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 performed by Data Scientist? (CO1, K1) 1

(a) Define the question

(b) Create reproducible code

(c) Challenge results

(d) All of the mentioned

1-b. How many principles of analytical graphs exist? (CO 1, K1) 1

(a) 3

(b) 4

(c) 6

(d) None of the mentioned

1-c. The primary purpose of a histogram in data analysis. (CO2, K2) 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. The statistical measure is used to quantify the strength and direction of a linear relationship between two variables. (CO2, K2) 1

(a) Skewness

(b) Median

(c) Correlation

(d) Standard Deviation

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. R-Square is a measure commonly used in: (CO3, K3) 1

(a) Data cleaning

(b) Data transformation

(c) Data compression

(d) Regression analysis

1-g. Which technique is commonly used for identifying redundancy among variables by measuring their linear relationships? (CO4, K4) 1

(a) Principal Component Analysis (PCA)

(b) Factor Analysis (FA)

(c) Correlation analysis

(d) Linear Discriminant Analysis (LDA)

1-h. Factor Analysis (FA) is used for: (CO4, K4) 1

(a) Data transformation

(b) Dimensionality reduction

(c) Identifying latent factors in data

(d) Correlation analysis

1-i. Which chart is most suitable for showing the distribution of a single continuous variable? (CO5, K3) 1

(a) Bar chart

(b) Line chart

(c) Scatter plot

(d) Histogram

1-j. Which data visualization tool is commonly used for creating interactive and dynamic visualizations? (CO5, K3) 1

(a) Microsoft Excel

(b) Tableau

(c) Photoshop

(d) Notepad

2. Attempt all parts:-

2.a. What are Data Science Tools and Technologies ? (CO1, K1) 2

2.b. Differentiate between the median and the mode as measures of central 2

	tendency. (CO2,K2)	
2.c.	Describe the R-Square statistic and its application in regression analysis.(CO3, K3)	2
2.d.	Write the primary goal of variable selection in a data analysis. (CO4, K4)	2
2.e.	Explain the "IF" function in Tableau calculations. (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.	Explain the concept of a normal distribution and its characteristics. How it is used in statistical analysis? (CO2,K2)	6
3-d.	Define skewness and discuss its implications in data analysis. (CO2,K2)	6
3.e.	Describe the different types of data attributes and their importance in data analysis. (CO3,K3)	6
3.f.	Explain the concept of the interquartile range (IQR) and how it helps identify outliers? (CO4,K4)	6
3.g.	Discuss the role of color in data visualization. How can color be used to enhance or mislead data interpretation? Provide examples. (CO5,K3)	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,K1)	10
4-b.	Explain Equality, diversity and inclusion in data science ? (CO1,K1)	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,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 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 are robust statistical methods, and how can they help in handling outliers	10

without removing data points? (CO4,K4)

7-b. Explain the challenges posed by missing data in a dataset and discuss various strategies for handling missing data. (CO4,K4) 10

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

8-a. How do you create a table calculation in Tableau, and what is its role in enhancing visualizations? (CO5,K3) 10

8-b. Describe the main features and advantages of using Tableau for data visualization. How does it differ from other tools? (CO5,K3) 10