CBSE | DEPARTMENT OF SKILL EDUCATION

DATA SCIENCE (SUBJECT CODE - 844)

MARKING SCHEME FOR CLASS XII (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.	Total Quality Management (TQM)	NCERT	4	98	1
ii.	c) I,III,IV	CBSE Study Material	3	14	1
iii.	Assertion & Reason is correct, R is correct explanation of A.	NCERT	02	49	1
iv.	d) Text Colour	NCERT	3	72	1
v.	Assertion & Reason is correct, R is correct explanation of A.	CBSE Study Material	02	58	1
vi.	a) Self doubt	NCERT	4	104	1
Q. 2	**Answer any 5 out of the given 6 questions (1 x 5 = 5 marks)**				
i.	Data privacy is only about secure data storage	CBSE	1	5	1
ii.	Box plot	CBSE	2	12	1
iii.	Result fruit	CBSE	3	18	1
iv.	Both of the mentioned	CBSE	4	27	1
v.	Continuous	CBSE	5	37	1
vi.	Unsupervised learning needs human intervention	CBSE	7	50	1
Q. 3	**Answer any 5 out of the given 6 questions (1 x 5 = 5 marks)**				
i.	All of the above	CBSE	1	5	1
ii.	Both scatter plot and histogram can be used	CBSE	2	12	1
iii.	True	CBSE	3	18	1

iv.	False	CBSE	4	27	1
v.	Regression can be done for categorical variables.	CBSE	5	34	1
vi.	more flexible	CBSE	6	42	1
Q. 4	**Answer any 5 out of the given 6 questions (1 x 5 = 5 marks)**				
i.	False	CBSE	1	5	1
ii.	Pair plots	CBSE	2	12	1
iii.	Decision Trees can only handle numerical variables.	CBSE	3	17	1
iv.	K-Nearest Neighbors can be used for clustering	CBSE	4	27	1
v.	True	CBSE	5	37	1
vi.	True	CBSE	6	42	1
Q. 5	**Answer any 5 out of the given 6 questions (1 x 5 = 5 marks)**				
i.	Data traceability	CBSE	1	2	1
ii.	Removing Duplicates	CBSE	2	10	1
iii.	Regression trees are used when the dependent variable is continuous. Classification trees are used when the dependent variable is categorical	CBSE	3	16	1
iv.	Both Regression and Classification	CBSE	4	27	1
v.	False	CBSE	5	37	1
vi.	Continuous	CBSE	6	44	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	Interpersonal means dealing with relationships. It is between two or more persons. Interpersonal skills thus mean the competencies required to work with other people. Example- Having a positive attitude, generally implies being optimistic about situations, interactions and oneself. People with a positive attitude are hope full and see the best in difficult situations. However if you have negative feelings, it is good to talk about it with someone you trust and find out why you are feeling like that.	NCERT	4	92	2
Q. 7	The steps to add the shapes are: 1. Click the arrow next to the Symbol Shapes optionon the Drawingbar. 2. Select the desired shape. 3. Click and drag on the slide to draw the shape. 4. Similarly, you can use the options available on the Drawing bar to draw manyother shapes, arrows, callouts, etc.	NCERT	3	74	2

Q. 8	a) =sum(C3:E3) b) Sorting	NCERT NCERT	3 3	52 58	2
Q. 9	Openness and consciousness factor inculcates individual personality with creativity, curious active adventurous nature. It also enhances self discipline and time management skills.	NCERT	02	48	2
Q. 10	a) Barriers b) Venture c) Risk				2
Answer any 4 out of the given 6 questions in 20 – 30 words each (2 x 4 = 8 marks)					
Q. 11	Data Governance covers the following aspects. • Data Quality • Data Security • Data Architecture • Data Integration and Interoperability • Data Storage	CBSE	1	2	2
Q. 12	Missing data is also a common issue with datasets. Many machine learning algorithms do not work well with missing data and so this must be handled well during the cleaning stage. The two common techniques to handle missing data is to either remove that row of data or to insert a value that is quite close to the mean or mode of the variable that is missing.	CBSE	2	11	2
Q. 13	A Decision tree is a diagrammatic representation of the decision-making process and has a tree-like structure. Each internal node in the decision tree denotes a question on choosing a particular class. Every branch represents the outcome of the test, and each leaf node holds a class label.	CBSE	3	15	2
Q. 14	There are two important characteristics of K-NN. • Lazy Learning – K-NN follows the principle of lazy learning. It does not have a specific training phase where it learns about the data. It uses all the training data while performing a classification operation. • Non-parametric Learning – K-NN is a non-parametric algorithm as it does not assume anything about the distribution of the data. So KNN does not have to find any parameter for the distribution of data. The only hyperparameter that KNN has is K, and that is provided by the user to the model.	CBSE	4	24	2

Q. 15	Mean Absolute Error measures the average magnitude of the errors in a set of predictions, without considering their direction. The Root Mean Square Deviation is used to determine how close the observed points are to the model's predicted values. Mathematically, the Root Mean Square Deviation is the square root of the variance of the residuals.	CBSE	5	35	2
Q. 16	Multiple Linear Regression uses multiple independent variables to predict the outcome of a dependent variable. For example, effects of age, weight, and height on cholesterol levels of an individual. Here, age, weight, and height are independent variables, and cholesterol level is the dependent variable because it is dependent on the factors age, height, and weight.	CBSE	6	42	2
Answer any 3 out of the given 5 questions in 50– 80 words each (4 x 3 = 12 marks)					
Q. 17	CCPA – California Consumer Privacy Act California passed the CCPA on June 28, 2018 and it went into effect on January 1, 2020. The CCPA is a landmark legislation designed to protect consumer data. The CCPA provides residents living in the state of California with the right to request businesses: To disclose to them what personal information the businesses have about them and what they intend to do with it • To request businesses to delete their personal information • To request businesses not to sell their personal information	CBSE	1	4	4
Q. 18	Univariate analysis can be considered as the easiest form of data analysis where we only analyze only one variable from the entire dataset. Since we deal with only one variable, we do not have to worry about causes or relationships. The main purpose of the univariate analysis is to describe the data and find patterns that exist within it. For univariate analysis, we pick up a variable from the dataset and try to analyze it in depth. One example of a variable in the univariate analysis might be "revenue". Another might be "height". For univariate analysis, we would not look at these two variables at the same time, nor would we look at the relationship between them. Multivariate analysis is a more complex form of a statistical analysis technique and is used to analyze more than two variables in the data set. One of the ways to do multivariate analysis is Bivariate analysis. It refers to the	CBSE	2	8	4

	analysis of two or more variables in the dataset. It is usually carried out between the target variable and another feature of the dataset. The main objective is to find out if there is a relationship between two different variables. Bivariate analysis is usually done by using graphical methods like scatter plots, line charts, and pair plots. Bivariate analysis is also a good way to measure the correlations between the two variables. For example – in a market survey we may be looking to analyze the relationship between price and sales of a product to see if there is any relationship				
Q. 19	To create a decision tree, you can follow the steps below. 1. Think about your main objective for which you are creating the decision tree. The main decision that you are trying to make should be placed at the very top of the decision tree. Therefore, the main objective should be the "root" of the entire diagram. 2. Next, you need to draw the branches and leaf nodes. For every possible decision, stemming from the root make a branch. One root or node can have two or more branches. At the end of the branches, attach the leaf nodes. The leaf nodes should represent the results of each decision. If another decision has to be made, draw a square leaf node. If the outcome is not quite certain, you should draw a circular node. 3. Finally, you need to calculate the probability of success of each decision being made. While creating the decision tree, it is essential to do some research, so that you can predict the probability of each decision. To do this research, you may examine old data or assess previous projects. Once you calculate the expected value of each decision in tree, put the values on the branches.	CBSE	3	17	4
Q. 20	Cross-Validation refers to a technique in which we reserve a particular portion of a dataset on which we do not train the model. After the training is over, we test the resulting model on this portion of the data before finalizing it. The steps involved in cross-validation are as follows - 1. Reserve a small portion of the data set called validation data. 2. Train the model using the remaining dataset 3. Test the model on the validation data set and check its accuracy.	CBSE	4	25	4

Q. 21	1. Recommendation Engines: Many websites selling products use recommendation engines to predict what products a customer is likely to purchase. This is done by using past purchase behavior data and unsupervised learning techniques which can help to discover trends in the data. These predictions are then used to make add-on recommendations relevant to a particular customer during the checkout process. 2. Medical imaging: Unsupervised machine learning provides essential features to medical imaging devices, such as image detection, classification, and segmentation, used in radiology and pathology to diagnose patients quickly and accurately. 3. Anomaly detection: Anomaly detection is also an important application for unsupervised learning. Anomalies can be useful for detecting fraud in financial systems or other security applications. Unsupervised learning models can go through large amounts of raw data and find unusual data points within a dataset. These unusual data points can then be analyzed manually to see if there has indeed been a fraud or security breach. 4. Customer personas: Customer personas are used to understand common purchasing habits and purchasing times for customers of a product. With the help of unsupervised learning, organizations can build better buyer persona profiles. This, in turn, enables them to align their sales and ads to such customer segments more appropriately.	CBSE	7	48	4