Dr. D.Y. Patil Institute of Technology, Pimpri, Pune
Department of Artificial Intelligence and Data Science
Academic Year: 2023-24 (Semester II)

Year: TE AI&DS   **Course: Data Science**

Unit 1 and 2 Question Bank

BLOOMS Level (1-6)	Q. No.	Questions	Marks
3	1.	What is data science? Identify three areas or domains in which data science is being used and describe how.	4
2	2.	How data science relate to and differ from Information Science?	4
2	3.	Explain Business intelligence versus Data Science?	5
2	4.	How data science relate to and differ from Business intelligence?	4
2	5.	Differentiate Structured and Unstructured Data	4
2	6.	Discussed the obstacles in unstructured data in detail	4
1	7.	List out characteristic of Open Data and different data collection methods?	4
1	8.	List the data Storage Format and Explain any two of them.	4
3	9.	Select and discuss the reason of the appropriate data storage format for personal digital assistants (PDAs), and smart watches, Comment on your answer in detail.	4
2	10.	Why data wrangling is important for data science?	4
2	11.	Explain Type of Missing Data with Example	4
2	12.	Discuss 4 conventional Methods to deal with missing data.	4
3	13.	Compare Next Observation Carried Backward and Linear Interpolation techniques to deal with missing data	4
2	14.	Discuss 3 Steps of Data Integration.	3
1	15.	List and explain various Data Discretization techniques	5
1	16.	Write a short on Data Normalization.	4
2	17.	Explain dimensionality reduction with relation to forward and backward selection	4
3	18.	Comment on the statement "Variance of the particular feature is zero."	4
2	19.	What is population and how it is different from a sample?	2
3	20.	Explain Nominal Data with Example. Also mention the Data Visualization techniques for nominal data.	4
4	21.	With reference to Absolute zero, Compare and differentiate Interval and Ratio scale data type of attributes.	5
3	22.	How to decide the type of data attributes? You can use flowchart to support your comments.	4
4	23.	A pizza outlet overview its weekly sales. They sold 57 cheese pizzas, 63 pasta pizzas, 53 veggies pizzas, 68 cottage cheese pizzas, and 56 max cheese pizzas. Find the mean of all the pizzas sold by them.	5
3	24.	The marks obtained by 40 students out of 50 in a class are given below in the table. Find the mode of the above data.	5
	25.		
3	26.	The following frequency distribution table shows the monthly consumption of electricity of 68 consumers of a locality. Find the median of the given data.	6

Marks (in $)	42	36	30	45	50
Number of Students	7	10	13	8	2

	27.	Monthly consumption of electricity (in units)	Number of consumers	
		65 – 85	4	
		85 – 105	5	
		105 – 125	13	
		125 – 145	20	
		145 – 165	14	
		165 – 185	8	
		185 – 205	4	
3	28.	With reference skewness of data, explain the empirical relation between mean, mode and median.		4
3	29.	With the suitable example, Comment on the statement "The range is influence by outliers"		4
3	30.	Here are the 19 scores listed out. 5, 7, 10, 15, 19 ,21, 21, 22, 22, 23, 23, 23, 23, 23, 24,24,24, 24, 25 Calculate 1.5*IQR for below the first quartile and above the third quartile. How many data points can low outliers or above outliers?		6
3	31.	For the given numbers find out variation and standard deviation. Also discuss that how variation and standard deviation are related to each other? 4, 34, 11, 12, 2, and 26		6
4	32.	A man is known to speak the truth 2 out of 3 times. He throws a die and reports that the number obtained is a four. Apply Bayes Theorem and Find the probability that the number obtained is actually a four.		6
1	33.	Explain the Chi-Square and t -test hypothesis testing with an example		6
2	34.	Given that the card randomly draw from a pack of 52 cards is a Face card, what is the probability that it is a Queen?		4
1	35.	Define hypothesis. What is hypothesis testing?		4
2	36.	What is the difference between Null and Alternative hypothesis. Give one example of each		4
3	37.	A researcher has exam results for a sample of students who took a training course for a national exam. The researcher wants to know if trained students score above the national average of 850. i) Define Null Hypothesis and Alternative Hypothesis. ii) Is it one tail or two tail hypothesis? Comment on your answer.		4