

Marking Scheme
Strictly Confidential
(For Internal and Restricted use only)
Secondary School Examination, 2026 (Xth)
SUBJECT NAME : DATA SCIENCE (Q.P. CODE /Set No. 419/106)

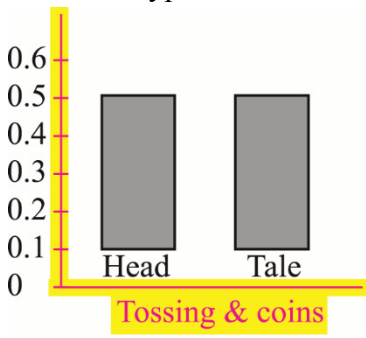
General Instructions: -

1	You are aware that evaluation is the most important process in the actual and correct assessment of the candidates. A small mistake in evaluation may lead to serious problems which may affect the future of the candidates, education system and teaching profession. To avoid mistakes, it is requested that before starting evaluation, you must read and understand the spot evaluation guidelines carefully.
2	“Evaluation policy is a confidential policy as it is related to the confidentiality of the examinations conducted, evaluation done and several other aspects. Its leakage to public in any manner could lead to derailment of the examination system and affect the life and future of millions of candidates. Sharing this policy/document to anyone, publishing in any magazine and printing in Newspaper/Website, etc. may invite action under various rules of the Board and IPC.”
3	Evaluation is to be done as per instructions provided in the Marking Scheme. It should not be done according to one’s own interpretation or any other consideration. Marking Scheme should be strictly adhered to and religiously followed. However, while evaluating, answers which are based on latest information or knowledge and/or are innovative, they may be assessed for their correctness otherwise and due marks be awarded to them. In Class-X, while evaluating two competency-based questions, please try to understand given answer and even if reply is not from marking scheme but correct competency is enumerated by the candidate, due marks should be awarded.
4	The Marking scheme carries only suggested value points for the answers. These are in the nature of Guidelines only and do not constitute the complete answer. The students can have their own expression and if the expression is correct, the due marks should be awarded accordingly.
5	The Head-Examiner must go through the first five answer books evaluated by each evaluator on the first day, to ensure that evaluation has been carried out as per the instructions given in the Marking Scheme. If there is any variation, the same should be zero after deliberation and discussion. The remaining answer books meant for evaluation shall be given only after ensuring that there is no significant variation in the marking of individual evaluators.
6	Evaluators will mark (✓) wherever answer is correct. For wrong answer CROSS ‘X’ be marked. Evaluators will not put right (✓) while evaluating which gives an impression that answer is correct and no marks are awarded. This is most common mistake which evaluators are committing.
7	If a question has parts, please award marks on the right-hand side for each part. Marks awarded for different parts of the question should then be totaled up and written in the left-hand margin and encircled. This may be followed strictly.
8	If a question does not have any parts, marks must be awarded in the left-hand margin and encircled. This may also be followed strictly.
9	If a student has attempted an extra question, answer of the question deserving more marks should be retained and the other answer scored out with a note “Extra Question” .
10	No marks to be deducted for the cumulative effect of an error. It should be penalized only once.
11	A full scale of marks _____ (example 0 to 80/70/60/50/40/30 marks as given in Question Paper) has to be used. Please do not hesitate to award full marks if the answer deserves it.

12	Every examiner has to necessarily do evaluation work for full working hours i.e., 8 hours every day and evaluate 20 answer books per day in main subjects and 25 answer books per day in other subjects (Details are given in Spot Guidelines). This is in view of the reduced syllabus and number of questions in question paper.
13	<p>Ensure that you do not make the following common types of errors committed by the Examiner in the past :-</p> <ul style="list-style-type: none"> • Leaving answer or part thereof unassessed in an answer book. • Giving more marks for an answer than assigned to it. • Wrong totaling of marks awarded on an answer. • Wrong transfer of marks from the inside pages of the answer book to the title page. • Wrong question wise totaling on the title page. • Wrong totaling of marks of the two columns on the title page. • Wrong grand total. • Marks in words and figures not tallying/not same. • Wrong transfer of marks from the answer book to online award list. • Answers marked as correct, but marks not awarded. (Ensure that the right tick mark is correctly and clearly indicated. It should merely be a line. Same is with the X for incorrect answer.) • Half or a part of answer marked correct and the rest as wrong, but no marks awarded.
14	While evaluating the answer books if the answer is found to be totally incorrect, it should be marked as cross (X) and awarded zero (0) Marks.
15	Any unassessed portion, non-carrying over of marks to the title page, or totaling error detected by the candidate shall damage the prestige of all the personnel engaged in the evaluation work as also of the Board. Hence, in order to uphold the prestige of all concerned, it is again reiterated that the instructions be followed meticulously and judiciously.
16	The Examiners should acquaint themselves with the guidelines given in the “Guidelines for Spot Evaluation” before starting the actual evaluation.
17	Every Examiner shall also ensure that all the answers are evaluated, marks carried over to the title page, correctly totaled and written in figures and words.
18	The candidates are entitled to obtain photocopy of the Answer Book on request on payment of the prescribed processing fee. All Examiners/Additional Head Examiners/Head Examiners are once again reminded that they must ensure that evaluation is carried out strictly as per value points for each answer as given in the Marking Scheme.

MARKING SCHEME
DATA SCIENCE (Subject Code-419)
(PAPER CODE: 106) (M1060419)

Q.No.	EXPECTED OUTCOMES/VALUE POINTS		Marks
	SECTION – A : OBJECTIVE TYPE QUESTIONS		
1.	Answer any 4 out of the given 6 questions on Employability Skills.		1X4=4
(i)	(A) Feedback	Unit-1 Page.4	1
(ii)	(D) To promote personal growth and decision making	Unit-2 Page.43	1
(iii)	(C) Worrying about stressors	Unit-2 Page.42	1
(iv)	(C) Temporary files	Unit-3 Page.77	1
(v)	(A) Innovation	Unit-4 Page.93	1
(vi)	(B) Environmental degradation	Unit-5 Page.105	1
2.	Answer any 5 out of the given 6 questions.		1x5=5
(i)	Data-based subsetting is used when : (A) When all the rows are selected (B) Subsetting is based on specific data conditions (C) The entire dataset is verified (D) Columns and rows are randomly selected	Unit-1 Page.3	1
Ans.	(B) Subsetting is based on specific data conditions (1 mark for correct answer)		
(ii)	What is the main purpose of the statistical problem-solving process ? (A) To prepare detailed report (B) To collect online data (C) To answer investigative questions using data (D) To visualize charts	Unit-2 Page.18	1
Ans.	(C) To answer investigative questions using data (1 mark for correct answer)		
(iii)	What is bias in data science ? (A) A type of data visualization (B) The process of representing data (C) A method of data analysis (D) A deviation from the expected outcome in data	Unit-3 Page.27	1
Ans.	(D) A deviation from the expected outcome in data (1 mark for correct answer)		

(iv)	Assertion (A): Percentiles help compare scores across different tests. Reason (R) : Percentiles represent the proportion of values at or below a certain data point. (A) Both (A) and (R) are true and (R) is the correct explanation of (A) (B) Both (A) and (R) are true, but (R) is not the correct explanation of (A) (C) (A) is true, but (R) is false (D) (A) is false, but (R) is true	Unit-4 Page.40	1
Ans.	(A) Both (A) and (R) are true and (R) is the correct explanation of (A) (1 mark for correct answer)		
(v)	State true or false Data collected for analysis should never interfere with human will.	Unit-5 Page.49	1
Ans.	True (1 mark for correct answer)		
(vi)	While storing data in a storage device, it is a good practice to _____ the data, so that hackers cannot read your data. (A) Copy (B) take photographs of (C) encrypt (D) format	Unit-5 Page.49	1
Ans.	(C) encrypt (1 mark for correct answer)		
3.	Answer any 5 out of the given 6 questions		1x5=5
(i)	Assertion (A) : Median is a more effective measure of central tendency where there are outliers in the data set. Reason (R) : Median is the average of all the values in a dataset. (A) Both (A) and (R) are true and (R) is the correct explanation of (A) (B) Both (A) and (R) are true, but (R) is not the correct explanation of (A) (C) (A) is true, but (R) is false (D) (A) is false, but (R) is true	Unit-1 Page.6	1
Ans.	(C) (A) is true, but (R) is false (1 mark for correct answer)		
(ii)	Name the type of distribution represented by the following graph. 	Unit-2 Page.17	1
Ans.	Uniform Distribution (1 mark for correct answer)		

(iii)	A researcher only sees the results they want to see in their study. This is an example of : (A) Confirmation Bias (B) Linearity Bias (C) Recall Bias (D) Selection Bias	Unit-3 Page.28	1																
Ans.	(A) Confirmation Bias (1 mark for correct answer)																		
(iv)	The Z-score represents : (A) The most common value of data (B) The percentage of values below a score (C) The number of standard deviations a data point is from the mean (D) The total sum of the data points	Unit-4 Page.39	1																
Ans.	(C) The number of standard deviations a data point is from the mean (1 mark for correct answer)																		
(v)	In a dataset, Q1=20, Q3=50. What is the IQR ? (A) 30 (B) 20 (C) 50 (D) 70	Unit-4 Page.41	1																
Ans.	(A) 30 (1 mark for correct answer)																		
(vi)	Which of the following is not an ethical guideline for data analysis ? (A) Ensure accurate and reliable sources of data (B) Maintain privacy and confidentiality (C) Use only relevant data for analysis (D) Data Analysts may select data based on their personal preferences.	Unit-5 Page.48	1																
Ans.	(D) Data Analysts may select data based on their personal preferences. (1 mark for correct answer)																		
4.	Answer any 5 out of the given 6 questions.		1x5=5																
(i)	A survey was done on number of people consuming tea or coffee during the day and the data was put together in the format given below : <table border="1" data-bbox="236 1368 790 1525"> <thead> <tr> <th>Beverage</th><th>Morning</th><th>Evening</th><th>Total</th></tr> </thead> <tbody> <tr> <td>Coffee</td><td>43</td><td>56</td><td>99</td></tr> <tr> <td>Tea</td><td>60</td><td>72</td><td>132</td></tr> <tr> <td>Total</td><td>103</td><td>128</td><td>231</td></tr> </tbody> </table> What is the name of this type of table ? (A) Subset Table (B) Two-way frequency table (C) Two-way relative frequency table (D) Standard Deviation Table	Beverage	Morning	Evening	Total	Coffee	43	56	99	Tea	60	72	132	Total	103	128	231	Unit-1 Page.3	1
Beverage	Morning	Evening	Total																
Coffee	43	56	99																
Tea	60	72	132																
Total	103	128	231																
Ans.	(B) Two-way frequency table (1 mark for correct answer)																		
(ii)	Which of the following is NOT a statistical investigative question ? (A) How fast can a ten-year old child run ? (B) Do children who have proper breakfast run faster ? (C) How does sleep affect the performance of a child ? (D) What time did the child take to run 200m ?	Unit-2 Page.19	1																
Ans.	(D) What time did the child take to run 200m ? (1 mark for correct answer)																		

(iii)	A survey is conducted to find the average height of students in a city. Instead of measuring all students, a researcher takes samples of size 40 from different schools and calculates the individual mean of each sample. Thereafter the mean of these individual sample means is calculated. It is noticed that the histogram of sample mean heights of students resembles the normal distribution. Which statistical concept is demonstrated here ? (A) Probability (B) Linear Bias (C) Central Limit Theorem (D) Data Evaluation	Unit-3 Page.30	1
Ans.	(C) Central Limit Theorem (1 mark for correct answer)		
(iv)	Which of the following correctly matches quartiles ? (A) Q1-25%, Q2-50%, Q3-75%, Q4-100% (B) Q1-10%, Q2-20%, Q3-30%, Q4-50% (C) Q1-33%, Q2-66%, Q3-100%, Q4-99% (D) Q1-50%, Q2-75%, Q3-100%, Q4-25%	Unit-4 Page.41	1
Ans.	(A) Q1-25%, Q2-50%, Q3-75%, Q4-100% (1 mark for correct answer)		
(v)	The value of this statistical term that describes the position of a point in terms of its distance from the mean, when it is measured in standard deviation units. Name this term.	Unit-4 Page.39	1
Ans.	Z-score (1 mark for correct answer)		
(vi)	What is the main aim of a data governance framework ? (A) To increase profit margins (B) To standardize, integrate, protect and store data (C) To delete duplicate records only (D) To design data visualizations	Unit-5 Page.48	1
Ans.	(B) To standardize, integrate, protect and store data (1 mark for correct answer)		
5.	Answer any 5 out of the given 6 questions.		1x5=5
(i)	Which of the following is a real-life application of standard deviation ? (A) Plotting data graphs in Excel (B) Calculating average rainfall only (C) Measuring students' performance spread in a test (D) Finding the number of columns in a spreadsheet	Unit-1 Page.9	1
Ans.	(C) Measuring students' performance spread in a test (1 mark for correct answer)		
(ii)	"Data collection designs must acknowledge variability in data. Few methods are used to reduce and detect variability in data". Which of the following method can be used ? (A) Statistical Process Control (B) Probability (C) Distribution (D) Event	Unit-2 Page.19	1
Ans.	(A) Statistical Process Control (1 mark for correct answer)		

(iii)	Assertion (A) : Biased data can lead to inaccurate predictive models. Reason (R) : The predictive models consider only the data that is fed into the system for training it. (A) Both (A) and (R) are true and (R) is the correct explanation of (A). (B) Both (A) and (R) are true, but (R) is not the correct explanation of (A). (C) (A) is true, but (R) is false. (D) (A) is false, but (R) is true.	Unit-3 Page.27-28	1
Ans.	(A) Both (A) and (R) are true and (R) is the correct explanation of (A). (1 mark for correct answer)		
(iv)	Which of the following is true about the process of data merging ? (A) All data sources being merged are always grouped in similar manner. (B) There is a lot of difference between multiple data sources. (C) All data sources are created with same objective and at same time. (D) No correction is required on the data merged from different data sources.	Unit-4 Page.36	1
Ans.	(B) There is a lot of difference between multiple data sources. (1 mark for correct answer)		
(v)	The following statements are with respect to one-to-one join : (i) Each row in one table is linked to a single row in another table. (ii) 'A student can register in multiple courses' is a valid example of this join. (iii) The linking between single row of two tables is done using a 'key column'. (iv) The key field used to link the tables is designed to contain unique values. (v) 'A student can have only one student Id' is a valid example of this join. Which of the given statement are correct ? (A) (i), (ii) and (iii) (B) (i), (iii), (iv) and (v) (C) (ii), (iv) and (v) (D) (ii), (iii), (iv) and (v)	Unit-4 Page.38	1
Ans.	(B) (i), (iii), (iv) and (v) (1 mark for correct answer)		
(vi)	Private data collected from a person with consent should : (A) Be freely shared with anyone who needs it (B) Be made available for data analysis (C) Always be handled with confidentiality (D) Never be audited under any circumstance	Unit-5 Page.48-49	1
Ans.	(C) Always be handled with confidentiality (1 mark for correct answer)		
SECTION – B : SUBJECTIVE TYPE QUESTION			
Answer any 3 out of the given 5 questions on Employability Skills Answer each question in 20-30 words.			2x3=6
6.	Linguistic barriers :-	Unit-1 Page.22	2
Ans.	The inability to communicate using a language is known as language barrier to communication. They are the most common barriers, which cause misunderstandings and misinterpretations between people.		

7. Ans.	Steps :- (1) Understand your emotions - Observe your behaviour (2) Rationalise - Not to take decision abruptly (3) Practise – Do meditation and yoga to keep yourself calm	Unit-2 Page.44	2
8. Ans.	(a) Sometimes we get mails from companies who are advertising a product or trying to attract customers to their website such mails are called SPAM. (b) We should never respond to SPAM and delete it on regular basis.	Unit-3 Page.7	1 1
9. Ans.	(i) every business needs to be unique and special (ii) A person needs lot of money to start a business (iii) person having a big business is an entrepreneur (iv) entrepreneurs are born not made (any 2)	Unit-4 Page.**	2
10. Ans.	(1) Save energy by switching off lights and fans when not in use. (2) Use natural light as much as possible. (3) Use energy efficient lights (LED bulbs) and appliances.	Unit-5 Page.**	2
Answer any 4 out of the given 6 questions in 20-30 words each			2x4=8
11.	Consider the following dataset : [5, 10, 2, 1, 20, 6, 15] Find the mean and median.		2
Ans.	Arrange the data in ascending order : [1, 2, 5, 6, 10, 15, 20] Mean = $(1+2+5+6+10+15+20)/7 = 59/7 = 8.42$ Median = Middle value = 6 (1 mark each for correct mean and median)	Unit-1 Page.7	
12.	“The data can be discrete or continuous.” Give one difference between discrete and continuous data. Also give one example of each.		2
Ans.	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. (1 mark for each correct difference ½ mark each for example)	Unit-2 Page.18	
13.	Define selection bias. When does it occur ?		2
Ans.	Selection bias is a type of bias that usually occurs when a model itself influences the creation of data that is used to train it. It occurs when the sample data that is gathered is not representative of the true future population of cases that the model will see. (1 mark for each correct part)	Unit-3 Page.28	

14.	What do you infer if the value of Z-score is positive or negative ? Give example to support your answer.		2
Ans.	<p>If the value of Z-score is positive, the raw score is higher than the mean average while a negative Z-score tells us that the raw score is below the mean average.</p> <p>(½ mark for each correct inference 1 mark for correct example)</p>	Unit-4 Page.40	
15.	Explain Many-to-Many join with an example.		2
Ans.	<p>A many-to-many relationship is said to occur when multiple records in one table are related to multiple records of other table. For example, a many-to-many relationship exists between students and courses. A student can register for multiple courses and a course can have multiple students.</p> <p>(1 mark for explanation 1 mark for example) (Any correct example may be marked)</p>	Unit-4 Page.38	
16.	<p>A company is analyzing customer purchase data. The analyst decides to discard all records from customers under 18 without justification, to make the dataset smaller.</p> <p>(a) Identify the ethical issue in this case. (b) Should the company soft delete the data while discarding it ? Why/Why not ?</p>		2
Ans.	<p>(a) Data was discarded without justification, leading to bias and lack of transparency. (b) No. Because if we soft delete the data gets deleted from its original location but is stored in a temporary folder.</p> <p>(1 mark for each correct part)</p>	Unit-5 Page.49, 50	
Answer any 3 out of the given 5 questions in 50-80 words each.			4x3=12
17.	Define Mean Absolute Deviation (MAD). Explain the steps to calculate MAD taking the following dataset as an example. [10, 12, 14, 16, 18]		4
Ans.	<p>Steps to calculate MAD :</p> <p>(1) Find the mean of the dataset. (2) Find the absolute difference between each value and the mean. (3) Take the average of these absolute differences.</p> <p>Sample Data : [10, 12, 14, 16, 18] Mean =14 Absolute differences : [4, 2, 0, 2, 4] MAD = (4+2+0+2+4)/5 = 2.4</p> <p>(1mark for correct definition of MAD ½ mark each for correct step ½ mark each for calculating correct mean, absolute differences and MAD)</p>	Unit-1 Page.8	

18.	What do you mean by distribution in data science ? How is it different from probability ? Explain with the help of the event – Tossing the coin.		4								
Ans.	<p>Distribution in data science is a method which shows the probable values for a variable and how often they occur.</p> <p>It is different from probability as probability gives us the mathematical calculations while distribution helps us to actually visualize what is happening underneath. The distribution does not just consists of input values that can be seen but made up of all possible values.</p> <p>For example :</p> <p>Event : Tossing the Coin</p> <p>Consider a coin that has two sides. When you toss the coin, the probability of getting a head or tail is 50% for each or half and half. Any other probability is zero. In the event - Tossing a coin, the probable values are Head and Tail. The probability for this even is given below :</p> <table><tr><th>Outcome</th><th>Probability</th></tr><tr><td>Head</td><td>0.5</td></tr><tr><td>Tail</td><td>0.5</td></tr><tr><td>All Else</td><td>0</td></tr></table> <p>The probability of getting a head is 0.5, a tail is 0.5 and any other probability is 0. The sum of probabilities should be equal to 1.</p> <p>(1 mark for correct definition of distribution 1 mark for correct difference between distribution and probability 2 marks for difference between the two with reference to the given event)</p>	Outcome	Probability	Head	0.5	Tail	0.5	All Else	0	Unit-2 Page.16-17	
Outcome	Probability										
Head	0.5										
Tail	0.5										
All Else	0										
19.	<p>(i) Why is Central Limit Theorem (CLT) is considered important in statistics ?</p> <p>(ii) With reference to CLT, if the sample size of data increases, will the error increase or decrease ?</p> <p>(iii) Mention any two real world applications of CLT.</p>		4								
Ans.	<p>(i) It helps in making reliable predictions and a good estimate of the population mean using small sample sizes</p> <p>(ii) Decrease</p> <p>(iii) Applications :</p> <ul style="list-style-type: none">Opinion polls before elections to estimate public preference.Estimating average household income in a city without surveying everyone.	Unit-3 Page.30-33									

20.	(i) What are deciles ? (ii) Consider the following dataset of 15 numbers. [77, 60, 63, 36, 54, 57, 36, 72, 55, 51, 32, 56, 33, 42, 55] Calculate the Data Position and decile value for D ₁ and D ₅ .		4																																
Ans.	(i) Deciles divide a dataset into 10 equal parts. They are denoted as D ₁ , D ₂ ,, D ₉ (D ₁₀ is the maximum value). Each decile corresponds to 10% of the ordered data. (1 mark for correct definition) (ii) Step 1 : Sort the data in ascending order as shown : <table><tr><td>Sr. No.</td><td>Digit</td></tr><tr><td>1</td><td>32</td></tr><tr><td>2</td><td>33</td></tr><tr><td>3</td><td>36</td></tr><tr><td>4</td><td>36</td></tr><tr><td>5</td><td>42</td></tr><tr><td>6</td><td>51</td></tr><tr><td>7</td><td>54</td></tr><tr><td>8</td><td>55</td></tr><tr><td>9</td><td>55</td></tr><tr><td>10</td><td>56</td></tr><tr><td>11</td><td>57</td></tr><tr><td>12</td><td>60</td></tr><tr><td>13</td><td>63</td></tr><tr><td>14</td><td>72</td></tr><tr><td>15</td><td>77</td></tr></table> Calculate Data Position and decile for D ₁ Formula : $Di = \frac{i \times (n+1)^{th}}{10}$ To calculate D ₁ : i = 1 n = 15 $D_1 = \frac{1 \times (15+1)^{th}}{10}$ = 16/10 = 1.6 th data, i.e data between digit numbers 1 and 2 which is 32 + 0.6 x (33–32) = 32 + 0.6 = 32.6 To calculate D ₅ : i = 5 n = 15 $D_1 = \frac{5 \times (15+1)^{th}}{10}$ = 80/10 = 8.0 th data, i.e data between at digit number 8 which is 55 (1mark for correctly forming the table with digit numbers and sorted values ½ each for correct data position and value for D ₁ ½ for correct data position and value for D ₅)	Sr. No.	Digit	1	32	2	33	3	36	4	36	5	42	6	51	7	54	8	55	9	55	10	56	11	57	12	60	13	63	14	72	15	77	Unit-4 Page.42	
Sr. No.	Digit																																		
1	32																																		
2	33																																		
3	36																																		
4	36																																		
5	42																																		
6	51																																		
7	54																																		
8	55																																		
9	55																																		
10	56																																		
11	57																																		
12	60																																		
13	63																																		
14	72																																		
15	77																																		

21.	Why is it important to discard data properly once its use is over ? Explain the three methods to discard physical copies of confidential data.		4
Ans.	<p>It's important to discard data properly once its use is over as it : Prevents unauthorized access and misuse of confidential data.</p> <p>Methods to discard physical copies of confidential data are :</p> <p>(i) Shredding the documents : A shredder is used to shred these documents. No one should be able to recover the information from the shredded documents.</p> <p>(ii) Cutting up the documents : If there is a single page or file to discard, cutting can be an appropriate method to discard. The file or paper should be cut into small pieces to make sure that no sensitive information is readable.</p> <p>(iii) Burning the documents – It is an effective way to discard the documents as it makes sure that documents that are burnt cannot be recovered again.</p> <p>(1 mark for correct importance 1 mark each for correct method)</p>	Unit-5 Page.49-50	