Printed Page:-04 Subject Code:- AEC0713

Roll. No:

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA

(An Autonomous Institute Affiliated to AKTU, Lucknow)

B.Tech

SEM: VII - THEORY EXAMINATION (2024 - 2025)

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. Which of the following is effective way of checking validity of data analysis? (CO1, K1) 1

(a) Re-run the analysis

(b) Review the code

(c) Check the sensitivity

(d) All of the mentioned

1-b. Which of the following step is not required in data analysis? (CO1,K1) 1

(a) Synthesize results

(b) Create reproducible code

(c) Interpret results

(d) None of the mentioned

1-c. Which type of data is organized in a well-defined structure with fixed fields and records? (CO2 , k2) 1

(a) Structured data

(b) Semi-structured data

(c) Unstructured data

(d) Numeric data

1-d. _____is not a measure of central tendency. (CO2, K2) 1

(a) Mean

.

(b) Median
(c) Variance
(d) Mode

1-e. What does KDD stand for in the context of data mining? (CO3, K3) 1

(a) Knowledge Discovery in Data
(b) Knowledge Design and Development
(c) Knowledge Data Determination
(d) Knowledge Database Deployment

1-f. What is the primary purpose of data integration in data preprocessing? (CO3,K3) 1

(a) Reducing data dimensionality
(b) Combining data from different sources
(c) Converting data to a standard format
(d) Cleaning noisy data

1-g. Which of the following is a technique for handling missing data by replacing missing values with the mean of the available data? (CO4 ,K4) 1

(a) Deletion
(b) Imputation
(c) Interpolation
(d) Transformation

1-h. What is an outlier in a dataset? (CO4 ,K4) 1

(a) A common data point
(b) A data point that conforms to the data distribution
(c) A data point that deviates significantly from the majority of data
(d) A missing data point

1-i. Which color scheme is best for emphasizing differences in data values? (CO5,K3) 1

(a) Monochromatic
(b) Analogous
(c) Complementary
(d) Triadic

1-j. What is the primary purpose of a radar chart? (CO5 ,K3) 1

(a) Comparing the performance of multiple categories over time
(b) Showing the distribution of a single variable
(c) Visualizing hierarchical data structures
(d) Displaying the relationship between multiple variables with a radial layout

2. Attempt all parts:-

2.a. Describe the Datafication process. (CO1, K1) 2

2.b. What is the difference between structured, semi-structured, and unstructured data? Provide examples for each type. (CO2, K2) 2

.

2.c.	Discuss the significance of the p-value in statistical analysis. (CO3,K3)	2
2.d.	What are outliers, and how can they be identified in a dataset? (CO4, K4)	2
2.e.	What is the primary purpose of data visualization? (CO5,K3)	2

SECTION-B 30

3. Answer any five of the following:-

3-a.	Why is data science important? Brief the Impact of Data Science on Society. (CO1,K1)	6
3-b.	What are the challenges in Data science? Write difference between Data Science vs Data Analytics. (CO1,K1)	6
3-c.	What is the margin of error in statistical analysis, and how is it calculated in hypothesis testing? (CO2 ,K2)	6
3-d.	Explain the different types of data: structured, semi-structured, and unstructured. Provide examples of each and discuss how they are used in Data Science. (CO2, K2)	6
3.e.	How can data visualization techniques enhance the understanding of data and its patterns? Provide examples of when and why different visualization methods should be used.(CO3,K3)	6
3.f.	Explain the importance of multivariate exploratory data analysis (EDA), and provide an example of a technique or visualization used in multivariate EDA. (CO4,K4)	6
3.g.	Discuss the role of the "DATEDIFF" function in Tableau calculations. How is it used to analyze date-related data? (CO5, K3)	6

SECTION-C 50

4. Answer any one of the following:-

4-a.	Describe the Data Science lifecycle. Discuss each step involved and its significance in building data-driven models. (CO1,K1)	10
4-b.	Explain the evolution of Data Science. How has it transformed from traditional data analysis to the modern, integrated field of Data Science? (CO1,K1)	10

5. Answer any one of the following:-

5-a.	Explain the concepts of covariance and correlation. How do they differ, and what do they indicate about the relationship between two variables?. (CO2,K2)	10
5-b.	Explain the concept of "transactional data" and provide an example of an industry where transactional data is commonly used. (CO2,K2)	10

6. Answer any one of the following:-

6-a.	Explain the concept of data cleaning and the various techniques used to handle missing values, noisy data, and inconsistent data. How do these cleaning steps improve the quality of the dataset for further analysis and modeling? (CO3,K3)	10
6-b.	How does the KDD process help in transforming raw data into actionable knowledge? Describe the stages involved in KDD and provide examples of how each stage contributes to discovering hidden patterns in the data. (CO3,K3)	10

7. Answer any one of the following:-

7-a.	What is dimensionality reduction, and why is it important? Discuss techniques such as Principal Component Analysis (PCA), Factor Analysis (FA), and Linear Discriminant Analysis (LDA) for reducing the complexity of large datasets. (CO4, K4)	10
7-b.	Explain the significance of data wrangling in the context of acquiring and preparing data from the web or the internet. Discuss the role of APIs in this process. (CO4,K4)	10

8. Answer any one of the following:-

8-a.	Describe the different Tableau calculation features, including SUM, AVG, and aggregate functions. How can you create custom calculations and fields in Tableau, and how do these calculations enhance your visualizations? (CO5,K3)	10
8-b.	Define the "WINDOW_" function in Tableau calculations and describe how it can be used to perform calculations within specific data windows. (CO5,k3)	10

.