

No. of Printed Pages : 5

MCS-062

**MASTER OF SCIENCE (DATA
SCIENCE AND ANALYTICS)
(MSCDSA)**

Term-End Examination

June, 2026

MCS-062 : INTRODUCTION TO DATA SCIENCE

Time : 3 Hours

Maximum Marks : 100

Note : (i) *Question no. 1 is compulsory.*

(ii) *Attempt any **three** from the rest.*

1. (a) Explain the functional blocks of data science with the help of a suitable block diagram. 5
- (b) List the properties of data and describe any *two* of them. 5

- (c) Describe the term 'Data Acquisition' in context of data science. Also, list the various hardware components of any Data Acquisition System (DAQ). Briefly discuss the utility of any *one* component. 5
- (d) What is the role of sampling in the process of Data Acquisition (DAQ) ? How does 'Nyquist Criterion' relates to sampling ? Explain. 5
- (e) What do you understand by Data preprocessing ? List the phases of Data preprocessing, and describe any *one* phase with suitable example. 5
- (f) Calculate Mean, Median, Mode and Standard deviation for the data given below also comment on the skewness of the data : 5

1, 5, 7, 2, 3, 1, 8, 2, 6, 4, 5, 7, 9, 4, 6

- (g) A sample of 100 students is drawn from the population of students who appeared for some exam. The observed mean is 3.2 with standard deviation of 0.4. Based on the given data we wish to test the claim that average score of student is 3.0. State Null hypothesis and Alternate hypothesis, and use Z-test to test the hypothesis (given $Z = \pm 1.96$ @ 5% significance level). 5
- (h) What is a Histogram ? How does it differ from Bargraph ? What does histogram show ? Elaborate. 5
2. (a) What is Data Handling ? List and explain the tasks involved in Data Handling. 6
- (b) What is a time series data ? Why is it important in forecasting ? Briefly discuss the trend and seasonality in time series data, with suitable example for each. 7

- (c) Describe distributed file system. Also, explain HDFS and Map Reduce. 7
3. (a) What is a Data stream ? Briefly discuss the issues and challenges of Data streams. 7
- (b) Briefly describe the following : 6
- (i) SQL Databases
 - (ii) Big data formats
 - (iii) XML
- (c) Explain the following with suitable example for each : 7
- (i) Predictive data analysis
 - (ii) Prescriptive data analysis
4. (a) Explain the term 'Crowd Sourcing'. Briefly discuss the relevance of 'Crowd Sourcing' in data science. 7
- (b) Briefly describe the following levels of measurement of data : 6
- (i) Nominal
 - (ii) Ordinal
 - (iii) Interval
 - (iv) Ratio

- (c) What is Programmable Logic Controller (PLC) ? Explain the key characteristics of PLC. Also, discuss the role of PLC in data science. 7

5. Write short notes on the following : $4 \times 5 = 20$

- (i) Web Scraping and its utility in data science
- (ii) Noisy data and approaches to remove noise from data
- (iii) Data integration and issues of data integration
- (iv) Common misconceptions of Data Analysis

x x x x x