(Please write your

End-Term Examination
(CBCS)(SUBJECTIVE TYPE)(OffLine)
Course Name: B. Tech(CSE-AI/AIML), **Semester:II**
(May, 2024)

Subject Code: BAI 104	Subject: Introduction To Data Science
Time :3 Hours	Maximum Marks :60

Note:Q. 1 is compulsory. Attempt one question each from the Units I, II, III & IV.

		(2.5*8=20)	
Q1			
	(a)Why is ethical consideration important in data science, and what are some potential consequences of unethical practices?		
	(b)Discuss the different aspects of privacy that data scientists should consider when working with sensitive data.		
	(c) Following code, which shows the sales for ABC Kitchen, an on-campus bakery: product_sales = pd.DataFrame({'Monday': [np.nan, 152, 94,np.nan], 'Tuesday': [45, 109, 114,np.nan], 'Wednesday': [np.nan,np.nan,np.nan,np.nan], 'Thursday': [29, 85, 132, np.nan], 'Friday': [63, 143, 101, 15], 'Saturday': [87, 211, 186, 25]}, index=['Cupcakes', 'Pie Slices', 'Brownies', 'Cookies']) Write a program in Pandas to perform the following task. 1.For each day of the week, find the number of missing data points. 2.For each product, find the number of missing data points. 4.Show the days of the week on which at least 3 different products were sold. 5.Find the product(s) that were sold on Thursday and Friday.		
	(d)Write the output for the following program with explanation in two to three lines. (i) import numpy as np import pandas as pd s = pd.Series([1, 2, np.nan, 4, 5]) s.fillna(3) print (s.sum()) (ii) import pandas as pd df = pd.DataFrame([[1, 2], [3, 4]], columns=['x' , 'y']) print (df.to_csv())		
	(e)(i) What are figures and subplots, and how do they help in organizing multiple visualizations within a single plot? (ii)When is a bar chart more suitable than other types of visualizations like line plots or histograms?		
	(f)You have a dataset containing daily temperature data for a city over the past year. Each data point consists of the date and the corresponding temperature. Create a line plot to visualize the temperature trend over time. Customize the plot with appropriate labels, title, and color scheme. Additionally, annotate the plot to highlight the highest and lowest temperatures.		
	(g)Discuss the importance of interactivity in Tableau dashboards. How can you incorporate actions, filters, and parameters to enhance the user experience		

	(h)Explain the concept of a Pareto chart in Tableau and how it is used for analysis.		
	UNIT-I		
Q2	(a)In what ways has the evolution of data science been influenced by advancements in hardware technology, and how does this impact the scalability and efficiency of data analysis? (b) Imagine a data science project where the available data is incomplete and noisy(form of errors, outliers, or inconsistencies). How can data scientists leverage techniques such as data imputation and noise reduction to improve the quality of their analysis and mitigate the risk of misleading conclusions?	(10)	
Q3	(a)How do you effectively define goals in a data science project, and why is this step crucial for project success? (b) Provide examples of applications of data science in various industries, such as healthcare, finance, and marketing.	(10)	
	UNIT-II		
Q4	(a)Mr. XYZ maintains a LinkedIn database of his profile connections every month. The database includes columns such as email_id, name, company, etc., where email_id serves as a unique identifier. Write a program in Pandas to list all connections who joined in the month of May compared to the previous month, April (b) The syntax for concatenating DataFrames is pd.concat([d1, d2, d3]). In this concat() function, we are passing a list. Can we pass a dictionary instead of a list? If yes, explain with an example. If not, give a reason	(6+4=10)	
Q5	The Human Resources department is maintaining an employee database. Below are a few observations from the DataFrame 'df'. Write a Pandas program to address the following questions. Emp_ID,F_Name,L_Name,Post,Project,Rating,Salary,Sex,Doj 198944,Archit,Chandra,ASE,Unallocated,2,20000,M,9/2/2005 201601,Amit,Pathak,ASE,Unallocated,3,21000,M,6/4/2010 198966,Pankaj,Taneja,ASE,Agilent,4,22000,M,8/1/2015 113786,Ajay,Choudhary,ITA,Agilent,5,30000,M,9/1/2015 145678,Abhishek,Bansal,ITA,Agilent,5,30000,M,4/4/2012 198312,Divya,Saxena,ASE,Ultimatix,2,20000,F,3/6/2008 198945,Sankalp,Srivastava,ASE,Amex,2,20000,M,11/1/2007 187310,Shikha,Kaushal,ITA,Amex,3,18000,F,6/4/2010 107178,Dhiren,Sahu,VP,Citi,5,300000,M,9/2/2005 (i)Write a Pandas program to find the number of employees who joined in each year? (ii)Write the Data Frame (df) to an Excel file, with each year's data on a separate sheet within the Excel file. (iii)Write a custom function to find the difference between the maximum and minimum values of a variable. Then apply this custom function to find the maximum and minimum salary differences for each post	(10)	

	(iv)Write Pandas code in two different ways to display entries where the post is either 'ASE' or 'ITA', using the isin() function and without using it (v)Write a Pandas program to filter rows based on the condition that the post has a total salary over Rs70,000		
	UNIT-III		
Q6	IGDTUW is maintaining a database (df) of students with details like roll no, name, height, weight, total marks, etc. Write Matplotlib code to plot a suitable chart to find if there is any association between height and weight. Additionally, write Matplotlib code to plot a suitable chart to identify outliers in marks."	(10)	
Q7	Consider you have a dataset containing monthly sales data for multiple products sold by a retail store over the past few years. Each data point consists of the date, the product name, and the corresponding sales amount. Your task is to create a line plot using Matplotlib to visualize the trend of sales over the months for a specific product chosen by the user. Allow the user to input the product name interactively. Additionally, customize the plot to include appropriate labels for the x-axis, y-axis, and title. Add a reference line representing the average sales over the months for the selected product. Finally, save the plot as a PNG file named 'sales_trend_product.png' for future reference."	(10)	
	UNIT-IV		
Q8	(a)Explain the difference between a filter and a parameter, in Tableau. Provide examples of scenarios where each one would be useful. (b)Demonstrate how to create a calculated field in Tableau to calculate the profit margin of a product based on its sales and cost.	(6+4=10)	
Q9	(a)How can you create a dynamic filter in Tableau that allows users to select a specific range of values? Describe the steps involved. (b) ABC Online Superstore maintains a database named 'retail' for daily transaction details such as product ID, segment, order ID, profit, sales, etc. Write Tableau code to create a calculated field based on the following condition: If the sum of profit is greater than 0, assign the label 'profit'; otherwise, assign the label 'loss'."	(10)	