

## CMR ENGINEERING COLLEGE: : HYDERABAD

## UGC AUTONOMOUS

## II-B.TECH-II-Semester End Examinations (Supply) –June- 2025

## FUNDAMENTAL OF DATA SCIENCE

## (AI&amp;DS)

[Time: 3 Hours]

[Max. Marks: 70]

**Note:** This question paper contains two parts A and B.

Part A is compulsory which carries 20 marks. Answer all questions in Part A.

Part B consists of 5 Units. Answer any one full question from each unit. Each question carries 10 marks and may have a, b, c as sub questions.

**PART-A****(20 Marks)**

1. a) What is faceting in data science? [2M]
- b) How to retrieve data in data science? [2M]
- c) What is Frequency Distribution? [2M]
- d) Define mean, median and mode. [2M]
- e) Explain the use of pandas in Python. [2M]
- f) Differentiate between data and dataframe. [2M]
- g) What is normal distribution? [2M]
- h) What is difference between simple linear and multiple linear regressions? [2M]
- i) What is Seaborn? [2M]
- j) Which tool is best for data visualization? [2M]

**PART-B****(50 Marks)**

2. Explain the benefits of Data science. Differentiate between Data Science and Data Engineering. [10M]

**OR**

3. Explain in detail Data Cleaning Process. [10M]

4. Explain in detail different types of Graphs used in frequency distribution. [10M]

**OR**

5. Calculate Range, Variance, Standard Deviation and InterQuartile range for below dataset. [10M]

72    110    134    190    238    287    305    324

6. Explain various operations on numpy arrays. [10M]

**OR**

7. Consider a dataset containing information about students' exam scores in three subjects: Math, Science, and History. Each student is identified by a unique "StudentID," and the data is given in the form of a dictionary: [10M]

python

data = {

'StudentID': [101, 102, 103, 104, 105, 106, 107, 108],

'Math': [85, 78, 92, 65, 90, 78, 88, 75],

'Science': [90, 85, 88, 75, 82, 80, 90, 92],

'History': [75, 92, 80, 78, 85, 88, 70, 78]

}

Create a DataFrame using the given data and perform the following operations:

i) Calculate the average score for each subject (Math, Science, History).

ii) Calculate the overall average score across all three subjects.

iii) Group the data by the overall average score (you can use binning based on ranges, e.g., 70-79, 80-89, etc.), and for each group, calculate the percentage of students who scored above 80 in each subject.

8. Define scatter plot and types of scatter plot correlation. [10M]  
Draw a scatter plot for the given data that shows the number of games played and scores obtained in each instance.

No. of games	3	5	2	6	7	1	2	7	1
Scores	80	90	75	80	90	50	65	85	40

**OR**

- 9.a) Explain the concept of a Z-Score and its importance in standardizing data. [5M]  
b) Explain the relationship between Z-Scores and the standard normal distribution. [5M]

10. Demonstrate the usage of histograms for data exploration and explain its attributes. [10M]

**OR**

11. Explain about [10M]  
i) Line plots.  
ii) Scatter plots.  
iii) Text and annotation.

\*\*\*\*\*