

# Python数据分析实战

## 第十四课 pandas数据查询

### 本节课程目标

- 从DataFrame中查询Series 1.查询一列 2.查询多列 3.查询一行 4.查询多行
- pandas常用数据查询

- 1.使用单个label值查询数据
- 2.使用值列表批量查询
- 3.使用数值区间进行范围查询
- 4.使用条件表达式查询
- 5.调用函数查询

### 从DataFrame中查询Series

```
#准备一个字典, value是数据列表
import pandas as pd
data = {'state':['zhangsan','lisi','wangwu','zhaoliu'],
        'year':[2000,2001,2002,2003],
        'pop':[1.3,1.5,1.7,1.9]}
#将数据传到DataFrame, 得到一个DataFrame对象
df = pd.DataFrame(data)
```

df

```
.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}
```

	state	year	pop
0	zhangsan	2000	1.3
1	lisi	2001	1.5
2	wangwu	2002	1.7
3	zhaoliu	2003	1.9

```
#查看下当前DataFrame的一些属性  
df.dtypes
```

```
state      object  
year       int64  
pop        float64  
dtype: object
```

```
df.columns
```

```
Index(['state', 'year', 'pop'], dtype='object')
```

```
df.index
```

```
RangeIndex(start=0, stop=4, step=1)
```

查询一列，结果是pd.series

```
df['year']
```

```
0    2000
1    2001
2    2002
3    2003
Name: year, dtype: int64
```

```
type(df['year'])
```

```
pandas.core.series.Series
```

## 查询多列，结果是一个pd.DataFrame

```
df[['year', 'pop']]
```

```
.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}
```

	year	pop
0	2000	1.3
1	2001	1.5
2	2002	1.7
3	2003	1.9

```
type(df[['year', 'pop']])
```

```
pandas.core.frame.DataFrame
```

## 查询一行，结果是Series

```
df.loc[1]#查询index为1的行
```

```
state    lisi
year     2001
pop       1.5
Name: 1, dtype: object
```

```
type(df.loc[1])
```

```
pandas.core.series.Series
```

## 查询多行，结果是一个DataFrame

```
df.loc[1:2]#采用切片的方式返回多行，包左也包右
```

```
.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}
```

	state	year	pop
1	lisi	2001	1.5
2	wangwu	2002	1.7

```
type(df.loc[1:2])
```

```
pandas.core.frame.DataFrame
```

## pandas数据查询

- 1.数值
- 2.列表
- 3.区间
- 4.条件
- 5.函数

### pandas查询数据的几种方法

- 1.df.loc方法，根据行、列的标签值查询
- 2.df.iloc方法，根据行、列的数字位置查询（这个用的不是很多）
- 3.df.where方法
- 4.df.query方法

**#读取数据**

```
df = pd.read_csv('./datas/beijing_tianqi_2018.csv')  
df.head()
```

```
.dataframe tbody tr th {  
    vertical-align: top;  
}  
  
.dataframe thead th {  
    text-align: right;  
}
```

	ymd	bWendu	yWendu	tianqi	fengxiang	fengli	aqi	aqiInfo	aqiLevel
0	2018-01-01	3°C	-6°C	晴~多云	东北风	1-2级	59	良	2
1	2018-01-02	2°C	-5°C	阴~多云	东北风	1-2级	49	优	1
2	2018-01-03	2°C	-5°C	多云	北风	1-2级	28	优	1
3	2018-01-04	0°C	-8°C	阴	东北风	1-2级	28	优	1
4	2018-01-05	3°C	-6°C	多云~晴	西北风	1-2级	50	优	1

```
df.index
```

```
RangeIndex(start=0, stop=365, step=1)
```

```
df.set_index('ymd',inplace=True)
```

```
df.index
```

```
Index(['2018-01-01', '2018-01-02', '2018-01-03', '2018-01-04', '2018-01-05',
      '2018-01-06', '2018-01-07', '2018-01-08', '2018-01-09', '2018-01-10',
      ...,
      '2018-12-22', '2018-12-23', '2018-12-24', '2018-12-25', '2018-12-26',
      '2018-12-27', '2018-12-28', '2018-12-29', '2018-12-30', '2018-12-31'],
      dtype='object', name='ymd', length=365)
```

```
df.head()
```

```
.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}
```

	bWendu	yWendu	tianqi	fengxiang	fengli	aqi	aqiInfo	aqiLevel
ymd								
2018-01-01	3°C	-6°C	晴~多云	东北风	1-2级	59	良	2
2018-01-02	2°C	-5°C	阴~多云	东北风	1-2级	49	优	1
2018-01-03	2°C	-5°C	多云	北风	1-2级	28	优	1
2018-01-04	0°C	-8°C	阴	东北风	1-2级	28	优	1
2018-01-05	3°C	-6°C	多云~晴	西北风	1-2级	50	优	1

```
df.loc[:, 'bWendu'] = df['bWendu'].str.replace('°C', '').astype('int32')
```

```
#检查下
df.head()
```

```
.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}
```

	bWendu	yWendu	tianqi	fengxiang	fengli	aqi	aqiInfo	aqiLevel
ymd								
2018-01-01	3	-6°C	晴~多云	东北风	1-2级	59	良	2
2018-01-02	2	-5°C	阴~多云	东北风	1-2级	49	优	1
2018-01-03	2	-5°C	多云	北风	1-2级	28	优	1
2018-01-04	0	-8°C	阴	东北风	1-2级	28	优	1
2018-01-05	3	-6°C	多云~晴	西北风	1-2级	50	优	1

```
#同理
df.loc[:, 'yWendu'] = df['yWendu'].str.replace('°C', '').astype('int32')

df.head()
```

```
.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}
```



	bWendu	yWendu	tianqi	fengxiang	fengli	aqi	aqiInfo	aqiLevel
ymd								
2018-01-01	3	-6	晴~多云	东北风	1-2级	59	良	2
2018-01-02	2	-5	阴~多云	东北风	1-2级	49	优	1
2018-01-03	2	-5	多云	北风	1-2级	28	优	1
2018-01-04	0	-8	阴	东北风	1-2级	28	优	1
2018-01-05	3	-6	多云~晴	西北风	1-2级	50	优	1

#检测下数据类型，看是否也已经替换  
df.dtypes

df.head()

## 1.使用单个label值查询数据

特点：行或者列，都可以传入单个值，实现精确匹配

```
df.loc['2018-01-03','bWendu']
```

2

```
df.loc['2018-01-03',['bWendu','yWendu']]
```

```
bWendu      2
yWendu     -5
Name: 2018-01-03, dtype: object
```

## 2.使用值列表批量查询

```
df.loc[['2018-01-03','2018-01-06','2018-02-03'],['bWendu','yWendu']]
```

```
.dataframe tbody tr th {  
    vertical-align: top;  
}  
  
.dataframe thead th {  
    text-align: right;  
}
```

	bWendu	yWendu
ymd		
2018-01-03	2	-5
2018-01-06	2	-5
2018-02-03	0	-9

## 3.使用数值区间进行范围查询

注意：区间的范围包左又包右

```
#行index  
df.loc['2018-01-03':'2018-01-06','bWendu']
```

```
ymd  
2018-01-03    2  
2018-01-04    0  
2018-01-05    3  
2018-01-06    2  
Name: bWendu, dtype: int32
```

#列

```
df.loc['2018-01-03','bWendu':'aqi']
```

```
bWendu      2
yWendu     -5
tianqi      多云
fengxiang   北风
fengli      1-2级
aqi         28
Name: 2018-01-03, dtype: object
```

```
df.loc['2018-01-03':'2018-01-09','bWendu':'aqi']
```

```
.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}
```

	bWendu	yWendu	tianqi	fengxiang	fengli	aqi
ymd						
2018-01-03	2	-5	多云	北风	1-2级	28
2018-01-04	0	-8	阴	东北风	1-2级	28
2018-01-05	3	-6	多云~晴	西北风	1-2级	50
2018-01-06	2	-5	多云~阴	西南风	1-2级	32
2018-01-07	2	-4	阴~多云	西南风	1-2级	59
2018-01-08	2	-6	晴	西北风	4-5级	50
2018-01-09	1	-8	晴	西北风	3-4级	34

4.使用条件表达式查询

bool列表的长度等于行数或者列数

```
df.loc[df['yWendu'] < -10,:]
```

```
.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}
```

	bWendu	yWendu	tianqi	fengxiang	fengli	aqi	aqiInfo	aqiLevel
ymd								
2018-01-23	-4	-12	晴	西北风	3-4级	31	优	1
2018-01-24	-4	-11	晴	西南风	1-2级	34	优	1
2018-01-25	-3	-11	多云	东北风	1-2级	27	优	1
2018-12-26	-2	-11	晴~多云	东北风	2级	26	优	1
2018-12-27	-5	-12	多云~晴	西北风	3级	48	优	1
2018-12-28	-3	-11	晴	西北风	3级	40	优	1
2018-12-29	-3	-12	晴	西北风	2级	29	优	1
2018-12-30	-2	-11	晴~多云	东北风	1级	31	优	1

```
df['yWendu'] < -10
```

ymd

2018-01-01	False
2018-01-02	False
2018-01-03	False
2018-01-04	False
2018-01-05	False
2018-01-06	False
2018-01-07	False
2018-01-08	False
2018-01-09	False
2018-01-10	False
2018-01-11	False
2018-01-12	False
2018-01-13	False
2018-01-14	False
2018-01-15	False
2018-01-16	False
2018-01-17	False
2018-01-18	False
2018-01-19	False
2018-01-20	False
2018-01-21	False
2018-01-22	False
2018-01-23	True
2018-01-24	True
2018-01-25	True
2018-01-26	False
2018-01-27	False
2018-01-28	False
2018-01-29	False
2018-01-30	False
	...
2018-12-02	False
2018-12-03	False
2018-12-04	False
2018-12-05	False
2018-12-06	False
2018-12-07	False
2018-12-08	False
2018-12-09	False
2018-12-10	False
2018-12-11	False
2018-12-12	False
2018-12-13	False
2018-12-14	False
2018-12-15	False
2018-12-16	False
2018-12-17	False
2018-12-18	False

```

2018-12-19    False
2018-12-20    False
2018-12-21    False
2018-12-22    False
2018-12-23    False
2018-12-24    False
2018-12-25    False
2018-12-26     True
2018-12-27     True
2018-12-28     True
2018-12-29     True
2018-12-30     True
2018-12-31    False
Name: yWendu, Length: 365, dtype: bool

```

```

#15~28, 晴天 优
# &
df.loc[(df['bWendu'] <= 28)&(df['yWendu'] >= 15) & (df['tianqi'] == '晴') &
(df['aqiLevel'] == 1),:]

```

```

.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}

```

	bWendu	yWendu	tianqi	fengxiang	fengli	aqi	aqiInfo	aqiLevel
ymd								
2018-09-07	27	16	晴	西北风	3-4级	22	优	1

## 5.调用函数查询

```
# def lambda
```

```
df.loc[lambda df:(df['bWendu'] <= 30)&(df['yWendu'] >= 15),:]
```

```

.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}

```

	bWendu	yWendu	tianqi	fengxiang	fengli	aqi	aqiInfo	aqiLevel
ymd								
2018-04-28	27	17	晴	西南风	3-4级	125	轻度污染	3
2018-04-29	30	16	多云	南风	3-4级	193	中度污染	4
2018-05-04	27	16	晴~多云	西南风	1-2级	86	良	2
2018-05-09	29	17	晴~多云	西南风	3-4级	79	良	2
2018-05-10	26	18	多云	南风	3-4级	118	轻度污染	3
2018-05-11	24	15	阴~多云	东风	1-2级	106	轻度污染	3
2018-05-12	28	16	小雨	东南风	3-4级	186	中度污染	4
2018-05-13	30	17	晴	南风	1-2级	68	良	2
2018-05-16	29	21	多云~小雨	东风	1-2级	142	轻度污染	3
2018-05-17	25	19	小雨~多云	北风	1-2级	70	良	2
2018-05-18	28	16	多云~晴	南风	1-2级	49	优	1
2018-05-19	27	16	多云~小雨	南风	1-2级	69	良	2
2018-	21	16	阴~小	东风	1-2级	54	良	2

05-20			雨					
2018-05-23	29	15	晴	西南风	3-4级	153	中度污染	4
2018-05-26	30	17	小雨~多云	西南风	3-4级	143	轻度污染	3
2018-05-28	30	16	晴	西北风	4-5级	178	中度污染	4
2018-06-09	23	17	小雨	北风	1-2级	45	优	1
2018-06-10	27	17	多云	东南风	1-2级	51	良	2
2018-06-11	29	19	多云	西南风	3-4级	85	良	2
2018-06-13	28	19	雷阵雨~多云	东北风	1-2级	73	良	2
2018-06-18	30	21	雷阵雨	西南风	1-2级	112	轻度污染	3
2018-06-22	30	21	雷阵雨~多云	东南风	1-2级	83	良	2
2018-07-08	30	23	雷阵雨	南风	1-2级	73	良	2
2018-07-09	30	22	雷阵雨~多云	东南风	1-2级	106	轻度污染	3
2018-07-10	30	22	多云~雷阵雨	南风	1-2级	48	优	1
2018-07-11	25	22	雷阵雨~大雨	东北风	1-2级	44	优	1
2018-07-12	27	22	多云	南风	1-2级	46	优	1
2018-07-13	28	23	雷阵雨	东风	1-2级	60	良	2
2018-07-17	27	23	中雨~雷阵雨	西风	1-2级	28	优	1
2018-07-24	28	26	暴雨~雷阵雨	东北风	3-4级	29	优	1



...	...	...	...	...	...	...	...	...
<b>2018-08-11</b>	30	23	雷阵雨 ~中雨	东风	1-2级	60	良	2
<b>2018-08-12</b>	30	24	雷阵雨	东南风	1-2级	74	良	2
<b>2018-08-14</b>	29	24	中雨~ 小雨	东北风	1-2级	42	优	1
<b>2018-08-16</b>	30	21	晴~多 云	东北风	1-2级	40	优	1
<b>2018-08-17</b>	30	22	多云~ 雷阵雨	东南风	1-2级	69	良	2
<b>2018-08-18</b>	28	23	小雨~ 中雨	北风	3-4级	40	优	1
<b>2018-08-19</b>	26	23	中雨~ 小雨	东北风	1-2级	37	优	1
<b>2018-08-22</b>	28	21	雷阵雨 ~多云	西南风	1-2级	48	优	1
<b>2018-08-24</b>	30	20	晴	北风	1-2级	40	优	1
<b>2018-08-27</b>	30	22	多云~ 雷阵雨	东南风	1-2级	89	良	2
<b>2018-08-28</b>	29	22	小雨~ 多云	南风	1-2级	58	良	2
<b>2018-08-30</b>	29	20	多云	南风	1-2级	47	优	1
<b>2018-08-31</b>	29	20	多云~ 阴	东南风	1-2级	48	优	1
<b>2018-09-01</b>	27	19	阴~小 雨	南风	1-2级	50	优	1
<b>2018-09-02</b>	27	19	小雨~ 多云	南风	1-2级	55	良	2
<b>2018-09-03</b>	30	19	晴	北风	3-4级	70	良	2
<b>2018-09-06</b>	27	18	多云~ 晴	西北风	4-5级	37	优	1

2018-09-07	27	16	晴	西北风	3-4级	22	优	1
2018-09-08	27	15	多云~晴	北风	1-2级	28	优	1
2018-09-09	28	16	晴	西南风	1-2级	51	良	2
2018-09-10	28	19	多云	南风	1-2级	65	良	2
2018-09-11	26	19	多云	南风	1-2级	68	良	2
2018-09-12	29	19	多云	南风	1-2级	59	良	2
2018-09-13	29	20	多云~阴	南风	1-2级	107	轻度污染	3
2018-09-14	28	19	小雨~多云	南风	1-2级	128	轻度污染	3
2018-09-15	26	15	多云	北风	3-4级	42	优	1
2018-09-17	27	17	多云~阴	北风	1-2级	37	优	1
2018-09-18	25	17	阴~多云	西南风	1-2级	50	优	1
2018-09-19	26	17	多云	南风	1-2级	52	良	2
2018-09-20	27	16	多云	西南风	1-2级	63	良	2

64 rows × 8 columns