Name. _____		Printed Pages:01
Student Admn. No.: _____		

School of Computer Science & Engineering
Back Paper Examination Even Semester (Non - Graduating Batches) – June 2024
[Programme: B.Tech]   [Semester:IV]

Course Title: Data Sciences	Max Marks: 100
Course Code: E2UC405B / BTCS9212	Time: 3 Hrs.

Instructions:
1. *All questions are compulsory.*
2. *Assume missing data suitably, if any.*

		K Level	COs	Marks
SECTION-A (15 Marks)		**5 Marks each**		
1.	Define data science. What is the role of data science in business, medical research, healthcare, education, social media, technology and financial institutions?	K1	CO1	5
2.	Define population and sample with a suitable example.	K1	CO1	5
3.	In a large number of parts manufactured by a machine , the mean number of defectives in a sample of 20 is 2. Out of 1000 such samples , how many would be expected to contain atleast 3 defective parts.	K2	CO2	5
SECTION-B (40 Marks)		**10 Marks each**		
4.	What is cross-validation, and why is it important in machine learning?	K1	CO3	10
5.	What are the different types of data normalization techniques? Explain any two with examples.	K2	CO2	10
6.	Discuss the importance of data visualization in data science. Mention any two data visualization libraries in Python.	K3	CO4	10
7.	Using a dataset of your choice, create a Python script to generate a histogram and a scatter plot. Explain the insights you can derive from these plots.	K2	CO3	10
SECTION-C (45 Marks)		**15 Marks each**		
8.	Explain the k-means clustering algorithm. Describe how you would determine the optimal number of clusters.			15
9.	Given a dataset, describe how you would handle missing values. Provide a Python code snippet demonstrating one of the methods.	K4	CO2	15
10	a)Describe the CRISP-DM process model in data science. What are its main phases? b) Provide an example of a data science project and describe how you would apply the CRISP-DM methodology to it.	K5	CO3	15