modern versions of agent causation are largely restricted to purposeful agents that are self-directed (teleology).

4. A correlation of zero means there is no relationship between two variables. In other words, as one variable moves one way, the other moves in another unrelated or uncertain direction.

5. A data quality assessment is a distinct phase within the data quality life-cycle that is used to verify the source, quantity and impact of any data items that breach pre-defined data quality rules.

B. 1. Three facts are as follows:-

a. Water freezes at 0 degrees Celsius under standard atmospheric pressure. Justification: This is a fact supported by scientific observation and experimentation. It has been consistently observed and verified in numerous experiments under various conditions.

b. The Earth orbits around the Sun. Justification: This fact is based on astronomical observations and mathematical calculations. It has been confirmed through centuries of

observations by astronomers and through advancements in technology like telescopes and space probes.

c. The chemical formula for water is H_2O. Justification: This fact is derived from the principles of chemistry and has been confirmed through experimentation and analysis. The atomic structure of water molecules has been extensively studied, and the composition of two hydrogen atoms and one oxygen atom in a water molecule is well-established and universally accepted in the scientific community.

2. Causation is the act or process of causing something to happen or exist. Causation is the relationship between an event or situation and a possible reason or cause.

   Example 1: Exercise and Weight Loss

   Causation is the relationship between exercise and weight loss. Regular physical exercise, combined with a balanced diet, is known to contribute to weight loss and maintenance of a healthy weight. Engaging in physical activity increases energy expenditure and promotes the burning of calories, leading to a negative energy balance in the body. Over time, this calorie deficit results in the reduction of body weight and fat mass. In this case, exercise (the cause) leads to the effect of weight loss.

   Example 2: Deforestation and Global Warming

   Cutting down trees reduces the Earth's ability to absorb carbon dioxide, leading to an increase in greenhouse gases and causing global warming.

3. Causal relationships describe the cause-and-effect connections between two or more variables or events. In a causal relationship, changes in one variable cause changes in another variable.

   Example 1: Rainfall and Plant Growth

   A causal relationship exists between rainfall and plant growth. When plants receive an adequate amount of water through rainfall, they can absorb nutrients from the soil more effectively, undergo photosynthesis, and grow. Conversely, insufficient rainfall can lead to drought conditions, causing plants to wilt and eventually die. Here, the presence or absence of rainfall (the cause) directly influences the growth of plants (the effect).

   Example 2: Education Level and Income

   There is a causal relationship between education level and income. Research consistently shows that individuals with higher levels of education tend to earn higher incomes compared to those with lower levels of education. Higher education provides individuals with specialized skills, knowledge, and qualifications that are valued in the job market, leading to better job opportunities and higher-paying careers. Therefore, education level (the cause) has a direct impact on income level (the effect).

4. The effects of different duration of light on the growth of radish seedlings: -

   a. **Optimal Growth with Continuous Light:** Radish seedlings exposed to 24 hours of light exhibited consistent growth. This treatment likely provided continuous energy through photosynthesis, promoting robust growth and development.

   b. **Mixed Growth with Alternating Light and Dark Periods:** Seedlings subjected to 12 hours of light followed by 12 hours of darkness showed mixed growth patterns. While they received periods of light for photosynthesis, the alternating light and dark periods

may have disrupted their circadian rhythms, potentially leading to fluctuations in growth rates.

c. **Limited Growth in Darkness:** Radish seedlings kept in 24 hours of darkness displayed limited growth. Without light, photosynthesis cannot occur, depriving the seedlings of energy needed for growth. This likely resulted in stunted growth and weaker seedlings compared to those receiving light exposure.

Continuous light exposure fosters optimal growth, while alternating light and dark periods may have mixed effects, and complete darkness inhibits growth.

5. The perception of time assessment reveals how individuals experience time during events, which can vary widely. Visual response latency refers to the delay between a stimulus and its perception, affecting reaction time. Human reaction time peaks around age 24, then slows due to neuronal loss. Males typically have faster reaction times than females across age groups. Time perception relates to species size and metabolic rate, and it tends to speed up as individuals age. Additionally, thrilling experiences, like a launch chamber ride, can cause time overestimation, as seen when riders estimate longer durations than the actual ride time.

   Examples for Perception of time assessment: -

   a. Two friends attend a concert together. One friend is thoroughly enjoying the music and dancing, while the other is feeling bored and checking the time frequently. Despite experiencing the same amount of time passing, their subjective perceptions of time differ significantly due to their differing levels of engagement in the event.

   b. During a science experiment, participants are shown a series of images on a computer screen and asked to press a button as soon as they see a specific image. Researchers measure the time it takes for the participants to respond, demonstrating the delay between the visual stimulus and the observer's perception and reaction.

Higher Order Thinking Skills

1. Yes, there is generally a correlation between speaking and writing skills, but it can vary depending on factors such as the individual's proficiency in the language, their exposure to it, and their educational background. In the case of someone using their native language, there tends to be a strong correlation between their speaking and writing abilities. This is because they have likely been exposed to the language from a young age, have had ample opportunities to practice both speaking and writing, and have likely received formal education in grammar and language usage. However, when it comes to individuals using a second language, the correlation between speaking and writing skills may not be as straightforward. Some people may be able to think and write in the language with relative ease, often avoiding grammatical mistakes through conscious effort and perhaps assistance from tools like grammar checkers. However, when speaking, they may struggle more with grammar and fluency due to factors such as lack of practice, nervousness, or differences in pronunciation and intonation. Nonetheless, with consistent practice and immersion in the language, individuals can improve both their speaking and writing skills over time, thus strengthening the correlation between the two.

2. To test the relationship between running outdoors and skin diseases while mitigating the impact of sun exposure, researchers could introduce an independent variable such as sunscreen usage. By comparing the occurrence of skin diseases among outdoor runners who consistently apply sunscreen versus those who do not, researchers can isolate the effect of sun exposure on skin health. Alternatively, researchers could explore the use of protective clothing, such as long-sleeved shirts, hats, or UV-protective fabrics, as an independent variable. By comparing the incidence of skin diseases between runners who wear protective clothing and those who do not, researchers can assess the effectiveness of these measures in reducing the risk of sun-related skin issues. Additionally, researchers could consider conducting the study during different times of the day when UV radiation is lower, or in locations with less intense sunlight, to further minimize the impact of sun exposure on the results. Overall, by incorporating these alternative variables and strategies, researchers can better understand the relationship between outdoor running and skin diseases while controlling for the confounding effects of sun exposure.
3. Some factors influencing reaction time:
    i. Fatigue: Tiredness slows down cognitive processing, leading to delayed responses.
    ii. Stress: High stress levels impair cognitive function, increasing reaction time.
    iii. Attentional Focus: Distractions or divided attention can slow response speed.
    iv. Experience and Expertise: Practice improves reaction time through enhanced skills and automation.
    v. Mood: Positive moods speed up responses, while negative moods may slow them down.
    vi. Medication and Substance Use: Certain drugs can impair reaction time.
    vii. Physical Health: Better health and fitness correlate with faster reactions.
    viii. Temperature and Environment: Extreme conditions can affect neural processing speed.
    ix. Nutrition and Hydration: Proper diet and hydration support optimal brain function.
    x. Genetics: Genetic factors play a role in neural processing speed.

Applied Project

Do it yourself.

3. Forecasting on Data

Unsolved Exercise

Section A
(Objective Type Questions)

A.

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

B.

1. trial assessments
2. naturalistic observation, participant observation, controlled observation
3. Sir Richard Doll
4. experimental
5. data, data organising, and coordination.

C.

1. False	2. True	3. True	4. True	5. False

Section B
(Subjective Type Questions)

A.

1. Forecasting is an approach that takes data and predicts the future value of the data, looking at its distinctive trends. Forecasting depends on the data that is collected, usually from both the past and the present, followed by the market trends analysis and the development of the description of the actions to follow. Examples: Weather Forecasting, Sales Forecasting
2. Forecasting provides companies an edge over their competitors. Forecasting helps in: -
   a. Gaining valuable insight
   b. Learning from past mistakes
   c. Decreasing the life cycle costs
   d. Establishing a New
   e. Formulating Plans
   f. Estimating Financial Needs
   g. Facilitating Managerial Decisions
3. Four main types of forecasting methods that financial analysts are likely to use:
   a. Perform financial forecasting
   b. Reporting, and operational metrics tracking
   c. Analyse financial data
   d. Create financial models use to predict future revenues
4. The names of more popular data collection methods are observations, Questionnaires and surveys, Interviews, Documents and records, Focus groups, Personal narratives.

5. The advantages of observational study are as follows:
   a. Simplicity
   b. Perfection
   c. Good assessment of real-world applicability
   d. Economical
   e. It helps identify prospective customers for a business
   f. Helps study rare events

   Disadvantages of observational study are as follows:
   a. Time commitment
   b. An event's non-occurrence
   c. Observer negligence
   d. Biased opinion
   e. Mediocre as an evidence
   f. Biased
   g Bogus answers of respondents
   h. Cost
   i. More than one control group
   j. Cannot show cause and effect relationship

6. Short notes
   a. Biased opinion: The chances of an unfair conclusion increase significantly in cases where an expert has not performed the study's analysis.
   b. More than one control group: In general discussions of observational studies, the possibility of using more than one control group has often been briefly mentioned. Many observational studies have even used two control groups. However, a second control group can be of little value—even though it happens in worst of circumstances.
   c. A retrospective observational study is a type of study where researchers analyze data that has already been collected from past events or records. In this type of study, researchers do not intervene or manipulate any variables; instead, they observe and analyze data to identify patterns, associations, or correlations between variables.

B. 1. Advantages of forecasting in detail.
   a. **Gaining valuable insight:** Looking at past and real-time data is a pre-requisite to predict future demand through forecasting. This will, in turn, help anticipate demand fluctuations more effectively. Also, it will give you an understanding of your company's health and provide you with an opportunity to make necessary amendments.
   b. **Learning from past mistakes:** You do not go back to square one after each forecast. Even if your prediction was completely off the mark, you now know where to begin.You can easily analyse why things didn't happen the way you predicted. This will help you

improve your predicting techniques. You can also reflect on your past achievements, as introspection can be a powerful driver of company growth.

c. **Decreasing the life cycle costs:** If demand forecasting is done the right way, it will help you modify your processes to multiply your efficiency all along the supply chain. Anticipating what and when customers will demand aids in reducing excess inventory and increasing gross profitability.

d. **Establishing a New Business:** A number of business forecasts are required at the time of starting a new business. One has to predict the demand for the product, the caliber of competitors, their share in the market, sources of raising finance, etc. Precision in such forecasts is therefore valuable for the success of a business.This will in turn help the operations of the business to run smoothly, thereby minimising the chances of any failure.

e. **Formulating Plans:** Forecasting provides a logical basis for preparing plans. It plays a major role in managerial planning and supplies the necessary information. The future assessment of various factors is essential for preparing plans. In fact, planning without forecasting is an impossibility. Henry Fayol has rightly observed that the entire plan of an enterprise is made up of a series of plans called forecasts.

f. **Estimating Financial Needs:** Adequate capital is a necessity for every business. Correct estimates of financial requirements are there to prevent businesses from suffering from inadequate or excess capital. Forecasting of sales, expenses, revenue, etc. helps predict financial needs. Financial planning is based on systematic forecasting.

Facilitating Managerial Decisions: It paves the way for sound managerial decisions about personnel, materials, and sales, etc., by providing a rational foundation for anticipating and judging the nature of future business operations.

2. Disadvantages of Forecasting.

a. **Forecasts are never 100% accurate:** Obviously, it's hard to predict the future. Despite having a great technique in place and an expert panel at your disposal, your forecasts might not be accurate. This is primarily because products and markets are volatile in nature. Sometimes there is no one specific reason for a surge in demand as many factors come into play.

b. **It can be time-consuming and resource-intensive:** Forecasting calls for a lot of data gathering, data organising, and coordination. Companies typically hire a team of demand planners who are in charge of coming up with the forecast. But in order to do this well, demand planners need considerable input from the sales and marketing teams. Also, it's not uncommon for processes to be manual and labour-intensive, thus making them time-consuming. But if there is the right technology in place, this is much less of an issue.

c. **It can also be costly:** Employing a team of demand planners is a significant investment. When you add to that the cost of using cutting-edge quality instruments, the upfront fees might quickly pile up. Investing in modern technologies,prime talent, and sound

forecasting methods, on the other hand, is nothing more than that. If investment is done correctly, it will yield positive results.

d. **Forecasting can be dangerous:** Forecasts become a focus for companies and governments, mentally limiting their range of actions by presenting the short-to long-term future as pre-determined.

e. **Not every situation can be predicted:** This is one of the disadvantages of demand forecasting. For example, severe weather could affect product or material supply accessibility or transportation logistics.

f. **Financial Forecasting Inefficiencies and Lack of Data Credibility:** From insufficient information to disconnected data within the forecast, a number of forecasts have credibility issues. Often the forecasts are unable to tell the true story of where the business is going.

3. Observational studies are conducted because sometimes it is not possible, ethical, or feasible to perform trial assessments.

   The main reasons are:

   a. **Ethical concerns:** It is unethical to randomly expose subjects to harmful treatments (e.g., exposing people to polluted air to study its effects). Thus, observational studies are preferred.

   b. **Cost and time limitations:** Trial assessments may require large sums of money and long durations. When resources are limited, observational studies are a better option.

   c. **Random assignment challenges:** In some scenarios, random assignment of subjects is impractical or impossible. Observational studies allow researchers to collect data without altering subject conditions.

4. Sometimes, it may not be possible to perform trial assessments. In such cases, you need to rely upon observational study for data collection. The reasons behind this are as follows:

   a. In trial assessments, the subject is assigned to a random treatment and control group. However, it is unethical to expose the subject to arbitrary treatment in specific scenarios. Thus, observational studies are preferred over trial assessments. For example, purposefully exposing a subject to polluted air to observe the health issues that come to the forefront is unethical.

   b. Some of the trial assessments may require large sums of money or time to execute. There may be occasions when such large sums of money cannot be arranged. In such scenarios, it will be a better idea to drop the idea of performing trial assessments, and give the observational study a priority.

   c. A trial assessment cannot be performed in some scenarios, as it becomes unfeasible to assign a subject to a group randomly.

5. • **Experimental Studies:** In a controlled experiment, we assign people or things to groups and apply some treatment to one of the groups, while the other group does not receive the treatment. Experimental studies are ones where researchers introduce an intervention

and study the effects. Experimental studies are usually randomized, meaning the subjects are grouped by chance.

- **Observational Studies:** An observational study can be defined as a procedure in which the subjects are just observed and the results are then noted. During the investigation, nobody tries to interfere with the subject to affect the outcome. In an observational study, we measure or survey members of a sample without trying to affect them. Observational studies are ones where researchers observe the effect of a risk factor, diagnostic test, treatment or other intervention without trying to change who is or isn't exposed to it. Cohort studies and case control studies are two types of observational studies.

6. Observational data is best used when one of the following situations occurs:
   a. While collecting sensitive information, when you don't trust your participants to be truthful with their self-reporting.
   b. When you need to understand the how or what of a research question.
   c. When you need robust data to describe consumer habits.
   d. When behaviour in a natural setting is important to your research question.
   e. When behaviour in a controlled setting is critical to your research question.
   f. When you are concerned that self-reported data about behaviours will not be the same as actual actions, despite being unintentional.
   g. When you need more information about a particular research question to form a complete and error-free survey in all respects.
   h. If you are in any of the above stated research states, then you might need an observational study.
7. An observational study can be of two types:

   Direct observation: In this type of observation, the information is collected using sensory organs. Through direct observation, you can document activities, habits, and physical aspects of a situation without depending on people's willingness or ability to respond correctly to questions.

   Indirect observation: In this type of observation, some data related to past behaviour or habits is used to infer what occurred during an event. This can be explained using the following example: looking at packaging usage in a fast food restaurant to understand what flavoured drinks sell the most.
8. There are three different types of methodologies to conduct observational research: controlled observations, naturalistic observations, and participant observations.
   - **Controlled Observation**

     Controlled observations are usually a structured examination that takes place in a psych lab. The researcher has a question in mind and controls many of the variables, including

participants, observation place, time of the study, conditions surrounding the research, etc. While conducting this observation, the researcher will often create codes that signify different types of behaviour. That way, instead of writing an exhaustive report, they can classify behaviour into different categories and evaluate the data easily.

- **Naturalistic Observation**

  Naturalistic observation is a research methodology used by market researchers to study the behaviours of participants in a natural environment. There are generally no fixed or predetermined behavioural codes. Rather, the researcher will take notes diligently and code the data thereafter.

- **Participant Observation**

  This is a type of naturalistic observation wherein market researchers observe participants in their natural habitat. Market researchers will involve themselves too in the environment.

9. The disadvantages of observational studies are as follows:
   - **Time commitment:** The amount of time spent on an observational study may not always be justified by the findings.
   - **An event's non-occurrence:** Certain events are uncertain and may not occur in the presence of an observer.
   - **Observer negligence:** The observer may occasionally fail to report important observational details.
   - **Biased opinion:** The chances of an unfair conclusion increase significantly in cases where an expert has not performed the study's analysis.
   - **Mediocre as an evidence:** Observational studies are considered inferior to experimental studies, are more prone to prejudice, and cannot be used to exhibit causality. Observational studies can be either retrospective (using existing data) or prospective (collecting new data).
   - **Biased:** Some of the weaknesses are human bias because the observer is a human with his/her own prejudice or bias. Moreover, one cannot know what the subject is thinking, the subject's perspective or decision-making process.
   - **Bogus answers of respondents:** Following are the factors on which survey data may depend: Respondents may not feel motivated to provide precise and honest answers. Respondents may not be fully aware of their reasons or justifications for any given answer because of lack of memory on the subject, or even boredom.
   - **Cost:** Cost is one of the limitations of the observation method. Under most scenarios, observational data is more expensive to obtain than other survey data. The observer remains idle while observing the events. This idle time further adds to the cost.
   - **More than one control group:** In general discussions of observational studies, the possibility of using more than one control group has often been briefly mentioned. Many observational studies have even used two control groups. However, a second control group can be of little value—even though it happens in worst of circumstances.

- **Cannot show cause and effect relationship:** Observational studies cannot demonstrate that the associations identified represent cause-and-effect relationships. However, observational studies can only demonstrate that notable associations exist between predictors and outcome variables.

Higher Order Thinking Skills

1. A monthly family budget is a fundamental tool for financial planning, offering a forward-looking projection of income and expenditure over a specific period. This forecast, however, is not without its challenges. While it aims to predict future financial flows based on historical data and current circumstances, it must contend with inherent uncertainties. Fluctuations in income sources, unexpected expenses, and changes in economic conditions all pose risks to the accuracy of the forecast. Moreover, budgeting extends beyond mere number-crunching; it requires disciplined behavior, effective communication within the family, and adaptability to changing circumstances. Despite its limitations, a well-designed budget provides a framework for decision-making, enabling families to prioritize spending, allocate resources effectively, and work towards their financial goals. Regular monitoring and review are crucial to ensure that the budget remains relevant and responsive to evolving needs and circumstances. Thus, while a monthly family budget serves as a valuable planning tool, its effectiveness hinges on factors such as flexibility, communication, and ongoing oversight.
2. When observing the food-seeking behavior of rats, the choice between conducting the study in the wild or in a laboratory setting presents distinct advantages and considerations. In the wild, researchers can observe rats in their natural habitat, providing insights into their behavior under real-world conditions. This approach offers a more ecologically valid understanding of their food-seeking strategies, influenced by factors such as competition, predation, and food availability. However, conducting observations in the wild may present challenges in controlling variables and ensuring consistency across observations. Conversely, laboratory settings offer controlled environments where researchers can manipulate variables and standardize conditions. This allows for precise experimentation and the isolation of specific factors influencing food-seeking behavior. Despite the advantages of laboratory studies, ethical considerations regarding animal welfare must be carefully addressed. Ultimately, the choice between wild and laboratory observation depends on research objectives, ethical concerns, and the need for ecological validity versus experimental control. Both approaches contribute valuable insights to our understanding of rat behavior and its implications.
3. The study aimed to understand the outcomes of hospitalized patients with COVID-19 in Mumbai City, India, during a specific period. Given that Mumbai had the highest number of COVID-19 cases in India during the study period, it targeted a population significantly affected by the pandemic. However, whether the study achieved its target depends on various factors, including the sample size, representativeness of the sample, and the study's objectives.
    - **Sample Size:** The study included 689 patients admitted to Nair Hospital & TN Medical College during the specified period. If this sample size is deemed sufficient to draw

meaningful conclusions about the outcomes of COVID-19 patients in Mumbai City, then the study could be considered to have achieved its target in terms of sample size.

- **Representativeness:** The extent to which the sample represents the broader population of COVID-19 patients in Mumbai City is crucial. If the patients included in the study are diverse in terms of demographics, severity of illness, and other relevant factors, then the findings may be more generalizable to the population. However, if the sample is biased or not representative of the larger population, the study's conclusions may be limited in their applicability.
- **Study Objectives:** The study aimed to understand the outcomes of hospitalized COVID-19 patients in Mumbai City. If the findings provide valuable insights into the mortality rates, recovery rates, or other relevant outcomes among this population, then the study can be considered successful in achieving its objectives.

In summary, the study's success in achieving its target depends on factors such as the adequacy of the sample size, the representativeness of the sample, and the extent to which it addresses its stated objectives. Evaluating these factors can help determine the study's overall impact and relevance to understanding the outcomes of COVID-19 patients in Mumbai City.

Applied Project

Do it yourself.

4. Randomisation

Unsolved Exercise

Section A
(Objective Type Questions)

A.

1. d	2. a	3. b	4. b	5. a
6. a	7. c	8. b	9. d	10. a

B.

1. XML
2. Microphone
3. Confidence
4. Information bias
5. Cluster sampling

C.

1. True	2. False	3. True	4. False	5. True

Section B
(Subjective Type Questions)

A. 1. **Primary Data**

Primary data is a type of data that is collected directly from first-hand sources like interviews, surveys, experiments, etc. by researchers. Primary data is generally collected from the

source—where the data initially originates from. This data is considered the premier data in research.

Secondary Data

Secondary data is the data that has already been collected in the past through primary sources and made readily accessible for researchers so that they can use it for their own research.

2. a. Observations b. Questionnaires and surveys c. Interviews
   d. Documents and records e. Focus groups f. Personal narratives
3. This data about the population is summarised via statistics known as "descriptive statistics". Descriptive statistics provide information about our immediate group of data. Numerical descriptors like mean and standard deviation for continuous data types (like income) and frequency and percentage for categorical data (like race) are very useful.
4. The three main measures of central tendency are Mean, Median and Mode.
5. • **Phenomenology:** Focuses on understanding and interpreting the subjective experiences and perspectives of individuals regarding a particular phenomenon.
   • **Grounded Theory:** Involves developing theories or explanations grounded in empirical data collected through systematic observation and analysis.
   • **Ethnography:** Involves immersive and in-depth study of cultures or social groups within their natural settings. Ethnographers seek to understand the cultural meanings, behaviors, and practices of the group under study.
   • **Narrative Inquiry:** Involves the study of human experiences and actions through the analysis of stories or narratives.
   • **Case Study:** Involves in-depth exploration and analysis of a particular case or phenomenon within its real-life context.
6. A survey is one of the research methods used for collecting data. In a survey, the subjects are usually people. The survey process involves asking various types of questions to a section of people (sample) for the desired information through a questionnaire. Surveys collect information from a targeted group of people or participants about their opinions, behaviours, suggestions, experiences, knowledge, etc

   The things to keep in mind while creating a survey:

   a. Define the purpose of the survey
   b. Make every question count
   c. Keep it short and simple
   d. Ask direct questions
   e. Ask one question at a time
   f. Avoid leading and biased questions
   g. Speak your respondent's language

h. Avoid using grids or matrices for responses

i. Consider adding incentives

g. Take your survey for a test drive

7. Surveys can be composed of two types of questions:

- Open-ended questions
- Close-ended questions

Open-ended questions can be about any comment/review or any suggestion for improvement.

- How can you bring a positive change in the society?
- Tell me about your relationship with your manager.

Closed-ended questions have a limited set of possible answers, such as:

- Do you have a backache?
- Do you prefer an online or an in-person lecture?

8. a. Clearly define the purpose of the survey and the specific information you want to gather.

b. Determine the demographic characteristics and attributes of the respondents you need to reach.

c. Choose an online survey platform (e.g., SurveyMonkey, Google Forms, Qualtrics) that fits your needs.

d. Design the Survey, Create and Format the Survey.

e. Distribute the Survey using Email Invitations or Social Media.

f. Collect Responses by Monitoring Responses And Ensuring Anonymity.

g. Analyze the Data and Interpret the Results.

h. Report Findings, Follow-Up and thus Reflect and Improve upon the shortcomings.

9. We might think that surveys are the same as questionnaires since these two terms are used interchangeably. But a questionnaire is any written set of questions. On the other hand a survey is both the set of questions and the process of gathering, aggregating, and analysing the responses to the questions asked.

10. • **Open-ended Questions**

Open-ended questions can be about any comment/review or any suggestion for improvement. These questions permit someone to give a free-form answer that cannot be answered with a "yes" or "no" response, or with a static response. The respondents can answer open-ended questions in their own words.

Examples of open-ended questions

Share with us what has been troubling you.

What do you feel about our customer service?

- **Close-ended Questions**

Closed-ended questions have a limited set of possible answers, such as:

A, B, C, or All of the above

None of the above

Yes/No

- **Star rating system**

The choice of answers from which to select is fixed in close-ended questions.

As a thumb rule, the type of question you choose depends on what you are trying to achieve; for instance, asking a closed-ended question when you want answers that can be plotted on a graph and used to show trends and percentages. For example, answers to the closed-ended question "Do you trust the information on this website?" will help you understand the section of people who find your website trustworthy.

Examples of closed ended questions

Should I take this as your final answer?

Are you excited to learn more about data science?

11. a. Strengths of Survey Methods

Following are the strengths of the survey methods:

- **High Representativeness:** Generally, surveying involves a large number of people, giving a better description of the relative characteristics of the general population involved in the study. Surveys are able to extract data that are near to the exact attributes of the large population as compared to other methods of data collection.
- **Economical:** The researcher conducting a survey needs to pay only for the production of survey questionnaires. While, other data collecting methods such as focus groups and personal interviews require researchers to pay more.
- **Convenient:** Nowadays, there are a number of ways through which surveys can be administered to the participants; online surveys, paper surveys, telephonic surveys, kiosk surveys, interviews, focus groups, etc. As such, researchers are able to collect data from people around the globe conveniently.
- **Easy to analyse and visualize data:** Since most surveys are quantitative in nature, third-party software tools can easily represent the results obtained from surveying via reports, charts, tables, etc. Thus, it becomes easier to analyse and visualize data.

b. Weaknesses of Survey Methods

Following are the weaknessess of the survey methods:

- **Lack of accessibility:** Surveys might be unsuitable for individuals who have a visual or hearing impairment. Some individuals might not have the prerequisite knowledge to answer the survey. There might even be areas with low literacy rates. To avoid this potential disadvantage, these issues must be considered during the planning stages of the research project.

- **Not a platform for contentious issues:** Controversial questions may not be accurately and honestly answered by few people due to reasons like prejudice. The truth behind these controversies may not be revealed as accurately as when using other data collection methods like focus groups and interviews. Some questions might not be appropriate to ask as these may instigate feelings of hatred.
- **Inflexibility:** The researcher is stuck with only one method of surveying. The design of surveying cannot be changed throughout the process of data collection.

12. A population is the entire group that you want to draw conclusions about. A sample is the specific group that you will collect data from. The size of the sample is always less than the total size of the population.
13. The most commonly used sampling Techniques are:

    Simple Random Sampling, Stratified Sampling, Systematic Sampling, Convenience Sampling, Quota Sampling, Cluster Sampling, Purposive Sampling (purposeful)

    The two most popular sampling techniques are purposeful and convenience sampling because they align the best across nearly all qualitative research designs.

B. 1. a. **Simple Random Sampling**

Advantages

Following are the advantages of simple random sampling:

- **Unbiased:** Since individuals are chosen at random from the larger group set, each individual in the large population has the same probability of being selected. In most cases, this creates a balanced subset that carries a huge capacity to represent the larger group as a whole.
- **Simple to apply:** This method involves dividing larger groups into smaller subgroups based on any attributes they share. The individuals are randomly selected and this doesn't require any special skills. Thereafter, there are no additional steps.
- **No prerequisite knowledge needed:** Researchers don't need to have any prior information or knowledge about the larger population.

Disadvantages

Following are the disadvantages of simple random sampling:

- **Time consuming:** Sample random sampling is time-consuming especially when a full list of a large population is not accessible or available. In that case, researchers might have to look for other sources of information. A list of population subset can be used to recreate the entire population list, but this might take time.
- **Costly:** The third-party data providers often charge for providing database or lists associated with population or its subsets.

b. **Stratified Sampling**

Advantages

Stratified random sampling has advantages when compared to simple random sampling. These advantages are:

- **Accurately Reflects Population Studied:** Stratified random sampling accurately reflects the population being studied because researchers are stratifying the entire population before applying random sampling methods. In short, it ensures each subgroup within the population receives proper representation within the sample. As a result, stratified random sampling provides better coverage of the population since the researchers have control over the subgroups to ensure all of them are represented in the sampling.
- **Good representation of attributes of population:** With simple random sampling, there isn't any guarantee that any particular subgroup or type of person is chosen. In our earlier example of the university students, using simple random sampling to procure a sample of 100 from the population might result in the selection of only 20 male undergraduates or only 20% of the total population. Also, 25 female graduate students might be selected (25% of the population) resulting in under-representation of male undergraduates and over-representation of female graduate students. Any errors in the representation of the population have the potential to diminish the accuracy of the study.

Disadvantages

Stratified random sampling also presents researchers with a disadvantage. Here is the disadvantage:

- **Can't be used in all studies:** Unfortunately, this method of research cannot be used in every study. The method's disadvantage is that several conditions must be met for it to be used properly. Researchers must identify every member of the population being studied and classify each of them into one, and only one, subpopulation. As a result, stratified random sampling is disadvantageous when researchers can't confidently classify every member of the population into a subgroup. Also, finding an exhaustive and definitive list of an entire population can be challenging.

c. **Systematic Sampling**

Advantages

Following are the advantages of systematic sampling:

- **Evenly distributed samples:** In systematic sampling, each individual is selected at a fixed distance, creating evenly distribution of collection of subjects. Since it is highly structured, it produces a more authentic representation of the entire population. As such, the results are easier to analyse.
- **Simplicity:** Researchers are able to build, analyse and manage such samples easily because of its basic structure. The formula to select sample subsets is known prior, as such choosing the initial subject randomly is what is required. Thereafter, the selection process follows a fixed pattern, until the desired sample group is complete.

- **Unbiased:** With systematic sampling, each participant is at a fixed distance from each other. Since there is a fixed interval, researchers have no control over which individuals are chosen for sampling. This helps reduce the chances for bias, favoritism, errors, and data manipulation.

Disadvantages

Following are the disadvantages of systematic sampling:

- **Possibility of uneven selection:** Since systematic sampling draws conclusions from a subset of a population, results might not be completely precise. Systematic sampling depends on a numbering system to select sample participants. Responses of some participants are not included, so the results can't be complete. Therefore, the researchers will always miss feedback, leading to a new finding.
- **Prediction of patterns:** If the population being surveyed is small, the integer pattern used to choose samples can be easily predicted. This can cause bias and lead to erroneous responses among the participants.
- **Dependability of outcome on population count:** It is the initial count of the population on which the outcome is dependent. After all, that's the number that is divided by the desired sample size to determine the fixed interval for selecting sample. When the population isn't measurable or available, a close approximation is required. If the population is smaller or larger than its actual number, this can produce wrong results.

d. **Convenience Sampling**

Advantages

Following are the advantages of convenience sampling:

- **Time saving:** This method is used to quickly collect data. Since it follows a simple approach to draw sample from that part of the population that is easier to find, data collection takes minimal time. Convenience sampling is not just easy to use, but also has other research advantages.
- **Convenient to use:** This sample is used as the last resort when researchers cannot have an access to the list of all the people in a population. It is generally chosen as it is convenient in terms of location, accessibility, etc.

Disadvantages

Following are the disadvantages of convenience sampling:

- **Biased:** This method of sampling can be very biased as the researchers may be tempted taken in the perspectives or responses of only those people they know.
- **Lack of credibility:** The results cannot be generalised due to lack of representation of the population. There can be sampling errors as well.

e. **Quota Sampling**

Advantages

Following are the advantages of quota sampling:

- **Meets specific needs:** Quota sampling is useful when the time frame to conduct a survey is limited, the research budget is very tight, or survey accuracy is not the priority. For example, job interviewers with a limited time frame to hire specific types of individuals can use quota sampling.
- **Simple:** It is an easy process to carry out and decipher information once the sampling is done. It also enhances the representation of any particular group within the population thereby preventing over-representation of groups. This sampling technique saves time and resources, making it easier to implement. Moreover, interpreting the responses is more straightforward and takes fewer resources.
- **Accurately represents the whole population:** Quota sampling is all about taking into consideration taking population samples. Since researchers use particular quotas, they can prevent over or underrepresentation of population.

Disadvantages

Following are the disadvantages of quota sampling:

- **Random selection not allowed:** Since quota sampling doesn't use random selection, detection of sampling error by researcher becomes difficult.
- **Not all traits taken into consideration:** In quota sampling, only the pre-determined traits of the population are taken into consideration by researchers. For example, in a research that involves creating samples based on gender and income, it might not represent other traits like age, race, or religion.
- **Biased:** In quota sampling, it is usually up to the researchers to decide who is sampled. Unknowingly or knowingly, researchers may sample a population on the basis of convenience, cost, etc., leading to bias.

f. **Cluster Sampling**

Advantages

Following are the advantages of cluster sampling:

- **Convenient:** By choosing large samples, researchers can increase accessibility to various clusters.
- **Data accuracy:** Since there can be large samples in each cluster, loss of precision in information per person can be compensated.
- **Economical:** This method of sampling requires fewer resources since only certain groups from the entire population are selected for the sampling process. It is economical as compared to other sampling techniques as it requires fewer administrative and travel expenses.

Disadvantages

Following are the disadvantages of cluster sampling:

- **Samples are biased:** This sampling technique is susceptible to biases. Incase clusters representing the entire population were formed with preconceived notions or prejudice, it is obvious that outcomes associated with the whole population would be biased too.
- **Not always applicable:** Cluster sampling requires people to be classified as a unit instead of an individual. This is a major drawback with this sampling technique.

g. **Purposive Sampling**

Advantages

Following are the advantages of purposive sampling:

- **Cost Effective:** The researcher selects the best-fit participants for the systematic investigation as per his understanding. This also saves time.
- **Precise research results:** Since the researcher collects qualitative responses, it leads to better insights and more accurate research results. Since the researcher gathers information from the best-fit participants, the results are not out of context. As such it also lowers the chances of any error.
- **Proper representation:** This technique of sampling ensures proper representation of the population when the survey has full knowledge of its composition and is free from any prejudice or bias.

Disadvantages

Following are the disadvantages of purposive sampling:

- **Not suitable for large samples:** This method is unsuitable for the large samples where the size of both the population and the sample being studies is large enough.
- **Prone to bias:** In purposive sampling, a sample is created initially and this depends on the judgement of the researcher. When the judgements are ill-conceived, then this problem becomes a huge disadvantage that can effect result. This can only be minimised when there is elicitation, accepted criteria, or a theoretical framework in place.

2. **Five Common Types of Sampling Errors**

Following are the five commonly used types of sampling errors:

a. **Population Specification Error:** This error occurs when the researcher does not understand who they should survey. For example, imagine a survey about breakfast cereal consumption.

b. **Sample Frame Error:** A frame error occurs when the wrong sub-population is used to select a sample.

c. **Selection Error:** This occurs when respondents self-select their participation in the study—only those who are interested respond. Selection error can be controlled by

going extra lengths to get participation. A typical survey process includes initiating pre-survey contact requesting cooperation, actual surveying, and post-survey follow-up. If a response is not received, a second survey request follows, and perhaps interviews using alternate modes such as telephone or person-to-person.

d. **Non-Response:** Non-response error occurs when almost the entire data for a sampling unit are missing. This can occur if the respondent is unreachable or temporarily missing, the respondent is unable to participate or refuses to participate in the survey, or if the dwelling is vacant.

e. **Sampling Size Errors:** These errors occur because of variation in the number or representativeness of the sample that responds. Sampling errors can be controlled by careful sample designs, large samples, and multiple contacts to assure representative response.

3. Bias means preference or prejudice for or against some thing.

   A sampling method is biased if it systematically prefers some outcomes over others.

   Sampling bias occurs in practice as it is practically impossible to ensure perfect randomness in sampling. If the degree of misrepresentation is small, then the sample can be treated as a reasonable approximation to a random sample

   A basic example of this bias is when a person refers to an individual by his or her occupation, such as 'doctor' or 'data scientist' and it is assumed that individual is male. Males, however, are not free from gender bias. Teachers, especially those who teach younger children, are often assumed to be women.

4. **Types of Sampling Bias**

   There are five main types of bias in research. Let us know each of them.

   - **Sampling bias:** Sampling bias is an error based bias in the way the survey respondents are chosen. This bias occurs when a survey sample is not completely random. In other words, if specific individuals are more or less likely to be selected as a sample for your research, chances are high that a sample selection bias may occur.
   - **Non-response bias:** Chances are high that there will be individuals who are disinterested or unable to take the survey. This causes discrepancy between the respondents and non-respondents, resulting in the so-called nonresponse bias.
   - **Response bias:** Merely responding to the survey is not enough, respondents must also provide accurate and honest responses. Less-than-truthful survey responses can come as an outcome of both conscious and subconscious cognitive factors.
   - **Question order bias:** The order of both questions and answers could cause survey respondents to provide biased answers. Sometimes, the initial questions of a survey could influence the answers respondents give to the subsequent questions later on in the survey.

- **Information bias:** Information bias is caused due to misrepresentation of truthfulness that occurs during the collection, handling, or analysis of data in a research study.
- **Healthy User Bias:** Occurs when the subjects who participate in a study are healthier than the general population, leading to skewed results.

  Example: Clinical trials for a new drug that mainly enroll individuals who are already in good health.

5. **Avoiding Sampling Bias**

   Using careful research design and sampling procedures can help you avoid sampling bias.

   Here are three ways to avoid sampling bias:

   Use simple random sampling. Probably the most effective method researchers use to prevent sampling bias is through simple random sampling where samples are selected strictly by chance.

   Use Stratified Random Sampling.

   Avoid Asking the Wrong Questions.

6. a. How amazing was your experience with our customer service team?(Question order bias)

      Correct question: Rate your experience with our customer service team.

   b. What problems did you have with the launch of this new product?(Question order bias)

      Correct question: Would you share the good and bad features of this new product?

   c. How do we compare to our competitors? (Response bias)

      Correct question: Where do consider yourself better than your competitors and in what areas do you want to work?

   d. Do you always use product X for your cleaning needs?(Non- response bias)

      Correct question: Which product do you use for your cleaning needs?

7. In statistics, a confidence interval is an educated guess about some characteristic of the population. A confidence interval contains an initial estimate plus or minus a margin of error (the amount by which you expect your results to vary, if a different sample were taken).

   A confidence interval shows how much uncertainty there is with any particular statistics. Confidence intervals are often used with a margin of error. Confidence intervals give us a range of plausible values for some unknown value based on results from a sample.

   A confidence interval is a range of values that is likely to contain an unknown population parameter.The confidence level represents the theoretical ability of the analysis to produce accurate intervals if you are able to assess many intervals and you know the value of the population parameter.

   Increasing the confidence level increases the error bound, making the confidence interval wider. Decreasing the confidence level decreases the error bound, making the confidence interval narrower.

In statistics, it is used to indicate the probability, with which the estimation of the location of a statistical parameter (e.g. an arithmetic mean) in a sample survey is also true for the population. In surveys, confidence levels of 90/95/99% are frequently used.

The width of a confidence interval depends upon two things:

a. **Diversity within the population of interest:** If all the population values were nearly the same, then we will have less variation. The estimate will be close to the actual population. Therefore, the confidence interval, in this case, will be small. But a more diverse population will lead to a more diverse sample. Different samples taken from the same population will vary more. There will be not be much assurance if the mean of the sample will be closer to the population mean. As such, the confidence interval will be large. So greater diversity in the population leads to a wider confidence interval.

b. **The sample size affects the width of the confidence interval:** With a small sample, we do not have much reference to base our inference. Small samples will be different from one another, causing a wider confidence interval. Whereas, in larger sample size, the effect of a few unusual values is evened out by the other values in the sample. Larger samples will be much similar. The effective sampling error is minimised with larger samples. Larger samples produce more information and estimates that a researcher is confident about, leading to a narrower confidence interval.

8. Of many examples of sensors used in healthcare, Oxygen Concentrator are one of the most important.

   Patient benefits: Oxygen concentrators are used to help patients with respiratory illnesses or lung disease. Patients who have difficulties in absorbing oxygen into the bloodstream benefit from the oxygen concentrator's enhanced response time, minimized oxygen waste, and improved portability.

   Sensors:

   Airflow – measures air and oxygen flow so the correct amount is delivered to the patient.

   Oxygen – measures and controls the oxygen concentration level of the air mixture delivered to the patient.

   Pressure – monitors patient breathing, detects when the patient inhales and exhales, allowing air/oxygen to be delivered more efficiently and effectively.

   Temperature – monitors and controls the temperature of the air delivered to the patient.

   Humidity – monitors and controls the moisture content of the air delivered to the patient.

   Magnetic – improves motor efficiency, reducing power consumption, noise and vibration.

9. A growing number of intelligent information-aware devices or smart devices, like sensors, actuators, smartphones, smart wristbands, tablets, devices based on readers' Radio Frequency Identification (RFID) and Machine-to-Machine (M2M), have led to the exponential growth

of generated-data volume. A smart device is any type of instrument or machine that has its own computing proficiency.

10. Data is collected through smart devices. These devices are managed through an Internet based technology known as Internet of Things (In short IoT.). Well-known examples of IoT devices include smart speakers like Amazon Alexa or Google Home, smartwatches like the Apple Watch, Internet-connected baby monitors, video doorbells, toys, etc.

    Sensor is a device that responds to a physical stimulus (such as heat, light, sound, pressure, magnetism, or a particular motion) and transmits a resulting impulse for automatically controlled actuators.

    Internal sensors collect data from IoT consumer devices, such as security systems, smart appliances, smart TVs, and wearable health meters. Data are collected from commercial devices, as well, including commercial security systems, traffic monitoring devices, and weather tracking systems.

    IoT connects a variety of sensors, alarms, cameras, lights, and microphones to provide 24/7 security—all of which can be controlled from a smart phone. For example, the Ring doorbell camera security system permits users to see, hear, and speak to visitors at their door via a computer, tablet, or mobile phone.

11. The statement "Data collection via sensors requires the least amount of human involvement" can be justified through several key points that highlight the automated and autonomous nature of sensor-based data collection:

    a. **Automation and Continuous Monitoring**: Sensors are designed to automatically collect data without the need for human intervention. Once installed and configured, they can continuously monitor and record data over extended periods. This contrasts with manual data collection methods, which require human presence and involvement at regular intervals.

    b. **Remote Data Collection**: Sensors can collect data from remote or hazardous locations where human access is difficult or dangerous. For instance, environmental sensors can monitor air and water quality in remote areas, while industrial sensors can track conditions inside machinery or at high elevations, all without needing humans to be physically present.

    c. **Real-Time Data Acquisition**: Sensors provide real-time data collection and transmission, enabling immediate access to data without delay. This eliminates the need for manual data entry or transcription, which can be time-consuming and prone to errors.

    d. **Precision and Consistency**: Sensors offer high levels of precision and consistency in data collection, reducing the variability and potential inaccuracies associated with human error. This leads to more reliable and accurate datasets.

    e. **Scalability**: Sensor networks can be scaled to cover large areas or multiple data points simultaneously. Managing a large-scale data collection effort manually would require

significant human resources, whereas sensors can perform this task efficiently with minimal oversight.

f. **Integration with Technology**: Sensors can be integrated with other technologies, such as Internet of Things (IoT) platforms, to enable automated data processing, analysis, and decision-making. This further reduces the need for human intervention by automating the entire data collection and analysis pipeline.

g. **Maintenance and Upkeep**: Modern sensors are designed to be low-maintenance, often requiring minimal human involvement for upkeep. Many sensors can self-calibrate or notify maintenance personnel when attention is needed, reducing the need for regular human check-ins.

By leveraging these capabilities, sensor-based data collection systems significantly minimize the need for human involvement, making them an efficient and reliable method for gathering data across various applications.

12. **The Advantages of Secondary Data**

Ease of access: Secondary data sources are easily accessible. The Internet has changed the way secondary research works. A lot of information is available just with the click of a mouse.

Economical: Secondary sources are usually completely free or at a minimal cost. It saves not only your money but also your effort. In comparison with primary research, where you have to design and conduct a whole primary study process from the beginning, secondary research allows you to collect data without spending much money.

Time-saving: It is a matter of a few Google searches to find a source of data, therefore secondary data is time-saving.

Reanalysing: It enables you to generate new insights from previous analysis. Reanalysing old data can bring unexpected new understandings and points of view or even new relevant inferences.

Longitudinal analysis: With secondary data, a longitudinal analysis can be done. This means the studies are performed spanning over a large period of time. A longitudinal study is a research design that involves repeated observations of the same variables (e.g., people) over short or long periods of time (i.e., using longitudinal data). This can help you determine varying trends. It allows you to compare data over a long period of time.

Versatility: Secondary data research can also be carried out by people who are not familiar with the different data collection methods. This means it can be collected by anyone.

A big source of data collection: It is the richest type of data available, with a diversity of sources and topics.

13. XML (Extensible MarkUp Language)

Web, e-business, and portable applications. XML provides a standard method to access information, making it easier for applications and devices of all kinds to use, store, transmit, and display data.XML stores data in plain text format. This provides a software and hardware

independent way of storing, transporting, and sharing data. XML also makes it easier to expand or upgrade to new operating systems, new applications, or new browsers, without losing data.

A programming language consists of grammar rules and its own vocabulary which is used to create computer programs. These programs instruct the computer to perform specific tasks. XML does not qualify to be a programming language as it does not perform any computation or algorithms. It is usually stored in a simple text file and is processed by special software that is capable of interpreting XML

14. Following are the characteristic features of XML:
    - It stands for Extensible Markup Language. It is a markup language similar to HTML.
    - It is not a replacement for HTML.
    - It was designed to store and transport data.
    - It is designed to be self-descriptive.
    - This language is a W3C recommendation.

Higher Order Thinking Skills:

1. Do it yourself.
2. Accessing higher-order thinking skills involves engaging in cognitive processes that go beyond mere memorization and recall of information. These skills include analysis, evaluation, synthesis, and creation, and they are essential for problem-solving, critical thinking, and innovation. Here are several strategies to access and develop higher-order thinking skills:
    a. **Ask Open-Ended Questions:**
        - Pose questions that require more than a yes or no answer. These questions should encourage exploration, explanation, and elaboration. For example, instead of asking "What is the capital of France?" ask "How does the geography of France influence its political and economic structure?"
    b. **Encourage Critical Thinking:**
        - Challenge assumptions and evaluate arguments. Encourage questioning the validity of sources, considering alternative perspectives, and identifying biases. For example, critically analyzing a news article by considering the author's perspective and potential biases.

Applied Project

Do it yourself.

5. Introduction to RStudio

Unsolved Exercise

Section A
(Objective Type Questions)

A. 1. c 2. c 3. a 4. c

B. 1. True 2. True 3. False 4. False

Section B
(Subjective Type Questions)

A. 1.

RStudio	R language
1. RStudio is a free and open-source Integrated Development Environment (IDE), used to develop programs for statistical computing using R language.	1. R language is a programming language. The R language is mainly used by data miners and statisticians for statistical analysis, graphics representation and visualising data
2. It provides a variety of robust tools and a platform that helps you develop programs easily.	2. R has become one of the most useful programming languages among the researchers, data analysts, and statisticians for retrieving, analyzing, visualising, and presenting data.

2. Learning R as a programming language for data science has several distinct characteristics that make it particularly suitable for this field. Here are the key characteristics:
   - It is an interpreted programming language which means it does not require a compiler.
   - It is a well-developed, simple, open-source, powerful, and highly extensible programming language.
   - It contains a variety of tools for plotting, viewing history, debugging and managing your workspace.
   - It has vast variety of built-in libraries and effective data handling facilities.
   - It is a machine-independent programming language, which means it supports the cross-platform operations. Therefore, it is used on many different operating systems.
   - It allows us to perform multiple complex mathematical operations using single command.
   - It is an object-oriented programming language.
3. RStudio is a powerful integrated development environment (IDE) specifically designed for working with the R programming language. Here are some compelling reasons why RStudio is widely used by data scientists and analysts:

a. **Scripting Ease**: RStudio makes it easy to write and manage R scripts. When you create a new script, the windows adjust automatically so you can view both your script and the results in the console. The tab key even provides syntax options as you write.

b. **Environment Management:** The "Environment" window in RStudio displays all stored objects, including data,

c. **Working Directory and File Access:** RStudio simplifies setting your working directory and accessing files on your computer. You can navigate folders, view files, and set the working directory seamlessly.

d. **Graphics Accessibility:** RStudio provides an intuitive graphics window where you can easily switch between plots, adjust plot sizes, and export or copy plots for other documents. While it's not as flexible as R's graphical devices, it covers most users' needs.

e. **Community Support:** RStudio has a large and active community, offering abundant resources, support, and packages.

f. **User-Friendly Interface:** RStudio's interface is beginner-friendly, making it an excellent choice for those learning R.

4. **Environment Tab:** It is the top-right portion of the RStudio application window. It shows a list of the variables and functions present in the current R program.

   Files Tab: It is the bottom-right portion of the RStudio application window, which allows you to open, delete, and rename files.

B. 1. Following are the steps to quit from RStudio session:

   Step 1: Click on the File menu. A drop-down menu appears.

   Step 2: Select the Quit Session option. The current session will quit.

   Following are the steps to close the RStudio:

   Step 1: Click on the File menu. A drop-down menu appears.

   Step 2: Select the Close option. The RStudio application window will close

2. Working directory is a folder from where RStudio reads and where it saves files. We can find the current working directory of RStudio by using the getwd() function, which stands for get working directory in the following way:

   getwd()

   "C:/Users/documents" ← current working directory

   It is possible to use the R language function setwd(), which stands for set working directory, to set the working directory. The path of the directory that you want to set as working directory is passed as a string parameter in the setwd() function.

   For Windows, the command might look like:

   ```
   setwd("c:/Documents/my/working/directory")
   ```

3. To create a new R script file, you can either select File → New → R Script or click on the icon and then select **R Script**, from the drop down menu, or simply press Ctrl + Shift + N.

6. Programming with R

Unsolved Exercise

Section A
(Objective Type Questions)

A.

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

B. 1. Numeric 2. Lists 3. Multiple elements
4. Character, Complex 5. Dataframe

C. 1. False 2. False 3. True 4. False 5. False

Section B
(Subjective Type Questions)

A. 1. Manipulate the R list elements

In R, we can *add, delete, or update* elements of the list. The addition or deletion of the elements can only be done at the end of the list. However, an update can be performed on any element in the list.

Let's now discuss how to manipulate the R list elements with the help of an example.

Example

Enter the following code snippet:

```
#Creating a list containing a vector, a matrix and a list.
data_list <- list(c("Jan","Feb","Mar"),
matrix(c(1,2,3,4,-1,9), nrow = 2),list("Red",12.3))
#Give names to the elements in the list.
names(data_list) <- c("Month", "Matrix", "Misc")
#Add an element at the end of the list.
data_list[4] <- "New element"
print(data_list[4])
#Remove the last element.
data_list[4] <- NULL
print(data_list[4])
```

```
#Update the 3rd Element.
data_list[3] <- "updated element"
```

2. Atomic vectors, are of six types:
   - logical
   - integer
   - double (for floating point numbers with double precision)
   - complex
   - character
   - raw

   Integer and double vectors are collectively known as numeric vectors.

B. 1.
```
#Create a multi-element vector containing the first ten natural numbers
natural_numbers <- 1:10
#Print the vector
print(natural_numbers)
```

2. (a) [8,4,7,5]

   (b) [-4,2,-7,5]

   (c) [12,3,0,0]

   (d) Division by zero is undefined

3.
```
#Create two 2x3 matrices.
matrix1 <- matrix(c(2, 9, -1, 4, 2, 5), nrow = 2)
print(matrix1)
matrix2 <- matrix(c(4, 3, 0, 9, 3, 4), nrow = 2)
print(matrix2)
#Multiply the matrices.
result <- matrix1 * matrix2
cat("Result of multiplication","\n")
print(result)
#Divide the matrices
result <- matrix1 / matrix2
cat("Result of division"c,"\n")
print(result)
```

Higher Order Thinking Skills:

Do it yourself.

7. Coding for Data Science Visualisation using R-Studio

Unsolved Exercise

Section A
(Objective Type Questions)

A.

1. c	2. a	3. d	4. c	5. d
6. d	7. b	8. d	9. a	10. b

B.

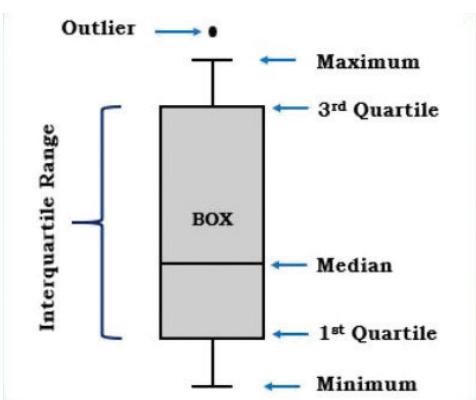
1. Trim	2. Trend	3. Plot()	4. Pie	5. xlab

C.

1. False	2. True	3. False	4. True	5. False

Section B
(Subjective Type Questions)

A.

1. Stacked bar chart extends the standard bar chart from looking at numeric values across one categorical variable to two. Each bar in a standard bar chart is divided into a number of sub-bars stacked end to end, each one corresponding to a level of the second categorical variable. Stacked Bar Graphs are used to show how a larger category is divided into smaller categories and what the relationship of each part has on the total amount.
2. The main advantages of a histogram are its simplicity and versatility. It can be used in many different situations to offer an insightful look at frequency distribution. For example, it can be used in sales and marketing to develop the most effective pricing plans and marketing campaigns.
3. A trend line is a straight line that best represents the points on a scatterplot. The trend line may go through some points but need not go through them all. The trend line is used to show the pattern of the data. The Trendline is the best line that fits through the points.
4. 

5. **Mean**

   The mean (or average) is the most popular measure of central tendency. It can be used with both discrete and continuous data, though it is mostly used with continuous data. The mean is equal to the sum of all the values in the data set divided by the number of values in that data set.

   Median

   The median is the middle value in a data set that has been arranged in order of magnitude. The median, unlike the mean, is less affected by outliers and skewed data

   Mode

   The mode is the most frequently occurring value in a data set. It represents the highest bar in a bar chart or histogram. You can, therefore, usually think of mode as being the most popular option among the given options.

6. The median is the middle value in a data set that has been arranged in order of magnitude. The median, unlike the mean, is less affected by outliers and skewed data. Suppose we have to calculate the median for the data below:

   60 50 85 51 32 11 51 50 82 43 91

   We first need to rearrange that data into an order of magnitude (smallest first):

   11 32 43 50 50 51 51 60 82 85 91

   The median value '51' is represented by a middle mark. It is the middle mark since there are five scores that lie before it and five scores that lie after it. Here, you have an odd number of values, so you can easily find the middle value, but if you have an even number of values, we can take the middle two values and find their average.

   Let us look at the example below:

   60 50 85 51 32 11 51 50 82 43

   We again rearrange that data into an order of magnitude (smallest first):

   11 32 43 50 50 51 51 60 82 85

   The average of the 5th and 6th values in this data set gives you the value for your median, which is 50.5.

B. 1.
```
# Creating the data for the chart
H <- c(31,28,31,30,31,30,31,31,30,31,30,31)
M <- c("Jan","Feb","Mar","Apr","May","Jun","Jul","Aug","Sept","Oct
","Nov","Dec")
# Giving the chart file a name
png(file = "barchart_months_days.png")
# Plotting the bar chart
```

```
barplot(H,names.arg=M,xlab=" year 2021 ",ylab="number of
days",col="orange",
main="Monthwise Days chart",border="red")
# Saving the file
dev.off()
```

2.
```
Create the input vectors
months <- c("Jan", "Feb", "Mar", "Apr", "May", "Jun")
store_A <- c(32, 34, 35, 28, 37, 45)
store_B <- c(28, 22, 34, 36, 25, 20)
colors <- c("green", "orange")
#Create the matrix correctly
sales <- matrix(c(store_A, store_B), ncol = 2, byrow = TRUE)
#Create side-by-side bar chart
barplot(sales,
        beside = TRUE,
        col = colors,
        legend.text = c("Store A", "Store B"),
        names.arg = months,
        main = "Washing Machines Sold by Store A and Store B (Jan-
        Jun 2021)",
        xlab = "Month",
        ylab = "Number of Washing Machines Sold")
```

3.
```
vector <- c(7,6,3,8,6,5,7,1,9,2,7,6)
#Mean
mean_value <- mean(vector)
cat("Mean:", mean_value, "\n")
#Median
median_value <- median(vector)
cat("Median:", median_value, "\n")
#Mode
mode_value <- as.numeric(names(sort(table(vector), decreasing = TRUE)
[1]))
cat("Mode:", mode_value,"\n")
```

4. **Pie chart/Graph**

   A pie chart is a circular statistical graphic that is divided into slices to illustrate numerical proportions. It depicts a special chart that uses "pie slices" where each sector shows the relative sizes of data. A circular chart, cut in the form of radii into segments describing relative frequencies or magnitudes, is also known as a circle graph.

   Bar chart/Graph

   A bar chart is a pictorial representation of data that presents categorical data with rectangular bars whose heights or lengths are proportional to the values that they represent. In other words, it is the pictorial representation of a dataset. These data sets contain the numerical values of variables that represent the length or height.

   Line charts

   Line charts are usually used to identify trends in data. A line graph is a chart that is used to display information in the form of a series of data points. It uses points and lines to represent change over time. Line graphs are created by plotting different points on their X and Y coordinates, then connecting them with a line from beginning to end. This line graph represents different values as it can move up and down based on the suitable variable.

   Histogram

   A histogram is a graphical representation that organises a group of data points into user-specified ranges. Similar in appearance to a bar graph, the histogram condenses a data series into an easily interpreted visual by taking many data points and grouping them into logical ranges or bins.

   Scatter plots

   Scatter plots are dispersion graphs built to represent the data points of variables (generally two, but can also be three). The main use of a scatter plot in R is to visually check if there is some relation between numeric variables.

   Scatterplots show many points plotted in the Cartesian plane. Each point represents the values of two variables. One variable is chosen in the horizontal axis and another in the vertical axis.

   Box plot

   A box plot is a graphical technique for summarising a set of data on an interval scale. Boxplots are extensively used in descriptive data analysis. Boxplots are a measure of how well distributed the data is in a data set. It divides the data set into three quartiles.

5. The basic syntax for creating pie chart in R is:

```
pie(x, labels, radius, main, col, clockwise)
```

Parameters:

x: This parameter is a vector that contains the numeric values that are used in the pie chart.

labels: This parameter gives the description of the slices in the pie chart.

radius: This parameter is used to indicate the radius of the circle in the pie chart(a value between -1 and +1).

main: This parameter represents the title of the pie chart.

clockwise: This parameter contains the logical value which indicates whether the slices are drawn clockwise or in an anti-clockwise direction.

col: This parameter gives colours to the pie in the graph.

6. R language supports two ways to plot a bar plot:
   i. Vertical Bars
   ii. Horizontal Bars

   Horizontal bar chart. The barplot() function has a field named "horiz" keep that field value "TRUE" to create the bar plot horizontally. By default the Bar Chart remains vertical.

Higher Order Thinking Skills

1. In R, the inbuilt dataset Orange contains growth data of orange trees. To create a scatter plot:

   • Using the normal plot function in R:

```
# Load the Orange dataset
data(Orange)
# Normal scatter plot using plot()
plot(Orange$age, Orange$circumference,
     xlab = "Age (days)",
     ylab = "Circumference (mm)",
     main = "Scatter Plot: Age vs Circumference")
```

   • Using ggplot2 library:

```
# Load required library
library(ggplot2)
# Scatter plot using ggplot2
ggplot(Orange, aes(x = age, y = circumference)) +
  geom_point() +
  labs(title = "Scatter Plot: Age vs Circumference",
       x = "Age (days)",
       y = "Circumference (mm)")
```

2. Plotting a histogram for Circumference of Orange dataset:

   A histogram shows how the circumference values are distributed.

   - Using base R's hist() function:

```
# Histogram of Circumference
hist(Orange$circumference,
     xlab = "Circumference (mm)",
     main = "Histogram of Orange Circumference",
     col = "lightblue",
     border = "black")
```

3. Plotting a line chart for Time vs Demand using BOD dataset:

   The BOD dataset records biochemical oxygen demand over time.

```
# Load the BOD dataset
data(BOD)
# Line chart for Time vs Demand
plot(BOD$Time, BOD$Demand,
     type = "o",
     col = "blue",
     xlab = "Time (days)",
     ylab = "Demand (mg/L)",
     main = "Line Chart: Time vs Demand")
```

Applied Project

Do it yourself.