NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA

(An Autonomous Institute Affiliated to AKTU, Lucknow)

MCA

SEM: III - THEORY EXAMINATION (20..... - 20.....)

Subject: Advanced Concepts of 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.   Balancing objectives with data integrity includes:(CO1, K1)   1

      (a)   Increasing data volume

      (b)   Ensuring strategic alignment and risk assessment

      (c)   Reducing costs

      (d)   Speeding up data collection

1-b.   Data validation in spreadsheets prevents:(CO1, K1)   1

      (a)   Incorrect data entry

      (b)   Data entry speed

      (c)   Manual proofreading

      (d)   Ignoring errors

1-c.   Noisy data often presents:(CO2, K1 K2)   1

      (a)   Perfectly accurate information

      (b)   Errors and fluctuations

      (c)   Consistent patterns

      (d)   Redundant information

1-d.   Detection of noisy data can be achieved through:(CO2, K1 K2)   1

      (a)   Data visualization

      (b)   Random sampling

      (c)   Ignoring errors

(d) Data normalization

1-e. A key benefit of using visualizations is:(CO3, K1 K2)   1

    (a) Complexity

    (b) Clarity

    (c) Length

    (d) Ambiguity

1-f. The purpose of a legend in a chart is to:(CO3, K1 K2)   1

    (a) Add colors

    (b) Provide context

    (c) Explain symbols

    (d) None of the above

1-g. Which file format can be imported directly into Power BI? (CO4, K1 K2 K3)   1

    (a) .txt

    (b) .docx

    (c) .jpg

    (d) .mp3

1-h. The Query Editor in Power BI is used for:(CO4, K1 K2 K3)   1

    (a) Data visualization

    (b) Data modeling

    (c) Data transformation and cleaning

    (d) Report sharing

1-i. Big Data has evolved primarily due to advancements in:(CO5, K1 K2)   1

    (a) Cloud computing

    (b) Data storage

    (c) Technology and connectivity

    (d) All of the above

1-j. One common myth about Big Data is that:(CO5, K1 K2)   1

    (a) It can only be used by large companies

    (b) It always requires advanced technology

    (c) It is just about large volumes of data

    (d) It provides accurate predictions

2. Attempt all parts:-

2.a. Define data integrity and explain its significance in decision-making processes.   2

2.b. In a dataset with missing values, explain how this issue can impact data analysis.   2

2.c. Define data visualization.   2

2.d. Define "measures" in the context of Power BI.   2

2.e. List one of the 6Vs that characterize Big Data.   2

.

SECTION-B 30

3. Answer any <u>five</u> of the following:-

3-a.	Explain the concept of data integrity and its critical role in organizational success. Include examples of how data integrity impacts decision-making.(CO1, K1)	6
3-b.	Discuss the benefits of clean data in enhancing operational efficiency. Provide examples of how clean data supports better decision-making.(CO1, K1)	6
3-c.	Outline the importance of exploring raw data before analysis. Discuss techniques that can be used for effective data exploration.(CO2, K1 K2)	6
3-d.	Explain the importance of correlation coefficients in understanding relationships between variables. Discuss how they guide decision-making.(CO2, K1 K2)	6
3.e.	Evaluate the role of dynamic visualizations in data analysis. How do they differ from static visualizations, and what advantages do they offer?(CO3, K1 K2)	6
3.f.	Explain the steps involved in importing data from flat files into Power BI.(CO4, K1 K2 K3)	6
3.g.	Describe the different sources from which Big Data can be generated.(CO5, K1 K2)	6

SECTION-C 50

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

4-a.	Evaluate various strategies for managing insufficient data during research studies. Discuss the trade-offs associated with each strategy, highlighting their effectiveness.(CO1, K1)	10
4-b.	Discuss the significance of cross-verifying data against trusted sources. Analyze how this practice strengthens data reliability and enhances organizational credibility.(CO1, K1)	10

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

5-a.	Describe the data cleaning process in detail, emphasizing its significance in preparing datasets for analysis and decision-making. Include key steps and methodologies involved.	10
5-b.	Detail the steps involved in detecting and resolving data value conflicts in datasets. Discuss the significance of addressing these conflicts to maintain data quality and integrity.	10

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

6-a.	Identify and explain the key elements of art that should be considered when designing data visualizations. How do these elements enhance the visual appeal and clarity?	10
6-b.	Identify different types of data visualizations and discuss their best use cases. How does the choice of visualization type impact data analysis and communication?	10

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

7-a.	Describe the key features of Power BI and their significance in data analysis.	10
7-b.	Analyze the different views available in Power BI Desktop and their respective	10

functionalities.

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

8-a.   Explain the concept of Big Data and its significance in the modern economy.   10

8-b.   Define data discovery and describe its importance in the context of Big Data   10
         analytics.

.