

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA
(An Autonomous Institute Affiliated to AKTU, Lucknow)

MBA

SEM: II - THEORY EXAMINATION (2025 - 2026)

Subject: Introduction to Data Science

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:-

- | | | |
|-------|---|---|
| (1.1) | A dataset containing text, images, and videos mainly represents which feature of Big Data?(CO1-K1) | 1 |
| (a) | Volume | |
| (b) | Velocity | |
| (c) | Variety | |
| (d) | Value | |
| (1.2) | Secure access to organizational data is commonly maintained through which mechanism?(CO1-K2) | 1 |
| (a) | Backup tools | |
| (b) | Storage systems | |
| (c) | Visualization | |
| (d) | Authentication | |
| (1.3) | Tweets containing hashtags, emojis, and free text are commonly classified under which category of data?(CO2-K1) | 1 |
| (a) | Structured data | |
| (b) | Unstructured data | |
| (c) | Numeric data | |
| (d) | Spatial data | |
| (1.4) | Converting textual information into numerical values for analysis is mainly related to which process?(CO2-K1) | 1 |
| (a) | Data deletion | |
| (b) | Data storage | |

- (c) Data transformation
- (d) Data copying
- (1.5) Conversion of selected datasets into suitable mining format mainly supports which stage in KDD?(CO3-K1) 1
- (a) Data storage
- (b) Data selection
- (c) Data cleaning
- (d) Data transformation
- (1.6) Conversion of categorical variables into dummy variables mainly reflects which transformation process?(CO3-K1) 1
- (a) Data cleaning
- (b) Encoding
- (c) Data reduction
- (d) Data storage
- (1.7) Identification of hidden latent variables influencing observed data mainly represents which analytical method?(CO4-K1) 1
- (a) PCA
- (b) FA
- (c) LDA
- (d) Clustering
- (1.8) Scatter plot matrices used for relationship analysis mainly belong to which analytical category?(CO4-K1) 1
- (a) Histogram
- (b) Scatter matrix
- (c) Box plot
- (d) Bar chart
- (1.9) Identification of outliers and quartile spread in datasets is commonly supported using which visualization tool?(CO5-K1) 1
- (a) Scatter plot
- (b) Box plot
- (c) Histogram
- (d) Heatmap
- (1.10) Visualization of intensity variations across matrix-based values mainly uses which chart type?(CO5-K1) 1
- (a) Line plot
- (b) Pie chart
- (c) Scatter plot
- (d) Heat map

2. Attempt all parts:-

- (2.1) Describe the evolution of data science from traditional data processing systems to modern analytical approaches (CO1-K2) 2

- | | | |
|-------|--|---|
| (2.2) | Demonstrate the process of filtering irrelevant attributes from datasets and explain its importance (CO2-K2) | 2 |
| (2.3) | Describe the role of preprocessing in preparing datasets for predictive modeling (CO3-K2) | 2 |
| (2.4) | Analyze the use of scatter plots in identifying relationships and correlations between variables (CO4-K2) | 2 |
| (2.5) | Describe the importance of color selection in visualization and its impact on interpretation (CO5-K2) | 2 |

SECTION-B 30

Answer any one of the following:-

- | | | |
|-------|--|---|
| (3.1) | Discuss ethical considerations in data science, including bias, fairness, transparency, and accountability (CO1-K3). | 6 |
| (3.2) | Provide a detailed explanation of the role of data science in marketing analytics and customer behavior analysis (CO1-K3). | 6 |

Answer any one of the following:-

- | | | |
|-------|--|---|
| (3.3) | Apply classification techniques in spam detection systems and evaluate their effectiveness in filtering unwanted communication (CO2-K4). | 6 |
| (3.4) | Analyze the integration of multiple heterogeneous datasets and discuss strategies to ensure consistency and usability (CO2-K4). | 6 |

Answer any one of the following:-

- | | | |
|-------|--|---|
| (3.5) | Explain the concept of data preprocessing in detail by discussing its objectives, key steps, and its critical role in improving data quality and analytical outcomes in real-world scenarios (CO3-K3). | 6 |
| (3.6) | Explain the importance of data transformation in the KDD process and illustrate how it prepares data for effective mining (CO3-K3). | 6 |

Answer any one of the following:-

- | | | |
|-------|--|---|
| (3.7) | Examine the importance of selecting the optimal number of principal components, and analyze methods such as scree plots and explained variance (CO4-K5). | 6 |
| (3.8) | Critically analyze Factor Analysis (FA) by explaining its objective of identifying latent variables and its underlying assumptions (CO4-K5). | 6 |

Answer any one of the following:-

- | | | |
|--------|---|---|
| (3.9) | Critically examine the use of bar plots in categorical data analysis by explaining their structure, interpretation, advantages, and limitations in comparison to other visual tools (CO5-K3). | 6 |
| (3.10) | Analyze the effectiveness of different graphical tools in Microsoft Excel, such as column charts, line charts, Pivot Charts, and slicers, in exploring and interpreting large datasets. Discuss how these tools help in identifying trends, comparing categories, and supporting decision-making. (CO5-K3). | 6 |

SECTION-C 50

4. Answer any one of the following:-

- | | | |
|-------|---|----|
| (4.1) | A marketing firm uses data science to target customers more effectively. (a) Analyze how data science enhances marketing strategies. (b) Evaluate its impact on customer engagement (CO1-K2). | 10 |
|-------|---|----|

- (4.2) A firm uses big data for customer insights but struggles with variety. (a) Analyze the role of variety in big data. (b) Evaluate its challenges in processing (CO1-K2). 10
5. Answer any one of the following:-
- (5.1) A health app collects step counts, heart rates, and activity labels. (a) Analyze the role of numeric and categorical data in health analytics. (b) Evaluate how data transformation improves model performance (CO2-K4). 10
- (5.2) A company analyzes sales trends across regions. (a) Analyze spatial data importance. (b) Evaluate manipulation techniques (CO2-K4). 10
6. Answer any one of the following:-
- (6.1) A hospital launches a data-driven diagnosis system, but patient records come with missing test results, handwritten notes converted to text, and inconsistent units (mg/dL vs mmol/L). (a) Analyze the preprocessing steps required to make this dataset reliable for analysis, clearly explaining handling of missing values and inconsistencies. (b) Evaluate the risks of skipping these preprocessing steps in a clinical decision-making system (CO3-K3). 10
- (6.2) A university analyzes student performance data but finds missing attendance records and inconsistent grading scales. (a) Analyze preprocessing steps to handle missing and inconsistent data. (b) Evaluate how these issues affect predictive modeling of student outcomes (CO3-K3). 10
7. Answer any one of the following:-
- (7.1) A healthcare dataset contains two patient groups but features are highly overlapping. (a) Analyze how LDA can create discriminant functions to separate groups. (b) Evaluate scenarios where LDA may fail (CO4-K5). 10
- (7.2) A firm performs univariate EDA but ignores multivariate interactions. (a) Analyze consequences. (b) Evaluate need for correlation analysis (CO4-K5). 10
8. Answer any one of the following:-
- (8.1) A marketing team uses scatter plots to analyze ad spend vs sales but struggles to interpret overlapping data points. (a) Analyze how scatter plots help identify relationships and limitations such as overplotting. (b) Evaluate techniques like transparency or clustering to improve interpretation (CO5-K3). 10
- (8.2) A marketing firm presents campaign results using inconsistent chart types. (a) Analyze importance of choosing appropriate visualizations. (b) Evaluate consequences of mismatched chart selection (CO5-K3). 10