Code No.: AD404PC   R20   H.T.No. [ ][ ][8][R][ ][ ][ ][ ][ ]

CMR ENGINEERING COLLEGE: : HYDERABAD
UGC AUTONOMOUS
II–B.TECH–II–Semester End Examinations (Supply) - December- 2025
FUNDAMENTALS OF DATA SCIENCE
(AI&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.

<u>**PART-A**</u>   **(20 Marks)**

1. a) Write about facets of data.   [2M]
   b) What are the goals of data science?   [2M]
   c) List different types of frequency distribution.   [2M]
   d) Define variance and standard deviation.   [2M]
   e) What are pandas and NumPy?   [2M]
   f) Which method is used to merge two data frames and write the syntax?   [2M]
   g) Define z-score.   [2M]
   h) Explain the concept of a Z-Score and its importance in standardizing data. [2M]
   i) What are main libraries used for data visualization in python?   [2M]
   j) What principles guide your approach to data visualization?   [2M]

<u>**PART-B**</u>   **(50 Marks)**

2. Explain in detail about build the model in data science.   [10M]
                                  OR
3. Give an overview of the data science process.   [10M]

4. What are levels of measurement in data and explain in detail.   [10M]
                                  OR
5. Define and differentiate between Interquartile Range (IQR) and Standard Deviation (SD). [10M]
    For below data set find IQR and SD Dataset: 1, 4, 8, 11, 13, 17, 19, 19, 20, 23, 24, 24, 25,
    28, 29, 31, 32.

6. Explain different ways of creating 1D, 2D and 3D arrays in numpy.   [10M]
                                  OR
7. Consider a dataset containing information about students' exam scores in three subjects: [10M]
    Math, Science, and History. Each student is identified by a unique "StudentID." The data
    is given in the form of a dictionary:
    python
    data = {
        'StudentID': [101, 102, 103, 104, 105],
        'Math': [85, 78, 92, 65, 90],
        'Science': [90, 85, 88, 75, 82],
        'History': [75, 92, 80, 78, 85]
    }
    Create a DataFrame using the given data and perform the following operations:
    i) Calculate the average score of each student across all three subjects and add a new
    column "Average" to the DataFrame.
    ii) Sort the DataFrame based on the "Average" column in descending order and display
    the top two rows with the highest average scores.
    iii) Calculate the percentage of students who scored above 80 in each subject and display
    the results in a new DataFrame with columns "Subject" and "Percentage.

8. What is Correlation Coefficient and calculate correlation coefficient for below data. [10M]

x	50	51	52	53	54
y	3.1	3.2	3.3	3.4	3.5

OR

9.a) What is a normal distribution, and what are its defining characteristics? [5M]
   b) How does a normal distribution differ from other types of distributions commonly observed in data? [5M]

10. Explain about concept of scatter plots and its applications. [10M]

OR

11. Explain about three-dimensional plotting with example. [10M]

****