.

Printed Page:-4 **Subject Code:- CMBIE0215**

Roll. No:

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA
(An Autonomous Institute Affiliated to AKTU, Lucknow)
MBA IEV
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 data science workflow often begins with _____ and ends with insight generation (CO1-K1) 1

(a) Data collection
(b) Data deletion
(c) Printing
(d) Storage

(1.2) A role involving interpretation of data insights is performed by _____ (CO1-K1) 1

(a) Data analyst
(b) Network admin
(c) Hardware engineer
(d) Typist

(1.3) A system mapping earthquake locations based on coordinates uses _____ (CO2-K1) 1

(a) Numeric data
(b) Structured data
(c) Social data
(d) Spatial data

(1.4) A grouping of survey responses into categories such as satisfied or dissatisfied is called (CO2-K1) 1

(a) Numeric data
(b) Categorical data
(c) Spatial data

(d) Structured data

(1.5) A preprocessing phase selecting useful attributes is _____ (CO3-K1) 1

(a) Data cleaning
(b) Data storage
(c) Feature selection
(d) Data reporting

(1.6) A KDD stage involving _____ of patterns ensures meaningful results (CO3-K1) 1

(a) Evaluation
(b) Data storage
(c) Data cleaning
(d) Data reporting

(1.7) A dataset requiring simultaneous analysis of multiple variables applies _____ (CO4-K1) 1

(a) Univariate EDA
(b) Multivariate EDA
(c) PCA
(d) LDA

(1.8) A multivariate analysis examines _____ among variables (CO4-K1) 1

(a) Single variable
(b) Interdependence
(c) PCA
(d) LDA

(1.9) A _____ helps compare distribution of dataset with theoretical distribution (CO5-K1) 1

(a) Box plot
(b) Histogram
(c) Q-Q plot
(d) Heatmap

(1.10) A visualization package in Python includes _____ (CO5-K1) 1

(a) TensorFlow
(b) Pandas
(c) Matplotlib
(d) Flask

2. Attempt all parts:-

(2.1) Describe the challenges related to data privacy and security in big data environments (CO1-K2) 2

(2.2) Analyze spatial data applications in navigation systems and mapping tools (CO2-K4) 2

(2.3) Describe feature selection and explain its importance in improving model performance (CO3-K3) 2

.

(2.4)	Analyze the role of covariance matrix in PCA computation (CO4-K5)	2
(2.5)	Describe the concept of heat maps and explain how they represent correlation among variables (CO5-K3)	2

SECTION-B 30

Answer any one of the following:-

(3.1)	Examine the importance of scalability in big data systems and explain how distributed computing supports large-scale data processing (CO1-K2).	6
(3.2)	Examine the role of data science in financial services, particularly in fraud detection, risk assessment, and algorithmic trading (CO1-K2).	6

Answer any one of the following:-

(3.3)	Analyze the process of merging multiple datasets from different sources, discussing challenges such as inconsistency, redundancy, and missing values, along with solutions (CO2-K4).	6
(3.4)	Apply encoding techniques such as label encoding and one-hot encoding in categorical data transformation, and evaluate their impact on model interpretation (CO2-K4).	6

Answer any one of the following:-

(3.5)	Explain various imputation techniques such as mean, median, and model-based imputation, and evaluate their effectiveness in different scenarios (CO3-K3).	6
(3.6)	Analyze the impact of poor data quality on business decisions and explain how preprocessing mitigates such risks (CO3-K3).	6

Answer any one of the following:-

(3.7)	Evaluate the performance of LDA compared to other classification techniques in terms of accuracy, interpretability, and efficiency (CO4-K5).	6
(3.8)	Evaluate the combined application of PCA, FA, and LDA in advanced analytical workflows and explain how these techniques complement each other in dimensionality reduction, feature extraction, and classification. (CO4-K5).	6

Answer any one of the following:-

(3.9)	Evaluate the importance of line plots in forecasting and predictive analytics, especially in business and financial contexts (CO5-K3).	6
(3.10)	Evaluate the importance of color schemes, labeling, and layout in designing effective and interpretable visualizations (CO5-K3).	6

SECTION-C 50

4. Answer any one of the following:-

(4.1)	A logistics company wants to optimize delivery routes using real-time and historical data. (a) Examine how data science contributes to route optimization. (b) Evaluate the role of real-time analytics in improving operational efficiency (CO1-K2).	10
(4.2)	A streaming platform wants to improve user engagement through personalized content recommendations. (a) Analyze how data science supports personalization strategies. (b) Evaluate the impact of such systems on user retention (CO1-K2).	10

5. Answer any one of the following:-

(5.1)	A bank introduces a digital loan approval system but faces inconsistent decisions due to mixed data formats such as numeric credit scores, categorical employment	10

types, and text-based financial histories. (a) Analyze the different data types involved and explain their significance in decision-making. (b) Evaluate appropriate transformation and encoding techniques to prepare this data for modeling (CO2-K4).

(5.2) A logistics company uses vehicle location data, delivery times, and driver comments to optimize routes. (a) Analyze the role of spatial and textual data in operational decisions. (b) Evaluate how these can be transformed and integrated for route optimization models (CO2-K4). 10

6. Answer any one of the following:-

(6.1) A weather forecasting system processes noisy sensor data. (a) Analyze smoothing techniques. (b) Evaluate their importance (CO3-K3). 10

(6.2) A firm integrates multiple structured datasets. (a) Analyse the challenges the company is going to face and define the pipeline to integrate all the datasets. (b) Evaluate solutions to the challenges faced by the company (CO3-K3). 10

7. Answer any one of the following:-

(7.1) A psychology research firm is studying human personality traits using survey data but finds overlapping variables across traits. (a) Analyze how Factor Analysis helps identify underlying constructs. (b) Evaluate the limitations of FA when variables are weakly correlated (CO4-K5). 10

(7.2) A firm reduces dimensionality but loses insight. (a) Analyze trade-offs. (b) Evaluate strategies (CO4-K5). 10

8. Answer any one of the following:-

(8.1) A data analyst uses box plots to compare salaries across departments but management misunderstands outliers as errors. (a) Analyze how box plots represent spread and outliers. (b) Evaluate how misinterpretation of visualization can affect business decisions (CO5-K3). 10

(8.2) A dashboard misleads due to scale manipulation. (a) Analyze issue. (b) Evaluate ethical concerns (CO5-K3). 10