SRM VALLIAMMAI ENGINEERING COLLEGE

(An Autonomous Institution)

SRM Nagar, Kattankulathur – 603 203

DEPARTMENT OF INFORMATION TECHNOLOGY

(Common to Artificial Intelligence and Data Science)

QUESTION BANK

III SEMESTER

AD3361 – DATA EXPLORATION AND VISUALIZATION

Regulation – 2023

Academic Year 2025 – 2026 (ODD)

Prepared by

Ms.S.Jonisha, Assistant Professor (O.G)/IT

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

Ms.M.Kanmani, Assistant Professor (O.G)/IT

Mr.B.Yogeshkumar, Assistant Professor (O.G) /AI&DS

SRM VALLIAMMAI ENGINEERING COLLEGE

(An Autonomous Institution)

SRM Nagar, Kattankulathur – 603 203.

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

DEPARTMENT OF INFORMATION TECHNOLOGY

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	List the widely used data collection tools.	**BTL1**	**Remember**
3	What are the major stages in the data science life cycle?	**BTL1**	**Remember**
4	What is a Likert Scale?	**BTL1**	**Remember**
5	Give some examples of data transformation activities	**BTL1**	**Remember**
6	List the visual tools used to analyze univariate data.	**BTL1**	**Remember**
7	What are the two arguments required when merge () is used?	**BTL1**	**Remember**
8	What are the three phases of the pandas groupby() function?	**BTL1**	**Remember**
9	List the software tools available for EDA.	**BTL1**	**Remember**
10	What is a run sequence plot?	**BTL1**	**Remember**
11	List down the popular data analysis approaches.	**BTL1**	**Remember**
12	What is cluster map used for?	**BTL1**	**Remember**
13	Distinguish between numerical data and categorical data.	**BTL2**	**Understand**
14	Why do we need data transformation?	**BTL2**	**Understand**
15	Why do we use lag plot?	**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	Write a note on data aggregation using an aggregate function in pandas.	**BTL2**	**Understand**
21	Write a note on 'split data into groups'.	**BTL2**	**Understand**
22	What is the purpose of reshaping and pivoting in data transformation?	**BTL2**	**Understand**
23	What does join () do in pandas?	**BTL2**	**Understand**
24	How is Exploratory Data Analysis (EDA) different from Classical and Bayesian analysis?	**BTL2**	**Understand**

PART-B			
1	Discuss the role of data wrangling in data science. Explain various techniques involved in cleaning, transforming, and enriching raw data. (16)	**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	Explain the various phases of data science. Analyze how each phase contributes to the overall data science process (16)	**BTL4**	**Analyze**
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 data science life cycle. (16)	**BTL5**	**Evaluate**
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**	**BTL Level**	**Competence**
1	What is visualization?	**BTL1**	**Remember**
2	What is Matplotlib used for?	**BTL1**	**Remember**
3	Why do we import Matplotlib in Python?	**BTL1**	**Remember**
4	Write a short note on Pyplot.	**BTL1**	**Remember**
5	Give any four objects in Basemap.	**BTL1**	**Remember**
6	What is the use of Basemap in Python?	**BTL1**	**Remember**
7	What are the plots drawn by Seaborn?	**BTL1**	**Remember**
8	List any four types of error bars.	**BTL1**	**Remember**
9	What is Matplotlib Basemap?	**BTL1**	**Remember**
10	What is Seaborn?	**BTL1**	**Remember**
11	What is the use of Python Seaborn library?	**BTL1**	**Remember**
12	Write a short note on Matplotlib.	**BTL1**	**Remember**
13	Write a note on format strings.	**BTL2**	**Understand**
14	What do you understand by annotating text?	**BTL2**	**Understand**
15	What are error bars used for?	**BTL2**	**Understand**
16	Write a note on three-dimensional plotting.	**BTL2**	**Understand**
17	What is the scatter () function in Python?	**BTL2**	**Understand**
18	What does plot function do in Matplotlib?	**BTL2**	**Understand**
19	What is the difference between Matplotlib and Pyplot?	**BTL2**	**Understand**
20	What is matplotlib. pyplot?	**BTL2**	**Understand**
21	What are scatter plots?	**BTL2**	**Understand**
22	Discuss various visual aids for EDA.	**BTL2**	**Understand**
23	What is the use of density plot?	**BTL2**	**Understand**
24	What is the significance of density and contour plots in EDA?	**BTL2**	**Understand**

PART-B			
1	(i) Explain in detail about importing Matplotlib. (7) (ii) Explain the Pyplot library and plt function. (9)	**BTL3**	**Apply**
2	Discuss various simple plots for univariate data analysis. (16)	**BTL3**	**Apply**
3	Explain categorical variables and plotting with categorical variables. (16)	**BTL3**	**Apply**
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	Describe the various distribution modules of Seaborn for visualization. Consider a sample application to illustrate. (16)	**BTL4**	**Analyze**
7	Discuss how Seaborn helps to visualize statistical relationships. Illustrate with code and example. (16)	**BTL4**	**Analyze**
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)	**BTL5**	**Evaluate**
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)	**BTL5**	**Evaluate**
12	What is Basemap in Python? How can you visualize geographic data like maps and country borders using Basemap? Give an example. (16)	**BTL5**	**Evaluate**
13	Evaluate the impact of different types of data transformation techniques on data quality and analysis outcomes. (16)	**BTL5**	**Evaluate**
14	(i) Compare and contrast Matplotlib and Seaborn for data visualization. (5) (ii) Discuss the visualizing errors in detail. (5) Explain contour plots in detail with example program. (6)	**BTL5**	**Evaluate**
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 a numerical summary?	**BTL1**	**Remember**
4	What is range scaling?	**BTL1**	**Remember**
5	What is normalization or min-max scaling?	**BTL1**	**Remember**
6	List the steps involved in univariate analysis.	**BTL1**	**Remember**
7	What are the three components of an observed time series?	**BTL1**	**Remember**
8	What is ADF test?	**BTL1**	**Remember**
9	What is ARIMA model?	**BTL1**	**Remember**
10	What are smoothing methods?	**BTL1**	**Remember**
11	State the two types of time series.	**BTL1**	**Remember**
12	What is detrending?	**BTL1**	**Remember**
13	Why is scaling necessary for algorithms such as SVM and KNN?	**BTL2**	**Understand**
14	Write a short note on standardization or z-score normalization.	**BTL2**	**Understand**
15	What is the difference between stock and flow series?	**BTL2**	**Understand**
16	Write a short note on trend stationary time series.	**BTL2**	**Understand**
17	Write a short note on inferential statistics.	**BTL2**	**Understand**
18	Give some hypothesis testing mechanisms.	**BTL2**	**Understand**
19	Write a short note on univariate tables.	**BTL2**	**Understand**
20	Define inequality in the context of data distribution.	**BTL2**	**Understand**
21	What are summary statistics?	**BTL2**	**Understand**
22	Write a short note on descriptive statistics.	**BTL2**	**Understand**
23	Write a short note on quartiles.	**BTL2**	**Understand**
24	Why should time series be stationary?	**BTL2**	**Understand**

PART-B			
1	(i) Explain in detail, Scaling and standardization. (9) (ii) Explain feature scaling in detail. (7)	**BTL3**	**Apply**
2	Explain numerical summaries level and spread used in univariate. (16)	**BTL3**	**Apply**
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 how stationarity can be checked in a time series. (16)	**BTL4**	**Analyze**
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	Discuss the need for and advantages of standardizing variables before applying machine learning algorithms like KNN and SVM. Include examples and Python code to demonstrate the process. (16)	**BTL5**	**Evaluate**
13	How, when, and why should you normalize/standardize/rescale your data? (16).	**BTL5**	**Evaluate**
14	Give a brief note on (i) One-Sample Sign Test (ii) Wilcoxon Signed Rank	**BTL5**	**Evaluate**
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	Write a detailed note on (i) ARIMA models. (9) (ii)Time series data visualization. (7)	**BTL6**	**Create**

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 are the three possible combinations of variables for bivariate analysis?	**BTL1**	**Remember**

6	What are the possible types of correlation between two variables in bivariate analysis?	**BTL1**	**Remember**
7	What is cross-tabulation?	**BTL1**	**Remember**
8	What is elaboration?	**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	Write some key features of joblib.	**BTL1**	**Remember**
13	Compare correlation and causality.	**BTL2**	**Understand**
14	Write a short note on percentage tables.	**BTL2**	**Understand**
15	State the limitations of elaboration.	**BTL2**	**Understand**
16	Briefly explain the crosstab function in Python.	**BTL2**	**Understand**
17	What is the purpose of a contingency table?	**BTL2**	**Understand**
18	What are the benefits of Python batch processing?	**BTL2**	**Understand**
19	Write a short note on bivariate analysis using scatter plots.	**BTL2**	**Understand**
20	Briefly explain bivariate analysis resistant lines.	**BTL2**	**Understand**
21	Briefly explain a bivariate table.	**BTL2**	**Understand**
22	Write a short note on simple linear regression.	**BTL2**	**Understand**
23	Briefly explain scatter plots.	**BTL2**	**Understand**
24	Write a short note on joblib in Python?	**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	Write Python code to find percentage table and contingency table or the above example table and explain it. (16)	**BTL3**	**Apply**
4	What are data transformations in bivariate analysis? (16)	**BTL3**	**Apply**
5	Discuss the best practices for designing scatter plots. (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 involved and discuss the insights gained from this approach.	**BTL3**	**Apply**
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	There are three types of scatter plots or charts: U-shaped, linear and exponential. Which one is the most significant for managing data bias? Discuss. (16)	**BTL4**	**Analyze**

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	Evaluate the effectiveness of using scatter plots for bivariate analysis by discussing its strengths and limitations. (16)	**BTL5**	**Evaluate**
13	Evaluate the use of percentage tables in bivariate analysis with examples. (16)	**BTL5**	**Evaluate**
14	Explain in detail about relationship between two variables in the percentage tables. (16)	**BTL5**	**Evaluate**
15	Explain the use of contingency tables in bivariate analysis with example. (16)	**BTL6**	**Create**
16	Design and implement a Python batch processing system for bivariate analysis. Explaining each step and its importance. (16)	**BTL2**	**Create**
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 some of the criteria for data quality.	**BTL1**	**Remember**
3	Write a short note on auto-correlation	**BTL1**	**Remember**
4	When is the Durbin-Watson test used?	**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	Write a short note on covariate independence of one variable in multivariate analysis.	**BTL1**	**Remember**
10	List the goals of multivariate analysis.	**BTL1**	**Remember**
11	List the applications of multivariate analysis.	**BTL1**	**Remember**
12	What is an ARIMA model?	**BTL1**	**Remember**
13	What is factor analysis?	**BTL2**	**Understand**

14	What does it mean for a time series to be stationary?	**BTL2**	**Understand**
15	What are the methods for estimating ARIMA models?	**BTL2**	**Understand**
16	What are dependence techniques?	**BTL2**	**Understand**
17	What is canonical correlation analysis?	**BTL2**	**Understand**
18	Write a short note on multivariate analysis of variance.	**BTL2**	**Understand**
19	What is cluster analysis?	**BTL2**	**Understand**
20	When is the Durbin-Watson test used?	**BTL2**	**Understand**
21	Write a short note on the cross-correlation function.	**BTL2**	**Understand**
22	Brief about multidimensional scaling?	**BTL2**	**Understand**
23	Brief about multivariate outliers.	**BTL2**	**Understand**
24	How do dependence techniques differ from interdependence techniques?	**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	Discuss in detail, the interdependence techniques. (16)		
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	What is TSA analysis? Explain ARIMA, smooth-based, and moving average. (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**