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.

Q1 Answer the following 20

a) What is the difference between data science and data analytics?

b) What are Type I and Type –II errors? Give examples.

c) Brief about SMOTE.

d) What do you mean by Time Series Decomposition?

Q2 a) Describe the terms: cross-validation, K-fold cross-validation, leave-1 out and Bootstrapping. 10

b) Explain the data science process in detail. 10

Q3 a) What are outliers? Explain different outlier detection methods. 10

b) Calculate the coefficient of correlation for the following data with Karl Pearson's method. 10

X	10	20	30	40	50	60	70	80	90	100
Y	2	4	8	5	10	15	14	20	22	50

Q4 a) Find Bowley's coefficient of skewness of the following series. 10

Size	4	4.5	5	5.5	6	6.5	7	7.5	8
F	10	18	22	25	40	15	10	8	7

b) Explain the Auto Regressive Integrated Moving Average (ARIMA) model in detail. 10

Q5 a) Brief about ANOVA and its types. How it is different from a t-test? 10

b) What is Hypothesis testing? Explain the steps involved in Hypothesis testing with an example. 10

Q6 **Write a note on any TWO :** 20

i. Data Visualization techniques
ii. Univariate Exploration and Multivariate Exploration
iii. House price Prediction or Fraud Detection

****