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

**SELECT THE CORRECT ANSWER**

Top of Form

data

meta data of distributed files

configuration file for mapper

client information

Bottom of Form

Correct Option:B

**EXPLANATION**

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

What is the point of entry of a spark application?

**SELECT THE CORRECT ANSWER**

Top of Form

pyspark

RDD

SparkContext (sc)

pyspark driver

Bottom of Form

Correct Option:C

**EXPLANATION**

SparkContext is the point of entry of a spark application and must be available for execution of a program

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

**SELECT THE CORRECT ANSWER**

Top of Form

disk

memory

cache

temp directory

Bottom of Form

Correct Option:A

**EXPLANATION**

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

--------------------------------------------------------------------------------------------------------------------------------------

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

**SELECT THE CORRECT ANSWER**

Top of Form

100 MB

64 MB

1 GB

150 MB

Bottom of Form

Correct Option:B

**EXPLANATION**

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

--------------------------------------------------------------------------------------------------------------------------------------

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

**SELECT THE CORRECT ANSWER**

Top of Form

DataNodes

Data stacks

Switches

Nodes

Bottom of Form

Correct Option:A

**EXPLANATION**

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

--------------------------------------------------------------------------------------------------------------------------------------

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

**SELECT THE CORRECT ANSWER**

Top of Form

lxml.html

lxml.xml

html.parser

html5lib

Bottom of Form

Correct Option:B

**EXPLANATION**

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

--------------------------------------------------------------------------------------------------------------------------------------

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

**SELECT THE CORRECT ANSWER**

Top of Form

a list

a key -value

tuples

an array

Bottom of Form

Correct Option:B

**EXPLANATION**

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

--------------------------------------------------------------------------------------------------------------------------------------

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

**SELECT THE CORRECT ANSWER**

Top of Form

enumerate

concat

zip

attach

Bottom of Form

Correct Option:C

**EXPLANATION**

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

--------------------------------------------------------------------------------------------------------------------------------------

Which of the following objects of BeautifulSoup is NOT editable?

**SELECT THE CORRECT ANSWER**

Top of Form

Tag

Navigable String

BeautifulSoup

Parser

Bottom of Form

Correct Option:B

**EXPLANATION**

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

--------------------------------------------------------------------------------------------------------------------------------------

def keyword is used to create a:

**SELECT THE CORRECT ANSWER**

Top of Form

variable

string

function

Integer

Bottom of Form

Correct Option:C

**EXPLANATION**

A function is created using def keyword in Python

--------------------------------------------------------------------------------------------------------------------------------------

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

**SELECT THE CORRECT ANSWER**

Top of Form

Series

DataFrame

Panel

Panelnd

Bottom of Form

Correct Option:B

**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.

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])

**SELECT THE CORRECT ANSWER**

Top of Form

Vstack

Hsplit

Vsplit

Hstack

Bottom of Form

Correct Option:B

**EXPLANATION**

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

--------------------------------------------------------------------------------------------------------------------------------------

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

**SELECT THE CORRECT ANSWER**

Top of Form

Web scraping

Web indexing

Advanced indexing

Web automation

Bottom of Form

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.

--------------------------------------------------------------------------------------------------------------------------------------

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

**SELECT THE CORRECT ANSWER**

Top of Form

XML

HTML

Neither A or B

Both A and B

Bottom of Form

Correct Option:B

**EXPLANATION**

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

--------------------------------------------------------------------------------------------------------------------------------------

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?

**SELECT THE CORRECT ANSWER**

Top of Form

Concat

Add

Combine

All of the above

Bottom of Form

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.

--------------------------------------------------------------------------------------------------------------------------------------

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

**SELECT THE CORRECT ANSWER**

Top of Form

.next\_sibling and .previous\_sibling

.contents and .children

.parents and .parent

.next\_element and .previous\_element

Bottom of Form

**EXPLANATION**

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

--------------------------------------------------------------------------------------------------------------------------------------

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

**SELECT THE CORRECT ANSWER**

Top of Form

Np.nan()

Fillna()

Dropna()

All of the above

Bottom of Form

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.

--------------------------------------------------------------------------------------------------------------------------------------

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

**SELECT THE CORRECT ANSWER**

Top of Form

Histogram

Pie chart

Scatter plot

Bar pot

Bottom of Form

Correct Option:B

**EXPLANATION**

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

--------------------------------------------------------------------------------------------------------------------------------------

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

**SELECT THE CORRECT ANSWER**

Top of Form

returns the list with all values

searches the entire document

returns only the first match value

returns an empty list if nothing is matched

Bottom of Form

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.

--------------------------------------------------------------------------------------------------------------------------------------

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

**SELECT THE CORRECT ANSWER**

Top of Form

string

list

function

All of the above

Bottom of Form

Correct Option:D

**EXPLANATION**

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

--------------------------------------------------------------------------------------------------------------------------------------

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

**SELECT THE CORRECT ANSWER**

Top of Form

ha

va

xy

arrowprops

Bottom of Form

Correct Option:C

**EXPLANATION**

xy indicates the position of the arrow.

--------------------------------------------------------------------------------------------------------------------------------------

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

**SELECT THE CORRECT ANSWER**

Top of Form

Supervised learning: classification

Unsupervised learning: clustering

Unsupervised learning: dimensionality reduction

Supervised learning: regression

Bottom of Form

Correct Option:B

**EXPLANATION**

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

--------------------------------------------------------------------------------------------------------------------------------------

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

**SELECT THE CORRECT ANSWER**

Top of Form

5

10

15

20

Bottom of Form

Correct Option:B

**EXPLANATION**

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

--------------------------------------------------------------------------------------------------------------------------------------

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

**SELECT THE CORRECT ANSWER**

Top of Form

40% test and 60% train

30% test and 70% train

25% test and 75% train

50% test and 50% train

Bottom of Form

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.

--------------------------------------------------------------------------------------------------------------------------------------

Which of the following is NOT the object of BeautifulSoup?

**SELECT THE CORRECT ANSWER**

Top of Form

Tag

Navigable String

Comment

Parser

Bottom of Form

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.

--------------------------------------------------------------------------------------------------------------------------------------

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

**SELECT THE CORRECT ANSWER**

Top of Form

feature

attribute

label

observation

Bottom of Form

Correct Option:C

**EXPLANATION**

Label or response is not present in unsupervised learning.

--------------------------------------------------------------------------------------------------------------------------------------

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

**SELECT THE CORRECT ANSWER**

Top of Form

inv

transpose

sum

trace

Bottom of Form

Correct Option:D

**EXPLANATION**

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

--------------------------------------------------------------------------------------------------------------------------------------

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

**SELECT THE CORRECT ANSWER**

Top of Form

array([[ 0., 0., 0.], [ 0., 0., 0.], [ 0., 0., 0.]])

array([ 0., 0., 0.])

array([[ 0., 0., 0.]])

array ([[ 3,3,3]])

Bottom of Form

Correct Option:A

**EXPLANATION**

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

--------------------------------------------------------------------------------------------------------------------------------------

Which of the following code lines will create an array?

**SELECT THE CORRECT ANSWER**

Top of Form

>>>a=np.array(5,8,12,65)

>>>a=np.array([5,8,12,65])

Both

None

Bottom of Form

Correct Option:B

**EXPLANATION**

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

--------------------------------------------------------------------------------------------------------------------------------------

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

**SELECT THE CORRECT ANSWER**

Top of Form

the test is independent of expected frequencies

the test is independent of observed frequencies

Observed Frequency = Expected Frequency

Observed Frequency ? Expected Frequency

Bottom of Form

Correct Option:C

**EXPLANATION**

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

--------------------------------------------------------------------------------------------------------------------------------------

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

**SELECT THE CORRECT ANSWER**

Top of Form

mean and median

median and standard deviation

mean and mode

mean and standard deviation

Bottom of Form

Correct Option:D

**EXPLANATION**

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

--------------------------------------------------------------------------------------------------------------------------------------

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

**SELECT THE CORRECT ANSWER**

Top of Form

chooses a hypothesis incorrectly

fails to reject the null hypothesis when it is false

a test statistic is incorrect

rejects the null hypothesis when it is true

Bottom of Form

Correct Option:D

**EXPLANATION**

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

--------------------------------------------------------------------------------------------------------------------------------------

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

**SELECT THE CORRECT ANSWER**

Top of Form

0,1,2

1,2,3

3,3,3

a,b,c

Bottom of Form

Correct Option:C

**EXPLANATION**

This is a scalar input to create the Pandas series.

--------------------------------------------------------------------------------------------------------------------------------------

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

**SELECT THE CORRECT ANSWER**

Top of Form

Descriptive statistics

Analytical statistics

Inferential statistics

Graphical statistics

Bottom of Form

Correct Option:A

**EXPLANATION**

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

--------------------------------------------------------------------------------------------------------------------------------------

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

**SELECT THE CORRECT ANSWER**

Top of Form

Pie-chart

Histogram

Heat map

Contour plot

Bottom of Form

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.

--------------------------------------------------------------------------------------------------------------------------------------

How can you shallow copy the data in NumPy?

**SELECT THE CORRECT ANSWER**

Top of Form

view()

in method

is method

= (or assignments)

Bottom of Form

Correct Option:A

**EXPLANATION**

view() is used to shallow copy the data.

--------------------------------------------------------------------------------------------------------------------------------------

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

**SELECT THE CORRECT ANSWER**

Top of Form

ax.plot(x, y, lw=2)

plt.setp(line, lw=2)

line.lw = 2

None of these

Bottom of Form

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.

--------------------------------------------------------------------------------------------------------------------------------------

How can you change the default plot size?

**SELECT THE CORRECT ANSWER**

Top of Form

plt.rcParam['figure.figsize'] = (w, h)

Set 'figure.figsize' in .matplotlibrc file

Both A and B

Can't change the default plot size

Bottom of Form

Correct Option:C

**EXPLANATION**

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

--------------------------------------------------------------------------------------------------------------------------------------

Which of these subpackage does not belong to SciPy?

**SELECT THE CORRECT ANSWER**

Top of Form

Cluster

linalg

sklearn

optimize

Bottom of Form

Correct Option:C

**EXPLANATION**

sklearn is a library for machine learning.

--------------------------------------------------------------------------------------------------------------------------------------

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

**SELECT THE CORRECT ANSWER**

Top of Form

style=

linestyle=

color=

alpha=

Bottom of Form

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.

--------------------------------------------------------------------------------------------------------------------------------------

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

**SELECT THE CORRECT ANSWER**

Top of Form

%matplot inline

%matplotlib inline

import matplotlib

import matplotlib.pyplot

Bottom of Form

Correct Option:B

**EXPLANATION**

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

--------------------------------------------------------------------------------------------------------------------------------------

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

**SELECT THE CORRECT ANSWER**

Top of Form

hspace, vspace

left, right

top, bottom

lef, bottom

Bottom of Form

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

--------------------------------------------------------------------------------------------------------------------------------------

Tuple object does not support item assignment.

**SELECT THE CORRECT ANSWER**

Top of Form

False

True

Bottom of Form

Correct Option:B

**EXPLANATION**

Tuple does not support any item assignment.

Which of the following is contained in the NumPy library?

**SELECT THE CORRECT ANSWER**

Top of Form

N-dimensional array object

Tools for integrating C/C++ and Fortran code

Fourier transform

All of the above

Bottom of Form

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 numerious mathematical operations including Fourier transform.

--------------------------------------------------------------------------------------------------------------------------------------

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

**SELECT THE CORRECT ANSWER**

Top of Form

add

update

remove

do all of the above

Bottom of Form

Correct Option:D

**EXPLANATION**

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

--------------------------------------------------------------------------------------------------------------------------------------

What is the ouput of a and b? Given: a = 9/2 b = 5.2/2

**SELECT THE CORRECT ANSWER**

Top of Form

a = 4.5, b =2.6

a = 4, b =2.6

a = 4.5, b =2

a = 4, b =2

Bottom of Form

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.

--------------------------------------------------------------------------------------------------------------------------------------

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

**SELECT THE CORRECT ANSWER**

Top of Form

Concat

Merge

Join

Add

Bottom of Form

Correct Option:A

**EXPLANATION**

To combine datasets, the Concat function of Pandas is utilized

--------------------------------------------------------------------------------------------------------------------------------------

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

**SELECT THE CORRECT ANSWER**

Top of Form

Series

Data Structure

Panel

All of the above

Bottom of Form

Correct Option:C

**EXPLANATION**

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

--------------------------------------------------------------------------------------------------------------------------------------

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

**SELECT THE CORRECT ANSWER**

Top of Form

Array[10, 10, 10, 10]

Array [10]

Array [11-1, 12-2, 13-3, 14-4]

Invalid array

Bottom of Form

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.

--------------------------------------------------------------------------------------------------------------------------------------

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

**SELECT THE CORRECT ANSWER**

Top of Form

Rank 2

Rank 3

Rank 1

Invalid numpy array

Bottom of Form

Correct Option:A

**EXPLANATION**

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

--------------------------------------------------------------------------------------------------------------------------------------