

K. J. Somaiya Institute of Technology, Sion, Mumbai-22
(Autonomous College Affiliated to University of Mumbai)

May-June 2026

(B. Tech) Program: Computer Engineering Scheme I/II/IIB/III: IIB

Regular Examination: LY Semester: VIII

Course Code: CEDLC8021 and Course Name: Applied Data Science

Date of Exam: 21/05/26

Duration: 02.5 Hours

Max. Marks: 60

Instructions:

- (1) All questions are compulsory.
- (2) Draw neat diagrams wherever applicable.
- (3) Assume suitable data, if necessary.

Q. No.	Question	Max. Mark	CO	BT level													
Q 1	Solve any two questions out of three: (05 marks each)	10															
a)	Describe the different steps used in data preprocessing in data science.		CO1	Un													
b)	Explain the structure of Box plot and Violin plot with a suitable example.		CO3	Un													
c)	Describe the use of t-test in data science. Mention the importance of significance level.		CO5	Un													
Q 2	Solve any two questions out of three: (05 marks each)	10															
a)	Describe the properties of inverse of a Matrix and Find the inverse of following Matrix using Adjoint Method $\begin{vmatrix} 5 & 3 & 1 \\ 0 & 3 & 4 \\ 0 & 0 & 6 \end{vmatrix}$		CO2	Ap.													
b)	Define Central limit theorem and standard error. The exam scores for a large university population are skewed, with a mean of 84 and a standard deviation of 4. If a random sample of (n = 35) students is chosen, what is the probability that the mean score of this sample is less than 83?		CO4	Un													
c)	Apply least square optimization on following data to Find the number of placements for the year 2026. <table><tr><td>X (Academic Year)</td><td>2020</td><td>2021</td><td>2022</td><td>2023</td><td>2024</td><td>2025</td></tr><tr><td>Y.(Students Placed)</td><td>35</td><td>48</td><td>39</td><td>62</td><td>89</td><td>65</td></tr></table>		X (Academic Year)	2020	2021	2022	2023	2024	2025	Y.(Students Placed)	35	48	39	62	89	65	CO6
X (Academic Year)	2020	2021	2022	2023	2024	2025											
Y.(Students Placed)	35	48	39	62	89	65											

K. J. Somaiya Institute of Technology, Sion, Mumbai-22
(Autonomous College Affiliated to University of Mumbai)

May-June 2026

(B. Tech) Program: Computer Engineering Scheme I/II/IIB/III: IIB

Regular Examination: LY Semester: VIII

Course Code: CEDLC8021 and Course Name: Applied Data Science

Date of Exam: 21/05/26

Duration: 02.5 Hours

Max. Marks: 60

Q.3	Solve any two questions out of three. (10 marks each)	20																		
a)	Explain few applications of Matrix Factorization in data science [2M] Apply QR Decomposition on following Matrix. [8M] $\begin{bmatrix} 1 & 2 & 4 \\ 0 & 0 & 5 \\ 0 & 3 & 6 \end{bmatrix}$		CO2	Ap.																
b)	Explain Skewness and Kurtosis of data. [4M] Find the estimates of location and variability for following: [6M] <table><tr><td>Class</td><td>10 - 20</td><td>20 - 30</td><td>30 - 40</td><td>40-50</td><td>50-60</td><td>60-70</td><td>70-80</td></tr><tr><td>Freque ncy</td><td>12</td><td>15</td><td>20</td><td>35</td><td>15</td><td>5</td><td>3</td></tr></table>		Class	10 - 20	20 - 30	30 - 40	40-50	50-60	60-70	70-80	Freque ncy	12	15	20	35	15	5	3	CO3	Ap.
Class	10 - 20		20 - 30	30 - 40	40-50	50-60	60-70	70-80												
Freque ncy	12	15	20	35	15	5	3													
c)	Explain why the concepts of sampling distributions and selection bias are important in data science? [4M] Apply suitable type of distributions on following cases.[6M] i. As only 5 students came to attend the class today, find the probability for exactly 6 students to attend the classes tomorrow. ii. A coin is tossed 15 times. What is the probability of getting exactly 9 heads? iii. Assume that, you usually get 2 phone calls per hour. calculate the probability, that a phone call will come within the next hour.	CO4	Ap.																	
Q.4	Solve any two questions out of three. (10 marks each)	20																		
a)	Describe the terms: Descriptive, Diagnostic, Predictive and Prescriptive analysis in Data Science[4M] Apply the principles of Joint and Conditional probability on Following: [6M] 1. A deck has 52 cards. What is the joint probability of drawing an Ace, not replacing it, and then drawing a King? 2. Out of 100 students, 40 like Math (M), 25 like Physics (P), and 15 like both. If a student is chosen, what is the probability they like Physics given they like Math? 3. A family has 2 children. Given that one of the children is a boy, what is the probability that the other child is also a boy?		CO1	Ap																

K. J. Somaiya Institute of Technology, Sion, Mumbai-22
(Autonomous College Affiliated to University of Mumbai)

May-June 2026

(B. Tech) Program: Computer Engineering Scheme I/II/IIB/III: IIB
Regular Examination: LY Semester: VIII

Course Code: CEDLC8021 and Course Name: Applied Data Science

Date of Exam: 21/05/26

Duration: 02.5 Hours

Max. Marks: 60

b) Describe the significance of p-value and Confidence level in hypothesis testing [2M]

The illness caused by a virus in a city concerning some restaurant inspectors is not consistent with their evaluations of cleanliness of restaurants. In order to investigate this possibility, the director has five restaurant inspectors to grade the cleanliness of three restaurants. The results are shown below.

Inspectors	Restaurants		
	I	II	III
1	71	55	84
2	65	57	86
3	70	65	77
4	72	69	70
5	76	64	85

Apply two way ANOVA test with 5% significance level. Using the results justify about the hypothesis you claimed.[8M]

CO5

Un,
Ap.

c) How to evaluate a Machine Learning model w.r.t. Data science?

Explain how following metrics are useful in Data Science. Give examples wherever required.

- Accuracy and Precision
- Sensitivity and Specificity;
- ROC Curve and AUC;
- Regression Metrics: MSE and MAE

CO6

Un
