

How can I select only numeric columns in Pandas?

Authored by
stats writer

June 27, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I select only numeric columns in Pandas?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=155041>

Pandas is a popular Python library used for data manipulation and analysis. It offers various functions to select specific columns from a dataset. To select only numeric columns in Pandas, one can use the "select_dtypes()" function and specify the data type as "number". This will filter out all the non-numeric columns and return only the columns with numerical data. This method is efficient and convenient when dealing with large datasets and helps in performing mathematical operations and statistical analysis on only the relevant columns.

Select Only Numeric Columns in Pandas

You can use the following basic syntax to select only numeric columns in a pandas DataFrame:

```
import pandas as pd
import numpy as np
```

```
df.select_dtypes(include=np.number)
```

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

Example: Select Only Numeric Columns in Pandas

Suppose we have the following pandas DataFrame that contains 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
```

We can use the following syntax to select only the numeric columns in the DataFrame:

```
import numpy as np
```

```
#select only the numeric columns in the DataFrame
```

```
df.select_dtypes(include=np.number)
```

```
points assists rebounds
```

```
0 18 5 11
1 22 7 8
2 19 7 10
3 14 9 6
4 14 12 6
5 11 9 5
6 20 9 9
7 28 4 12
```

Notice that only the three numeric columns have been selected - points, assists, and rebounds.

We can verify that these columns are numeric by using the `dtypes()` function to display the data type of each variable in the DataFrame:

```
#display data type of each variable in DataFrame
df.dtypes
```

```
team object
```

```
points int64
```

```
assists int64
```

```
rebounds int64
```

```
dtype: object
```

From the output we can see that team is an object (i.e. string) while points, assists, and rebounds are all numeric.

Note that we can also use the following code to get a list of the numeric columns in the DataFrame:

```
#display list of numeric variables in DataFrame  
df.select_dtypes(include=np.number).columns.tolist()
```

This allows us to quickly see the names of the numeric variables in the DataFrame without seeing their actual values.

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