



School of Engineering and Technology

First Year Department (CSE/CSE-AIML/CSE-AIDS/CSE-DS/ECE/EEE)

BTech - Semester-I

Question bank

Course Title: Fundamentals of Data Science

Course Code: 24BTELY107

SL NO	QUESTIONS	M AR KS	COs	BL
1.	What is Data Science?	2	CO1	R
2.	What are the needs of Data Science?	2	CO1	R
3.	List out the various field, where data science are used?	2	CO1	U
4.	What is data cleaning?	2	CO1	U
5.	What are various forms of data used in data science?	2	CO1	U
6.	List out the steps involved in data science process.	2	CO1	U
7.	List out the types of Big Data.	2	CO1	R
8.	List out the types of Metadata.	2	CO1	R
9.	Define data mining.	2	CO1	U
10.	List out the basic requirements of Data Scientist.	2	CO1	R
11.	Define structured data.	2	CO1	R
12.	Define unstructured data.	2	CO1	R
13.	What is outliers?	2	CO1	U
14.	Define Data preparation.	2	CO1	R
15.	Mention the techniques used for sampling.	2	CO1	U
16.	What is Exploratory data analysis?	2	CO1	U
17.	What are the key aspects of Exploratory data analysis?	2	CO1	A
18.	List out the types of Exploratory Data Analysis.	2	CO1	U
19.	What are Descriptive Statistics?	2	CO1	A
20.	List out the characteristics of Big data.	2	CO1	U
21.	Illustrate the types of sampling in detail.	5	CO1	U
22.	Explain the various tools used for Data Science process.	5	CO1	A
23.	Discuss about the various types of data.	5	CO1	U
24.	Differentiate between structured and unstructured data.	5	CO1	2
25.	Explain in detail about Meta data, its types with example.	10	CO1	2
26.	Discuss in detail about Statistical Inference with its types.	10	CO1	U
27.	Explain the different types of population in detail.	10	CO1	U
28.	Illustrate the concept and characteristics of Data warehousing in detail.	10	CO1	U
29.	Illustrate the concept of Big Data with applications in detail.	20	CO1	R
30.	Explain the roles and responsibilities of a Data Scientist.	20	CO1	R
31.	Illustrate the concept of population and sample in detail.	20	CO1	U
32.	Explain the various steps involved in Data Science process in detail.	20	CO1	U
33.	Analyze the various factors of housing market in detail.	20	CO1	A
34.	Explain the various job profiles of Data Science in detail.	20	CO1	U



School of Engineering and Technology

First Year Department (CSE/CSE-AIML/CSE-AIDS/CSE-DS/ECE/EEE)

Module - 2

SL NO	QUESTIONS	M AR KS	COs	BL
35.	Define Probability.	2	CO2	R
36.	Discuss any two terms in Probability.	2	CO2	R
37.	Mention the types of Descriptive statistics.	2	CO2	R
38.	List some applications of conditional probability	2	CO2	R
39.	What is Questionnaire Method?	2	CO2	R
40.	Define Bayes Theorem.	2	CO2	R
41.	Define measure of variability.	2	CO2	R
42.	Write formula for regression.	2	CO2	R
43.	Define Data Munging.	2	CO2	R
44.	What is data enrichment?	2	CO2	R
45.	Define data transformation.	2	CO2	R
46.	What is quality assurance in data?	2	CO2	R
47.	Write down definition of mean with formula.	2	CO2	R
48.	Define Mode and Median.	2	CO2	R
49.	Explain the types of correlation	5	CO2	R
50.	Difference between correlation and regression.	5	CO2	R
51.	Using the computation formula for the sum of squares, calculate the population standard deviation and sample deviation for the score? a)1,3,7,2,0,4,3,7, b)10,8,5,0,1,7,9,2,1	5	CO2	A
52.	Compute the mean, median and mode for the following data sets a)45, 55, 60, 60, 63, 63, 63, 63, 65, 65, 70 b)26.9, 26.3, 28.7, 27.4, 26.6, 27.4, 26.9, 26.9	5	CO2	A
53.	Describe Data cleaning process.	5	CO2	R
54.	a) Define crowdsourcing b) Explain the types of crowdsourcing	5	CO2	R
55.	Explain about the Primary data collection methods.	5	CO2	U
56.	Explain any two types of Descriptive statistics.	5	CO2	U
57.	Write a short note on Measures of Central Tendency in statistics.	5	CO2	U
58.	Write a short note on Data munging.	5	CO2	U
59.	Differentiate between Correlation and Regression.	10	CO2	R
60.	a) Describe about correlation in detail with its types. b) Explain any one application of correlation.	10	CO2	R
61.	Mention the similarities between Correlation and Regression	10	CO2	R
62.	Explain Conditional Probability. i. Formula ii. Example	10	CO2	R
63.	Explain Bayes theorem with a suitable example in detail.	20	CO2	U
64.	Explain Data Cleaning process. (all 8 steps)	20	CO2	U



School of Engineering and Technology

First Year Department (CSE/CSE-AIML/CSE-AIDS/CSE-DS/ECE/EEE)

65.	Discuss data collection methods in detail.	20	CO2	U
66.	Discuss Crowdsourcing in data science.	20	CO2	U
67.	a) Describe how crowdsourcing can be used to collect data for a machine learning project. b) Provide an example and discuss the potential challenges and solutions.	20	CO2	A

Module 3

SL N O	QUESTIONS	MA RK S	COs	B L
68.	Define scoring and Ranking.	2	CO3	R
69.	Explain Z-score in statistics?	2	CO3	U
70.	What does a z-score of 0 mean?	2	CO3	U
71.	Why are z-scores useful in data analysis?	2	CO3	U
72.	What is Normalization?	2	CO3	R
73.	Define Sampling.	2	CO3	R
74.	Define Distribution.	2	CO3	R
75.	What is outlier detection?	2	CO3	U
76.	Define Standard Deviation.	2	CO3	U
77.	Define Normalization and Outlier detection.	2	CO3	R
78.	What is the importance of p-value?	2	CO3	U
79.	Write down Characteristics of Z-Score.	5	CO3	R
80.	How is a z-score calculated?	5	CO3	A
81.	What is z-score and why are z-scores useful?	5	CO3	U
82.	Define and explain Normalization.	5	CO3	U
83.	What is Null Hypothesis Testing?	5	CO3	A
84.	What is min-max scaling? Give one example.	5	CO3	A
85.	Define binomial distribution.	5	CO3	A
86.	Illustrate the concept of population and sample in detail.	5	CO3	U
87.	Brief overview of the steps involved in developing scoring systems.	10	CO3	U
88.	Write a note on scoring and ranking with example.	10	CO3	A
89.	Explain Z-Score with formula and example.	10	CO3	A
90.	a) Brief characteristics of Z-Score. b) Write use of z-score in data science.	10	CO3	R
91.	Explain the statistical significance in terms of Null hypothesis, permutation test and P-values.	10	CO3	R
92.	Write a note on Sampling and Distribution.	10	CO3	U
93.	Write use of z-score in data science and explain with example.	20	CO3	A
94.	Explain Normalization with example	20	CO3	A
95.	Write a note on Permutation test and P-value in detail.	20	CO3	R
96.	Explain in detail Types of Statistical Significance Tests.	20	CO3	U
97.	Explain Statistical distribution types with examples.	20	CO3	A



School of Engineering and Technology

First Year Department (CSE/CSE-AIML/CSE-AIDS/CSE-DS/ECE/EEE)

Module-4

SL NO	QUESTIONS	MARKS	COs	BL
98.	What are the principles in data visualization.	2	CO4	R
99.	What are the steps involved in data visualization?	2	CO4	R
100.	Name few tools used for creating data visualizations.	2	CO4	R
101.	What are the 12 key design principles for effective data visualization?	2	CO4	U
102.	What is univariate analysis?	2	CO4	U
103.	How can you choose the right chart for your data visualization needs?	2	CO4	U
104.	What is the primary function of a scatter plot?	2	CO4	U
105.	What are the elements in Box Plot.	2	CO4	U
106.	What is the main goal of Exploratory Data Analysis (EDA)?	2	CO4	U
107.	Why is it important to check for missing values during EDA?	2	CO4	U
108.	Describe the role of a heatmap in multivariate analysis.	2	CO4	U
109.	What is a deterministic model?	2	CO4	U
110.	Difference between a descriptive model and a prescriptive model.	2	CO4	U
111.	What is a baseline model in machine learning?	2	CO4	R
112.	What does cross-validation help to assess in machine learning models?	2	CO4	U
113.	Define mean squared error (MSE).	2	CO4	R
114.	What is the Zero Rule (ZeroR) in machine learning?	2	CO4	R
115.	How does a Random Classifier function in machine learning?	2	CO4	U
116.	What are evaluation models in machine learning, and how are they used?	2	CO4	U



School of Engineering and Technology

First Year Department (CSE/CSE-AI ML/CSE-AIDS/CSE-DS/ECE/EEE)

117.	What is a confusion matrix and how is it used to evaluate machine learning models?	2	CO4	U
118.	Explain the important of Data Visualization in detail?	5	CO4	U
119.	Explain the steps for performing Exploratory Data Analysis.	5	CO4	U
120.	Describe how to choose the appropriate chart type for a given dataset and analysis goal. What factors would you consider, and why?	5	CO4	A
121.	Draw the block diagram of basic model evaluation environment and explain briefly.	5	CO4	A
122.	Explain the difference between deterministic and stochastic models. Provide an example for each model.	5	CO4	A
123.	Analyse how color, typography, and layout contribute to the overall aesthetic of a data visualization. Provide examples of good and bad practices for each element.	10	CO4	A
124.	Discuss in detail about the Taxonomy of Models.	10	CO4	U
125.	Explain the process of cross-validation in model evaluation. Why is it important, and how does it contribute to the robustness of a model?	10	CO4	A
126.	Discuss in detail about the Random Classifier in Model Evaluation.	10	CO4	U
127.	Discuss the role of univariate, bivariate, and multivariate analyses in Exploratory Data Analysis (EDA).	20	CO4	U
128.	Describe the important of charts for data visualization and explain each chart in detail.	20	CO4	U
129.	Explain the concept of a baseline model in machine learning. Why are baseline models important, and how are they used to evaluate the performance of more complex models?	20	CO4	A



School of Engineering and Technology

First Year Department (CSE/CSE-AIIML/CSE-AIDS/CSE-DS/ECE/EEE)

Module 5

SL NO	QUESTIONS	MARKS	COs	BL
130.	Define Data Engineering	2	CO5	R
131.	What are the key responsibilities of a data engineer?	2	CO5	U
132.	What is Map Reduce? Explain its basic concept.	2	CO5	U
133.	What are the steps involved in a Map Reduce solution?	2	CO5	U
134.	Name any two popular tools used for social network analysis.	2	CO5	R
135.	What are the nodes and edges in a social network graph?	2	CO5	R
136.	Define Degree Centrality and Betweenness Centrality.	2	CO5	R
137.	Define clustering in the context of graphs.	2	CO5	U
138.	What is graph partitioning?	2	CO5	U
139.	What are the applications of Graph Partitioning.	2	CO5	R
140.	Name two popular tools used for social network analysis and briefly describe their features.	2	CO5	U
141.	Explain the concept of social networks as graphs.	2	CO5	U
142.	Discuss the role of data engineers in managing and transforming the data.	5	CO5	U
143.	Explain how Map Reduce helps in solving the Word Frequency Problem.	5	CO5	A
144.	Discuss the significance of clustering in social network graphs.	5	CO5	U
145.	Describe a method used for partitioning graphs.	5	CO5	U
146.	Explain the steps involved in Clustering.	5	CO5	A
147.	Illustrate with an example how clustering can be used to identify communities within a social network graph.	5	CO5	A
148.	Explain the role of the Map function in the Map Reduce framework and provide an example.	5	CO5	A



School of Engineering and Technology

First Year Department (CSE/CSE-AIML/CSE-AIDS/CSE-DS/ECE/EEE)

149.	Explain the concept of graph partitioning and its importance in large-scale data analysis.	10	CO5	U
150.	Describe in detail the architecture and working of the Map Reduce model. Include an example to illustrate your points.	10	CO5	A
151.	Discuss the different techniques of graph clustering and their applications in social network analysis.	10	CO5	A
152.	a) Explain the concept of edge betweenness centrality and its significance in social network analysis. b) Describe how it can be used to detect important connections within a network.	20	CO5	U
153.	Write short note on Social networks as graphs with its basic components, concept, applications and tools of network analysis.	20	CO5	U

Bloom's taxonomy:

R-Remembering: can the student recall or remember the information?

U-Understanding: can the student explain ideas or concepts?

A-Applying: can the student use the information in a new way?