ODD SEMESTER EXAMINATION, 2023 – 24
IInd Year B.Tech. – CS&E/EE
INTRODUCTION TO DATA SCIENCE

Duration: 3:00 Hrs.   **Max Marks: 100**

Note: - Attempt all questions. All Questions carry equal marks. In case of any ambiguity or missing data, the same may be assumed and state the assumption made in the answer.

Q 1.	Answer any four parts of the following.	5x4=20
	a) What is Data Science, and why is it important in today's digital age?	
	b) Explain the importance of Exploratory Data Analysis (EDA) in the Data Analytics Process and how it aids in decision-making.	
	c) Differentiate between Feature Generation and Feature Selection in the context of data analysis, and explain their roles in enhancing model performance.	
	d) Describe data visualization. Write two tools commonly used in data visualization and share a brief idea for creating impactful visualizations using these tools.	
	e) What skills and qualities do you think will be crucial for the next generation of Data Scientists, considering the evolving landscape of technology and data?	
	f) Define the term 'Data Analytics' and provide two examples of how organizations can benefit from employing data analytics.	
Q 2.	Answer any four parts of the following.	5x4=20
	a) Name various sectors where Data Science is extensively applied, and explain any three with example.	
	b) Compare and contrast two graphical techniques commonly used in Exploratory Data Analysis (EDA), highlighting their respective strengths and applications.	
	c) Explain two Feature Selection algorithms, and provide scenarios where each algorithm might be particularly useful.	
	d) How can storytelling be integrated into data visualization, and what tools support this narrative-driven approach?	
	e) Explain evolution of Data Science, highlighting key milestones and advancements that have shaped the field over the years.	
	f) How does the Matplotlib library contribute to Data Science in Python. Give an example of its application?	
Q 3.	Answer any two parts of the following.	10x2= 20
	a) What are the key steps involved in the Data Analytics Process? Briefly explain each step?	
	b) Discuss how both Quantitative Techniques and Graphical Techniques play integral roles in the EDA process. Provide examples of each technique and explain how they contribute to uncovering patterns, trends, and insights in a given dataset.	

	c) Discuss the critical considerations and challenges related to privacy and security in the field of data science. Provide examples and recommendations on how data scientists can address these concerns to ensure responsible and ethical data handling practices.	
Q 4.	Answer any two parts of the following. a) Describe the steps you would take to create a visualization for a complex dataset, and explain the considerations for choosing the appropriate visualization type. b) Demonstrate any five real-world examples of applications of Data Science and explain how they have positively impacted their respective industries. c) What is the purpose of using measures of central tendency in Exploratory Data Analysis (EDA)? Provide examples of three common measures of central tendency. Create a python program to find mean, median and mode using statistics module for given data. X= [4, 8, 6, 5, 3, 2, 8, 9, 2, 5].	10x2= 20
Q 5.	Answer any two parts of the following. a) Discuss the ethical considerations related to data privacy and security in the context of Data Science. Propose measures that organizations can adopt to address these ethical concerns. b) Explain the various data science terminology: - i. Data Analyst ii. Data Governance iii. Data Wrangling iv. Artificial Intelligence v. Business Intelligence (BI) c) What are the fundamental principles of effective data visualization, and why are they important in conveying information? Create a python program to visualize data through histogram.	10x2= 20

****