**Pandas**

1 读写csv文件

写入数据：to\_csv()函数

语法：pd.series / dataframe.to\_csv(path,sep,index,header,encoding)

path：指定写入文件的路径

sep：指定数据的分隔符，默认为‘，’

index：指定是否有行索引，true / false

header：指定作为行数据的列名，默认为0，1，2……

encoding：指定数据的编码格式，如：GB2312，UTF-8，UTF-8-SIG

读取数据：read\_csv()

语法：pd.read\_csv(path,index\_col,header,names,nrows)

path：指定读取文件的路径

index\_col：指定是否有行索引，true / false

header：指定某行数据(数据内部)作为列名

names：用于指定结果的列名列表，传入列表

nrows：指定读取文件的前几行

2 读写excel文件

暂略

3 series数据类型：一维的数据列表

语法：pd.serise(data,index,dype)

series数据由序列数据和与之相关的索引两部分组成

data：用于传入的数据，list，ndarray

index：索引，必须与data数据的长度相同，默认0～N

dtype：指定数据的类型

用index属性获取索引值，用values属性获取数据

例如：pd.index，pd.values

4 dataframe数据类型：二维数据表，由键和值组成，每一个键对应一列，字典

语法：pd.dataframe(data,index,columns,dtype)

data：传入的数据，list，dictionary，ndarray

index：用于指定行索引，默认数字，从0开始，传入列表

columes：用于指定列索引，默认数字，从0开始，传入列表

dtype：指定数据的类型

常见属性、方法：

1 pd.values：用于获取ndarray类型的对象

2 pd.index：用于获取行索引

3 pd.columns：用于获取列索引

4 pd.head(n)：用于显示前n行数据，默认显示前5行数据

5 pd.tail(n)：用于显示后n行数据，默认显示后5行数据

6 dataframe.sample(n/frac，random\_state)

n：随机抽取n行数据

frac：按比例抽取frac\*m行数据

random\_state：设置随机数种子

7 loc()：用于按索引的名称选取数据，当执行切片操作时是前后全闭

语法：dataframe.loc(行索引名称或条件，列索引名称)

例如：

data.loc[:,[‘姓名’]](取单行，记得加[]，不然没索引值)，data.loc[:,’姓名’:’年龄’]

out： out：

8 iloc()：用于按索引的位置选取数据，当执行切片操作时是前闭后开

语法：dataframe.iloc(行索引位置，列索引位置)

例如：

data.iloc[:,0](注意索引从0开始，而不是从1开始) data.iloc[;,0:1]

out： out：