JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

Make-up Examination-Nov-2025

COURSE CODE (CREDITS): 22M1WCI136 (3) MAX. MARKS: 25

COURSE NAME: Data Visualization

COURSE INSTRUCTORS: Dr. Ramesh Narwal MAX. TIME: 1 Hour 30 Min

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

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

Q.No	Question	CO	Marks
Q1	A data classifier correctly predicts spam emails with a probability of 0.8. (a) Model this as a Binomial distribution for 10 emails and calculate the probability of exactly 7 correct classifications. (b) Compare this with the Bernoulli distribution for a single trial.	3	5
Q2	Describe the process of extracting and joining data from three formats – CSV (sales.csv), JSON (inventory.json), and XLS (supplier.xls) – into one standardized DataFrame. Include key steps like schema alignment, type conversion, and error handling.	2	5
Q3	You are asked to visualize company performance metrics: revenue, profit margin, and employee count across five regions. (a) Suggest three visualization types (for example bubble chart, heat map, box plot) and justify each. (b) Discuss how color, size, and 3D representation can affect interpretation accuracy. (c) Suggest one visualization that could mislead viewers, explaining why.	4	5
Q4	In a dataset of advertising spend (X) and sales (Y), the regression equation is $Y = 5 + 1.2X$. (a) Predict sales for $X = 20$. (b) Compute the residual if the actual sales were 30. (c) Discuss how multicollinearity could distort the interpretation of coefficients.	1	5
Q5	You are given a CSV dataset of customer transactions containing missing values, duplicate rows, and inconsistent date formats (for example "2025/11/10", "10-11-25", "Nov 10, 2025"). (a) Explain how you would detect and correct these inconsistencies using Python (Pandas). (b) Discuss how automated data validation pipelines can improve data quality in large organizations.	2	5