**Daily Report**

4th March, 2018

1. **Series**

它包含一个数组数据和一个与数组相互关联的标签，即索引

可以通过values和index来获取series中的数组和索引对象

当创建带有索引时，如：

In [8]: obj2 = Series([4, 7, -5, 3], index=['d', 'b', 'a', 'c'])

In [9]: obj2

Out[9]:

d 4

b 7

a -5

c 3

In [10]: obj2.index

Out[10]: Index([d, b, a, c], dtype=object)

否则会创建默认索引，即0-（N-1）

创建也可以通过将python字典转化为Series的方法

Series的一个重要功能为在算术运算中自动对齐不同索引的数据

1. **Insnull和notnull函数**

能使用该函数检测数据丢失

1. **Dataframe**

构建一个Dataframe，最常用的方法是用一个相等长度列表的字典或Numpy数组，当不对列的顺序进行规定时，它的索引会自动分配，并且对列进行了排序；

当对列的顺序设定时，则产生的Dataframe将按照设定的顺序排列，如：

In [39]: DataFrame(data, columns=['year', 'state', 'pop'])

Out[39]:

year state pop

0 2000 Ohio 1.5

1 2001 Ohio 1.7

当设定的排序中若出现在数据中不存在的值时，它的结果将表示为NA

Dataframe中的任一列都可以通过字典记法或属性来检索

Dataframe中的任一行也可以通过位置或名字来检索，如：**frame2.ix**['three']

可以通过纯量或单个数组给无数值的列进行赋值，当采用纯量时，所有的数值都将等于纯量；当采用数组时，所赋的值的长度必须和Dataframe的长度相匹配；当采用Series赋值时，能够通过索引，精确地给索引处的值赋值

给一个不存在的列赋值，将创建一个新的列；使用del功能，能够删除列

1. **嵌套的字典格式**

当被传到Dataframe时，外部键将被解释为列索引，内部键将被解释为行索引，如：

In [57]: pop = {'Nevada': {2001: 2.4, 2002: 2.9},

....: 'Ohio': {2000: 1.5, 2001: 1.7, 2002: 3.6}}

In [58]: frame3 = DataFrame(pop)

In [59]: frame3

Out[59]:

Nevada Ohio

2000 NaN 1.5

2001 2.4 1.7

2002 2.9 3.6

1. 通过对index和colums的name命名，可以使其也在Dataframe中显示出来，如：

In [64]: frame3.index.name = 'year'; frame3.columns.name = 'state'

In [65]: frame3

Out[65]:

state Nevada Ohio

year

2000 NaN 1.5

2001 2.4 1.7

2002 2.9 3.6

1. 索引对象是不可变的
2. **索引的方法和属性**

|  |
| --- |
| 捕获 |

1. 可以使用（对象.reindex（[索引]））的方法，使数据符合新的索引，若新的索引值不存在，就引入缺失数据值；也可以使用fill\_value=‘’的方法，使缺失的值变为理想的值
2. 为了对时间序列这样的数据排序，当重建索引的时候可能想要对值进行内插或填充。 method 选项可以是你做到这一点，使用一个如 ffill 的方法来向前填充值，如：

In [84]: obj3 = Series(['blue', 'purple', 'yellow'], index=[0, 2, 4])

In [85]: obj3.reindex(range(6), method='ffill')

Out[85]:

0 blue

1 blue

2 purple

3 purple

4 yellow

5 yellow

reindex 的 method（内插）选项：

ffill或pad 前向（或进位）填充

bfill或backfill 后向（或进位）填充

1. **Drop**

该方法能够返回一个新的对象并从坐标轴中删除指定的一个或多个值

1. Pandas最重要的特性之一是在具有不同索引的对象间进行算术计算的能力。当把对象相加时，如果有任何的索引相同的话，在结果中将会把各自的索引联合起来。在索引不同时，引入NA值