



DreambigCareer

## Data Science Python Interview Questions and Answers

- 1) How can you build a simple logistic regression model in Python?
- 2) How can you train and interpret a linear regression model in SciKit learn?
- 3) Name a few libraries in Python used for Data Analysis and Scientific computations.  
NumPy, SciPy, Pandas, SciKit, Matplotlib, Seaborn
- 4) Which library would you prefer for plotting in Python language: Seaborn or Matplotlib?  
Matplotlib is the python library used for plotting but it needs lot of fine-tuning to ensure that the plots look shiny. Seaborn helps data scientists create statistically and aesthetically appealing meaningful plots. The answer to this question varies based on the requirements for plotting data.
- 5) What is the main difference between a Pandas series and a single-column DataFrame in Python?
- 6) Write code to sort a DataFrame in Python in descending order.
- 7) How can you handle duplicate values in a dataset for a variable in Python?
- 8) Which Random Forest parameters can be tuned to enhance the predictive power of the model?
- 9) Which method in `pandas.tools.plotting` is used to create scatter plot matrix?

Scatter\_matrix

10) How can you check if a data set or time series is Random?

To check whether a dataset is random or not use the lag plot. If the lag plot for the given dataset does not show any structure then it is random.

11) Can we create a DataFrame with multiple data types in Python? If yes, how can you do it?

12) Is it possible to plot histogram in Pandas without calling Matplotlib? If yes, then write the code to plot the histogram?

13) What are the possible ways to load an array from a text data file in Python? How can the efficiency of the code to load data file be improved?

`numpy.loadtxt ()`

14) Which is the standard data missing marker used in Pandas?

NaN

15) Why you should use NumPy arrays instead of nested Python lists?

16) What is the preferred method to check for an empty array in NumPy?

17) List down some evaluation metrics for regression problems.

18) Which Python library would you prefer to use for Data Munging?

Pandas

19) Write the code to sort an array in NumPy by the nth column?

Using `argsort ()` function this can be achieved. If there is an array `X` and you would like to sort the `n`th column then code for this will be `x[x [: n-1].argsort ()]`

20) How are NumPy and SciPy related?

21) Which python library is built on top of matplotlib and Pandas to ease data plotting?

Seaborn

21) Which plot will you use to access the uncertainty of a statistic?

Bootstrap

22) What are some features of Pandas that you like or dislike?

23) Which scientific libraries in SciPy have you worked with in your project?

24) What is pylab?

A package that combines NumPy, SciPy and Matplotlib into a single namespace.

26) Which python library is used for Machine Learning?

SciKit-Learn

]

Basic Python Programming Interview Questions

27) How can you copy objects in Python?

The functions used to copy objects in Python are-

- 1) `Copy.copy ()` for shallow copy
- 2) `Copy.deepcopy ()` for deep copy

However, it is not possible to copy all objects in Python using these functions. For instance, dictionaries have a separate copy method whereas sequences in Python have to be copied by 'Slicing'.

28) What is the difference between tuples and lists in Python?

Tuples can be used as keys for dictionaries i.e. they can be hashed. Lists are mutable whereas tuples are immutable - they cannot be changed. Tuples should be used when the order of elements in a sequence matters. For example, set of actions that need to be executed in sequence, geographic locations or list of points on a specific route.

29) What is PEP8?

PEP8 consists of coding guidelines for Python language so that programmers can write readable code making it easy to use for any other person, later on.

30) Is all the memory freed when Python exits?

No it is not, because the objects that are referenced from global namespaces of Python modules are not always de-allocated when Python exits.

31) What does `_init_.py` do?

`_init_.py` is an empty py file used for importing a module in a directory. `_init_.py` provides an easy way to organize the files. If there is a module `maindir/subdir/module.py`, `_init_.py` is placed in all the directories so that the module can be imported using the following command-

```
import maindir.subdir.module
```

32) What is the different between `range ()` and `xrange ()` functions in Python?

`range ()` returns a list whereas `xrange ()` returns an object that acts like an iterator for generating numbers on demand.

33) How can you randomize the items of a list in place in Python?

`Shuffle (lst)` can be used for randomizing the items of a list in Python

34) What is a pass in Python?

Pass in Python signifies a no operation statement indicating that nothing is to be done.

35) If you are gives the first and last names of employees, which data type in Python will you use to store them?

You can use a list that has first name and last name included in an element or use Dictionary.

36) What happens when you execute the statement `mango=banana` in Python?  
A name error will occur when this statement is executed in Python.

37) Write a sorting algorithm for a numerical dataset in Python.

38) Optimize the below python code-

```
word = 'word'
```

```
print word.__len__()
```

Answer: `print 'word'._len_()`

39) What is monkey patching in Python?

Monkey patching is a technique that helps the programmer to modify or extend other code at runtime. Monkey patching comes handy in testing but it is not a good practice to use it in production environment as debugging the code could become difficult.

40) Which tool in Python will you use to find bugs if any?

Pylint and Pychecker. Pylint verifies that a module satisfies all the coding standards or not. Pychecker is a static analysis tool that helps find out bugs in the course code.

41) How are arguments passed in Python- by reference or by value?



The answer to this question is neither of these because passing semantics in Python are completely different. In all cases, Python passes arguments by value where all values are references to objects.