Code No.: R22AI504PC

R22	H.T.No.			8	R					

CMR ENGINEERING COLLEGE: : HYDERABAD
UGC AUTONOMOUS
III–B.TECH–I–Semester End Examinations (Regular) - December- 2024
DATA ANALYTICS USING R
(CSM)

[Time: 3 Hours] **[Max. Marks: 60]**

Note: This question paper contains two parts A and B.
Part A is compulsory which carries 10 marks. Answer all questions in Part A.
Part B consists of 5 Units. Answer any one full question from each unit. Each question carries 10 marks.

PART-A **(10 Marks)**

1. a) Explain how sensors generate data. [1M]
   b) List two sources of data commonly used in analytics. [1M]
   c) Define data analytics and Name two common tools used in the analytics environment. [1M]
   d) What is the significance of data modeling in analytics? [1M]
   e) Write an R code snippet to fit a simple linear regression model using the lm() function. [1M]
   f) Compare the ordinary least squares method with robust regression techniques. [1M]
   g) Compare RMSE and MAE. [1M]
   h) Identify a real-world scenario where object segmentation can be applied in image processing. [1M]
   i) Why Data Visualization is required? [1M]
   j) What is a scatter plot? [1M]

PART-B **(50 Marks)**

2. Define an outlier and explain how to identify and treat them with an example. [10M]

OR

3.a) Describe the difference between data preprocessing and data processing with examples. [5M]
   b) Explain the characteristics of good quality data. [5M]

4.a) In detail explain various applications of data analytics in various business domains [5M]
   b) Fill the missing values using Mean/Median/Mode Imputation. [5M]

ID	Age	Salary	Dept.
1	25	50000	Sales
2	30	--	HR
3	--	45000	Sales
4	28	48000	HR

OR

5. Create a relational model to manage an e-commerce system. The system tracks customers, products, and orders? [10M]

6. Explain the difference between underfitting and overfitting? Linear Regression and logistic regression models. [10M]

OR

7.a) Explain the concept of the "Blue Property" (Best Linear Unbiased Estimator) with reference to regression models. [5M]

b) Compare and contrast the use of R-squared in linear regression versus logistic regression. [5M]

8.a) Compare supervised and unsupervised learning. [4M]

b) Explain STL approach with an example. [6M]

OR

9. Explain the ARIMA model and its approach with an example using R. [10M]

10. Explain about the Hierarchical Visualization Techniques. [10M]

OR

11. Explore the Visualizing the Complex Data and Relations. [10M]

****