Roll No. | | | | | | | | | | |

ANNA UNIVERSITY (UNIVERSITY DEPARTMENTS)

B.E. / B.Tech. END SEMESTER EXAMINATIONS, NOVEMBER / DECEMBER 2024

MINOR DEGREE ON DATA SCIENCE

FIFTH SEMESTER

CSM507: FOUNDATIONS OF DATA SCIENCE WITH PYTHON

(Regulations 2019)

Time: 3 Hours **Answer ALL Questions** **Max. Marks: 100**

CO 1	To understand fundamentals and the process of data science.
CO 2	To comprehend different types and representation of data and analyze them.
CO 3	To apply inferential techniques to extrapolate information from the available data.
CO 4	To utilize the Python libraries for Data Wrangling.
CO 5	To interpret data and present it using visualization libraries in Python.

BL – Bloom's Taxonomy Levels

(L1 - Remembering, L2 - Understanding, L3 - Applying, L4 - Analysing, L5 - Evaluating, L6 - Creating)

PART- A (10 x 2 = 20 Marks)

(Answer all Questions)

Q. No	Questions	Marks	CO	BL
1	Outline the need for data preparation.	2	CO1	L1
2	Differentiate between supervised learning and unsupervised learning.	2	CO1	L1
3	Consider the data for analysis includes the attribute age in years. The age values for the data tuples in increasing order are 13, 15, 16, 19, 20, 21 and 23. Compute the mean of the data.	2	CO2	L2
4	Consider the data for analysis includes the attribute age in years. The age values for the data tuples in increasing order are 13, 15, 16, 19, 20, 21 and 23. Compute the midrange of the data.	2	CO2	L2
5	Outline probability with an example.	2	CO3	L1
6	What is chi-square test?	2	CO3	L1
7	Write a Python program to accept two distinct numbers find the greatest and display the result.	2	CO4	L2

8	Write a Python program to accept a number, find whether the number is an even number or not and display the result.	2	CO4	L2
9	What is Matplotlib?	2	CO5	L1
10	Write a note on Seaborn.	2	CO5	L1

PART- B (5 x 13 = 65 Marks)

(Answer all Questions)

Q. No	Questions	Marks	CO	BL
11 (a)	What is data science? Outline the steps in the data science process with a diagram.	13	CO1	L1
	OR			
11 (b)	How to analyze data using exploratory data analysis? Outline with an example.	13	CO1	L1
12 (a)	"Machine learning models use many types of data" Outline the types of data with an example.	13	CO2	L1
	OR			
12 (b)	What Is a normal distribution? Outline the properties of normal distribution.	13	CO2	L1
13 (a)	What is Analysis of Variance? Outline the steps in Analysis of Variance for one factor with an example.	13	CO3	L2
	OR			
13 (b)	Outline the steps in Bayesian classification with an example..	13	CO3	L2
14 (a)	Write a Python program to compute mean, median, mode, standard deviation and variance with an example. Validate the correctness of the program with an example.	13	CO4	L2
	OR			
14 (b)	What is data wrangling? How to do data wrangling using Python? Outline with an example and code snippets.	13	CO4	L2
15 (a)	Define a line plot and a scatter plot. Outline visualization using line plot and scatter plot with an example.	13	CO5	L2
	OR			

15 (b)	Define a density plot and a contour plot. Outline visualization using density plot and contour plot with an example.	**13**	**CO5**	**L2**

PART- C (1 x 15 = 15 Marks)

(Q.No.16 is compulsory)

Q. No	Question	Marks	CO	BL
16.	The values of x and their corresponding values of y are shown in the table below:			

x	*1*	*2*	*3*	*3*	*4*	*5*	*6*	*7*
y	*2*	*3*	*4*	*5*	*5*	*7*	*7*	*8*

	Find the least square regression line y = ax + b.	**15**	**CO2**	**L3**