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 ouput 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

## 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

## 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

## Q8.Tuple object does not support item assignment.

    A. 1
    B. 0

## Q9.The plt.subplots_adjust() method uses _____,_____ to move around subplots on the grid.

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

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

A. %matplot inline
B. %matplotlib inline
C. import matplotlib
D. import matplotlib.pyplot

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

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

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

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

Q13.How can you change the default plot size?

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

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

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

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

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

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

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]])

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

   A. enumerate
   B. concat

C. zip

D. attach

Q44.Dictionary stores a value as _____.

A. a list

B. a key -value

C. tuples

D. an array

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

A. lxml.html

B. lxml.xml

C. html.parser

D. html5lib

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

A. DataNodes

B. Data stacks

C. Switches

D. Nodes

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

A. 100 MB

B. 64 MB

C. 1 GB

D. 150 MB

Q48.In MapReduce, the output is written to _____.

A. disk

B. memory

C. cache

D. temp directory

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

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

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