MASTER OF SCIENCE
DATA SCIENCE AND ANALYTICS
(MSCDSA)

MSCDSA/ASSIGN/SEMESTER-I

ASSIGNMENTS

(January – 2026 & July – 2026)

MCS-061, MCS-062, MCS-063, MCS-207, MCSL-064, MCSL-065

CONTENTS

Important Notes

1. Submit your assignments to the Coordinator of your Study Centre on or before the due date.

2. Assignment submission before due dates is compulsory to become eligible for appearing in corresponding Term End Examinations. For further details, please refer to MSCDSA Programme Guide.

3. To become eligible for appearing the Term End Practical Examination for the lab courses, it is essential to fulfill the minimum attendance requirements as well as submission of assignments (on or before the due date). For further details, please refer to the MSCDSA Programme Guide.

4. The viva voce is compulsory for the assignments. For any course, if a student submitted the assignment and not attended the viva-voce, then the assignment is treated as not successfully completed and would be marked as ZERO.

Course Code	:	MCS-062
Course Title	:	**Introduction to Data Science**
Assignment Number	:	**MSCDSA(I)/062/Assign/2026**
Maximum Marks	:	**100**
Weightage	:	**30%**
Last Dates for Submission	:	**30th April, 2026 (for January session)**
		31st October, 2026 (for July session)

There are 25 questions in this assignment which carries 80 marks. Question No. 1 to 20 are of 3 marks each, remaining question no. 21 to 25 are of 4 marks each. Rest 20 marks are for viva-voce.

Answer all the questions. Questions are application oriented. Some questions require numerical calculations. You may use illustrations and diagrams to enhance the explanations. Include the screen layouts also along with your assignment responses. Please go through the guidelines regarding assignments given in the Programme Guide for the format of presentation.

Q1: Identify a real-life problem from healthcare or banking and explain how the Data Science lifecycle can be applied to solve it.

Q2: A company wants to predict future sales instead of analyzing past trends. Should they use Data Analytics or Data Science? Justify.

Q3: Classify the following data: (a) Customer feedback text, (b) Monthly temperature readings, (c) Employee ID numbers. Mention data type and scale.

Q4: A dataset contains values: 20, 25, 30, 35, 40. Identify whether the data is discrete or continuous and calculate the range.

Q5: Suggest suitable data acquisition tools for collecting air quality data in a smart city project and justify your choice.

Q6: Explain how sampling and signal conditioning are used in a temperature monitoring system.

Q7: List the steps you would follow to clean and integrate sales data collected from two different e-commerce platforms.

Q8: Normalize the values 200, 400, 600, 800 using Min–Max normalization.

Q9: Perform Exploratory Data Analysis (EDA) for a student marks dataset. What insights can be obtained?

Q10: Calculate mean and standard deviation for the dataset: 10, 12, 15, 18, 20.

Q11: A bank wants to predict loan default. Is this inferential or predictive analysis? Explain the reasoning.

Q12: A sample mean is 52, population mean is 50, and standard deviation is 4 for n = 16. Compute the Z-value.

Q13: Which visualization technique will you use to compare monthly sales across two years and why?

Q14: Draw a histogram for the following data: 5, 7, 8, 10, 12, 15, 18, 20 and interpret it.

Q15: Using Excel, explain how frequency distribution and central tendency help in summarizing student performance.

Q16: A dataset shows X = [1,2,3,4,5] and Y = [2,4,6,8,10]. Calculate the correlation coefficient and interpret.

Q17: Differentiate between RDBMS and NoSQL databases with an example from an e-commerce system.

Q18: Explain the 5V characteristics of Big Data with one real-world example.

Q19: Explain data stream mining and its importance in real-time traffic monitoring systems.

Q20: Compare Python, R, Tableau, and Power BI for data analysis and visualization use cases.

Q21: Given the data set: 2, 4, 4, 6, 8, 10

    a. Calculate the Mean, Median, and Mode.
    b. Find the Variance and Standard Deviation of the data.
    c. Determine the Interquartile Range (IQR).
    d. Briefly explain what each measure tells you about the data.

Q22: suppose we have a sample of 25 students with a sample mean score of 75 and a sample standard deviation of 10. We wish to test whether the mean score differs from 70. State the Null Hypothesis and the alternate hypothesis. Use t-test to test whether the null hypothesis will be accepted or not. Given, 24 degrees of freedom, the critical t-value at a 5% level is approximately 2.064.

Q23: Explain Logistic regression, and use it to Predict pass (1) or fail (0) based on study hours. The relevant data related to the study hours and result(pass/fail) is given below:
                Study Hours (X): 2, 4, 6, 8, 10;
                   Pass/Fail (Y): 0, 0, 1, 1, 1

Q24: Describe the utility of time series forecasting in data science. Also, compare the Moving Average and Exponential Smoothing techniques for time series forecasting. Also, perform the following tasks:

    a. Given the Sales for 5 months as: [100, 120, 130, 110, 140] compute the 3-month moving average for month 6.
    b. Write and explain the Formula for Exponential smoothing

Q25: A company wants to compare the variability in salaries between two departments: Department A and Department B. The company randomly samples 10 employees from each department and records their salaries (in $1000s):
        · Department A Salaries (Sample 1): 45, 48, 50, 52, 49, 46, 51, 47, 50, 48
        · Department B Salaries (Sample 2): 55, 58, 54, 60, 57, 53, 59, 56, 55, 58

The company wants to know if there is a significant difference in the variances of salaries between the two departments. Use F-test to compare the variances and test the null hypothesis.

Given the Critical values ($\alpha = 0.05$, two-tailed)

$$F_{upper} = F_{0.975}(9,9) = 4.026 \&$$
$$F_{lower} = F_{0.025}(9,9) = 0.248.$$

Decision rule: Reject H₀ if $F < 0.248$ or $F > 4.026$.