DataFrames

DataFrames are a bunch of Series objects put together to share the same index.

```
import pandas as pd
import numpy as np
```

Create DataFrame

A pandas DataFrame can be created using various inputs like –

Lists dict Series Numpy ndarrays Another DataFrame

Create a DataFrame from Lists

The DataFrame can be created using a single list or a list of lists.

───────────────────── + Code ──── + Text ─────────────────────

```
a = [10,20,30,40,50]
B = pd.DataFrame(a)
print(B)
# B[0]
```

```
        0
    0  10
    1  20
    2  30
    3  40
    4  50
```

```
data = [['Alex',10],['Bob',12],['Clarke',13]]
df = pd.DataFrame(data, columns=['Name','Age'])
# print(df)
df['Age']
```

```
    0    10
    1    12
    2    13
    Name: Age, dtype: int64
```

```
data = [['Alex',10],['Bob',12],['Clarke',13]]
df = pd.DataFrame(data,columns=['Name','Age'],dtype=float)
print(df)
```

```
        Name    Age
```

```
0    Alex   10.0
1     Bob   12.0
2  Clarke   13.0
```

Create a DataFrame from Dict of Lists/ndarrays

All the ndarrays must be of same length. If index is passed, then the length of the index should
equal to the length of the arrays.

```python
a = {'Name':['Tom', 'Jack', 'Steve', 'Ricky'],'Age':[28,34,29,42]}
# print(type(a))
df = pd.DataFrame(a)
print(df)
```

```
    Name  Age
0    Tom   28
1   Jack   34
2  Steve   29
3  Ricky   42
```

```python
a = {'Name':np.array(['Tom', 'Jack', 'Steve', 'Ricky']),'Age':np.array([28,34,29,42])}
df = pd.DataFrame(a)
print(df)
```

```
    Name  Age
0    Tom   28
1   Jack   34
2  Steve   29
3  Ricky   42
```

```python
a = {'Name':['Tom', 'Jack', 'Steve', 'Ricky'],'Age':[28,34,29,42]}
df = pd.DataFrame(a, index=['rank1','rank2','rank3','rank4'])
print(df)
```

```
        Name  Age
rank1    Tom   28
rank2   Jack   34
rank3  Steve   29
rank4  Ricky   42
```

```python
df = pd.DataFrame(np.random.randn(5,4),index='A B C D E'.split(),columns='W X Y Z'.split())
# 'A B C D E'.split()
df
```

	W	X	Y	Z
A	-0.877680	1.239049	0.618773	0.622085
B	0.046040	-0.800241	0.953133	0.591053

```python
# df
# np.random.randn(5,4)
'A B C D E'.split()
```

    ['A', 'B', 'C', 'D', 'E']

Create a DataFrame from Dict of Series

```python
d = {'one' : pd.Series([1, 2, 3], index=['a', 'b', 'c']),
    'two' : pd.Series([1, 2, 3, 4], index=['a', 'b', 'c', 'd'])}
df = pd.DataFrame(d)
df
```

	one	two
a	1.0	1
b	2.0	2
c	3.0	3
d	NaN	4

Selection and Indexing

Let's learn the various methods to grab data from a DataFrame

[ ] ↳ 33 cells hidden

More Index Details

[ ] ↳ 9 cells hidden