



VIVEKANANDHA COLLEGE OF ENGINEERING FOR WOMEN
 [AUTONOMOUS INSTITUTION AFFILIATED TO ANNA UNIVERSITY, CHENNAI]
 Elayampalayam – 637 205, Tiruchengode, Namakkal Dt., Tamil Nadu.

Question Paper Code: 130016

B.E. / B.Tech. DEGREE END-SEMESTER EXAMINATIONS – NOV. / DEC. 2025

Fifth Semester

Computer Science and Technology

U23CTV13 – EXPLORATORY DATA ANALYTICS

(Regulation 2023)

Time: Three Hours

Maximum: 100 Marks

Answer ALL the questions

Knowledge Levels (KL)	K1 – Remembering	K3 – Applying	K5 - Evaluating
	K2 – Understanding	K4 – Analyzing	K6 - Creating

PART – A

(10 x 2 = 20 Marks)

Q.No.	Questions	Marks	KL	CO
1.	Describe the importance of EDA in the data science life cycle.	2	K1	CO1
2.	Arrange the following steps of EDA in correct order: a) Data Cleaning b) Data Collection c) Data Visualization d) Feature Engineering	2	K1	CO1
3.	Justify how line charts assist in identifying trends over time.	2	K2	CO2
4.	State the primary purpose of using visual aids in Exploratory Data Analysis.	2	K1	CO2
5.	Outline the difference between concatenation and merging operations.	2	K2	CO3
6.	Write the importance of random sampling in unbiased data analysis.	2	K1	CO3
7.	Recall one method used to measure trend in time series analysis.	2	K1	CO4
8.	Demonstrate how mean, median, and mode indicate data concentration.	2	K2	CO4
9.	List the essential steps in data preprocessing.	2	K1	CO5
10.	State two key metrics that determine model performance.	2	K1	CO5

PART – B

(5 x 13 = 65 Marks)

Q.No.	Questions	Marks	KL	CO
11.	a) Describe the key steps involved in EDA. How does each step contribute to better model performance? (OR)	13	K1	CO1
	b) Define classical data analysis and list key differences between exploratory data analysis and Bayesian data analysis.	13	K1	CO1
12.	a) Describe, with an appropriate example, how duplicates and outliers can be identified and removed from a given dataset. (OR)	13	K1	CO2
	b) Explain the concept of Data Refactoring and its importance in EDA.	13	K2	CO2
13.	a) Discuss the steps involved in Data Transformation before visualization with an example. (OR)	13	K2	CO3
	b) Discuss the significance of outlier detection in improving model quality.	13	K2	CO3
14.	a) Explain the different types of kurtosis and their interpretation. (OR)	13	K2	CO4
	b) Describe the concept of univariate, bivariate, and multivariate analysis with examples.	13	K2	CO4
15.	a) Demonstrate the steps of data preparation to a dataset and demonstrate how data cleaning, feature encoding, and scaling are implemented before training a machine-learning model. (OR)	13	K3	CO5
	b) Illustrate how to compare model accuracy scores and select the optimal model programmatically.	13	K3	CO5

PART – C

(1 x 15 = 15 Marks)

Q.No.	Questions	Marks	KL	CO																
16. a)	Given the following dataset summary for monthly sales (in \$ 000) of three products:			CO4																
	<table border="1" style="margin-left: auto; margin-right: auto; border-collapse: collapse;"> <thead> <tr> <th style="padding: 5px;">Month</th> <th style="padding: 5px;">Product A</th> <th style="padding: 5px;">Product B</th> <th style="padding: 5px;">Product C</th> </tr> </thead> <tbody> <tr> <td style="padding: 5px;">Jan</td> <td style="padding: 5px;">50</td> <td style="padding: 5px;">60</td> <td style="padding: 5px;">45</td> </tr> <tr> <td style="padding: 5px;">Feb</td> <td style="padding: 5px;">55</td> <td style="padding: 5px;">65</td> <td style="padding: 5px;">47</td> </tr> <tr> <td style="padding: 5px;">Mar</td> <td style="padding: 5px;">62</td> <td style="padding: 5px;">68</td> <td style="padding: 5px;">49</td> </tr> </tbody> </table>	Month	Product A	Product B	Product C	Jan	50	60	45	Feb	55	65	47	Mar	62	68	49			
Month	Product A	Product B	Product C																	
Jan	50	60	45																	
Feb	55	65	47																	
Mar	62	68	49																	
	i. Apply appropriate measures of central tendency and dispersion to compare sales stability across products.	8	K2																	
	ii. Recommend which product exhibits more consistent sales and justify using the computed statistics.	7	K2																	
	(OR)																			
b)	Consider a dataset of student performance containing numeric features (marks, study hours, attendance) and categorical features (gender, stream).			CO2																
	i. Apply an EDA workflow including data loading, transformation, and visualization using Python (NumPy/Pandas/Seaborn).	8	K2																	
	ii. Interpret the insights gained and suggest data-driven recommendations for improving academic outcomes.	7	K2																	