

欢迎继续我们的学习。

在这一课的内容中，我们将学习Pandas中的数据输入和输出。

这里主要会介绍四种格式的数据输入：

1.CSV

2.Excel

3.HTML

4.SQL

在此之前，我们需要使用pip或conda安装以下的module:

1.conda install sqlalchemy

2.conda install lxml

3.conda install html5lib

4.conda install BeautifulSoup4

接下来打开Anaconda和Jupyternotebook，创建一个新的notebook，命名为DataInOut

输入以下代码：

```
In [1]: import pandas as pd
```

```
In [2]: pwd
```

```
Out[2]: '/Users/eseedo/Documents/PythonProjects'
```

在输入pwd之后可以看到项目所在的文件夹路径。

在继续之前，请在这个文件夹中准备好四种不同格式的文件，

example.csv,example.xls,example.htm,example.sql

接下来看看如何打开和读取csv格式的文件。

首先看看example.csv文件中的内容：

	A	B	C	D
1	a	b	c	d
2	12	213	22	34
3	4	5	6	8
4	3	23	10	11
5	90	8	2	16
6				

在jupyter notebook中输入以下代码：

```
In [5]: pd.read_csv('example.csv')
```

Out[5]:

	a	b	c	d
0	12	213	22	34
1	4	5	6	8
2	3	23	10	11
3	90	8	2	16

只需一行代码，就轻松的打开并读取了example.csv中的内容。
实际上，Pandas可以读取多种格式的文件：

Out[5]:

- pd.read_clipboard
- pd.read_csv
- pd.read_excel
- pd.read_feather
- pd.read_fwf
- pd.read_gbq
- pd.read_hdf
- pd.read_html
- pd.read_json
- pd.read_msgpack

```
In [ ]: pd.read
```

接下来看看如何编辑csv文件中的内容。

```
In [6]: df = pd.read_csv('example.csv')

In [7]: df.to_csv('My_output',index=False)

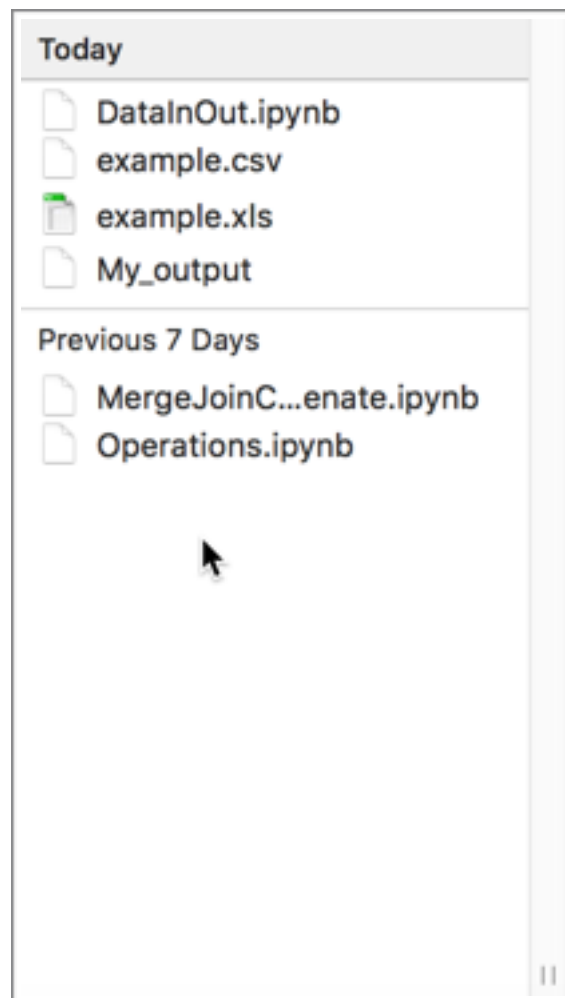
In [8]: pd.read_csv('My_output')

Out[8]:
```

	a	b	c	d
0	12	213	22	34
1	4	5	6	8
2	3	23	10	11
3	90	8	2	16

在以上代码中，我们将example.csv的内容读取到df这个dataframe中，然后将其输出到My_output.csv文件，最后读取其中的内容。

注意，此时如果查看项目文件夹，会看到文件夹中多了My_output文件。



接下来看看Pandas中如何读取和输出excel文件。

需要说明的是，Pandas只能读取excel文件中的数据，至于公式、图片等更复杂的对象就无能为力了。

在Pandas中，会将其中的每个sheet看作一个dataframe。

```
In [10]: pd.read_excel('example.xls', sheet_name='Sheet1')
```

Out[10]:

	a	b	c	d
0	12	213	22	34
1	4	5	6	8
2	3	23	10	11
3	90	8	2	16

在以上代码中，需要在sheet_name中指定要读取的是哪一个sheet。另外Pandas也支持xlsx格式的文件。

输出还是一行代码可以解决的问题~

```
In [11]: df.to_excel('sample2.xlsx', sheet_name='NewSheet')
```

Today

- DataInOut.ipynb
- example.csv
- example.xls
- My_output
- sample2.xlsx**

Previous 7 Days

- MergeJoinC...enate.ipynb
- Operations.ipynb

接下来看看如何读取html中的数据。

```
In [25]: data = pd.read_html('https://www.fdic.gov/bank/individual/failed/banklist.html')
```

```
In [28]: data[0].head()
```

Out[28]:

	Bank Name	City	ST	CERT	Acquiring Institution	Closing Date	Updated Date	Loss Share Type	Agreement Terminated	Ter Dat
0	The Woodbury Banking Company	Woodbury	GA	11297	United Bank	August 19, 2016	August 30, 2016	NaN	NaN	NaN

需要注意的是，读取html中的数据并非每次都能成功，因为有些页面做了相应的限制。

最后，让我们来了解下如何读取和写入sql。

```
In [22]: from sqlalchemy import create_engine
```

```
In [23]: engine = create_engine('sqlite:///memory:')
```

```
In [25]: df.to_sql('my_table',engine)
```

```
In [26]: sqldf = pd.read_sql('my_table',con = engine)
```

```
In [27]: sqldf
```

Out[27]:

	index	a	b	c	d
0	0	12	213	22	34
1	1	4	5	6	8
2	2	3	23	10	11
3	3	90	8	2	16

关于对sql的操作，这里就不展开讲了。

在后续的学习和实践中，我们主要涉及到的还是读写csv格式的文件。

好了，本课的内容就到这里了。

关于Pandas的知识，我们先介绍到这里。
我们下一课再见~

讨论群-笨猫学编程QQ群：
375143733

知乎专栏：
<https://zhuanlan.zhihu.com/kidscoding>

新浪博客：
<http://blog.sina.com.cn/eseedo>

Github：
<https://github.com/eseedo>

个人网站：
<http://icode.ai/>