Name: **Enrolment No:**	

UPES

Supplementary Examination, December 2023

Course: Data Visualization for Analytics **Semester: III**

Program: B.Tech (CSE -Spl. Buss. Analytics & Opt.) **Time : 03 hrs.**

Course Code: CSBA2007 **Max. Marks: 100**

Instructions:

SECTION A
(5Qx4M=20Marks)

S. No.		Marks	CO
Q 1	Define geospatial data and its uses.	**4**	**CO1**
Q 2	Define bivariate analysis. List out any 2 methods used for univariate and bivariate analysis.	**2+2**	**CO2**
Q 3	What is a hypothesis in statistics? Explain the difference between the null hypothesis and the alternative hypothesis.	**2+2**	**CO3**
Q 4	Differentiate bar chart and pie chart? When would you choose one over the other to represent data?	**2+2**	**CO4**
Q 5	Imagine a factory that produces widgets. Out of all the widgets produced, 10% are defective. The quality control system of the factory can accurately identify 90% of the defective widgets but also incorrectly flags 5% of the non-defective widgets as defective. If a randomly selected widget is found to be defective by the quality control system, what is the probability that it is actually defective?	**4**	**CO3**

SECTION B
(4Qx10M= 40 Marks)

Q 6	Explain rules for better visualizations with examples.	**10**	**CO1**
Q 7	Explore and discuss five different visual forms commonly used in data representation and their practical applications.	**10**	**CO1**
Q 8	Describe five possible causes or factors contributing to outliers within a dataset. Discuss any 2 algorithms used for outlier analysis.	**5+5**	**CO2**

Q 9	Consider a study analyzing the relationship between hours spent studying and exam scores obtained by a group of students. For a sample of five students, the hours spent studying (in hours) and their respective exam scores (out of 100) are as follows: Student A: Studied for 3 hours, scored 60 Student B: Studied for 5 hours, scored 75 Student C: Studied for 2 hours, scored 50 Student D: Studied for 6 hours, scored 80 Student E: Studied for 4 hours, scored 70 Calculate the Karl Pearson's coefficient of correlation to determine the relationship between study hours and exam scores for this group of students.	**10**	**CO2**
	OR		
	a) Discuss the types of Discrete Probability Distributions. b) Consider a scenario where a fair coin is flipped 8 times. Each flip results in either heads (H) or tails (T). Assuming the probability of getting heads in a single flip is p=0.5, calculate the probability of getting exactly 5 heads in these 8 coin flips according to the Bernoulli distribution.	**5+5**	**CO2**
SECTION-C (2Qx20M=40 Marks)			
Q 10	a) Explain the utility of boxplots in detecting outliers. Describe the distinct components of a boxplot and their relevance in outlier identification. b) Construct a boxplot for the given dataset and pinpoint any outlier values: 25, 40, 18, 35, 10, 55, 30, 28, 22, 20, 28, 26	**10+10**	**CO3**
Q 11	A study was conducted to analyze the relationship between advertising spending and product sales for a company across 12 months. The company's monthly advertising expenditure (in ₹) and corresponding product sales (in ₹) data were collected: Monthly Advertising Expenditure: January: ₹25,000 February: ₹28,000 March: ₹22,000 April: ₹30,000 May: ₹26,000 June: ₹32,000	**8+8+4=20**	**CO4**

	July: ₹29,000 August: ₹33,000 September: ₹27,000 October: ₹31,000 November: ₹34,000 December: ₹35,000 Monthly Product Sales: January: ₹80,000 February: ₹85,000 March: ₹75,000 April: ₹90,000 May: ₹82,000 June: ₹95,000 July: ₹88,000 August: ₹98,000 September: ₹81,000 October: ₹92,000 November: ₹96,000 December: ₹1,00,000 a) Calculate the correlation coefficient between the monthly advertising expenditure and product sales. b) Write a python code to visualize the relationship between advertising spending and product sales using an appropriate graph. c) Analyze the correlation result and the plot to determine the strength and direction of the relationship between advertising expenditure and product sales.		
	OR		
	a) Explore various types of dashboards commonly used in different domains and elucidate their applications. b) Provide a comprehensive step-by-step guide to creating a dashboard, detailing the necessary stages and key considerations involved in its development.	**10+10**	**CO4**