Reg. No. : | E | N | G | G | T | R | E | E | . | C | O | M |

Question Paper Code : 70004

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

Third Semester

Artificial Intelligence and Data Science

AD 3301 – DATA EXPLORATION AND VISUALIZATION

(Regulations 2021)

Time : Three hours | Maximum : 100 marks

Answer ALL questions.

PART A — (10 × 2 = 20 marks)

1. What is meant by EDA?
2. How do you get cross tabulation?
3. What is the difference between MATLAB and matplotlib?
4. Is a histogram always a bar chart? Justify with your answer.
5. What is the main purpose of univariate analysis?
6. What is the mathematical mean of the following numbers? 10, 6,4, 4, 6, 4.
7. What are the three common methods for performing bivariate analysis?
8. Outline the difference between univariate and bivariate data.
9. Show the characteristics of multivariate analysis.
10. What is TSA in Statsmodel?

PART B — (5 × 13 = 65 marks)

11. (a) What is the primary purpose of EDA? What are the differences between EDA with classical and Bayesian analysis? Discuss it in detail.

Or

(b) Explain various transformation techniques in EDA.

12. (a) How to over plot a line on a scatter plot in Python? Illustrate with code.

Or

(b) Discuss with how Seaborn helps to visualize the statistical relationships. Illustrate with code and example.

13. (a) Explain the 10 Essential Numerical Summaries in Statistics with example.

Or

(b) How, When, and Why Should You Normalize / Standardize / Rescale Your Data?

14. (a) What is a table of frequency values for a bivariate distribution? Explain What graph is used in the analysis of bivariate data?

Or

(b) How do you analyze a contingency table? Discuss.

15. (a) What is meant by time series data? Describe its four components.

Or

(b) What is the best way to visualize time series data? What patterns might appear when you plot the time series data?

PART C — (1 × 15 = 15 marks)

16. (a) What are the tools used for EDA? Give a case study on applying EDA in a real business scenario.

Or

(b) Discuss in detail about Data Cleaning (missing data, outliers detection and treatment).

08/02/2024 - FN.

Reg. No. : | | | | | | | | | | | | |

Question Paper Code : 20011

B.E./B.Tech. DEGREE EXAMINATIONS, NOVEMBER/DECEMBER 2023.

Third Semester

Artificial Intelligence and Data Science

AD 3301 – DATA EXPLORATION AND VISUALIZATION

(Regulations 2021)

Time : Three hours Maximum : 100 marks

Answer ALL questions.

PART A — (10 × 2 = 20 marks)

1. What does EDA mean in data?

2. State the purpose of data aggregation.

3. What is Matplot lib used for?

4. Differentiate plot vs subplot.

5. List the three main types of univariate analyses.

6. What is the purpose of smoothing a time series data?

7. Name the two main types of statistical testing in bivariate analysis.

8. Is bivariate qualitative or quantitative?

9. What is multivariate analysis?

10. What are the two common techniques used to perform dimension reduction?

PART B — (5 × 13 = 65 marks)

11. (a) Provide an explanation of the various EDA tools that are used for data analysis.

Or

(b) What is cross-tabulation and PivotTable? How to Build Pivot Table and Cross Tab Reports?

12. (a) Why matplotlib is used for data visualization? Which module of matpotlib is used for data visualization?

Or

(b) How do you Visualize a Three-Dimensional Function in python? Illustrate with a code.

13. (a) What is scaling and standardization? When and why to standardize a variable? Illustrate with suitable example.

Or

(b) Explain the Smoothing Techniques for time series data with suitable example.

14. (a) How do you analyze contingency tables? Give examples.

Or

(b) Discuss the best Practices for Designing Scatter Plots.

15. (a) What are the characteristics of multivariate analysis? How do you explain multivariate analysis?

Or

(b) What is TSA analysis? Explain ARIMA, smooth-based and moving average.

PART C — ($1 \times 15 = 15$ marks)

16. (a) Give a case study on univariate and multivariate analysis with example.

Or

(b) Describe the various distributions module of Seaborn for visualization. Consider a sample application to illustrate.
