

# 总结了这67个pandas函数，完美解决数据处理，拿来即用！

Python绿色通道 Yesterday

The following article is from 快学Python Author 黄伟呢



**快学Python**

Python可视化、自动化办公、数据分析、爬虫、Web开发！人生苦短，快学Python！

↑ 关注 + 星标 ，每天学Python新技能

后台回复【**大礼包**】送你Python自学大礼包



**Python绿色通道**

看Python干货，点击回复「大礼包」送你一份大礼，每天20个名额！

100篇原创内容

Official Account

作者：黄伟呢  
来源：快学Python

不管是业务数据分析，还是数据建模。数据处理都是及其重要的一个步骤，它对于最终的结果来说，至关重要。

今天，就为大家总结一下“Pandas数据处理”几个方面重要的知识，拿来即用，随查随查。

- 导入数据
- 导出数据
- 查看数据
- 数据选取
- 数据处理
- 数据分组和排序
- 数据合并

```
# 在使用之前，需要导入pandas库
```

```
import pandas as pd
```

## 导入数据

这里我为大家总结7个常见用法。

```
pd.DataFrame() # 自己创建数据框，用于练习

pd.read_csv(filename) # 从CSV文件导入数据

pd.read_table(filename) # 从限定分隔符的文本文件导入数据

pd.read_excel(filename) # 从Excel文件导入数据

pd.read_sql(query,connection_object) # 从SQL表/库导入数据

pd.read_json(json_string) # 从JSON格式的字符串导入数据

pd.read_html(url) # 解析URL、字符串或者HTML文件，抽取其中的tables表格
```

## 导出数据

这里为大家总结5个常见用法。

```
df.to_csv(filename) #导出数据到CSV文件

df.to_excel(filename) #导出数据到Excel文件

df.to_sql(table_name,connection_object) #导出数据到SQL表

df.to_json(filename) #以Json格式导出数据到文本文件

writer=pd.ExcelWriter('test.xlsx',index=False)
df1.to_excel(writer,sheet_name='单位')和writer.save(), 将多个数据帧写入同一个工作簿的多个sheet(工作表)
```

## 查看数据

这里为大家总结11个常见用法。

```
df.head(n) # 查看DataFrame对象的前n行

df.tail(n) # 查看DataFrame对象的最后n行

df.shape() # 查看行数和列数

df.info() # 查看索引、数据类型和内存信息

df.columns() # 查看字段（首行）名称

df.describe() # 查看数值型列的汇总统计

s.value_counts(dropna=False) # 查看Series对象的唯一值和计数

df.apply(pd.Series.value_counts) # 查看DataFrame对象中每一列的唯一值和计数

df.isnull().any() # 查看是否有缺失值

df[df[column_name].duplicated()] # 查看column_name字段数据重复的数据信息

df[df[column_name].duplicated()].count() # 查看column_name字段数据重复的个数
```

## 数据选取

这里为大家总结10个常见用法。

```
df[col] # 根据列名，并以Series的形式返回列

df[[col1,col2]] # 以DataFrame形式返回多列

s.iloc[0] # 按位置选取数据

s.loc['index_one'] # 按索引选取数据

df.iloc[0,:] # 返回第一行

df.iloc[0,0] # 返回第一列的第一个元素
```

```
df.loc[0,:] # 返回第一行（索引为默认的数字时，用法同df.iloc），但需要注意的是loc是按索引，iloc参数只接受数字

df.ix[:5],["col1","col2"]] # 返回字段为col1和col2的前5条数据，可以理解为loc和iloc的结合体。

df.at[5,"col1"] # 选择索引名称为5，字段名称为col1的数据

df.iat[5,0] # 选择索引排序为5，字段排序为0的数据
```

## 数据处理

这里为大家总结16个常见用法。

```
df.columns= ['a','b','c'] # 重命名列名（需要将所有列名列出，否则会报错）

pd.isnull() # 检查DataFrame对象中的空值，并返回一个Boolean数组

pd.notnull() # 检查DataFrame对象中的非空值，并返回一个Boolean数组

df.dropna() # 删除所有包含空值的行

df.dropna(axis=1) # 删除所有包含空值的列

df.dropna(axis=1,thresh=n) # 删除所有小于n个非空值的行

df.fillna(value=x) # 用x替换DataFrame对象中所有的空值，支持

df[column_name].fillna(x)

s.astype(float) # 将Series中的数据类型更改为float类型

s.replace(1,'one') # 用'one'代替所有等于1的值

s.replace([1,3],['one','three']) # 用'one'代替1，用'three'代替3

df.rename(columns=lambdax:x+1) # 批量更改列名

df.rename(columns={'old_name':'new_name'}) # 选择性更改列名

df.set_index('column_one') # 将某个字段设为索引，可接受列表参数，即设置多个索引
```

```
df.reset_index("col1") # 将索引设置为col1字段，并将索引新设置为0,1,2...

df.rename(index=lambdax:x+1) # 批量重命名索引
```

## 数据分组、排序、透视

这里为大家总结13个常见用法。

```
df.sort_index().loc[:5] # 对前5条数据进行索引排序

df.sort_values(col1) # 按照列col1排序数据，默认升序排列

df.sort_values(col2,ascending=False) # 按照列col1降序排列数据

df.sort_values([col1,col2],ascending=[True,False]) # 先按列col1升序排列，后按col2降序排列数据

df.groupby(col) # 返回一个按列col进行分组的Groupby对象

df.groupby([col1,col2]) # 返回一个按多列进行分组的Groupby对象

df.groupby(col1)[col2].agg(mean) # 返回按列col1进行分组后，列col2的均值,agg可以接受列表参数，agg([len,r

df.pivot_table(index=col1,values=[col2,col3],aggfunc={col2:max,col3:[ma,min]}) # 创建一个按列col1进行

df.groupby(col1).agg(np.mean) # 返回按列col1分组的所有列的均值,支持

df.groupby(col1).col2.agg(['min','max'])

data.apply(np.mean) # 对DataFrame中的每一列应用函数np.mean

data.apply(np.max,axis=1) # 对DataFrame中的每一行应用函数np.max

df.groupby(col1).col2.transform("sum") # 通常与groupby连用，避免索引更改
```

## 数据合并

这里为大家总结5个常见用法。

```
df1.append(df2) # 将df2中的行添加到df1的尾部
```

```
df.concat([df1,df2],axis=1,join='inner') # 将df2中的列添加到df1的尾部,值为空的对应行与对应列都不要
```

```
df1.join(df2.set_index(col1),on=col1,how='inner') # 对df1的列和df2的列执行SQL形式的join,默认按照索引连接
```

```
pd.merge(df1,df2,on='col1',how='outer') # 对df1和df2合并,按照col1,方式为outer
```

```
pd.merge(df1,df2,left_index=True,right_index=True,how='outer') # 与 df1.join(df2, how='outer')效果相同
```



## Python自学超级硬核资料

最后送大家一份Python学习大礼包，从Python基础，爬虫，数据分析Web开发等全套资料，吃透资料，你可以扔掉其他资料，这些资料都是视频，学起来非常友好

**Ps：都是视频学习资料，非常适合基础不好或者零基础的同学**

## 推荐阅读

1. 搞定，爬取公众号文章转换成PDF，自动邮件发送给自己！

2. 这可能是近 2 年进腾讯的最好机会！
3. 痛心！深圳程序员入职一月在家中身亡，已排除他杀，亲友称其常加班
4. 推荐60个相见恨晚的神器工具

看完记得关注@Python绿色通道

及时收看更多好文

↓↓↓



Python绿色通道

看Python干货，点击回复「大礼包」送你一份大礼，每天20个名额！

100篇原创内容

Official Account

点个 在看 你最好看

Read more

喜欢此内容的人还喜欢

MySQL 默认最大连接数是多少？

我是程序汪

手把手教你用Python求最大值和最小值

未闻Code

你一定要搞明白的C函数调用方式与栈原理

C语言与CPP编程