

How can I create a tuple from two columns in Pandas?

Authored by
stats writer

June 27, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I create a tuple from two columns in Pandas?*.

PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=154770>

Creating a tuple from two columns in Pandas involves using the "zip" function to combine the two columns into a single list, and then converting that list into a tuple using the "tuple" function. This process allows for the creation of a tuple with each element representing a pair of values from the two columns. This can be useful for organizing and manipulating data in a Pandas dataframe.

Create a Tuple from Two Columns in Pandas

You can use the following basic syntax to create a tuple from two columns in a pandas DataFrame:

```
df = list(zip(df.column1, df.column2))
```

This particular formula creates a new column called `new_column`, which is a tuple formed by `column1` and `column2` in the DataFrame.

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

Example: Create Tuple from Two 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': })
```

```
#view DataFrame
```

```
print(df)
```

```
team points assists
```

```
0 A 18 5
```

```
1 B 22 7
```

```
2 C 19 7
```

```
3 D 14 9
```

```
4 E 14 12
```

```
5 F 11 9
```

```
6 G 20 9
```

```
7 H 28 4
```

We can use the following syntax to create a new column called `points_assists`, which is a tuple formed by the values in the `points` and `assists` columns:

```
#create new column that is a tuple of points and assists  
columns
```

```
df = list(zip(df.points, df.assists))
```

```
#view updated DataFrame
```

```
print(df)
```

```
team points assists points_assists
```

```
0 A 18 5 (18, 5)
```

```
1 B 22 7 (22, 7)
```

```
2 C 19 7 (19, 7)
```

```
3 D 14 9 (14, 9)
```

```
4 E 14 12 (14, 12)
```

```
5 F 11 9 (11, 9)
```

```
6 G 20 9 (20, 9)
```

```
7 H 28 4 (28, 4)
```

The new column called `points_assists` is a tuple formed by the `points` and `assists` columns.

Note that you can also include more than two columns in a tuple if you'd like.

For example, the following code shows how to create a tuple that uses values from all three original columns in the DataFrame:

```
#create new column that is a tuple of team, points and  
assists columns
```

```
df = list(zip(df.team, df.points, df.assists))
```

```
#view updated DataFrame  
print(df)
```

```
team points assists all_columns
```

```
0 A 18 5 (A, 18, 5)
```

```
1 B 22 7 (B, 22, 7)
```

```
2 C 19 7 (C, 19, 7)
```

```
3 D 14 9 (D, 14, 9)
```

```
4 E 14 12 (E, 14, 12)
```

```
5 F 11 9 (F, 11, 9)
```

```
6 G 20 9 (G, 20, 9)
```

```
7 H 28 4 (H, 28, 4)
```

You can use this same basic syntax to create a tuple column with as many columns as you'd like.

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