

Reg. No. :

--	--	--	--	--	--	--	--	--	--	--	--	--

Question Paper Code:R2E04

B.E./B.Tech. DEGREE EXAMINATION, APRIL / MAY 2025

Second Semester

Artificial Intelligence and Data Science

R21UAD204- FOUNDATIONS OF DATA SCIENCE

(Regulations R2021)

Duration: Three hours

Maximum: 100 Marks

Answer ALL Questions

PART A - (10 x 2 = 20 Marks)

1. What is Data science? CO1- U
2. What is meant by Research Goal? CO1- U
3. List out the Data types of R Programming. CO1- U
4. Create a data frame with two columns: "Name" and "age" with at least three entries. Then display only the rows where age is greater than 35. CO2-App
5. Name two common types of clustering algorithms. CO1- U
6. What is Naive Bayes theorem? CO1- U
7. What does a correlation coefficient of +1 indicate? CO1- U
8. What is the difference between positive correlation and negative correlation? CO1- U
9. Define a box plot and mention its key components. CO1- U
10. How can you improve the readability of a data visualization? CO1- U

PART – B (5 x 16= 80 Marks)

11. (a) Briefly Explain about Data Science Process. CO1- U (16)
Or
(b) Write an R interactive data visualization dashboard using R (Shiny) or Python (Dash). The dashboard should allow users to filter data, change parameters, and visualize key insights dynamically. Explain the significance of interactivity in data-driven decision-making. CO1- U (16)

12. (a) Using R, apply control structures (e.g., if-else, for loops, and while loops) to solve a data manipulation problem. Write a script to automate a process such as filtering out certain data points or iterating through multiple datasets to calculate summary statistics. CO2- App (16)
- Or
- (b) Using R, apply control structures (e.g., if-else, for loops, and while loops) to solve a data manipulation problem. Write a script to automate a process such as filtering out certain data points or iterating through multiple datasets to calculate summary statistics. CO2- App (16)
13. (a) Briefly Explain the concept of clustering in data science. CO1- U (16)
- Or
- (b) Discuss the challenges of clustering high-dimensional data. Explain the concept of the "curse of dimensionality" and the techniques (e.g., dimensionality reduction) used to address these challenges. CO1- U (16)
14. (a) Using R, Write a program how to analyze the relationship between two continuous variables. Use appropriate visualizations (e.g., scatter plot) and statistical methods (e.g., correlation coefficient) to describe the strength and direction of the relationship. CO3- Ana (16)
- Or
- (b) Analyze a dataset using R to find relationships between time-dependent data points (e.g., stock prices or temperature data). Use time-series analysis techniques and visualizations (like line charts or moving averages) to explore trends, seasonality, and cyclic patterns. CO3- Ana (16)
15. (a) Explain the importance of data visualization in data science. Discuss different types of visualizations with examples. CO1- U (16)
- Or
- (b) Explain histograms and box plots in detail. Compare their uses in understanding data distribution. CO1- U (16)