

How can I remove special characters from a column in Pandas?"

Authored by
stats writer

June 26, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I remove special characters from a column in Pandas?"*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=153882>

The process of removing special characters from a column in Pandas involves utilizing the `str.replace()` function to replace the special characters with empty strings. This function can be applied to a specific column or the entire dataframe, depending on the desired outcome. Additionally, regular expressions can be used to target specific patterns of special characters. By removing these special characters, the data in the column can be cleaned and made more suitable for analysis or further manipulation.

Pandas: Remove Special Characters from Column

You can use the following basic syntax to remove special characters from a column in a pandas DataFrame:

```
df = df.str.replace('W', "", regex=True)
```

This particular example will remove all characters in my_column that are not letters or numbers.

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

Example: Remove Special Characters from Column 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' : })
```

```
#view DataFrame
```

```
print(df)
```

```
team points
```

```
0 Mavs$ 12
```

```
1 Nets 15
```

```
2 Kings!! 22
```

```
3 Spurs% 29
```

```
4 &Heat& 24
```

Suppose we would like to remove all special characters from values in the team column.

We can use the following syntax to do so:

```
#remove special characters from team column
```

```
df = df.str.replace('W', "", regex=True)
```

```
#view updated DataFrame
```

```
print(df)
```

```
team points
```

```
0 Mavs 12
```

1 Nets 15

2 Kings 22

3 Spurs 29

4 Heat 24

Notice that all special characters have been removed from values in the team column.

Note: The regex `W` is used to find all non-word characters, i.e. characters which are not alphabetical or numerical.

In this example, we replaced each non-word character with an empty value which is equivalent to removing the non-word characters.

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

[How to Replace NaN Values with Zeros in Pandas](#)