

BDTM: Big Data Tools for Managers

1. Big data is also changing at a faster rate,
 - a. **False**
 - b. True
2. Text Mining/Analysis can be used in:
 - a. Detecting spam model
 - b. Predicting stock Movements
 - c. News stories categorization
 - d. **All the above**
3. Semi-structured data, also known as having a schema-less or self-describing structure
 - a. False
 - b. **True**
4. Select wrong Elements of Big Data
 - a. Volume
 - b. Velocity
 - c. Variety
 - d. **VolPro**
5. _____ 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
6. In which database, data is a blend between structured and unstructured data formats?
 - a. Full-structured data
 - b. Partial-structured data
 - c. **Semi-structured data**
 - d. Uni-structured data
7. The process of breaking out long-form text into sentences and words called?
 - a. Stem
 - b. Cluster
 - c. Bag
 - d. **Tokens**
8. Text mining is being used by large media companies, to clarify information and to provide readers with greater search experiences.
 - a. **TRUE**
 - b. FALSE
 - c. Can be true or false
 - d. Can not say
9. Typical text mining tasks include?
 - a. text categorization
 - b. text clustering
 - c. entity relation modeling
 - d. **All of the above**

BDTM: Big Data Tools for Managers

10. Stemming: This refers to the process of separating the prefixes and suffixes from words to derive the root word form and meaning.
- a. **TRUE**
 - b. FALSE
 - c. Can be true or false
 - d. Can not say
11. Most of the data in the world is in fact unstructured
- a. False
 - b. **True**
12. The structured and unstructured data that is collected and analyzed in social media analytics programs includes all of the following, EXCEPT:
- a. Comments/reviews
 - b. Demographic information
 - c. **Transaction data**
 - d. Number of likes and followers
13. Which of the following techniques can be used for the purpose of keyword normalization, the process of converting a keyword into its base form(Multiple Choice)?
- a. **Lemmatization**
 - b. Levenshtein
 - c. **Stemming**
 - d. Soundex
14. _____ Python libraries used to perform text analysis
- a. **Spacy**
 - b. Pandas
 - c. Numpy
 - d. None
15. Tokenization refers to segmenting text into words, punctuations marks, numbers..etc
- a. **True**
 - b. False
16. Select the right statement to install spacy library
- a. pip download spacy
 - b. **pip install spacy**
 - c. download spacy
 - d. None
17. Big data refers to collection small and strictly structure type data.
- a. True
 - b. **False**

BDTM: Big Data Tools for Managers

18. Select the sources of data in Big Data eco systems (Multiple Choice).
- a. In/Out
 - b. Internal
 - c. External
 - d. None
19. Structured data can be defined as the data that has a defined repeating pattern.
- a. True
 - b. False
20. Is a new data challenge that requires leveraging existing systems differently, then we classify data as Big Data.
- a. True
 - b. False
21. Big Data usually unstructured and ~70-80% data available in unstructured format.
- a. False
 - b. True
22. Velocity is the speed in which data is process and becomes accessible.
- a. False
 - b. True
23. Variability refers to the data which keeps on changing constantly.
- a. True
 - b. False
24. Variety describes one of the biggest challenges of _____.
- a. Big data
 - b. Data science
 - c. Data integration
 - d. None of the mentioned above
25. The data that can be processed, stored, and retrieved in a fixed format called _____,
- a. Structured Data
 - b. Unstructured Data
 - c. Semi-structured Data
 - d. None of the mentioned above
26. Semi-structured data, also known as having a schema-less or self-describing structure
- a. False
 - b. True
27. Select correct Elements of Big Data
- a. Volume
 - b. Velocity
 - c. Variety
 - d. All

BDTM: Big Data Tools for Managers

28. Text mining is process of transforming free of textual data into a structure format.
- a. False
 - b. True
29. In which database, data is a blend between structured and unstructured data formats?
- a. Full-structured data
 - b. Partial-structured data
 - c. Semi-structured data
 - d. Uni-structured data
30. The process of breaking out long-form text into sentences and words called?
- a. Stem
 - b. Cluster
 - c. Bag
 - d. Tokens
31. Text mining is being used by large media companies, to clarify information and to provide readers with greater search experiences.
- a. TRUE
 - b. FALSE
 - c. Can be true or false
 - d. Can not say
32. Typical text mining tasks include?
- a. text categorization
 - b. text clustering
 - c. entity relation modeling
 - d. All of the above
33. Stemming: This refers to the process of separating the prefixes and suffixes from words to derive the root word form and meaning.
- a. TRUE
 - b. FALSE
 - c. Can be true or false
 - d. Can not say
34. Most of the data in the world is in fact unstructured
- a. False
 - b. True
35. The structured and unstructured data that is collected and analyzed in social media analytics programs includes all of the following, EXCEPT:
- a. Comments/reviews
 - b. Demographic information
 - c. Transaction data
 - d. Number of likes and followers

BDTM: Big Data Tools for Managers

36. _____ Python libraries used to perform text analysis
- a. **Spacy**
 - b. Pandas
 - c. Numpy
 - d. None
37. Tokenization refers to segmenting text into words, punctuations marks, numbers..etc
- a. **True**
 - b. False
38. Select the right statement to install spacy library
- a. pip download spacy
 - b. **pip install spacy**
 - c. download spacy
 - d. None
39. Big data refers to collection large data and strictly structure type data.
- a. True
 - b. **False**
40. Select the sources of data in Big Data eco systems (Multiple Choice).
- a. In/Out
 - b. **Internal**
 - c. **External**
 - d. None
41. Structured data can be defined as the data that has a defined repeating pattern.
- a. **True**
 - b. False
42. Is a new data challenge that requires leveraging existing systems differently, then we classify data as Big Data.
- a. **True**
 - b. False
43. Big Data usually in structured format and ~70-80% data available in structured format.
- a. **False**
 - b. True
44. SQL Stands for
- a. Structured Language
 - b. Simple Query Language
 - c. **Structured Query Language**
 - d. None

BDTM: Big Data Tools for Managers

45. Which data manipulation command is used to combines the records from one or more tables?
- a. SELECT
 - b. PROJECT
 - c. PRODUCT
 - d. JOIN
46. The command used to delete a particular record in a Table
- a. UPDATE TABLE
 - b. DELETE FROM
 - c. TRUNCATE COLUMN
 - d. DROP
47. Which of the following is a legal expression in SQL?
- a. SELECT NULL FROM SALES;
 - b. SELECT * FROM SALES;
 - c. SELECT * FROM SALES WHEN PRICE = NULL;
 - d. SELECT # FROM SALES;
48. The virtual table that its created by data from the result of an SQL 'Select' statement is called
- a. View
 - b. Synonym
 - c. Sequence
 - d. Transaction
49. Which statement used to update/change the value of any columns
- a. MODIFY statement
 - b. UPDATE statement
 - c. ALTER statement
 - d. None
50. Which of the following commands can be used to delete a tables/relation if it is not needed anymore?
- a. UPDATE
 - b. ALTER
 - c. DROP
 - d. DELETE
51. Having clause can be use without GROUP BY
- a. TRUE
 - b. FALSE
52. Which of the following represents basic SQL DML statements
- a. Drop, Update, Delete
 - b. Insert, Alter, Delete
 - c. Insert, View, Create
 - d. Insert, Update, Delete
53. WHERE & HAVING clause, both used to filter out the data
- a. TRUE
 - b. FALSE

BDTM: Big Data Tools for Managers

54. Numbers, Strings and DateTime are most common data types in SQL
- a. **TRUE**
 - b. FALSE
55. In SQL – the function AVG, MIN, MAX, SUM, COUNT are called as _____
- a. adjunct function
 - b. set operation
 - c. scaler operation
 - d. **aggregate function**
56. Statement used to get unique records from the Table
- a. UNIQUE
 - b. **DISTINCT**
 - c. UNION
 - d. None
57. Tuples & Rows are same in Table
- a. **TRUE**
 - b. FALSE
58. Statement used to display table structures in MySQL
- a. **DESCRIBE**
 - b. STR
 - c. DISPLAY
 - d. None
59. DELETE statement without where clause delete all the records from Table
- a. **TRUE**
 - b. FALSE
60. Which join used get the common records from both the tables (table 1 & table 2).
- a. LEFT JOIN
 - b. **INNER JOIN**
 - c. CROSS JOIN
 - d. None
61. _____ operator is used to combine the result-set of two or more SELECT statements.
- a. **UNION**
 - b. JOIN
 - c. MERGE
 - d. None
62. HAVING clause is equivalent to WHERE clause but HAVING used with only GROUP BY clause.
- a. FALSE
 - b. **TRUE**

BDTM: Big Data Tools for Managers

63. _____ operators used to specify multiple possible values for a column while defining condition with WHERE clause.
- a. LIKE
 - b. **IN**
 - c. AND
 - d. OR
64. Which statement used to delete all the records from table
- a. DELETES
 - b. **TRUNCATE**
 - c. REMOVE
 - d. DROP
65. UPDATE statement without WHERE clause update all records for selected columns.
- a. FALSE
 - b. **TRUE**
66. SQL Stands for
- a. Structured Language
 - b. Simple Query Language
 - c. **Structured Query Language**
 - d. None
67. Asterisk (*) in SELECT statement indicates that Query will bring only few columns from the Table
- a. **FALSE**
 - b. TRUE
68. The command used to delete a particular record in a Table
- a. UPDATE TABLE
 - b. **DELETE FROM**
 - c. TRUNCATE COLUMN
 - d. DROP
69. Which of the following is a legal expression in SQL?
- a. SELECT NULL FROM students;
 - b. **SELECT * FROM students;**
 - c. SELECT * FROM students WHEN marks = NULL;
 - d. SELECT # FROM students;
70. The virtual table that its created by data from the result of an SQL 'Select' statement is called
- a. **View**
 - b. Synonym
 - c. Sequence
 - d. Transaction

BDTM: Big Data Tools for Managers

71. Which statement used to INSERT the value for Table
- a. **INSERT statement**
 - b. TRIGGER statement
 - c. ALTER statement
 - d. None
72. Which of the following commands can be used to delete a tables/relation if it is not needed anymore?
- a. UPDATE
 - b. ALTER
 - c. **DROP**
 - d. DELETE
73. Numeric data can be inserted without Single/Double quotes with INSERT statement.
- a. **TRUE**
 - b. FALSE
74. Which of the following represents basic SQL statements?
- a. Drop, Update, Delete
 - b. Insert, Alter, Delete
 - c. Insert, View, Create
 - d. **Select, Insert, Update, Delete**
75. WHERE & HAVING clause, both used to filter out the data
- a. **TRUE**
 - b. FALSE
76. Numbers, Strings and DateTime are most common data types in SQL
- a. **TRUE**
 - b. FALSE
77. In SQL – the function AVG, MIN, MAX, SUM, COUNT are called as _____
- a. adjunct function
 - b. set operation
 - c. scaler operation
 - d. **aggregate function**
78. Statement used to get unique records from the Table
- a. UNIQUE
 - b. **DISTINCT**
 - c. UNION
 - d. None
79. Columns & Attributes are same in Table
- a. **TRUE**
 - b. FALSE

BDTM: Big Data Tools for Managers

80. Statement used to display table structures in MySQL
- a. **DESCRIBE**
 - b. STR
 - c. DISPLAY
 - d. None
81. DELETE statement without where clause delete all the records from Table
- a. **TRUE**
 - b. FALSE
82. Which join used get all the records from right table and matching records from the left table (table 1 & table 2).
- a. LEFT JOIN
 - b. **RIGHT JOIN**
 - c. CROSS JOIN
 - d. None
83. _____ operator is used to combine the result-set of two or more SELECT statements and duplicates allowed.
- a. **UNION ALL**
 - b. JOIN
 - c. MERGE
 - d. None
84. HAVING clause is equivalent to WHERE clause but HAVING used with only GROUP BY clause.
- a. FALSE
 - b. **TRUE**
85. _____ operators used to specify multiple possible values for a column while defining condition with WHERE clause.
- a. LIKE
 - b. **IN**
 - c. AND
 - d. OR
86. Which statement used to delete all the records from table
- a. DELETES
 - b. **TRUNCATE**
 - c. REMOVE
 - d. DROP
87. UPDATE statement without WHERE clause update all records for selected columns.
- a. FALSE
 - b. **TRUE**

BDTM: Big Data Tools for Managers

88. Function used to load the package in memory/project of R is
- import()
 - load.packages()
 - library()**
 - None
89. install() is used to install the package in the memory/R environment?
- True
 - False**
90. R Studio has basically _____ window sessions.
- 2/Two
 - 4/Four**
 - 3/Three
 - 1/One
91. Function used to get the project working directory/folder?
- cd()
 - pwd()
 - show()
 - getwd()**
92. function used to create vectors in R?
- t()
 - c()**
 - class()
 - None
93. Which command is correct to create numeric (1,2,3,4,5,6) vectors in R
- c('1','2','3','4','5','6')
 - c(1,2,3,4,5,6)**
 - c(1L,2L,3L,4L,5L,6L)
 - None
94. Function used show the class of basic data types?
- print()
 - class()**
 - show()
 - c()
95. Operator used to create range of values i.e. 1:10?
- Colon (:) operator**
 - Special operator
 - Semicolon(;) operator
 - None
96. Function used to create sequence of number with specified increment/decrement values?
- seq(from=, to=, by=)**
 - seq(from=, to=, length.out=)
 - seq(from=, to=, increment_by=)
 - None
97. Special variable/function **letters** used to get the small letter alphabets vectors directly?
- True**
 - False

BDTM: Big Data Tools for Managers

98. Let's assume `num <- c(10,20,30,40,50)` is numeric vector, need to access the 2nd element from vector which statement is correct?

- a. `num[2]`
- b. `num[1:2]`
- c. `num[]`
- d. None

99. Function used to sort the vector elements?

- a. `c()`
- b. `sort()`
- c. `print()`
- d. None

100. Function used to get absolute values of the vectors?

- a. `print()`
- b. `abs()`
- c. `prod()`
- d. `sum()`

101. Operators to check the elementwise conditions in R?

- a. Condition
- b. Relational operators
- c. Filter
- d. None

102. Matrix data structures has only one indexes?

- a. True
- b. False

103. Function used to get the dimension of Matrix/DataFrame?

- a. `nrow()`
- b. `ncol()`
- c. `dim()`
- d. None

104. Operator used to access the variable/column of DataFrame?

- a. `$(dollar)`
- b. `#{hash}`
- c. `.(dot)`
- d. None

105. Matrix element are arranged on column basis by specifying `byrow=TRUE`?

- a. True
- b. False

106. Combining two same matrixing by row using `rbind()` function in R?

- a. True
- b. False

BDTM: Big Data Tools for Managers

107. List is collections of all types of data & data structures in R?

- a. **True**
- b. False

108. True Matrix multiplication perform with special operator

- a. **%*%**
- b. %/%
- c. %in%
- d. None

109. Function used to transpose of the Matrix/DataFrame in R?

- a. **t()**
- b. T()
- c. c()
- d. None

110. Function used to read comma separated values in R?

- a. **read.csv()**
- b. read_csv()
- c. read.xls()
- d. None

111. Function to display the structure of DataFrame/List?

- a. **str()**
- b. summary()
- c. describe()
- d. None

112. Function to get quick summary of DataFrame?

- a. describe()
- b. **summary()**
- c. table()
- d. None

113. Function to split the Dataframe based on specified column name?

- a. **split()**
- b. subset()
- c. get()
- d. None

114. tail() functions to display records/observation from Top of the dataframe?

- a. True
- b. **False**

115. The Following function is used to print first 6 records of the dataset

- a. **head**
- b. tail
- c. start
- d. initial

BDTM: Big Data Tools for Managers

116. Function used to display the value of any variables (Multiple Selection)?

- a. **print()**
- b. **View()**
- c. display()
- d. get()

117. Python is a general purpose interpreted, and high-level programming language.

- a. **True**
- b. False

118. Python statement looks like English, this indicates python code is more readable and looks like English statement.

- a. False
- b. **True**

119. Python is only available for Windows PC/Laptop operating system.

- a. **False**
- b. True

120. Identify the software name which used to write python code and projects.

- a. RStudio
- b. **Jupyter Notebook**
- c. MS Word
- d. None

121. Python reserved words that cannot be used for variable name/ function name and reserved words are in lowercase always.

- a. False
- b. **True**

122. Identify invalid variable names

- a. **2myvar = "John"**
- b. Myvar = "John"
- c. MYVAR = "John"
- d. None

123. _____ function used to print the value of variables in Python

- a. **Print**
- b. Display
- c. View
- d. None

124. Python used _____ approach to indicate the block of codes.

- a. **Indentation**
- b. Brackets
- c. Block
- d. None

125. Select the quotations supported by Python

- a. Single

BDTM: Big Data Tools for Managers

- b. Double
 - c. Triple
 - d. All of the above
- 126. _____ symbol represents comment line in Python
 - a. +
 - b. /* */
 - c. #
 - d. None
- 127. To download pandas library, select the right statements
 - a. pip install pandas
 - b. pip download pandas
 - c. pip update pandas
 - d. None
- 128. _____ functions to check type of data variable is holding
 - a. type()
 - b. check()
 - c. print()
 - d. None
- 129. _____ (if..else) statement used for checking condition with Python.
 - a. Conditional statement
 - b. Control Statement
 - c. Loop statement
 - d. None
- 130. _____ keyword used to define functions in Python.
 - a. If
 - b. def
 - c. fun
 - d. None
- 131. Select common data structures available in Python (Multiple selection)
 - a. Matrix
 - b. List
 - c. Tuple
 - d. Sets
- 132. _____ data structure is immutable.
 - a. Tuple
 - b. List
 - c. Dictionary
 - d. None
- 133. Set data structures always contains unique elements.
 - a. False
 - b. True

BDTM: Big Data Tools for Managers

Plus(+) operators used to combine elements of _____ data structures.

- c. Set
- d. Dictionary
- e. **List**
- f. None

134. Len function is common function to find the total length of elements in data structures.

- a. False
- b. **True**

135. _____ data structures contain element in form of keys: values.

- a. List
- b. **Dictionary**
- c. DataFrame
- d. None

136. _____ approach to access elements from List & Tuple data structures.

- a. **Indexing**
- b. Columns
- c. Name
- d. None

137. _____ function used to combine two set.

- a. Append
- b. Combine
- c. **Union**
- d. None

138. Identify the software name which used to write python code and projects.

- a. RStudio
- b. **Jupyter Notebook**
- c. MS Word
- d. None

139. Python reserved words that cannot be used for variable name/ function name and reserved words are in lowercase always.

- a. False
- b. **True**

140. Identify invalid variable names

- a. **My-var = "John"**
- b. Myvar = "John"
- c. MYVAR = "John"
- d. None

141. Python used _____ approach to indicates the block of codes.

- a. **Indentation**
- b. Brackets
- c. Block
- d. None

BDTM: Big Data Tools for Managers

142. Select the quotations support by Python

- a. Single
- b. Double
- c. Triple
- d. **All of the above**

143. _____ symbol represents comment line in Python

- a. +
- b. /* */
- c. **#**
- d. None

144. To download pandas library, select the right statements

- a. **pip install sklearn**
- b. pip download sklearn
- c. pip update sklearn
- d. None

145. _____ functions to check type of data variable is holding

- a. **type()**
- b. check()
- c. print()
- d. None

146. _____ statement used for writing conditional statement with Python.

- a. **if..else**
- b. For loop
- c. def
- d. None

147. _____ keyword used to define functions in Python.

- a. If
- b. **def**
- c. fun
- d. None

148. Select common data structures available in Python (Multiple selection)

- a. Matrix
- b. **List**
- c. **Tuple**
- d. **Sets**

149. _____ data structure is mutable (Multiple selection).

- a. Tuple
- b. **List**
- c. **Dictionary**
- d. **Set**

BDTM: Big Data Tools for Managers

150. Set data structures always contain unique elements.

- a. False
- b. **True**

151. Plus(+) operators are used to combine elements of _____ data structures.

- a. Set
- b. Dictionary
- c. **List**
- d. None

152. Len function is a common function to find the total length of elements in data structures.

- a. False
- b. **True**

153. _____ data structures contain elements in the form of keys: values.

- a. List
- b. **Dictionary**
- c. DataFrame
- d. None

154. _____ approach to access elements from List & Tuple data structures.

- a. **Indexing**
- b. Columns
- c. Name
- d. None

155. _____ function is used to add or update values in Dictionary.

- a. **update**
- b. Combine
- c. Union
- d. None

156. _____ brackets are used to create set data structures.

- a. **{ }**
- b. []
- c. ()
- d. None

157. Data Structures/Collections are useful containers to store and manipulate lists of homogeneous or heterogeneous elements.

- a. False
- b. **True**

158. _____ statement is used to add library/package in Python code.

- a. Load
- b. **import**
- c. Add
- d. None

159. Variable names can only contain alphanumeric characters and underscores (A-z, 0-9 and _)

- a. False
- b. **True**

BDTM: Big Data Tools for Managers

===== Set-2 =====

1. _____ is a collection of data that is used in volume, yet growing exponentially with time
 - a. Big databases
 - b. Big DBMS
 - c. Big Data Files
 - d. **Big Data**
2. Which of the following are the Benefits of Big Data Processing?
 - a. Businesses can utilize outside intelligence while taking decision
 - b. Better operational efficiency
 - c. Improve customer server
 - d. **All of the above**
3. Big data analysis does the following except?
 - a. Spread data
 - b. **Analyze data**
 - c. Organizes data
 - d. Collect data
4. Which of the following is true about big data?
 - a. Big data can be processed using traditional techniques
 - b. **Big data refers to data sets that are at least a petabyte in size**
 - c. Big data analysis does not involve reporting and data mining techniques
 - d. Big data has low velocity meaning that it is generated slowly
5. How many V's of Big Data
 - a. 2
 - b. 3
 - c. 5
 - d. **4**
6. Choose the primary characteristics of big data among the following
 - a. Value
 - b. Variety
 - c. Volume
 - d. **All of the above**
7. Database Management System is software used to store and retrieve the databases.
 - a. **True**
 - b. False
8. Which is DBMS software used to manipulated data & databases.
 - a. **MySQL**
 - b. SQL
 - c. Data Studio
 - d. None

BDTM: Big Data Tools for Managers

9. Identify correct Full form of RDBMS
 - a. Row Database Management System
 - b. Relational Database Management System**
 - c. Right Database Management System
 - d. None
10. Records are also known as Tuples in RDBMS
 - a. True**
 - b. False
11. Which join used get all the records from right table and matching records from the left table (table 1 & table 2).
 - a. LEFT JOIN
 - b. RIGHT JOIN**
 - c. CROSS JOIN
 - d. None
12. _____ operator is used to combine the result-set of two or more SELECT statements and duplicates allowed.
 - a. UNION ALL**
 - b. JOIN
 - c. MERGE
 - d. None
13. Which statement used to insert the value for a Table
 - c. INSERT statement**
 - d. TRIGGER statement
 - e. ALTER statement
 - f. None
14. Which join used get the common records from both the tables (table 1 & table 2).
 - a. LEFT JOIN
 - b. INNER JOIN**
 - c. CROSS JOIN
 - d. None
15. Numeric data can be inserted without Single/Double quotes with INSERT statement.
 - a. TRUE**
 - b. FALSE
16. Which statement used to delete all the records from table
 - a. DELETES
 - b. TRUNCATE**
 - c. REMOVE

BDTM: Big Data Tools for Managers

- d. DROP
- 17. Which of the following represents four basic SQL statements?
 - a. Drop, Update, Delete
 - b. Insert, Alter, Delete
 - c. Insert, View, Create
 - d. **Select, Insert, Update, Delete**
- 18. DELETE statement without where clause delete all the records from Table
 - a. **TRUE**
 - b. FALSE
- 19. _____ operator is used to combine the result-set of two or more SELECT statements and duplicates allowed.
 - a. **UNION ALL**
 - b. JOIN
 - c. MERGE
 - d. None
- 20. _____ operators used to specify multiple possible values for a column while defining condition with WHERE clause.
 - a. LIKE
 - b. **IN**
 - c. AND
 - d. OR
- 21. R Studio has basically _____ window sessions.
 - a. 2/Two
 - b. **4/Four**
 - c. 3/Three
 - d. 1/One
- 22. Function used show the value of variable.
 - a. **print()**
 - b. class()
 - c. show()
 - d. c()
- 23. Operator used to create range of values i.e. 11:20?
 - a. **Colon (:) operator**
 - b. Special operator
 - c. Semicolon(;) operator
 - d. None
- 24. True Matrix multiplication perform with special operator
 - a. **%*%**
 - b. %/%
 - c. %in%

BDTM: Big Data Tools for Managers

- d. None
- 25. Function used to display the data in Excel like format
 - a. `print()`
 - b. `display()`
 - c. **`View()`**
 - d. None
- 26. Operator used to access the variable/column of DataFrame?
 - a. **`$(dollar)`**
 - b. `$(hash)`
 - c. `.(dot)`
 - d. None
- 27. Function used to transpose of the Matrix/DataFrame in R?
 - a. **`t()`**
 - b. `T()`
 - c. `c()`
 - d. None
- 28. Functions used to count the frequency of Factor/Character data?
 - a. **`table()`**
 - b. `freq()`
 - c. `count()`
 - d. Nonw
- 29. Function to get quick summary of DataFrame?
 - a. `describe()`
 - b. **`summary()`**
 - c. `table()`
 - d. None
- 30. Function to subset the data from Dataframe based on conditions?
 - a. `split()`
 - b. **`subset()`**
 - c. `get()`
 - d. None
- 31. which command is correct to create Integer (10,20,30,40,50,60) vectors in R
 - e) `c('10','20','30','40','50','60')`
 - f) `c(10,20,30,40,50,60)`
 - g) **`c(10L,20L,30L,40L,50L,60L)`**
 - h) None

BDTM: Big Data Tools for Managers

32. Function used to read comma separated values in R?
- a. **read.csv()**
 - b. read_csv()
 - c. read.xls()
 - d. None
33. To define comments in Python which operator/symbol used
- a. //
 - b. **#**
 - c. /*...*/
 - d. None
34. Select the Quotation which Python supports (Multiple selection)
- a. '
 - b. "
 - c. **""" or '''**
 - d. " " "
35. Python used Indentation to define a code block, like other programming used {} (brackets)
- a. **True**
 - b. False
36. Read Excel file in Pandas, which functions used to read file and store as DataFrame
- a. **read_excel**
 - b. read.excel
 - c. read_file
 - d. None
37. Python reserved words are basic building blocks of the Python programming language
- a. False
 - b. **True**
38. A variable name can only contain alpha-numeric characters and underscores (A-z, 0-9, and _)
- a. **True**
 - b. False
39. Identify the valid statement for installing pandas packages.
- a. **pip install pandas**
 - b. python install pandas
 - c. pip setup pandas
 - d. None
40. Function which used to display summary of Pandas DataFrame
- a. summary
 - b. **describe**
 - c. head
 - d. None

BDTM: Big Data Tools for Managers

41. `value_counts()` function gives the frequency count of any selected columns from Pandas DataFrame.
- a. **True**
 - b. False
42. In Python, Dictionaries are immutable
- a. **False**
 - b. True
43. Set data structure allow to store unique values in Python.
- a. False
 - b. **True**
44. Most common used graphics library to display graphs with Pandas DataFrame
- a. Seaborn
 - b. Ploty
 - c. **Matplotlib**
 - d. None
45. What arithmetic operators cannot be used with Python strings?
- a. +
 - b. **-**
 - c. *
 - d. All of them mentioned
46. Strings are immutable in Python, which means a string cannot be modified.
- a. **True**
 - b. False
47. The process of breaking out long-form text into sentences and words called?
- a. Stem
 - b. Cluster
 - c. Bag
 - d. Bag
 - e. **Tokens**
48. Identify the most common libraries used for performing text mining/analysis in Python? (Multiple choice)
- a. NLTK
 - b. Spacy
 - c. Sklearn
 - d. **All of the above**
49. To return the length of string `s` what command do we execute? (assume `str` is string variable)
- a. **`str.len()`**
 - b. `len(str)`
 - c. `size(str)`
 - d. `str.size()`

BDTM: Big Data Tools for Managers

50. Write the output of the following code:

```
>> L = [1,2,3,4,5,[6,7,8]]  
>> print(L[5])
```

- a. **[6,7,8]**
- b. Error
- c. 6,7,8
- d. 5

51. How many V's of Big Data

- a. 2
- b. 3
- c. 5
- d. **4**

52. Data in _____ bytes size called big data

- a. **Peta**
- b. Giga
- c. Tera
- d. Meta

53. Total forms of big data is _____

- a. 1
- b. 2
- c. **3**
- d. 4

54. Choose the primary characteristics of big data among the following

- a. Value
- b. Variety
- c. Volume
- d. **All of the above**

55. Point out the wrong statement:

- a. **Non-Relational databases require that schemas be defined before you can add data**
- b. NoSQL databases are built to allow the insertion of data without a predefined schema.
- c. NewSQL database are built to allow the insertion of data without a predefined schema.
- d. All of the options.

56. Volume is one of the characteristics of big data. What does Volume refer to?

- a. The hard disk or server capacity it can hold
- b. The data that can be processed
- c. The number of YouTube/Netflix videos that existed
- d. **The amount of data in variety of formats**

57. Data is a collection of a _____

- a. Large Textual contents of files
- b. Various files in Folders
- c. **Small piece of information**
- d. None

BDTM: Big Data Tools for Managers

58. Database is an organized collection of data so that it can be easily manage and accessible.
- a. **True**
 - b. False
59. DBMS stands for
- a. **Database Management System**
 - b. Databases Management Software
 - c. Data Manipulation Software
 - d. None
60. Columns/Attributes/Field all are same in RDBMS
- a. **True**
 - b. False
61. Which data manipulation command is used to combines the records from one or more tables?
- a. SELECT
 - b. PROJECT
 - c. PRODUCT
 - d. **JOIN**
62. Having clause can be use without GROUP BY
- a. True
 - b. **False**
63. In SQL – the function AVG, MIN, MAX, SUM, COUNT are called as_____
- a. Adjunct function
 - b. Set operation
 - c. Scaler operation
 - d. **Aggregate function**
64. Statement used to display table structures in MySQL
- a. **DESCRIBE**
 - b. STR
 - c. DISPLAY
 - d. None
65. Which join used get the common records from both the tables (table 1 & table 2).
- a. LEFT JOIN
 - b. **INNER JOIN**
 - c. CROSS JOIN
 - d. None
66. HAVING clause is equivalent to WHERE clause but HAVING used with only GROUP BY clause.
- a. FALSE
 - b. **TRUE**

BDTM: Big Data Tools for Managers

67. Which statement used to delete all the records from table
- DELETES
 - TRUNCATE**
 - REMOVE
 - DROP
68. Which of the following is a legal expression in SQL?
- SELECT NULL FROM Employee;
 - SELECT * FROM Employee;**
 - SELECT , FROM Employee WHEN City = NULL;
 - SELECT # FROM Employee;
69. DELETE statement without where clause delete all the records from Table
- TRUE**
 - FALSE
70. Statement used to get unique records from the Table
- UNIQUE
 - DISTINCT**
 - UNION
 - None
71. R Studio has basically _____ window sessions.
- 2/Two
 - 4/Four**
 - 3/Three
 - 1/One
72. Function used to create sequence of number with specified increment/decrement values?
- seq(from=, to=, by=)**
 - seq(from=, to=, length.out=)
 - seq(from=, to=, increment_by=)
 - None
73. Function used to get absolute values of the vectors?
- print()
 - abs()**
 - prod()
 - sum()
74. Operator used to access the variable/column of DataFrame?
- \$(dollar)**
 - #{hash}
 - .(dot)
 - None
75. Function used to transpose of the Matrix/DataFrame in R?
- t()**
 - T()

BDTM: Big Data Tools for Managers

- c. `c()`
 - d. None
76. Function to display the structure of DataFrame/List?
- a. **`str()`**
 - b. `summary()`
 - c. `describe()`
 - d. None
77. Function to get quick summary of DataFrame?
- a. `describe()`
 - b. **`summary()`**
 - c. `table()`
 - d. None
78. `tail()` functions to display records/observation from Top of the DataFrame?
- a. True
 - b. **False**
79. which command is correct to create numeric (1,2,3,4,5,6) vectors in R
- i) `c('1','2','3','4','5','6')`
 - j) **`c(1,2,3,4,5,6)`**
 - k) `c(1L,2L,3L,4L,5L,6L)`
 - l) None
80. Function used to read comma separated values in R?
- a. **`read.csv()`**
 - b. `read_csv()`
 - c. `read.xls()`
 - d. None
81. Python is a general-purpose interpreted, interactive, object-oriented, and high-level programming language
- a. **True**
 - b. False
82. Select the Quotation which Python supports (Multiple selection)
- a. Single
 - b. Double
 - c. **Triple**
 - d. Four
83. Which of the following is used to define a block of code in Python language?
- a. **Indentation**
 - b. Key
 - c. Brackets
 - d. All of the mentioned

BDTM: Big Data Tools for Managers

84. To define comments in Python which operator/symbol used
- a. //
 - b. #**
 - c. /*...*/
 - d. None
85. Read CSV file in Pandas, which functions used to read file and store as DataFrame
- a. read_csv**
 - b. read.csv
 - c. read_file
 - d. None
86. Python reserved words can not be used as Variable/Functions names
- a. False
 - b. True**
87. A variable name can only contain alpha-numeric characters and underscores (A-z, 0-9, and _)
- a. True**
 - b. False
88. _____ is package manager in Python to download & install Python packages.
- a. download
 - b. pip**
 - c. install
 - d. None
89. Function which used to display summary of Pandas DataFrame
- a. summary
 - b. describe**
 - c. head
 - d. None
90. _____ Function to get the frequency of categorical/labeled variable.
- a. count
 - b. freq
 - c. value_counts**
 - d. None
91. 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
92. In Python, Dictionaries are immutable
- a. False**
 - b. True

BDTM: Big Data Tools for Managers

93. What will be the output of the following code snippet?

```
a = [1, 2, 3]
a = tuple(a)
a[0] = 2
print(a)
```

- a. [2,2,3]
- b. (2,2,3)
- c. (1,2,3)
- d. Error**

94. To visualize graphs in Pandas which is most common library used

- a. Ploty
- b. Matplotlib**
- c. Seaborn
- d. None

95. What arithmetic operators cannot be used with Python strings?

- a. +
- b. -**
- c. *
- d. All of them mentioned

96. Strings are immutable in Python, which means a string cannot be modified.

- a. True**
- b. False

97. A _____ is a sequence of observations over a certain period.

- a. Theano
- b. Scikit
- c. Time Series**
- d. Tensor

98. The process of breaking out long-form text into sentences and words called?

- a. Stem
- b. Cluster
- c. Bag
- d. Bag
- e. Tokens**

99. Typical text mining tasks include?

- a. Text Categorization
- b. Text Clustering
- c. Entity Relation Modelling
- d. All of the above**

BDTM: Big Data Tools for Managers

100. To return the length of string `s` what command do we execute? (assume `s` is string variable)

- a. `s.len()`
- b. `len(s)`
- c. `size(s)`
- d. `s.size()`

101. In the early 60s, technology witnessed problem with velocity or real-time data assimilation. This inspired the evolution of _____ technology.

- a. Database
- b. Files
- c. Big Data
- d. None

102. Is Big Data usually unstructured and qualitative in nature?

- a. True
- b. False

103. Data in _____ bytes size called big data

- a. Peta (1 PB = 1000TB)
- b. Giga (1 GB = 1000MB)
- c. Tera (1 TB = 1000GB)
- d. Mega Byte (1 MB = 1000KB)

104. What kind of data big data can manage _____ (Multiple Choice)?

- a. Structured
- b. Unstructured
- c. Semi-structured
- d. Mix-data

105. Database is an organized collection of data so that it can be easily manage and accessible.

- a. True
- b. False

106. Choose the primary characteristics of big data among the following

- a. Value
- b. Variety
- c. Volume
- d. All of the above

107. Volume is one of the characteristics of big data. What does Volume refer to?

- a. The hard disk or server capacity it can hold
- b. The data that can be processed
- c. The number of YouTube/Netflix videos that existed
- d. The amount of data in variety of formats

BDTM: Big Data Tools for Managers

108.To define Data, which statement is right from option?

- a. Large Textual contents of files
- b. **Small piece of information**
- c. Various files in Folders
- d. None

109.RDBMS stands for

- a. **Relational Database Management System**
- b. Raw Databases Management System
- c. Relation Data Manipulation Software
- d. None

110.Columns/Attributes/Field all are same in RDBMS

- a. **True**
- b. False

111._____ tools & software used to write database related programming.

- a. **MySQL**
- b. Notepad
- c. Jupyter
- d. None

112.To display data from students table, choose correct statement from below option.

- a. SELECT # FROM students;
- b. SELECT + FROM students WHERE Null;
- c. **SELECT * FROM students;**
- d. None

113.Which data manipulation command is used to combines the records from one or more tables?

- a. SELECT
- b. PROJECT
- c. PRODUCT
- d. **JOIN**

114.In SQL – the function AVG, MIN, MAX, SUM, COUNT are called as_____

- a. Adjunct function
- b. Set operation
- c. Scaler operation
- d. **Aggregate function**

115.Statement used to display table structures in MySQL

- a. **DESCRIBE**
- b. STR
- c. DISPLAY
- d. None

BDTM: Big Data Tools for Managers

116. Which join used to get all the records from left table and matching records from right table.

- a. **LEFT JOIN**
- b. INNER JOIN
- c. CROSS JOIN
- d. None

117. UPDATE statement will update all the records if WHERE clause is not specified.

- a. FALSE
- b. **TRUE**

118. Which statement is used to remove a complete table from MySQL?

- a. DELETES
- b. **TRUNCATE**
- c. REMOVE
- d. DROP

119. DELETE statement without WHERE clause deletes all the records from Table

- a. **TRUE**
- b. FALSE

120. Statement used to get unique records from the Table

- a. UNIQUE
- b. **DISTINCT**
- c. UNION
- d. None

121. Select software required to write & run R Code. (Multiple choice)

- a. **R**
- b. Notebook
- c. **R Studio**
- d. Visual Studio

122. R Studio has basically _____ window sessions.

- a. Two
- b. Five
- c. Three
- d. **Four**

123. Select most common data structures available in R (Multiple choice)

- a. **Vector**
- b. **Matrix**
- c. **DataFrame**
- d. Integer

124. Function used to display values of the vectors or any variables in R?

- a. **print()**
- b. abs()
- c. prod()
- d. sum()

BDTM: Big Data Tools for Managers

125. Operator used to access the variable/column of DataFrame?

- a. `$`(dollar)
- b. `#`(hash)
- c. `.`(dot)
- d. None

126. Function used to transpose of the Matrix/DataFrame in R?

- a. `t()`
- b. `T()`
- c. `c()`
- d. None

127. Function to display top records from R DataFrame?

- a. `head()`
- b. `print()`
- c. `tail()`
- d. None

128. haven package/library used to read SPSS/SAV file in R Code.

- a. `True`
- b. `False`

129. which command is correct to create character ('A','B','C','D','E','F') vectors in R

- a. `c('A','B','C','D','E','F')`
- b. `c(A,B,C,D,E,F)`
- c. `c(AL,BL,CL,DL,EL,FL)`
- d. None

130. Function used to read comma separated values in R?

- a. `read.csv()`
- b. `read_csv()`
- c. `read.xls()`
- d. None

131. Most common data types in Python programming language (Multiple choice)

- a. `int`
- b. `float`
- c. `bool`
- d. `str`

132. Python is a general-purpose interpreted, interactive, object-oriented, and high-level programming language

- a. `True`
- b. `False`

BDTM: Big Data Tools for Managers

133. Software used to run Python code.

- a. **Jupyter notebook**
- b. R Studio
- c. phpMyAdmin
- d. None

134. Select the Quotation which Python supports (Multiple selection)

- a. **Single**
- b. **Double**
- c. **Triple**
- d. Four

135. Which statement python used to define sets of conditional statements

- a. **if..else**
- b. having
- c. where
- d. None

136. Which of the following is used to define a block of code in Python language?

- a. **Indentation**
- b. Key
- c. Brackets
- d. All of the mentioned

137.

138. To define comments in Python _____ operator/symbol used

- a. //
- b. **#**
- c. /*....*/
- d. None

139. Read CSV file in Pandas, which functions used to read file and store as DataFrame

- a. read_csv
- b. **read.csv**
- c. read_file
- d. None

140. Python reserved words are python keyword which can be used as Variable/Functions names

- a. **False**
- b. True

141. A variable name can only contain alpha-numeric characters and underscores (A-z, 0-9, and _)

- a. **True**
- b. False

142. _____ is package manager in Python to download & install Python packages.

- a. download
- b. **pip**
- c. install
- d. None

BDTM: Big Data Tools for Managers

143. To define/create Function in Python which keyword _____ is used.

- a. **def**
- b. describe
- c. colon
- d. None

144. _____ Pandas Function to get the frequency of categorical/labeled variable.

- a. count
- b. freq
- c. **value_counts**
- d. None

145. Loop is useful for iterating elements of List/Set/Tuple

- a. False
- b. **True**

146. 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

147. In Python, Dictionaries are immutable

- a. **False**
- b. True

148. Sets always contains repeated value.

- a. False
- b. **True**

149. To visualize graphs in Pandas which is most common library used

- a. Ploty
- b. **Matplotlib**
- c. Seaborn
- d. None

150. len() functions to display the no of elements available in List/Tuple/Sets/Dictionary.

- a. False
- b. **True**