

Q1.What is the rank of the numpy array? `array([[ 0, 4, 2], [ 9, 3, 7]])`

- A. Rank 2
- B. Rank 3
- C. Rank 1
- D. Invalid numpy array

Correct Option : A

EXPLANATION : The array has 2 axes and hence its rank is 2.

Q2.Choose the correct output of the following program: `>>> a = np.array([11, 12, 13, 14])`  
`>>> b = np.array([1, 2, 3, 4]) >>> c = a - b >>>c`

- A. Array[10, 10, 10, 10]
- B. Array [10]
- C. Array [11-1, 12-2, 13-3, 14-4]
- D. Invalid array

Correct Option : A

EXPLANATION : Arrays a and b are of same dimensions, and one can be directly subtracted from the other. Thus, the answer will be the array of direct subtraction.

Q3.Which of the following data structures of Pandas can handle 3D data?

- A. Series
- B. Data Structure
- C. Panel
- D. All of the above

Correct Option : C

EXPLANATION : A panel is a data structure that handles 3D data.

Q4.To combine datasets, the \_\_\_\_\_ function of Pandas can be utilized.

- A. Concat
- B. Merge
- C. Join
- D. Add

Correct Option : A

EXPLANATION : To combine datasets, the Concat function of Pandas is utilized.

Q5.What is the output of a and b? Given:  $a = 9/2$   $b = 5.2/2$

- A.  $a = 4.5$ ,  $b = 2.6$
- B.  $a = 4$ ,  $b = 2.6$
- C.  $a = 4.5$ ,  $b = 2$
- D.  $a = 4$ ,  $b = 2$

Correct Option : A

EXPLANATION : In Python 3.0,  $9 / 2$  will return 4.5 and  $9 // 2$  will return 4 . The former is floating point division, and the latter is floor division, sometimes also called integer division. a is int/ int so the result will be an integer, b is float/ int so the result will be float.

Q6.A list is collection of values of multiple data types and can:

- A. add
- B. update
- C. remove
- D. do all of the above

Correct Option : D

EXPLANATION : A list is collection of values of multiple data types and can add, remove, and update the data elements.

Q7.Which of the following is contained in the NumPy library?

- A. N-dimensional array object
- B. Tools for integrating C/C++ and Fortran code
- C. Fourier transform
- D. All of the above

Correct Option : D

EXPLANATION : NumPy is the fundamental package for scientific computing with Python. It has support for N (multidimensional) array, can integrate with C/C++ platform, and can perform numerous mathematical operations including Fourier transform.

Q8.Tuple object does not support item assignment.

- A. 1
- B. 0

Correct Option : A

EXPLANATION : Tuple does not support any item assignment.

Q9.The plt.subplots\_adjust() method uses \_\_\_\_\_,\_\_\_\_\_ to move around subplots on the grid.

- A. hspace, vspace
- B. left, right
- C. top, bottom
- D. left, bottom

Correct Option : A

EXPLANATION : It uses hspace and wspace for subplots to move them around. wspace : The amount of width reserved for blank space between subplots hspace: The amount of height reserved for white space between subplots

Q10.Which of the following is used to display plots on the Jupyter notebook?

- A. `%matplotlib inline`
- B. `%matplotlib inline`
- C. `import matplotlib`
- D. `import matplotlib.pyplot`

Correct Option : B

EXPLANATION : `%matplotlib inline` is used to display plots on the Jupyter notebook.

Q11. Which parameter should you use to change a plot's transparency?

- A. `style=`
- B. `linestyle=`
- C. `color=`
- D. `alpha=`

Correct Option : D

EXPLANATION : A picture's transparency is controlled by a parameter called alpha channel. In matplotlib, the plots have a parameter, `alpha=` to change transparency.

Q12. Which of these subpackage does not belong to SciPy?

- A. Cluster
- B. linalg
- C. sklearn
- D. optimize

Correct Option : C

EXPLANATION : sklearn is a library for machine learning.

Q13. How can you change the default plot size?

- A. `plt.rcParams['figure.figsize'] = (w, h)`
- B. Set 'figure.figsize' in .matplotlibrc file
- C. Both A and B
- D. Can't change the default plot size

Correct Option : C

EXPLANATION : We can change the default options with an rc file or change them inside a program. When we set the options inside a program, the option settings will overwrite those in the rc file.

Q14. With `line, = ax.plot(x, y)`, how can you change the width of the line?

- A. `ax.plot(x, y, lw=2)`
- B. `plt.setp(line, lw=2)`
- C. `line.lw = 2`
- D. None of these

Correct Option : B

**EXPLANATION :** For any element in the plot, we can call `plt.setp()` to change its style. Although `ax.plot()` overwrites the line that gives a correct plot, it is bad for performance. However, `line.lw = 2` will cause an exception because there is no `lw` parameter for `line`.

Q15.How can you shallow copy the data in NumPy?

- A. `view()`
- B. `in` method
- C. `is` method
- D. `=` (or assignments)

**Correct Option : A**

**EXPLANATION :** `view()` is used to shallow copy the data.

Q16.If we want to check the distribution of a set of numerical data, what kind of chart should be used?

- A. Pie-chart
- B. Histogram
- C. Heap map
- D. Contour plot

**Correct Option : B**

**EXPLANATION :** A histogram is used to visualize the distribution, a pie chart for visualizing the category proportion, a heat map for visualizing the data as a color map, and a contour plot is used to visualize 3D values in a 2D space.

Q17.Numerical and pictorial information about variables is \_\_\_\_\_.

- A. Descriptive statistics
- B. Analytical statistics
- C. Inferential statistics
- D. Graphical statistics

**Correct Option : A**

**EXPLANATION :** The pictorial representation of data in the form of graphs, charts, and so on is called descriptive statistics.

Q18.What is the value of data for the following? `import pandas as pd`  
`series = pd.Series(3,index=['a','b','c'])`

- A. 0,1,2
- B. 1,2,3
- C. 3,3,3
- D. a,b,c

**Correct Option : C**

**EXPLANATION :** This is a scalar input to create the Pandas series.

Q19.In statistics, Type I error \_\_\_\_\_.

- A. chooses a hypothesis incorrectly
- B. fails to reject the null hypothesis when it is false
- C. a test statistic is incorrect

D. rejects the null hypothesis when it is true

Correct Option : D

EXPLANATION : In statistics, Type I error rejects the null hypothesis when it is true.

Q20.The "Bell Curve" normal distribution is fully characterized by:

- A. mean and median
- B. median and standard deviation
- C. mean and mode
- D. mean and standard deviation

Correct Option : D

EXPLANATION : The Bell curve is fully characterized by the mean (?) and standard deviation (?).

Q21.For Chi-Square test, no association of variables means:

- A. the test is independent of expected frequencies
- B. the test is independent of observed frequencies
- C. Observed Frequency = Expected Frequency
- D. Observed Frequency ? Expected Frequency

Correct Option : C

EXPLANATION : In a Chi-square test if Observed Frequency = Expected Frequency, it indicates no association

Q22.Which of the following code lines will create an array?

- A. `>>>a=np.array(5,8,12,65)`
- B. `>>>a=np.array([5,8,12,65])`
- C. Both
- D. None

Correct Option : B

EXPLANATION : A numpy array must be passed as a list to the array method in closed parentheses.

Q23.What is the output of the following code line? `>>>np.zeros((3,3))`

- A. `array([[ 0., 0., 0.], [ 0., 0., 0.], [ 0., 0., 0.]])`
- B. `array([ 0., 0., 0.])`
- C. `array([[ 0., 0., 0.]])`
- D. `array ([[ 3,3,3]])`

Correct Option : A

EXPLANATION : This command will create an array of zeros, with three rows and columns.

Q24.Which method will give you the sum of diagonal elements of an array matrix?

- A. `inv`
- B. `transpose`
- C. `sum`
- D. `trace`

Correct Option : D

EXPLANATION : The Trace() method helps you to perform sum on the diagonal data elements of the array.

Q25.Which of the following is NOT used in unsupervised machine learning?

- A. feature
- B. attribute
- C. label
- D. observation

Correct Option : C

EXPLANATION : Label or response is not present in unsupervised learning.

Q26.Which of the following is NOT the object of BeautifulSoup?

- A. Tag
- B. Navigable String
- C. Comment
- D. Parser

Correct Option : D

EXPLANATION : There are four types of objects in BeautifulSoup: Tag, Navigable String, BeautifulSoup, and Comment. Parser is not an object in BeautifulSoup.

Q27.In machine learning, for the train\_test\_split method, what is the default split of the train dataset and test dataset?

- A. 40% test and 60% train
- B. 30% test and 70% train
- C. 25% test and 75% train
- D. 50% test and 50% train

Correct Option : C

EXPLANATION : The default split for train\_test\_split method of machine learning model training is 25% test dataset and 75% train dataset.

Q28.What will be the size of the numpy array for the following code? >>>np.linspace(5,15,10)

- A. 5
- B. 10
- C. 15
- D. 20

Correct Option : B

EXPLANATION : The linspace method in NumPy (start,end,number of elements) equally spaces data elements.

Q29.If data is categorical and has no labels (response), then it falls under \_\_\_\_\_.

- A. Supervised learning: classification
- B. Unsupervised learning: clustering

- C. Unsupervised learning: dimensionality reduction
- D. Supervised learning: regression

Correct Option : B

EXPLANATION : If data is categorical and has no labels (response), then it is a unsupervised learning clustering problem.

Q30.Which attribute is used to annotate the arrow position in a plot?

- A. ha
- B. va
- C. xy
- D. arrowprops

Correct Option : C

EXPLANATION : xy indicates the position of the arrow.

Q31.Identify a kind of filter to search a web tree using BeautifulSoup. Select all that apply.

- A. string
- B. list
- C. function
- D. All of the above

Correct Option : D

EXPLANATION : All of these filters can be used to search for information in a web tree.

Q32.The find() method in BeautifulSoup \_\_\_\_.

- A. returns the list with all values
- B. searches the entire document
- C. returns only the first match value
- D. returns an empty list if nothing is matched

Correct Option : C

EXPLANATION : The find() method searches only for the passed argument, returns only the first match value, and returns None when nothing is matched.

Q33.Which type of plot is most appropriate to show percentage or proportional data?

- A. Histogram
- B. Pie chart
- C. Scatter plot
- D. Bar pot

Correct Option : B

EXPLANATION : Pie charts are most appropriate to show percentage or proportional data.

Q34.Which of the following functions is used to manage missing data?

- A. Np.nan()
- B. Fillna()
- C. Dropna()

D. All of the above

Correct Option : D

EXPLANATION : All of these functions can be used to manage missing data. Use np.nan value to represent missing data. By default, it is not included in any computation. The dropna() function is used to drop rows that contain missing values. And fillna() function fills the missing data with other predetermined values.

Q35.Which of the following attributes is used to navigate sideways in BeautifulSoup?

- A. .next\_sibling and .previous\_sibling
- B. .contents and .children
- C. .parents and .parent
- D. .next\_element and .previous\_element

Correct Option : A

EXPLANATION : The attributes used to navigate sideways are .next\_sibling and .previous\_sibling.

Q36.Which facility can be used to combine the Series, DataFrame, and Panel objects with various kinds of set logic for the indexes and relational algebra functionality?

- A. Concat
- B. Add
- C. Combine
- D. All of the above

Correct Option : A

EXPLANATION : The Concat function from Pandas provides facilities to combine the Series, DataFrame, and Panel objects with various kinds of set logic for the indexes and relational algebra functionality.

Q37.If both HTML and XML parsers are available, which one will be chosen by BeautifulSoup for parsing?

- A. XML
- B. HTML
- C. Neither A or B
- D. Both A and B

Correct Option : B

EXPLANATION : BeautifulSoup chooses the best available parser for parsing a document. By default, it chooses an HTML parser.

Q38.Which of the following extracts unstructured data from the web and transforms it into a structured format?

- A. Web scraping
- B. Web indexing
- C. Advanced indexing
- D. Web automation

Correct Option : A



**EXPLANATION :** Web scraping extracts unstructured data on the web and transforms it into a structured format. Web scraping is also related to web automation.

Q39. Which function is applied to get the output array B from the input array A? Input array A : array([1, 2, 3]) Output array B : array([1]), array ([2]), array ([3])

- A. Vstack
- B. Hsplit
- C. Vsplit
- D. Hstack

**Correct Option : B**

**EXPLANATION :** Hsplit can split an array along its horizontal axis by specifying the number of equally shaped arrays to return.

Q40. A dataset is provided in the form of a table. Which data structure will you use to view specific columns and rows?

- A. Series
- B. DataFrame
- C. Panel
- D. Panelnd

E. Correct Option : B

F. **EXPLANATION :** Data viewing is one of the key applications of a DataFrame function. The data can also be viewed with a conditional approach, for example, certain columns only, certain rows only, or certain rows and columns.

Q41. def keyword is used to create a:

- A. variable
- B. string
- C. function
- D. Integer

**Correct Option : C**

**EXPLANATION :** A function is created using def keyword in Python.

Q42. Which of the following objects of BeautifulSoup is NOT editable?

- A. Tag
- B. Navigable String
- C. BeautifulSoup
- D. Parser

**Correct Option : B**

**EXPLANATION :** You cannot edit the Navigable String object but can convert it into a Unicode string using the function Unicode.

Q43. Which function should be used to pair the data elements of a list?

- A. enumerate
- B. concat

- C. zip
- D. attach

Correct Option : C

EXPLANATION : zip is a built-in sequence function and is used to pair the data elements of a list.

Q44.Dictionary stores a value as \_\_\_\_\_.

- A. a list
- B. a key -value
- C. tuples
- D. an array

Correct Option : B

EXPLANATION : dictionary is a basic data structure which stores a value in form of a key-value.

Q45.Which of the following is the only XML parser available in BeautifulSoup?

- A. lxml.html
- B. lxml.xml
- C. html.parser
- D. html5lib

Correct Option : B

EXPLANATION : Lxml.xml is the only XMLparser available, and it also depends on C.

Q46.\_\_\_\_\_ are nothing but the machines in cluster where the data is actually stored.

- A. DataNodes
- B. Data stacks
- C. Switches
- D. Nodes

Correct Option : A

EXPLANATION : DataNodes are nothing but the machines in cluster where the data is actually stored.

Q47.In a Hadoop cluster, what is the usual size of the data block?

- A. 100 MB
- B. 64 MB
- C. 1 GB
- D. 150 MB

Correct Option : B

EXPLANATION : 64 MB is the usual data block size on Hadoop cluster data nodes.

Q48.In MapReduce, the output is written to \_\_\_\_\_.

- A. disk
- B. memory
- C. cache
- D. temp directory

Correct Option : A

EXPLANATION : In MapReduce, all data is written to a disk on local data nodes.

Q49.What is the point of entry of a spark application?

- A. pyspark
- B. RDD
- C. SparkContext (sc)
- D. pyspark driver

Correct Option :

EXPLANATION

Q50.In a Hadoop set up, the NameNode contains the actual \_\_\_\_.

- A. data
- B. meta data of distributed files
- C. configuration file for mapper
- D. client information

Correct Option : B

EXPLANATION : The NameNode keeps the metadata information of the distributed files which reside on HDFS.