

6.1 - Columns

April 11, 2017

```
In [1]: import pandas as pd
```

```
In [2]: df = pd.DataFrame(index=range(2))
```

```
In [3]: def fill_df1(df):  
        mydf = df.copy()  
        for i in range(10000):  
            mydf[i] = [i, -i]  
        return mydf
```

```
In [4]: fill_df1(df)
```

```
   0    1    2    3    4    5    6    7    8    9    ...  9990  \  
0   0    1    2    3    4    5    6    7    8    9    ...  9990  
1   0   -1   -2   -3   -4   -5   -6   -7   -8   -9    ... -9990  
  
   9991  9992  9993  9994  9995  9996  9997  9998  9999  
0  9991  9992  9993  9994  9995  9996  9997  9998  9999  
1 -9991 -9992 -9993 -9994 -9995 -9996 -9997 -9998 -9999
```

```
[2 rows x 10000 columns]
```

```
In [5]: %timeit fill_df1(df)
```

```
1 loop, best of 3: 3.58 s per loop
```

```
In [6]: df2 = pd.DataFrame(index=range(2), columns=range(10000))
```

```
In [7]: %timeit fill_df1(df2)
```

```
1 loop, best of 3: 4.02 s per loop
```

```
In [8]: df3 = pd.DataFrame(-1, index=range(2), columns=range(10000))
```

```
In [9]: df3.dtypes
```

0	int64
1	int64
2	int64
3	int64
4	int64
5	int64
6	int64
7	int64
8	int64
9	int64
10	int64
11	int64
12	int64
13	int64
14	int64
15	int64
16	int64
17	int64
18	int64
19	int64
20	int64
21	int64
22	int64
23	int64
24	int64
25	int64
26	int64
27	int64
28	int64
29	int64
	...
9970	int64
9971	int64
9972	int64
9973	int64
9974	int64
9975	int64
9976	int64
9977	int64
9978	int64
9979	int64
9980	int64
9981	int64
9982	int64
9983	int64
9984	int64
9985	int64

```
9986    int64
9987    int64
9988    int64
9989    int64
9990    int64
9991    int64
9992    int64
9993    int64
9994    int64
9995    int64
9996    int64
9997    int64
9998    int64
9999    int64
dtype: object
```

```
In [10]: %timeit fill_df1(df3)
```

```
1 loop, best of 3: 858 ms per loop
```

```
In [11]: df4 = pd.DataFrame(-1.0, index=range(2), columns=range(10000))
```

```
In [12]: %timeit fill_df1(df4)
```

```
1 loop, best of 3: 3.73 s per loop
```

```
In [13]: df5 = pd.DataFrame(-1.0, index=range(2), columns=range(10000))
```

```
In [14]: df5._data.blocks
```

```
Out[14]: (FloatBlock: slice(0, 10000, 1), 10000 x 2, dtype: float64,)
```

```
In [15]: df5[0] = [0, 0]
```

```
In [16]: df5._data.blocks
```

```
Out[16]: (FloatBlock: slice(1, 10000, 1), 9999 x 2, dtype: float64,
          IntBlock: slice(0, 1, 1), 1 x 2, dtype: int64)
```

```
In [17]: df5[1] = [1, -1]
```

```
In [18]: df5._data.blocks
```

```
Out[18]: (FloatBlock: slice(2, 10000, 1), 9998 x 2, dtype: float64,
          IntBlock: slice(0, 1, 1), 1 x 2, dtype: int64,
          IntBlock: slice(1, 2, 1), 1 x 2, dtype: int64)
```

```
In [19]: df5[2] = [2, -2]
```

```
In [20]: df5._data.blocks
```

```
Out[20]: (FloatBlock: slice(3, 10000, 1), 9997 x 2, dtype: float64,  
          IntBlock: slice(0, 1, 1), 1 x 2, dtype: int64,  
          IntBlock: slice(1, 2, 1), 1 x 2, dtype: int64,  
          IntBlock: slice(2, 3, 1), 1 x 2, dtype: int64)
```

```
In [21]: m = df5.max()
```

```
In [22]: m
```

```
0      0.0  
1      1.0  
2      2.0  
3     -1.0  
4     -1.0  
5     -1.0  
6     -1.0  
7     -1.0  
8     -1.0  
9     -1.0  
10    -1.0  
11    -1.0  
12    -1.0  
13    -1.0  
14    -1.0  
15    -1.0  
16    -1.0  
17    -1.0  
18    -1.0  
19    -1.0  
20    -1.0  
21    -1.0  
22    -1.0  
23    -1.0  
24    -1.0  
25    -1.0  
26    -1.0  
27    -1.0  
28    -1.0  
29    -1.0  
...  
9970  -1.0  
9971  -1.0  
9972  -1.0  
9973  -1.0
```

```
9974 -1.0
9975 -1.0
9976 -1.0
9977 -1.0
9978 -1.0
9979 -1.0
9980 -1.0
9981 -1.0
9982 -1.0
9983 -1.0
9984 -1.0
9985 -1.0
9986 -1.0
9987 -1.0
9988 -1.0
9989 -1.0
9990 -1.0
9991 -1.0
9992 -1.0
9993 -1.0
9994 -1.0
9995 -1.0
9996 -1.0
9997 -1.0
9998 -1.0
9999 -1.0
dtype: float64
```

```
In [23]: df5._data.blocks
```

```
Out[23]: (FloatBlock: slice(3, 10000, 1), 9997 x 2, dtype: float64,
          IntBlock: slice(0, 3, 1), 3 x 2, dtype: int64)
```

```
In [ ]: def fill_df2(df):
        mydf = df.copy()
        for i in range(10000):
            mydf[i] = [i, -i]
            mydf.reset_index(drop=True, inplace=True)
        return mydf
```

```
In [ ]: df6 = pd.DataFrame(-1.0, index=range(2), columns=range(10000))
```

```
In [ ]: %timeit fill_df2(df6)
```

```
In [ ]: df7 = pd.DataFrame(-1, index=range(2), columns=range(10000))
```

```
In [ ]: %timeit fill_df2(df7)
```