

Reg. No. :

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

Question Paper Code : 11166

B.E./B.Tech. DEGREE EXAMINATIONS, APRIL/MAY 2026.

Third Semester

Computer Science and Engineering

CS3352 — FOUNDATIONS OF DATA SCIENCE

(Common to Computer Science and Engineering (Artificial Intelligence and Machine Learning)/Computer Science and Engineering (Cyber Security)/Computer Science and Engineering (Data Science)/Computer and Communication Engineering/Electrical and Computer Engineering/Information Technology)

(Regulations 2021)

Time : Three hours

Maximum : 100 marks

Answer ALL questions.

PART A — ($10 \times 2 = 20$ marks)

1. What is the purpose of data warehousing in the context of data science?
2. Mention any two basic statistical descriptions of data.
3. Differentiate between qualitative and quantitative data.
4. What is a z-score and what does it indicate?
5. What does the correlation coefficient indicate in quantitative data?
6. Define regression towards the mean.
7. What is a structured array in NumPy, and how is it different from a regular array?
8. How Pandas handle missing data?
9. What is `sns.histplot()` in Seaborn and how it differs from a standard Matplotlib histogram?
10. What is the purpose of the `legend()` function in Matplotlib, and how is it used?

PART B — ($5 \times 13 = 65$ marks)

11. (a) Explain the data science process in detail, covering each stage from defining research goals to building applications with neat flow diagram.

Or

- (b) Discuss the benefits and uses of data science. Also explain the different facets of data and the role of data mining.

12. (a) Explain the different types of variables and how data can be described using tables and graphs.

Or

- (b) Discuss how data is described using averages and measures of variability. Explain the concept of normal distribution with the help of z-scores.
13. (a) Explain the concept of correlation and regression. Describe how scatter plots and correlation coefficients are used to analyze relationships between variables.

Or

- (b) Discuss regression analysis in detail. Explain the least squares regression line, standard error of estimate, r^2 , and multiple regression equations.
14. (a) Discuss the basics of NumPy arrays and their key functionalities such as aggregations, computations, comparisons, and fancy indexing with suitable examples.

Or

- (b) Explain how data manipulation works in Pandas. Discuss indexing and selection, operating on data, handling missing data, and hierarchical indexing.
15. (a) Discuss the different types of plots available in Matplotlib and how they are used for visualizing data. Explain line plots, scatter plots, and density plots with examples.

Or

- (b) Discuss three-dimensional plotting, geographic data visualization with Basemap, and how Seaborn enhances visualization.

PART C — (1 × 15 = 15 marks)

16. (a) Calculate the correlation coefficient for the following data :

Subject	Age	Glucose level Y
1	42	98
2	23	68
3	22	73
4	47	79
5	50	88
6	60	82

Or

- (b) Consider a small dataset of 10 students marks in Maths, Physics, Chemistry, English and Computer Science. Write Pandas functions to perform following tasks.
- (i) sum and aggregation (5)
 - (ii) group using split-apply-combine process (5)
 - (iii) group by one and multi columns. (5)