

Pandas questions

07 June 2022 13:00

Mention the different types of Data Structures in Pandas?

Pandas provide two data structures, which are supported by the pandas library, **Series**, and **DataFrames**. Both of these data structures are built on top of the NumPy.

A **Series** is a one-dimensional data structure in pandas, whereas the **DataFrame** is the two-dimensional data structure in pandas.

From <<https://www.javatpoint.com/python-pandas-interview-questions>>

Define Series in Pandas?

A Series is defined as a one-dimensional array that is capable of storing various data types. The row labels of series are called the **index**. By using a '**series**' method, we can easily convert the list, tuple, and dictionary into series. A Series cannot contain multiple columns.

From <<https://www.javatpoint.com/python-pandas-interview-questions>>

How can we calculate the standard deviation from the Series?

The Pandas **std()** is defined as a function for calculating the standard deviation of the given set of numbers, DataFrame, column, and rows.

Series.std(axis=None, skipna=None, level=None, ddof=1, numeric_only=None, **kwargs)

From <<https://www.javatpoint.com/python-pandas-interview-questions>>

Example-2: Create a DataFrame from dict of ndarrays:

1. **import** pandas as pd
2. info = {'ID': [101, 102, 103], 'Department': ['B.Sc', 'B.Tech', 'M.Tech',]}
3. info = pd.DataFrame(info)
4. print (info)

Output:

	ID	Department
0	101	B.Sc
1	102	B.Tech
2	103	M.Tech

From <<https://www.javatpoint.com/python-pandas-interview-questions>>

Create a Series from dict:

We can also create a Series from dict. If the dictionary object is being passed as an input and the index is not specified, then the dictionary keys are taken in a sorted order to construct the index.

If index is passed, then values correspond to a particular label in the index will be extracted from the dictionary.

5. **import** pandas as pd
6. **import** numpy as np
7. info = {'x': 0., 'y': 1., 'z': 2.}
8. a = pd.Series(info)

9. print (a)

Output:

```
x    0.0
y    1.0
z    2.0
dtype: float64
```

From <<https://www.javatpoint.com/python-pandas-interview-questions>>

Create an empty DataFrame:

The below code shows how to create an empty DataFrame in Pandas:

10. # importing the pandas library

11. **import** pandas as pd

12. info = pd.DataFrame()

13. print (info)

Output:

```
Empty DataFrame
Columns: []
Index: []
```

From <<https://www.javatpoint.com/python-pandas-interview-questions>>

How will you add a column to a pandas DataFrame?

We can add any new column to an existing DataFrame. The below code demonstrates how to add any new column to an existing DataFrame:

14. # importing the pandas library

15. **import** pandas as pd

16. info = {'one': pd.Series([1, 2, 3, 4, 5], index=['a', 'b', 'c', 'd', 'e']),

17. 'two': pd.Series([1, 2, 3, 4, 5, 6], index=['a', 'b', 'c', 'd', 'e', 'f'])}

18.

19. info = pd.DataFrame(info)

20.

21. # Add a **new** column to an existing DataFrame object

22.

23. print ("Add new column by passing series")

24. info['three']=pd.Series([20,40,60],index=['a','b','c'])

25. print (info)

26. print ("Add new column using existing DataFrame columns")

27. info['four']=info['one']+info['three']

28. print (info)

Output:

Add new column by passing series

	one	two	three
a	1.0	1	20.0
b	2.0	2	40.0
c	3.0	3	60.0
d	4.0	4	NaN
e	5.0	5	NaN
f	NaN	6	NaN

Add new column using existing DataFrame columns

	one	two	three	four
a	1.0	1	20.0	21.0
b	2.0	2	40.0	42.0

From <<https://www.javatpoint.com/python-pandas-interview-questions>>

How to iterate over a Pandas DataFrame?

You can iterate over the rows of the DataFrame by using for loop in combination with an `iterrows()` call on the DataFrame.

From <<https://www.javatpoint.com/python-pandas-interview-questions>>

How can we convert DataFrame into an excel file?

We can export the DataFrame to the excel file by using the `to_excel()` function.

From <<https://www.javatpoint.com/python-pandas-interview-questions>>

How to Reset the index?

The Reset index of the DataFrame is used to reset the index by using the '`reset_index`' command. If the DataFrame has a MultiIndex, this method can remove one or more levels.

From <<https://www.javatpoint.com/python-pandas-interview-questions>>

Describe Data Operations in Pandas?

In Pandas, there are different useful data operations for DataFrame, which are as follows:

- **Row and column selection**

We can select any row and column of the DataFrame by passing the name of the rows and columns. When you select it from the DataFrame, it becomes one-dimensional and considered as Series.

- **Filter Data**

We can filter the data by providing some of the boolean expressions in DataFrame.

- **Null values**

A Null value occurs when no data is provided to the items. The various columns may contain no values, which are usually represented as NaN.

From <<https://www.javatpoint.com/python-pandas-interview-questions>>