

(3 Hours)

Total Marks: 80

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 A What is the difference between data science and data analytics? 10

B What is ANOVA? Brief about benefits of ANOVA technique. 10

Q2 A What are Type I and Type -II errors? Give examples. 10

B Explain the data science tasks with proper examples. 10

Q3 A Describe the terms: Cross Validation, K-fold cross validation, leave-1 out and Bootstrapping. 10

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

X	15	18	20	28	34
Y	40	42	46	50	52

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

Profit (in crores)	4-8	8-12	12-16	16-20	20-24
No. of films	4	10	15	8	3

B What are the pros and cons of an Auto Regressive Integrated Moving Average (ARIMA) model? Explain with proper examples. 10

Q5 A Explain the steps to build a product recommendation model in detail. 10

B What is Hypothesis testing? Write about the different types of Hypothesis testing. 10

Q6 Write a note on any FOUR : 20

- A. Data Visualization
- B. Applications of Data Science
- C. Data Exploration
- D. Taxonomy of time series forecasting methods
- E. Outlier detection methods