|| Jai Sri Gurudev |

Sri AdichunchanagiriShikshana Trust (R)

SJB INSTITUTE OF TECHNOLOGY

SJBIT

Question Bank

Module 1

Subject Name: Exploratory Data Analytics

Subject Code: 23CSE422

By

Faculty Name: Mrs. Shilpashree S

Designation: Assistant Professor

Semester: IV

Department of Computer Science & Engineering

Aca. Year: Even Sem /2024-25

Module 1 Question Bank with Scheme of Evaluation(SoE)

Introduction to EDA: Exploratory Data Analysis Fundamentals

Sl. No.	Question with SoE	Marks
1.	**Describe the key phases of data analysis, emphasizing the role of Exploratory Data Analysis (EDA).**	**(10 Marks)**
SoE	**1. Introduction to Data Analysis Phases (2 Marks):** ● Briefly introduce the concept of data analysis and its significance in data science. ● Mention that data analysis includes several phases, each crucial for extracting insights from data. **2. Explanation of Each Key Phase (4 Marks):** ● **Data Requirements (0.5 Marks):** Define what data is needed and how to determine what types of data are required. ● **Data Collection (0.5 Marks):** Explain how data is gathered from various sources and stored in the correct format. ● **Data Processing (0.5 Marks):** Describe preprocessing steps that prepare the data for further analysis. ● **Data Cleaning (0.5 Marks):** Explain the importance of data cleaning, including techniques like error, duplicate, and missing value detection. ● **Modeling and Algorithms (0.5 Marks):** Outline the role of modeling in understanding relationships between variables. ● **Data Product (0.5 Marks):** Mention how the outcome of data analysis leads to creating data products. ● **Communication (0.5 Marks):** Discuss how the results are shared with stakeholders using visualization tools. **3. Emphasis on Exploratory Data Analysis (2 Marks):** ● Focus on the importance of EDA in understanding data. ● Describe methods used in EDA like descriptive statistics and visualizations to uncover patterns and trends. ● Example: How EDA is used to detect anomalies, outliers, and relationships within the data.	

	4. Real-life Example (1 Mark): ● Provide a relevant example from the textbook, such as the sensor data for a dementia health application, to illustrate the importance of the phases. **5. Conclusion (1 Mark):** ● Summarize the role of EDA in data analysis and its impact on the overall process. --- **Marks Breakdown:** 1. **Introduction to Phases**: 2 Marks 2. **Explanation of Phases**: 4 Marks 3. **Emphasis on EDA**: 2 Marks 4. **Real-life Example**: 1 Mark 5. **Conclusion**: 1 Mark	
2.	**What is the significance of Exploratory Data Analysis (EDA) in data mining? Describe its key components and how it helps in understanding and visualizing data.**	**10 Marks**
SoE	1. **Introduction to EDA (2 Marks):** ○ Define Exploratory Data Analysis and its significance. ○ Briefly explain its role as the first step in data mining. 2. **Key Components of EDA (4 Marks):** ○ Summarizing Data (1 Mark): Explain how data summarization is done (e.g., mean, median, standard deviation). ○ Statistical Analysis (1 Mark): Discuss the role of statistical analysis in EDA. ○ Data Visualization (1 Mark): Describe the types of visualizations used in EDA. ○ Tools for EDA (1 Mark): Mention Python libraries used for EDA (e.g., Pandas, SciPy, Matplotlib, Plotly). 3. **EDA in Data Mining (2 Marks):** ○ Explain how EDA uncovers insights and guides subsequent steps in the data mining process. ○ Discuss how it helps data scientists to form hypotheses and decide on models. 4. **Example or Application of EDA (1 Mark):** ○ Provide a practical example of EDA in a specific field (e.g., marketing, economics, or engineering). 5. **Conclusion (1 Mark):**	

	○ Summarize the significance of EDA and its role in preparing data for further analysis --- **Marks Breakdown:** ● **Introduction to EDA:** 2 Marks ● **Key Components (Summarizing, Statistical Analysis, Visualization, Tools):** 4 Marks ● **EDA's Role in Data Mining:** 2 Marks ● **Example/Application of EDA:** 1 Mark ● **Conclusion:** 1 Mark	
3.	**Describe the key steps involved in Exploratory Data Analysis (EDA) and explain the importance of each step.**	**10 Marks**
SoE	1. **Introduction to the Steps in EDA (2 Marks):** ○ Provide a brief overview of the steps in the EDA process. ○ Mention the four key steps involved: Problem Definition, Data Preparation, Data Analysis, and Development and Representation of Results. 2. **Explanation of Each Step (6 Marks):** ○ Problem Definition (1.5 Marks): Define the step and explain its significance, including tasks like objective definition, deliverables, and cost/benefit analysis. ○ Data Preparation (1.5 Marks): Describe the data preparation process, including tasks like defining data sources, cleaning the data, and transforming it. ○ Data Analysis (1.5 Marks): Discuss the tasks in data analysis, including summarization, finding correlations, developing models, and evaluating them. ○ Development and Representation of Results (1.5 Marks): Explain the importance of presenting results effectively and the techniques used for graphical representation. 3. **Example/Application of Each Step (1 Mark):** ○ Provide a real-life or practical example of how each of these steps is applied in a typical data analysis project. 4. **Conclusion (1 Mark):** ○ Summarize the importance of the structured steps in EDA for successful data analysis. --- **Marks Breakdown:** ● **Introduction to Steps in EDA: 2 Marks**	

	● **Explanation of Each Step: 6 Marks** ● **Example/Application of Each Step: 1 Mark** ● **Conclusion: 1 Mark**	
4.	**Define numerical and categorical data. Provide examples for each, highlighting the difference between discrete and continuous numerical data.**	**10 Marks**
SoE	**For Question 1: "Define numerical and categorical data. Provide examples for each, highlighting the difference between discrete and continuous numerical data. (10 Marks)"** 1. **Introduction to Numerical and Categorical Data (2 Marks):** ○ Clearly define both numerical and categorical data. (1 Mark each) 2. **Explanation of Numerical Data (2 Marks):** ○ Define discrete and continuous data with appropriate examples. (1 Mark each) 3. **Explanation of Categorical Data (2 Marks):** ○ Define categorical data and give examples for binary and polytomous variables. (1 Mark each) 4. **Difference Between Discrete and Continuous Data (2 Marks):** ○ Clearly explain how discrete and continuous data differ. (2 Marks) 5. **Conclusion (2 Marks):** ○ Provide a brief summary of the significance of distinguishing between numerical and categorical data in analysis. (2 Marks) **Marks Breakdown:** ● **Definitions: 4 Marks** ● **Examples: 4 Marks** ● **Conclusion: 2 Marks**	
5.	**Explain the concept of categorical data. Describe binary and polytomous variables with examples.**	**10 Marks**
SoE	1. **Introduction to Categorical Data (2 Marks):** ○ Provide a clear definition of categorical data. (2 Marks) 2. **Explanation of Binary Variables (3 Marks):** ○ Define binary variables and give relevant examples. (1.5 Marks) ○ Explain how binary data is used in analysis. (1.5 Marks) 3. **Explanation of Polytomous Variables (3 Marks):**	

	○ Define polytomous variables and provide examples. (1.5 Marks) ○ Explain the use of polytomous variables in analysis. (1.5 Marks) 4. **Conclusion (2 Marks):** ○ Summarize the role of binary and polytomous variables in data analysis. (2 Marks) **Marks Breakdown:** ● **Definitions: 5 Marks** ● **Examples: 4 Marks** ● **Conclusion: 1 Mark**	
6.	**Explain the four types of measurement scales in statistics. Provide examples for each scale and discuss the type of data that each scale is best suited for.**	**10 Marks**
SoE	**Evaluation Scheme:** ● **Introduction to Measurement Scales: 1 Mark** ● **Explanation of Nominal Scale: 2 Marks** ● **Explanation of Ordinal Scale: 2 Marks** ● **Explanation of Interval Scale: 2 Marks** ● **Explanation of Ratio Scale: 2 Marks** ● **Conclusion/Summary: 1 Mark**	
	Alternate Questions on " Measurement Scale"	
	1. **How would you visualize and analyze nominal data? Discuss the importance of understanding the type of data in deciding which computations and models can be applied.** **Explanation of Nominal Data** **Analysis Techniques for Nominal Data (e.g., Frequency, Proportion)** **Suitable Visualizations (Pie Chart, Bar Chart)** **Importance of Understanding Data Types** **Conclusion**	

	2. **Discuss the key differences between ordinal and interval scales. Provide examples where each scale is used and explain how data is analyzed differently based on the scale.** **Explanation of Ordinal Scale** **Explanation of Interval Scale** **Key Differences Between Ordinal and Interval Scales** **Examples and Applications** 3. **Describe how the ratio scale allows for more comprehensive statistical analysis compared to other scales. Provide real-life examples where ratio scales are used.** **Explanation of Ratio Scale** **Discussion on Comprehensive Statistical Analysis** **Examples of Ratio Scale Usage** **Conclusion**	
7.	**Compare and contrast the Classical, Exploratory Data Analysis (EDA), and Bayesian data analysis approaches. Highlight their execution steps and key differences.**	**10 Marks**
SoE	**Introduction to Data Analysis Approaches (2 Marks):** ● Briefly define the three approaches. **Detailed Steps of Each Approach (3 Marks):** ● Classical: Mention all steps. ● EDA: Highlight the swapping of model imposition and data analysis. ● Bayesian: Emphasize the use of prior distributions. **Comparison of the Approaches (4 Marks):** ● Model handling, flexibility, and use cases. ● Highlight differences in focus and methodologies. **Conclusion (1 Mark):**	

	• Summarize their relevance for different types of data analysis.	
	Alternate Question	
	1. **Explain the role of prior probability distribution in Bayesian data analysis. How does this differ from the approaches used in Classical and EDA methods?** **Definition of Prior Probability Distribution:** a. **Explain prior beliefs and their significance in Bayesian analysis.** **Role in Bayesian Data Analysis:** b. **How prior knowledge is combined with observed data.** **Comparison with Classical and EDA Approaches:** c. **Highlight the absence of prior knowledge in these approaches.** **Conclusion:** d. **Summarize the strengths of the Bayesian approach.**	
8.	**Explain the prerequisites for performing Exploratory Data Analysis (EDA) using Python and its libraries. Highlight why Python is suitable for EDA.**	**10 Marks**
SoE	**Introduction (2 Marks):** Define EDA and its role in data analysis. **Why Python? (3 Marks):** Discuss Python's libraries (NumPy, pandas, etc.) and community support. **Prerequisites (3 Marks):** List Python programming skills and familiarity with key libraries. **Conclusion (2 Marks):** Summarize Python's suitability for EDA.	
9.	**Write a Python script to create a 3D NumPy array, display its properties (shape, size, strides), and perform arithmetic operations on it.**	**10 Marks**

SoE	**Create 3D Array (2 Marks):** **Code snippet for creating and printing a 3D array.** **Display Properties (4 Marks):** **Code and explanation for displaying shape, size, and strides.** **Arithmetic Operations (3 Marks):** **Code and explanation for addition, subtraction, etc., on arrays.** **Output (1 Mark):** **Include printed results of the code.**	
10.	**Describe broadcasting rules in NumPy with examples, and explain why mismatched dimensions may lead to errors.**	**10 Marks**
SoE	**Introduction (2 Marks):** **Define broadcasting and its importance in NumPy.** **Rules (3 Marks):** **Explain the three rules with examples.** **Error Handling (3 Marks):** **Discuss why errors occur when rules are not met.** **Conclusion (2 Marks):** **Summarize with practical implications of broadcasting.**	
11.	**Perform the following tasks using the pandas library in Python:** **a. Write a Python code to set default display options for pandas and print its version.** **b. Explain how to create a DataFrame using a dictionary. Provide an example.** **c. Write code to load the UCI dataset into a pandas DataFrame and display the first 10 rows.** **d. Define and demonstrate how to select:** **- A specific row** **- A range of rows** **- Columns 3-5 for every other row.** **e. 5. Describe how to style a pandas DataFrame to highlight maximum and minimum values in different colors.**	**12 Marks**

SoE	**Set up pandas environment (1 marks):** ● 0.5 mark for correctly setting the default display options for pandas. ● 0.5 mark for printing the pandas version. **Create a DataFrame (1 marks):** ● 0.5 mark for explaining how to create a DataFrame using a dictionary. ● 0.5 mark for providing a correct example. **Load and explore data (2 marks):** ● 1 mark for correctly writing the code to load the UCI dataset. ● 1 mark for displaying the first 10 rows of the dataset. **Data selection (4 marks):** ● 1 mark for demonstrating how to select a specific row. ● 1 mark for demonstrating how to select a range of rows. ● 2 marks for demonstrating how to select columns 3-5 for every other row. **Styling DataFrames (4 marks):** ● 2 marks for correctly describing the styling process to highlight maximum and minimum values. ● 2 marks for demonstrating the process with appropriate code.	