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

Question Paper Code : 20865

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

Third/Fifth Semester

Computer Science and Engineering

CS 3352 – FOUNDATIONS OF DATA SCIENCE

(Common to : Computer Science and Engineering (Artificial Intelligence and Machine Learning / Computer Science and Engineering (Cyber Security) / Computer and Communication Engineering / Electronics and Instrumentation Engineering / Instrumentation and Control Engineering and Information Technology)

(Regulations 2021)

Time : Three hours Maximum : 100 marks

Answer ALL questions.

PART A — (10 × 2 = 20 marks)

1. What is Structured data?

2. Give an overview of common errors?

3. Explain the types of data.

4. Define median with example.

5. Define multiple regressions.

6. Define regression towards the mean.

7. What are the key properties of Pearson correlation coefficient?

8. Summarize some built – in Pandas aggregations?

9. Explain partial sort.

10. Give a summary about the comparison operators.

PART B — (5 × 13 = 65 marks)

11. (a) Explain the different facets of data with example.

Or

(b) Explain in detail about the cleansing, integrating, transforming data and build a model.

12. (a) (i) Explain Normal curve and Z-score. (6)

(ii) Using standard normal curve table, find the proportion of the total area identified with the following statements. (7)

(1) above a z score of 1.80

(2) between the mean and a z score of 1.65

(3) between z scores of 0 and –1.96

Or

(b) (i) Describe the types of variable. (6)

(ii) Suppose a hospital tested the age and body fat data for randomly selected adults with the following result :

Age	23	27	39	49	50	52	54	56	57	58	60
%fat	9.5	17.8	31.4	27.2	31.2	34.6	42.5	33.4	30.2	34.1	41

Draw the boxplots for age. (7)

13. (a) (i) Explain scatter plot. (6)

(ii) Describe range and variance. (7)

Or

(b) (i) Explain the correlation coefficient. (6)

(ii) Explain how the least squares equation which is used to minimize the total of all squared prediction errors with example. (7)

14. (a) Explain grouping in python with example.

Or

(b) Explain the following in python

(i) Data indexing (6)

(ii) Operation on missing data (7)

15. (a) Explain the different types of joins in python.

Or

(b) Explain various features of Matplotlib platform used for data visualization and illustrate its challenges.

PART C — (1 × 15 = 15 marks)

16. (a) Describe in detail about pivot table.

Or

(b) Find the following for the given data set :

Mean, Median, Mode, Variance, Standard Deviation and Skewness.

Marks	0–10	10–20	20–30	30–40	40–50	50–60	60–70	70–80
No. of Students	10	40	20	0	10	40	16	14