Reg. No. : E N G G T R E E . C O M

Question Paper Code : 50007

B.E./B.Tech. DEGREE EXAMINATIONS, APRIL/MAY 2024

Third Semester

Artificial Intelligence and Data Science

AD 3301 – DATA EXPLORATION AND VISUALIZATION

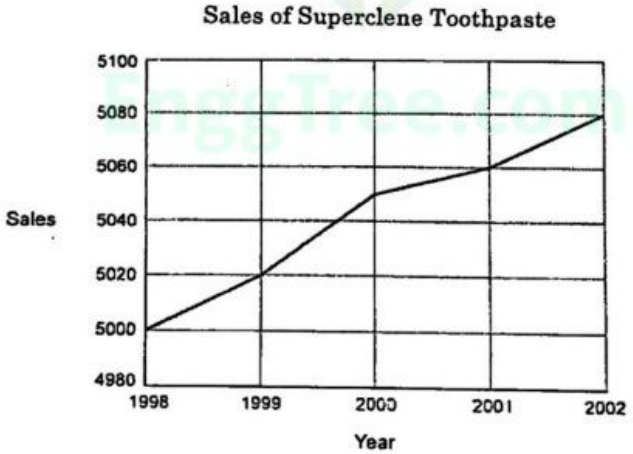
(Regulations 2021)

Time : Three hours Maximum : 100 marks

Answer ALL questions.

PART A — (10 × 2 = 20 marks)

1. Mention the key responsibilities of a data analyst.
2. Name some of the best tools used for data analysis and data visualization.
3. List the software and hardware components required for data visualization.
4. Draw and label a rough contour plot of the joint probability density function. When $P = -0.4$ $\rho = -0.4$.
5. Difference between normalized scaling and standardized scaling.
6. Illustrate important steps to be followed in preparing a base map.
7.



The diagram represents the sales of Superclene toothpaste over the last few years. Give a reason why it is misleading.

8. How do you find the correlation of a scatter plot?

9. Define least square method in time series.

10. List the techniques used in smoothing time series.

PART B — (5 × 13 = 65 marks)

11. (a) (i) Discuss about Descriptive Statistics in exploratory analysis. (7)

(ii) Explain in detail about data transformation Techniques. (6)

Or

(b) (i) Explain in detail about Comparative Statistics in Exploratory analysis. (6)

(ii) Discuss in detail about the practical use of Pivot Table in data science with suitable example. (7)

12. (a) (i) Define line plot. With an example, explain how to create a line plot to visualize the trend. (6)

(ii) The following table gives the lifetime of 400 neon lamps. Draw the histogram for the below data. (7)

Lifetime (in hours)	Number of lamps
300–300	14
400–500	56
500–600	60
600–700	86
700–800	74
800–900	62
900–1000	48

Or

(b) (i) Explain in detail about 3D Data Visualization, its components and its working flow with suitable example. (6)

(ii) Discuss in detail about text and annotation. (7)

13. (a) (i) Does universe frequency distribution have variable? Justify in detail. (7)

(ii) Explain in detail about scaling and standardizing. (6)

Or

(b) Imagine you are working on a time series dataset. Your manager has asked you to build a highly accurate model. You started to build two types of models which are given below. (13)

Model 1 : Decision Tree model

Model 2 : Time series regression model

At the end of evaluation of these two models, you found that model 2 is better than model 1. What could be the possible reason for your inference.

14. (a) (i) Discuss in detail about contingency table with example. (7)

(ii) Explain in detail about percentage table with suitable example. (6)

Or

(b) 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 belong to with neat illustration. (13)

No. of games	3	5	2	6	7	1	2	7	1	7
Scores	80	90	75	80	90	50	65	85	40	100

15. (a) (i) Explain the main components of time series data. Which of these would be most prevalent in data relating to unemployment? (6)

(ii) Suppose, you are a data scientist at Times of India and you observed the views on the articles increases during the month of Jan–Mar. Whereas the views during Nov–Dec decreases. Does the above statement represent seasonality? Justify your answer. (7)

Or

(b) Suppose the following data represent total revenues. (13)

(in millions of constant 1995 dollars) by a car rental agency over the 11 year period 1990 to 2000;

4.0, 5.0, 7.0, 6.0, 8.0, 9.0, 5.0, 2.0, 3.5, 5.5, 6.5

Compute the 5-year moving averages for this annual time series.

PART C — (1 × 15 = 15 marks)

16. (a) Using the following data, calculate the following forecasts : naïve, 3 period moving average, 4-period moving average, 3-2-1 weighted moving average, 1-4-5 weighted moving average, a=.1 exponential smoothing, and a=.8 exponential smoothing. Round all forecasts to whole numbers.

Period	1	2	3	4	5	6	7	8	9	10
Actual	974	766	727	849	693	655	854	742	717	852

Or

(b) Compute the average seasonal movement for the following series and justify your answer :

Year	Quarterly Production			
	I	II	III	IV
2002	3.5	3.8	3.7	3.5
2003	3.6	4.2	3.4	4.1
2004	3.4	3.9	3.7	4.2
2005	4.2	4.5	3.8	4.4
2006	3.9	4.4	4.2	4.6