How do I use Pandas to find the unique values in a column and sort them?

Authored by

stats writer

June 24, 2024

Pandas is a popular Python library used for data analysis and manipulation. One of its useful features is the ability to find unique values in a column and sort them. To do this, first import the Pandas library and read in your data as a Pandas DataFrame. Then, use the unique() function on the desired column to retrieve all the unique values. Finally, use the sort_values() function to sort the values in ascending or descending order. This approach allows for efficient identification of distinct values in a dataset and enables further analysis and processing.

Pandas: Find Unique Values in Column and Sort Them

You can use the following basic syntax to find the unique values in a column of a pandas DataFrame and then sort them:

df.drop_duplicates().sort_values()

This will return a pandas Series that contains each unique value in a column sorted in ascending order.

To instead sort the unique values in descending order, use ascending=False:

df.drop_duplicates().sort_values(ascending=False)

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

Example: Find Unique Values in Pandas Column and Sort Them

Suppose we have the following pandas DataFrame:

```
import pandas as pd

#create DataFrame
df = pd.DataFrame({'team': ,
'points': })

#view DataFrame
print(df)
```

team points
0 A 5
1 A 5
2 A 9
3 A 12
4 A 12
5 B 5
6 B 10
7 B 13
8 B 13
9 B 19

We can use the following syntax to get the unique values from the points column and then sort them in

ascending order:

#get unique values in points column and sort them
df.drop_duplicates().sort_values()

0 5
2 9
6 10
3 12
7 13
9 19
Name: points, dtype: int64

The output displays each of the unique values in the points column sorted in ascending order:

5910121319

We can also get the unique values in the points column sorted in descending order by specifying ascending=False within the sort_values() function:

#get unique values in points column and sort them in descending order
df.drop_duplicates().sort_values(ascending=False)

9 19

7 13

3 12

6 10

2 9

0 5

Name: points, dtype: int64

The output displays each of the unique values in the points column sorted in descending order:

1913121095

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