

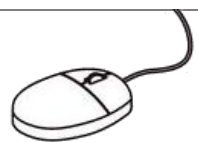


南通师范高等专科学校
Nantong Normal College

Python
数据分析与应用

pandas数据处理之
数据准备

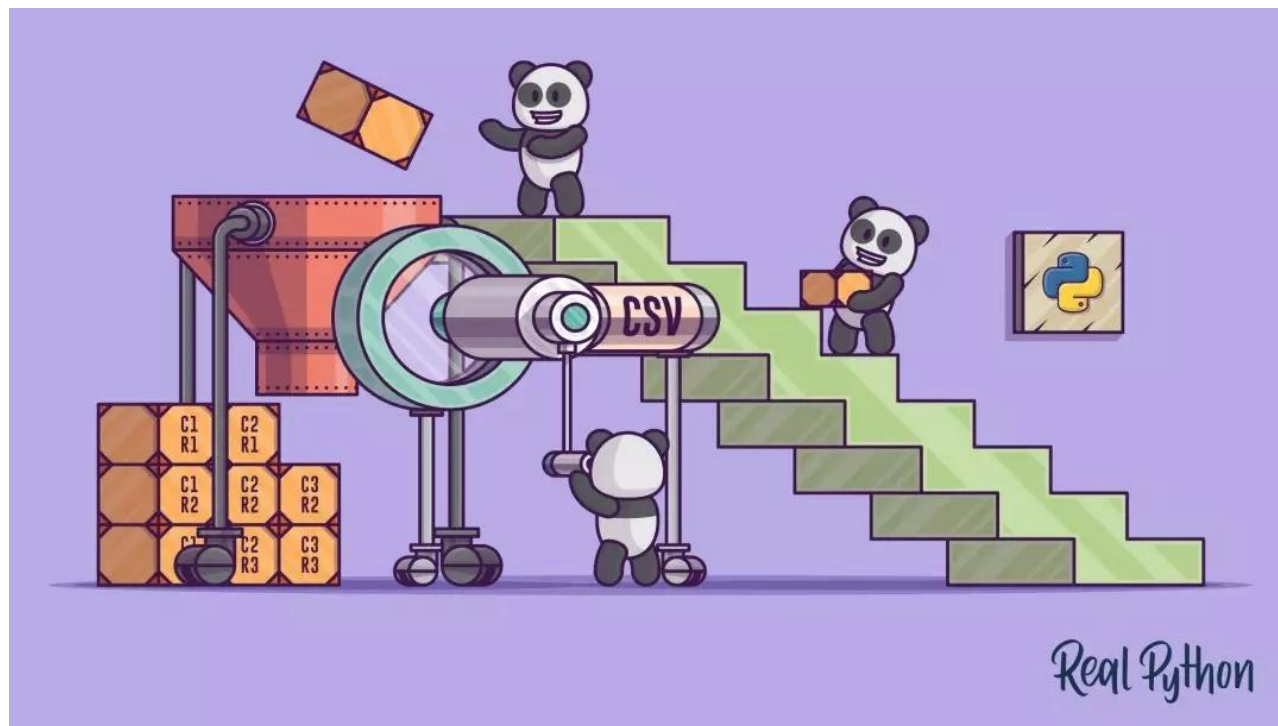
执教：朱亚林



数据导入



- 数据存在的形式多种多样，有文件（TXT、CSV、Excel）和数据库（MySQL、Access、SQL Server）等形式。
- 数据分析离不开数据，面对如此多种类的格式该怎样导入呢？



使用read_table()函数来读取文本文件。

```
read_table(filepath,sep='0\t',header='infer',names=None,index_col=None,dtype=None,encoding=utf-8,engine=None,nrows=None)
```

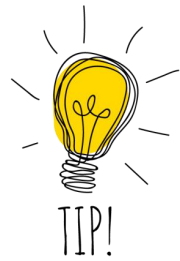
参数名称	说明
filepath	string, 读取文件的路径
sep	string, 代表分隔符。默认为制表符。
header	int或sequence。指定某行数据为列名。默认infer为自动识别。
names	array。表示列名。
index_col	int,sequence或False。表示索引列的位置。
dtype	dict。代表写入数据的类型。
engine	C或Python。代表数据解析引擎。默认为C。
nrows	int。表示读取前n行。

试试课本上的案例！



■ TXT都会导入了，CSV就是换个名字而已！函数名read_csv()。

不讲了，自己悟！



如果非要有点不一样，那就是他们的**默认切割符号**不一样，

read_table默认是'\t'(也就是tab)切割数据集的；

read_csv 默认是','(也就是逗号)切割数据集的！



- 丨 Excel文件的导入需要使用`read_excel()`函数。
- 丨 可接收的文件类型为“xls”和“xlsx”。
- 丨 除了参数里多了一个`sheetname`(指定Excel工作簿中工作表的位置), 其他都一样。

其实还是不想讲，自己悟！

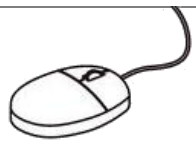


这个悟不了!

- 丨 要读取MySQL数据库，需要导入MySQLdb模块
- 丨 需要连接MySQL数据库
- 丨 使用read_sql()函数

```
>>> import pandas as pd
>>> import MySQLdb
>>>
con=MySQLdb.connect(host='localhost',user='root',passwd='pwd',db='dbname',port=5029,charset='utf8')
>>> data=pd.read_sql("select * from books;",con=con)
>>> connection.close()
```

EXAMPLE



数据导出



- 数据的导出是指将处理后的数据保存到具体文件中，这些常见的文件类型与导出对象一致，有文本、Excel文件和数据库等等。



- 导出方式与导入类似，对于结构化的数据，可以使用`to_csv()`函数来实现以CSV文件格式存储。

```
to_csv(filepath, sep=',', na_rep="", columns=None, header=True, index=True,  
mode='w', index_label=None, encoding=None)
```



to_csv()函数的语法一般如下：

```
to_csv(filepath, sep=',', na_rep="", columns=None, header=True, index=True, mode='w', index_label=None, encoding=None)
```

参数名称	说明
filepath	string, 写入文件的路径
sep	string, 代表分隔符。默认为制表符。
na_rep	string, 代表缺失值。默认为"
columns	list, 表示是写出的列名。
header	boolean。表示是否将列名写出。默认为True
index	boolean。表示是否将行名（索引）写出。默认为True
index_label	sequence。表示索引名。默认为None
mode	string。代表数据写入模式。默认为w
encoding	string。代表存储文件的编码格式。默认为None



- 构建一个DataFrame，并将它以csv的格式写出。



- 导出为Excel文件可使用`to_excel()`函数。该函数的使用方法与`to_csv()`函数基本一致。区别之处在于：
 - `to_excel()`函数没有`sep`参数
 - `to_excel()`函数增加了`sheetname`参数
- 尝试将上个练习中的DataFrame以xlsx的格式导出

