Total No. of Questions: 6 *Total No. of Printed Pages:3*

Enrollment No.....

Faculty of Engineering
End Sem Examination Dec-2023
CS3EL13 Data Science

Programme: B.Tech. Branch/Specialisation: CSE All

Duration: 3 Hrs. **Maximum Marks: 60**

Note: All questions are compulsory. Internal choices, if any, are indicated. Answers of Q.1 (MCQs) should be written in full instead of only a, b, c or d. Assume suitable data if necessary. Notations and symbols have their usual meaning.

Q.1 i. What is the primary focus of data science? **1**

(a) Managing hardware and software
(b) Extracting insights from data
(c) Building websites
(d) Conducting market research

ii. Why is data science important in today's business world? **1**

(a) It increases hardware sales.
(b) It helps reduce data storage costs.
(c) It enables data-driven decision-making.
(d) It simplifies project management.

iii. What is a sample space in probability theory? **1**

(a) The set of all possible outcomes of an experiment
(b) A specific outcome of an experiment
(c) The likelihood of an event occurring
(d) The set of events that are mutually exclusive

iv. Which of the following distributions is commonly used to model the heights of people in a population? **1**

(a) Uniform distribution (b) Exponential distribution
(c) Normal distribution (d) Poisson distribution

v. In the Data Science Process, what comes after exploratory data analysis if you are following a structured approach? **1**

(a) Model building (b) Data collection
(c) Data preprocessing (d) Data cleaning

vi. What is the primary goal of Exploratory Data Analysis (EDA)? **1**
(a) To make predictions about the future
(b) To visualize and summarize data to understand its characteristics
(c) To build complex machine learning models
(d) To perform statistical inference

vii. What are the basic principles of data visualization? **1**
(a) Making visualizations as complex as possible
(b) Using as many colors as possible
(c) Maximizing data ink ratio
(d) Avoiding the use of labels and legends

viii. What is the primary purpose of data visualization in data science? **1**
(a) To complicate data analysis
(b) To make data look aesthetically pleasing
(c) To convey complex data in an understandable and insightful manner
(d) To replace statistical analysis

ix. In the context of data science, why is Python a popular choice as a programming language? **1**
(a) Python is known for its slow execution speed.
(b) Python has limited libraries for data manipulation.
(c) Python is user-friendly and has a wide range of libraries for data science tasks.
(d) Python is primarily used for web development.

x. Which Python library is commonly used for scientific computing and data analysis tasks, including linear algebra, optimization, and statistics? **1**
(a) Matplotlib (b) PyBrain (c) SciPy (d) Pylearn

Q.2 i. Define data science. **2**

ii. Explain the historical development of data science and highlight key milestones in its evolution. **3**

iii. Describe the primary components of data science and their respective roles in the data analysis process. **5**

OR iv. Provide an overview of different data science techniques and their practical applications. **5**

Q.3 i. Define the sample space and an event in the context of probability theory. **3**

ii. Differentiate between discrete and continuous random variables, providing examples of each. **7**

OR iii. Explain the concepts of descriptive, predictive, and prescriptive statistics, and provide examples of each. **7**

Q.4 i. What are some basic tools used in EDA? How do they aid in data exploration? **4**

ii. Describe the Data Science Process, highlighting the role of EDA within this process. **6**

OR iii. Why is it important to document and communicate the results of EDA in a case study or data analysis project? **6**

Q.5 i. What are the fundamental principles of data visualization? Why are they important in the field of data science? **4**

ii. Why is it important to choose appropriate visual encoding techniques when creating data visualizations? **6**

OR iii. Create own datasets as an exercise and create visualization using box plot, scatter plot and histogram. **6**

Q.6 Attempt any two:

i. How does Matplotlib contribute to data science? What types of visualizations can be created using this library? **5**

ii. Explain the primary advantage of NoSQL databases over traditional relational databases. **5**

iii. Explain the significance of Python as a data science tool. **5**

******

Marking Scheme
Data Science-CS3EL13(T)

Q.1	i)	b) Extracting insights from data		**1**
	ii)	c) It enables data-driven decision-making.		**1**
	iii)	a) The set of all possible outcomes of an experiment		**1**
	iv)	c) Normal distribution		**1**
	v)	a) Model building		**1**
	vi)	b) To visualize and summarize data to understand its characteristics		**1**
	vii)	c) Maximizing data ink ratio		**1**
	viii)	c) To convey complex data in an understandable and insightful manner		**1**
	ix)	c) Python is user-friendly and has a wide range of libraries for data science tasks.		**1**
	x)	c) SciPy		**1**
Q.2	i.	Explanation	(As per explanation)	
	ii.	Components	2 Marks	
		Roles	3 Marks	
	iii.	The primary components of Data Science	3 Marks	**3&2**
		Their respective ……process.	2 Marks	
OR	iv.	Provide an …… techniques	3 Marks	**3&2**
		Their practical applications	2 Marks	
Q.3	i.	Sample ……. theory.	(1 Mark*3)	
	ii.	Discrete	3.5 Marks	
		Continuous random variables	3.5 Marks	
OR	iii.	The concepts of descriptive…. of each.	(Any Three)	
Q.4	i.	What are some basic tools used in EDA,	2 Marks	**2&2**
		How do they aid in data exploration	2 Marks	
	ii.	Describe the Data Science Process	3 Marks	**3&3**
		Highlighting the role of EDA within this process.	3 Marks	
OR	iii.	Document….. ……..project (Explanation)	6 Marks	
Q.5	i.	Principle	2 Marks	
		Importance	2 Marks	
	ii.	Each techniques	1 Mark	
OR	iii.	Create own datasets as an exercise	2 Marks	**2,2,2**
		Create visualization using box plot, scatter plot	2 Marks	
		Histogram.	2 Marks	
Q.6				
	i.	Matplotlib contribute to data science	2 Marks	
		Types ……..library	3 Marks	
	ii.	Primary …… ……databases.	(1 Mark*5)	
	iii.	The significance of Python …… tool.	(1 Mark*5)	

******