```python
import pandas as pd

import numpy as np

# Series
labels = ['a', 'b', 'c']
my_list = [10,20,30]
arr = np.array(my_list)
d = {
    'a' : 10,
    'b' : 20,
    'c' : 30
}

pd.Series(data = my_list)
```

```
0    10
1    20
2    30
dtype: int64
```

```python
pd.Series(data = my_list, index=labels)
```

```
a    10
b    20
c    30
dtype: int64
```

```python
pd.Series(d)
```

```
a    10
b    20
c    30
dtype: int64
```

```python
ser1 = pd.Series(data = [1,2,3,4], index = ['USA', 'Germany', 'USSR', 'Japan'])
ser1
```

```
USA        1
Germany    2
USSR       3
Japan      4
dtype: int64
```

```python
ser2 = pd.Series(data = [1,2,5,4], index = ['USA', 'Germany', 'Italy', 'Japan'])
ser2
```

```
USA        1
Germany    2
Italy      5
```

```
Japan          4
dtype: int64

ser1 + ser2

Germany     4.0
Italy       NaN
Japan       8.0
USA         2.0
USSR        NaN
dtype: float64
```

# DataFrames

```python
import pandas as pd
import numpy as np
from numpy.random import randn

np.random.seed(101)

df = pd.DataFrame(randn(5,4), index = ['A', 'B', 'C', 'D', 'E'],
columns = ['W', 'X', 'Y', 'Z'])
df
```

```
          W         X         Y         Z
A  2.706850  0.628133  0.907969  0.503826
B  0.651118 -0.319318 -0.848077  0.605965
C -2.018168  0.740122  0.528813 -0.589001
D  0.188695 -0.758872 -0.933237  0.955057
E  0.190794  1.978757  2.605967  0.683509
```

```python
df['W']
```

```
A     2.706850
B     0.651118
C    -2.018168
D     0.188695
E     0.190794
Name: W, dtype: float64
```

```python
df[['W', 'Z']]
```

```
          W         Z
A  2.706850  0.503826
B  0.651118  0.605965
C -2.018168 -0.589001
D  0.188695  0.955057
E  0.190794  0.683509
```

```python
df.W
```

```
A    2.706850
B    0.651118
C   -2.018168
D    0.188695
E    0.190794
Name: W, dtype: float64
```

```python
df['NEW'] = df['W'] + df['Z'] # IMMUTABLE FEATURE
```

```python
df
```

```
          W          X          Y          Z        NEW
A  2.706850   0.628133   0.907969   0.503826   3.210676
B  0.651118  -0.319318  -0.848077   0.605965   1.257083
C -2.018168   0.740122   0.528813  -0.589001  -2.607169
D  0.188695  -0.758872  -0.933237   0.955057   1.143752
E  0.190794   1.978757   2.605967   0.683509   0.874303
```

```python
df = df.drop('NEW', axis=1) # MUTABLE FEATURE
```

```python
df
```

```
          W          X          Y          Z
A  2.706850   0.628133   0.907969   0.503826
B  0.651118  -0.319318  -0.848077   0.605965
C -2.018168   0.740122   0.528813  -0.589001
D  0.188695  -0.758872  -0.933237   0.955057
E  0.190794   1.978757   2.605967   0.683509
```

```python
df > 0
```

```
       W      X      Y      Z
A   True   True   True   True
B   True  False  False   True
C  False   True   True  False
D   True  False  False   True
E   True   True   True   True
```

```python
df[df > 0]
```

```
          W          X          Y          Z
A  2.706850   0.628133   0.907969   0.503826
B  0.651118        NaN        NaN   0.605965
C       NaN   0.740122   0.528813        NaN
D  0.188695        NaN        NaN   0.955057
E  0.190794   1.978757   2.605967   0.683509
```

```python
df
```

```
          W          X          Y          Z
A  2.706850   0.628133   0.907969   0.503826
B  0.651118  -0.319318  -0.848077   0.605965
```

```
C -2.018168  0.740122  0.528813 -0.589001
D  0.188695 -0.758872 -0.933237  0.955057
E  0.190794  1.978757  2.605967  0.683509
```

```python
df[ df['W']>0 ]
```

```
          W         X         Y         Z
A  2.706850  0.628133  0.907969  0.503826
B  0.651118 -0.319318 -0.848077  0.605965
D  0.188695 -0.758872 -0.933237  0.955057
E  0.190794  1.978757  2.605967  0.683509
```

```python
df[df['W']>1][['Y', 'Z']]
```

```
          Y         Z
A  0.907969  0.503826
```

```python
df
```

```
          W         X         Y         Z
A  2.706850  0.628133  0.907969  0.503826
B  0.651118 -0.319318 -0.848077  0.605965
C -2.018168  0.740122  0.528813 -0.589001
D  0.188695 -0.758872 -0.933237  0.955057
E  0.190794  1.978757  2.605967  0.683509
```

```python
states = ['CA', 'NY', 'WY', 'OR', 'CO']
df['States'] = states
```

```python
df
```

```
          W         X         Y         Z States
A  2.706850  0.628133  0.907969  0.503826     CA
B  0.651118 -0.319318 -0.848077  0.605965     NY
C -2.018168  0.740122  0.528813 -0.589001     WY
D  0.188695 -0.758872 -0.933237  0.955057     OR
E  0.190794  1.978757  2.605967  0.683509     CO
```

```python
df.set_index('States') # MUTABLE
```

```
              W         X         Y         Z
States
CA     2.706850  0.628133  0.907969  0.503826
NY     0.651118 -0.319318 -0.848077  0.605965
WY    -2.018168  0.740122  0.528813 -0.589001
OR     0.188695 -0.758872 -0.933237  0.955057
CO     0.190794  1.978757  2.605967  0.683509
```

```python
df.set_index('States', inplace=True)
```

```python
df
```

```
              W          X          Y          Z
States
CA      2.706850   0.628133   0.907969   0.503826
NY      0.651118  -0.319318  -0.848077   0.605965
WY     -2.018168   0.740122   0.528813  -0.589001
OR      0.188695  -0.758872  -0.933237   0.955057
CO      0.190794   1.978757   2.605967   0.683509
```

```python
# Group Index
outside = ['G1', 'G1', 'G1', 'G2', 'G2', 'G2']
inside = [1,2,3,1,2,3]
# zip
hier_index = list(zip(outside, inside))
hier_index = pd.MultiIndex.from_tuples(hier_index)
df = pd.DataFrame(np.random.randn(6,2), index=hier_index, columns =
['A', 'B'])
df
```

```
             A          B
G1 1   0.302665   1.693723
   2  -1.706086  -1.159119
   3  -0.134841   0.390528
G2 1   0.166905   0.184502
   2   0.807706   0.072960
   3   0.638787   0.329646
```

```python
df.loc['G1']
```

```
          A          B
1   0.302665   1.693723
2  -1.706086  -1.159119
3  -0.134841   0.390528
```

```python
df.loc['G1'].loc[1]
```

```
A    0.302665
B    1.693723
Name: 1, dtype: float64
```

```python
df
```

```
             A          B
G1 1   0.302665   1.693723
   2  -1.706086  -1.159119
   3  -0.134841   0.390528
G2 1   0.166905   0.184502
   2   0.807706   0.072960
   3   0.638787   0.329646
```