

SRM VALLIAMMAI ENGINEERING COLLEGE

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

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

QUESTION BANK



VII SEMESTER

PAD 107-EXPLORATORY DATA ANALYSIS

Regulation – 2023

Academic Year 2026 – 2027 (ODD)

Prepared by

Dr..A.Vidhya(Sr.G)/CSE

Ms. V. Prema (Sr.G)/CSE

Ms.R.Ranjani (O.G)/CSE



SRM VALLIAMMAI ENGINEERING COLLEGE

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

DEPARTMENT OF ARTIFICIAL INTELLIGENCE & DATA SCIENCE



QUESTION BANK

SUBJECT : Exploratory Data Analysis

YEAR/SEM : IV Year /VII Sem

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 .			
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 cross tab function in pandas.	BTL2	Understand
20	Describe the concept of reshaping and pivoting data.	BTL2	Understand
21	What data transformation techniques are intended to achieve.	BTL2	Understand
22	State the purpose of merging database-style datasets.	BTL2	Understand
23	Write the general steps involved in performing EDA.	BTL2	Understand

24	Compare Classical data analysis and Bayesian data analysis.	BTL2	Understand
PART-B			
1	Define data science. Explain data science tools, uses, and applications of data science. (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 EDA USING PYTHON
Data Manipulation using Pandas – Pandas Objects – Data Indexing and Selection – Operating on Data – Handling Missing Data – Hierarchical Indexing – Combining datasets – Concat, Append, Merge and Join – Aggregation and grouping – Pivot Tables – Vectorized String Operations.

PART – A			
Q.No	Questions	BT Level	Competence
1	What is Pandas and what is its role in data manipulation?	BTL1	Remember
2	Name the three fundamental Pandas data structures.	BTL1	Remember
3	Define a Pandas Series.	BTL1	Remember
4	Define a Pandas DataFrame.	BTL1	Remember
5	What is a Pandas Index object?	BTL1	Remember
6	List the indexer attributes used for data selection in Pandas.	BTL1	Remember
7	How is missing data represented in Pandas?	BTL1	Remember
8	Name the methods used to detect and remove missing values.	BTL1	Remember
9	Define hierarchical (multi-level) indexing.	BTL1	Remember
10	List the four operations used to combine datasets in Pandas.	BTL1	Remember
11	Name the Pandas method used to build pivot tables.	BTL1	Remember
12	What are vectorized string operations in Pandas?	BTL1	Remember
13	Explain the difference between a Pandas Series and a DataFrame.	BTL2	Understand
14	Explain how a Pandas Series differs from a NumPy array.	BTL2	Understand
15	Differentiate between the loc and iloc indexers.	BTL2	Understand
16	Explain index alignment when operating on Pandas data.	BTL2	Understand
17	Explain how Pandas handles missing data using NaN and None.	BTL2	Understand
18	Differentiate between the dropna() and fillna() methods.	BTL2	Understand
19	Explain the concept of hierarchical indexing with an example.	BTL2	Understand
20	Differentiate between concatenation and merging of datasets.	BTL2	Understand
21	Explain the difference between append and concat.	BTL2	Understand
22	Explain the split-apply-combine idea behind the groupby operation.	BTL2	Understand
23	Explain how a pivot table summarizes data.	BTL2	Understand
24	Explain why vectorized string operations are preferred over explicit Python loops.	BTL2	Understand
PART-B			
1	Given raw data, demonstrate how you would create Pandas Series and DataFrame objects and access their elements using the loc and iloc indexers.	BTL3	Apply
2	Demonstrate how you would detect, remove and fill missing values in a given dataset.	BTL3	Apply
3	Apply hierarchical indexing to represent and access multi-dimensional data within a single DataFrame.	BTL3	Apply
4	Given two related datasets, demonstrate how concat, merge and join would each be applied to combine them, and show the outcome of each.	BTL3	Apply
5	Apply groupby and aggregation to summarize a dataset by one or more categorical variables.	BTL3	Apply
6	Analyse the different data indexing and selection methods in Pandas and examine when each is appropriate.	BTL4	Analyze

7	Differentiate between the four dataset-combining operations (concat, append, merge, join), analysing the situation in which each is used.	BTL4	Analyze
8	Analyse how index alignment affects arithmetic operations performed on Pandas objects.	BTL4	Analyze
9	Analyse the split-apply-combine strategy that underlies the groupby operation.	BTL4	Analyze
10	Compare pivot tables with groupby aggregation, analysing their similarities and differences.	BTL5	Evaluate
11	Evaluate the advantages of Pandas over NumPy for data manipulation tasks.	BTL5	Evaluate
12	Assess the different strategies for handling missing data and justify when each is appropriate.	BTL5	Evaluate
13	Justify the use of hierarchical indexing for representing higher-dimensional data in a DataFrame.	BTL5	Evaluate
14	Evaluate the trade-offs between merge/join and concatenation when combining datasets.	BTL5	Evaluate
15	Justify why vectorized string operations improve both efficiency and code clarity.	BTL6	Create
16	Design a complete data-manipulation pipeline in Pandas for a raw dataset — covering loading, indexing, handling missing data, combining sources, and aggregating — and specify the methods used at each step.	BTL6	Create
17	Propose a Pandas-based solution to clean, combine and summarize multiple messy datasets into a single analysis-ready DataFrame, including grouped summaries and a pivot table.	BTL6	Create

UNIT III UNIVARIATE ANALYSIS			
Introduction to Single variable: Distributions and variables – Numerical Summaries of Level and Spread – Scaling and Standardizing – Inequality			
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	How charts are useful for univariate analysis?	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	What is mean by inequality in a distribution?	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	Compare Scaling and standardizing, analyzing when each is appropriate. (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 which measure of central tendency best summarizes a skewed distribution and justify . (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	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	Design a complete univariate analysis plan for chosen variable (Age (Continuous Variable)) specifying the distribution examination. (16)	BTL6	Create
17	Propose a method to summarize and compare the distributions of two Variables measured on different scales, constructing the necessary standardized measures and inequality indicators. (16)	BTL6	Create

UNIT IV BIVARIATE ANALYSIS			
Relationship between Two variables – Percentage Tables – Analyzing Contingency Tables - Handling Several Batches – Scatterplots and Resistant Lines			
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	Competence

		Level	
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	BTL4	Analyze

	constructed, and their significance in data analysis.		
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




