

END TERM EXAMINATION

FOURTH SEMESTER [BCA] JULY 2023

Paper Code: BCAT-212

Subject: Introduction to Data Science

Time: 3 Hours

Maximum Marks: 75

Note: Attempt five question in all including Q No 1 which is compulsory.**Select one question from each unit.**

Q1 Attempt any five parts:

(5×3=15)

- a) What is a Series and how is it different from a 1-D array, a list, and a dictionary.
- b) Differentiate Supervised and Unsupervised learning techniques
- c) Briefly explain Bar and whisker plot with example.
- d) Give use of pivot_table() in pandas by giving example.
- e) Specify any two python Libraries and their applications.
- f) Explain skewness and its importance.
- g) Given a Dataframe with rainfall data (day of the week and rainfall inches), write code to find the minimum, maximum, Q1, Q2, Q3, mean and standard deviation of rainfall.

UNIT-I

- Q2 a) Explain the roles and responsibilities of any six Data Science professionals. (6.5)
- b) Explain four major tasks in data pre-processing (6)
- Q3 a) What is Data Science Lifecycle? Explain all stages with diagram. (6.5)
- b) Describe any five data collection strategies. (6)

UNIT-II

- Q4 a) What are the various types of Data in Statistics? Explain with example (6.5)
- b) Explain and give formula for the following Measures of Variability of data: (6)
- i. Standard Deviation.
 - ii. Inter-quartile range
 - iii. Mean Absolute Deviation
 - iv. Variance
- Q5 a) Explain linear regression and write the code to plot true line by taking an example. (6.5)
- b) List various types of graph/chart available in the pyplot of matplotlib library for data visualization. Explain any two of them in brief. (6)

UNIT-III

- Q6 Give 4 ways of creating Numpy arrays (12.5)
- Give the code or syntax to Perform the following operation on two 2D numpy array array1 and array2 and 1D array array3.
- a. Add array1 and array2
 - b. Find sum of array1 elements over a given axis.
 - c. Find product of array2 elements over a given axis.
 - d. Change the dimension of an array3 to 2D.
 - e. Transpose the array created in part d.
 - f. Display 2 rows and third column of 2D array array1.

P.T.O.

BCAT-212
D.

- g. Join two 2D array along row.
- h. Convert array2 to 1D array.
- i. Split an array 1 into multiple subarrays

- Q7 Give 4 ways of creating series (12.5)
- a) Write python code to create the following series
 - 101 Harsh
 - 102 Arun
 - 103 Ankur
 - 104 Harpal
 - 105 Divya
 - 106 Jeet
 - b) Show details of 1st 3 employees using head function
 - c) Show details of last 3 employees using tail function
 - d) Show details of 1st 3 employees without using head function
 - e) Show details of last 3 employees without using tail function
 - f) Show value of index no 102.
 - g) Show 2nd to 4th records.
 - h) Show values of index no=101,103,105.
 - i) Show details of "Arun"

UNIT-IV

- Q8 Give 4 ways to create data frame (12.5)

Create a dataframe to store data for 10 students

| Name | Age | Semester I marks out of 600 | Semester II marks out of 500 | Attendance |
|------|-----|-----------------------------|------------------------------|------------|
| | | | | |

Write program to perform following operations on above dataframe:

- a. Display details of students who scored more than 560 marks in semester I
- b. Display details of students who scored less than 250 marks in semester II
- c. Display details of student who scored minimum marks in semester II
- d. Display details of student who scored maximum marks in semester II
- e. Display details of students whose attendance is more than 75.
- f. Display details of students whose attendance is less than 50.
- g. Insert 2 new records in dataframe
- h. Add a new column corresponding to percentage of marks of both semester.
- i. Add a new column corresponding to grades:-

| | |
|---------------------|-------|
| Both sem percentage | Grade |
| >=90 | O |
| >=75 to <90 | A+ |
| >=60 to <75 | A |
| >=50 to <60 | B+ |
| >=40 to <50 | B |
| < 40 | F |

P.T.O.

Q9

Create the following DataFrame Sales containing year-wise sales figures for five salespersons in INR. Use the years as column labels, and salesperson names as row labels. (12.5)

| | 2014 | 2015 | 2016 | 2017 |
|----------------|-------|-------|-------|-------|
| Madhu | 100.5 | 12000 | 20000 | 50000 |
| Kusum | 150.8 | 18000 | 50000 | 60000 |
| Kinshuk | 200.9 | 22000 | 70000 | 70000 |
| Ankit | 30000 | 30000 | 10000 | 80000 |
| Shruti | 40000 | 45000 | 12500 | 90000 |

- Display the row labels of Sales.
- Display the column labels of Sales.
- Display the dimensions, shape, size and values of Sales.
- Display the last two rows of Sales.
- Display the first two columns of Sales.
- Change the DataFrame Sales such that it becomes its transpose.
- Add data to Sales for salesman Sumeet where the sales made are [196.2, 37800, 52000, 78438] in the years [2014, 2015, 2016, 2017] respectively.
- Delete the data for the year 2014 from the DataFrame Sales.
- Update the sale made by Shruti in 2017 to 100000.
- Write the values of DataFrame Sales to a comma-separated file SalesFigures.csv on the disk. Do not write the row labels and column labels.
- Change the name of the salesperson Ankit to Vivaan and Kinshuk to Shailesh.
- Delete the data for salesman Madhu from the DataFrame Sales.

BCAT-212
P3/3