

Environment: Ubuntu

If you do not have pip installed on your system. Use the following command in the terminal:

```
$ sudo apt-get install python-pip python-dev build-essential
```

What is pip? - pip is a package management system used to install and manage software packages written in Python. [Source: Wikipedia]

The two python packages that we'll be using to perform web scraping are:

- **BeautifulSoup:**

Python package used for parsing HTML and XML documents.

```
$ pip install beautifulsoup4
```

- **Requests:**

Requests is a python library designed to ping a webpage or portal for information.

```
$ pip install requests
```

EXAMPLE 1: Travel Site

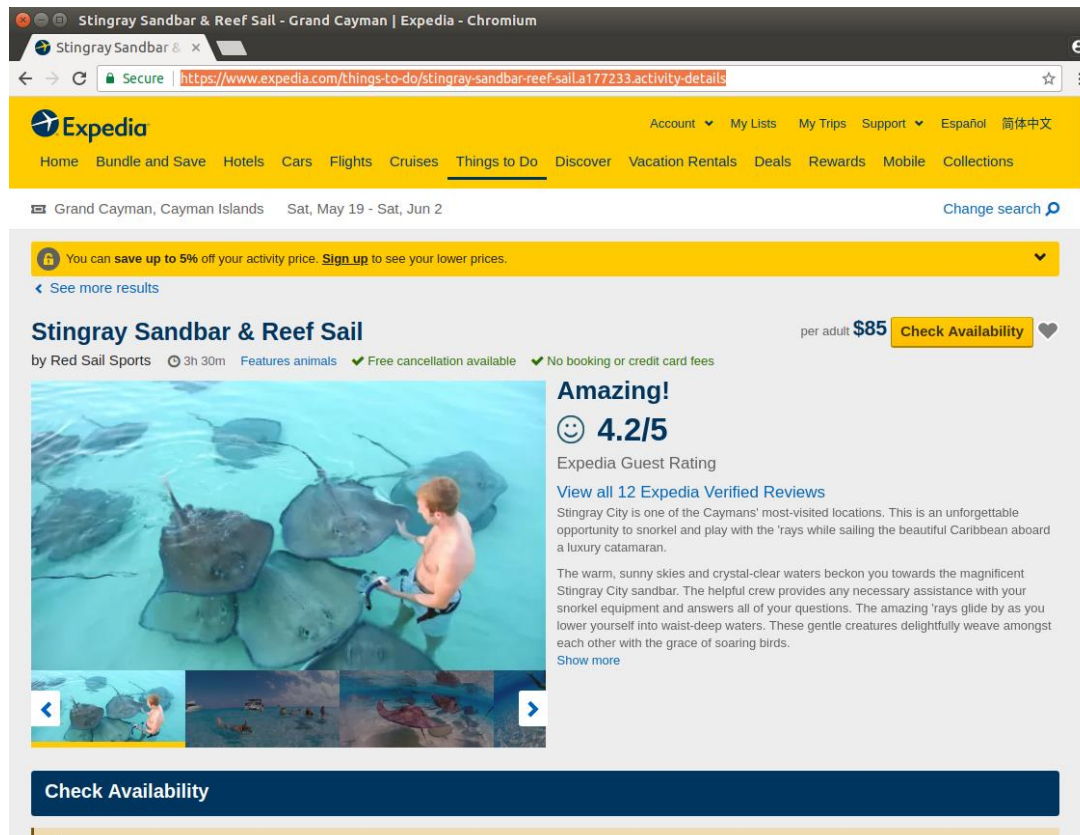
In this example we'll be scraping some sample data from a travel site, Expedia.

```
IPython 2.4.1 -- An enhanced Interactive Python.
?      -> Introduction and overview of IPython's features.
%quickref -> Quick reference.
help    -> Python's own help system.
object? -> Details about 'object', use 'object??' for extra details.

In [1]: import requests

In [2]: from bs4 import BeautifulSoup
```

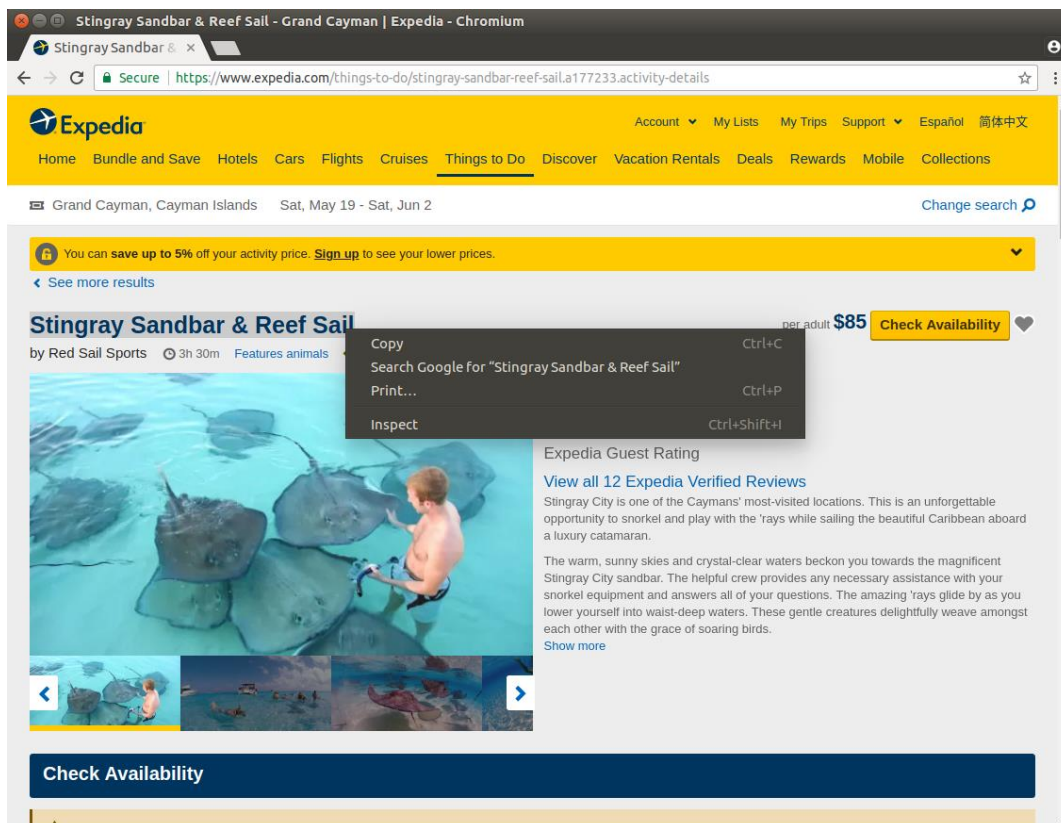
The screen shot below shows a webpage of some popular activity at a destination:



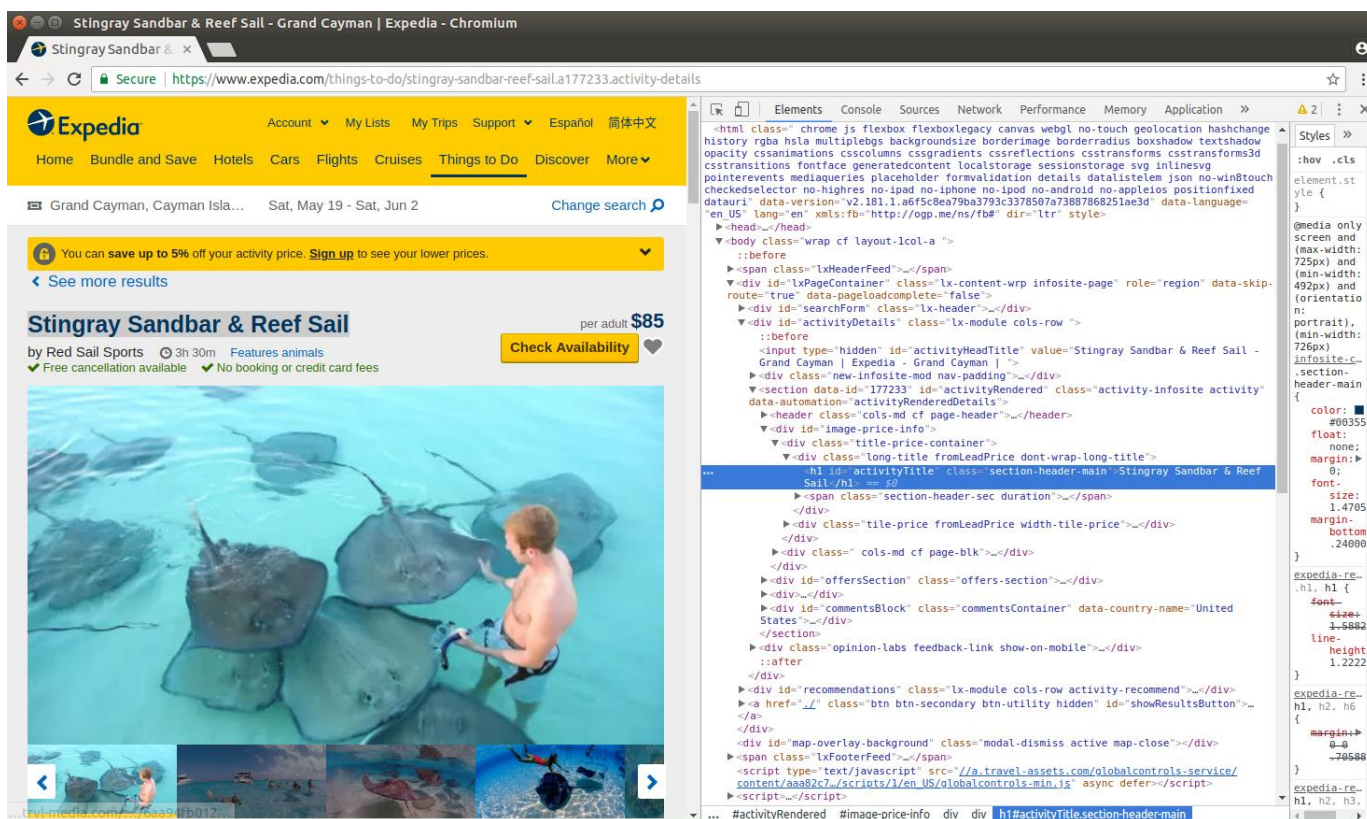
The screenshot displays the Expedia website interface. At the top, there's a yellow navigation bar with the Expedia logo and links for Account, My Lists, My Trips, Support, Español, and 简体中文. Below this is a search bar and a promotional banner stating "You can save up to 5% off your activity price. Sign up to see your lower prices." The main content area features a large image of a person snorkeling with stingrays, a rating of 4.2/5, and a "Check Availability" button. The page also includes a detailed description of the activity and a "Show more" link.

```
In [3]: webpage = requests.get("https://www.expedia.com/things-to-do/stingray-sandbar-reef-sail.a177233.activity-details")
In [4]: print(webpage.status_code)
200
```

A status_code of 200 indicates that the page was downloaded successfully.



Right-click on the main header and click on inspect.



Parse this page using the BeautifulSoup library. Then print out the main heading by searching its class/id tag.

```
In [5]: soup = BeautifulSoup(webpage.content, 'html.parser')

In [6]: mainHeading = soup.find(class_="section-header-main")

In [7]: print(mainHeading.prettify())
<h1 class="section-header-main" id="activityTitle">
  Stingray Sandbar & Reef Sail
</h1>

In [8]: print(mainHeading.get_text())
Stingray Sandbar & Reef Sail
```

The screenshot shows the Expedia website interface. The top navigation bar includes the Expedia logo and links for Account, My Lists, My Trips, Support, Español, and 简体中文. The main search bar displays 'Grand Cayman, Cayman Isla...' and 'Sat, May 19 - Sat, Jun 2'. Below the search bar, a promotional banner offers a 5% discount. The activity 'Stingray Sandbar & Reef Sail' is highlighted, showing it is by Red Sail Sports, lasts 3h 30m, and includes features like free cancellation and no booking or credit card fees. The price is listed as \$85 per adult. A 'Check Availability' button is present. The right side of the image shows the browser's developer tools, displaying the HTML structure of the page, including the main heading and the price information.

Let's try to find the cost of this activity per adult.

```
In [9]: costPerAdult = soup.find(id="activityFromPrice")

In [10]: print(costPerAdult.get_text())
$85
```



EXAMPLE 2: Weather Forecast

Like the previous, this example shows how we can extract some weather forecast data of a specific region:

7-Day Forecast for Latitude 37.77°N and Longitude 122.41°W (Elev. 131 ft) - Chromium

7-Day Forecast for | x

https://forecast.weather.gov/MapClick.php?textField1=37.78&textField2=122.42#WwCblZ-YW4E



NATIONAL WEATHER SERVICE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

HOMEFORECAST▼PAST WEATHER▼SAFETY▼INFORMATION▼EDUCATION▼NEWS▼SEARCH▼ABOUT▼

Go

[View Location Examples](#)

Your local forecast office is
San Francisco Bay Area/Monterey, CA

[En Español](#) [Share](#) [Facebook](#) [Twitter](#) [Email](#) [Pinterest](#)

Current conditions at
SAN FRANCISCO DOWNTOWN (SFOC1)
Lat: 37.77056°N Lon: 122.42694°W Elev: 150ft.

NA

58°F
14°C

Humidity 67%

Wind Speed NA

Barometer NA


Dewpoint 47°F (8°C)


Visibility NA


Last update 19 May 1:43 pm PDT


More Information:
[Local Forecast Office](#)
[More Local Wx](#)
[3 Day History](#)
[Mobile Weather](#)
[Hourly Weather Forecast](#)


Extended Forecast for
San Francisco CA


This Afternoon

Mostly Sunny and Breezy


Tonight

Partly Cloudy and Breezy then Mostly Cloudy


Sunday

Decreasing Clouds and Breezy


Sunday Night

Partly Cloudy and Breezy then Mostly Cloudy

Monday

Mostly Sunny and Breezy

Monday Night

Partly Cloudy and Breezy then Mostly Cloudy

Tuesday

Partly Sunny

Tuesday Night

Mostly Cloudy

Wednesday

Mostly Sunny

7-Day Forecast for Latitude 37.77°N and Longitude 122.41°W (Elev. 131 ft) - Chromium

7-Day Forecast for | x

https://forecast.weather.gov/MapClick.php?textField1=37.78&textField2=122.42#WwCblZ-YW4E

Current conditions at
SAN FRANCISCO DOWNTOWN (SFOC1)
Lat: 37.77056°N Lon: 122.42694°W Elev: 150ft.

NA

58°F
14°C

Humidity 67%

Wind Speed NA

Barometer NA


Dewpoint 47°F (8°C)


Visibility NA


Last update 19 May 1:43 pm PDT


More Information:
[Local Forecast Office](#)
[More Local Wx](#)
[3 Day History](#)
[Mobile Weather](#)
[Hourly Weather Forecast](#)


Extended Forecast for
San Francisco CA


This Afternoon

Mostly Sunny and Breezy
High: 60 °F

Tonight

Partly Cloudy and Breezy then Mostly Cloudy
Low: 53 °F

Sunday

Decreasing Clouds and Breezy
High: 61 °F

Sunday Night

Partly Cloudy and Breezy then Mostly Cloudy
Low: 52 °F

Monday

Mostly Sunny and Breezy
High: 63 °F

Monc Nigl

Partly C and Br then M Clou
Low: 5

Detailed Forecast

This Afternoon
Mostly sunny, with a high near 60. Breezy, with a west wind around 25 mph, with gusts as high as 32 mph.

Tonight
Increasing clouds, with a low around 53. Breezy, with a west wind 21 to 25 mph, with gusts as high as 32 mph.

Sunday
Cloudy, then gradually becoming mostly sunny, with a high near 61. Breezy, with a west wind 18 to 23 mph, with gusts as high as 30 mph.

Sunday Night
Increasing clouds, with a low around 52. Breezy, with a west wind 18 to 23 mph, with gusts as high as 30 mph.

Elements Console Sources Network Performance Memory Application Security Audits

```
<!DOCTYPE html>
<html class="no-js flexbox flexboxlegacy canvas canvastext webgl no-touch geolocation postmessage websqlindexdb hashchange history draganddrop websockets rgba hsla multiplebgs backgroundsize borderimage borderradius boxshadow textshadow opacity cssanimations csscolumns cssgradients cssreflections csstransforms csstransforms3d csstransitions fontface generatedcontent video audio localstorage sessionstorage webworkers applicationcache svg inlinesvg smil svgclippaths" style="
  <!-- Start Forecastsearch -->
  <div class="fcst-search"></div>
  <!-- end Forecastsearch -->
  <link rel="stylesheet" type="text/css" href="/css/topnews.css">
  <div id="news-items"></div>
  <script type="text/javascript"></script>
  <!-- PageFormat-Land -->
  <script language="javascript"></script>
  
  <div id="quicklinks"></div>
  <!-- Current Conditions -->
  <div id="current-conditions" class="panel panel-default">
    <!-- Current Conditions header row -->
    <div class="panel-heading">
      <div>
        <b>Current conditions at</b>
        <h2 class="panel-title">SAN FRANCISCO DOWNTOWN (SFOC1)</h2>
        <span class="smallText"></span>
      </div>
    </div>
    <div class="panel-body" id="current-conditions-body">
      <!-- before -->
      <!-- Graphic and temperatures -->
      <div id="current-conditions-summary" class="pull-left">
        <p class="myforecast-current">NA</p>
        <p class="myforecast-current-lrg">58°F</p>
        <p class="myforecast-current-sm">14°C</p>
      </div>
      <div id="current_conditions_detail" class="pull-left"></div>
      <div id="current_conditions_station"></div>
    </div>
  </div>
  <!-- /current-conditions-body -->
  <!-- after -->
```



```

In [1]: import requests

In [2]: from bs4 import BeautifulSoup

In [3]: webpage = requests.get("https://forecast.weather.gov/MapClick.php?textField1=37.78&textField2=-122.42#.WwCblZ-YW4E")

In [4]: print(webpage.status_code)
200

In [5]: soup = BeautifulSoup(webpage.content, 'html.parser')

In [6]: location = soup.find(class_="panel-title").get_text()

In [7]: print(location)
SAN FRANCISCO DOWNTOWN (SFOC1)

In [8]: forecast = soup.find(id="current_conditions-summary").get_text()

In [9]: print(forecast)

NA
58°F
14°C

```

7-Day Forecast for Latitude 37.77°N and Longitude 122.41°W (Elev. 131 ft) - Chromium

7-Day Forecast for L x

← → C <https://forecast.weather.gov/MapClick.php?textField1=37.78&textField2=-122.42#.WwCyup-YW4F>

NA

58°F

14°C

Humidity 69%

Wind Speed NA

Barometer NA

Dewpoint 48°F (9°C)

Visibility NA

Last update 19 May 3:43 pm PDT

More Information:

[Local Forecast Office](#)

[More Local Wx](#)

[3 Day History](#)

[Mobile Weather](#)

[Hourly Weather Forecast](#)

Extended Forecast for
San Francisco CA

Tonight	Sunday	Sunday Night	Monday	Monday Night	Tuesday	Tuesday Night
Partly Cloudy and Breezy then Mostly Cloudy	Decreasing Clouds and Breezy	Partly Cloudy and Breezy then Mostly Cloudy	Decreasing Clouds and Breezy	Partly Cloudy and Breezy then Mostly Cloudy	Partly Sunny	Mostly Cloudy
Low: 53 °F	High: 61 °F	Low: 52 °F	High: 63 °F	Low: 53 °F	High: 63 °F	Low: 53 °F

Detailed Forecast

Period	Forecast
Tonight	Increasing clouds, with a low around 53. Breezy, with a west wind 21 to 25 mph, with gusts as high as 32 mph.
Sunday	Cloudy, then gradually becoming mostly sunny, with a high near 61. Breezy, with a west wind 18 to 23 mph, with gusts as high as 30 mph.
Sunday Night	Increasing clouds, with a low around 52. Breezy, with a west wind 18 to 23 mph decreasing to 11 to 16 mph after midnight. Winds could gust as high as 30 mph.
Monday	Cloudy, then gradually becoming mostly sunny, with a high near 63. Breezy, with a west wind 7 to 12 mph increasing to 18 to 23 mph in the morning. Winds could gust as high as 30 mph.
Monday Night	Mostly cloudy, with a low around 53. Breezy, with a west southwest wind 16 to 22 mph, with gusts as high as 29 mph.
Tuesday	Partly sunny, with a high near 63.

```

In [13]: extendedForecast = soup.find_all(class_="tombstone-container")
In [14]: tonight = extendedForecast[0]

```

```

In [16]: print(tonight.get_text())

Tonight
Partly Cloudyand Breezythen MostlyCloudyLow: 53 °F

In [17]: period_name = tonight.find(class_="period-name").get_text()

In [18]: short_desc = tonight.find(class_="short-desc").get_text()

In [19]: temp = tonight.find(class_="temp").get_text()

In [20]: print(period_name)
Tonight

In [21]: print(short_desc)
Partly Cloudyand Breezythen MostlyCloudy

In [22]: print(temp)
Low: 53 °F

```

```

In [24]: seven_day = soup.find(id="seven-day-forecast")
In [25]: period_names = [x.get_text() for x in seven_day.select(".tombstone-container .period-name")]
In [26]: short_descs = [y.get_text() for y in seven_day.select(".tombstone-container .short-desc")]
In [27]: temps = [z.get_text() for z in seven_day.select(".tombstone-container .temp")]
In [28]: import pandas as pd
In [29]: thisWeekWeather = pd.DataFrame({"period-name": period_names,
....:                                  "short-desc": short_descs,
....:                                  "temp": temps})
In [30]: thisWeekWeather
Out[30]:

```

	period-name	short-desc	temp
0	Tonight	Partly Cloudyand Breezythen MostlyCloudy	Low: 53 °F
1	Sunday	DecreasingClouds andBreezy	High: 61 °F
2	SundayNight	Partly Cloudyand Breezythen MostlyCloudy	Low: 52 °F
3	Monday	DecreasingClouds andBreezy	High: 63 °F
4	MondayNight	Partly Cloudyand Breezythen MostlyCloudy	Low: 53 °F
5	Tuesday	Partly Sunny	High: 63 °F
6	TuesdayNight	Mostly Cloudy	Low: 54 °F
7	Wednesday	Mostly Sunny	High: 63 °F
8	WednesdayNight	Partly Cloudy	Low: 54 °F

Reference:

- DATAQUEST: Python Web Scraping Tutorial using BeautifulSoup
[<https://www.dataquest.io/blog/web-scraping-tutorial-python/>]
- freeCodeCamp: How to scrape websites with Python and BeautifulSoup
[<https://medium.freecodecamp.org/how-to-scrape-websites-with-python-and-beautifulsoup-5946935d93fe>]