

## VIVA VOCE

1. What is Python? What are the benefits of using Python?  
**Ans.** Python is a programming language with objects, modules, threads, exceptions and automatic memory management. The benefits of Python are that it is simple and easy, portable, extensible, built-in data structure and is open source.
2. What is pickling and unpickling?  
**Ans.** Pickle module accepts any Python object and converts it into a string representation and dumps it into a file by using dump function. This process is called pickling. The process of retrieving original Python objects from the stored string representation is called unpickling.
3. How is Python interpreted?  
**Ans.** Python language is an interpreted language. Python program runs directly from the source code. It converts the source code that is written by the programmer into an intermediate language, which is again translated into machine language that has to be executed.
4. What are numpy arrays?  
**Ans.** A numpy array is a table of elements, all of the same type, indexed by a list/tuple of integers. Numpy arrays are also called ndarrays.
5. Why are numpy arrays used over lists?  
**Ans.** Numpy arrays have contiguous memory allocation. Thus, a same array stored as list will require more space as compared to arrays.
  - They are more speedy to work with and, hence, are more efficient than lists.
  - They are more convenient to deal with.
6. What do you understand by Array slicing?  
**Ans.** Array slicing refers to the process of extracting a subset of elements from an existing array and returning the result as another array, possibly in a different dimension from the original.
7. Is Python numpy better than lists?  
**Ans.** Yes, numpy array are better than the lists. We use Python numpy array instead of a list because of the following three reasons:
  - (a) Less Memory
  - (b) Fast
  - (c) Convenient
8. What is Pandas Series?  
**Ans.** Series is a one-dimensional labelled data structure capable of holding data of any type (integer, string, float, Python objects, etc.) accessed using respective data labels, called its index.
9. Who is the main author of Pandas?  
**Ans.** Wes McKinney
10. What is dataframe?  
**Ans.** Dataframe is a two-dimensional data structure with heterogeneous data, usually represented in tabular format. The data is represented in rows and columns. Each column represents an attribute and each row represents a record. Features of dataframe are:
  - Potentially, columns are of different types.
  - Size is mutable.
  - Labelled axes (rows and columns).
  - Can perform arithmetic operations on rows and columns.
11. How can we fill missing values in dataframe?  
**Ans.** For filling missing values, we can use fillna() method. *For example,*  

```
df = df1.fillna(0)
```
12. Which is the standard data missing marker used in Pandas?  
**Ans.** NaN

- 13.** Differentiate between aggregate function and transform().
- Ans.** The aggregation function `agg()` returns a reduced version of the data by producing one summary result per group. On the other hand, `transform()` returns the transformed version of the summary data by repeating rows for the group to make it take the shape as that of the full data and thus the result of transform can be combined with the dataframes easily.
- 14.** How is memory managed in Python?
- Ans.** Python memory is managed by Python private heap space. All Python objects and data structures are located in a private heap. The programmer does not have access to this private heap and the interpreter takes care of this Python private heap.
- The allocation of Python heap space for Python objects is done by Python memory manager. The core API gives access to some tools for the programmer to code.
  - Python also has an inbuilt garbage collector, which recycles all the unused memory, frees up memory, and makes it available to the heap space.
- 15.** What is lambda in Python?
- Ans.** It is a single expression, anonymous function (without any name) often used as inline function.
- 16.** What is a quartile? How is it related to quantile? How do you generate it in Pandas?
- Ans.** Quartiles Q1, Q2 and Q3 are three points that divide a distribution into 4 parts. The 4 quantiles are called quartiles. With a quantile, the values of the distribution do not need to be multiples of 0.25 while quartiles are multiples of 0.25.
- In Pandas, we generate these with the function `quantile()`.
- 17.** What is pivoting? Which functions of Pandas support pivoting?
- Ans.** Data pivoting is a summarizing technique to rearrange the columns and rows in a report so as to view data from different perspectives.
- Pandas library makes available two functions for pivoting—the `pivot()` and `pivot_table()` functions.
- 18.** What is the difference between `pivot()` and `pivot_table()` functions?
- Ans.** `pivot()` and `pivot_table()` both perform data pivoting on a data set. But with `pivot()`, if there are multiple entries for a column value pointing to same index (row), it leads to an error.
- The `pivot_table()` pivots the data by aggregating it; thus, it can work with duplicate entries.
- 19.** What is `groupby()` in Pandas?
- Ans.** The `groupby()` function rearranges data into groups based on some criteria and stores the rearranged data in a new `groupby` object. You can apply aggregate functions on the `groupby` object using `agg()`.
- 20.** What are the rules for local and global variables in Python?
- Ans. Local variables:** If a variable is assigned a new value anywhere within a function's body, it is assumed to be local.
- Global variables:** Those variables that are only referenced inside a function are implicitly global.
- 21.** What is the purpose of PYTHONPATH environment variable?
- Ans.** PYTHONPATH has a role similar to PATH. This variable tells the Python interpreter where to locate the module files imported into a program. It should include the Python source library directory and the directories containing Python source code.
- 22.** Why is the return keyword used in Python?
- Ans.** The purpose of a function is to receive the inputs and return some output.
- The `return` is a Python statement which we can use in a function for sending a value back to its calling function or the operating system.
- 23.** How do you debug a program in Python? Is it possible to step through the Python code?
- Ans.** Yes, we can use the Python debugger (`pdb`) to debug any Python program. If we start a program using `pdb`, then it lets us even step through the code.

- 24.** What do you understand by Python modules?
- Ans.** A file containing Python definitions and statements is called a Python module. So naturally, the filename is the module name which is appended with the suffix **.py**.
- 25.** What is the difference between del keyword and clear() function?
- Ans.** The difference between del keyword and clear() function is that while del keyword removes one element at a time, clear function removes all the elements.
- 26.** Which library is used for plotting in Python language?
- Ans.** Matplotlib is the library used for plotting in Python language.
- 27.** Explain reindexing in Pandas.
- Ans.** Reindexing means to conform dataframe to a new index placing NA/NaN in locations having no value in the previous index. It changes the row labels and column labels of a dataframe.
- 28.** What is Pandas used for?
- Ans.** This library is written for the Python programming language for performing operations like data manipulation, data analysis, etc. The library provides various operations as well as data structures to manipulate time series and numerical tables.
- 29.** How can we create an empty dataframe?
- Ans.** Using function dataframe().
- 30.** What is the difference between apply() and applymap() in Pandas?
- Ans.** apply() is used to apply a function along an axis (across rows or columns) of the dataframe while applymap() is used to apply a function to a dataframe element-wise.
- 31.** What are subplots in matplotlib?
- Ans.** Subplots are grid of plots within a single figure. Subplots can be plotted using subplots() function from matplotlib.pyplot module.
- 32.** Which function is used to draw a histogram?
- Ans.** hist()
- 33.** What is meant by broadcasting in numpy?
- Ans.** The term broadcasting refers to the ability of numpy to treat arrays of different shapes during arithmetic operations. If the dimensions of two arrays are dissimilar, element-to-element operations are not possible. However, operations on arrays of non-similar shapes are still possible in numpy, because of the broadcasting capability. The smaller array is broadcast to the size of the larger array so that they have compatible shapes.
- 34.** State the purpose of using zip() in Pandas.
- Ans.** zip() function creates the objects and that can be used to produce a single item at a time. This function can create Pandas dataframes by merging two lists.
- 35.** What is primary key?
- Ans.** Primary key is a combination of columns that uniquely identifies a row in a relational table.
- 36.** What is candidate key?
- Ans.** All possible combinations of columns that can possibly serve as the primary key are called candidate keys.
- 37.** What is foreign key?
- Ans.** A combination of columns where values are derived from primary key of some other table is called the foreign key of the table in which it is contained.
- 38.** What is alternate key?
- Ans.** A candidate key that is not serving as a primary key is called an alternate key.
- 39.** What is MYSQL?
- Ans.** MYSQL is an open source RDBMS that relies on SQL for processing the data in the database. The database is available for free under the terms of General Public License (GPL).

**40.** What is RDBMS?

**Ans.** Relational Database Management System (RDBMS) facilitates access, security and integrity of data and eliminates data redundancy. *For example*, MYSQL, Oracle, Microsoft Sql Server, etc.

**41.** What is the use of drop command?

**Ans.** Drop command is used to delete tables. *For example*, Drop Table Orders. Delete commands are used to delete rows of a table.

**42.** What do you understand by NOT NULL constraint?

**Ans.** This constraint ensures that the null values are not permitted on a specified column. This constraint can be defined at the column level and not at the table level.

**43.** What is the significance of COUNT?

**Ans.** It is used to count the number of values in a given column or number of rows in a table. *For example*, Select count (Roll No) from students.

**44.** How can we delete a table in MYSQL?

**Ans.** We can delete a table in MYSQL using the drop command.

**45.** How can we delete a record in MYSQL?

**Ans.** We can delete a record in MYSQL using the delete command.

**46.** How can we add a record and add a column in a table?

**Ans.** We can add a record by using insert command and we can add a column through the alter table command.

**47.** Give any two differences between GET and POST submission methods of HTML form.

**Ans.**

GET METHOD	POST METHOD
All form data is encoded into the URL appended to the action URL as query string parameters.	Form data appears within the message body of the HTTP request.
Parameters remain in browser history, hence cannot be used to send password-like sensitive information.	Parameters are not saved in browser history, hence can be used to send sensitive information.
Can be bookmarked.	Cannot be bookmarked.
Easier to hack for script kiddies.	More difficult to hack.
Can be cached.	Cannot be cached.

**48.** Give the necessary command to incorporate SQL interface within Python.

**Ans.** import MySQLdb or import mysql-connector

**49.** What is Django?

**Ans.** Django is a high-level Python web framework that encourages rapid development and clean, pragmatic design. Developed by a fast-moving online news operation, Django was designed to handle two challenges: the intensive deadlines of a newsroom and the stringent requirements of the experienced web developers who wrote it. It lets us build high-performing, elegant web applications quickly.