pandas

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

pd.__version__
```

Out[1]:
```
'0.23.4'
```

Series

In [2]:
```
ser = pd.Series(['张三', '李四', '王五'])
ser
```

Out[2]:
```
0    张三
1    李四
2    王五
dtype: object
```

In [3]:
```
ser = pd.Series(['张三', '李四', '王五'], list(range(1,4)))
ser
```

Out[3]:
```
1    张三
2    李四
3    王五
dtype: object
```

In [4]:
```
ser = pd.Series(['张三', '李四', '王五'], index=list(range(1,4)))
ser
```

Out[4]:
```
1    张三
2    李四
3    王五
dtype: object
```

In [5]:
```
ser[2]
```

Out[5]:
```
'李四'
```

In [6]:
```
ser[2] = 'Jack'
```

In [7]:
```
ser
```

Out[7]:
```
1      张三
2    Jack
3      王五
dtype: object
```

```
In [8]:  type(ser)

Out[8]:  pandas.core.series.Series
```

```
In [9]:  ser.values

Out[9]:  array(['张三', 'Jack', '王五'], dtype=object)
```

```
In [10]:  type(ser.values)

Out[10]:  numpy.ndarray
```

```
In [11]:  ser2 = pd.Series([19, 18, 20], index=range(1, 4))
          ser2

Out[11]:  1    19
          2    18
          3    20
          dtype: int64
```

```
In [12]:  ser2 + 1

Out[12]:  1    20
          2    19
          3    21
          dtype: int64
```

```
In [13]:  ser2[ser2%2==0]

Out[13]:  2    18
          3    20
          dtype: int64
```

```
In [14]:  data = {'beijing':12000, 'shanghai':11500, 'guangzhou':10800}
          data

Out[14]:  {'beijing': 12000, 'guangzhou': 10800, 'shanghai': 11500}
```

```
In [15]:  ser3 = pd.Series(data)
```

```
In [16]:  ser3

Out[16]:  beijing      12000
          shanghai     11500
          guangzhou    10800
          dtype: int64
```

```
In [17]:  ser3['beijing']

Out[17]:  12000
```

```
In [18]:  'shanghai' in ser3

Out[18]:  True
```

In [19]:

```
ser4 = pd.Series(data, index=['shanghai', 'beijing', 'chongqing'])
ser4
```

Out[19]:

```
shanghai     11500.0
beijing      12000.0
chongqing        NaN
dtype: float64
```

In [22]:

```
ser4.to_dict()
```

Out[22]:

```
{'beijing': 12000.0, 'chongqing': nan, 'shanghai': 11500.0}
```

In [23]:

```
ser4.tolist()
```

Out[23]:

```
[11500.0, 12000.0, nan]
```

In [24]:

```
ser4.to_json()
```

Out[24]:

```
'{"shanghai":11500.0,"beijing":12000.0,"chongqing":null}'
```

In [25]:

```
ser4.to_frame()
```

Out[25]:

	0
shanghai	11500.0
beijing	12000.0
chongqing	NaN

DataFrame

In [27]:

```
import numpy as np
data = np.arange(100, 109).reshape(3, -1)
data
```

Out[27]:

```
array([[100, 101, 102],
       [103, 104, 105],
       [106, 107, 108]])
```

In [28]:

```
df = pd.DataFrame(data)
```

In [29]:

```
df
```

Out[29]:

	0	1	2
0	100	101	102
1	103	104	105
2	106	107	108

In [31]:

```
data = {
    'name': ['jack','marry','lily'],
    'age': [19, 18, 20, 33],
    'height': [167, 160, 164]
}
df = pd.DataFrame(data)
df
```

```
------------------------------------------------------------------
---------
ValueError                                Traceback (most recent c
all last)
<ipython-input-31-b1c7f38ae0fb> in <module>()
      4     'height': [167, 160, 164]
      5 }
----> 6 df = pd.DataFrame(data)
      7 df

/Library/Frameworks/Python.framework/Versions/3.6/lib/python3.6/si
te-packages/pandas/core/frame.py in __init__(self, data, index, co
lumns, dtype, copy)
    346                                  dtype=dtype, copy=copy)
    347         elif isinstance(data, dict):
--> 348             mgr = self._init_dict(data, index, columns,
dtype=dtype)
    349         elif isinstance(data, ma.MaskedArray):
    350             import numpy.ma.mrecords as mrecords

/Library/Frameworks/Python.framework/Versions/3.6/lib/python3.6/si
te-packages/pandas/core/frame.py in _init_dict(self, data, index,
columns, dtype)
    457             arrays = [data[k] for k in keys]
    458
--> 459         return _arrays_to_mgr(arrays, data_names, index,
columns, dtype=dtype)
    460
    461     def _init_ndarray(self, values, index, columns, dtype=
None, copy=False):

/Library/Frameworks/Python.framework/Versions/3.6/lib/python3.6/si
te-packages/pandas/core/frame.py in _arrays_to_mgr(arrays, arr_nam
es, index, columns, dtype)
   7354     # figure out the index, if necessary
   7355     if index is None:
-> 7356         index = extract_index(arrays)
   7357
   7358     # don't force copy because getting jammed in an ndarra
y anyway

/Library/Frameworks/Python.framework/Versions/3.6/lib/python3.6/si
te-packages/pandas/core/frame.py in extract_index(data)
   7400             lengths = list(set(raw_lengths))
   7401             if len(lengths) > 1:
-> 7402                 raise ValueError('arrays must all be same
length')
   7403
   7404             if have_dicts:

ValueError: arrays must all be same length
```

```
In [32]: data = {
             'name': ['jack','marry','lily'],
             'age': [19, 18, 20],
             'height': [167, 160, 164]
         }
         df = pd.DataFrame(data)
         df
```

Out[32]:

	name	age	height
0	jack	19	167
1	marry	18	160
2	lily	20	164

```
In [34]: df.columns

Out[34]: Index(['name', 'age', 'height'], dtype='object')
```

```
In [35]: df.columns = ['username', 'age', 'height']
         df
```

Out[35]:

	username	age	height
0	jack	19	167
1	marry	18	160
2	lily	20	164

```
In [37]: df = pd.DataFrame(data, columns=['name', 'age', 'height', 'email'])
         df
```

Out[37]:

	name	age	height	email
0	jack	19	167	NaN
1	marry	18	160	NaN
2	lily	20	164	NaN

```
In [38]: df.index

Out[38]: RangeIndex(start=0, stop=3, step=1)
```

In [39]:

```
df = pd.DataFrame(data, columns=['name', 'age', 'height', 'email'],
index=range(1,4))
df
```

Out[39]:

	name	age	height	email
1	jack	19	167	NaN
2	marry	18	160	NaN
3	lily	20	164	NaN