Maximum Marks: 30	Semester: January 2024 – April 2024 Examination: In-Semester Examination		Duration :1h 15 min
Programme code: h54 Programme: Honors- Data Science and Analytics		Class: SY	Semester: IV (SVU 2020)
Name of the Constituent College: K. J. Somaiya College of Engineering		Name of the department: COMP/EXCP	
Course Code: 116h54C401	Name of the Course: Applied Data Science		

Question No.		Max. Marks
Q1	**Solve any one of the following (10 marks each):**	10
	Draw and explain the Data Science Process life cycle. Explain the skill sets required for applied data science.	
	OR	
	Explain the following terms with examples a) Skewness & Kurtosis b) Homoscedasticity & Heteroskedasticity	
Q2	**Solve any two of the following (5 marks each):**	10
	Most graduate schools of business require applicants for admission to take the Graduate Management Admission Council's GMAT examination. Scores on the GMAT are roughly normally distributed with a mean of 527 and a standard deviation of 112. What is the probability of an individual scoring above 500 on the GMAT?	
	OR	
	On average, every one out of 10 telephones is found busy. Six telephone numbers are selected at random. Use binomial distribution. i) Find the probability that four of them will be busy. ii) Find the probability that less than three will be busy.	
	OR	
	Solve the following numericals based on similarity measures: a)Calculate Euclidean distance between following two points-A (2,4,5) and B(1,7,9). b)Calculate Jaccard distance for the following two sets	

	A{1,2,3,4} , B{3,4,5,6}	
Q3	Under what assumptions can we apply linear regression to a given set of variables? Find the equation for the linear regression line of the blood glucose values versus the age values as given in the following table:	10

Age (x)	Glucose (y)
43	99
21	65
25	79
42	75
57	87
59	81