How can I create a scatter plot using multiple columns in Pandas?

Authored by

stats writer

June 25, 2024

Creating a scatter plot using multiple columns in Pandas involves using the Pandas library to import and organize data into a dataframe, selecting the desired columns for the plot, and using the plot function to display the data points as a scatter plot with the selected columns as the x and y variables. This allows for visualizing the relationship between multiple variables in a dataset and identifying any patterns or correlations.

Pandas: Create Scatter Plot Using Multiple Columns

You can use the following basic syntax to create a scatter plot using multiple columns in a pandas DataFrame:

```
import pandas as pd

#create scatter plot of A vs. B
ax1 = df.plot(kind='scatter', x='A', y='B', color='r')

#add scatter plot on same graph of C vs. D
ax2 = df.plot(kind='scatter', x='C', y='D', color='g', ax=ax1)
```

This particular example creates a scatter plot using columns A and B, then overlays another scatter plot on the same graph using columns C and D.

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

Example: Create Pandas Scatter Plot Using Multiple Columns

Suppose we have the following pandas DataFrame that shows the points and assists for various basketball players on teams A and B:

```
import pandas as pd

#create DataFrame
df = pd.DataFrame({'A_assists': ,
'A_points': ,
'B_assists': ,
'B_points': })
#view DataFrame
print(df)
```

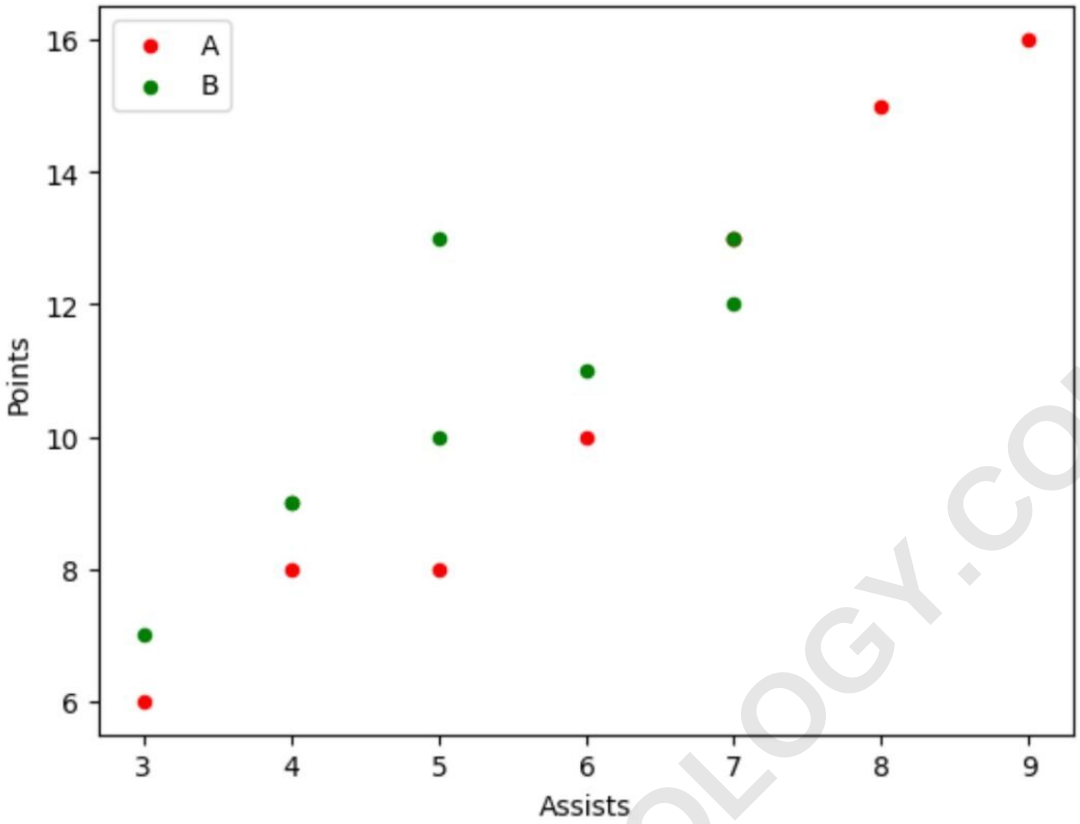
```
A_assists A_points B_assists B_points
0 3 6 3 7
1 4 8 4 9
2 5 8 4 9
3 6 10 5 13
4 7 13 5 10
5 7 13 6 11
6 8 15 7 12
7 9 16 7 13
```

We can use the following syntax to create a scatter plot using columns A_assists and A_points, then overlay another scatter plot on the same graph using columns B_assists and B_points:

```
#create scatter plot of A_assists vs. A_points
ax1=df.plot(kind='scatter', x='A_assists', y='A_points', color='r', label='A')

#add scatter plot on same graph using B_assists vs. B_points
ax2=df.plot(kind='scatter', x='B_assists', y='B_points', color='g', label='B', ax=ax1)

#specify x-axis and y-axis labels
ax1.set_xlabel('Assists')
ax1.set_ylabel('Points')
```



The end result is a scatter plot that contains the values in the columns A_assists and A_points in red and the values in the columns B_assists and B_points in green.

Note #1: The label argument specifies the label to use in the legend of the plot.

Note #2: In this example, we used two groups of columns to plot two scatter plots on the same graph. However, you could use ax3, ax4, etc. to add as many columns as you'd like to the scatter plot.

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