2017年3月26日　　23:51

不得不说DataFrame现在很火，现在已经有很多库都是基于DataFrame写的,而且它用起来也很方便，读excel只需要一行代码，想当初xlrd可是让我头疼了好久，所以对于用Python处理大数据的人来说，pandas是必须要了解的。对于一个数据处理工具来说，读写是最基本的，下面是我最近整理的关于pandas一些基本本操作。

pandas入门

In [2]:
```
import numpy as np
import pandas as pd
```

create DataFrame

1 create by dictionary

In [33]:
```
df1=pd.DataFrame({"one":np.random.randn(4),"two":np.linspace(1,4,4),"three":["zhangsan","lisi",999.99,1]})
```

In [30]:
```
#each row or columns could have different data type
#keys of dictionary become columns of DataFrame
#df1 has no index,default from 0
df1
```

Out[30]:

	one	three	two
0	1.329260	zhangsan	1.0
1	-0.079364	lisi	2.0
2	-1.414885	999.99	3.0
3	0.858778	1	4.0

In [5]:
```
#set index
index=pd.date_range("20161001",periods=4)
df1.set_index(index,inplace=True)
df1
```

Out[5]:

	one	three	two
2016-10-01	0.771654	zhangsan	1.0
2016-10-02	0.608155	lisi	2.0
2016-10-03	1.166374	999.99	3.0
2016-10-04	1.083655	1	4.0

2 create by array

In [290]:
```
a=np.random.randn(4,6)
df2=pd.DataFrame(np.random.randn(4,6),index=index,columns=list("ABCDEF"))
```

In [291]:
```
df2
```

Out[291]:

	A	B	C	D	E	F
2016-10-01	-1.265130	0.560846	-0.024106	-1.185406	1.185334	-0.500951
2016-10-02	1.401373	0.164059	0.938154	-0.865296	0.359899	0.632356
2016-10-03	-0.089939	0.796111	-2.062924	0.259829	0.463562	0.550072
2016-10-04	-1.071990	-0.262464	-0.418685	-0.925290	-1.451162	-0.268825

3 create by ohter DataFrame

In [145]:
```
df3=df2[["A","B","C"]].copy()
```

In [146]:
```
df3
```

Out[146]:

	A	B	C
2016-10-01	1.206268	2.472697	-1.381334
2016-10-02	0.125169	0.027393	0.180191
2016-10-03	-0.133640	-0.512497	1.227997
2016-10-04	-0.008833	-0.393885	0.560344

In [162]:
```
#slice by row
df3=df2[0:3]
```

In [163]:
```
df3
```

Out[163]:

	A	B	C	D	E	F
2016-10-01	999.000000	2.472697	-1.381334	-0.439292	0.813846	-0.194123
2016-10-02	0.125169	0.027393	0.180191	-0.011879	-0.271406	0.311124
2016-10-03	-0.133640	-0.512497	1.227997	-0.558309	0.266703	-1.309833

In [172]:
```
#use row label slice
df3=df2["2016-10-01":"2016-10-04"]
```

Out[172]:

	A	B	C	D	E	F
2016-10-01	999.000000	2.472697	-1.381334	-0.439292	0.813846	-0.194123
2016-10-02	0.125169	0.027393	0.180191	-0.011879	-0.271406	0.311124
2016-10-03	-0.133640	-0.512497	1.227997	-0.558309	0.266703	-1.309833
2016-10-04	-0.008833	-0.393885	0.560344	-0.062900	1.679855	-1.476096

In [185]:
```
#slice by columns
df3=df2.iloc[:,0:5]
df3
```

Out[185]:

	A	B	C	D	E
2016-10-01	999.000000	2.472697	-1.381334	-0.439292	0.813846
2016-10-02	0.125169	0.027393	0.180191	-0.011879	-0.271406
2016-10-03	-0.133640	-0.512497	1.227997	-0.558309	0.266703
2016-10-04	-0.008833	-0.393885	0.560344	-0.062900	1.679855

Read DataFrame

1 read all data

In [186]:
```
df3
```

Out[186]:

	A	B	C	D	E
2016-10-01	999.000000	2.472697	-1.381334	-0.439292	0.813846
2016-10-02	0.125169	0.027393	0.180191	-0.011879	-0.271406
2016-10-03	-0.133640	-0.512497	1.227997	-0.558309	0.266703
2016-10-04	-0.008833	-0.393885	0.560344	-0.062900	1.679855

In [189]:
```
df3.iloc[:,:]
```

Out[189]:

	A	B	C	D	E
2016-10-01	999.000000	2.472697	-1.381334	-0.439292	0.813846
2016-10-02	0.125169	0.027393	0.180191	-0.011879	-0.271406
2016-10-03	-0.133640	-0.512497	1.227997	-0.558309	0.266703
2016-10-04	-0.008833	-0.393885	0.560344	-0.062900	1.679855

In [191]:
```
df3.loc[:,:]
```

Out[191]:

	A	B	C	D	E
2016-10-01	999.000000	2.472697	-1.381334	-0.439292	0.813846
2016-10-02	0.125169	0.027393	0.180191	-0.011879	-0.271406
2016-10-03	-0.133640	-0.512497	1.227997	-0.558309	0.266703
2016-10-04	-0.008833	-0.393885	0.560344	-0.062900	1.679855

In [192]:
```
df3.ix[:,:]
```

Out[191]:

	A	B	C	D	E
2016-10-01	999.000000	2.472697	-1.381334	-0.439292	0.813846
2016-10-02	0.125169	0.027393	0.180191	-0.011879	-0.271406
2016-10-03	-0.133640	-0.512497	1.227997	-0.558309	0.266703
2016-10-04	-0.008833	-0.393885	0.560344	-0.062900	1.679855

In [192]:
```
df3.ix[:,:]
```

Out[192]:

	A	B	C	D	E
2016-10-01	999.000000	2.472697	-1.381334	-0.439292	0.813846
2016-10-02	0.125169	0.027393	0.180191	-0.011879	-0.271406
2016-10-03	-0.133640	-0.512497	1.227997	-0.558309	0.266703
2016-10-04	-0.008833	-0.393885	0.560344	-0.062900	1.679855

2 read by row

In [216]:
```
df3.loc["2016-10-02"]
```

Out[216]:
```
A    0.125169
B    0.027393
C    0.180191
D   -0.011879
E   -0.271406
Name: 2016-10-02 00:00:00, dtype: float64
```

In [226]:
```
df3.loc["2016-10-02":"2016-10-05"]
```

Out[226]:

	A	B	C	D	E
2016-10-02	0.125169	0.027393	0.180191	-0.011879	-0.271406
2016-10-03	-0.133640	-0.512497	1.227997	-0.558309	0.266703
2016-10-04	-0.008833	-0.393885	0.560344	-0.062900	1.679855

In [238]:
```
df3.iloc[1:4]
```

Out[238]:

	A	B	C	D	E
2016-10-02	0.125169	0.027393	0.180191	-0.011879	-0.271406
2016-10-03	-0.133640	-0.512497	1.227997	-0.558309	0.266703
2016-10-04	-0.008833	-0.393885	0.560344	-0.062900	1.679855

In [227]:
```
df3[:3]
```

Out[227]:

	A	B	C	D	E
2016-10-01	999.000000	2.472697	-1.381334	-0.439292	0.813846
2016-10-02	0.125169	0.027393	0.180191	-0.011879	-0.271406
2016-10-03	-0.133640	-0.512497	1.227997	-0.558309	0.266703

In [228]:
```
df3.head(3)
```

Out[238]:

	A	B	C	D	E
2016-10-02	0.125169	0.027393	0.180191	-0.011879	-0.271406
2016-10-03	-0.133640	-0.512497	1.227997	-0.558309	0.266703
2016-10-04	-0.008833	-0.393885	0.560344	-0.062900	1.679855

In [227]:
```
df3[:3]
```

Out[227]:

	A	B	C	D	E
2016-10-01	999.000000	2.472697	-1.381334	-0.439292	0.813846
2016-10-02	0.125169	0.027393	0.180191	-0.011879	-0.271406
2016-10-03	-0.133640	-0.512497	1.227997	-0.558309	0.266703

In [228]:
```
df3.head(3)
```

Out[228]:

	A	B	C	D	E
2016-10-01	999.000000	2.472697	-1.381334	-0.439292	0.813846
2016-10-02	0.125169	0.027393	0.180191	-0.011879	-0.271406
2016-10-03	-0.133640	-0.512497	1.227997	-0.558309	0.266703

3 read by col

In [240]:
```
df3["A"]
```

Out[240]:
```
2016-10-01    999.000000
2016-10-02      0.125169
2016-10-03     -0.133640
2016-10-04     -0.008833
Freq: D, Name: A, dtype: float64
```

In [242]:
```
df3[["A","C","D"]]
```

Out[242]:

	A	C	D
2016-10-01	999.000000	-1.381334	-0.439292
2016-10-02	0.125169	0.180191	-0.011879
2016-10-03	-0.133640	1.227997	-0.558309
2016-10-04	-0.008833	0.560344	-0.062900

In [275]:
```
df3.iloc[:,0]
```

Out[275]:
```
2016-10-01    1111.000000
2016-10-02       0.125169
2016-10-03      -0.133640
2016-10-04      -0.008833
Freq: D, Name: A, dtype: float64
```

In [246]:
```
df3.loc[:,"A"]
```

Out[246]:
```
2016-10-01    999.000000
2016-10-02      0.125169
2016-10-03     -0.133640
2016-10-04     -0.008833
Freq: D, Name: A, dtype: float64
```

4 read by pixel

In [248]:
```
df3["A"][0]
```

Out[248]:
```
999.0
```

In [253]:
```
df3["A"]["2016-10-01"]
```

Out[253]:
```
999.0
```

In [255]:
```
df3.iat[0,0]
```

Out[255]:
```
999.0
```

set value for DataFrame

1 set a pixel

In [267]:
```
df2["A"][0]
```

Out[267]:
```
99.989999999999995
```

In [271]:
```
df2["A"][0]=1111
```

In [272]:
```
df2["A"][0]
```

Out[272]:
```
1111.0
```

In [278]:
```
df2.iloc[0,0]=2222
```

In [279]:
```
df2["A"][0]
```

Out[279]:
```
2222.0
```

In [282]:
```
df2.loc["A","2016-10-01"]=3333
```

In [283]:
```
df2["A"][0]
```

Out[283]:
```
2222.0
```

2 set values by row

In [293]:
```
df2["A"]
```

Out[293]:
```
2016-10-01   -1.265130
2016-10-02    1.401373
2016-10-03   -0.089939
2016-10-04   -1.071990
Freq: D, Name: A, dtype: float64
```

In [294]:
```
A=["good","great","fine","nice"]
df2["A"]=A
```

In [296]:
```
df2
```

Out[296]:

	A	B	C	D	E	F
2016-10-01	good	0.560846	-0.024106	-1.185406	1.185334	-0.500951
2016-10-02	great	0.164059	0.938154	-0.865296	0.359899	0.632356
2016-10-03	fine	0.796111	-2.062924	0.259829	0.463562	0.550072
2016-10-04	nice	-0.262464	-0.418685	-0.925290	-1.451162	-0.268825

In [307]:
```
AC=[["a","b"],["aa","bb"],["aaa","bbb"],["aaaa","bbbb"]]
df2[["A","C"]]=AC
```

In [308]:
```
df2
```

Out[308]:

	A	B	C	D	E	F
2016-10-01	a	0.560846	b	-1.185406	1.185334	-0.500951
2016-10-02	aa	0.164059	bb	-0.865296	0.359899	0.632356
2016-10-03	aaa	0.796111	bbb	0.259829	0.463562	0.550072
2016-10-04	aaaa	-0.262464	bbbb	-0.925290	-1.451162	-0.268825

In [309]:
```
df2[["D","E"]]="NAN"
```

In [310]:
```
df2
```

Out[310]:

	A	B	C	D	E	F
2016-10-01	a	0.560846	b	NAN	NAN	-0.500951
2016-10-02	aa	0.164059	bb	NAN	NAN	0.632356
2016-10-03	aaa	0.796111	bbb	NAN	NAN	0.550072
2016-10-04	aaaa	-0.262464	bbbb	NAN	NAN	-0.268825

In [315]:
```
df2.iloc[:,2:4]="========"
```

In [316]:
```
df2
```

Out[316]:

	A	B	C	D	E	F
2016-10-01	a	0.560846	========	========	NAN	-0.500951
2016-10-02	aa	0.164059	========	========	NAN	0.632356
2016-10-03	aaa	0.796111	========	========	NAN	0.550072
2016-10-04	aaaa	-0.262464	========	========	NAN	-0.268825

In [320]:
```
df2.loc[:,"A":"C"]="****"
```

In [321]:
```
df2
```

Out[321]:

	A	B	C	D	E	F
2016-10-01	****	****	****	========	NAN	-0.500951
2016-10-02	****	****	****	========	NAN	0.632356
2016-10-03	****	****	****	========	NAN	0.550072
2016-10-04	****	****	****	========	NAN	-0.268825

In [324]:
```
df2.ix[:,3:5]=",,,"
```

In [325]:
```
df2
```

Out[325]:

	A	B	C	D	E	F
2016-10-01	****	****	****	,,,	,,,	-0.500951
2016-10-02	****	****	****	,,,	,,,	0.632356
2016-10-03	****	****	****	,,,	,,,	0.550072
2016-10-04	****	****	****	,,,	,,,	-0.268825

In [331]:
```
df2.ix[:,"E":"F"]="999"
```

In [332]:
```
df2
```

Out[332]:

	A	B	C	D	E	F
2016-10-01	****	****	****	...	999	999
2016-10-02	****	****	****	...	999	999
2016-10-03	****	****	****	...	999	999
2016-10-04	****	****	****	...	999	999

set values by col

- set values by col and by row are the same , not detail description

Insert data

In [334]:
```
df1
```

Out[334]:

	one	three	two
2016-10-01	0.135058	zhangsan	1.0
2016-10-02	0.263112	lisi	2.0
2016-10-03	-1.064534	999.99	3.0
2016-10-04	-0.245670	1	4.0

In [342]:
```
#inster a columns
df1.insert(1,"four",[111,222,333,444])
df1
```

Out[342]:

	one	four	three	two
2016-10-01	0.135058	111	zhangsan	1.0
2016-10-02	0.263112	222	lisi	2.0
2016-10-03	-1.064534	333	999.99	3.0
2016-10-04	-0.245670	444	1	4.0

In [34]:
```
df1
```

Out[34]:

	one	three	two
0	-1.006160	zhangsan	1.0
1	1.808572	lisi	2.0
2	1.237810	999.99	3.0
3	-2.362585	1	4.0

In [46]:
```
#insert a row
row={"one":1,"three":11,"two":111}
df1.iloc[1]=row
df1.loc[3]=row
```

In [47]:
```
df1
```

Out[47]:

	one	three	two
0	-1.00616	zhangsan	1
1	1	11	111
2	1.23781	999.99	3
3	1	11	111

In [49]:
```
#insert data in the tail
#insert a row
row1={"one":"good","two":"nice","three":"great"}
df1.append(row1,ignore_index=True)
```

Out[49]:

	one	three	two
0	-1.00616	zhangsan	1
1	1	11	111
2	1.23781	999.99	3
3	1	11	111
4	good	great	nice

In [97]:
```
df1
```

Out[97]:

	one	three	two
0	-1.00616	zhangsan	1
1	1	11	111
2	1.23781	999.99	3
3	1	11	111

In [92]:
```
row2={1:"AAA",2:"BBB",3:"CCC",0:"DDD"}
df4=df1.T.append(row2,ignore_index=True).T
```

In [94]:
```
df4.columns=list("one","two","three","four")
```

In [98]:
```
df4
```

Out[98]:

	A	B	C	D
0	-1.00616	zhangsan	1	DDD
1	1	11	111	AAA
2	1.23781	999.99	3	BBB
3	1	11	111	CCC

From <http://blog.csdn.net/xtfge0915/article/details/52938740>