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: IV - THEORY EXAMINATION (2023 - 2024)

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 performed by Data Scientist? (CO1) 1

(a) Define the question
(b) Create reproducible code
(c) Challenge results
(d) All of the mentioned

1-b. Which of the following can be used for data analysis model? (CO1) 1

(a) CRAN
(b) CPAN
(c) CTAN
(d) All of the mentioned

1-c. The primary purpose of a histogram in data analysis. (CO2) 1

(a) To determine the mode of a dataset
(b) To measure central tendency
(c) To calculate the standard deviation
(d) To visualize the distribution of data

1-d. Which of the following is an example of spatial data? (CO2) 1

(a) Temperature measurements
(b) Stock market prices
(c) Geographic coordinates

(d) Social media posts

1-e. Inconsistent data in a dataset may arise due to: (CO3) 1

(a) Data duplication
(b) Outliers
(c) Missing values
(d) Data compression

1-f. Which of the following is an example of numerosity reduction in data preprocessing? (CO3) 1

(a) Principal Component Analysis (PCA
(b) Feature selection
(c) Data clustering
(d) Histogram equalization

1-g. What is the primary objective of handling missing data in a dataset? (CO4) 1

(a) Reducing data storage space
(b) Improving data visualization
(c) Ensuring data quality and analysis accuracy
(d) Enhancing data security

1-h. Which technique is commonly used for identifying redundancy among variables by measuring their linear relationships? (CO4) 1

(a) Principal Component Analysis (PCA)
(b) Factor Analysis (FA)
(c) Correlation analysis
(d) Linear Discriminant Analysis (LDA)

1-i. Which chart is most suitable for showing the distribution of a single continuous variable? (CO5) 1

(a) Bar chart
(b) Line chart
(c) Scatter plot
(d) Histogram

1-j. What type of chart is used to show the relationship between two variables in a two-dimensional space? (CO5) 1

(a) Pie chart
(b) Radar chart
(c) Scatter plot
(d) Heat map

2. Attempt all parts:-

2.a. What is the Need for Data Science ? (CO1) 2

2.b. Define variance and discuss its importance in measuring data variability. (CO2) 2

2.c.	Discuss the differences between a bar chart and a pie chart for data representation. (CO3)	2
2.d.	What is the primary objective of handling missing data in a dataset.(CO4)	2
2.e.	Explain the key difference between a dimension and a measure in Tableau calculations. (CO5)	2

SECTION-B 30

3. Answer any five of the following:-

3-a.	Write Brief History of Data Science. (CO1)	6
3-b.	Difference between Data Science vs Data Analytics. (CO1)	6
3-c.	Explain the concept of a normal distribution and its characteristics. How it is used in statistical analysis? (CO2)	6
3-d.	How can you import data from a CSV file into R or Python, and what are the benefits of using these programming languages for data manipulation? (CO2)	6
3.e.	Describe the different types of data attributes and their importance in data analysis. (CO3)	6
3.f.	Explain what is Principle Component Analysis (PCA) with any one of its applications.(CO4)	6
3.g.	Describe the process of converting a dimension into a measure in Tableau calculations. In what situations might this transformation be necessary? (CO5)	6

SECTION-C 50

4. Answer any one of the following:-

4-a.	How to solve a problem in Data Science using Machine learning algorithms? (CO1)	10
4-b.	Explain Equality, diversity and inclusion in data science ? (CO1)	10

5. Answer any one of the following:-

5-a.	Explain the differences between structured, semi-structured, and unstructured data. Provide examples for each type. (CO2)	10
5-b.	Explain the central limit theorem and its implications in statistical analysis. How does it affect the distribution of sample means? (CO2)	10

6. Answer any one of the following:-

6-a.	Describe the Knowledge Discovery in Databases (KDD) process. (CO3)	10
6-b.	What is R-Square in regression analysis, and what are its limitations? (CO3)	10

7. Answer any one of the following:-

7-a.	Discuss the differences between univariate and multivariate EDA and provide examples of how EDA can reveal insights and patterns in datasets.(CO4)	10
7-b.	Explain decomposition, trend and seasonality analysis, and autoregressive modeling.(CO4)	10

8. Answer any one of the following:-

8-a.	What are the key principles of effective data visualization, and how can they be	10

applied to enhance data communication? (CO5)

8-b. How do you create a table calculation in Tableau, and what is its role in enhancing visualizations? (CO5) 10