

(28)

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -3 EXAMINATIONS- 2026

B.Tech-VI Semester (CSE/IT)

COURSE CODE (CREDITS): 25B1WCI642 (3)

MAX MARKS: 35

COURSE NAME: Introduction to Data Analytics and Visualization

COURSE INSTRUCTOR: Dr Nancy Singla

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

(c) Use of calculator is allowed

Q.No	Question	CO	Marks
Q1	(a) State one difference between data science and traditional computer science?	[CO1]	1*5=5
	(b) Which Python library is primarily used for numerical computations and matrix operations, and which one is widely used for data manipulation and analysis?	[CO2]	
	(c) What is the difference between stemming and lemmatization?	[CO3]	
	(d) State one application of Box Plot in data visualization.	[CO4]	
	(e) What is the rank of a matrix?	[CO6]	
Q2	(a) Explain the components of Data Science and how they interact in solving real-world problems.	[CO1]	2*3=6
	(b) Describe the importance of data cleansing in the data processing pipeline.	[CO3]	
	(c) Write a short note on Chi-square Test and its applications.	[CO5]	
Q3	(a) Solve the following system of equations using Matrix Inversion: $2x+y=5, x-y=1$	[CO6]	3*2=6
	(b) Compute the eigenvalues of the matrix: $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$		
Q4	(a) Derive the formula for LU Decomposition of a 3x3 matrix.	[CO6]	3*2=6
	(b) Show how Principal Component Analysis (PCA) reduces dimensionality using eigenvalue decomposition.		
Q5	(a) How would you design a data wrangling pipeline for unstructured text data (e.g., customer reviews)?	[CO3]	3*2=6
	(b) Critically evaluate the use of HeatMaps vs Scatter Plots for representing correlation in large datasets.	[CO4]	
Q6	An airline has shared flight delay data (flight number, departure time, delay minutes, route). As a data analyst, design a Python-based solution to improve scheduling. Your task should: - Clean and prepare the dataset - Perform statistical analysis (average delay, variance, correlation between route and delay) - Visualize results using at least two chart types Implement using Pandas, NumPy, Matplotlib	[CO3] [CO5] [CO4] [CO2]	6