

19BDH101

REG.NO.....

END SEMESTER EXAM – NOVEMBER 2019

I M.Sc BIG DATA ANALYTICS – I SEMESTER

**SUBJECT: DATA WAREHOUSE AND DATAMINING**

**Code: 19BDH101**

Time: 3 Hrs

Max. Marks: 70

**PART - A**

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

**8X2=16**

- a) Data mining is also known as knowledge discovery. Justify.
- b) Define Null rule?
- c) What is Data visualization?
- d) What is PCA?
- e) What is a Data Model?
- f) Define Association rules?
- g) Define Data classification?
- h) What is the need for cluster analysis?
- i) What are frequent patterns?
- j) What are outliers?

**PART - B**

**II. Answer any SIX of the following:**

**6X4=24**

1. Explain why Data warehouse is necessary?
2. Write a note on Multidimensional data mining.
3. Differentiate between OLTP & OLAP?
4. Explain correlation analysis using lift.
5. Discuss the mining methodology that mines colossal patterns.
6. What is Information gain? Explain in detail.
7. Write the characteristics of Partitioning methods.
8. Discuss any 4 Data mining applications.

**PART - C**

**Answer any FIVE questions:**

**5X6=30**

1. Differentiate between Data warehousing & Data base. Explain their similarities.
2. Explain Range, Quartile & Interquartile range with examples.
3. A database has 5 transactions. Let min sup=60% & Min conf = 80%  
TID Items bought  
T100 {M, O, N, K, E, Y}  
T200 {D, O, N, K, E, Y}  
T300 {M, A, K, E}  
T400 {M, U, C, K, Y}

Find all frequent items using Apriori and FP-Grwoth respectively. Compare the efficiency of 2 mining processes.

4. Explain discriminative frequent pattern based classification.
5. Explain any 3 Data mining types in detail.
6. Explain various types of outliers and explain them briefly.
7. Suppose that values for a given set of data is grouped into intervals. Then, intervals and corresponding frequencies are as follows :

Age	Frequency
1-5	200
6-15	450
16-20	300
21-50	1500
51-80	700
81-110	44

Compute an approximate median value for the data.

\*\*\*\*\*

END SEMESTER EXAM – NOVEMBER 2019

I M.Sc BIG DATA ANALYTICS – I SEMESTER

**SUBJECT: PYTHON PROGRAMMING FOR DATA SCIENCE**

Code: 19BDH102

Time: 3 Hrs

Max. Marks: 70

**PART - A****1. Answer any EIGHT of the following:****8X2=16**

- a) What are magic commands in IPython.
- b) Write the use of % magic command.
- c) Describe NumPy
- d) Define structural data in NumPy.
- e) What is Pandas series object.
- f) What is hierarchical indexing?
- g) Define data aggregation.
- h) Define Time Series.
- i) Define contour plot.
- j) Define line plot.

**PART - B****2. Answer any SIX of the following:****6X4=24**

1. Write a note on reshaping of arrays in NumPy.
2. Explain NumPy's Fancy Indexing with an example.
3. Write a note on operating on Data in Panads.
4. Explain how to remove duplicates in a given Data Frame.
5. Write a note on Pivot tables and explain.
6. Write a brief note on frequencies in time series.
7. Explain any four data frame plot arguments.
8. Explain Scatter plot.

**PART - C****3. Answer any FIVE questions:****5X6=30**

1. What are the different NumPy standard data types.
2. Explain Boolean arrays and operators.
3. What are the different ways to combine data in panda's object? Explain.
4. Explain the various types of data transformation used in the Data Frames.
5. Write a note on filtering and transformation.
6. Expain the concept of "Shifting" in time series with suitable example.
7. Explain plot function in matplotlib.

\*\*\*\*\*

END SEMESTER EXAM – NOVEMBER 2019

I.M.Sc BIG DATA ANALYTICS – I SEMESTER

**SUBJECT: STATISTICS FOR DATA SCIENCE & DATA VISUALIZATION**

Code: 19BDH103

Time: 3 Hrs

Max. Marks: 70

**PART - A****1. Answer any TEN of the following:****10X2=20**

- Explain the principle of inclusion and exclusion.
- Define Power set and write it for the Set  $A = \{1,3,5,8\}$ .
- What is Eigen Value and Eigen vector of a matrix?
- What is tabulation? How it is different from classification?
- If mean=30, mode=25, find median.
- What do you mean by inclusive and exclusive class intervals.
- Distinguish between two way classification and dichotomy.
- Under what conditions Binomial distribution tends to normal distribution?
- What is MGF? Give the formula for it.
- If  $E(X) = 6$ ,  $E(X^2) = 40$ , Find  $V(5X-4)$
- What is the need of data visualization?
- What are the limitations of diagrams and graphs?

**PART - B****Answer any FOUR of the following:****4X5=20**

- Explain the properties of relation with example.
- In how many ways can the letters of the word REARRANGE be arranged? How many of these begin with R and end with E? In how many arrangements of all A's occur together?
- Explain negative and positive Skewness calculate quartile deviation and coefficient of Skewness from the following: Median = 18.8 inches  $Q_1 = 14.6$  inches,  $Q_3 = 5.2$  inches.
- Explain Chi Square test for independence of attributes.
- If a random variable X takes the values 10 and 20 with respective probabilities 0.6 and 0.4, find  $E(X)$  and  $SD(X)$ .
- How to build a graph in ggplot2 with desired color and scale?

**PART - C****Answer any THREE questions:****3X10=30**

- Examine the consistency of the set of evaluations.  $5X-3Y+7Z=4$ ,  $3X+26Y+2Z=9$ ,  $7X+2Y+10Z=5$ . 5M.
  - Write properties of determinant of a matrix. 5M.

2. a) Given  $A = \{1, 3, 5, 9, 10, 18\}$ ,  $B = \{3, 5, 10\}$ ,  $C = \{9, 18, 3\}$ . Find 5M.

- i) Number of subsets of B ii)  $A - B$ .  
iii)  $A \cap (B \cup C)$  iv)  $A \cup (B - C)$  v)  $A \cap B \cap C$

b) Explain the different types of relation with example. 5M.

3. a) A can hit a target 2 times with 5 shots. B can hit it 3 times with 4 shots and C can hit it 5 times with 8 shots. If they fire at a Volley, What is the probability that at least one of them hits it? 5M.

b) What is classification of data? What are its functions? 5M.

4. a) Write the merits and demerits of standard deviation. 5M

b) Explain frequency distribution, class limits, class mark, width of a class and cumulative frequency with example. 5M

5. a) From the following data regarding heights of randomly selected boys and girls of SSLC class test, whether SSLC boys on an average are taller than SSLC girls. 5M

	Boys	Girls
Sample size	9	12
Mean height ( Cms.)	171	169
Standard deviation ( Cms.)	3	2

b) What are the different methods used to test the normality of data? 5M

\*\*\*\*\*

END SEMESTER EXAM – NOVEMBER 2019  
I.M.Sc. BIG DATA ANALYTICS – I SEMESTER  
SUBJECT: MANAGEMENT SCIENCE  
Code: 19BDS104.

Time: 3 Hrs

Max. Marks: 70

**PART - A**

1. Answer any EIGHT of the following:

8X2=16

- a) State any 2 properties of dual LPP.
- b) In a TP when do you say that a BPS is
  - i) degenerate and ii) non - degenerate
- c) Explain the importance of sensitivity analysis.
- d) What do you mean by balanced and unbalanced assignment problem.
- e) What are the common errors in drawing network diagrams?
- f) What is pay off table?
- g) Write any 2 advantages of simulation.
- h) Explain steady state and transient state.
- i) What is the indicator of multiple solution in the optional simplex method?
- j) What is travelling sales man problem?

**PART - B**

II. Answer any THREE questions.

8X3=24

1) a) Mr. Deshpande has decided to start a yoga class. If more than 10 members are enrolled, the hall booked for the session can accommodate maximum of 50 participants. A special personal counselling (if 10 minutes for each individuals who opt for it) is available after the yoga session and Mr. Deshpande has decided to spend maximum of 2 hours to this counselling session. The fees per day per person is ₹ 150 for those who opt for personal counselling and ₹100 for those who enrol to yoga class without personal counselling. Formulate the LPP to maximise per day income of Mr. Deshpande. ~~(5M+3M)~~ 8

2. a) Solve the following LPP using simplex method.

$$\text{Min } Z(x) = 3x_1 - 4x_2 + x_3$$

$$\text{Subjected to } x_1 + 2x_2 + x_3 \leq 20$$

$$x_2 + x_3 \leq 35$$

$$x_1 + x_2 - x_3 \leq 18 \quad x_1, x_2, x_3 \geq 0.$$

b) Explain Queue discipline.

(5M+3M)

3) Explain Hungarian method of solving an A.P.

(8M)

4) a) What are the objectives of time analysis in networking.

b) How Monte Carlo simulation can be applied in inventory control.

(6M+2M)

5) a) Explain explosive state.

b) Define basic solution. Find all basic solutions for the system :  $x_1 + 2x_2 + 3x_3 = 4$ 

$$(0, -2/5, 8/5)$$

$$2x_1 - x_2 + x_3 = 2$$

$$(2/5, 0, 6/5)$$

$$(8/5, 4/5, 0)$$

*Answer as per* Part - C

6) a) Explain the algorithm of simplex method. 3x10=30

b) Solve graphically

$$\begin{aligned} \text{Min } Z &= 3x + 5y \\ x + y &= 200 \\ x \leq 80, y \geq 60, x \geq 0, y \geq 0 \end{aligned}$$

(5M+5M)

7) a) Show that the following LP has unbounded solution,

$$\begin{aligned} \text{Max } Z &= 2x + y \\ x - y &\leq 10, \quad 2x - y \leq 10, \quad x, y \geq 0. \end{aligned}$$

b) Obtain the dual of primal

$$\begin{aligned} \text{Max } Z &= 6x_1 + 5x_2 + 4x_3 \\ \text{Subjected to } 3x_1 + 8x_2 + 4x_3 &= 15 \\ 4x_1 + 2x_2 - 7x_3 &= 10 \quad x_1, x_2, x_3 \geq 0. \end{aligned}$$

(5M+5M)

8) Solve the following TP

Factory	Warehouse			Available
	1	2	3	
A	15	26	13	350
B	3	7	8	100
C	9	4	3	110
Required	80	150	330	

(10M)

9) a) Describe NWCR to obtain initial B.F.S. to a T.P.

b) How to convert maximization AP into minimization AP?

Solve the assignment problem.

	1	2	3	4
1	12	30	21	15
2	18	23	9	31
3	43	25	24	21
4	23	30	28	14

(5M+5M)

10) A project has following characteristics.

Activity	Most optimistic time	Most pessimistic time	Most likely time
(1-2)	1	5	1.5
(2-3)	1	3	2
(2-4)	1	5	3
(3-5)	3	5	4
(4-5)	2	4	3
(4-6)	3	7	5
(5-7)	4	6	5
(6-7)	6	8	7
(7-8)	2	6	4
(7-9)	5	8	6
(8-10)	1	3	2
(9-10)	3	7	5

Construct a PERT network. Find critical path and variance of each event. Find the project duration at 95% probability.

(10M)

\*\*\*\*\*