

1. Digital data is the data that represents other forms of data using specific machine language that can be interpreted by human easily.
  - a. True
  - b. False
2. Identify the correct data sources from the given list (Multiple Choice)
  - a. CRM (Customer Relationship Management)
  - b. Business Partners
  - c. Microsoft Word Documents
  - d. Binary data
3. Select the sources of data in Big Data eco systems (Multiple Choice).
  - a. In/Out
  - b. Internal
  - c. External
  - d. None
4. Is a new data challenge that requires leveraging existing systems differently, then we classify data as Big Data.
  - a. True
  - b. False
5. Variety describes one of the biggest challenges of \_\_\_\_\_.
  - a. Big data
  - b. SQL
  - c. Data integration
  - d. None of the mentioned above
6. Big data analysis does the following except?
  - a. Spread data
  - b. Analyze data
  - c. Organizes data
  - d. Collect data
7. Major drawback of the file-based approach is that (Multiple Choice).
  - a. Repetition of data/information
  - b. Wastage of storage
  - c. Easy to read
  - d. None

8. Database supports \_\_\_\_\_ to perform database operation effectively.
- a. Command
  - b. Query Language
  - c. View
  - d. None
9. Identify the correct elements for the Database management system (Multiple Choice)
- a. Physical database
  - b. Database Engine
  - c. Schema
  - d. XAMPP
10. Cardinality is defined by the number of columns in a Table.
- a. True
  - b. False
11. Identify the incorrect Data Definition language (DDL) command in MySQL
- a. SELECT
  - b. CREATE
  - c. DROP
  - d. ALTER
12. VARCHAR data types allow to store variable length of character data in a Table.
- a. False
  - b. True
13. DEFAULT constraint will drop the repetitive values in Table
- a. True
  - b. False
14. Difference between TRUNCATE and DELETE is to delete the tables from MySQL Database.
- a. True
  - b. False
15. Find out the errors in MySQL queries "SELECT FROM employee\_details;"
- a. Column selection is missing
  - b. Can't say
  - c. Given query is correct
  - d. None

16. WHERE & HAVING used interchangeably to filter out in SELECT queries.
- True
  - False
17. View is physical copy of the table, and that will work exactly like Tables.
- False
  - True
18. Select possible JOIN in MySQL from the below given open (Multiple Choice)
- Mix join
  - Inner join
  - Left join
  - None
19. UPDATE & DELETE statement without WHERE will update only selected records.
- True
  - False
20. Identify the operators to combine two or more identical MySQL table
- UNION
  - UNION ALL
  - MERGE
  - None
21. setwd() functions to get the current working directory for R Project.
- False
  - True
22. Identify the commonly used variable assignment operators in R (Multiple Choice)
- =
  - <-
  - ==
  - +
23. Identify the R data structure which allows to combine mixed kind of data.
- Vector
  - List
  - Matrix
  - Data Frame

24. Find out the errors in given R code

```
fruits_vector <- c("Apple", "Banana", "Orange")  
num_list <- 1:50  
data <- list(fruits_vector, num_list)  
data[[1]][2]
```

- a. List creation is not correct
- b. Missing square brackets [ ]
- c. Wrong indexes
- d. Display output

25. Function to create a vectors in R

- a. Combine function
- b. Create function
- c. Make function
- d. None

26. cbind() function helps to combine \_\_\_\_\_ for two Data frame.

- a. Rows
- b. Columns
- c. Data Frame
- d. None

27. Vector data structure needed ROW INDEX and COLUMN INDEX to access the elements from a list

- a. True
- b. False

28. Dollar(\$) symbols used to access multiple columns together in R DataFrame.

- a. True
- b. False

29. table() & prop.table() function helps to calculate the frequency for the categorical values for selected column.

- a. True
- b. False

30. Summary function is commonly used function to print the summary of numeric, categorical, regression, time series data.

- a. False
- b. True

31. Identify the software name which used to write python code and projects.
- a. RStudio
  - b. Jupyter Notebook
  - c. Notepads
  - d. None
32. \_\_\_\_\_ is most common function to display the value of variable/object.
- a. View
  - b. print
  - c. display
  - d. get
33. Select the quotations support by Python programming
- a. Single (')
  - b. Double (")
  - c. Triple (""" or """)
  - d. All of the above
34. \_\_\_\_\_ symbol represent Python comment, and comment is used for documentation and annotation of Python code.
- a. +
  - b. /\* \*/
  - c. #
  - d. None
35. To download spacy library, select the right statements
- a. pip install spacy
  - b. pip download spacy
  - c. pip update spacy
  - d. None
36. \_\_\_\_\_ (if..else) statement used for checking condition with Python.
- a. Conditional statement
  - b. Control Statement
  - c. Loop statement
  - d. None
37. \_\_\_\_\_ keyword used to define functions in Python.
- a. If
  - b. def
  - c. fun
  - d. None

38. \_\_\_\_\_ data structure is immutable.
- a. Tuple
  - b. List
  - c. Dictionary
  - d. None
39. Set data structures always contains unique elements.
- a. False
  - b. True
40. \_\_\_\_\_ data structures contain element in form of keys: values.
- a. List
  - b. Dictionary
  - c. DataFrame
  - d. None
41. \_\_\_\_\_ is the process of transforming unstructured text into a structured format to identify meaningful patterns and new insights.
- a. Data mining
  - b. Text Mining
  - c. File Mining
  - d. Deep Mining
42. \_\_\_\_\_ Python libraries used to perform text analysis
- a. Spacy
  - b. Pandas
  - c. Numpy
  - d. None
43. Function which used to display summary of Pandas DataFrame
- a. summary
  - b. describe
  - c. head
  - d. None
44. To view no of rows & columns for Pandas data, which is best option to get (assume data is panda DataFrame)
- a. data.shape
  - b. print(data)
  - c. data
  - d. None

45. A \_\_\_\_\_ is a sequence of observations over a certain period.
- a. Theano
  - b. Scikit
  - c. Time Series
  - d. Tensor
46. \_\_\_\_\_ plot used to display analysis of two or more variables.
- a. Bar graph
  - b. Pair plot
  - c. Histogram
  - d. Line chart
47. \_\_\_\_\_ function used to rename the columns in pandas DataFrame, and always used axis=1 while performing operations.
- a. colname
  - b. rename
  - c. modify
  - d. alias
48. \_\_\_\_\_ functions used to get the summary, min, max, sum...etc
- a. Aggregate
  - b. Sum
  - c. Mix
  - d. Max
49. Most common function to display top 40 records in Python.
- a. tail(40)
  - b. display(40)
  - c. head(40)
  - d. print(40)
50. \_\_\_\_\_ functions to import comma separated values files in Python.
- a. read\_csv
  - b. read\_excel
  - c. read.csv
  - d. None