SRINIVAS UNIVERSITY

INSTITUTE OF ENGINEERING AND TECHNOLOGY

FIRST SEMESTER QUESTION BANK

COURSE CODE: BTA001 COURSE NAME: INTRODUCTION TO DATA SCIENCE CREDITS:04

Q. No	Objective Type of Questions (Single sentence answer questions, Multiple choice or fill in the blanks type) There should be more than 50 questions for 5 Module course and 50 questions for 5 Module courses (Compulsory 10 questions from each module)	Module or Unit No.	Mark for each question (1mark)	CO- can be any one of the below CO1,CO2, CO3, CO4
MODULE-1				
1	Which of the following best defines Data Science? A. Study of computer hardware B. Extraction of knowledge and insights from structured and unstructured data C. Only data visualization techniques D. Database management system Answer: B	1	1	CO1
2	Which discipline is NOT a core contributor to Data Science? A. Statistics	1	1	CO1

	B. Machine Learning C. Philosophy D. Computer Science **Answer:** C			
3	**The term “Data Science” was popularized during which decade?** A. 1970s B. 1980s C. 1990s D. 2000s **Answer:** D	**1**	**1**	**CO1**
4	**Which component of Data Science focuses on discovering hidden patterns in data?** A. Data Collection B. Data Storage C. Data Analysis D. Data Transmission **Answer:** C	**1**	**1**	**CO1**
5	**Which of the following is a key difference between Data Science and Business Analytics?** A. Data Science focuses only on past data B. Business Analytics does not use data C. Data Science emphasizes predictive and prescriptive modeling	**1**	**1**	**CO1**

	D. Business Analytics ignores visualization **Answer:** C			
6	**Which term refers to data that does not follow a predefined schema?** A. Structured data B. Semi-structured data C. Unstructured data D. Normalized data **Answer:** C	**1**	**1**	**CO1**
7	Which layer in Data Science architecture deals with data acquisition from various sources? A. Application layer B. Data ingestion layer C. Analytics layer D. Visualization layer **Answer:** B	**1**	**1**	**CO1**
8	Who typically occupies the top position in the Data Science hierarchy within an organization? A. Data Analyst B. Data Engineer C. Chief Data Officer (CDO) D. Business User **Answer:** C	**1**	**1**	**CO1**
9	Which of the following is a major challenge in Data Science projects? A. Excessive computational power	**1**	**1**	**CO1**

	B. Poor data quality C. Availability of tools D. Too much documentation **Answer:** B			
10	**Which industry widely applies Data Science for fraud detection?** A. Agriculture B. Healthcare C. Banking and Finance D. Education **Answer:** C	**1**	**1**	**CO1**
MODULE 2				
1.	**What is the primary importance of Microsoft Excel?** A. Creating animations B. Performing data analysis and calculations C. Web browsing D. Software development **Answer:** B	**2**	**1**	**CO1**
2	**Which feature in Excel is used to organize data in a structured format?** A. Pivot Table B. Chart C. Excel Table	**2**	**1**	**CO1**

	D. Macro **Answer:** C			
3	**Which symbol is used for multiplication in Excel formulas?** A. × B. # C. * D. % **Answer:** C	**2**	**1**	**CO1**
4	**To restrict the type of data entered into a cell, which Excel feature is used?** A. Conditional Formatting B. Data Validation C. Filtering D. Sorting **Answer:** B	**2**	**1**	**CO1**
5	**Which option is used to arrange data in ascending or descending order?** A. Grouping B. Subtotal C. Sorting D. Validation **Answer:** C	**2**	**1**	**CO1**
6	**Which Excel feature displays only records that meet specific criteria?** A. Sorting B. Filtering C. Subtotal	**2**	**1**	**CO1**

	D. Grouping **Answer:** B			
7	**Which symbol is used to start any formula in Excel?** A. # B. $ C. = D. @ **Answer:** C	**2**	**1**	**CO1**
8	**Which logical function returns TRUE if all conditions are satisfied?** A. OR B. NOT C. IF D. AND **Answer:** D	**2**	**1**	**CO1**
9	**Which chart type is best suited for showing trends over time?** A. Pie Chart B. Column Chart C. Line Chart D. Bar Chart **Answer:** C	**2**	**1**	**CO1**
10	**Which file format is commonly used to import text data into Excel?** A. .xml	**2**	**1**	**CO1**

	B. .xls C. .csv D. .accdb **Answer:** C			
MODULE 3				
1	**Which type of machine learning uses labeled training data?** A. Unsupervised learning B. Reinforcement learning C. Supervised learning D. Semi-supervised learning **Answer:** C	**3**	**1**	**CO1**
2	**Which of the following is a classification algorithm?** A. K-Means B. Decision Tree C. Apriori D. PCA **Answer:** B	**3**	**1**	**CO1**
3	**Which algorithm is commonly used for clustering?** A. Logistic Regression B. Naïve Bayes C. K-Means	**3**	**1**	**CO1**

	D. Linear Regression **Answer:** C			
4	**Which technique is used for feature selection?** A. Gradient Descent B. Principal Component Analysis (PCA) C. K-Nearest Neighbor D. K-Means **Answer:** B	**3**	**1**	**CO1**
5	**Bayes’ theorem is mainly used to calculate:** A. Mean B. Conditional probability C. Variance D. Correlation **Answer:** B	**3**	**1**	**CO1**
6	**In a Cartesian plane, how many axes are present?** A. One B. Two C. Three D. Four **Answer:** B	**3**	**1**	**CO1**
7	**Which of the following represents a linear equation?** A. $y = x^2$ B. $y = mx + c$ C. $y = \log x$	**3**	**1**	**CO1**

	D. $y = e^x$ Answer: B			
8	Which SQL command is used to retrieve data from a database? A. INSERT B. UPDATE C. SELECT D. DELETE Answer: C	3	1	CO1
9	Which SQL command category includes COMMIT and ROLLBACK? A. DDL B. DML C. TCL D. DCL Answer: C	3	1	CO1
10	Which tool is commonly used for data science tasks? A. MS Paint B. Jupyter Notebook C. Notepad D. Calculator Answer: B	3	1	CO1
MODULE 4				

1.	**Correlation measures the:** A. Causation between two variables B. Strength and direction of relationship between variables C. Difference between variables D. Distribution of data **Answer:** B	**4**	**1**	**CO1**
2.	**Which correlation coefficient value indicates a strong positive correlation?** A. －0.9 B. －0.1 C. 0 D. +0.9 **Answer:** D	**4**	**1**	**CO1**
3.	**Which regression technique is used for binary classification problems?** A. Linear Regression B. Polynomial Regression C. Logistic Regression D. Ridge Regression **Answer:** C	**4**	**1**	**CO1**
4.	**The Gaussian distribution is also known as:** A. Uniform distribution B. Binomial distribution C. Normal distribution	**4**	**1**	**CO1**

	D. Poisson distribution **Answer:** C			
5.	**Standardization converts data to have:** A. Mean = 1 and Variance = 0 B. Mean = 0 and Standard Deviation = 1 C. Mean = 1 and SD = 1 D. Mean = 0 and Variance = 0 **Answer:** B	**4**	**1**	**CO1**
6.	**Z-score represents:** A. Raw data value B. Probability value C. Number of standard deviations from the mean D. Variance of data **Answer:** C	**4**	**1**	**CO1**
7.	**According to the Central Limit Theorem, the sampling distribution of the mean approaches:** A. Uniform distribution B. Binomial distribution C. Normal distribution D. Exponential distribution **Answer:** C	**4**	**1**	**CO1**

8.	Markowitz Portfolio Optimization primarily focuses on: A. Maximizing risk B. Minimizing return C. Optimizing risk and return D. Eliminating variance Answer: C	**4**	**1**	**CO1**
9.	Standardizing x and y variables in linear regression helps to: A. Increase data size B. Simplify coefficient interpretation C. Remove correlation D. Increase error Answer: B	**4**	**1**	**CO1**
10.	Information gain in linear regression helps in: A. Increasing variance B. Feature selection and model improvement C. Reducing data size D. Data visualization Answer: B	**4**	**1**	**CO1**
MODULE 5				
1.	Which visualization technique is best suited to show the relationship between two continuous variables? A. Histogram	**5**	**1**	**CO1**

	B. Scatter plot C. Pie chart D. Map **Answer:** B			
2.	**Which measure represents the average value of continuous data?** A. Median B. Mode C. Mean D. Range **Answer:** C	**5**	**1**	**CO1**
3.	**Standard deviation is used to measure:** A. Central tendency B. Frequency C. Dispersion of data D. Percentage **Answer:** C	**5**	**1**	**CO1**
4.	**Which statistical measure is most appropriate for categorical data analysis?** A. Mean B. Standard deviation C. Frequency and percentage D. Variance **Answer:** C	**5**	**1**	**CO1**
5.	**Which Python data type is immutable?** A. List	**5**	**1**	**CO1**

	B. Dictionary C. Set D. Tuple **Answer:** D			
6.	**Which Python library is mainly used for numerical computations?** A. Pandas B. NumPy C. Matplotlib D. Scikit-Learn **Answer:** B	**5**	**1**	**CO1**
7.	**Which library is primarily used for data visualization in Python?** A. NumPy B. Pandas C. Matplotlib D. Scikit-Learn **Answer:** C	**5**	**1**	**CO1**
8.	**Which Python data structure stores data in key－value pairs?** A. List B. Tuple C. Set D. Dictionary **Answer:** D	**5**	**1**	**CO1**

9.	**Which library is used for data manipulation and analysis using DataFrames?** A. NumPy B. Pandas C. Matplotlib D. TensorFlow **Answer:** B	**5**	**1**	**CO1**
10.	Which Python library is widely used for implementing machine learning algorithms? A. Matplotlib B. NumPy C. Pandas D. Scikit-Learn Answer: D	**5**	**1**	**CO1**
Question NO.				
MODULE 1				
1.	Define Data Science and explain its scope in modern data-driven organizations.	**1**	**8**	**CO1, PO1**
2.	Describe the history and evolution of Data Science, highlighting key technological milestones.	**1**	**8**	**CO1, PO1**
3.	Explain the important terminologies used in Data Science such as big data, machine learning, data mining, and artificial intelligence.	**1**	**8**	**CO2, PO1**
4.	Discuss the basic framework and architecture of Data Science with suitable examples.	**1**	**8**	**CO2, PO1**
5.	Differentiate between Data Science and Business Analytics with respect to objectives, tools, and applications.	**1**	**8**	**CO3, PO2**

6.	Explain the importance of Data Science in today’s business world with real-world use cases.	**1**	**8**	**CO3, PO2**
7.	Describe the primary components of Data Science and explain the role of each component.	**1**	**8**	**CO2, PO1**
8	Explain the different users of Data Science in an organization and discuss the Data Science hierarchy.	**1**	**8**	**CO3,PO4**
9	Provide an overview of various Data Science techniques used for data analysis and decision-making.	**1**	**8**	**CO4,PO4**
10	Discuss the challenges and opportunities of Data Science in business analytics and explain its industrial applications across different sectors.		**8**	**CO4, PO2**
MODULE 2				
1.	Explain the importance of Microsoft Excel in data analysis and business applications.	**2**	**8**	**CO1,,PO5**
2.	Describe the steps involved in creating and managing Excel tables.	**2**	**8**	**CO2,PO5**
3.	Explain how to perform addition, subtraction, multiplication, and division in Excel with examples.	**2**	**8**	**CO2,PO5**
4.	Discuss Excel Data Validation and its role in maintaining data accuracy.	**2**	**8**	**CO2,PO5**
5.	Explain sorting, filtering, grouping, ungrouping, and subtotal operations in Excel.	**2**	**8**	**CO2,PO5**
6.	Introduce formulas and functions in Excel and explain their significance.	**2**	**8**	**CO3,PO5**
7.	Explain logical operators and conditional functions used in Excel with suitable examples.	**2**	**8**	**CO4,PO5**
8.	Describe different types of charts in Excel and explain how they help in data visualization.	**2**	**8**	**CO4,PO5**
9	Explain the procedure to import XML, CSV (Text), and MS Access data into Excel.	**2**	**8**	**CO4,PO5**
10	Discuss working with multiple worksheets and managing data across worksheets in Excel.	**2**	**8**	**CO4,PO5**
MODULE-3				

1.	Explain the different types of machine learning with suitable examples.	**3**	**8**	**CO1,PO1**
2.	List and explain machine learning algorithms used for classification, clustering, and feature selection.	**3**	**8**	**CO1,PO2**
3.	Explain probability theory and derive Bayes' theorem with an example.	**3**	**8**	**CO1,PO1**
4.	Define Bayes probability and explain its role in machine learning.	**3**	**8**	**CO1,PO2**
5.	Explain the Cartesian plane and equations of straight lines with graphical representation.	**3**	**8**	**CO2,PO1**
6	Explain the concept of exponents and their importance in data science computations.	**3**	**8**	**CO2,PO1**
7	Describe commonly used tools for data science and explain their applications.	**3**	**8**	**CO4,PO1**
8	Explain SQL and describe different SQL command categories: DDL, DML, DCL, TCL, and DQL with examples.	**3**	**8**	**CO3,PO5**
9	Demonstrate the use of SELECT, INSERT, UPDATE, and DELETE commands in SQL.	**3**	**8**	**CO3,PO5**
10	Explain the procedure to import SQL database data into Microsoft Excel.	**3**	**8**	**CO4,PO5**
MODULE-4				
1.	Define correlation and explain its types with suitable examples.	**4**	**8**	**CO1, PO1**
2.	Describe linear regression and explain its assumptions and applications.	**4**	**8**	**CO1, PO1**
3.	Explain logistic regression and compare it with linear regression.	**4**	**8**	**CO1, PO2**
4.	Explain the Gaussian (normal) distribution and discuss its properties.	**4**	**8**	**CO1, PO1**
5.	Explain the concept of standardization and its importance in data analysis.	**4**	**8**	**CO2, PO2**
6	Explain the standard normal probability distribution and demonstrate probability calculation using Excel.	**4**	**8**	**CO2,PO5**

7	Explain Z-scores and describe how probabilities are calculated using Z-score tables.	**4**	**8**	**CO2, PO2**
8	State and explain the Central Limit Theorem and its significance in statistics.	**4**	**8**	**CO2, PO4**
9	Explain Markowitz Portfolio Optimization and discuss the role of Gaussian algebra in finance.	**4**	**8**	**CO3,PO3**
10	Explain how standardization simplifies linear regression, including modeling error and information gain.	**4**	**8**	**CO4,PO4**
MODULE-5				
1.	Explain the importance of data visualization and describe scatter plots, charts, graphs, histograms, and maps with suitable examples.	**5**	**8**	**CO1, PO1**
2.	Explain descriptive statistics for continuous data, focusing on mean and standard deviation, and discuss their significance.	**5**	**8**	**CO2,PO1**
3.	Explain how frequency and percentage are used for analyzing categorical data with examples.	**5**	**8**	**CO2,PO2**
4.	Describe the basic concepts and features of Python that make it suitable for data science.	**5**	**8**	**CO2,PO2**
5.	Explain Python strings and lists, including their operations and applications in data analysis.	**5**	**8**	**CO3,PO1**
6	Describe tuples, sets, and dictionaries in Python and explain their differences with examples.	**5**	**8**	**CO4,PO5**
7	Explain the role of NumPy in data science and discuss its important features.	**5**	**8**	**CO4,PO5**
8	Describe the use of Pandas and Matplotlib libraries for data analysis and visualization.	**5**	**8**	**CO3,PO1**
9	Explain the purpose of the Scikit-Learn library and discuss its role in implementing machine learning models.	**5**	**8**	**CO4,PO5**
10	Describe the general steps involved in implementing a machine learning model using Python libraries.	**5**	**8**	**CO4,PO5**

Signature of the subject teacher Signature of the Course Co-Ordinator Signature of the Dean