How can specific columns in a Pandas DataFrame be converted to a NumPy array?

Authored by

stats writer

June 26, 2024

Pandas DataFrame is a widely used data structure in Python for data analysis and manipulation. It consists of rows and columns similar to a spreadsheet. In certain situations, it may be necessary to convert specific columns of a DataFrame to a NumPy array for further analysis or processing. This can be achieved by using the "values" attribute of the DataFrame, which returns a NumPy array containing the values of the specified columns. This conversion allows for efficient data handling and computation as NumPy arrays are optimized for numerical operations. Additionally, it allows for seamless integration with other scientific Python libraries. The process of converting specific columns to a NumPy array can be easily performed using built-in methods in Pandas, providing a versatile solution for data manipulation and analysis.

Pandas: Convert Specific Columns to NumPy Array

You can use the following methods to convert specific columns in a pandas DataFrame to a NumPy array:

Method 1: Convert One Column to NumPy Array

column_to_numpy = df.to_numpy()

Method 2: Convert Multiple Columns to NumPy Array

columns_to_numpy = df].to_numpy()

The following examples show how to use each method in practice with the following pandas DataFrame:

import pandas as pd

#create DataFrame

```
df = pd.DataFrame({'team': ,
'points': ,
'assists': ,
'rebounds': })

#view DataFrame
print(df)

team points assists rebounds
0 A 18 5 11
1 B 22 7 8
2 C 19 7 10
3 D 14 9 6
4 E 14 12 6
5 F 11 9 5
6 G 20 9 9
7 H 28 4 12
```

Example 1: Convert One Column to NumPy Array

The following code shows how to convert the points column in the DataFrame to a NumPy array:

```
#convert points column to NumPy array
column_to_numpy = df.to_numpy()
```

#view result
print(column_to_numpy)

We can confirm that the result is indeed a NumPy array by using the type() function:

#view data type
print(type(column_to_numpy))

<class 'numpy.ndarray'>

Example 2: Convert Multiple Columns to NumPy Array

The following code shows how to convert the team and assists columns in the DataFrame to a multidimensional NumPy array:

#convert team and assists columns to NumPy array
columns_to_numpy = df].to_numpy()

#view result
print(columns_to_numpy)

]

We can confirm that the result is indeed a NumPy array by using the type() function:

```
#view data type
print(type(columns_to_numpy))
```

```
<class 'numpy.ndarray'>
```

```
#view shape of array
print(columns_to_numpy.shape)
```

```
(8, 2)
```

We can see that the resulting NumPy array has 8 rows and 2 columns.

The following tutorials explain how to perform other common tasks in NumPy: