Name: **Enrolment No:**	

UPES

End Semester Examination, May 2024

Course: Exploratory Data Analysis **Semester: II**

Program: MCA **Time: 03 hrs.**

Course Code: CSDA7008P **Max. Marks: 100**

Instructions: Section A (Attempt All Questions), Section B (Attempt Four Questions), Section C (Attempt Two Questions)

SECTION A
(5Qx4M=20Marks)

S. No.		Marks	CO
Q 1	Describe the process of handling missing values in a dataset.	**4**	**CO1**
Q 2	Create a histogram and interpret its distribution for a given dataset.	**4**	**CO1**
Q 3	Discuss the importance of feature engineering in data preparation.	**4**	**CO2**
Q 4	Illustrate the use of bar plots and scatter plots with an example.	**4**	**CO2**
Q 5	How do heatmaps assist in understanding data correlations?	**4**	**CO2**
	SECTION B **(4Qx10M= 40 Marks)**		
Q 6	Discuss the steps involved in data cleaning and preparation, providing examples of techniques used at each step.	**10**	**CO4**
Q 7	Evaluate the use of advanced visualization techniques like 3D plots and interactive visualizations in representing complex data.	**10**	**CO3**
Q 8	Implement LDA on a given dataset and discuss how it differs from PCA in terms of feature extraction.	**10**	**CO3**
Q 9	Create a geospatial visualization using a provided dataset and explain the insights you can derive from it.	**10**	**CO4**
Q 10	Define seasonality in time-series data and its significance.	**10**	**CO3**
	SECTION-C **(2Qx20M=40 Marks)**		
Q 11	Given a dataset, perform a complete EDA and present your findings with appropriate visualizations.	**20**	**CO5**
Q 12	Analyze the role of dimensionality reduction in multivariate analysis and its impact on data visualization. Write a python script for applying PCA in any dataset of your choice.	**20**	**CO4**

Q 13	Explain the concept of distribution analysis and how it aids in understanding the underlying patterns in data. Include examples of different types of plots used.	**20**	**CO5**