CBSE | DEPARTMENT OF SKILL EDUCATION

DATA SCIENCE (SUBJECT CODE 419)
MARKING SCHEME FOR CLASS X (SESSION 2022-2023)

Max. Time: 2 Hours   **Max. Marks: 50**

General Instructions:

1. Please read the instructions carefully.
2. This Question Paper consists of **21 questions** in two sections – Section A & Section B.
3. Section A has Objective type questions whereas Section B contains Subjective type questions.
4. **Out of the given (5 + 16 =) 21 questions, a candidate has to answer (5 + 10 =) 15 questions in the allotted (maximum) time of 2 hours.**
5. All questions of a particular section must be attempted in the correct order.
6. **SECTION A - OBJECTIVE TYPE QUESTIONS (24 MARKS):**
    i. This section has 05 questions.
    ii. There is no negative marking.
    iii. Do as per the instructions given.
    iv. Marks allotted are mentioned against each question/part.
7. **SECTION B – SUBJECTIVE TYPE QUESTIONS (26 MARKS):**
    i. This section contains 16 questions.
    ii. A candidate has to do 10 questions.
    iii. Do as per the instructions given.
    iv. Marks allotted are mentioned against each question/part.

SECTION A: OBJECTIVE TYPE QUESTIONS

Q. No.	QUESTION	Source Material (NCERT/PSSCIVE/ CBSE Study Material)	Unit/ Chap. No.	Page no. of source material	Marks
Q. 1	**Answer any 4 out of the given 6 questions on Employability Skills (1 x 4 = 4 marks)**				
i.	c) Adversity, Belief, Consequences				1
ii.	d) Remember				1
iii.	b) Information Communication Technology				1
iv.	d) Hover, Point and click, Drop and Drag, Double Click				1
v.	a) Diligent				1
vi.	a) Both i and iii				1
Q. 2	**Answer any 5 out of the given 6 questions (1 x 5 = 5 marks)**				
i.	Feature based	CBSE	1	3	1
ii.	Simple average	CBSE	1	6	1
iii.	Middle	CBSE	1	7	1
iv.	Average	CBSE	1	9	1
v.	Continuous Data	CBSE	2	21	1
vi.	Data should never institutionalize unfair biases	CBSE	5	56	1

Q. 3	Answer any 5 out of the given 6 questions (1 x 5 = 5 marks)				
i.	One to One join	CBSE	4	42	**1**
ii.	Increases	CBSE	3	36	**1**
iii.	Modularity bias	CBSE	3	32	**1**
iv.	Standard deviation	CBSE	1	9	**1**
v.	All of the mentioned	CBSE	5	57, 58	**1**
vi.	Data governance	CBSE	5	56	**1**
Q. 4	Answer any 5 out of the given 6 questions (1 x 5 = 5 marks)				
i.	Population mean	CBSE	4	45	**1**
ii.	Lies on the mean	CBSE	4	45	**1**
iii.	First quartile from the Third quartile	CBSE	4	47	**1**
iv.	Variability	CBSE	2	29	**1**
v.	all of the mentioned	CBSE	2	28	**1**
vi.	p percent of the observations fall at or below it	CBSE	4	52	**1**
Q. 5	Answer any 5 out of the given 6 questions (1 x 5 = 5 marks)				
i.	2.60	CBSE	4	52	**1**
ii.	6014.29	CBSE	1	16	**1**
iii.	2.89	CBSE	1	16	**1**
iv.	4/52	CBSE	2	27	**1**
v.	Can not be determined from the given information	CBSE	2	28	**1**
vi.	It is the collection of several experiments	CBSE	3	38	**1**

SECTION B: SUBJECTIVE TYPE QUESTIONS

Q. No.	QUESTION	Source Material (NCERT/PSSCIVE/ CBSE Study Material)	Unit/ Chap. No.	Page no. of source material	Marks
Answer any 3 out of the given 5 questions on Employability Skills in 20 – 30 words each (2 x 3 = 6 marks)					
Q. 6	a. Self-management is our ability to manage our behaviors, thoughts, and emotions in a conscious and productive way. b. the ability to adjust to different conditions or circumstances c. Time management is the process of organizing and planning how to divide your time between different activities.				**2**
Q. 7	Self motivated person always shows Commitment to goals. Example :: You set a goal to run for 20 minutes, but at 15 minutes you're exhausted. But you keep going and pushing yourself.				**2**
Q. 8	**Antivirus software is designed to detect, remove and prevent malware infections on a device or network**. Though specifically created to eliminate viruses, antivirus software can also aid against spyware, adware and other malicious software.				**2**
Q. 9	Need, want, demand seems to be very closely related words, however they all differ with respect to their meaning.				**2**

	Need - necessity. Something that is required to survive or to sustain. For example, food, water, shelter etc. **Want** - desires. Things without which we can survive, but we need them for higher satisfaction. For example, i want iphone 6S, though I can do without it. (Human wants are unlimited! As one want is satisfied, a new one is created.) **Demand** - willing and able to buy. Things that you want to buy/consume and your pocket allows its consumption, i.e you can afford it, thus you demand that thing. For example, I have Rs. 1000 in my pocket and I want to see a movie whose ticket is worth Rs 500, since I want it and I can afford it, I demand a movie ticket.				
Q. 10	An entrepreneur needs the society to purchase whatever they offer for them to survive. The society provides raw materials needed by an entrepreneur – The society provides raw materials to entrepreneurs for production of the goods that the same society needs.				**2**

Answer any 4 out of the given 6 questions in 20 – 30 words each (2 x 4 = 8 marks)

Q. 11	1. Grading tests 2. Evaluating a survey 3. Weather forecasting	CBSE	1	10, 11	**2**
Q. 12	Discrete Data is the data that takes only specified values. For example, if you give a test, you can either pass or fail. So, data is discrete in this case as it has only two specified outcomes. Continuous Data is the data that can take any value within a given range. This range can be either finite or infinite. For example, depth of an ocean, weight of a person or length of a road.	CBSE	2	21	**2**
Q. 13	Selection bias usually occurs when a model itself influences the creation of data that is used to train it. Selection bias is said to occur when the sample data that is gathered is not representative of the true future population of cases that the model will see.	CBSE	3	33	**2**
Q. 14	In Data Science, data merging is the process of combining two or more data sets into a single data frame. This process is necessary when we have raw data stored in multiple files or data tables, that we want to analyze all in one go.	CBSE	4	42	**2**

Q. 15	Once you are done with the job and you no longer need the user data, you can go ahead and clean out the data from the memory. Even while storing the data in your device, you can encrypt the data to make sure that even in the case of a data leak, hackers are not able to read your data. You can also format the computer drive/hardisk where the client confidential data was stored for a clean discarding.	CBSE	5	57	**2**
Q. 16	1. It gives us an opportunity to calculate the probability of a value occurring within a normal distribution. 2. Z-score allows us to compare two values that are from the different samples.	CBSE	4	46	**2**

Answer any 3 out of the given 5 questions in 50– 80 words each (4 x 3 = 12 marks)

Q. 17	1. Row based subsetting - In this method of subsetting, we take some rows from the top or bottom of the table. 2. Column based subsetting - Sometimes the original data set may contain large number of columns and all of them may not be necessary to perform the analysis. We then select specific columns from the dataset. 3. Data based subsetting - To subset the data based on specific data we use data-based subsetting.	CBSE	1	4	**4**
Q. 18	**1.** Formulate Statistical Investigative Questions - This can also be called as anticipating variability while beginning with the process. Formulating statistical investigative questions that anticipate variability leads to productive investigations. **2.** Collect/Consider the Data - This step can be called as acknowledging variability while designing for differences. Data collection designs must acknowledge variability in data. **3.** Analyze the Data - This step can also be called as accounting of variability while the distributions. When we analyze the data, we try to understand its variability. **4.** Interpret the Results - This step can also be called as allowing for variability while looking beyond the data. Statistical interpretations are made in the presence of variability and must take variability into account.	CBSE	2	21	**4**

Q. 19	The Central Limit Theorem states that distribution of sample approaches a normal distribution as the sample size gets larger irrespective of what is the shape of the population distribution. The Central Limit Theorem is a statistical theory stating that given a significantly large sample size from a population with finite variance, the mean of all samples from same set of population will be roughly equal to the mean of the population. Some practical implementations of the Central Limit Theorem include: 1. Voting polls estimate the count of people who support a particular election candidate. The results of news channels that come with confidence intervals are all calculated using the Central Limit Theorem. The Central Limit Theorem can also be used to calculate the mean family income for a specific region.	CBSE	3	34, 36	**4**
Q. 20	A percentile can be defined as the percentage of the total ordered observations at or below it. Therefore, pth percentile of a distribution is the value such that p percentage of the ordered observation falls at or below it. Quartiles of dataset partitions the data into four equal parts, with one-fourth of the data values in each part. The total of 100% is divided into four equal parts: 25%, 50%, 75% & 100%. Since the median is defined as the middlemost value in the observation, the median will have 50% of the observations at or below it. Just like quartiles, we have deciles. While quartiles sort the data into four quarters, deciles sort the data into ten equal parts: the 10th, 20th, 30th, 40th, 50th, 60th, 70th, 80th, 90th,100th. The higher the place in the decile ranking, the higher is the overall ranking.	CBSE	4	46, 47	**4**

| Q. 21 | 1. Protect Your Customer Privacy does not always mean confidentiality because private data may need to be audited based on the relevant requirements. 2. The private information that is shared should always be handled with confidentiality Third party companies share sensitive data, either financial, location related or medical. 3. Customers should always have a clear view of how their data is getting used or traded and should have the authority to manage the flow of their confidential information across enormous, third party systems. 4. Data should never interfere with human will: Data analytics can average out and at times, even discover who we are even before we make up our mind. 5. Data should never institutionalize unfair biases like sexism or racism. Analytical systems can absorb unconscious biases in a crowd and boost them with the help of training samples. | CBSE | 5 | 56 | **4** |