

21BDH101

Reg No : .....

**CHOICE BASED CREDIT SYSTEM**

**M.Sc. BIG DATA ANALYTICS FIRST SEMESTER DEGREE EXAMINATION  
MARCH 2022**

**Data Warehousing and Data Mining**

**Duration: 3 Hrs**

**Max Marks:70**

**PART - A**

**I. Answer any EIGHT of the following:**

**(2×8= 16)**

- a) What is the role of Data Mining in Web Search Engines?
- b) How can we convey data to users effectively?
- c) List out the different ways in which OLAP-style analysis can be fused with data mining techniques.
- d) Give an example for frequent itemset mining and demonstrate it.
- e) Compare OLAP systems and statistical databases.
- f) What are the components of belief networks?
- g) Demonstrate the working of prepruning approach.
- h) What do you mean by Bayesian Classification?
- i) Why global detection is important? Describe with an example.
- j) Define a data matrix.

**PART - B**

**Answer any FOUR questions :**

**(6×4= 24)**

2. What is Data preprocessing? Explain the various data preprocessing techniques.
3. Illustrate FP growth algorithm.
4. Demonstrate data cube materialization.
5. Explain the challenges in clustering high dimensional data.
6. Describe Bayesian belief networks.
7. How does data mining help in financial data analysis?

**PART - C**

**Answer any THREE questions :**

**(10×3= 30)**

8. Describe the steps involved in data mining when viewed as a knowledge discovery process with an illustration.
9. Explain three areas of basic statistical descriptions.
10. Define metadata. Explain the contents of a metadata repository with a neat diagram.
11. What are the applications of pattern mining?
12. Explain the techniques to improve classification accuracy.

\*\*\*\*\*

21BDH102

Reg No : .....

**CHOICE BASED CREDIT SYSTEM**

**M.Sc. BIG DATA ANALYTICS FIRST SEMESTER DEGREE EXAMINATION**

**MARCH 2022**

**Python Programming for Data Analytics**

**Duration:3 Hrs**

**Max Marks:70**

---

**PART - A**

**I. Answer any EIGHT of the following:**

**(2×8= 16)**

- a) What do you mean by a recursive function?
- b) What are dunder methods?
- c) What is negative indexing in Numpy arrays?
- d) List any four merge function arguments in Pandas.
- e) Mention any four significant features of the Pandas library.
- f) Write the Datetime format specification for:(i) %Y (ii) %d
- g) What is the function of the rolling() attribute of Series and DataFrame objects?
- h) Give an example to depict how to set ticks along x axis and y axis.
- i) Write the Python code to find the factorial of a number using recursion.
- j) What is the function of Seaborn package ?

**PART - B**

**Answer any FOUR questions :**

**(6×4= 24)**

2. Explain NumPy's sort() function with its syntax.
3. What are the different ways to indicate the missing data? Explain.
4. Write a note on filtering and transformation.
5. How can you create different Subplot Sizes in Matplotlib? Explain.
6. What is a package? Describe how packages are used from another program.
7. Explain scatter plots with plt.scatter.

## PART - C

Answer any THREE questions :

(10×3= 30)

8. (a) Explain any 3 NumPy's Array creation functions with examples. (6)  
(b) What is the significance of using broadcasting? Explain the rules of broadcasting with examples. (4)
9. (a) Explain any six built-in string methods. (6)  
(b) With suitable examples explain vectorization of string functions in Pandas. (4)
10. (a) Explain the different parameters of pandas crosstab() function. (5)  
(b) Describe multilevel pivot tables. (5)

11. (a) Consider the given array:

```
x = np.array([1, 2, 3, 4, 5])
```

Write the expected output of the following operations on the array:

(i)  $x \leq 3$  (ii)  $(2 * x) == (x ** 2)$  (iii)  $x \geq 4$

(b) Given the following Series:

A	100
B	200
C	300
D	400
E	500

Write the command to create above Series and then double the value in series and store in another series named series2. (6+4)

12. (a) What are sorted and unsorted indices? Explain. (6)  
(b) Explain Pandas Multiply Indexed Series. (4)

\*\*\*\*\*

21BDH103

Reg No : .....

**CHOICE BASED CREDIT SYSTEM**

**M.Sc. BIG DATA ANALYTICS FIRST SEMESTER DEGREE EXAMINATION**

**MARCH 2022**

**Quantitative Techniques for Data Science and Data Visualisation**

**Duration: 3 Hrs**

**Max Marks:70**

---

**PART - A**

**I. Answer any EIGHT of the following:**

**(2×8= 16)**

- a) Define inverse of the matrix.
- b) Define orthogonal matrix with an example.
- c) What are independent events? Give example.
- d) Name different measures of dispersion.
- e) Define permutation and combination.
- f) What is open end class interval? Give example.
- g) State the conditions for Binomial distribution tends towards normal distribution.
- h) State the addition theorem of expectations for two random variables.
- i) List the predefined ggplot color palette.
- j) Define brainstorming.

**PART - B**

**Answer any FOUR questions :**

**(6×4= 24)**

2. Derive the general form of solving system of equations.
3. Explain negative and positive skewness also indicate it using curves. Calculate coefficient of skewness from the following:  
Median=18.8inches, Q1=14.6inches, Q3=5.2inches.
4. Three factories produce light bulbs to supply to the market. Factory A produces 20%, 50% of the tools are produced in factory B and 30% in factory C. 2% of the bulbs produced in factory A, 1% of the bulbs produced in factory B and 3% of the bulbs produced in factory C are defective. A bulb is selected at random in the market and found to be defective. What is the probability that this bulb was produced by factory B?

5. Define set. Elaborate operation's on sets.
6. Explain K-S test.
7. Explain various types of representation of data.

### PART - C

**Answer any THREE questions :**

**(10×3= 30)**

8. Define Cartesian product. Write the important results of Cartesian product.
9. Give the empirical relation between mean, median and mode. Discuss the merits and demerits of Arithmetic mean, Median and Mode.
10. For a large sample,
  - a) Explain the procedure of testing whether the population proportion has a given value.
  - b) Explain the procedure of testing equality of proportions of two populations.
11. According to a theory in genetics, the population of beans of 4 types A, B, C and D are in the ratio 9:3:3:1. In an experiment among 1600 beans, the frequency of beans of each of the above 4 types were 882, 313, 287, 118. Test the results support the theory.
12. Elaborate the guiding principle of data visualization.

\*\*\*\*\*

**CHOICE BASED CREDIT SYSTEM****M.Sc. BIG DATA ANALYTICS FIRST SEMESTER DEGREE EXAMINATION****MARCH 2022****Optimization Techniques****Duration: 3 Hrs****Max Marks:70****PART - A****I. Answer any EIGHT of the following:****(2×8= 16)**

- a) What is optimization?
- b) What is initial basic solution to simplex problem?
- c) What is quadratic programming?
- d) State any three models of queuing theory.
- e) Define seperable programming.
- f) What do you mean by decision tree analysis?
- g) What is model?
- h) Write any two limitations of simulation?
- i) Mention the nature of Operations Research.
- j) Define slack variable.

**PART - B****Answer any FOUR questions :****(6×4= 24)**

2. What are the applications of Simulation Technique? Explain.
3. What are the advantages and limitations of LPP?
4. A retired person wants to invest up to an amount of Rs. 30,000 in fixed income securities. His broker recommends investing in two bounds. Bound a yielding 7% and bond B yielding 10%. After some consideration. He decides to decides to invest at the most Rs. 12,000 in bond B and at least Rs. 6,000 in bond A. He also wants the amount invested in bond A to be at least equal to the amount invested in bond B. What should the broker recommend if the investor wants to maximize his return on investment? Solve graphically.

5. Write an explanatory note on Kuhn Tucker Conditions.
6. Write a note on LPP and Simulation Models.
7. Explain the differences between linear programming and non-linear programming.

**PART - C**

**Answer any THREE questions :**

**(10×3= 30)**

8. Solve the below LPP model under Graphical solution:

$$\text{Max } Z = 50x_1 + 80x_2$$

$$\text{Subject to: } x_1 \leq 80$$

$$x_2 \leq 60$$

$$5x_1 + 6x_2 \leq 600$$

$$x_1 + 2x_2 \leq 160$$

$$x_1, x_2 \geq 0$$

9. Define Operations Research. Explain its characteristics.
10. "Optimization problem can be linear, nonlinear, geometric and quadratic programming problems". Justify.
11. Using the Simplex method obtain the first two tables for the given problem.

$$\text{Max } Z = 3X_1 + 2X_2$$

$$\text{Sub to } X_1 + X_2 \leq 4$$

$$X_1 - X_2 \leq 2$$

$$\text{And } X_1, X_2 \geq 0$$

12. Obtain the dual for following LPP

$$\text{Min } Z = 7x_1 + 3x_2 + 8x_3 \text{ s.t.c}$$

$$8x_1 + 2x_2 + x_3 \geq 3$$

$$3x_1 + 6x_2 + 4x_3 \leq -4$$

$$4x_1 + x_2 + 5x_3 = 1$$

$$x_1, x_2 \geq 0$$

\*\*\*\*\*