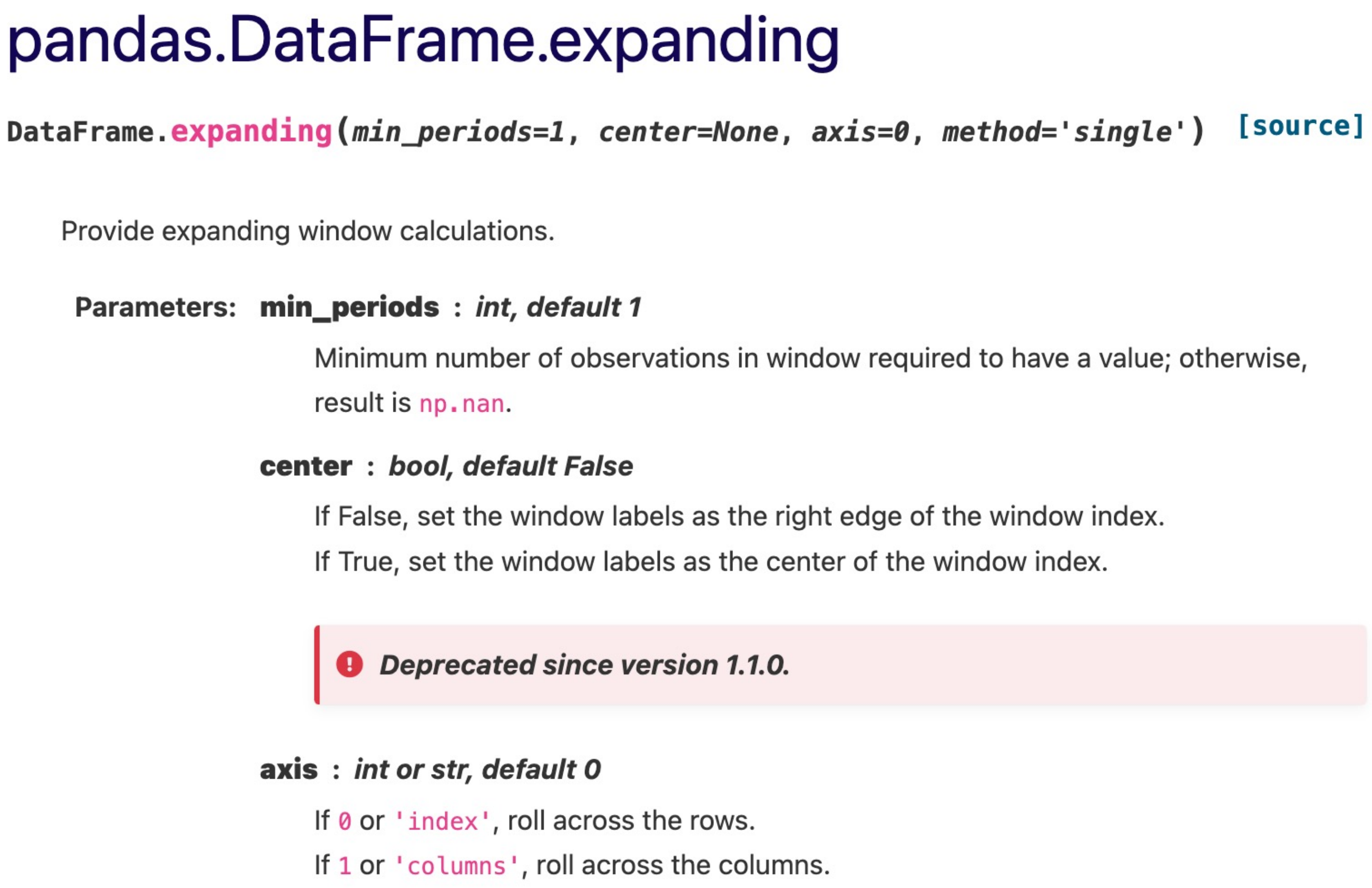
Expanding window features: part 3

Window features

Expanding window with Pandas



pandas.DataFrame.expanding

`DataFrame.expanding(min_periods=1, center=None, axis=0, method='single')` [source]

Provide expanding window calculations.

Parameters: **min_periods** : *int, default 1*

Minimum number of observations in window required to have a value; otherwise, result is `np.nan`.

center : *bool, default False*

If False, set the window labels as the right edge of the window index.

If True, set the window labels as the center of the window index.

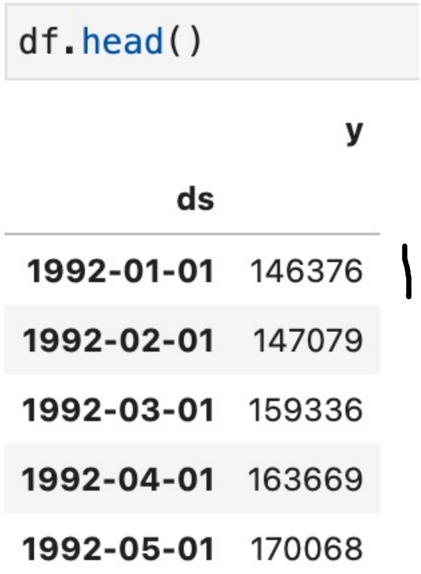
Deprecated since version 1.1.0.

axis : *int or str, default 0*

If `0` or `'index'`, roll across the rows.

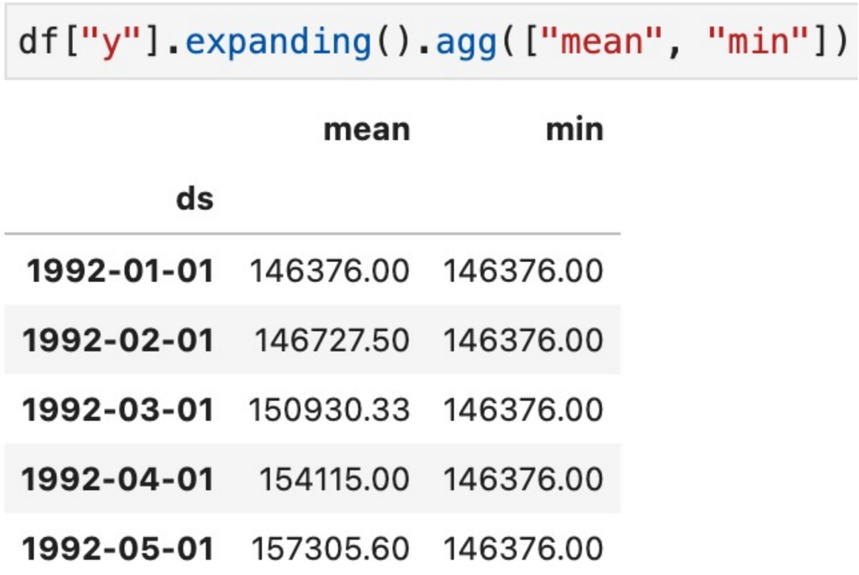
If `1` or `'columns'`, roll across the columns.

Expanding window with Pandas



```
df.head()
```

	y
ds	
1992-01-01	146376
1992-02-01	147079
1992-03-01	159336
1992-04-01	163669
1992-05-01	170068

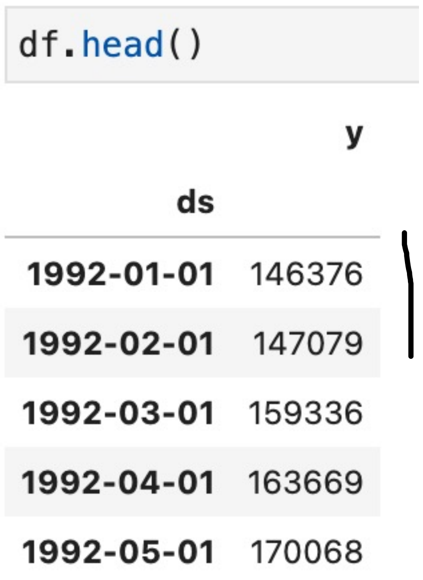


```
df["y"].expanding().agg(["mean", "min"])
```

	mean	min
ds		
1992-01-01	146376.00	146376.00
1992-02-01	146727.50	146376.00
1992-03-01	150930.33	146376.00
1992-04-01	154115.00	146376.00
1992-05-01	157305.60	146376.00

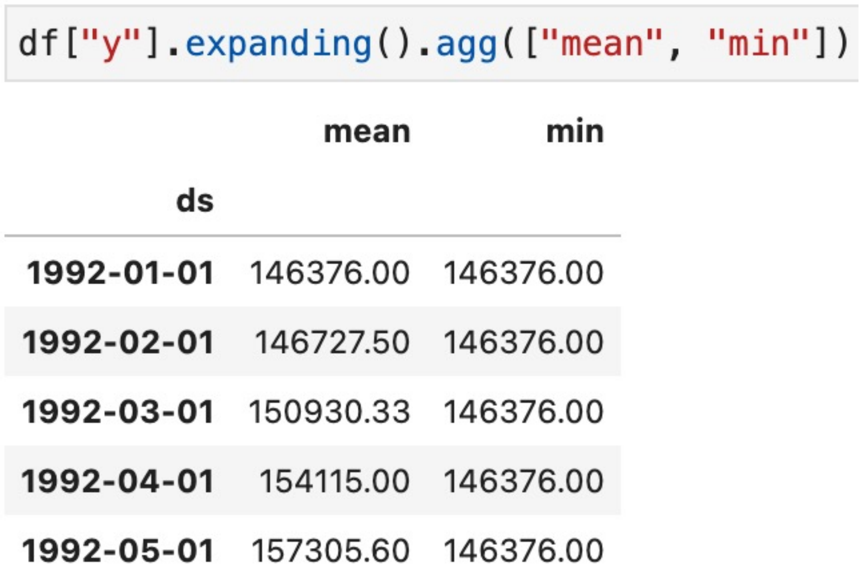
The `expanding` method by default assigns the rolling statistics to the edge of the window.

Expanding window with Pandas



```python
df.head()
```

ds	y
1992-01-01	146376
1992-02-01	147079
1992-03-01	159336
1992-04-01	163669
1992-05-01	170068

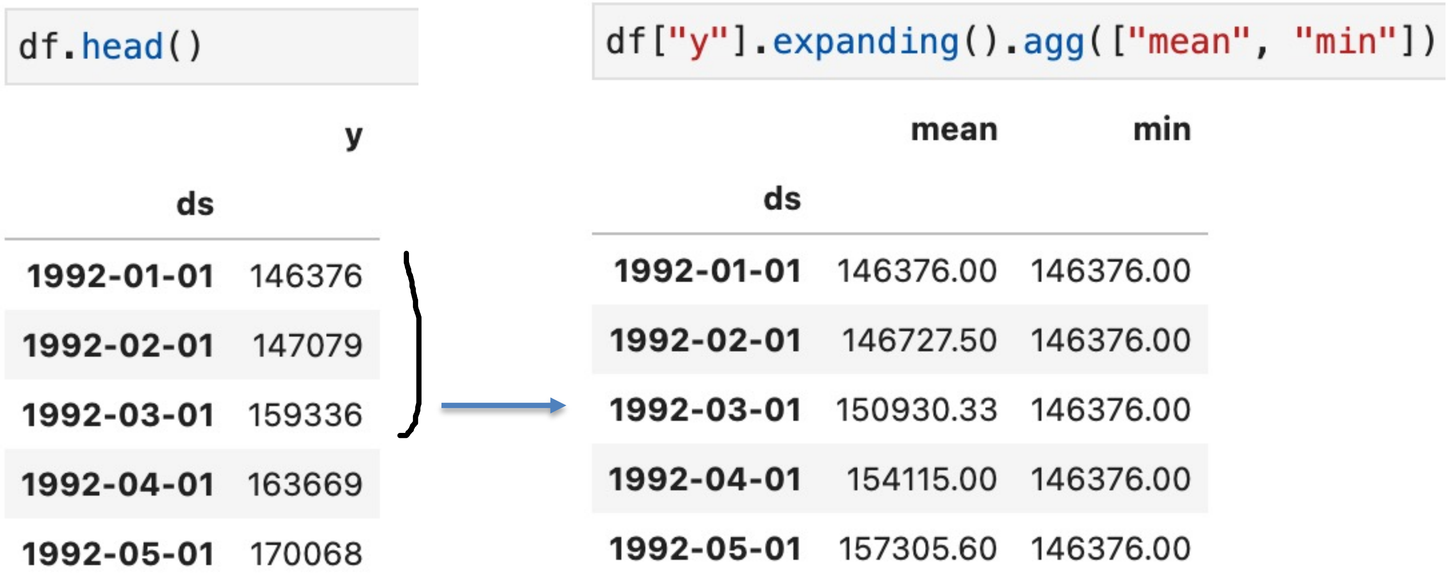


```python
df["y"].expanding().agg(["mean", "min"])
```

ds	mean	min
1992-01-01	146376.00	146376.00
1992-02-01	146727.50	146376.00
1992-03-01	150930.33	146376.00
1992-04-01	154115.00	146376.00
1992-05-01	157305.60	146376.00

The `expanding` method by default assigns the rolling statistics to the edge of the window.

Expanding window with Pandas



```
df.head()
```

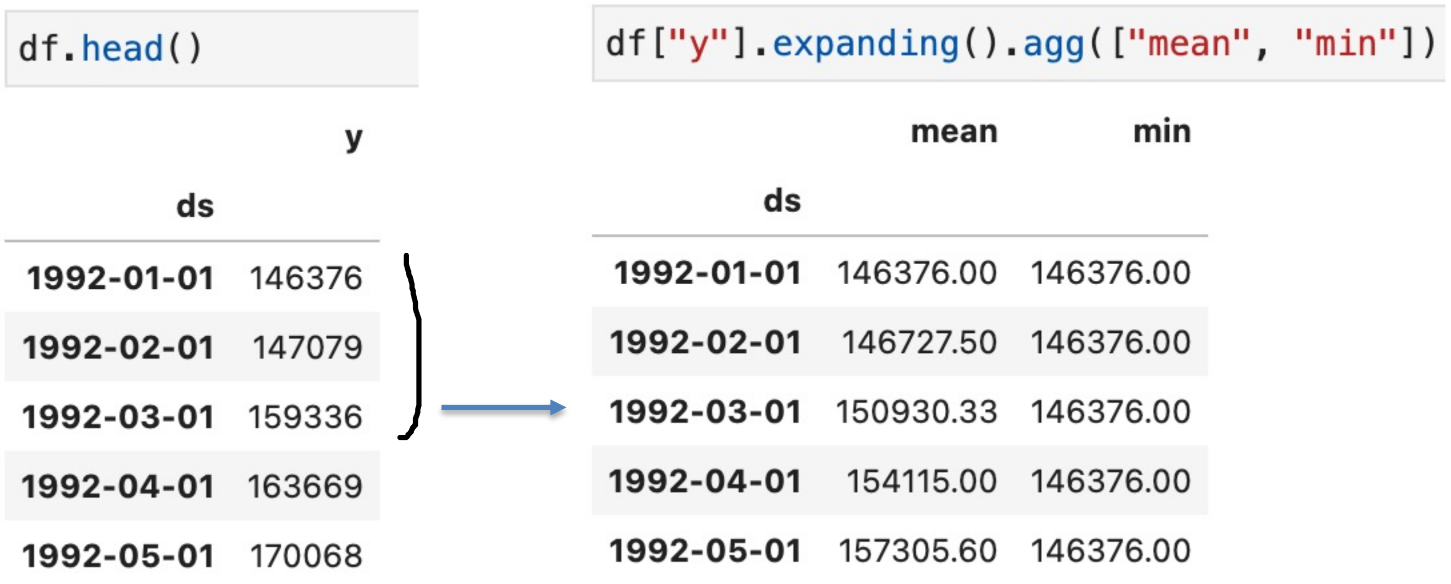
ds	y
1992-01-01	146376
1992-02-01	147079
1992-03-01	159336
1992-04-01	163669
1992-05-01	170068

```
df["y"].expanding().agg(["mean", "min"])
```

ds	mean	min
1992-01-01	146376.00	146376.00
1992-02-01	146727.50	146376.00
1992-03-01	150930.33	146376.00
1992-04-01	154115.00	146376.00
1992-05-01	157305.60	146376.00

The `expanding` method by default assigns the rolling statistics to the edge of the window.

Expanding window with Pandas



```
df.head()
```

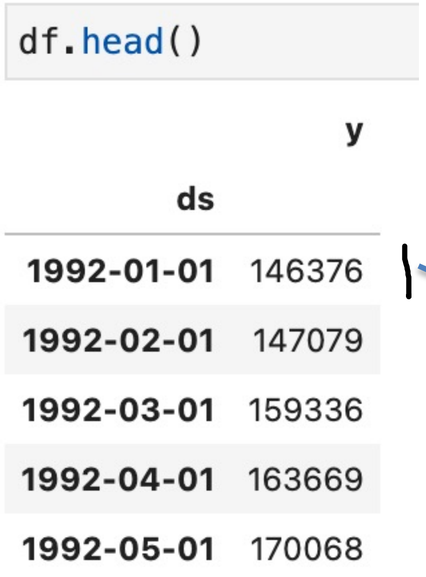
	y
ds	
1992-01-01	146376
1992-02-01	147079
1992-03-01	159336
1992-04-01	163669
1992-05-01	170068

```
df["y"].expanding().agg(["mean", "min"])
```

	mean	min
ds		
1992-01-01	146376.00	146376.00
1992-02-01	146727.50	146376.00
1992-03-01	150930.33	146376.00
1992-04-01	154115.00	146376.00
1992-05-01	157305.60	146376.00

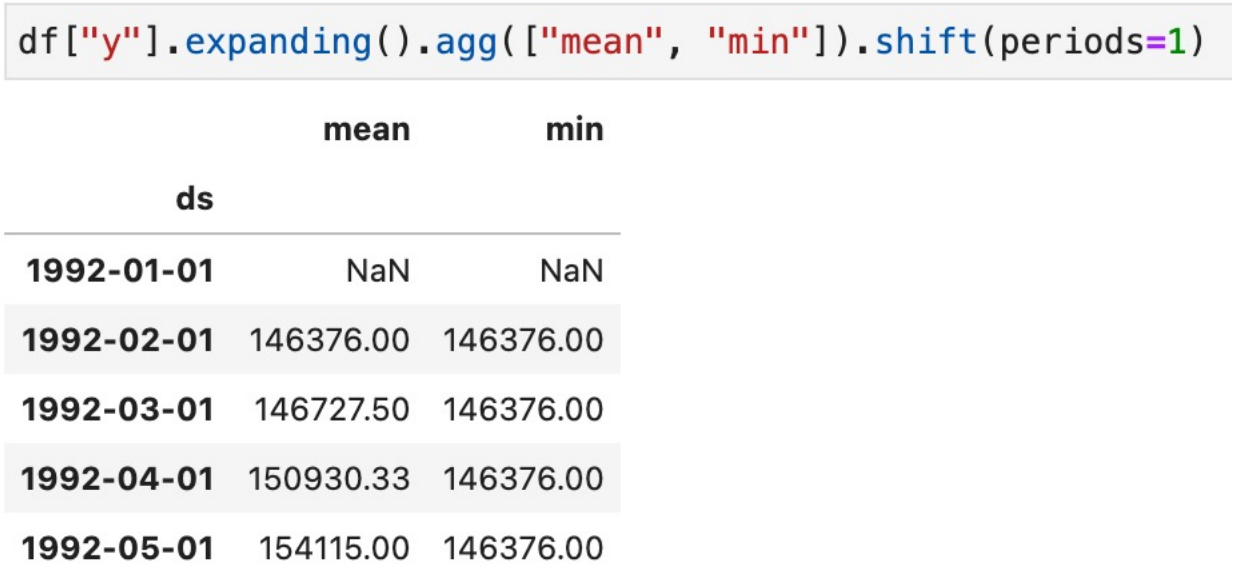
For forecasting we want features which only use information that we will have at predict time (i.e, the past). So we want to shift the output of the row down by one to avoid data leakage.

Expanding window with Pandas



```
df.head()
```

	y
ds	
1992-01-01	146376
1992-02-01	147079
1992-03-01	159336
1992-04-01	163669
1992-05-01	170068

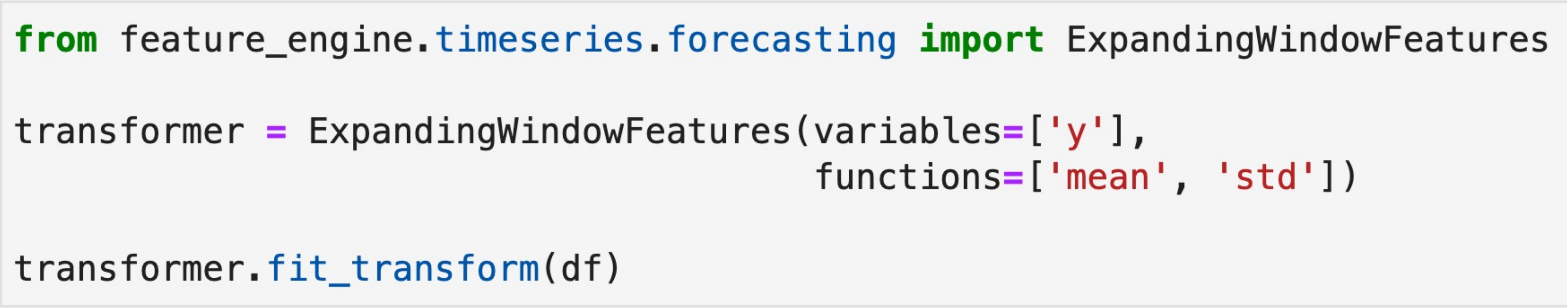


```
df["y"].expanding().agg(["mean", "min"]).shift(periods=1)
```

	mean	min
ds		
1992-01-01	NaN	NaN
1992-02-01	146376.00	146376.00
1992-03-01	146727.50	146376.00
1992-04-01	150930.33	146376.00
1992-05-01	154115.00	146376.00

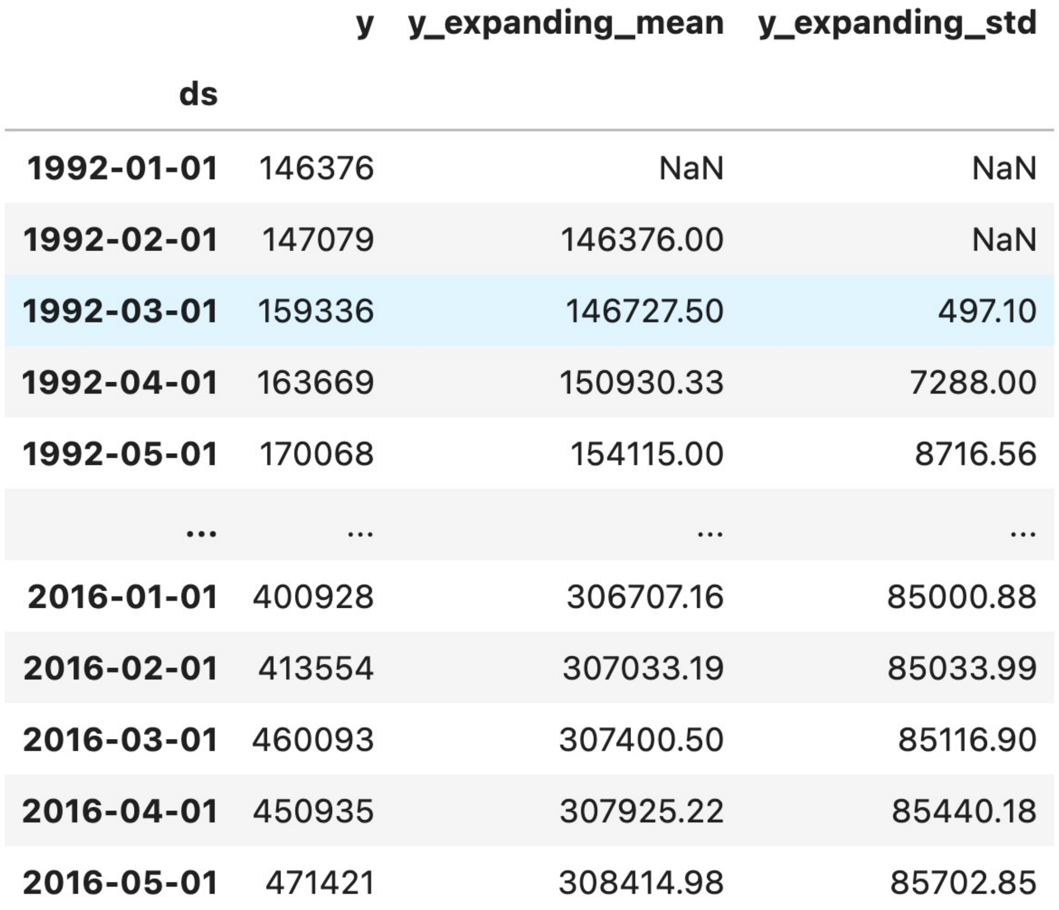
For forecasting we want features which only use information that we will have at predict time (i.e, the past). So we want to shift the output of the row down by one to avoid data leakage.

Expanding window with Feature-engine



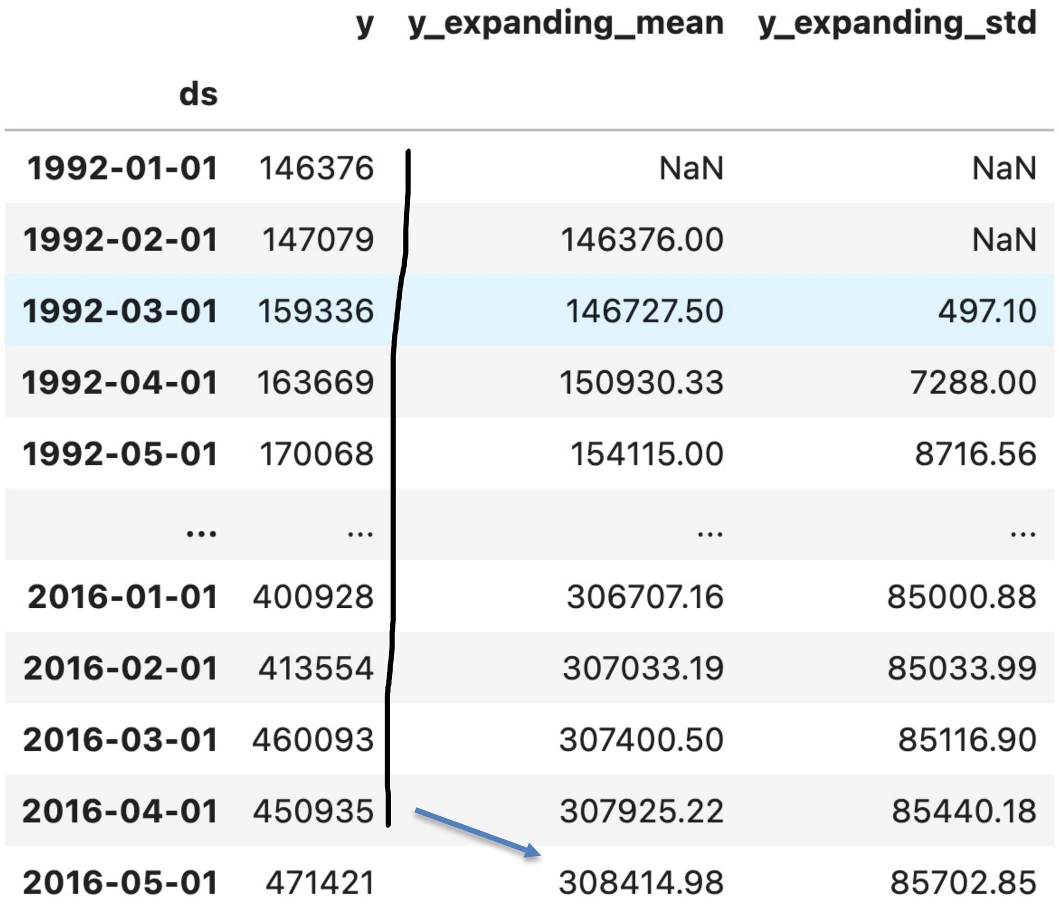
```
from feature_engine.timeseries.forecasting import ExpandingWindowFeatures

transformer = ExpandingWindowFeatures(variables=['y'],
                                      functions=['mean', 'std'])

transformer.fit_transform(df)
```

Expanding window with Feature-engine



	y	y_expanding_mean	y_expanding_std
ds			
1992-01-01	146376	NaN	NaN
1992-02-01	147079	146376.00	NaN
1992-03-01	159336	146727.50	497.10
1992-04-01	163669	150930.33	7288.00
1992-05-01	170068	154115.00	8716.56
...	...	...	...
2016-01-01	400928	306707.16	85000.88
2016-02-01	413554	307033.19	85033.99
2016-03-01	460093	307400.50	85116.90
2016-04-01	450935	307925.22	85440.18
2016-05-01	471421	308414.98	85702.85

Expanding window with Feature-engine



ds	y	y_expanding_mean	y_expanding_std
1992-01-01	146376	NaN	NaN
1992-02-01	147079	146376.00	NaN
1992-03-01	159336	146727.50	497.10
1992-04-01	163669	150930.33	7288.00
1992-05-01	170068	154115.00	8716.56
...	...	...	...
2016-01-01	400928	306707.16	85000.88
2016-02-01	413554	307033.19	85033.99
2016-03-01	460093	307400.50	85116.90
2016-04-01	450935	307925.22	85440.18
2016-05-01	471421	308414.98	85702.85

Summary

Expanding windows include all the data prior to the end of the window.

They can be helpful to capture cumulative effects and are used in target encoding.

Not used as often as rolling windows as more recent data tends to be more predictive.