

SRM VALLIAMMAI ENGINEERING COLLEGE

(An Autonomous Institution)
SRM Nagar, Kattankulathur – 603 203

DEPARTMENT OF ARTIFICIAL INTELLIGENCE & DATA SCIENCE

QUESTION BANK



III SEMESTER

AD3361 – DATA EXPLORATION AND VISUALIZATION

Regulation – 2023

Academic Year 2026 – 2027 (ODD)

Prepared by

Ms. V. Abarna, Assistant Professor (O.G)/AI&DS



SRM VALLIAMMAI ENGINEERING COLLEGE

(An Autonomous Institution)
SRM Nagar, Kattankulathur – 603 203.

DEPARTMENT OF ARTIFICIAL INTELLIGENCE & DATA SCIENCE



QUESTION BANK

SUBJECT : AD3361 – Data Exploration and Visualization
YEAR/SEM : II Year / III Semester

UNIT I EXPLORATORY DATA ANALYSIS			
EDA fundamentals – Understanding data science – Significance of EDA – Making sense of data - Comparing EDA with classical and Bayesian analysis – Software tools for EDA – Visual Aids for EDA – Data transformation techniques – merging database, reshaping and pivoting, Transformation techniques – Grouping Datasets – data aggregation – Pivot tables and cross- tabulations.			
PART – A			
Q.No	Questions	BT Level	Competence
1	Define Exploratory Data Analysis (EDA).	BTL1	Remember
2	Differentiate EDA from classical statistical analysis?	BTL2	Understand
3	What is the significance of EDA in data science?	BTL2	Understand
4	Give some examples of data transformation activities	BTL1	Remember
5	What are the two arguments required when merge () is used?	BTL1	Remember
6	What are the three phases of the pandas groupby() function?	BTL1	Remember
7	List the software tools available for EDA.	BTL1	Remember
8	What visual aids are commonly used in EDA?	BTL1	Remember
9	Differentiate between reshaping and pivoting in EDA?	BTL1	Understand
10	Define the term "data normalization" in EDA?	BTL1	Remember
11	What is the role of visual aids like violin plots in EDA?	BTL1	Understand
12	Why are pivot tables and cross-tabulations useful in summarizing data?	BTL1	Understand
13	Distinguish between numerical data and categorical data.	BTL2	Understand
14	Why do we need data transformation?	BTL2	Understand
15	Compare EDA with Bayesian analysis?	BTL2	Understand
16	What is the difference between merge(), join(), and concat()?	BTL2	Understand
17	Write a short note on the crosstab function in pandas.	BTL2	Understand
18	What are the reasons for using EDA?	BTL2	Understand
19	Write a short note on the crosstab function in pandas.	BTL2	Understand
20	Define data aggregation in EDA?	BTL2	Remember

21	Why is making sense of data important in EDA?	BTL2	Understand
22	What are the challenges of the process of data transformation?	BTL2	Understand
23	What does join () do in pandas?	BTL2	Understand
24	Compare Classical data analysis and Bayesian data analysis.	BTL2	Understand
PART-B			
1	(i) Explain the purpose of EDA in detail. (8) (ii) Illustrate the application of Pivot Tables. (8)	BTL3	Apply
2	Explain the steps involved in exploratory data analysis (EDA) and the significance of EDA. (16)	BTL3	Apply
3	What is cross-tabulation and pivot table? How do you build pivot table and cross tab reports? (16)	BTL3	Apply
4	Explain about merging databases using pandas. (16)	BTL3	Apply
5	(i) Write a detailed note on reshaping and pivoting. (9) (ii) Give an example Python's Pandas library program for creating a pivot table. (7)	BTL3	Apply
6	(i) Discuss about Descriptive Statistics in exploratory analysis. (8) (ii) Explain in detail about data transformation techniques. (8)	BTL3	Apply
7	What is the primary purpose of Exploratory Data Analysis (EDA)? Discuss the differences between EDA, classical analysis, and Bayesian analysis in detail. (16)	BTL4	Analyze
8	Explain data transformation and the different types of data transformation techniques. (16)	BTL4	Analyze
9	(i) Explain the common issues in data in detail. (7) (ii) Explain the data transformation technique. (9)	BTL4	Analyze
10	(i) Analyze the benefits of grouping datasets in data analysis. (7) (ii) Give an example program for grouping of data. (9)	BTL4	Analyze
11	Explain the role of Data Aggregation in detail. (16)	BTL3	Apply
12	Explain various types of Exploratory Data Analysis (EDA). (16)	BTL5	Evaluate
13	Evaluate the impact of different types of data transformation techniques on data quality and analysis outcomes. (16)	BTL5	Evaluate
14	Discuss in detail the process of data cleaning, including handling missing data, outlier detection, and treatment. Provide examples to illustrate each step. (16)	BTL5	Evaluate
15	What are the tools used for Exploratory Data Analysis (EDA)? Provide a detailed case study on applying EDA in a real business scenario. (16)	BTL6	Create
16	What do you mean by making sense of data? Explain the steps to make sense of data. (16)	BTL6	Create
17	Discuss various visual aids for Exploratory Data Analysis (EDA). (16)	BTL6	Create

UNIT II VISUALIZING USING MATPLOTLIB			
Importing Matplotlib – Simple line plots – Simple scatter plots – visualizing errors – density and contour plots – Histograms – legends – colors – subplots – text and annotations – customization - Three-dimensional plotting – Geographic Data with Base map – Visualization with seaborn.			
PART – A			
Q.No	Questions	BT Level	Competence
1	What is visualization?	BTL1	Remember
2	What is Matplotlib used for?	BTL1	Remember
3	What is Geographic Data Visualization with Basemap?	BTL1	Remember
4	Mention different ways colors can be applied to plots in Matplotlib.	BTL1	Remember
5	Define subplots in Matplotlib.	BTL1	Remember
6	What is the use of Basemap in Python?	BTL1	Remember
7	What are the plots drawn by Seaborn?	BTL1	Remember
8	Identify the purpose of using error bars in visualization.	BTL2	Understand
9	Define the term "Colormap" in Matplotlib	BTL1	Remember
10	What is Seaborn?	BTL1	Remember
11	What is the use of Python Seaborn library?	BTL1	Remember
12	Draw and label a rough contour plot of the joint probability density function. When $P=-0.4$ $\rho=-0.4$.	BTL2	Understand
13	Differentiate between Colors and Colormaps in Matplotlib.	BTL2	Understand
14	What do you understand by annotating text?	BTL2	Understand
15	Differentiate plot vs subplot.	BTL2	Understand
16	Define Three-Dimensional Plotting in Matplotlib.	BTL1	Remember
17	What is the scatter () function in Python?	BTL2	Understand
18	What does plot function do in Matplotlib?	BTL2	Understand
19	Define the term “Scatter Plot matrix” in Matplotlib.	BTL1	Remember
20	What is the advantage of using Seaborn for visualization?	BTL2	Understand
21	Differentiate between Line Plots and Contour Plots	BTL2	Understand
22	Is a histogram always a bar chart? Justify with your answer.	BTL2	Understand
23	What is the use of density plot?	BTL2	Understand
24	What is a legend in Matplotlib?	BTL2	Understand
PART-B			
1	(i) Explain in detail about importing Matplotlib. (6) (ii) Describe the customization options in Matplotlib. (10)	BTL3	Apply
2	(i) Explain in detail about three-dimensional data visualization, its components and its workflow with suitable example. (8) (ii) Discuss in detail about text and annotation. (10)	BTL3	Apply
3	An airline wants to display its national flight routes. Using Basemap, apply geographic plotting techniques to map all routes between major	BTL3	Apply

	cities. Explain how geographic visualization aids in route optimization. (16)																		
4	Discuss how Seaborn helps to visualize the statistical relationships. Illustrate with code and example. (16)	BTL3	Apply																
5	How to over plot a line on a scatter plot in Python? Illustrate with code. (16)	BTL3	Apply																
6	A retail company recorded monthly sales for three regions over a year. Apply Matplotlib to create a comparative visualization using line plots and scatter plots. Explain how your plotted visualizations help the company identify seasonal trends and region-wise differences. (16)	BTL3	Apply																
7	Explain how density plots and contour plot help in understanding data distribution. (16)	BTL3	Apply																
8	Discuss various customization options in scatter plots. (16)	BTL4	Analyze																
9	Explain why Matplotlib is used for data visualization? Which module of Matplotlib is used for data visualization? (16)	BTL4	Analyze																
10	Write a detailed note on visualizing errors (16)	BTL3	Apply																
11	(i) What are scatter plots? Explain scatter () methods and its syntax. (7) (ii) Explain the steps to plot error bars in Matplotlib and the syntax to plot error bars. (9)	BTL3	Apply																
12	Write a note on Basemap. Explain geographic data with Basemap. (16)	BTL3	Apply																
13	Define line plot. With an example, explain how to create a line plot to visualize the trend. (8) (ii) The following table gives the lifetime of 400 neon lamps. Draw the histogram for the below data. (8) <table><tr><th>Lifetime (in hours)</th><th>Number of lamps</th></tr><tr><td>300–400</td><td>14</td></tr><tr><td>400–500</td><td>56</td></tr><tr><td>500–600</td><td>60</td></tr><tr><td>600–700</td><td>86</td></tr><tr><td>700–800</td><td>74</td></tr><tr><td>800–900</td><td>62</td></tr><tr><td>900–1000</td><td>48</td></tr></table>	Lifetime (in hours)	Number of lamps	300–400	14	400–500	56	500–600	60	600–700	86	700–800	74	800–900	62	900–1000	48	BTL3	Apply
Lifetime (in hours)	Number of lamps																		
300–400	14																		
400–500	56																		
500–600	60																		
600–700	86																		
700–800	74																		
800–900	62																		
900–1000	48																		
14	(i) Compare and contrast Matplotlib and Seaborn for data visualization. (5) (ii) Discuss the visualizing errors in detail. (5) (iii) Explain contour plots in detail with example program. (6)	BTL3	Apply																
15	(i) Write a note on visualization with Seaborn. (7) (ii) Write a short note on (i) Density Plots (ii) Histograms (iii) Subplots. (9)	BTL6	Create																

16	Explain three-dimensional plotting with example programs. (16)	BTL6	Create
17	Create a Python program that demonstrates the use of Seaborn for creating a pair plot and explain its significance in data visualization. (16)	BTL6	Create

UNIT III UNIVARIATE ANALYSIS			
Introduction to Single variable: Distributions and variables – Numerical Summaries of Level and Spread – Scaling and Standardizing – Inequality – Smoothing Time Series.			
PART – A			
Q.No	Questions	BT Level	Competence
1	What is meant by univariate analysis?	BTL1	Remember
2	What is the primary purpose of univariate analysis?	BTL1	Remember
3	What is the mathematical mean of the following numbers? 10, 6, 4, 4, 6, 4.	BTL1	Remember
4	List the three main types of univariate analyses.	BTL1	Remember
5	What is normalization or min-max scaling?	BTL1	Remember
6	List the steps involved in univariate analysis.	BTL1	Remember
7	Define the Coefficient of Variation (CV)?	BTL1	Remember
8	Differentiate between Level and Spread in Data?	BTL2	Understand
9	Define Interquartile Range (IQR)?	BTL1	Remember
10	Define Skewness in a Distribution?	BTL1	Remember
11	What is Z-Score in Standardization?	BTL1	Remember
12	Define Single Variable?	BTL1	Remember
13	Differentiate between Scaling and Standardizing?	BTL2	Understand
14	Define Time Series Smoothing?	BTL1	Remember
15	Differentiate between Mean and Median?	BTL2	Understand
16	What is the Purpose of Scaling Variables?	BTL2	Understand
17	Give one example each of a discrete variable and a continuous variable.	BTL1	Remember
18	Define Outliers in a Dataset?	BTL1	Remember
19	What is the Purpose of Time Series Smoothing Techniques?	BTL2	Understand
20	Write a short note on quartiles	BTL2	Understand
21	Write a short note on descriptive statistics.	BTL2	Understand
22	What is the purpose of inequality measures in data analysis?	BTL2	Understand
23	Define Kernel Density Estimation (KDE).	BTL1	Remember
24	What is the role of a boxplot in data visualization?	BTL2	Understand
PART-B			
1	(i) Explain in detail, Scaling and standardization. (10) (ii) Explain feature scaling in detail. (6)	BTL3	Apply
2	Given the distribution of daily calorie intake of athletes, apply	BTL3	Apply

	numerical summaries and inequality measures to evaluate dietary balance. Explain how measures of spread and inequality reveal variations in consumption habits. (16)		
3	Apply time series analysis technique to demonstrate their usefulness in data analysis. Provide examples from various domains such as finance, weather forecasting, or sales forecasting. (16)	BTL3	Apply
4	What is scaling and standardization? When and why to standardize a variable? Illustrate with a suitable example. (16)	BTL3	Apply
5	Explain the smoothing techniques for time series data with a suitable example. (16)	BTL3	Apply
6	(i) Explain univariate statistics in detail. (7) (ii) Discuss with examples how univariate analysis is helpful. (9)	BTL4	Analyze
7	Analyze the role of variables and distribution in univariate analysis in detail. (16)	BTL4	Analyze
8	Explain numerical summaries level and spread used in univariate. (16)	BTL3	Apply
9	Distinguish between Univariate and Multivariate analysis with suitable example. (16)	BTL4	Analyze
10	(i) What are univariate tables? (4) (ii) Explain in detail the steps involved in conducting univariate analysis. (12)	BTL4	Analyze
11	Evaluate the importance of smoothing methods in time series analysis. Discuss different smoothing techniques and their applications with suitable examples. (16)	BTL5	Evaluate
12	Evaluate the effectiveness of various methods for performing univariate analysis compare and contrast these methods. Providing examples to illustrate their strength and weakness. (16)	BTL5	Evaluate
13	How, when, and why should you normalize/standardize/rescale your data? (16).	BTL5	Evaluate
14	Given a dataset of students' exam scores from two different departments, apply scaling and standardization techniques. Compare the transformed datasets and discuss how these transformations affect interpretation and comparability. (16)	BTL3	Apply
15	Explain the 10 Essential Numerical Summaries in statistics with example. (16)	BTL6	Create
16	What is a time series? Explain types, properties, and decomposition of time series. (16)	BTL6	Create
17	(i) Describe the role of Interquartile Range (IQR) in data analysis. (8) (ii) Explain the importance of outlier detection in data analysis. (8)	BTL3	Apply

UNIT IV BIVARIATE ANALYSIS			
Relationship between Two variables – Percentage Tables – Analyzing Contingency Tables - Handling Several Batches – Scatterplots and Resistant Lines – Transformations.			
PART – A			
Q.No	Questions	BT Level	Competence
1	What is meant by bivariate analysis?	BTL1	Remember
2	List the three common ways to perform bivariate analysis.	BTL1	Remember
3	What are the types of bivariate analysis?	BTL1	Remember
4	List the two main types of variables.	BTL1	Remember
5	What is the role of Scatterplots in identifying patterns?	BTL2	Understand
6	What are the possible types of correlation between two variables in bivariate analysis?	BTL1	Remember
7	What is cross-tabulation?	BTL1	Remember
8	Name the two main types of statistical testing in bivariate analysis.	BTL1	Remember
9	What is a contingency table?	BTL1	Remember
10	What is a marginal proportion in a contingency table?	BTL1	Remember
11	What is batch processing?	BTL1	Remember
12	Is bivariate qualitative or quantitative?	BTL1	Remember
13	Compare correlation and causality.	BTL2	Understand
14	Write a short note on percentage tables.	BTL2	Understand
15	State the purpose of analyzing the relationship between two variables.	BTL2	Understand
16	What is the crosstab function in Python?	BTL1	Remember
17	What is the purpose of a contingency table?	BTL2	Understand
18	What are the benefits of Python batch processing?	BTL2	Understand
19	What is the purpose of resistant lines in outlier prone data?	BTL2	Understand
20	State the purpose of resistant lines in bivariate analysis.	BTL2	Understand
21	Define a bivariate table.	BTL1	Remember
22	Write a short note on simple linear regression.	BTL2	Understand
23	Outline the difference between univariate and bivariate data.	BTL2	Understand
24	Mention the insights that batch-wise analysis reveals in datasets.	BTL2	Understand
PART-B			
1	Discuss the common ways to perform bivariate analysis in detail. (16)	BTL3	Apply
2	Discuss various types of bivariate analysis. (16)	BTL3	Apply
3	(i) Discuss in detail about contingency table with example. (8) (ii) Explain in detail about percentage table with suitable example. (8)	BTL3	Apply
4	What are data transformations in bivariate analysis? (16)	BTL3	Apply
5	Explain how patterns and trends are identified through scatterplot analysis? (16)	BTL3	Apply
6	Apply the method of handling several batches of data to analyze relationships between variables in a large dataset. Detail the steps	BTL3	Apply

	involved and discuss the insights gained from this approach.																								
7	What is a table of frequency values for a bivariate distribution? Explain what graph is used in the analysis of bivariate data? (16)	BTL4	Analyze																						
8	Analyze how resistant lines are used in bivariate analysis. (16)	BTL4	Analyze																						
9	How will you analyze contingency tables? Illustrate with an example. (16)	BTL4	Analyze																						
10	A retail company collects data on customers' age groups and their preferred product categories. Using this dataset, (i) Construct appropriate percentage tables. (6) (ii) Develop contingency tables and analyze the relationships between the two categorical variables. (6) (iii) Recommend actionable insights that could support marketing decisions. (4)	BTL3	Apply																						
11	Analyze the role of resistant lines in scatterplots for understanding the relationship between two numerical variables. Use examples to illustrate how resistant lines can help identify trends and outliers.	BTL4	Analyze																						
12	Draw a scatter plot for the given data that shows the number of games played and scores obtained in each instance. With this plot explain scatter plot correlation and its types and justify which type of correlation it belongs to with neat illustration. (16) <table border="1"><tr><td>No.of games</td><td>3</td><td>5</td><td>2</td><td>6</td><td>7</td><td>1</td><td>2</td><td>7</td><td>1</td><td>7</td></tr><tr><td>Scores</td><td>80</td><td>90</td><td>75</td><td>80</td><td>90</td><td>50</td><td>65</td><td>85</td><td>40</td><td>100</td></tr></table>	No.of games	3	5	2	6	7	1	2	7	1	7	Scores	80	90	75	80	90	50	65	85	40	100	BTL4	Analyze
No.of games	3	5	2	6	7	1	2	7	1	7															
Scores	80	90	75	80	90	50	65	85	40	100															
13	Explain in detail about relationship between two variables in the percentage tables. (16)	BTL4	Analyze																						
14	A social media analysis team studies the relationship between time spent online and number of posts shared, using data collected from three user categories, viz. students, professionals, creators. (i) Build percentage tables for categorical comparisons. (6) (ii) Create scatterplots for each category and fit resistant lines. (6) (iii) Apply transformation to reduce skewness and explain how the transformed relationship helps in understanding engagement behavior. (4)	BTL3	Apply																						
15	Explain the use of contingency tables in bivariate analysis with example. (16)	BTL6	Create																						
16	Explain the challenges and strategies in handling several batches in data analysis. (16)	BTL4	Analyze																						
17	Discuss in detail about scatter plot and resistant lines in bivariate analysis. (16)	BTL6	Create																						

UNIT V MULTIVARIATE AND TIME SERIES ANALYSIS			
Introducing a Third Variable – Casual Explanations – Three-Variable Contingency Tables and Beyond – Longitudinal Data – Fundamentals of TSA – Characteristics of time series data – Data Cleaning – Time-based indexing – Visualizing – Grouping – Resampling.			
PART – A			
Q.No	Questions	BT Level	Competence
1	What is meant by multivariate analysis?	BTL1	Remember
2	List the basic cleaning steps in time series data.	BTL1	Remember
3	What is the role of grouping in visualizing time series data?	BTL2	Understand
4	Define causal explanations in data analysis?	BTL1	Remember
5	List the five common relationships among three variables in a statistical model.	BTL1	Remember
6	State the four reasons for which multivariate analyses are used.	BTL1	Remember
7	What are the two types of multivariate analysis techniques?	BTL1	Remember
8	List the types of time series analysis.	BTL1	Remember
9	State the fundamentals of Time Series Analysis (TSA).	BTL1	Remember
10	Differentiate between up-sampling and down-sampling.	BTL2	Understand
11	List the applications of multivariate analysis.	BTL1	Remember
12	Differentiate Between Two-Variable and Three-Variable Contingency Tables?	BTL2	Understand
13	What is the importance of data cleaning in time series analysis?	BTL2	Understand
14	What is the significance of longitudinal data in analysis?	BTL2	Understand
15	Define Characteristics of Time Series Data?	BTL1	Remember
16	Define the term "Third Variable" in data analysis?	BTL1	Remember
17	Differentiate Between Cross-Sectional and Longitudinal Data?	BTL2	Understand
18	What are the Characteristics of Seasonality in Time Series Data?	BTL2	Understand
19	What is Resampling in the Context of Time Series Analysis?	BTL1	Remember
20	What is the purpose of time-based indexing in data analysis?	BTL2	Understand
21	Define the concept of resampling frequency in time series analysis?	BTL1	Remember
22	Define Time Series Analysis (TSA)	BTL1	Remember
23	What is the purpose of a three-variable contingency table?	BTL2	Understand
24	What is the role of causal explanations in predictive modeling?	BTL2	Understand
PART-B			
1	What is meant by time series data? Describe its four components. (16)	BTL3	Apply
2	Discuss the various dependence techniques used in multivariate analysis, explaining each technique briefly. (16)	BTL3	Apply
3	What are the characteristics of multivariate analysis? How do you explain multivariate analysis? (16)	BTL3	Apply
4	How is Time Series Analysis used in weather (climate) forecasting? Illustrate with an example. (16)	BTL3	Apply
5	How do you analyze contingency table? Give examples. (16)	BTL3	Apply

6	Analyze how resampling alters the interpretation of time series data using a stock market dataset. Discuss how grouping by week or month and using suitable visualization helps uncover long-term trends and volatility patterns. (16)	BTL4	Analyze
7	Provide a detailed description of causal explanations in statistics, including examples to illustrate the concept.(16)	BTL4	Analyze
8	Describe what three-variable contingency tables are, how they are constructed, and their significance in data analysis.	BTL4	Analyze
9	Discuss common relationships among three variables in a statistical model. (16)	BTL4	Analyze
10	Analyze a longitudinal dataset tracking patients' blood pressure reading over several months. Explain how the repeated-measures structure supports casual reasoning, trend identification and forecasting. Compare its analytical advantages to a single-time- point health survey. (16)	BTL4	Analyze
11	Analyze the characteristics of time series data with suitable example. (16)	BTL4	Analyze
12	Differentiate univariate and multivariate time series analysis in detail with suitable example. (16)	BTL4	Analyze
13	Define longitudinal data and explain its importance in statistical analysis, including examples of its application. (16)	BTL5	Evaluate
14	Discuss the patterns that can appear in time series plots, including trends, seasonality, cycles, and irregularities, and analyze their implications for data analysis and forecasting. (16)	BTL5	Evaluate
15	Evaluate the effectiveness and applications of contingency tables in data analysis. (16)	BTL5	Evaluate
16	Design a comprehensive study utilizing time series analysis to forecast sales for a retail company. (16)	BTL6	Create
17.	Provide a comprehensive explanation of data cleaning, outlining the steps involved, common techniques used, and the importance of data cleaning in ensuring data quality and reliability. (16)	BTL6	Create


