

**Dr. Babasaheb Ambedkar Open University**  
**Term End Examination July – 2026**

|              |                                |            |                      |
|--------------|--------------------------------|------------|----------------------|
| Course       | : BCA                          | Date       | : 14-07-2026         |
| Subject Code | : BCAMI-404                    | Time       | : 03:00pm to 05:15pm |
| Subject Name | : Fundamentals of Data Science | Duration   | : 02.15 Horse        |
|              |                                | Max. Marks | : 70                 |

---

**Section A**

**Answer the following (Attempt any three) (30)**

1. Explain various types of data and differentiate between structured and un-structured data.
2. Write detailed note on Applications of Data Science.
3. Explain in details about Data Transformation.
4. Explain measure of central tendency with suitable examples.
5. Write detailed note on programming languages used for data science.

**Section B**

**Answer the following (Attempt any four) (20)**

1. Write short note on Importance of Data Science.
2. Explain Data Modelling in brief.
3. Differentiate: Primary data and Secondary Data.
4. Define feature engineering. Explain why it is one of the most important steps in machine learning, with at least three reasons.
5. Explain in brief about Types of Data Integration
6. Explain Importance of Data Visualization in Data Science

**Section C**

**Part – A (Multiple Choice Questions) (10)**

- 1 Which of the following is NOT a key component of Data Science?  
A Data Collection                      B Data Visualization  
C Networking                            D Model Evaluation
- 2 Which component selects only the most relevant variables for model building?  
A Data Collection                      B Feature Selection  
C Deployment                            D Monitoring
- 3 Which type of data has a predefined structure and is commonly stored in relational databases?  
A Structured                              B Semi-structured  
C Unstructured                            D Multimedia
- 4 Which Big Data characteristic refers to the speed at which data is generated?  
A Variety                                  B Volume  
C Value                                      D Velocity

- Which measure represents the middle value of an ordered dataset?  
A Mean  
B Median  
C Mode  
D Variance
- Which branch of mathematics measures the likelihood of an event?  
A Algebra  
B Calculus  
C Probability  
D Geometry
- Which visualization is most suitable for showing trends over time?  
A Pie Chart  
B Histogram  
C Scatter Plot  
D Line Chart
- Which programming language is most widely used in Data Science?  
A COBOL  
B Python  
C Pascal  
D BASIC
- Which Python library is commonly used for numerical computations?  
A NumPy  
B Django  
C OpenCV  
D Tkinter
- What is the primary advantage of Jupyter Notebook?  
A Compiles C programs  
B Executes only SQL queries  
C Combines code, output, and documentation in one document  
D Stores only images

### Part – B (Do as Directed)

**(10)**

**State whether the following statements are true or false**

- 1 Data Science combines programming, statistics, and domain knowledge to extract meaningful insights from data.
- 2 The Data Mining era was the earliest stage in the evolution of Data Science.
- 3 Deployment is the first component performed in a Data Science project.
- 4 Ordinal data have categories without any logical order.
- 5 One-hot encoding is commonly used for converting categorical variables into numerical form.
- 6 Data normalization increases the number of records in a dataset.
- 7 Probability values are always greater than 1.
- 8 A scatter plot is useful for identifying relationships between two numerical variables.
- 9 Pandas is primarily used for data manipulation and analysis in Python.
- 10 Jupyter Notebook allows users to combine executable code, text, and graphical output in a single document.

\*\*\*\*\*