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

May-June 2025
(B. Tech) Program: Computer Engineering Scheme I/II/IIB/III: III
Regular Examination: LY Semester: VIII
Course Code: CEDLC8021 and Course Name: Applied Data Science
21/05/25 Duration: 02.5 Hours Max. Marks: 60

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

Q. No.	Question	Max. Marks	CO	BT level
Q 1	**Solve any two questions out of three: (05 marks each)**	**10**		
a)	Describe the different data types used in data science and possible types of errors that can occur while using these data types.		CO1	Un
b)	Explain how and why follwing tools are used in data science to handle categorical data: **Contingency tables and Hexagonal binning.**		CO3	Un
c)	Describe steps used in Chi square test with a suitable example.		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 | 1 2 3 | | 0 2 4 | | 0 0 5 |		CO2	Ap.
b)	Explain the different types of sampling techniques used in statistical data science with their advantages and drawbacks.		CO4	Un
c)	**Apply** least square optimization on following data to Find the number of placements for the year 2025.		CO6	Ap.

Seat No.:

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

May-June 2025
(B. Tech) Program: Computer Engineering Scheme I/II/IIB/III: [redacted]
Regular Examination: LY Semester: VIII
Course Code: CEDLC8021 and Course Name: Applied Data Science
Date of Exam: 21/05/25 Duration: 02.5 Hours Max. Marks: 60

	X(Academic Year)	**2020**	**2021**	**2022**	**2023**	**2024**			
	Y(No. of Student Placed)	**32**	**55**	**72**	**49**	**65**			
Q.3	**Solve any two questions out of three. (10 marks each)**						**20**		
a)	**Explain** few applications of Matrix Factorization in data science [3M] **Apply** QR Decomposition on following Matrix. [7M] | 1 2 4 | | 0 0 5 | | 0 3 6 |							CO2	Ap.
b)	**Explain** Skewness and Kurtosis of data. [4M] **Find** the estimates of location and variability for following: [6M]							CO3	Ap.
	Class	Frequency							
	10 - 20	15							
	20 - 30	25							
	30 - 40	20							
	40 - 50	12							
	50 - 60	8							
	60 - 70	5							
	70 - 80	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?							CO4	Ap.

Seat No.:

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

May-June 2025
(B. Tech) Program: Computer Engineering Scheme I/II/IIB/III:
Regular Examination: LY Semester: VIII
Course Code: CEDLC8021 and Course Name: Applied Data Science
Date of Exam: 21/05/25 Duration: 02.5 Hours Max. Marks: 60

	iii. Assume that, you usually get 2 phone calls per hour. calculate the probability, that a phone call will come within the next hour.			
Q.4	**Solve any two questions out of three. (10 marks each)**	**20**		
a)	**Describe** different types of data analysis [4M] With suitable examples describe the following libraries used for data visualization in Data Science [6M] i. matplotlib ii. seaborn iii. ggplot		CO1	Un
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. **Apply** two way ANOVA test with 5% significance level. [8M]		CO5	Un, Ap.
c)	Explain how following metrics are useful in Data Science. i. Confusion Matrix; ii. Precision and Recall; iii. Sensitivity and Specificity; iv. ROC Curve and AUC; v. Regression Metrics: MSE and MAE		CO6	Un

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

****

Seat No.: