

In [5]:

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

In [6]:

```
data=requests.get("https://realpython.github.io/fake-jobs/")
```

In [7]:

```
data
```

Out[7]:

```
<Response [200]>
```

In [8]:

```
data.text
```

Out[8]:

```
'<!DOCTYPE html>\n<html>\n  <head>\n    <meta charset="utf-8">\n    <meta name="viewport" content="width=device-width, initial-scale=1">\n    <title>Fake Python</title>\n    <link rel="stylesheet" href="https://cdn.jsdelivr.net/npm/bulma@0.9.2/css/bulma.min.css">\n  </head>\n  <body>\n    <section class="section">\n      <div class="container mb-5">\n        <h1 class="title is-1">\n          Fake Python\n        </h1>\n        <p class="subtitle is-3">\n          Fake Jobs for Your Web Scraping Journey\n        </p>\n      </div>\n      <div class="container">\n        <div id="ResultsContainer" class="columns is-multiline">\n          <div class="column is-half">\n            <div class="card">\n              <div class="card-content">\n                <div class="media">\n                  <div class="media-left">\n                    <figure class="image is-48x48">\n                      \n                    </figure>\n                  </div>\n                  <div class="media-content">\n                    <h2 class="title is-5">Senior Python Developer</h2>\n                    <h3 class="subtitle is-6 company">Payne, Roberts and Davis</h3>\n                  </div>\n                </div>\n                <div class="content">\n                  <p class="location">\n                    Stewarthurv. AA\n                  </p>\n                </div>\n              </div>\n            </div>\n          </div>\n        </div>\n      </div>\n    </body>\n  </html>
```

In [9]:

```
soup=BeautifulSoup(data.content,"html.parser")
```

In [10]:

soup

Out[10]:

```
<!DOCTYPE html>

<html>
<head>
<meta charset="utf-8"/>
<meta content="width=device-width, initial-scale=1" name="viewport"/>
<title>Fake Python</title>
<link href="https://cdn.jsdelivr.net/npm/bulma@0.9.2/css/bulma.min.css"
rel="stylesheet"/>
</head>
<body>
<section class="section">
<div class="container mb-5">
<h1 class="title is-1">
    Fake Python
    </h1>
<p class="subtitle is-3">
    Fake Jobs for Your Web Scrapping Journey
```

In [11]:

soup.select('p')

Out[11]:

```
[<p class="subtitle is-3">
    Fake Jobs for Your Web Scrapping Journey
    </p>,
<p class="location">
    Stewartbury, AA
    </p>,
<p class="is-small has-text-grey">
<time datetime="2021-04-08">2021-04-08</time>
</p>,
<p class="location">
    Christopherville, AA
    </p>,
<p class="is-small has-text-grey">
<time datetime="2021-04-08">2021-04-08</time>
</p>,
<p class="location">
    Port Ericaburgh, AA
    </n>.
```

In [12]:

```
list=[]
for i in range(0,51):
    list.append(soup.select('p'))
```

In [13]:

```
list
```

Out[13]:

```
[[<p class="subtitle is-3">
    Fake Jobs for Your Web Scrapping Journey
    </p>,
  <p class="location">
    Stewartbury, AA
    </p>,
  <p class="is-small has-text-grey">
    <time datetime="2021-04-08">2021-04-08</time>
    </p>,
  <p class="location">
    Christopherville, AA
    </p>,
  <p class="is-small has-text-grey">
    <time datetime="2021-04-08">2021-04-08</time>
    </p>,
  <p class="location">
    Port Ericaburgh, AA
    </n>.
```

In [21]:

```
results=soup.find(id='ResultsContainer')
```

In [23]:

```
print(results)
```

```
<div class="columns is-multiline" id="ResultsContainer">
  <div class="column is-half">
    <div class="card">
      <div class="card-content">
        <div class="media">
          <div class="media-left">
            <figure class="image is-48x48">
              
            </figure>
          </div>
          <div class="media-content">
            <h2 class="title is-5">Senior Python Developer</h2>
            <h3 class="subtitle is-6 company">Payne, Roberts and Davis</h3>
          </div>
        </div>
        <div class="content">
          <p class="location">
            Stewartbury, AA
            </p>
          </div>
        </div>
      </div>
    </div>
  </div>
</div>
```


In [32]:

```
for i in job_elements:  
    print(i)
```

```
<div class="card-content">  
<div class="media">  
<div class="media-left">  
<figure class="image is-48x48">  
  
</figure>  
</div>  
<div class="media-content">  
<h2 class="title is-5">Senior Python Developer</h2>  
<h3 class="subtitle is-6 company">Payne, Roberts and Davis</h3>  
</div>  
</div>  
<div class="content">  
<p class="location">  
    Stewartbury, AA  
    </p>  
<p class="is-small has-text-grey">  
<time datetime="2021-04-08">2021-04-08</time>  
</p>  
</div>
```

In [39]:

```
for i in job_elements:  
    title=i.find("h3")  
    print(title) # you can do it for anything as h2 class p and many more
```

<h3 class="subtitle is-6 company">Payne, Roberts and Davis</h3>
<h3 class="subtitle is-6 company">Vasquez-Davidson</h3>
<h3 class="subtitle is-6 company">Jackson, Chambers and Levy</h3>
<h3 class="subtitle is-6 company">Savage-Bradley</h3>
<h3 class="subtitle is-6 company">Ramirez Inc</h3>
<h3 class="subtitle is-6 company">Rogers-Yates</h3>
<h3 class="subtitle is-6 company">Kramer-Klein</h3>
<h3 class="subtitle is-6 company">Meyers-Johnson</h3>
<h3 class="subtitle is-6 company">Hughes-Williams</h3>
<h3 class="subtitle is-6 company">Jones, Williams and Villa</h3>
<h3 class="subtitle is-6 company">Garcia PLC</h3>
<h3 class="subtitle is-6 company">Gregory and Sons</h3>
<h3 class="subtitle is-6 company">Clark, Garcia and Sosa</h3>
<h3 class="subtitle is-6 company">Bush PLC</h3>
<h3 class="subtitle is-6 company">Salazar-Meyers</h3>
<h3 class="subtitle is-6 company">Parker, Murphy and Brooks</h3>
<h3 class="subtitle is-6 company">Cruz-Brown</h3>
<h3 class="subtitle is-6 company">Macdonald-Ferguson</h3>
<h3 class="subtitle is-6 company">Williams, Peterson and Rojas</h3>
<h3 class="subtitle is-6 company">Smith and Sons</h3>
<h3 class="subtitle is-6 company">Moss, Duncan and Allen</h3>
<h3 class="subtitle is-6 company">Gomez-Carroll</h3>
<h3 class="subtitle is-6 company">Manning, Welch and Herring</h3>
<h3 class="subtitle is-6 company">Lee, Gutierrez and Brown</h3>
<h3 class="subtitle is-6 company">Davis, Serrano and Cook</h3>
<h3 class="subtitle is-6 company">Smith LLC</h3>
<h3 class="subtitle is-6 company">Thomas Group</h3>
<h3 class="subtitle is-6 company">Silva-King</h3>
<h3 class="subtitle is-6 company">Pierce-Long</h3>
<h3 class="subtitle is-6 company">Walker-Simpson</h3>
<h3 class="subtitle is-6 company">Cooper and Sons</h3>
<h3 class="subtitle is-6 company">Donovan, Gonzalez and Figueroa</h3>
<h3 class="subtitle is-6 company">Morgan, Butler and Bennett</h3>
<h3 class="subtitle is-6 company">Snyder-Lee</h3>
<h3 class="subtitle is-6 company">Harris PLC</h3>
<h3 class="subtitle is-6 company">Washington PLC</h3>
<h3 class="subtitle is-6 company">Brown, Price and Campbell</h3>
<h3 class="subtitle is-6 company">Mcgee PLC</h3>
<h3 class="subtitle is-6 company">Dixon Inc</h3>
<h3 class="subtitle is-6 company">Thompson, Sheppard and Ward</h3>
<h3 class="subtitle is-6 company">Adams-Brewer</h3>
<h3 class="subtitle is-6 company">Schneider-Brady</h3>
<h3 class="subtitle is-6 company">Gonzales-Frank</h3>
<h3 class="subtitle is-6 company">Smith-Wong</h3>
<h3 class="subtitle is-6 company">Pierce-Herrera</h3>
<h3 class="subtitle is-6 company">Aguilar, Rivera and Quinn</h3>
<h3 class="subtitle is-6 company">Lowe, Barnes and Thomas</h3>
<h3 class="subtitle is-6 company">Lewis, Gonzalez and Vasquez</h3>
<h3 class="subtitle is-6 company">Taylor PLC</h3>
<h3 class="subtitle is-6 company">Oliver, Jones and Ramirez</h3>
<h3 class="subtitle is-6 company">Rivera and Sons</h3>
<h3 class="subtitle is-6 company">Garcia PLC</h3>
<h3 class="subtitle is-6 company">Johnson, Wells and Kramer</h3>
<h3 class="subtitle is-6 company">Gonzalez LLC</h3>
<h3 class="subtitle is-6 company">Morgan, White and Macdonald</h3>
<h3 class="subtitle is-6 company">Robinson-Fitzpatrick</h3>
<h3 class="subtitle is-6 company">Waters, Wilson and Hoover</h3>
<h3 class="subtitle is-6 company">Hill LLC</h3>
<h3 class="subtitle is-6 company">Li-Gregory</h3>
<h3 class="subtitle is-6 company">Fisher, Ryan and Coleman</h3>
<h3 class="subtitle is-6 company">Stewart-Alexander</h3>

<h3 class="subtitle is-6 company">Abbott and Sons</h3>
<h3 class="subtitle is-6 company">Bryant, Santana and Davenport</h3>
<h3 class="subtitle is-6 company">Smith PLC</h3>
<h3 class="subtitle is-6 company">Patterson-Singh</h3>
<h3 class="subtitle is-6 company">Martinez-Berry</h3>
<h3 class="subtitle is-6 company">May, Taylor and Fisher</h3>
<h3 class="subtitle is-6 company">Bailey, Owen and Thompson</h3>
<h3 class="subtitle is-6 company">Vasquez Ltd</h3>
<h3 class="subtitle is-6 company">Leblanc LLC</h3>
<h3 class="subtitle is-6 company">Jackson, Ali and Mckee</h3>
<h3 class="subtitle is-6 company">Blankenship, Knight and Powell</h3>
<h3 class="subtitle is-6 company">Patton, Haynes and Jones</h3>
<h3 class="subtitle is-6 company">Wood Inc</h3>
<h3 class="subtitle is-6 company">Collins Group</h3>
<h3 class="subtitle is-6 company">Flores-Nelson</h3>
<h3 class="subtitle is-6 company">Mitchell, Jones and Olson</h3>
<h3 class="subtitle is-6 company">Howard Group</h3>
<h3 class="subtitle is-6 company">Kramer-Edwards</h3>
<h3 class="subtitle is-6 company">Berry-Houston</h3>
<h3 class="subtitle is-6 company">Mathews Inc</h3>
<h3 class="subtitle is-6 company">Riley-Johnson</h3>
<h3 class="subtitle is-6 company">Spencer and Sons</h3>
<h3 class="subtitle is-6 company">Camacho-Sanchez</h3>
<h3 class="subtitle is-6 company">Oliver and Sons</h3>
<h3 class="subtitle is-6 company">Eaton PLC</h3>
<h3 class="subtitle is-6 company">Stanley-Frederick</h3>
<h3 class="subtitle is-6 company">Bradley LLC</h3>
<h3 class="subtitle is-6 company">Parker, Goodwin and Zavala</h3>
<h3 class="subtitle is-6 company">Kim-Miles</h3>
<h3 class="subtitle is-6 company">Moreno-Rodriguez</h3>
<h3 class="subtitle is-6 company">Brown-Ortiz</h3>
<h3 class="subtitle is-6 company">Hartman PLC</h3>
<h3 class="subtitle is-6 company">Brooks Inc</h3>
<h3 class="subtitle is-6 company">Washington-Castillo</h3>
<h3 class="subtitle is-6 company">Nguyen, Yoder and Petty</h3>
<h3 class="subtitle is-6 company">Holder LLC</h3>
<h3 class="subtitle is-6 company">Yates-Ferguson</h3>
<h3 class="subtitle is-6 company">Ortega-Lawrence</h3>
<h3 class="subtitle is-6 company">Fuentes, Walls and Castro</h3>

In [40]:

```
for i in job_elements:  
    title=i.find("h3")  
    print(title.text)
```

Payne, Roberts and Davis
Vasquez-Davidson
Jackson, Chambers and Levy
Savage-Bradley
Ramirez Inc
Rogers-Yates
Kramer-Klein
Meyers-Johnson
Hughes-Williams
Jones, Williams and Villa
Garcia PLC
Gregory and Sons
Clark, Garcia and Sosa
Bush PLC
Salazar-Meyers
Parker, Murphy and Brooks
Cruz-Brown
Macdonald-Ferguson
Williams, Peterson and Rojas
Smith and Sons
Moss, Duncan and Allen
Gomez-Carroll
Manning, Welch and Herring
Lee, Gutierrez and Brown
Davis, Serrano and Cook
Smith LLC
Thomas Group
Silva-King
Pierce-Long
Walker-Simpson
Cooper and Sons
Donovan, Gonzalez and Figueroa
Morgan, Butler and Bennett
Snyder-Lee
Harris PLC
Washington PLC
Brown, Price and Campbell
Mcgee PLC
Dixon Inc
Thompson, Sheppard and Ward
Adams-Brewer
Schneider-Brady
Gonzales-Frank
Smith-Wong
Pierce-Herrera
Aguilar, Rivera and Quinn
Lowe, Barnes and Thomas
Lewis, Gonzalez and Vasquez
Taylor PLC
Oliver, Jones and Ramirez
Rivera and Sons
Garcia PLC
Johnson, Wells and Kramer
Gonzalez LLC
Morgan, White and Macdonald
Robinson-Fitzpatrick
Waters, Wilson and Hoover
Hill LLC
Li-Gregory
Fisher, Ryan and Coleman
Stewart-Alexander

Abbott and Sons
Bryant, Santana and Davenport
Smith PLC
Patterson-Singh
Martinez-Berry
May, Taylor and Fisher
Bailey, Owen and Thompson
Vasquez Ltd
Leblanc LLC
Jackson, Ali and McKee
Blankenship, Knight and Powell
Patton, Haynes and Jones
Wood Inc
Collins Group
Flores-Nelson
Mitchell, Jones and Olson
Howard Group
Kramer-Edwards
Berry-Houston
Mathews Inc
Riley-Johnson
Spencer and Sons
Camacho-Sanchez
Oliver and Sons
Eaton PLC
Stanley-Frederick
Bradley LLC
Parker, Goodwin and Zavala
Kim-Miles
Moreno-Rodriguez
Brown-Ortiz
Hartman PLC
Brooks Inc
Washington-Castillo
Nguyen, Yoder and Petty
Holder LLC
Yates-Ferguson
Ortega-Lawrence
Fuentes, Walls and Castro

In [41]:

```
for i in job_elements:  
    title=i.find("h3")  
    print(title.text.strip())
```

Payne, Roberts and Davis
Vasquez-Davidson
Jackson, Chambers and Levy
Savage-Bradley
Ramirez Inc
Rogers-Yates
Kramer-Klein
Meyers-Johnson
Hughes-Williams
Jones, Williams and Villa
Garcia PLC
Gregory and Sons
Clark, Garcia and Sosa
Bush PLC
Salazar-Meyers
Parker, Murphy and Brooks
Cruz-Brown
Macdonald-Ferguson
Williams, Peterson and Rojas
Smith and Sons
Moss, Duncan and Allen
Gomez-Carroll
Manning, Welch and Herring
Lee, Gutierrez and Brown
Davis, Serrano and Cook
Smith LLC
Thomas Group
Silva-King
Pierce-Long
Walker-Simpson
Cooper and Sons
Donovan, Gonzalez and Figueroa
Morgan, Butler and Bennett
Snyder-Lee
Harris PLC
Washington PLC
Brown, Price and Campbell
Mcgee PLC
Dixon Inc
Thompson, Sheppard and Ward
Adams-Brewer
Schneider-Brady
Gonzales-Frank
Smith-Wong
Pierce-Herrera
Aguilar, Rivera and Quinn
Lowe, Barnes and Thomas
Lewis, Gonzalez and Vasquez
Taylor PLC
Oliver, Jones and Ramirez
Rivera and Sons
Garcia PLC
Johnson, Wells and Kramer
Gonzalez LLC
Morgan, White and Macdonald
Robinson-Fitzpatrick
Waters, Wilson and Hoover
Hill LLC
Li-Gregory
Fisher, Ryan and Coleman
Stewart-Alexander

Abbott and Sons
Bryant, Santana and Davenport
Smith PLC
Patterson-Singh
Martinez-Berry
May, Taylor and Fisher
Bailey, Owen and Thompson
Vasquez Ltd
Leblanc LLC
Jackson, Ali and McKee
Blankenship, Knight and Powell
Patton, Haynes and Jones
Wood Inc
Collins Group
Flores-Nelson
Mitchell, Jones and Olson
Howard Group
Kramer-Edwards
Berry-Houston
Mathews Inc
Riley-Johnson
Spencer and Sons
Camacho-Sanchez
Oliver and Sons
Eaton PLC
Stanley-Frederick
Bradley LLC
Parker, Goodwin and Zavala
Kim-Miles
Moreno-Rodriguez
Brown-Ortiz
Hartman PLC
Brooks Inc
Washington-Castillo
Nguyen, Yoder and Petty
Holder LLC
Yates-Ferguson
Ortega-Lawrence
Fuentes, Walls and Castro

In [44]:

```
python_jobs = results.find_all("h2", string="python")
```

In [45]:

```
python_jobs
```

Out[45]:

```
[]
```

In [46]:

```
python_jobs = results.find_all(  
    "h2", string=lambda text: "python" in text.lower()  
)
```

In [47]:

```
python_jobs
```

Out[47]:

```
[<h2 class="title is-5">Senior Python Developer</h2>,  
<h2 class="title is-5">Software Engineer (Python)</h2>,  
<h2 class="title is-5">Python Programmer (Entry-Level)</h2>,  
<h2 class="title is-5">Python Programmer (Entry-Level)</h2>,  
<h2 class="title is-5">Software Developer (Python)</h2>,  
<h2 class="title is-5">Python Developer</h2>,  
<h2 class="title is-5">Back-End Web Developer (Python, Django)</h2>,  
<h2 class="title is-5">Back-End Web Developer (Python, Django)</h2>,  
<h2 class="title is-5">Python Programmer (Entry-Level)</h2>,  
<h2 class="title is-5">Software Developer (Python)</h2>]
```

In [57]:

```
for i in python_jobs:  
    print(i.text)
```

```
Senior Python Developer  
Software Engineer (Python)  
Python Programmer (Entry-Level)  
Python Programmer (Entry-Level)  
Software Developer (Python)  
Python Developer  
Back-End Web Developer (Python, Django)  
Back-End Web Developer (Python, Django)  
Python Programmer (Entry-Level)  
Software Developer (Python)
```

In []: