

How can I drop a specific column while importing a CSV file using Pandas?

Authored by
stats writer

June 25, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I drop a specific column while importing a CSV file using Pandas?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=152331>

Pandas is a popular open-source library in Python used for data manipulation and analysis. One of its key features is the ability to import and export data from various sources, including CSV files. When importing a CSV file using Pandas, there may be instances where you want to drop a specific column from the dataset. This can be easily achieved by using the "drop" function, which allows you to specify the name of the column you want to remove. By using this function, you can efficiently manipulate your data and remove unnecessary columns, saving time and improving the accuracy of your analysis.

Pandas: Drop Specific Column when Importing CSV File

You can use the following basic syntax to drop a specific column when importing a CSV file into a pandas DataFrame:

```
df = pd.read_csv('basketball_data.csv', usecols=lambda x: x != 'rebounds')
```

This particular example will read each column from a CSV file called `basketball_data.csv` into a pandas DataFrame except for the column called `rebounds`.

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

Example: Drop Specific Column when Importing CSV File in Pandas

Suppose we have the following CSV file called `basketball_data.csv`:

```
1 |team,points,rebounds
2 |A, 22, 10
3 |B, 14, 9
4 |C, 29, 6
5 |D, 30, 2
```

We can use the following syntax to import the CSV file into pandas and drop the column called rebounds when importing:

```
import pandas as pd
```

```
#import all columns except 'rebounds' into DataFrame
```

```
df = pd.read_csv('basketball_data.csv', usecols=lambda  
x: x != 'rebounds')
```

```
#view resulting DataFrame
```

```
print(df)
```

```
team points
```

```
0 A 22
```

1 B 14

2 C 29

3 D 30

Notice that the rebounds column was dropped when we imported the CSV file into pandas.

If you would like to drop multiple columns when importing, you can use the not in operator as follows:

```
import pandas as pd
```

```
#import all columns except 'team' and 'rebounds' into DataFrame
```

```
df=pd.read_csv('basketball_data.csv', usecols=lambda x: x not in )
```

```
#view resulting DataFrame
```

```
print(df)
```

```
points
```

```
0 22
```

```
1 14
```

```
2 29
```

```
3 30
```

Notice that the team and rebounds columns were both dropped when we imported the CSV file into pandas.

Note that you can include as many column names as you'd like in the list following the not in operator to drop as many columns as you'd like when importing a CSV file.

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

ARABPSYCHOLOGY.COM