Name: Enrolment No:	

UPES

End Semester Examination, December 2024

Programme Name : M.Tech **Semester : 1**
Course Name : Data Visualization **Time : 03 hrs.**
Course Code : CSDS7003 **Max. Marks: 100**
Nos. of page(s) : 2

Instructions: Please attempt according to the time provided and given weightage.
Total 3 Sections: Section A, Section B and Section C

SECTION A
(20 Marks) 5 Questions – Each 4 Marks-No Choice-Attempt all Questions

S.No.	Question	Marks	CO
Q1	Describe the steps involved in hypothesis testing. Why is it important to set a significance level before conducting a hypothesis test?	4	CO1
Q2	What is DAX, and how does it enhance data analysis in Power BI? Describe some commonly used DAX functions	4	CO4
Q3	What are outliers? How do you identify them in a box plot.	4	CO3
Q4	Explain the difference between Pandas Data Frames and NumPy arrays. In what situations would you prefer one over the other?	4	CO3
Q5	Given the following dataset of exam scores: 75, 82, 89, 91, 85, 78, 88, calculate the mean, median, mode, variance, and standard deviation	4	CO1
SECTION B **(40 Marks) 5 Questions-Each 10 Marks-One Choice-Attempt any 4 questions out of 5**			
Q6	Explain the difference between a null hypothesis and an alternative hypothesis. Why is the significance level important in hypothesis testing?	10	CO1
Q7	Two groups of students are given different study materials. Group A (n = 10) scores: 78, 82, 79, 81, 76, 85, 77, 84, 80, 83. Group B (n = 10) scores: 74, 75, 77, 76, 78, 72, 73, 74, 76, 75. Use a two-sample t-test to determine if there is a significant difference in average scores between the two groups at a 5% significance level.	10	CO2
Q8	What are the key building blocks of Power BI, and how do they contribute to effective data analysis and visualization?	10	CO4
Q9	What is a chi-square test, and in which situations is it used? Describe the difference between a chi-square test of independence and a chi-square goodness-of-fit test.	10	CO3
Q10	Give python code for a z-test based hypothesis testing to determine if a sample scores mean differs significantly from a population mean. Sample scores = [70, 80, 75, 85, 90, 72, 78, 82, 88, 76,74, 91, 85, 84, 79, 81, 77, 80, 83, 86,89, 91, 73, 75, 87, 78, 80, 82, 76, 74] Population mean = 75, Population standard deviation = 10 Note: use NumPy and SciPy libraries	10	CO3

SECTION C
(40 Marks) 2 Questions -Each 20 Marks - No Choice-Attempt all Questions

<table>
<tr><td>Q11</td><td>A study is conducted to assess the impact of three different fertilizers on crop yield. The yields (in kg) for each fertilizer type are recorded as follows:
• Fertilizer A: 30, 32, 28, 34, 29
• Fertilizer B: 40, 42, 41, 39, 38
• Fertilizer C: 25, 27, 26, 28, 29
Using a one-way ANOVA test at a 5% significance level, determine if there is a statistically significant difference in the mean crop yield among the three fertilizer groups.</td><td>20</td><td>CO3</td></tr>
<tr><td>Q12</td><td>A health researcher is interested in understanding the relationship between smoking status and exercise habits. The researcher collects data from a sample of 150 individuals, categorizing them into two groups based on smoking status (smoker, non-smoker) and two groups based on exercise habits (active, inactive). The data collected is summarized in the following contingency table:
<table>
<tr><td></td><td>Active</td><td>Inactive</td></tr>
<tr><td>Smoker</td><td>30</td><td>45</td></tr>
<tr><td>Non-Smoker</td><td>40</td><td>35</td></tr>
</table>
Using a chi-square test for independence, determine whether there is a significant association between smoking status and exercise habits at a significance level of 0.05.
Note: State clearly Null-Alternate hypothesis and conclusion</td><td>20</td><td>CO1</td></tr>
</table>