06-Data-Input-and-Output

October 19, 2022

Copyright Pierian Data

For more information, visit us at www.pieriandata.com

NOTE: Typically we will just be either reading csv files directly or using pandas-datareader to pull data from the web. Consider this lecture just a quick overview of what is possible with pandas (we won't be working with SQL or Excel files in this course)

1 Data Input and Output

This notebook is the reference code for getting input and output, pandas can read a variety of file types using its pd.read_ methods. Let's take a look at the most common data types:

```
[1]: import numpy as np
     import pandas as pd
```

1.1 CSV

Comma Separated Values files are text files that use commas as field delimeters. Unless you're running the virtual environment included with the course, you may need to install xlrd and openpyxl. In your terminal/command prompt run:

```
conda install xlrd
conda install openpyxl
```

Then restart Jupyter Notebook. (or use pip install if you aren't using the Anaconda Distribution)

1.1.1 CSV Input

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

```
[2]:    a  b  c  d
     0  0  1  2  3
```

```
1    4    5    6    7
2    8    9   10   11
3   12   13   14   15
```

1.1.2 CSV Output

```
[3]: df.to_csv('example.csv',index=False)
```

1.2 Excel

Pandas can read and write MS Excel files. However, this only imports data, not formulas or images. A file that contains images or macros may cause the .read_excel()method to crash.

1.2.1 Excel Input

```
[4]: pd.read_excel('Excel_Sample.xlsx',sheet_name='Sheet1')
```

```
[4]:     a    b    c    d
     0   0    1    2    3
     1   4    5    6    7
     2   8    9   10   11
     3  12   13   14   15
```

1.2.2 Excel Output

```
[5]: df.to_excel('Excel_Sample.xlsx',sheet_name='Sheet1')
```

1.3 HTML

Pandas can read table tabs off of HTML. Unless you're running the virtual environment included with the course, you may need to install lxml, htmllib5, and BeautifulSoup4. In your terminal/command prompt run:

```
conda install lxml
conda install html5lib
conda install beautifulsoup4
```

Then restart Jupyter Notebook. (or use pip install if you aren't using the Anaconda Distribution)

1.3.1 HTML Input

Pandas read_html function will read tables off of a webpage and return a list of DataFrame objects:

```
[6]: df = pd.read_html('http://www.fdic.gov/bank/individual/failed/banklist.html')
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
/tmp/ipykernel_12074/1593670774.py in <module>
----> 1 df = pd.read_html('http://www.fdic.gov/bank/individual/failed/banklist.
 ↪html')

~/anaconda3/envs/tsa_course/lib/python3.7/site-packages/pandas/io/html.py in↵
 ↪read_html(io, match, flavor, header, index_col, skiprows, attrs, parse_dates, ↵
 ↪tupleize_cols, thousands, encoding, decimal, converters, na_values,↵
 ↪keep_default_na, displayed_only)
    985                          decimal=decimal, converters=converters,↵
 ↪na_values=na_values,
    986                          keep_default_na=keep_default_na,
--> 987                          displayed_only=displayed_only)


~/anaconda3/envs/tsa_course/lib/python3.7/site-packages/pandas/io/html.py in↵
 ↪_parse(flavor, io, match, attrs, encoding, displayed_only, **kwargs)
    813                 break
    814         else:
--> 815             raise_with_traceback(retained)
    816
    817     ret = []


~/anaconda3/envs/tsa_course/lib/python3.7/site-packages/pandas/compat/__init__.
 ↪py in raise_with_traceback(exc, traceback)
    402             if traceback == Ellipsis:
    403                 _, _, traceback = sys.exc_info()
--> 404             raise exc.with_traceback(traceback)
    405         else:
    406             # this version of raise is a syntax error in Python 3


ValueError: No tables found
```

```
[8]: df[0].head()
```

```
[8]:                                         Bank Name          City  ST   CERT  \
     0             Washington Federal Bank for Savings       Chicago  IL  30570
     1     The Farmers and Merchants State Bank of Argonia       Argonia  KS  17719
     2                             Fayette County Bank    Saint Elmo  IL   1802
     3   Guaranty Bank, (d/b/a BestBank in Georgia & Mi…    Milwaukee  WI  30003
     4                                   First NBC Bank   New Orleans  LA  58302
```

	Acquiring Institution	Closing Date	Updated Date
0	Royal Savings Bank	December 15, 2017	February 21, 2018
1	Conway Bank	October 13, 2017	February 21, 2018
2	United Fidelity Bank, fsb	May 26, 2017	July 26, 2017
3	First-Citizens Bank & Trust Company	May 5, 2017	March 22, 2018
4	Whitney Bank	April 28, 2017	December 5, 2017

2 Great Job!