

How can I extract a number from a string in Pandas?

Authored by
stats writer

June 25, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I extract a number from a string in Pandas?*.

PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=152205>

Extracting a number from a string in Pandas refers to the process of isolating and retrieving a numerical value from a text data column using the Pandas library. This can be achieved by using various methods such as regular expressions, string manipulation functions, or built-in Pandas functions. The extracted number can then be used for further data analysis or manipulation purposes. This feature is particularly useful when dealing with datasets that contain mixed data types, where numbers may be embedded within strings.

Extract Number from String in Pandas

You can use the following basic syntax to extract numbers from a string in pandas:

```
df.str.extract('(d+)')
```

This particular syntax will extract the numbers from each string in a column called my_column in a pandas DataFrame.

Note: When using a regular expression, d represents "any digit" and + stands for "one or more."

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

Example: Extract Number from String in Pandas

Suppose we have the following pandas DataFrame that contains information about the sales of various

products:

```
import pandas as pd
```

```
#create DataFrame
```

```
df = pd.DataFrame({'product': ,  
'sales': })
```

```
#view DataFrame
```

```
print(df)
```

```
product sales
```

```
0 A33 18
```

```
1 B34 22
```

```
2 A22 19
```

```
3 A50 14
```

```
4 C200 14
```

```
5 D7 11
```

```
6 A9 20
```

```
7 A13 28
```

Suppose we would like to extract the number from each string in the product column.

We can use the following syntax to do so:

```
#extract numbers from strings in 'product' column  
df.str.extract('(d+)')
```

0

0 33

1 34

2 22

3 50

4 200

5 7

6 9

7 13

The result is a DataFrame that contains only the numbers from each row in the product column.

For example:

The formula extracts 33 from the string A33 in the first row. The formula extracts 34 from the string B34 in the first row. The formula extracts 22 from the string A22 in the first row.

And so on.

If you'd like, you can also store these numerical values in a new column in the DataFrame:

#extract numbers from strings in 'product' column and store them in new column

```
df = df.str.extract('(d+)')
```

#view updated DataFrame

```
print(df)
```

```
product sales product_numbers
```

```
0 A33 18 33
```

```
1 B34 22 34
```

```
2 A22 19 22
```

```
3 A50 14 50
```

```
4 C200 14 200
```

```
5 D7 11 7
```

```
6 A9 20 9
```

```
7 A13 28 13
```

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