Pandas总结

|  |  |  |  |
| --- | --- | --- | --- |
|  | 方法 | 作用 | 页码 |
| Series | 1. 传入list 2. 传入字典 | 创建 | 116 |
|  | Obj.values | 获得其值，类型为array |  |
|  | Obj.index | 获得其索引 |  |
|  | Obj[index] | 获得索引为index的单值 | 117 |
|  | isnull(),notnull() | 检测是否存在缺失值 |  |
|  | Obj.reindex(list) | 重新索引，根据list重新排列，还可以传入method方法 | 127 |
|  | Obj.sort\_index() | 对索引进行排序 |  |
|  | Obj.order() | 对值进行排序 |  |
| DataFrame | 1. 传入字典 2. 传入list | 创建  DataFrame(data,index,columns) |  |
|  | Frame[columns]  Frame.columns | 获得columns列的值，类型为Series | 124,130 |
|  | Frame.ix[index] | 获得index行的值，类型为Series |  |
|  | Frame.values | 获得data的值，类型为array |  |
|  | Frame.index  frame.columns | 获得行、列索引 |  |
|  | isnull(),notnull() | 检测是否存在缺失值 |  |
|  | Frame.reindex(list) | 重新索引, 根据list重新排列，默认是对列重新索引，可以传入columns=来对行进行重新索引，同时也可以传入method | 128 |
|  | Frame.ix[] | 对行和列进行重新索引，注意这里是方括号 | 128 |
|  | Frame.drop() | 返回一个在指定轴上删除了指定值的新对象，默认是对轴0 | 129 |
|  | +,- | 两个frame的并集和交集 | 133 |
|  | Frame.apply | 函数的映射 | 137 |
|  | frame.sort\_index(by) | 通过将一个或多个列的名字传递给by，可以实现根据一个或多个列中的值进行排序 | 139 |
|  | frame.sum()  frame.describe() | 对数据进行汇总 | 144 |
|  |  |  |  |
| 处理缺失值 | dropna() | 直接抛弃存在NaN的列或行，可以设置阈值来调整 | 148 |
|  | fillna() | 对NaN的值进行填充 |  |

ix,loc,iloc,row,irow的区别

<https://blog.csdn.net/hecongqing/article/details/61927615>

文件读入

|  |  |  |  |
| --- | --- | --- | --- |
|  | 作用 | 备注 | 页码 |
| Pd.read\_csv() | 读取以逗号分隔的文件 | 可以读取文件的部分行或列 | 163 |
| Frame.to\_csv() | 将数据写到一个以逗号分隔的文件中 |  |  |

数据规整化

|  |  |  |  |
| --- | --- | --- | --- |
|  | 作用 | 备注 | 页码 |
| merge(df1,df2,on,how) | 根据一个或多个键将不同的DataFrame中的**行**连接起来。 | 默认是inner连接（交集），还可以是左连接，右连接和外连接（并集）。  参数left\_index,right\_index是利用索引做连接键。 | 190 |
| Concat([df1,df2],axis) | 沿着一条轴将多个对象堆叠在一起。 |  | 198 |

Merge，concat的区别

<https://www.cnblogs.com/bawu/p/7701810.html>

当没有索引时、concat不管列名，直接加到一起，可以加到后面、也可以加到右边，axis=0为加到后面，axis=1为加到右边，左边的数据结构没有变，变的是右边数据结构。

数据聚合与分组运算

|  |  |  |  |
| --- | --- | --- | --- |
| groupby | 根据参数对数据进行分组 | 可以支持迭代 | 265  268 |
|  |  |  |  |
|  |  |  |  |

df.groupby(‘key1’)[‘data1’]与df[‘data1’].groupby(df.[‘key1’])等价