

Total Marks 80

(3 Hours)

NB

- 1) Question **number 1** is compulsory
- 2) Attempt **any three** out of the remaining **five questions**.
- 3) Assume suitable data if **necessary** and justify the assumptions.
- 4) Figures to the **right** indicate full marks

- Q1 **Attempt any four** 20
- a) Explain in brief the objectives of Data Exploration
 - b) Explain in brief the taxonomy of time series forecasting
 - c) What are the outliers in the dataset? State the reasons for the outliers occurring in the dataset
 - d) Explain validation techniques bootstrap and cross-validation
 - e) State the importance of Data Visualization. State the purpose of scatter plots, quartile plots, bubble charts, density chart
- Q2 a) Given data of 10 companies. Find out the type of correlation between advertisement expenses and sales volume using Karl Pearson's coefficient of correlation method 10

Company	1	2	3	4	5	6	7	8	9	10
Advt expenses	11	13	14	16	16	15	15	14	13	13
Sales volume	50	50	55	60	65	65	65	60	60	50

- b) Explain the data science process in detail 10
- Q3 a) Explain the density-based outlier detection approach 10
- b) Explain SMOTE in detail 10
- Q4 a) Explain the working of the Auto Regressive Integrated Moving Average Model 10
- b) The data given shows salary packages (in lakhs) offered after a campus interview. Find the coefficient of skewness using Bowley's method. 10

Salary	4-8	8-12	12-16	16-20	20-24
No of Candidates	4	10	15	8	3

- Q5 a) What are the attributes of time series decomposition? Explain the classical decomposition technique 10
- b) In certain food experiment to compare two types of baby foods A and B, the following results of the increase in weight (lbs) we observed in 8 children as follows 10

Food A	49	53	51	52	47	50	52	53
Food B	52	55	52	53	50	54	54	53

Examine the significance of the increase in weight of children due to food B. (Given t-value at $\alpha=0.05$ is 2.365)

- Q6 a) Explain how the time-series approach is used to forecast the demand for a product. 10
- b) Explain how predictive modelling can be applied to the House price prediction recommendation 10