

How can I skip specific columns when importing an Excel file in Pandas?

Authored by
stats writer

June 25, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I skip specific columns when importing an Excel file in Pandas?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=152302>

When importing an Excel file in Pandas, it is possible to skip specific columns by specifying the column names or indices to be excluded. This can be achieved by using the "usecols" parameter and passing a list of columns or indices to skip. This feature allows for more efficient and targeted data import, as unwanted columns can be ignored and not loaded into the data frame.

Pandas: Skip Specific Columns when Importing Excel File

You can use the following basic syntax to skip specific columns when importing an Excel file into a pandas DataFrame:

```
#define columns to skip
```

```
skip_cols =
```

```
#define columns to keep
```

```
keep_cols =
```

```
#import Excel file and skip specific columns
```

```
df = pd.read_excel('my_data.xlsx', usecols=keep_cols)
```

This particular example will skip columns in index positions 1 and 2 when importing the Excel file called my_data.xlsx into pandas.

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

Example: Skip Specific Columns when Importing Excel File into Pandas

Suppose we have the following Excel file called `player_data.xlsx`:

	A	B	C	D	E	F
1	team	points	rebounds	assists		
2	A	24	8	5		
3	B	20	12	3		
4	C	15	4	7		
5	D	19	4	8		
6	E	32	6	8		
7	F	13	7	9		
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						

We can use the following syntax to import this file into a pandas DataFrame and skip the columns in index positions 1 and 2 (the points and rebounds columns) when importing:

```
#define columns to skip
```

```
skip_cols =
```

```
#define columns to keep
```

```
keep_cols =
```

```
#import Excel file and skip specific columns
```

```
df = pd.read_excel('player_data.xlsx',  
usecols=keep_cols)
```

```
#view DataFrameprint(df)
```

```
team assists
```

```
0 A 5
```

```
1 B 3
```

```
2 C 7
```

```
3 D 8
```

```
4 E 8
```

```
5 F 9
```

Notice that all columns in the Excel file except for the columns in index positions 1 and 2 (the points and rebounds columns) were imported into the pandas DataFrame.

Note that this method assumes you know how many columns are in the Excel file beforehand.

Since we knew that there were 4 total columns in the file, we used range(4) when defining which columns we wanted to keep.

Note: You can find the complete documentation for the pandas read_excel() function .

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