

Reg. No. :

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

Question Paper Code : 90044

B.E./B.Tech. DEGREE EXAMINATIONS, NOVEMBER/DECEMBER 2022.

Fifth Semester

Artificial Intelligence and Data Science

AD 8502 – DATA EXPLORATION AND VISUALIZATION

(Regulations 2017)

Time : Three hours

Maximum : 100 marks

Answer ALL questions.

PART A — (10 × 2 = 20 marks)

1. Write the similarities and differences between R and Python.
2. Define Gaussian distribution
3. What are the alternative measures of Inequality?
4. List down the goals of transformation.
5. Define preprocessing.
6. List some of the popular techniques of Data visualization.
7. What is Tidy Data?
8. Why are `pivot_longer()` and `pivot_wider()` not perfectly symmetrical?
9. Both `unite()` and `separate()` have a `remove` argument. What does it do? Why would you set it to `FALSE`?
10. Give the function that is used to measure the covariation between variables using different symbol sizes.

PART B — (5 × 13 = 65 marks)

11. (a) Exemplify in detail to produce bar charts, pie charts and histogram using SPSS. (13)

Or

- (b) (i) Explain in detail about smoothing time series techniques. (8)
(ii) Write short notes on Gini coefficient under inequality. (5)
12. (a) Explain in detail about logistic regression models and interpret the coefficients in logistic regression models under three variable contingency table. (13)

Or

- (b) Write short notes on the following (13)
(i) Lone parents in scatterplots and resistant line
(ii) Linear relationships in scatterplots and resistant line
(iii) Log Transformation
13. (a) Explain in detail about the stages of Data Visualization with a neat diagram. (13)

Or

- (b) (i) What is correlation? What are the types of correlation? How scatter plot is used for this? (7)
(ii) What is the significance of time series data? Give some application that are suitable for generating time series data. (6)
14. (a) What is a Tidy Data? Explain in detail about the steps involved in cleaning the data. (13)

Or

- (b) (i) How to load a tab delimited file with read table? (4)
(ii) How to load data in R programme. Write a program in R to load and visualize the data using any three charts. (9)
15. (a) Discuss in detail how to add Code chunks and Text to an R Markdown File. (13)

Or

- (b) (i) How geographic data are visualized using gmap? Explain. (7)
(ii) How to create a map in Rstudio? Explain. (6)

PART C — (1 × 15 = 15 marks)

16. (a) (i) What is scaling? How is it useful in Machine learning domain? List the various types of scaling with example. (10)

- (ii) Create a dataset with at least 3 features contributing to the calculation of rental value. Create minimum 10 rows of data. List down the challenges, in acquiring this data from real time. (5)

Or

- (b) (i) How decision trees are formed? How will you visualize the tree? List the advantages and limitations of tree data structure. (8)
(ii) Write a recursive program in R to print the n_c value given the values of n and r . (7)