

How can I select rows between two dates in Pandas?

Authored by
stats writer

June 27, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I select rows between two dates in Pandas?*.

PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=154639>

Pandas is a popular Python library used for data analysis and manipulation. It offers a variety of functions and methods for selecting and filtering data in a dataframe. One useful feature is the ability to select rows between two specific dates. This can be achieved by using the "between" function, which allows the user to specify a start and end date range. It also offers the flexibility to include or exclude the specified dates. This function is useful for tasks such as analyzing time series data, where selecting data within a certain time period is essential. By using the "between" function in Pandas, users can efficiently and accurately extract the desired data from a dataframe based on specific date criteria.

Pandas: Select Rows Between Two Dates

You can use the following syntax to select rows between two specific dates in a pandas DataFrame:

```
df
```

This particular example selects all rows in the DataFrame between 2022-01-02 and 2022-01-06.

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

Example: Select Rows Between Two Dates in Pandas

Suppose we have the following pandas DataFrame:

```
import pandas as pd
```

```
#create DataFrame
```

```
df = pd.DataFrame({'date':  
pd.date_range(start='1/1/2022', periods=8),  
'sales': ,  
'returns': })
```

```
#view DataFrame
```

```
print(df)
```

```
date sales returns
```

```
0 2022-01-01 18 5
```

```
1 2022-01-02 20 7
```

```
2 2022-01-03 15 7
```

```
3 2022-01-04 14 9
```

```
4 2022-01-05 10 12
```

```
5 2022-01-06 9 3
```

```
6 2022-01-07 8 2
```

```
7 2022-01-08 12 4
```

We can use the following syntax to select only the rows that fall between the date 2022-01-02 and 2022-01-06:

```
#select all rows where date is between 2022-01-02 and  
2022-01-06
```

```
df
```

```
date sales returns
```

```
1 2022-01-02 20 7
```

```
2 2022-01-03 15 7
```

```
3 2022-01-04 14 9
```

```
4 2022-01-05 10 12
```

```
5 2022-01-06 9 3
```

Notice that only the rows between the dates 2022-01-02 and 2022-01-06 are selected.

If you'd like, you can also define the start and end dates outside of the `between()` function:

```
#define start and end dates
```

```
start_date = '2022-01-02'
```

```
end_date = '2022-01-06'
```

```
#select all rows where date is between start and end  
df
```

```
date sales returns
```

```
1 2022-01-02 20 7
```

```
2 2022-01-03 15 7
```

```
3 2022-01-04 14 9
```

```
4 2022-01-05 10 12
```

5 2022-01-06 9 3

This produces the same result.

Note that if your date column is not in a recognizable datetime format, you may first need to use the following code to convert it to a datetime format:

```
df = pd.to_datetime(df)
```

Once you've done this, you can proceed to use the `between()` function to select rows between specific dates.