NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA

(An Autonomous Institute Affiliated to AKTU, Lucknow)

B.Tech

SEM: VI - THEORY EXAMINATION (2022-2023)

Subject: Data Analytics

Time: 3 Hours   Max. Marks: 100

General Instructions:

IMP: *Verify that you have received the question paper with the correct course, code, branch etc.*

1. *This Question paper comprises of **three Sections -A, B, & C.** It consists of Multiple Choice Questions (MCQ's) & Subjective type questions.*

2. *Maximum marks for each question are indicated on right -hand side of each question.*

3. *Illustrate your answers with neat sketches wherever necessary.*

4. *Assume suitable data if necessary.*

5. *Preferably, write the answers in sequential order.*

6. *No sheet should be left blank. Any written material after a blank sheet will not be evaluated/checked.*

SECTION A   20

1. Attempt all parts:-

1-a.   A collection of information about a related topic is referred to as a_____   1
        (CO1)

        (a) Visualisation

        (b) Analysis

        (c) Conclusion

        (d) Data

1-b.   A column is a _____ representation of data.(CO1)   1

        (a) Horizontal

        (b) Diagonal

        (c) Vertical

        (d) Top

1-c.   A course is being rated by students on a scale of 1 star to 5 stars. What is the   1
        level of measurement being used in this case?(CO2)

        (a) Ordinal

(b) Ratio

(c) Nominal

(d) Interval

1-d.   Inconsistent Data is (CO2)   1

(a) Strucured Data

(b) Un structured Data

(c) Semi Structured Data

(d) None of the above

1-e.   To remove noise and inconsistent data......is needed. (CO3)   1

(a) Data Cleaning

(b) Data Transformation

(c) Data Reduction

(d) Data Integration

1-f.   Multiple data sources may be combined is called as........(CO3)   1

(a) Data Reduction

(b) Data Cleaning

(c) Data Integration

(d) Data Transformation

1-g.   Find the outlier s in the given dataset: 21, 15, 18,11,13,71 (CO4)   1

(a) 11

(b) 21

(c) 71

(d) 18

1-h.   PCA is a........... (CO4)   1

(a) Non linear method

(b) Linear method

(c) Continuous method

(d) Repeated method

1-i.   Which of these is not a datatype in Tableau? (CO5)   1

(a) Integer

(b) Character

(c) Float

(d) String

1-j.   The most popular data visualization library in python is _____(CO5)   1

        (a) matinfolib

        (b) matplotlib

        (c) matpiplib

        (d) pip

2. Attempt all parts:-

2.a.   What is structured data? (CO1)   2

2.b.   What does categorical data mean? (CO2)   2

2.c.   Explain Data Quality Assessment. (CO3)   2

2.d.   Why do we perform exploratory data analysis? (CO4)   2

2.e.   Define Tree Map? (CO5)   2

<div align="center">

SECTION B   **30**

</div>

3. Answer any <u>five</u> of the following:-

3-a.   Explain the role of Data Science in various fields.(CO1)   6

3-b.   Explain why is Big Data Important?(CO1)   6

3-c.   Differetiate between categorical data and numerical data. (CO2)   6

3-d.   Explain the advantages of data manipulation. (CO2)   6

3.e.   Explain data cleaning and why is it important?(CO3)   6

3.f.   Why organizations use time series analysis? Explain 3 examples of time series   6
        analysis. (CO4)

3.g.   Differentiate parameters and filters in Tableau.(CO5)   6

<div align="center">

SECTION C   **50**

</div>

4. Answer any <u>one</u> of the following:-

4-a.   Explain the three types of Big data. Describe the features of Big Data in   10
        detail.(CO1)

4-b.   Explain Data Science phases and lifecycle and write the names of tools for Data   10
        Science. (CO1)

5. Answer any <u>one</u> of the following:-

5-a.   Explain various types of data in detail. (CO2)   10

5-b.   Explain the key requirements for becoming a Data Analyst. (CO2)   10

6. Answer any <u>one</u> of the following:-

6-a.   Describe noisy data. Which techniques are used to handle noisy data? (CO3)   10

6-b.   Explain about Data Integration Hierarchy. (CO3)   10

7. Answer any <u>one</u> of the following:-

7-a.   Can PCA be used to reduce the dimensionality of a highly nonlinear dataset? Explain various limitations of PCA. (CO4)   10

7-b.   Differentiate between data processing, data preprocessing and data wrangling. (CO4)   10

8. Answer any <u>one</u> of the following:-

8-a.   Explain the uses of calculated fields? What is the main difference between calculated fields and table calculations in Tableau? (CO5)   10

8-b.   Describe Tableau Server. Discuss its components. (CO5)   10