



**PRESIDENCY
UNIVERSITY**

BENGALURU

School of Computer Science and Engineering

QUESTION BANK

CSE3190- Fundamentals of Data Analytics

SL. NO	Questions.	marks
1.	Define Data Analysis. Mention the different types of Data Analysis.	2 Marks
2.	List out the difference between data and information.	2 Marks
3.	Identify 10 V's of data.	2 Marks
4.	List out the difference between Qualitative data and Quantitative data.	2 Marks
5.	Name the command used for Accessing Subgroups in Data Frames	2 Marks
6.	Define Data Analysis.	2 Marks
7.	List the command used to add a column and row a to a data frame.	2 Marks
8.	Define the vector of integers from 2 to 10 in R.	2 Marks
9.	List two features of R studio	2 Marks
10.	Demonstrate the program in R to display the "Hello World".	2 Marks
11.	What is Data analysis? And provide a real-world example for it.	2 Marks
12.	List the different types of digital data with examples.	2 Marks
13.	Classify the following variables based on their scale of measurement: Movie ratings, Political parties, Profit, Time.	2 Marks
14.	Describe the steps involved in creating a variable in R, providing a specific example to illustrate each step.	2 Marks
15.	Which are some common causes of Missing Data.	2 Marks
16.	Explain the different sources of data?	5 Marks
17.	Discuss the concept of relational operators in R. Provide example.	5 Marks
18.	Review the concept of logical operators in R. Provide example.	5 Marks
19.	Differentiate between structured & unstructured data	5 Marks

20.	Explain the measures of central tendency (mean, median, and mode) and compute them for the following dataset: 3, 4, 5, 6, 7, 7, 7, 8, 8, 9	5 Marks
21.	Summarize , the basic data types available in R and provide examples for each.	5 Marks
22.	Describe nominal, ordinal and interval scales of measurement in data analysis.	5 Marks
23.	Discuss the rules to declare variables in R? Provide example for valid and invalid variable.	5 Marks
24.	Summarize the concept of data frames in R and provide an example.	5 Marks
25.	Illustrate , how to access and update elements in dataframes.	5 Marks
26.	Describe the concept of list in R and provide an example.	5 Marks
27.	Demonstrate , how to access and update elements in lists.	5 Marks
28.	Apply the cbind() and rbind() functions to combine the following multiple vectors in R. Display the result. vector1 <- c(1, 2, 3) vector2 <- c(4, 5, 6)	4 Marks
29.	Explain the following functions. i) is.na() ii) sum(is.na()) iii) anyNA() iv) complete.cases() v) colSums(is.na()) and rowSums(is.na())	6 Marks
30.	Apply the cbind() and rbind() functions combine the following multiple matrices in R. Display the result. matrix1 <- matrix(1:6, nrow = 3, ncol = 2) matrix2 <- matrix(7:12, nrow = 3, ncol = 2)	4 Marks
31.	Compute and normalize following group of data: 1000,2000,3000,9000 using min-max normalization by setting min:0 and max:1	6 Marks
32.	Differentiate between Data and Information	5 Marks
33.	Demonstrate the creation of data frame consists of Roll Numbers of 5 students along with their names and Mid Term and End Term Marks	5 Marks
34.	Summarize the key distinctions between structured and unstructured data.	5 Marks
35.	Illustrate the creation of data frame consists of Employee ID of 10 employees along with their names and their department in and the reporting manager	5 Marks
36.	Describe the human-generated unstructured data.	7 Marks
37.	Illustrate the loading of the cars.csv into a dataframe called Cars_Info without the header.	3 Marks
38.	Explain the necessary steps for Data preprocessing.	7 Marks

39.	Demonstrate the data type of each column in cars.csv with output.	3 Marks
40.	Explain the types of Digital Data.	4 Marks
41	Interpret the output of the following commands on the Cars_Info i) nrow(Cars_Info) ii) sum(is.na(Cars_Info)) iii) anyNA(Cars_Info)	6 Marks
42	Discuss about the main sources of Data	
43	Apply your knowledge to analyze the output of the following commands on the Cars_Info dataset. i) rowSums(is.na(Cars_Info)) ii) class(Cars_Info) iii) col(is.na(data))	
42	E-commerce Business Analysis A fictional e-commerce company, ShopSmart, is facing challenges with declining sales over the last few months. To address this issue, ShopSmart's management wants to use data analytics to better understand the situation and decide on actionable steps i) Discuss how the four types of data analytics (descriptive, diagnostic, predictive, and prescriptive) were applied to help ShopSmart understand and address its declining sales. (6 marks) ii) Explain the key insights derived from diagnostic and prescriptive analytics, and how these insights influenced the company's strategy to improve sales. (4 marks)	6 Marks 4 Marks
43	Hospital Readmission Reduction Program A large hospital, CityCare Medical Center, is struggling with high rates of patient readmissions within 30 days of discharge, particularly for patients with heart failure. The hospital administration aims to reduce readmission rates by using data analytics to understand the situation and improve patient outcomes. i) Describe the role of each of the four types of data analytics (descriptive, diagnostic, predictive, and prescriptive) in helping CityCare Medical Center reduce patient readmissions.(6marks) ii) Summarize specific insights were gained from the use of diagnostic and predictive analytics, and how did these insights guide the hospital's decision-making? (4 marks)	6 Marks 4 Marks

44	Differentiate between Data and Information	3Marks
45	List and explain Many V's of Data	7Marks
46	Differentiate between qualitative and quantitative variables, providing real-world examples.	5Marks
47	Create a data frame named students with columns for "Name", "Age", and "Marks". Add some sample data and print the data frame.(Minimum 5 rows)	5Marks
48	List out some of the human generated structured data	4Marks
49	Using R, Create a vector called Exam having minimum of 7 numbers, and write the code of how you would calculate mean, median and mode for the given vector.	6Marks
50	Outline the key steps involved in data preparation	5Marks
51	List the different commands used in finding missing values in the given dataset.	5Marks
52	List and explain some of the main sources of data.	4Marks
53	Using R, create two 3x3 matrices filled with numbers between 1 and 10. Combine these two matrices using the cbind() and rbind() functions. Explain the difference between the resulting matrices.	6Marks
54	Demonstrate how to import a excel file inside R studio inside a dataframe called data without any headers and separated by a comma.	4Marks
55	Create two dataframes called data1 and data2 having students details and combine both dataframes into a single matrix.	6Marks
56	List the different methods to access elements from a given List.	3Marks
57	Write an R script to create a function named calculator that takes two numbers and an operator as input. The function should perform the specified operation and return the result. The valid operators should be "+", "-", "*", and "sqrt".	8Marks
58	<p>Student Performance Analysis: The school administration is concerned about the declining overall academic performance of students. They want to analyze student data to identify areas for improvement and implement targeted interventions.</p> <p>i) Explain how the four types of data analysis will be used to analyze school performance. and,</p> <p>ii) Develop a comprehensive plan to improve student performance, utilizing predictive and prescriptive analytics. What strategies can be implemented based on data-driven insights?</p>	<p>6Marks</p> <p>4Marks</p>

cars.csv stored as Cars_Info

Model	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21	6	160	110	3.9	2.62	16.46	0	1	4	4
Mazda RX4 Wag	21	6	160	110	3.9	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.32	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.44	17.02	0	0	3	2
Valiant	18.1	6	225	105	2.76	3.46	Na	1	0	3	1
Duster 360	14.3	8	360	245	3.21	3.57	15.84	0	0	3	4
Merc 240D	24.4	4	146.7	62	3.69	3.19	20	1	0	4	2
Merc 230	22.8	4	140.8	95	3.92	3.15	22.9	1	0	4	2