ST. JOSEPH'S COLLEGE (AUTONOMOUS), BANGALORE – 27
BA/BSc/BCom/BBA – IV SEMESTER
SEMESTER EXAMINATION – APRIL 2022
(EXAM CONDUCTED IN JULY 2022)
BCADAOE4121: DATASCIENCE USING R

Time: 1.5 hours   **Maximum marks: 35 marks**

This paper contains five printed pages

Answer all questions from the following   **1*35 = 35 Marks**

1. R files has an extension………………..
   a. .R
   b. .S
   c. .Rp
   d. .c
2. Data frames can be created by using the function
   a. Data. Sets ()
   b. Function ()
   c. C ()
   d. Data. Frames ()
3. Which is the basic data structure of R containing the same type of data?
   a. Functions
   b. Array
   c. Vector
   d. List
4. Identify the term related to analytics.
   a. Analytics about explaining past events
   b. Analytics about explaining present events
   c. Analytics about predicting future events
   d. None

5. Identify the pipeline operator which allows you to perform string operations left to right form
   a. %>%>
   b. %>%
   c. >%>%
   d. None of the above
6. Which of the following is TRUE for a vector in R
   a. It is a homogenous 1-dimensional data structure
   b. It is a heterogenous 1-dimensional data structure
   c. It is a homogenous 2-dimensional data structure
   d. It is a heterogenous 2-dimensional data structure
7. Which function cannot be used to import a CSV file in R
   a. read.table()
   b. read.csv()
   c. read_excel()
   d. none of the above
8. ………………. function is used to test objects if they are NaN
   a. as.nan()
   b. s.nan()
   c. is.nan()
   d. none of the above
9. …………………..code is used for scatter plot in R
   a. scatter.plain()
   b. scatter ()
   c. plot()
   d. None of these
10. For the population y<-c (1,2,3,4,5) what is the R command to find the mean ?
   a. mean(y)
   b. mean{y}
   c. means(y)
   d. mean[y]
11. What is the role of exploratory graphs in data analysis?
   a. They are made for formal presentations
   b. They are typically made very quickly
   c. Axes, legends, and other details are clean and exactly detailed
   d. They are used in place of formal modelling
12. If P value is more than 0.05 then it is _____
   a. Null Hypothesis
   b. Alternate Hypothesis
   c. positive correlation
   d. No correlation
13. Which of the following is the characteristic of Processed Data?
   a. Data is not ready for analysis
   b. All steps should be noted
   c. Hard to use for data analysis
   d. None of the mentioned

14. Which algorithm is used to check the correlation between two continuous values?
    a. Anova
    b. Karl pearson
    c. Chi square test
    d. None
15. Which of the following is used to find the variance of all values?
    a. var ()
    b. sd ()
    c. mean ()
    d. anova ()
16. Which of the following is one of the keys in data science skills?
    a. Statistics
    b. Machine learning
    c. Visualization
    d. All the above
17. Which of the following is considered data science?
    a. Case study
    b. Qualitative analysis
    c. Qualitative analysis
    d. Forecasting
18. Which one of the following functions is can be used to list the installed packages in R?
    a. library[]
    b. library()
    c. library{}
    d. None
19. Data science is the process of diverse set of data through_____
    a. organizing data
    b. processing data
    c. analysing data
    d. All of the mentioned
20. Which of the following step is performed by data scientist after acquiring the data?
    a. Data Cleansing
    b. Data Integration
    c. Data replication
    d. All the mentioned

21. If we have a precision value of 1 or 100% when all positive samples are classified as positive
    a. True
    b. False
22. Identify the task performed by Data Scientist?
    a. Define the question
    b. Create reproducible code
    c. Challenge results
    d. All of the mentioned

23. _____uses data on some object to predict values for other object?
    a. Inferential
    b. Exploratory
    c. Predictive
    d. None of the mentioned

24. Logistic regression is for predicting_____
    a. Continuous variable
    b. Categorical Variable
    c. Continuous and categorical
    d. None

25. Give example for unstructured data
    a. Excel
    b. Mail
    c. All the mentioned
    d. None of the mentioned

26. _____ is not an example of Machine Learning.
    a. symbolic reasoning
    b. Forecasting
    c. Client retention
    d. Image recognition

27. Identify the true statement.
    a. Raw data is original source of data
    b. Pre-processed data is original source of data
    c. Raw data is the data obtained after processing steps
    d. None

28. How do you handle missing or corrupted data in a dataset?
    a. Drop missing rows or columns
    b. Replace missing values with mean/median/mode
    c. Assign a unique category to missing values
    d. All the mentioned

29. Head and tail functions in R will show ……………..rows
    a. 4
    b. 5
    c. 6
    d. 7

30. ……………… shows many observations that are present with respect to the data like maximum, minimum, median, etc.
    a. Histogram
    b. Boxplot
    c. Pie chart
    d. None of these

31. Correlation values will vary from _____
    a. 0 to 1
    b. -1 to 1
    c. infinity to 1

d. none
32. Give example for structured data
    a. Excel
    b. Mail
    c. All the mentioned
    d. None of the mentioned
33. Which algorithm is used to check the correlation between continuous and categorical values?
    a. Anova
    b. Karl pearson
    c. Chi square test
    d. None
34. Which of the following is not a part of data science process?
    a. Discovery
    b. Model Planning
    c. Communication Building
    d. Operationalize
35. Which of the following are correct component for data science?
    a. Data Engineering
    b. Advanced Computing
    c. Domain expertise
    d. All the mentioned