JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

Supplementary Examination- 2026

B.Tech-V Semester (CSE/IT)

COURSE CODE (CREDITS): 20B1WCI531 (3) MAX. MARKS: 75

COURSE NAME: FOUNDATIONS FOR DATA SCIENCE AND VISUALIZATION

COURSE INSTRUCTORS: RBT MAX. TIME: 2 Hours

Note: *(a) All questions are compulsory.*

(b) The candidate is allowed to make Suitable numeric assumptions wherever required for solving problems. Scientific calculator allowed.

Q.No	Question	CO	Marks
Q1	a) Define Data Science. Explain the data science lifecycle with a neat diagram. b) Write Python code to: i) Read a CSV file ii) Display first 5 rows iii) Find mean and standard deviation of a numeric column c) Explain the importance of data visualization in data science.	1	5 + 5 + 5 = 15
Q2	a) How does Pandas help in data cleaning and analysis? b) Explain NumPy, SciPy, and Matplotlib and their role in data science. c) What is multicollinearity. Discuss the effects of Multicollinearity	2	5 + 5+ 5 = 15
Q3	The determination of the shear strength of spot welds is relatively difficult, where as measuring the weld diameter of spot welds is relatively simple. As a result, it would be advantageous if shear strength could be predicted from a measurement of weld diameter. The data are as follows: (Use $t_{0.025,8} = 2.306$) **Shear Strength (psi)** — **Weld Diameter (.0001 in.)** 370 — 400 780 — 800 1,210 — 1,250 1,560 — 1,600 1,980 — 2,000 2,450 — 2,500 3,070 — 3,100 3,550 — 3,600 3,940 — 4,000 3,950 — 4,000	6	15

	(a) Draw a scatter diagram. (b) Find the least squares estimates of the regression coefficients. (c) Test the hypothesis that the slope of the regression line is equal to 1 at the .05 level of significance. (d) Estimate the expected value of shear strength when the weld diameter is .2500. (e) Find a prediction interval such that, with 95 percent confidence, the value of shear strength corresponding to a weld diameter of .2250 inch will be contained in it. (f) Plot the standardized residuals. (g) Does the plot in part (f) support the assumptions of the model?		
Q4	a) Describe any four common visualization techniques and their use cases. b) Explain how data visualization helps in exploratory data analysis (EDA). c) With the help of an example show how to test the statistical significance of the fitted equation (Use ANOVA approach).	4	5+ 5 + 5 = 15
Q5	a) Differentiate between Linear Regression and Logistic Regression. b) Verify that Sum of squares residuals $= \frac{S_{xx}S_{YY} - S_{xY}^2}{S_{xx}}$ c) The following data relate the number of units of a good that were ordered as a function of the price of the good at six different locations. How many units do you think would be ordered if the price were 25?	6	5+ 5+ 5 = 15

Number ordered	88	112	123	136	158	172
Price	50	40	35	30	20	15