

```
In [1]: #!/usr/bin/env python
# coding: utf-8

#
# - Python web scraping libraries you need for the course and how to install them
# - How to extract URLs from one webpage.
# - How to extract other text data pieces from one webpage.
# - How to crawl multiple webpages and extract data from each of them.
# - How to handle navigation links and move to *next* pages.
# - How to save your scraped data into a CSV file.
# - And finally, a quick overview about *other* popular web scraping frameworks.

from bs4 import BeautifulSoup
import requests
import pandas as pd
```

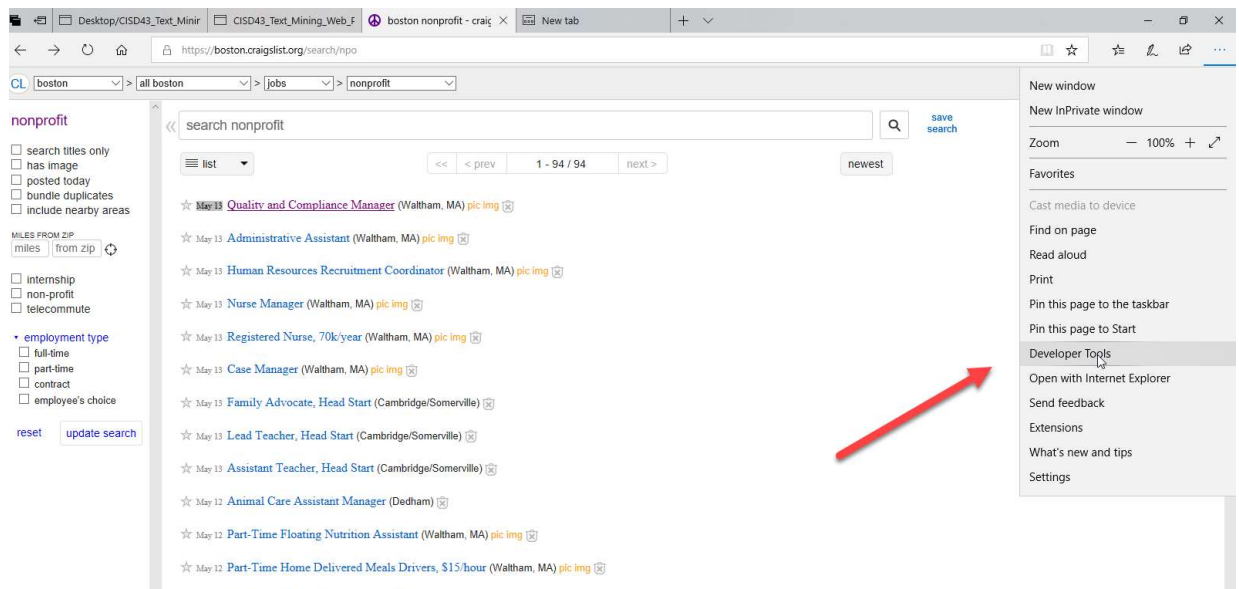
```
In [2]: url = "https://boston.craigslist.org/search/npo"

# Create a dictionary
d = {'key': 'value'}
print(d)

# Update the dictionary
d['new key'] = 'new value'
print(d)

{'key': 'value'}
{'key': 'value', 'new key': 'new value'}
```

## Using developer tools in the ie web browser to find classes used in the web page



CL

**nonprofit**

☐ search titles only  
☐ has image  
☐ posted today  
☐ bundle duplicates  
☐ include nearby areas

MILES FROM ZIP

☐ internship  
☐ non-profit  
☐ telecommute

▼ employment type  
☐ full-time  
☐ part-time  
☐ contract  
☐ employee's choice

search nonprofit

list << < prev 1 - 94 / 94 next > newest

★ May 13 [Quality and Compliance Manager \(Waltham, MA\)](#) pic img

★ May 13 [Administrative Assistant \(Waltham, MA\)](#) pic img

★ May 13 [Human Resources Recruiter \(Waltham, MA\)](#) pic img

★ May 13 [Nurse Manager \(Waltham, MA\)](#) pic img

★ May 13 [Registered Nurse, 70k/year \(Waltham, MA\)](#) pic img

★ May 13 [Case Manager \(Waltham, MA\)](#) pic img

★ May 13 [Family Advocate, Head Start \(Cambridge/Somerville\)](#) pic img

★ May 13 [Lead Teacher, Head Start \(Cambridge/Somerville\)](#) pic img

★ May 13 [Assistant Teacher, Head Start \(Cambridge/Somerville\)](#) pic img

★ May 12 [Animal Care Assistant Manager \(Dedham\)](#) pic img

★ May 12 [Part-Time Floating Nutrition Assistant \(Waltham, MA\)](#) pic img

★ May 12 [Part-Time Home Delivered Meals Drivers, \\$15/hour \(Waltham, MA\)](#) pic img

★ May 12 [Part-Time Dining Site Assistant, \\$15/hour \(Waltham, MA\)](#) pic img

Open in new tab  
 Open in new window  
 Open in new InPrivate window  
 Save target as  
 Copy link  
 Add to reading list  
 Search the web for "Quality and Compliance Manager"  
 Ask Cortana about "Quality and Compliance Manager"  
 Select all  
 Copy  
 View source  
 Inspect element

← → ↻ ↺

CL

**nonprofit**

☐ search titles only  
☐ has image  
☐ posted today  
☐ bundle duplicates  
☐ include nearby areas

MILES FROM ZIP

☐ internship  
☐ non-profit  
☐ telecommute

▼ employment type  
☐ full-time  
☐ part-time  
☐ contract  
☐ employee's choice

search nonprofit

list << < prev 1 - 94 / 94 next > newest

★ May 13 [Quality and Compliance Manager \(Waltham, MA\)](#) pic img

★ May 13 [Administrative Assistant \(Waltham, MA\)](#) pic img

★ May 13 [Human Resources Recruiter \(Waltham, MA\)](#) pic img

★ May 13 [Nurse Manager \(Waltham, MA\)](#) pic img

★ May 13 [Registered Nurse, 70k/year \(Waltham, MA\)](#) pic img

★ May 13 [Case Manager \(Waltham, MA\)](#) pic img

★ May 13 [Family Advocate, Head Start \(Cambridge/Somerville\)](#) pic img

★ May 13 [Lead Teacher, Head Start \(Cambridge/Somerville\)](#) pic img

★ May 13 [Assistant Teacher, Head Start \(Cambridge/Somerville\)](#) pic img

★ May 12 [Animal Care Assistant Manager \(Dedham\)](#) pic img

★ May 12 [Part-Time Floating Nutrition Assistant \(Waltham, MA\)](#) pic img

★ May 12 [Part-Time Home Delivered Meals Drivers, \\$15/hour \(Waltham, MA\)](#) pic img

★ May 12 [Part-Time Dining Site Assistant, \\$15/hour \(Waltham, MA\)](#) pic img

right click and select inspect element

Elements Console Debugger Network Performance Memory

```
<div class="global-header narrow"></div>
<form class="search-form" id="searchform" action="/search/npo">
  <div class="querybox"></div>
  <div class="search-options-container"></div>
  <div class="search-legend"></div>
  <div class="content" id="sortable-results">
    <section class="favlistsection"></section>
    <div class="open-map-view-button"></div>
    <div id="mapcontainer" data-arealonn="71.060303" data-arealat="42.358299"></div>
    <ul class="rows">
      <li class="result-row" data-pid="7123618539">
        <a class="result-image gallery" href="https://boston.craigslist.org/gbs/npo/d/waltham-quality-and-compliance-manager/7123618539.html" data-id="1:00808_IPvtgvd5Iu"></a>
        <p class="result-info">
          <span title="save this post in your favorites list" class="con icon-star" role="button"></span>
          <time title="Wed 13 May 04:25:50 PM" class="result-date" datetime="2020-05-13 16:25:50">May 13</time>
          <div class="result-title hdrlnk" href="https://boston.craigslist.org/gbs/npo/d/waltham-quality-and-compliance-manager/7123618539.html" data-id="7123618539">Quality and Compliance Manager</div>
          <span class="result-meta"></span>
        </p>
      </li>
      <li class="result-row" data-pid="7123618079"></li>
      <li class="result-row" data-pid="7123617743"></li>
      <li class="result-row" data-pid="7123617209"></li>
      <li class="result-row" data-pid="7123616929"></li>
      <li class="result-row" data-pid="7123616486"></li>
      <li class="result-row" data-pid="7123507216"></li>
      <li class="result-row" data-pid="7123504820"></li>
      <li class="result-row" data-pid="7123503013"></li>
      <li class="result-row" data-pid="7123049892"></li>
      <li class="result-row" data-pid="7123044906"></li>
      <li class="result-row" data-pid="7123044728"></li>
    </ul>
  </div>
```

right click and select inspect element

```

<div class="result-row" data-pid="7123618539">
  <a class="result-image gallery" href="https://boston.craigslist.org/gbs/npo/d/waltham-quality-and-compliance-manager/7123618539.html" data-id="1:00808_IPvkzgvdsIu"></a>
  <p class="result-info">
    <span title="save this post in your favorites list" class="icon icon-star" role="button"></span>
    <time title="Wed 13 May 04:25:50 PM" class="result-date" datetime="2020-05-13 16:25">May 13</time>
    <a class="result-title href" href="https://boston.craigslist.org/gbs/npo/d/waltham-quality-and-compliance-manager/7123618539.html" data-id="7123618539">Quality and Compliance Manager</a>
    <span class="result-meta"></span>
  </p>
</div>

```

location tag

Inspect element



https://boston.craigslist.org/search/npo

jobs nonprofit

search nonprofit

list prev 1 - 94 next newest

May 13 Quality and Compliance Manager (Waltham, MA) pic img

May 13 Administrative Assistant (Waltham, MA) pic img

May 13 Human Resources Recruiter (Waltham, MA) pic img

May 13 Nurse Manager (Waltham, MA) pic img

May 13 Registered Nurse, 70k/year (Waltham, MA) pic img

May 13 Case Manager (Waltham, MA) pic img

May 13 Family Advocate, Head Start (Cambridge/Somerville) pic img

May 13 Lead Teacher, Head Start (Cambridge/Somerville) pic img

May 13 Assistant Teacher, Head Start (Cambridge/Somerville) pic img

May 12 Animal Care Assistant Manager (Dedham) pic img

May 12 Part-Time Floating Nutrition Assistant (Waltham, MA) pic img

May 12 Part-Time Home Delivered Meals Drivers, \$15/hour (Waltham, MA) pic img

May 12 Part-Time Dining Site Assistant, \$15/hour (Waltham, MA) pic img

May 12 Elder Respite Services Transition (Waltham, MA) pic img

class used for location tag

```

<section class="favlistsection"></section>
<div class="open-map-view-button"></div>
<div id="mapcontainer" data-arealonn="-71.060303" data-arealat
  ="42.358299"></div>
<ul class="rows">
  <li class="result-row" data-pid="7123618539">
    <a class="result-image gallery" href="https://boston.craigs
      list.org/gbs/npo/d/waltham-quality-and-compliance-manager/712
      3618539.html" data-ids="1:00808_1Pvkzgv5IU"></a>
    <p class="result-info">
      <span title="save this post in your favorites list" class="
        con icon-star" role="button"></span>
      <time title="Wed 13 May 04:25:58 PM" class="result-date"
        datetime="2020-05-13 16:25">May 13</time>
      <a class="result-title hdrlnk" href="https://boston.craigs
        list.org/gbs/npo/d/waltham-quality-and-compliance-manager/712
        618539.html" data-id="7123618539">Quality and Compliance Ma
        nager</a>
      <span class="result-meta">
        <span class="result-hood">(Waltham, MA)</span>
        <span class="result-tags"></span>
        <span class="banish icon-trash" role="button"></span>
        <span class="unbanish icon-trash red" role="button"
          aria-hidden="true"></span>
        <a class="restore-link" href="https://boston.craigs
          list.org/gbs/npo/d/waltham-quality-and-compliance-manager/712
          618539.html" data-id="7123618539"></a>
      </span>
    </p>
  </li>
  <li class="result-row" data-pid="7123618079"></li>
  <li class="result-row" data-pid="7123617743"></li>
  <li class="result-row" data-pid="7123617209"></li>
  <li class="result-row" data-pid="7123616929"></li>
  <li class="result-row" data-pid="7123616486"></li>
  <li class="result-row" data-pid="7123507216"></li>
  <li class="result-row" data-pid="7123504820"></li>
  <li class="result-row" data-pid="7123503013"></li>
  <li class="result-row" data-pid="7123049892"></li>
  <li class="result-row" data-pid="7123044906"></li>

```

Not secure: https://boston.craigslist.org/gbs/npo/d/waltham-quality-and-compliance-manager/7123618539.html

CL boston > jobs > boston/camb/brook > nonprofit

prev next

reply below favorite hide flag

Posted about 3 hours ago

Quality and Compliance Manager (Waltham, MA)

Exciting opportunity for the unique professional who has great clinical skills and enjoys working with data.

POSITION DESCRIPTION

Springwell is seeking a professional who has great clinical skills and enjoys working with data. The position has a variety of responsibilities:

- hands on tasks such as conducting record reviews, reviewing reports, and collaborating with systems that monitor standards.
- identifying and implementing standards, improve the delivery of programs,
- providing leadership around the development and process of annual surveys and record reviews,

307 Waverley Oaks Rd Suite 205

compensation: Competitive salary

employment type: full-time

```

<html class="js canvas draggable fileAPI geolocat
  a picture pushState placeholder no-touchCapable t
  ><head></head>
  <body class="posting desktop">
    <script type="text/javascript"><!-- function
      <section class="page-container">
        <div class="bglogo"></div>
        <div class="tryapp"></div>
        <header class="global-header wide"></header>
        <header class="global-header narrow"></hea
          <section class="body">
            <header class="dateReplyBar"></header>
            <h2 class="postingtitle"></h2>
            <section class="userbody">
              <figure class="iw oneimage"></figure>
              <div class="mapAndAttrs"></div>
              <section id="postingbody">
                <div class="print-information print-q
                  <p>Exciting opportunity for the u.</p>
                  <p>POSITION DESCRIPTION</p>
                  <p>Springwell is seeking an indiv.</p>
                  <ul></ul>
                  <p>QUALIFICATIONS</p>
                  <p>Our ideal candidate is incredi.</p>
                  <ul></ul>
                  <p>GENEROUS BENEFITS</p>
                  <ul></ul>
                  <p>COMPANY PROFILE</p>
                  <p>At Springwell, we are serious </p>
                  <p>We are also serious about our </p>
                  <p>We believe in:</p>

```

CL boston > jobs > post account  
boston/camb/brook > nonprofit

reply below favorite hide flag Posted about 3 hours ago print

## Quality and Compliance Manager (Waltham, MA)

Exciting opportunity for the unique professional who has the skills and enjoys working with data.

### POSITION DESCRIPTION

Springwell is seeking an individual to oversee quality improvement activities throughout the agency. The position has a variety of responsibilities that will consist of:

- hands on tasks such as running and reviewing reports, and conducting record reviews,
- collaborating with Program Managers to implement and maintain systems that monitor compliance with standards,
- identifying and creating practices that improve the delivery of programs,
- providing leadership around the development and process of annual surveys and record reviews,
- organizing and leading the Quality Improvement Committee,
- developing an annual Quality Improvement Plan, and
- tracking and analyzing data to support quality outcomes.

307 Waverley Oaks Rd Suite 205  
(google map)

compensation: Competitive salary  
employment type: full-time  
non-profit organization

for position description

```
<html class="js canvas draggable fileAPI geolocation hashChange matchMedia picture pushState placeholder no-touchCapable transitions">
  <head></head>
  <body class="posting desktop">
    <script type="text/javascript"><!-- function C(k){return(../script>
    <section class="page-container">
      <div class="bglogo"></div>
      <div class="tryapp"></div>
      <header class="global-header wide"></header>
      <header class="global-header narrow"></header>
      <section class="body">
        <header class="dateReplyBar"></header>
        <h2 class="postingtitle"></h2>
        <section class="userbody">
          <figure class="iw oneimage"></figure>
          <div class="mapAndAttrs"></div>
          <section id="postingbody">
            <div class="print-qr-container print-qr-code-container"></div>
            <p>Exciting opportunity for the u.</p>
            <p>POSITION DESCRIPTION</p>
            <p>Springwell is seeking an indiv.</p>
            <ul></ul>
            <p>QUALIFICATIONS</p>
            <p>Our ideal candidate is incre.</p>
            <ul></ul>
            <p>GENEROUS BENEFITS</p>
            <ul></ul>
            <p>COMPANY PROFILE</p>
            <p>At Springwell, we are serious </p>
            <p>We are also serious about our </p>
            <p>We believe in:</p>
            <ul></ul>
            <p></p>
            <p></p>
            <p>Keywords: Waltham, Newton, Bro.</p>
```

```

In [3]: npo_jobs = {}
        job_no = 0
        while True:

            response = requests.get(url)
            data = response.text
            soup = BeautifulSoup(data, 'html.parser')
            jobs = soup.find_all('p',{'class':'result-info'})

            for job in jobs:

                title = job.find('a',{'class':'result-title'}).text
                location_tag = job.find('span',{'class':'result-hood'})
                location = location_tag.text[2:-1] if location_tag else "N/A"
                date = job.find('time', {'class': 'result-date'}).text
                link = job.find('a', {'class': 'result-title'}).get('href')

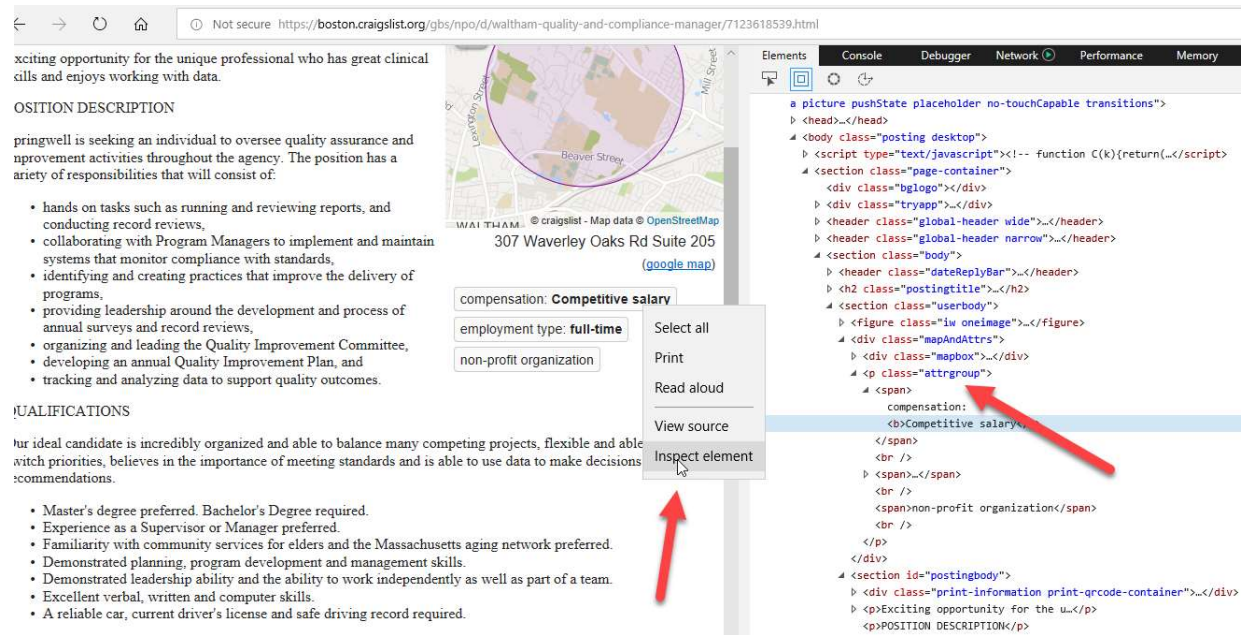
                job_response = requests.get(link)
                job_data = job_response.text
                job_soup = BeautifulSoup(job_data, 'html.parser')
                job_description = job_soup.find('section',{'id':'postingbody'}).text
                job_attributes_tag = job_soup.find('p',{'class':'attrgroup'})
                job_attributes = job_attributes_tag.text if job_attributes_tag else "N/A"

                job_no+=1
                npo_jobs[job_no] = [title, location, date, link, job_attributes, job_desc

#         print('Job Title:', title, '\nLocation:', location, '\nDate:', date, '\n

url_tag = soup.find('a',{'title':'next page'})
if url_tag.get('href'):
    url= 'https://boston.craigslist.org' + url_tag.get('href')
    print(url)
else:
    break

```



```
In [4]: print("Total Jobs:", job_no)
npj_jobs_df = pd.DataFrame.from_dict(npj_jobs, orient = 'index', columns = ['Job Title', 'Location', 'Date', 'Link', 'Job Attributes'])

npj_jobs_df.head()
```

Total Jobs: 116

Out[4]:

	Job Title	Location	Date	Link	Job Attributes
1	Quality and Compliance Manager	Waltham, MA	May 13	https://boston.craigslist.org/gbs/npod/waltha...	\ncompensation: Competitive salary\n\nemployment type:..
2	Administrative Assistant	Waltham, MA	May 13	https://boston.craigslist.org/gbs/npod/waltha...	\ncompensation: Competitive pay\n\nemployment ..
3	Human Resources Recruitment Coordinator	Waltham, MA	May 13	https://boston.craigslist.org/gbs/npod/waltha...	\ncompensation: Competitive\n\nemployment type..
4	Nurse Manager	Waltham, MA	May 13	https://boston.craigslist.org/gbs/npod/waltha...	\ncompensation: Competitive salary\n\nemployment..
5	Registered Nurse, 70k/year	Waltham, MA	May 13	https://boston.craigslist.org/gbs/npod/waltha...	\ncompensation: \$70k annually\n\nSalary\n\nemploy..

```
In [6]: npj_jobs_df.to_csv('data/npj_jobs.csv')
```

In [ ]: