

How can I slice a Pandas DataFrame into chunks?

Authored by
stats writer

June 25, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I slice a Pandas DataFrame into chunks?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=152532>

The process of dividing a Pandas DataFrame into smaller sections is known as slicing. This can be achieved by using the built-in functions and methods available in Pandas. By specifying the desired size or number of chunks, the DataFrame can be divided into equal or uneven sections. This allows for better data manipulation and analysis, as well as easier handling of large datasets. Slicing a DataFrame into chunks is a useful technique in data science and can be easily implemented using Pandas.

Slice Pandas DataFrame into Chunks

You can use the following basic syntax to slice a pandas DataFrame into smaller chunks:

```
#specify number of rows in each chunk  
n=3#split DataFrame into chunks  
list_df = [df.iloc[i:i+n]]
```

You can then access each chunk by using the following syntax:

```
#access first chunklist_df[0]
```

The following example shows how to use this syntax in practice.

Example: Split Pandas DataFrame into Chunks

Suppose we have the following pandas DataFrame with nine rows that contain information about various

basketball players:

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

8 I 23 11 10

We can use the following syntax to split the DataFrame

into chunks where each chunk has 3 rows:

```
#specify number of rows in each chunk
```

```
n=3#split DataFrame into chunks
```

```
list_df = [df.iloc[i:i+n]]
```

We can then use the following syntax to access each chunk:

```
#view first chunk
```

```
print(list_df[0])
```

```
team points assists rebounds
```

```
0 A 18 5 11
```

```
1 B 22 7 8
```

```
2 C 19 7 10
```

```
#view second chunk
```

```
print(list_df[1])
```

```
team points assists rebounds
```

```
3 D 14 9 6
```

```
4 E 14 12 6
```

```
5 F 11 9 5
```

```
#view third chunk
```

```
print(list_df)
```

```
team points assists rebounds
```

```
6 G 20 9 9
```

```
7 H 28 4 12
```

```
8 I 23 11 10
```

Notice that each chunk contains three rows, just as we specified.

Note that in this example we used a DataFrame with only nine rows as a simple example.

In practice, you'll likely be working with a DataFrame with hundreds of thousands or even millions of rows.

You can use the same syntax that we used in this example to split your DataFrame into chunks of specific sizes.