Roll No. [ ][ ][ ][ ][ ][ ][ ][ ][ ][ ][ ][ ]

PRESIDENCY UNIVERSITY
BENGALURU

Department of Research & Development
Mid - Term Examinations - SEPTEMBER 2024

Odd Semester: Ph.D. Course Work		**Date**: 28 /09/2024	
Course Code: CSE5008		**Time**: 2:00pm – 3:30pm	
Course Name: **Programming in Data Science**		**Max Marks**: 50	
Department: Computer Science and Engineering		**Weightage**: 25%	

Instructions:

(i) Read all questions carefully and answer accordingly.

(ii) Do not write anything on the question paper other than roll number.

Part A

Answer ALL the Questions. Each question carries 5 marks.		4Qx5M=20M
1	Compare and contrast descriptive statistics and inferential statistics. Provide an example of when each type would be appropriately used in data analysis.	**5 Marks**
2	Define dimensionality reduction and describe two popular techniques used for this purpose. How can reducing dimensionality improve the performance of machine learning models?	**5 Marks**
3	Describe the difference between reshaping and resizing arrays in NumPy. Provide suitable code snippets.	**5 Marks**
4	Explain how NumPy handles missing values in an array and provide an example demonstrating the use of np.nan.	**5 Marks**

Part B

Answer ALL Questions. Each question carries 15 marks.		2QX15M=30M
5	The Data Science methodology involves several key stages, from problem identification to feedback. Explain the steps involved in the Data Science process, starting from problem formulation to data collection, preparation, modeling, and evaluation. Further, describe the significance of deploying a model and gathering feedback for continuous improvement, using an example to illustrate the practical application of this methodology.	**15 Marks**
6	Handling data efficiently is essential in data analysis. Discuss how Pandas facilitates descriptive statistics, sorting, and dealing with missing data. Provide examples	**15 Marks**

| | demonstrating the use of Pandas statistical functions, window functions, and how missing data is handled in DataFrames. Additionally, explain how groupby, aggregation and joining operations are used in data manipulation. | |