

PANDAS PART 6


Window Functions

- Cumulative / Expanding window
- Rolling window

A window represents one or more column values of a dataframe. Window functions are functions that take data from a window and do calculations on that data. Window functions are used mainly in Time series analysis for finding trends and patterns in the data.

What is a window and window function?

A	Cumulative SUM	Rolling SUM
1	1	NaN
3	4	4
2	6	5
5	11	7

 **window = 2**

```
[1]: import numpy as np
import pandas as pd
```

```
[41]: # create a data frame with 2 cols
df = pd.DataFrame({'a': [1,2,3,4,5,6,7,8],
                   'b': [10,np.nan, np.nan, 20, np.nan, np.nan, 30, 30]})
df
```

```
[41]:
```

	a	b
0	1	10.0
1	2	NaN
2	3	NaN
3	4	20.0
4	5	NaN
5	6	NaN
6	7	30.0
7	8	30.0

```
[42]: # find cumulative sum of column 'a'
df['a_cumsum'] = df['a'].cumsum()
df
# similarly calculate cumprod(), cummin(), cummax()
```

[42]:

	a	b	a_cumsum
0	1	10.0	1
1	2	NaN	3
2	3	NaN	6
3	4	20.0	10
4	5	NaN	15
5	6	NaN	21
6	7	30.0	28
7	8	30.0	36

```
[43]: # rolling functions
# find rolling sum on col 'a' for window=2
obj = df['a'].rolling(window=2)
df['a_roll_sum'] = obj.sum()
df
```

[43]:

	a	b	a_cumsum	a_roll_sum
0	1	10.0	1	NaN
1	2	NaN	3	3.0
2	3	NaN	6	5.0
3	4	20.0	10	7.0
4	5	NaN	15	9.0
5	6	NaN	21	11.0
6	7	30.0	28	13.0
7	8	30.0	36	15.0

```
[44]: # find rolling count for col 'b' for window=2
# by default window size and min_periods are same in future versions
obj = df['b'].rolling(window=2, min_periods=2)
df['b_roll_count'] = obj.count()
df
# similarly use min(), max(), mean(), std(), var()
```

```
[44]:
```

	a	b	a_cumsum	a_roll_sum	b_roll_count
0	1	10.0	1	NaN	NaN
1	2	NaN	3	3.0	1.0
2	3	NaN	6	5.0	0.0
3	4	20.0	10	7.0	1.0
4	5	NaN	15	9.0	1.0
5	6	NaN	21	11.0	0.0
6	7	30.0	28	13.0	1.0
7	8	30.0	36	15.0	2.0

```
[45]: # expanding functions are like cumulative functions
obj = df['a'].expanding(min_periods=1)
df['a_exp_sum'] = obj.sum()
df
# similary try with mean(),
```

```
[45]:
```

	a	b	a_cumsum	a_roll_sum	b_roll_count	a_exp_sum
0	1	10.0	1	NaN	NaN	1.0
1	2	NaN	3	3.0	1.0	3.0
2	3	NaN	6	5.0	0.0	6.0
3	4	20.0	10	7.0	1.0	10.0
4	5	NaN	15	9.0	1.0	15.0
5	6	NaN	21	11.0	0.0	21.0
6	7	30.0	28	13.0	1.0	28.0
7	8	30.0	36	15.0	2.0	36.0

```
[46]: # find mean on expanding window
obj = df['a'].expanding(min_periods=1 )
df['a_exp_mean'] = obj.mean()
df
# Similarly try max(), min(), std(), var()
```

[46]:

	a	b	a_cumsum	a_roll_sum	b_roll_count	a_exp_sum	a_exp_mean
0	1	10.0	1	NaN	NaN	1.0	1.0
1	2	NaN	3	3.0	1.0	3.0	1.5
2	3	NaN	6	5.0	0.0	6.0	2.0
3	4	20.0	10	7.0	1.0	10.0	2.5
4	5	NaN	15	9.0	1.0	15.0	3.0
5	6	NaN	21	11.0	0.0	21.0	3.5
6	7	30.0	28	13.0	1.0	28.0	4.0
7	8	30.0	36	15.0	2.0	36.0	4.5