

Register No : _____

THE KAVERY ENGINEERING COLLEGE
(An Autonomous Institution)

B.E/B .TECH DEGREE END SEMESTER THEORY EXAMINATIONS, NOV /DEC 2024

Third Semester

Information Technology

CS3352 - Foundations of Data Science

(Common to Computer Science and Engineering)

Regulations - 2021

Duration : 3 Hrs

Maximum Marks : 100 Marks

PART - A (ANSWER ALL QUESTIONS) (Marks 10*2=20)

<i>Q.No</i>	<i>Questions</i>	<i>BLT</i>	<i>CO</i>	<i>Marks</i>
1.	List the difference between structured data and unstructured data.	AN	1	2
2.	How do measuring central tendency?	AN	1	2
3.	Compare continuous and discrete variables with an example.	UN	2	2
4.	Define misleading graph.	RE	2	2
5.	Consider Helen sent 10 greeting cards to her friends and she received back 8 cards, what is the kind of relationship it is? Brief on it.	AN	3	2
6.	What is causation?	RE	3	2
7.	Define data wrangling.	RE	4	2
8.	What is pivot table?	RE	4	2
9.	What is the purpose of error bar function in Matplotlib? Give an example.	RE	5	2
10.	Define data visualization.	RE	5	2

PART - B (ANSWER ALL QUESTIONS) (Marks 5*13 = 65)

<i>Q.No</i>	<i>Questions</i>	<i>BLT</i>	<i>CO</i>	<i>Marks</i>
11.	a i Explain the steps in data process with a suitable diagram and example.	UN	1	6
	ii What is data warehousing? Outline the architecture of data warehousing with neat diagram	UN	1	7
	Or			
	b i How do you set the research goal, retrieving data and data preparation process in data science process?	UN	1	8
	ii Explain the benefits of data science.	UN	1	5
12.	a i Calculate the sum of square population standard deviation of the given x data value 13,10,11,7,9,11,9.	AP	2	7
	ii Calculate the sample standard deviation for the given data 7,3,1,0,4.	AP	2	6

Or

- b Suppose the IQ score have a bell-shaped distribution with a mean of 100 and standard deviation of 15 then calculate the following:
 (i) what percentage of people should have an IQ score between 85 and 115?
 (ii) What percentage of people should have an IQ score between 70 and 130?

AP 2 13

13. a Calculate the correlation coefficient for the heights of fathers(X) and their sons(y) with the data presented below:

AN 3 13

X	66	68	68	70	71	72	72
y	68	70	69	72	72	72	74

Or

- b i Examine the different types of regression analysis in detail.
 ii Inspect how correlation coefficient can be calculated for the quantitative data?

UN 3 7

UN 3 6

14. a Outline the basics of Numpy arrays with example.

UN 4 13

Or

- b Explain the hierarchical indexing with examples.

UN 4 13

15. a Explain about various visualization charts like line plots, scatter plots and histograms using Matplotlib with an example.

UN 5 13

Or

- b Explain in detail about 3-Dimensional plotting with an example.

UN 5 13

*PART - C (Marks 1*15 = 15)*

Q.No	Questions	BLT	CO	Marks
16.	a Assume that an $r=0.30$ describe the relationship between education level and estimate number of hours spent reading each work:	EV	3	15

Education level(X)	Weekly Reading Time(Y)
$\bar{X}=13$	$\bar{Y}=8$
$SSX=25$	$SSY=50$

- i) Determine the least square equation for predicting weekly report time from education level.
 ii) Faith's education level is 15. What is her predicted reading time?
 iii) Keegan's education level is 11. What is his predicted reading time?
 iv) Calculate the standard error estimate based on $n=35$ pairs of observation.
 v) Supply a rough interpretation of standard error estimate.

Or

- b Compose text and image annotations in python. Annotate an image of a cat with the label "domestic house cat".

EV 4 15