PRESIDENCY UNIVERSITY
BENGALURU

School Of Computer Science and Engineering & Information Science

End-Term Examinations, Aug 2024

Odd Semester: 2023 - 24

Course Code: CSE2027

Course Name: Fundamentals of Data Analytics

Department: B.Tech. Computer Science and Engineering

Date: 19-08-2024

Time:9:30 am – 12:30 pm

Max Marks: 100

Weightage: 50%

Instructions:
(i) Read the all questions carefully and answer accordingly.
(ii) Do not write any matter on the question paper other than roll number.

Q. No	Questions	Marks	CO	RBT
1	a. Define the term "data analytics." What are the primary goals of data analytics?	4	CO1	L1
	b. Compare difference between structured and unstructured data types with example.	6	CO1	L2
	c. Big data is often characterized by the 10 V's. Imagine you are working for a multinational retail company that wants to leverage big data to improve its operations. Describe how each of these 10 V's could be applied or considered in the context of managing and utilizing big data within the company. Use specific examples to illustrate the potential applications and challenges associated with each V.	10	CO1	L3

OR

Q. No	Questions	Marks	CO	RBT
2	a. Explain data analysis-types.	4	CO1	L1
	b. Describe central tendency with suitable example for each.	6	CO1	L2
	c. A company is analyzing customer data and wants to normalize the income data using min-max normalization. The data contains the following customer income values : 15,25,35, 45, 55,65,75, 85,95,105. Normalize these income values to a range of [0, 1] and explain the steps involved in the min-max normalization process.	10	CO1	L3

Q. No	Questions	Marks	CO	RBT
3	a. What are descriptive and inferential statistics? Provide suitable examples to illustrate each type.	4	CO2	L1
	b. Describe various sampling techniques, providing appropriate examples for each type.	6	CO2	L2
	c. Suppose we want to determine if the average score of girls in an exam exceeds 600, but we do not have information about the variance or standard deviation of the scores. Describe how you would perform a t-test in this scenario, using a	10	CO2	L3

random sample of 10 girls' scores and choosing a significance level (α) of 0.05 for hypothesis testing.
Sample Data (Girl Scores): 587, 602, 627, 610, 619, 622, 605, 608, 596, 592

OR

4	a. What is contingency table and how it is constructed?	4	CO2	L1
	b. Explain the coefficient of variance (CV) and its significance in data analysis Given the following datasets, calculate the CV for each and interpret the results: Dataset A: [100, 150, 200, 250, 300]	6	CO2	L2
	c. Calculate the mean, variance, and standard deviation for the following dataset, which represents the monthly expenses (in dollars) of a household over a year: 48.50, 87.40, 19.98, 59.74, 40.87, 105.51, 40.80, 23.10, 98.10, 60.54, 64.54, 48.01.Analyze these measures to understand the household's spending patterns.	10	CO2	L3

5	a. Explain the methods of primary data collection.	4	CO3	L1
	b. Discuss different types of observational methods .	6	CO3	L2
	c. Demonstrate various interview methods, and discuss their respective advantages and disadvantages.	10	CO3	L3

OR

6	a. List out the main difference between Survey and experiment.	4	CO3	L1
	b. Discuss the main aspects of questionnaire , schedule and enumerators.	6	CO3	L2
	c. Explain the concept of correlation coefficients, What are the different types of correlation and with suitable example describe how they can be used to measure relationship between two variables.	10	CO3	L3

7	a. Explain how can tables be used to organize data interactively, and how do they help with data organization and retrieval?	4	CO4	L1
	b. How can you effectively create dashboards that are presentation-ready and turn real-world data into actionable business insights?	6	CO4	L2
	c. What are some examples of built-in charts in Excel and how can they be used effectively to represent different types of data?	10	CO4	L3

OR

8	a. What are some key factors to consider when making forecasts for market trends in a volatile industry?	4	CO4	L1
	b. How can Pivot Tables be utilized for data analysis in Excel? Discuss the process of setting up a Pivot Table, the types of data insights that can be gained?	6	CO4	L2
	c. Discuss the importance of interpretation and report writing in data analysis. Explain the key components that should be included in a well-structured	10	CO4	L3

	report, including how to present data findings, conclusions, and recommendations effectively. Provide an example of how data interpretation can influence decision-making in a business context.			

9	a. What is logistic regression and explain the role of the sigmoid function in logistic regression?	4	CO5	L1
	b. Discuss the concept of classification in machine learning, and provide an example to illustrate how it is used.	6	CO5	L2
	c. Find the straight line that best fit the following data using OLS method as well as to calculate MSE. X=5,10,15,20,25 Y=16,19,23,26,30	10	CO5	L3

OR

10	a. List out the difference between linear and non linear regression.	4	CO5	L1
	b. What is a confusion matrix, and how are performance metrics derived from it used to evaluate a model's effectiveness?	6	CO5	L2
	c. Apply your knowledge of selecting prediction models. Imagine you are a data scientist responsible for predicting patient readmission rates in a hospital. Explain how you would choose the most appropriate prediction model for this task, considering the techniques and criteria you would use for evaluation.	10	CO5	L3