

How can I compare two columns in Pandas and what are some examples to do so?

Authored by
stats writer

April 27, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I compare two columns in Pandas and what are some examples to do so?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=139852>

Pandas is a popular Python library used for data analysis. It provides a variety of functions and methods for comparing data, including the ability to compare two columns in a dataset. To compare two columns in Pandas, the user can use the "equals()" method or the "==" operator. This will return a boolean value indicating whether the two columns are equal or not. Other comparison methods such as "greater than", "less than", and "not equals" can also be used to compare columns.

For example, if we have a dataset with two columns, "Age" and "Income", we can use the "equals()" method to compare these two columns and check if they are equal. If we want to find the rows where the "Age" column is greater than the "Income" column, we can use the ">" operator. Similarly, we can use the "!=" operator to find the rows where the "Age" column is not equal to the "Income" column. These comparison methods can provide valuable insights into the data and help in making data-driven decisions.

Compare Two Columns in Pandas (With Examples)

Often you may want to compare two columns in a Pandas DataFrame and write the results of the comparison to a third column.

You can easily do this by using the following syntax:

conditions=

choices=

df=np.select(conditions, choices, default)

Here's what this code does:

conditions are the conditions to check for between the two columns
choices are the results to return based on

the conditionsnp.select is used to return the results to the new column

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

Example: Compare Two Columns in Pandas

Suppose we have the following DataFrame that shows the number of goals scored by two soccer teams in five different matches:

```
import numpy as np
import pandas as pd

#create DataFrame
df = pd.DataFrame({'A_points': ,
'B_points': })

#view DataFrame
df
```

	A_points	B_points
0	1	4
1	3	5
2	3	2

3 3 3

4 5 2

We can use the following code to compare the number of goals by row and output the winner of the match in a third column:

```
#define conditions
```

```
conditions = > df,
```

```
df < df]
```

```
#define choices
```

```
choices =
```

```
#create new column in DataFrame that displays results  
of comparisons
```

```
df = np.select(conditions, choices, default='Tie')
```

```
#view the DataFrame
```

```
df
```

```
A_points B_points winner
```

```
0 1 4 B
```

```
1 3 5 B
```

```
2 3 2 A
```

3 3 3 Tie

4 5 2 A

The results of the comparison are shown in the new column called *winner*.

Notes

Here are a few things to keep in mind when comparing two columns in a pandas DataFrame:

The number of conditions and choices should be equal. The default value specifies the value to display in the new column if none of the conditions are met. Both NumPy and Pandas are required to make this code work.

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