Python for Economics

Zhenhua He
Richard Lawrence
Morning session, 10/1/2021
Table of Contents

This course is divided into numbered lessons

17. Web Scraping with BeautifulSoup

- Advanced exercise: Candlestick plot
- Wrap-up
Lesson 17
Web Scraping with BeautifulSoup

Use Python libraries to extract data from a website
Applying Lesson Learning Objectives

After this lesson, you will know:

- What are regular expressions
- What is BeautifulSoup library and how to use it

During this lesson, you will use:

- What is web scraping
- HTML Basics
- What is requests library and how to use it
What is web scraping

- **Scraping**: a process of extracting (from various means), copying and screening of data.
- **Web Scraping**: scraping or extracting data or feeds from the web (like from web-pages or websites)
Why does one scrape the web?

- Collect data for Research
- Sales and Marketing
- Products prices & popularity comparison
- …
Regular Expression

- A sequence of characters forms a search **pattern**
- `findall()`: Check if a string contains the pattern
- Python built-in module **re**
- Example

```python
import re

text = "I went to TAMU at 8 AM"

pattern = "AM"

result = re.findall(pattern, text)

["AM", "AM"]
```
Regular Expression

re.compile()

- convert a pattern string into a pattern object
- **Object** combines the pattern and `findall()` function
- Example

  ```python
  text = "I went to TAMU at 8 AM"
  pattern = re.compile("AM")
  result = pattern.findall(text)
  ```

- No need to enter the pattern string again
Regular Expression Examples and Exercises

Go to Google Classroom Lesson 17 “Regular Expressions”

Tasks

● Follow instructions for the examples
● Work on the exercises (due by 10/1 6:00 PM)
BeautifulSoup

- A Python library for pulling data out of HTML files.
- Works with a *parser* to provide idiomatic ways of navigating, searching, and modifying the source tree.
- Built-in parser available (we will use)
Review: JSON - JavaScript Object Notation

- A text format for storing data
- Language-independent
- The data to extract could be in JSON format.

JSON string examples:

'{"name":"Jack", "age":20, "major":"computer science"}"

'{"args": {}, "data": ", "files": {}, "form": { "soup": "hot soup"}, ... }'

From the Requests exercise we did previously
JSON module

- Python built-in module `json`
- `json.loads()`: converts JSON string to Python dictionary
- Example
  ```python
  import json
  text = "{"keys : values, ... ""
  dict = json.loads(text)
  ```
Today’s Exercise: Scraping a website

- Collect salary stats from https://www.salary.com/
- Salary stats for a single city
- Salary stats for multiple cities
- Save data in a file
Investigate salary stats
Analyze the salary stats
Collect the salary stats for Dallas, TX

<table>
<thead>
<tr>
<th>PERCENTILE</th>
<th>SALARY</th>
<th>LOCATION</th>
<th>LAST UPDATED</th>
</tr>
</thead>
<tbody>
<tr>
<td>10th</td>
<td>$47,610</td>
<td>Dallas,TX</td>
<td>August 27, 2021</td>
</tr>
<tr>
<td>25th</td>
<td>$52,800</td>
<td>Dallas,TX</td>
<td>August 27, 2021</td>
</tr>
<tr>
<td>50th</td>
<td>$58,500</td>
<td>Dallas,TX</td>
<td>August 27, 2021</td>
</tr>
<tr>
<td>75th</td>
<td>$65,200</td>
<td>Dallas,TX</td>
<td>August 27, 2021</td>
</tr>
<tr>
<td>90th</td>
<td>$71,300</td>
<td>Dallas,TX</td>
<td>August 27, 2021</td>
</tr>
</tbody>
</table>
Collect stats for multiple cities – structured URL

Senior accountant
Dallas, Texas
https://www.salary.com/research/salary/alternate/senior-accountant-salary/dallas-tx

Entry data analyst
Miami, Florida
https://www.salary.com/research/salary/alternate/entry-data-analyst-salary/miami-fl

Tax accountant entry
Washington, DC
https://www.salary.com/research/salary/alternate/tax-accountant-entry-salary/washington-dc

url_template = 'https://www.salary.com/research/salary/alternate/{{}}-salary/{{}'}
Goal: CSV file with collected data

<table>
<thead>
<tr>
<th>Title</th>
<th>Location</th>
<th>Description</th>
<th>1 to 5 of 5 entries</th>
<th>Filter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tax Accountant - Entry</td>
<td>New York, NY</td>
<td>The Tax Accountant - Entry helps with the preparation of local, state and federal level returns to be submitted within specified tax deadlines. Assists in the maintenance and preparation of tax-related items to include tax records, tax returns, tax schedules, and related tax reports. Being a Tax Accountant - Entry typically reports to a supervisor or a manager. May require a bachelor's degree. Being a Tax Accountant - Entry works on projects/matters of limited complexity in a support role. Work is closely managed. Working as a Tax Accountant - Entry typically requires 0-2 years of related experience.</td>
<td>57618</td>
<td>78900</td>
</tr>
<tr>
<td>Tax Accountant - Entry</td>
<td>Los Angeles, CA</td>
<td>The Tax Accountant - Entry helps with the preparation of local, state and federal level returns to be submitted within specified tax deadlines. Assists in the maintenance and preparation of tax-related items to include tax records, tax returns, tax schedules, and related tax reports. Being a Tax Accountant - Entry typically reports to a supervisor or a manager. May require a bachelor's degree. Being a Tax Accountant - Entry works on projects/matters of limited complexity in a support role. Work is closely managed. Working as a Tax Accountant - Entry typically requires 0-2 years of related experience.</td>
<td>54182</td>
<td>74200</td>
</tr>
<tr>
<td>Tax Accountant - Entry</td>
<td>Chicago, IL</td>
<td>The Tax Accountant - Entry helps with the preparation of local, state and federal level returns to be submitted within specified tax deadlines. Assists in the maintenance and preparation of tax-related items to include tax records, tax returns, tax schedules, and related tax reports. Being a Tax Accountant - Entry typically reports to a supervisor or a manager. May require a bachelor's degree. Being a Tax Accountant - Entry works on projects/matters of limited complexity in a support role. Work is closely managed. Working as a Tax Accountant - Entry typically requires 0-2 years of related experience.</td>
<td>50937</td>
<td>76155</td>
</tr>
<tr>
<td>Tax Accountant - Entry</td>
<td>Houston, TX</td>
<td>The Tax Accountant - Entry helps with the preparation of local, state and federal level returns to be submitted within specified tax deadlines. Assists in the maintenance and preparation of tax-related items to include tax records, tax returns, tax schedules, and related tax reports. Being a Tax Accountant - Entry typically reports to a supervisor or a manager. May require a bachelor's degree. Being a Tax Accountant - Entry works on projects/matters of limited complexity in a support role. Work is closely managed. Working as a Tax Accountant - Entry typically requires 0-2 years of related experience.</td>
<td>49219</td>
<td>73482</td>
</tr>
<tr>
<td>Tax Accountant - Entry</td>
<td>Phoenix, AZ</td>
<td>The Tax Accountant - Entry helps with the preparation of local, state and federal level returns to be submitted within specified tax deadlines. Assists in the maintenance and preparation of tax-related items to include tax records, tax returns, tax schedules, and related tax reports. Being a Tax Accountant - Entry typically reports to a supervisor or a manager. May require a bachelor's degree. Being a Tax Accountant - Entry works on projects/matters of limited complexity in a support role. Work is closely managed. Working as a Tax Accountant - Entry typically requires 0-2 years of related experience.</td>
<td>47401</td>
<td>70900</td>
</tr>
</tbody>
</table>
Examples and Exercises

Go to Google Classroom assignment Lesson 17 “Beautiful Soup”

Tasks

● Follow instructions for the examples
● Work on the exercises (due by 10/1 6:00 PM )
Lesson 18
Candlestick Plot

Use Matplotlib finance module (mplfinance) to plot stock price movement
Candlestick plot

![Candlestick plot image]
Candlestick Plot

The color of candlestick rectangle area depends on the relationship between the opening and closing price of a stock.

The color of the rectangle shows increase/decrease.
Candlestick Plot

Keyword arguments:

- type
- style
- volume (trade volume)
- mav (moving averages)
Examples and Exercises

Go to Google Classroom assignment Lesson 18 “Candlestick Plot”

Tasks

● Follow instructions for the examples
● Work on the exercises (due by 10/1 6:00 PM)
References

- https://www.w3schools.com/html/
- https://github.com/matplotlib/mplfinance