Advanced Python
Assignment – Web Pages – Access Web Page Data with pandas and Beautiful Soup – Yield Curves

This assignment has 2 parts. Each part involves using Python to get data from a web page with US Treasury interest rates, available at:
https://home.treasury.gov/resource-center/data-chart-center/interest-rates/TextView?type=daily_treasury_yield_curve&field_tdr_date_value=2023

You should submit 2 notebooks for this assignment, one for each part. Use several cells for each part. Name your notebook like: LastnameFirstnameAsn3<part>, for example, SmithJohnAsn3A. Include in each cell of your notebook a Python comment at the start of the cell, describing the cell's statements.

Part A – Accessing a Web Page Table Using pandas

Use pandas `read_html()` to access the Treasury yields from the page indicated above. In particular,

1) Call the `read_html()` function to get all HTML tables from the above web page. How many tables are returned?

2) From the list of returned DataFrames, set the DataFrame with the interest rates into its own variable.

3) For the DataFrame in step #3, set the Date column to be the row names (index). Display this DataFrame.

4) From the table of yields (one row for each date), extract the latest yields into a Python list. This is the last row of the table. (The particular date will depend on when you run your program.)

5) Plot the latest yield curve. This is a scatter chart with yields (in percent) on the Y axis, and time (in years) on the X axis. Time should be represented as a numeric list, e.g., time = [1/12, 2/12, …, 30]

6) In one chart, plot the time series of the 2 year, 5 year, and 10 year yields. These are all the values in the columns labeled '2 Yr', '5 Yr', and '10 Yr'.

Part B – Accessing a Web Page Table Using Beautiful Soup

Use requests and bs4 packages to access the Treasury yields from the page indicated above. In particular,

1) Use the requests package to read the data from the url indicated above, using its get() method

2) Create a BeautifulSoup instance from the data read by requests

2) Locate the <table> element in the page corresponding to the Treasury yields.

3) From the BeautifulSoup table Tag located in step 2, create a DataFrame.

4) Similar to part A step 7, plot the time series of the 2 year, 5 year, and 10 year yields.

5) From the BeautifulSoup data, locate and display the date displayed below the table on the web page. This is in a <p> tag with the class attribute 'page-watch-date'. Use the .text attribute so that the date without the enclosing <p> tag is displayed.

Appendix – Plotting notes

Scatter plots

There are several options for creating a simple scatter plot in Python. Here is a simple example using matplotlib, assuming x and y are lists of numbers, with the same number of elements:

```
import matplotlib.pyplot as plt
plt.scatter(x, y)
plt.show()
```

Line charts

pandas can plot 1 or more time series, taking the data from a DataFrame. Here is an example, assuming df is a DataFrame with columns 'A' and 'B':

```
df[['A','B']].plot()
```