.

Printed Page:-4 **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 (2025 - 2026)
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. Example of raw data is: (CO1,K1) 1

(a) original swath files generated from a sonar system
(b) initial time-series file of temperature values
(c) a real-time GPS-encoded navigation file
(d) all of the mentioned

1-b. Data analysis model can be used for : (CO1,K1) 1

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

1-c. The type of data can be represented using charts, graphs, or plots (CO2,K2) 1

(a) Numeric data
(b) Categorical data
(c) Graphical data
(d) High Dimensional Data

1-d. Identify the type of data that consists of distinct categories or labels and cannot be measured on a numerical scale.(CO2,K2) 1

(a) Numeric data
(b) Categorical data
(c) Graphical data
(d) High Dimensional Data

.

1-e. The selection of a relevant subset of variables during data preprocessing is referred to as _____. (CO3,K1) 1

(a) Data Integration
(b) Data Reduction
(c) Data Discretization
(d) Data Transformation

1-f. R-Square is a measure commonly used in: (CO3, K1) 1

(a) Data cleaning
(b) Data transformation
(c) Data compression
(d) Regression analysis

1-g. Identify a potential consequence of not addressing missing data in a data analysis project.(CO4,K2) 1

(a) Improved model accuracy
(b) Unbiased results
(c) Biased or inaccurate analysis
(d) Faster computation

1-h. Primary focus of time series analysis is: (CO4,K1) 1

(a) Analyzing text data
(b) Understanding data patterns collected over time
(c) Examining cross-sectional data
(d) Assessing geographic data

1-i. A heat map is used to visualize _____. (CO5,K1) 1

(a) Geographic data
(b) Hierarchical data
(c) Frequency or density of data points
(d) Time series data

1-j. The primary purpose of Datawrapper in data visualization is to _____.(CO5,K2) 1

(a) Creating custom visualizations
(b) Preparing data for machine learning models
(c) Generating static charts and maps
(d) Building 3D visualizations

2. Attempt all parts:-

2.a. Mention the different types of learning algorithms used in Data Science.(CO1,K1) 2

2.b. Define the mean as a measure of central tendency and explain its significance.(CO2,K2) 2

2.c. List the main steps involved in the data integration process. (CO3,K1) 2

2.d. Define Linear Discriminant Analysis (LDA) and state its primary objective in data analysis.(CO4,K2) 2

2.e. Differentiate between a dimension and a measure in Tableau calculations.(CO5,K2) 2

SECTION-B 30

3. Attempt all parts:-

3.a. Answer any one of the following:-

3.a.(i)	Describe the concept of Datafication and explain the future trends and potential applications of Data Science in various industries. (CO1,K4)	6
3.a.(ii)	Describe the Data Science lifecycle and discuss the key stages involved in a data project.(CO1,K2)	6

3.b. Answer any one of the following:-

3.b.(i)	Define margin of error in statistical analysis and explain its calculation in hypothesis testing. (CO2 ,K2)	6
3.b.(ii)	Define mean, median, and mode as measures of central tendency and explain when each should be used.(CO2,K2)	6

3.c. Answer any one of the following:-

3.c.(i)	Discuss how noisy data affects the quality of analysis and explain strategies for detecting and handling it. (CO3,K4)	6
3.c.(ii)	Explain the role of data visualization techniques in improving understanding of data patterns, and illustrate the appropriate use of different visualization methods with examples.(CO3,K4)	6

3.d. Answer any one of the following:-

3.d.(i)	Define the interquartile range (IQR) and describe its role in identifying outliers in a dataset. (CO4,K2)	6
3.d.(ii)	Differentiate between Principal Component Analysis (PCA) and Factor Analysis (FA) in dimensionality reduction.(CO4,K4)	6

3.e. Answer any one of the following:-

3.e.(i)	Describe how custom calculations and aggregate functions (SUM, AVG) are applied in Tableau to enhance data visualizations.(CO5,K2)	6
3.e.(ii)	Describe the process of filtering and pivoting data in Tableau and explain their importance in preparing data for visualization. (CO5,K2)	6

SECTION-C 50

4. Answer any one of the following:-

4-a.	Describe the applications of Data Science in various industries and illustrate its use through real-world cases such as Facebook, Netflix, Amazon, Uber, and AirBnB (CO1,K4)	10
4-b.	Explain the concept of Exploratory Data Analysis (EDA) and discuss various visualization techniques used to analyze and interpret data effectively.(CO1,K2)	10

5. Answer any one of the following:-

5-a.	Explain the different types of data used in Data Science, including structured, semi-structured, and unstructured data, and discuss their applications with examples.(CO2,K2)	10
5-b.	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

6. Answer any one of the following:-

6-a.	Explain how binning, clustering, and histograms help organize data. Describe how each method works, why it is useful, and give examples.(CO3,K2)	10

.

6-b.	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

7. Answer any one of the following:-

7-a.	Explain the differences between univariate and multivariate Exploratory Data Analysis (EDA), and illustrate how EDA helps reveal patterns and insights in datasets with examples.(CO4,K4)	10
7-b.	Explain different techniques for handling missing data and removing redundant variables. Discuss how these steps improve the quality of data analysis.(CO4,K2)	10

8. Answer any one of the following:-

8-a.	Explain techniques for manipulating data in Tableau, including using the Data Interpreter, structuring, sorting, filtering, and pivoting data. Discuss how these methods improve data quality and visualization effectiveness. (CO5,K4)	10
8-b.	Discuss the use of advanced visualization tools in Tableau, including the Detail and Size panels, customizing tooltips, and applying color formatting.(CO5,K2)	10