JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -1 EXAMINATION- 2025

M.Tech-I Semester (CSE)

COURSE CODE (CREDITS): 22M11CI112 (3) MAX. MARKS: 15

COURSE NAME: INTRODUCTION TO DATA SCIENCE

COURSE INSTRUCTORS: Dr Nancy Singla MAX. TIME: 1 Hour

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.	An airline authorities wants to reduce flight delays using predictive analytics. Briefly explain how you would approach this problem using six stages of the Data Science Lifecycle.	CO1	[3]

Q2.	Write the output of the following code snippets:
	(a)
	`def func(val, lst=[]):`
	`  lst.append(val)`
	`  return lst`
	`print(func(1))`
	`print(func(2))`
	`print(func(3, []))`
	`print(func(4))`
	(b)
	`data = {'x': [1, 2, {'a': 10, 'b': 20}], 'y': (3, 4, 5)}`
	`print(data['x'][2]['b'] + data['y'][1])`
	(c)
	`date_str <- "31-12-2023"`
	`date_obj <- as.Date(date_str, format = "%d-%m-%Y")`
	`formatted_date <- format(date_obj, "%B %d, %Y")`
	`print(formatted_date)`
		CO2	[1+1+1]

Q3.	You are working as a data analyst at a healthcare company. You receive a patient dataset with several columns, including age, blood pressure, cholesterol level, and diagnosis. However, you notice that some values in the dataset are missing. Explain three different strategies you could use to handle the missing data before building a predictive model.	CO3	[3]

Q4.	The number of phone text messages send by 11 different students is given below. 14, 25, 31, 36, 37, 41, 51, 52, 55, 79, 112. a) Find the lower quartile, the median and the upper quartile of the data. b) Identify the outlier(s) present in the data. c) Draw a suitably labelled box plot for this data, clearly indicating any outliers.	CO3	[1+1+1]

Q5.	You have a list of dictionaries representing sales data:
	
	`sales = [`
	`{'product': 'A', 'units_sold': 10, 'price': 5.0},`
	`{'product': 'B', 'units_sold': 3, 'price': 15.0},`
	`{'product': 'C', 'units_sold': 8, 'price': 7.5},`
	`{'product': 'D', 'units_sold': 0, 'price': 12.0}`
	`]`
	
	Write a python code to
	(a) Filter the list to only include products that sold more than 5 units.
	(b) Calculate the total revenue (units_sold * price) for the filtered products.
		CO2	[3]