

Lecture 8. Web Page and Crawler

Instructor: Luping Yu

April 23, 2024

Web Page

Before we start writing code, we need to understand a little bit about the structure of a web page.

When we visit a web page, our web browser makes a request to a web server. This request is called a `GET` request, since we're getting files from the server. The server then sends back files that tell our browser how to render the page for us. These files will typically include:

- HTML — the main content of the page.
- CSS — used to add styling to make the page look nicer.
- JS — Javascript files add interactivity to web pages.

After our browser receives all the files, it **renders** the page and displays it to us.

There's a lot that happens behind the scenes to render a page nicely, but we don't need to worry about most of it when we're web scraping. When we perform web scraping, we're interested in the main content of the web page, so we look primarily at the `HTML`.

What is HTML?

- HTML stands for Hyper Text Markup Language
 - HTML is the standard **markup language** for creating Web pages
 - HTML describes the structure of a Web page
 - HTML consists of a series of **elements**
 - HTML elements tell the browser how to display the content
 - HTML elements label pieces of content such as "this is a heading", "this is a paragraph", "this is a link", etc.
-

How to View HTML Source?

View HTML Source Code:

- Right-click in an HTML page and select "View Page Source" (in Chrome) or "View Source" (in Edge), or similar in other browsers. This will open a window containing the HTML source code of the page.

Chrome Developer Tools `F12` :

- More Tools ---> Developer Tools
- Chrome DevTools is a set of web developer tools built directly into the Google Chrome browser.

HTML THE SKELETON

HTML provides names (tags) to describe different types of content (elements) on your website —for example: `<header>`, `<link>`, `<div>`.

This allows your browser understand what it is reading and how to render it. While your browser can render HTML by itself, you can make it dynamic and beautiful using Javascript & CSS. So think of HTML as a skeleton that comes to life with Javascript & CSS.



A Simple HTML Document (a.k.a. "Page")

In [15]: %%HTML

```
<!DOCTYPE html>
<html>

  <head>
    <title>My page title</title>
  </head>

  <body>
    <h1>!</Helloh1>
    <p>My second paragraph.</p>
  </body>

</html>
```

Hello!

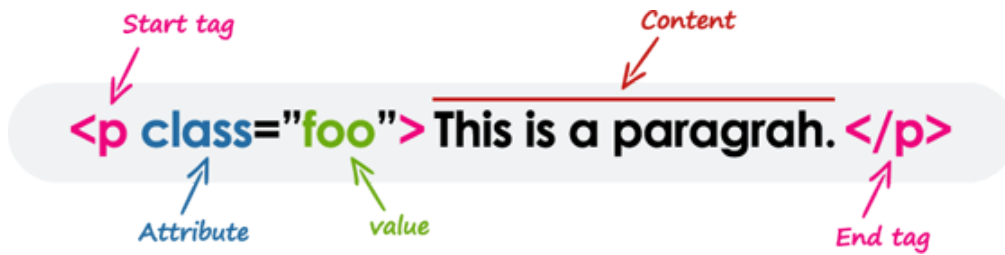
My second paragraph.

Example Explained

- The `<!DOCTYPE html>` declaration defines that this document is an HTML document
- The `<html>` element is the root element of an HTML page
- The `<head>` element contains meta information about the HTML page
- The `<title>` element specifies a title for the HTML page (which is shown in the browser's title bar or in the page's tab)
- The `<body>` element defines the document's body, and is a container for all the **visible** contents, such as headings, paragraphs, images, hyperlinks, tables, lists, etc.
- The `<h1>` element defines a large heading
- The `<p>` element defines a paragraph

HTML Element

An HTML element is defined by a **start tag**, some **content**, and an **end tag**:



The HTML **element** is everything from the start tag to the end tag:

In [2]: %%HTML

```
<h1>My First Heading</h1>
<p>My first paragraph.</p>
```

My First Heading

My first paragraph.

1. Empty HTML Elements

- Empty elements (also called self-closing or void elements) are not container tags.
- A typical example of an empty element, is the `
` element, which represents a line break. Some other common empty elements are ``, `<input>`, etc.

In [10]: %%HTML

```
<p>This paragraph contains <br> a line break.</p>

<input type="text" name="username">
```

This paragraph contains
a line break.



2. Nesting HTML Elements

- Most HTML elements can contain any number of further elements, which are, in turn, made up of tags, attributes, and content or other elements.
- The following example shows some elements nested inside the `<p>` element.

In [4]: %%HTML

```
<p>Here is some <b>bold</b> text.</p>
<p>Here is some <em>emphasized</em> text.</p>
<p>Here is some <mark>highlighted</mark> text.</p>
```

Here is some **bold** text.

Here is some *emphasized* text.

Here is some **highlighted** text.

3. HTML Links

HTML links are defined with the `<a>` tag:

- The link's destination is specified in the `href` attribute.
- `Attributes` are used to provide **additional information** about HTML elements.

In [5]: %%HTML

```
<a href="https://sm.xmu.edu.cn/">This is a link</a>
```

[This is a link](https://sm.xmu.edu.cn/)

4. Writing Comments in HTML

- Comments are usually added with the purpose of making the source code easier to understand.
- You can also comment out part of your HTML code for debugging purpose.
- An HTML comment begins with `<!--`, and ends with `-->`, as shown in the example below:

In [6]: %%HTML

```
<!-- This is an HTML comment -->  
<!-- This is a multi-line HTML comment  
      that spans across more than one line -->  
<p>This is a normal piece of text.</p>
```

This is a normal piece of text.

5. HTML Elements Types

The basic elements of an HTML page are:

- A text header, denoted using the `<h1>`, `<h2>`, `<h3>`, `<h4>`, `<h5>`, `<h6>` tags.
- A paragraph, denoted using the `<p>` tag.
- A link, denoted using the `<a>` (anchor) tag.
- A list, denoted using the `` (unordered list), `` (ordered list) and `` (list element) tags.
- An image, denoted using the `` tag
- A divider, denoted using the `<div>` tag
- A text span, denoted using the `` tag

Elements can be placed in two distinct groups: **block level** and **inline level** elements. The former make up the document's structure, while the latter dress up the contents of a block.

- A block element occupies 100% of the available width and it is rendered with a line break before and after. Whereas, an inline element will take up only as much space as it needs.

- The most commonly used block-level elements are `<div>`, `<p>`, `<h1>` through `<h6>`, `<form>`, ``, ``, ``, and so on. Whereas, the commonly used inline-level elements are ``, `<a>`, ``, ``, ``, ``, `<i>`, `<code>`, `<input>`, `<button>`, etc.
- The block-level elements should not be placed within inline-level elements. For example, the `<p>` element should not be placed inside the `` element.

HTML Attributes

Attributes define **additional characteristics or properties** of the element such as width and height of an image.

- Attributes are always specified in the start tag (or opening tag) and usually consists of name/value pairs like `name="value"`.
 - Some attributes are required for certain elements. For instance, an `` tag must contain a `src` and `alt` attributes.

Let's take a look at some examples of the attributes usages:

```
In [7]: %%HTML
Google</a>
<input type="text" value="Guido van Rossum">
```



Google

In the above example `src` inside the `` tag is an attribute and image path provided is its value. Similarly `href` inside the `<a>` tag is an attribute and the link provided is its value, and so on. There are some **general purpose** attributes, such as `id`, `title`, `class`, `style`, etc. that you can use on the majority of HTML elements.

1.The `id` Attribute

The `id` attribute is used to give a unique identifier to an element within a document. This makes it easier to select the element using CSS or JavaScript.

- The `id` of an element must be **unique** within a single document.
 - No two elements in the same document can be named with the same id.
 - Each element can have only one id.

```
In [11]: %%HTML
<input type="text" id="firstName">
<div id="container">Some content</div>
<p id="infoText">This is a paragraph.</p>
```

Some content

This is a paragraph.

2. The class Attribute

Like `id` attribute, the `class` attribute is also used to identify elements. But unlike `id`, the class attribute **does not have to be unique** in the document.

- This means you can apply the same class to multiple elements in a document.
- Any style rules that are written to that class will be applied to all the elements having that class.

In [12]: %%HTML

```
<input type="text" class="highlight">
<p class="highlight">This is a paragraph.</p>

<p>This is a paragraph.</p>
```

This is a paragraph.

This is a paragraph.

3. The style Attribute

HTML is quite limited when it comes to the presentation of a web page. It was originally designed as a simple way of presenting information.

CSS (Cascading Style Sheets) was introduced in December 1996 by the World Wide Web Consortium (W3C) to provide a better way to style HTML elements.



The `style` attribute allows you to specify CSS styling rules such as color, font, border, etc. directly within the element. Let's check out an example to see how it works:

In [13]: %%HTML

```
<p style="color: blue;">This is a paragraph.</p>
<div style="border: 1px solid red;">Some content</div>
```

This is a paragraph.

Some content

4. HTML Attributes Types

Each element can also have attributes - each element has a different set of attributes relevant to the element.

There are a few global elements, the most common of them are:

- `id` - Denotes the unique ID of an element in a page. Used for locating elements by using links, JavaScript, and more.
- `class` - Denotes the CSS class of an element.
- `style` - Denotes the CSS styles to apply to an element.

Crawler

Some websites offer data sets that are downloadable in CSV format, or accessible via an **Application Programming Interface (API)**. But many websites with useful data don't offer these convenient options.

Web scraping is the process of **gathering information** from the Internet. Even copying and pasting the lyrics of your favorite song is a form of web scraping!

The words "web scraping" usually refer to a process that involves automation. Some websites don't like it when automatic scrapers gather their data, while others don't mind.

Common tools:

- `requests`
- `BeautifulSoup`
- `Selenium`
- `Pandas.read_html()`

How Does Web Scraping Work?

When we scrape the web, we write code that sends a request to the server that's hosting the page we specified. The server will return the source code — HTML, mostly — for the page (or pages) we requested.

So far, we're essentially doing the same thing a web browser does — sending a server request with a specific URL and asking the server to return the code for that page.

But unlike a web browser, our web scraping code won't interpret the page's source code and display the page visually. Instead, we'll write some custom code that filters through the page's source code looking for specific elements we've specified, and extracting whatever content we've instructed it to extract.

For example, if we wanted to get all of the data from inside a table that was displayed on a web page, our code would be written to go through these steps in sequence:

- Request the content (source code) of a specific URL from the server.
- Download the content that is returned.
- Identify the elements of the page that are part of the table we want.
- Extract and (if necessary) reformat those elements into a dataset we can analyze or use in whatever way we require.

Python and `BeautifulSoup` have built-in features designed to make this relatively straightforward.

The requests library

Now that we understand the structure of a web page, it's time to get into the fun part: scraping the content we want!

The first thing we'll need to do to scrape a web page is to download the page. We can download pages using the Python `requests` library.

The requests library will make a `GET` request to a web server, which will download the HTML contents of a given web page for us.

Let's try downloading a simple [sample website](#):

```
In [14]: import requests

page = requests.get("https://dataquestio.github.io/web-scraping-pages/simple.html")

page
```

```
Out [14]: <Response [200]>
```

A `status_code` of 200 means that the page downloaded successfully.

We won't fully dive into status codes here, but a status code starting with a 2 generally indicates success, and a code starting with a 4 or a 5 indicates an error.

We can print out the HTML content of the page using the `content` property:

```
In [12]: page.content
```

```
Out [12]: b'<!DOCTYPE html>\n<html>\n    <head>\n        <title>A simple example page</title>\n    >\n    </head>\n    <body>\n        <p>Here is some simple content for this page.\n    </p>\n    </body>\n</html>'
```

Parsing a page with BeautifulSoup

As you can see above, we now have downloaded an HTML document.

We can use the `BeautifulSoup` library to parse this document, and extract the text from the `p` tag. If we want to extract a single tag, we can use the `find_all` method, which will find all the instances of a tag on a page.

We first have to import the library, and create an instance of the `BeautifulSoup` class to parse our document:

```
In [13]: from bs4 import BeautifulSoup
```



```
soup = BeautifulSoup(page.content, 'html.parser')
soup
```

Out[13]: <!DOCTYPE html>

```
<html>
<head>
<title>A simple example page</title>
</head>
<body>
<p>Here is some simple content for this page.</p>
</body>
</html>
```

In [14]: `soup.find_all('p')`

Out[14]: [`<p>Here is some simple content for this page.</p>`]

Note that `find_all` returns a list, so we'll have to loop through, or use list indexing, it to extract text:

In [15]: `soup.find_all('p')[0].get_text()`

Out[15]: 'Here is some simple content for this page.'

Searching for tags by class and id

We introduced classes and ids earlier, but it probably wasn't clear why they were useful.

`Classes` and `ids` are used by CSS to determine which HTML elements to apply certain styles to. But when we're scraping, we can also use them to specify the elements we want to scrape.

To illustrate this principle, we'll work with another [sample website](#):

In [16]: `page = requests.get("https://dataquestio.github.io/web-scraping-pages/ids_and_class")`
`soup = BeautifulSoup(page.content, 'html.parser')`
`soup`

```

Out[16]: <html>
<head>
<title>A simple example page</title>
</head>
<body>
<div>
<p class="inner-text first-item" id="first">
    First paragraph.
</p>
<p class="inner-text">
    Second paragraph.
</p>
</div>
<p class="outer-text first-item" id="second">
<b>
    First outer paragraph.
</b>
</p>
<p class="outer-text">
<b>
    Second outer paragraph.
</b>
</p>
</body>
</html>

```

Now, we can use the `find_all` method to search for items by class or by id. Let's look for any tag that has the class `outer-text` :

```
In [17]: soup.find_all(class_='outer-text')
```

```

Out[17]: [<p class="outer-text first-item" id="second">
<b>
    First outer paragraph.
</b>
</p>,
<p class="outer-text">
<b>
    Second outer paragraph.
</b>
</p>]

```

We can also search for elements by `id` :

```
In [18]: soup.find_all(id="first")
```

```

Out[18]: [<p class="inner-text first-item" id="first">
    First paragraph.
</p>]

```

Using CSS Selectors

We can also search for items using `CSS` selectors. These selectors are how the CSS language allows developers to specify HTML tags to style. Here are some examples:

- `p a` — finds all a tags inside of a p tag.
- `body p a` — finds all a tags inside of a p tag inside of a body tag.
- `html body` — finds all body tags inside of an html tag.
- `p.outer-text` — finds all p tags with a class of outer-text.
- `p#first` — finds all p tags with an id of first.
- `body p.outer-text` — finds any p tags with a class of outer-text inside of a body tag.

BeautifulSoup objects support searching a page via CSS selectors using the `select` method. We can use CSS selectors to find all the `p` tags in our page that are inside of a `div` like this:

```
In [19]: soup.select("div p")
```

```
Out[19]: [

---


```

Example 1. Downloading weather data

We now know enough to proceed with extracting information about the local weather from the National Weather Service website.

The first step is to find the page we want to scrape. We'll extract weather information about downtown San Francisco from [this page](#). The page has information about the extended forecast for the next week, including time of day, temperature, and a brief description of the conditions.

Specifically, let's extract data about the extended forecast.

1. Exploring page structure with Chrome DevTools

The first thing we'll need to do is inspect the page using `Chrome Devtools`. If you're using another browser, Firefox and Safari have equivalents.

You can start the developer tools in Chrome by clicking `View -> Developer -> Developer Tools`. Make sure the Elements panel is highlighted.

The elements panel will show you all the HTML tags on the page, and let you navigate through them. It's a really handy feature!

We can then scroll up in the elements panel to find the "outermost" element that contains all of the text that corresponds to the extended forecasts. In this case, it's a `div` tag with the id `seven-day-forecast`.

If we click around on the console, and explore the div, we'll discover that each forecast item (like "Tonight", "Thursday", and "Thursday Night") is contained in a `div` with the class `tombstone-container`.

2. Time to Start Scraping!

We now know enough to download the page and start parsing it. In the below code, we will:

- Download the web page containing the forecast.
- Create a BeautifulSoup class to parse the page.
- Find the div with id `seven-day-forecast`, and assign to `seven_day`
- Inside `seven_day`, find each individual forecast item.
- Extract and print the first forecast item.

```
In [20]: page = requests.get("https://forecast.weather.gov/MapClick.php?lat=37.7772&lon=-122")
soup = BeautifulSoup(page.content, 'html.parser')

soup
```

Out[20]: <!DOCTYPE html>

```
<html class="no-js">
<head>
<!-- Meta -->
<meta content="width=device-width" name="viewport"/>
<link href="http://purl.org/dc/elements/1.1/" rel="schema.DC"/><title>National Weather Service</title><meta content="National Weather Service" name="DC.title"><meta content="NOAA National Weather Service National Weather Service" name="DC.description"/><meta content="US Department of Commerce, NOAA, National Weather Service" name="DC.creator"/><meta content="" name="DC.date.created" scheme="ISO8601"/><meta content="EN-US" name="DC.language" scheme="DCTERMS.RFC1766"/><meta content="weather, National Weather Service" name="DC.keywords"/><meta content="NOAA's National Weather Service" name="DC.publisher"/><meta content="National Weather Service" name="DC.contributor"/><meta content="//www.weather.gov/disclaimer.php" name="DC.rights"/><meta content="General" name="rating"/><meta content="index, follow" name="robots"/>
<!-- Icons -->
<link href="/images/favicon.ico" rel="shortcut icon" type="image/x-icon"/>
<!-- CSS -->
<link href="css/bootstrap-3.2.0.min.css" rel="stylesheet"/>
<link href="css/bootstrap-theme-3.2.0.min.css" rel="stylesheet"/>
<link href="css/font-awesome-4.3.0.min.css" rel="stylesheet"/>
<link href="css/ol-4.6.4.css" rel="stylesheet" type="text/css"/>
<link href="css/mapclick.css" rel="stylesheet" type="text/css">
<!--[if lte IE 7]><link rel="stylesheet" type="text/css" href="css/bootstrap-ie7.css" /><![endif]-->
<!--[if lte IE 9]><link rel="stylesheet" type="text/css" href="css/mapclick-ie.css" /><![endif]-->
<link href="css/print.css" rel="stylesheet" type="text/css">
<link href="css/search.css" rel="stylesheet" type="text/css">
<!-- Javascript -->
<script src="js/lib/modernizr-2.8.3.js" type="text/javascript"></script>
<script src="js/lib/json3-3.3.2.min.js" type="text/javascript"></script>
<script src="js/lib/jquery-1.11.3.min.js" type="text/javascript"></script>
<script src="js/lib/jquery.hoverIntent-1.8.1.min.js" type="text/javascript"></script>
<script src="js/lib/bootstrap-3.2.0.min.js" type="text/javascript"></script>
<script src="js/lib/ol-4.6.4.js" type="text/javascript"></script>
<!--[if lte IE 8]><script type="text/javascript" src="js/respond.min.js"></script>
<![endif]-->
<script src="js/jquery.autocomplete.min.js" type="text/javascript"></script>
<script src="js/cfisurvey/cfi.js?v2" type="text/javascript"></script>
<script src="js/forecast.esri.js" type="text/javascript"></script>
<script src="js/forecast.search.js" type="text/javascript"></script>
<script src="js/forecast.openlayers.js" type="text/javascript"></script>
<script src="js/browserSniffer.js" type="text/javascript"></script>
<script id="_fed_an_ua_tag" src="https://dap.digitalgov.gov/Universal-Federated-Analytics-Min.js?agency=DOC&subagency=NOAA" type="text/javascript"></script>
<script type="text/javascript">
    // GoogleAnalyticsObject is defined in the federated analytics script,
    but PUA option not used as forecast UA needs sampleRate
    window[window['GoogleAnalyticsObject']]('create', 'UA-40768555-1', 'weather.gov', {'sampleRate': 6});
    window[window['GoogleAnalyticsObject']]('set', 'anonymizeIp', true);
    window[window['GoogleAnalyticsObject']]('require', 'linkid');
    window[window['GoogleAnalyticsObject']]('send', 'pageview');
</script>
<script type="javascript">
// ForeSee Staging Embed Script v2.01
// DO NOT MODIFY BELOW THIS LINE *****
;(function (g) {
    var d = document, am = d.createElement('script'), h = d.head || d.getElementsByTagName("head")[0], fsr = 'fsReady',
        aex = {
            "src": "//gateway.foresee.com/sites/weather-gov/production/gateway.min.js",
```

```

        "type": "text/javascript",
        "async": "true",
        "data-vendor": "fs",
        "data-role": "gateway"
    };
    for (var attr in aex) { am.setAttribute(attr, aex[attr]); } h.appendChild(am); g
[fsr] || (g[fsr] = function () { var aT = '__' + fsr + '_stk__'; g[aT] = g[aT] ||
[]; g[aT].push(arguments); });
})(window);
// DO NOT MODIFY ABOVE THIS LINE *****
</script>
</link></link></link></meta></head>
<body>
<main class="container">
<header class="row clearfix" id="page-header">
<a class="pull-left" href="//www.noaa.gov" id="header-noaa"></a>
<a class="pull-left" href="//www.weather.gov" id="header-ews"></a>
<a class="pull-right" href="//www.commerce.gov" id="header-doc"></a>
</header>
<nav class="navbar navbar-default row" role="navigation">
<div class="container-fluid">
<div class="navbar-header">
<button class="navbar-toggle collapsed" data-target="#top-nav" data-toggle="collap
se" type="button">
<span class="sr-only">Toggle navigation</span>
<span class="icon-bar"></span>
<span class="icon-bar"></span>
<span class="icon-bar"></span>
</button>
</div>
<div class="collapse navbar-collapse" id="top-nav">
<ul class="nav navbar-nav">
<li><a href="//www.weather.gov">HOME</a></li>
<li class="dropdown"><a class="dropdown-toggle" data-toggle="dropdown" href="htt
p://www.weather.gov/forecastmaps">FORECAST <span class="caret"></span></a><ul clas
s="dropdown-menu" role="menu"><li><a href="http://www.weather.gov">Local</a></li><
li><a href="http://digital.weather.gov">Graphical</a></li><li><a href="http://www.
aviationweather.gov/">Aviation</a></li><li><a href="http://www.weather.gov/marin
e">Marine</a></li><li><a href="http://water.noaa.gov">Rivers and Lakes</a></li><li
><a href="http://www.nhc.noaa.gov/">Hurricanes</a></li><li><a href="http://www.sp
c.noaa.gov/">Severe Weather</a></li><li><a href="http://www.weather.gov/fire">Fir
e Weather</a></li><li><a href="https://www.esrl.noaa.gov/gmd/grad/solcalc/sunrise.
html">Sun/Moon</a></li><li><a href="http://www.cpc.ncep.noaa.gov/">Long Range Fore
casts</a></li><li><a href="http://www.cpc.ncep.noaa.gov">Climate Prediction</a></li
><li><a href="https://www.swpc.noaa.gov/">Space Weather</a></li></ul> </li>
<li class="dropdown"><a class="dropdown-toggle" data-toggle="dropdown" href="http
s://www.weather.gov/wrh/climate/">PAST WEATHER <span class="caret"></span></a><ul
class="dropdown-menu" role="menu"><li><a href="https://www.weather.gov/wrh/climat
e/">Past Weather</a></li><li><a href="https://www.weather.gov/wrh/climate/">Heatin
g/Cooling Days</a></li><li><a href="https://www.weather.gov/wrh/climate/">Monthly
Temperatures</a></li><li><a href="https://www.weather.gov/wrh/climate/">Records</a
></li><li><a href="http://aa.usno.navy.mil/">Astronomical Data</a></li></ul> </li>
<li class="dropdown"><a class="dropdown-toggle" data-toggle="dropdown" href="htt
p://www.weather.gov/safety">SAFETY <span class="caret"></span></a><ul class="drowp
down-menu" role="menu"><li><a href="https://www.weather.gov/safety/flood">Floods</a
></li><li><a href="https://www.weather.gov/safety/tsunami">Tsunamis</a></li><li><a
href="https://www.weather.gov/safety/beachhazards">Beach Hazards</a></li><li><a hr
ef="https://www.weather.gov/safety/wildfire">Wildfire</a></li><li><a href="http
s://www.weather.gov/safety/cold">Cold</a></li><li><a href="https://www.weather.go
v/safety/tornado">Tornadoes</a></li><li><a href="https://www.weather.gov/safety/fo
g">Fog</a></li><li><a href="https://www.weather.gov/safety/airquality">Air Quality
</a></li><li><a href="https://www.weather.gov/safety/heat">Heat</a></li><li><a hre
f="https://www.weather.gov/safety/hurricane">Hurricanes</a></li><li><a href="http
s://www.weather.gov/safety/lightning">Lightning</a></li><li><a href="https://www.w

```

[eather.gov/safety/safeboating](https://www.weather.gov/safety/safeboating)>Safe BoatingRip CurrentsThunderstormsSpace WeatherSun (Ultraviolet Radiation)Safety CampaignsWindDroughtWinter Weather

<li class="dropdown">INFORMATION <ul class="dropdown-menu" role="menu">Wireless Emergency AlertsBrochuresWeather-Ready NationCooperative ObserversDaily BriefingDamage/Fatality/Injury StatisticsForecast ModelsGIS Data PortalNOAA Weather RadioPublicationsSKYWARN Storm SpottersStormReadyTsunamiReadyService Change Notices

<li class="dropdown">EDUCATION <ul class="dropdown-menu" role="menu">Be A Force of NatureNWS Education Home

<li class="dropdown">NEWS <ul class="dropdown-menu" role="menu">NWS NewsEventsPubs/Brochures/Booklets NWS Media Contacts

<li class="dropdown">SEARCH <ul class="dropdown-menu" role="menu"> <!-- Begin search code -->
<div id="site-search">
<form action="//search.usa.gov/search" method="get" style="margin-bottom: 0; margin-top: 0;">
<input name="v:project" type="hidden" value="firstgov"/>
<label for="query">Search For</label>
<input id="query" name="query" size="12" type="text"/>
<input type="submit" value="Go"/>
<p>
<input checked="checked" id="nws" name="affiliate" type="radio" value="nws.noaa.gov"/>
<label class="search-scope" for="nws">NWS</label>
<input id="noaa" name="affiliate" type="radio" value="noaa.gov"/>
<label class="search-scope" for="noaa">All NOAA</label>
</p>
</form>
</div>

<li class="dropdown">ABOUT <ul class="dropdown-menu" role="menu">About NWSOrganizationStrategic PlanCommitment to DiversityFor NWS EmployeesInternationalNational Centers<a href

```

="http://www.weather.gov/careers/">Careers</a></li><li><a href="http://www.weather.gov/contact">Contact Us</a></li><li><a href="https://w1.weather.gov/glossary">Glossary</a></li><li><a href="https://weather.gov/socialmedia">Social Media</a></li>
</ul> </li>
</ul>
</div>
</div>
</nav>
<div class="contentArea">
<!-- Start Forecastsearch -->
<div class="" id="fcst-search">
<form action="https://forecast.weather.gov/zipcity.php" class="form-inline" id="getForecast" method="get" name="getForecast" role="form">
<div id="getfcst-body">
<input class="form-control" id="inputstring" name="inputstring" placeholder="" type="text"/>
<input class="btn btn-default" id="btnSearch" name="btnSearch" type="submit" value="Go"/>
<div id="txtHelp"><a href="javascript:void(window.open('http://weather.gov/ForecastSearchHelp.html','locsearchhelp','status=0,toolbar=0,location=0,menubar=0,directories=0,resizable=1,scrollbars=1,height=500,width=530')).focus()";>View Location Examples</a></div>
</div>
<div id="txtError">
<div id="errorNoResults" style="display:none;">Sorry, the location you searched for was not found. Please try another search.</div>
<div id="errorMultipleResults" style="display:none;">Multiple locations were found. Please select one of the following:</div>
<div id="errorChoices" style="display:none;"></div>
<input id="btnCloseError" style="display:none" type="button" value="Close"/>
</div>
<div id="getfcst-head">
<p>Your local forecast office is</p>
<h3 id="getfcst-headOffice"></h3>
</div>
</form>
</div>
<!-- end Forecastsearch -->
<link href="/css/topnews.css" rel="stylesheet" type="text/css"/>
<div id="news-items">
<div id="topnews">
<div class="icon"></div>
<div class="body">
<h1 style="font-size: 11pt;">Critical Fire Weather Conditions Continue Across the Plains </h1>
<p>
Locally critical fire weather conditions continue for portions of the Northern/Central Plains and much of northern lower Michigan Monday evening. Red Flag Warnings are currently in effect. Dry conditions and gusty winds will persist across southern Colorado Tuesday.

<a href="http://www.wpc.ncep.noaa.gov/discussions/hpcdiscussions.php?disc=pmdspd" target="_blank">Read More &gt;</a>
</p>
</div>
</div>
</div>
<script type="text/javascript">(function ($) { var topnews = $("#topnews"); topnews.hide(); $.get("siteNews.php", {a:"mtr"},function(response){ if (response !== "false") topnews.replaceWith($(response)); topnews.show(); }); })(jQuery);</script><!-- PageFormat-Land -->
<script language="javascript">document.title = '7-Day Forecast 37.77N 122.41W';</script>
<div id="quickLinks">
<span class="lang-spanish"><a href="//forecast.weather.gov/MapClick.php?lat=37.772&lon=-122.4168&lg=sp">En Español</a></span>
</div>

```



```

<!-- Current Conditions -->
<div class="panel panel-default" id="current-conditions">
<!-- 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="smallTxt"><b>Lat: </b>37.77056°N<b>Lon: </b>122.42694°W<b>Elev: </b>1
50.0ft.</span>
</div>
</div>
<div class="panel-body" id="current-conditions-body">
<!-- Graphic and temperatures -->
<div class="pull-left" id="current_conditions_summary">
<p class="myforecast-current">NA</p>
<p class="myforecast-current-lrg">55°F</p>
<p class="myforecast-current-sm">13°C</p>
</div>
<div class="pull-left" id="current_conditions_detail">
<table>
<tr>
<td class="text-right"><b>Humidity</b></td>
<td>76%</td>
</tr>
<tr>
<td class="text-right"><b>Wind Speed</b></td>
<td>NA NA MPH</td>
</tr>
<tr>
<td class="text-right"><b>Barometer</b></td>
<td>NA</td>
</tr>
<tr>
<td class="text-right"><b>Dewpoint</b></td>
<td>47°F (8°C)</td>
</tr>
<tr>
<td class="text-right"><b>Visibility</b></td>
<td>NA</td>
</tr>
<tr>
<td class="text-right"><b>Last update</b></td>
<td>
                22 Apr 08:43 PM PDT
            </td>
</tr>
</table>
</div>
<div id="current_conditions_station">
<div class="current-conditions-extra">
<!-- Right hand section -->
<p class="moreInfo"><b>More Information:</b></p><p><a href="https://www.weather.gov/mtr" id="localWFO" title="San Francisco Bay Area/Monterey, CA"><span class="hideText">Local</span> Forecast Office</a><a href="https://www.weather.gov/wrh/LocalWeather?zone=CAZ006" id="moreWx">More Local Wx</a><a href="https://www.wrh.noaa.gov/mesowest/getobext.php?wfo=mtr&sid=SFOC1&num=72&raw=0">3 Day History</a><a href="//mobile.weather.gov/index.php?lat=37.7772&lon=-122.4168&unit=0&lg=english" id="mobileWxLink">Mobile Weather</a><a href="MapClick.php?lat=37.7772&lon=-122.4168&unit=0&lg=english&FcstType=graphical" id="wxGraph">Hourly <span class="hideText">Weather </span>Forecast</a></p> </div>
<!-- /current_conditions_station -->
</div>
<!-- /current-conditions-body -->
</div>
<!-- /Current Conditions -->
</div>
<!-- 7-Day Forecast -->
<div class="panel panel-default" id="seven-day-forecast">

```

```
<div class="panel-heading">
<b>Extended Forecast for</b>
<h2 class="panel-title">
    San Francisco CA    </h2>
</div>
<div class="panel-body" id="seven-day-forecast-body">
<div id="seven-day-forecast-container"><ul class="list-unstyled" id="seven-day-forecast-list"><li class="forecast-tombstone">
<div class="tombstone-container">
<p class="period-name">Tonight<br/><br/></p>
<p></p>
<p class="short-desc">Mostly Cloudy</p><p class="temp temp-low">Low: 53 °F</p></div></li><li class="forecast-tombstone">
<div class="tombstone-container">
<p class="period-name">Tuesday<br/><br/></p>
<p></p><p class="short-desc">Decreasing<br/>Clouds</p><p class="temp temp-high">High: 64 °F</p></div></li><li class="forecast-tombstone">
<div class="tombstone-container">
<p class="period-name">Tuesday<br/>Night</p>
<p></p><p class="short-desc">Mostly Cloudy</p><p class="temp temp-low">Low: 54 °F</p></div></li><li class="forecast-tombstone">
<div class="tombstone-container">
<p class="period-name">Wednesday<br/><br/></p>
<p></p><p class="short-desc">Partly Sunny</p><p class="temp temp-high">High: 64 °F</p></div></li><li class="forecast-tombstone">
<div class="tombstone-container">
<p class="period-name">Wednesday<br/>Night</p>
<p></p><p class="short-desc">Partly Cloudy</p><p class="temp temp-low">Low: 52 °F</p></div></li><li class="forecast-tombstone">
<div class="tombstone-container">
<p class="period-name">Thursday<br/><br/></p>
<p></p><p class="short-desc">Partly Sunny</p><p class="temp temp-high">High: 63 °F</p></div></li><li class="forecast-tombstone">
<div class="tombstone-container">
<p class="period-name">Thursday<br/>Night</p>
<p></p><p class="short-desc">Mostly Cloudy</p><p class="temp temp-low">Low: 54 °F</p></div></li><li class="forecast-tombstone">
<div class="tombstone-container">
<p class="period-name">Friday<br/><br/></p>
<p></p><p class="short-desc">Mostly Sunny</p><p class="temp temp-high">High: 64 °F</p></div></li><li class="forecast-tombstone">
<div class="tombstone-container">
<p class="period-name">Friday<br/>Night</p>
<p></p><p class="short-desc">Mostly Clear</p><p class="temp temp-low">Lo
```

```

w: 51 °F</p></div></li></ul></div>
<script type="text/javascript">
// equalize forecast heights
$(function () {
    var maxh = 0;
    $(".forecast-tombstone .short-desc").each(function () {
        var h = $(this).height();
        if (h > maxh) { maxh = h; }
    });
    $(".forecast-tombstone .short-desc").height(maxh);
});
</script> </div>
</div>
<!-- Everything between 7-Day Forecast and Footer goes in this row -->
<div class="row" id="floatingDivs">
<!-- Everything on the left-hand side -->
<div class="col-md-7 col-lg-8">
<!-- Detailed Forecast -->
<div class="panel panel-default" id="detailed-forecast">
<div class="panel-heading">
<h2 class="panel-title">Detailed Forecast</h2>
</div>
<div class="panel-body" id="detailed-forecast-body">
<div class="row row-odd row-forecast"><div class="col-sm-2 forecast-label"><b>Toni
ght</b></div><div class="col-sm-10 forecast-text">Mostly cloudy, with a low around
53. South southwest wind around 10 mph. </div></div><div class="row row-even row-f
orecast"><div class="col-sm-2 forecast-label"><b>Tuesday</b></div><div class="col-
sm-10 forecast-text">Mostly cloudy, then gradually becoming sunny, with a high nea
r 64. South southwest wind 8 to 14 mph, with gusts as high as 20 mph. </div></div>
<div class="row row-odd row-forecast"><div class="col-sm-2 forecast-label"><b>Tues
day Night</b></div><div class="col-sm-10 forecast-text">Mostly cloudy, with a low
around 54. South southwest wind 6 to 10 mph. </div></div><div class="row row-even
row-forecast"><div class="col-sm-2 forecast-label"><b>Wednesday</b></div><div clas
s="col-sm-10 forecast-text">Partly sunny, with a high near 64. South wind 6 to 10
mph. </div></div><div class="row row-odd row-forecast"><div class="col-sm-2 foreca
st-label"><b>Wednesday Night</b></div><div class="col-sm-10 forecast-text">Partly
cloudy, with a low around 52. West wind 7 to 10 mph. </div></div><div class="row r
ow-even row-forecast"><div class="col-sm-2 forecast-label"><b>Thursday</b></div><d
iv class="col-sm-10 forecast-text">Partly sunny, with a high near 63.</div></div><
div class="row row-odd row-forecast"><div class="col-sm-2 forecast-label"><b>Thurs
day Night</b></div><div class="col-sm-10 forecast-text">Mostly cloudy, with a low
around 54.</div></div><div class="row row-even row-forecast"><div class="col-sm-2
forecast-label"><b>Friday</b></div><div class="col-sm-10 forecast-text">Mostly sun
ny, with a high near 64.</div></div><div class="row row-odd row-forecast"><div cla
ss="col-sm-2 forecast-label"><b>Friday Night</b></div><div class="col-sm-10 foreca
st-text">Mostly clear, with a low around 51.</div></div><div class="row row-even r
ow-forecast"><div class="col-sm-2 forecast-label"><b>Saturday</b></div><div class
="col-sm-10 forecast-text">Sunny, with a high near 64.</div></div><div class="row
row-odd row-forecast"><div class="col-sm-2 forecast-label"><b>Saturday Night</b></
div><div class="col-sm-10 forecast-text">Mostly clear, with a low around 51.</div>
</div><div class="row row-even row-forecast"><div class="col-sm-2 forecast-label">
<b>Sunday</b></div><div class="col-sm-10 forecast-text">Sunny, with a high near 6
5.</div></div><div class="row row-odd row-forecast"><div class="col-sm-2 forecast-
label"><b>Sunday Night</b></div><div class="col-sm-10 forecast-text">Mostly clear,
with a low around 52.</div></div><div class="row row-even row-forecast"><div class
="col-sm-2 forecast-label"><b>Monday</b></div><div class="col-sm-10 forecast-tex
t">Sunny, with a high near 65.</div></div> </div>
</div>
<!-- /Detailed Forecast -->
<!-- Additional Forecasts and Information -->
<div class="panel panel-default" id="additional_forecasts">
<div class="panel-heading">
<h2 class="panel-title">Additional Forecasts and Information</h2>
</div>
<div class="panel-body" id="additional-forecasts-body">
<p class="myforecast-location"><a href="MapClick.php?zoneid=CAZ006">Zone Area Fore
cast for San Francisco County, CA</a></p>

```

```
<!-- First nine-ten links -->
<div id="linkBlockContainer">
<div class="linkBlock">
<ul class="list-unstyled">
<li><a href="//forecast.weather.gov/product.php?site=MTR&issuedby=MTR&product=AFD&format=CI&version=1&glossary=1">Forecast Discussion</a></li>
<li><a href="MapClick.php?lat=37.7772&lon=-122.4168&unit=0&lg=english&FcstType=text&TextType=2">Printable Forecast</a></li>
<li><a href="MapClick.php?lat=37.7772&lon=-122.4168&unit=0&lg=english&FcstType=text&TextType=1">Text Only Forecast</a></li>
</ul>
</div>
<div class="linkBlock">
<ul class="list-unstyled">
<li><a href="MapClick.php?lat=37.7772&lon=-122.4168&unit=0&lg=english&FcstType=graphical">Hourly Weather Forecast</a></li>
<li><a href="MapClick.php?lat=37.7772&lon=-122.4168&unit=0&lg=english&FcstType=digital">Tabular Forecast</a></li>
<!-- <li><a href="afm/PointClick.php?lat=37.7772&lon=-122.4168">Quick Forecast</a>
</li> -->
</ul>
</div>
<div class="linkBlock">
<ul class="list-unstyled">
<li><a href="//weather.gov/air/probe_aq_data.php?latitude=37.7772&longitude=-122.4168">Air Quality Forecasts</a></li>
<li><a href="MapClick.php?lat=37.7772&lon=-122.4168&FcstType=text&unit=1&lg=en">International System of Units</a></li>
<li><a href="//www.wrh.noaa.gov/forecast/wxtables/index.php?lat=37.7772&lon=-122.4168">Forecast Weather Table Interface</a></li>
</ul>
</div>
<!-- /First nine-ten links -->
<!-- Additional links -->
<div class="linkBlock"><ul class="list-unstyled"><li><a href="http://www.wrh.noaa.gov/mtr/versprod.php?pil=RR8&sid=RSA" target="_self">Hourly River Stages</a></li><li><a href="http://www.wrh.noaa.gov/mtr/versprod.php?pil=RR5&sid=RSA" target="_self">Hourly Rainfall</a></li></ul></div><div class="linkBlock"><ul class="list-unstyled"><li><a href="http://www.wrh.noaa.gov/mtr/wxlinks.php" target="_self">NWS Office Map</a></li><li><a href="http://www.wrh.noaa.gov/mtr/wx_calculator.php" target="_self">Weather Calculator</a></li></ul></div><div class="linkBlock"><ul class="list-unstyled"><li><a href="http://www.nws.noaa.gov/wtf/udaf/area/?site=mtr" target="_self">User Defined Area</a></li></ul></div>
</div> <!-- /linkBlockContainer -->
</div><!-- /additional-forecasts-body-->
</div> <!-- /additional_forecasts -->
</div> <!-- /Everything on the left-hand side -->
<!-- right-side-data -->
<div class="col-md-5 col-lg-4" id="right-side-data">
<div id="mapAndDescriptionArea">
<!-- openlayer map -->
<style>
#custom-search{
display: block;
position: relative;
z-index: 50;
top: 52px;
left: 60px;
}
#esri-geocoder-search{
display: block;
position: relative;
z-index: 50;
top: 52px;
left: 60px;
}
#emap{
```

```

margin-top:15px;
cursor:pointer;
height:370px;
width:100%;
border: 1px solid #ccc;
border-radius: 3px;
}
#switch-basemap-container{
}
#basemap-selection-form ul{
list-style: none;
margin: 0px;
}
#basemap-selection-form li{
float: left;
}
.disclaimer{
margin-top:350px;
margin-left: 5px;
z-index: 100;
position: absolute;
text-transform: none;
}
.esriAttributionLastItem{
text-transform: none;
}
.esriSimpleSlider div{
height:22px;
line-height:20px;
width:20px;
}
#point-forecast-map-label {
text-align:center;
font-weight:bold;
color:black;
}
@media (max-width: 767px) {
#emap{
margin-top:.5em;
height:270px;
}
.disclaimer{
margin-top:250px;
}
}
</style>
<!-- forecast-map -->
<div class="point-forecast-map">
<div class="point-forecast-map-header text-center">
<div id="toolbar">
<div id="switch-basemap-container">
<div id="basemap-selection-form" title="Choose a Basemap">
<div id="basemap-menu">
<select autocomplete="off" id="basemap-selected" name="basemap-selected" title="Basemap Dropdown Menu">
<option value="none">Select Basemap</option>
<option selected="" value="topo">Topographic</option>
<option value="streets">Streets</option>
<option value="satellite">Satellite</option>
<option value="ocean">Ocean</option>
</select>
</div>
</div>
<div id="point-forecast-map-label">
Click Map For Forecast
</div>
</div><!-- //switch-basemap-container -->

```

```
<div style="clear:both;"></div>
</div><!-- // #toolbar -->
</div><!-- //.point-forecast-map-header -->
<div id="emap">
<noscript><center><br/><br/><b>Map function requires Javascript and a compatible browser.</b></center></noscript>
<div class="disclaimer"><a href="http://www.weather.gov/disclaimer#esri">Disclaimer</a></div>
</div><!-- // #emap -->
<div class="point-forecast-map-footer">

</div><!-- //.point-forecast-map-footer -->
</div> <!-- //.point-forecast-map -->
<!-- //forecast-map -->
<!-- //openlayer map -->
<!-- About this Forecast -->
<div id="about_forecast">
<div class="fullRow">
<div class="left">Point Forecast:</div>
<div class="right">San Francisco CA<br/> 37.77°N 122.41°W (Elev. 131 ft)</div>
</div>
<div class="fullRow">
<div class="left"><a href="//www.weather.gov/glossary/index.php?word=Last+update" target="_blank">Last Update</a>: </div>
<div class="right">5:53 pm PDT Apr 22, 2024</div>
</div>
<div class="fullRow">
<div class="left"><a href="//www.weather.gov/glossary/index.php?word=forecast+valid+for" target="_blank">Forecast Valid</a>: </div>
<div class="right">9pm PDT Apr 22, 2024-6pm PDT Apr 29, 2024</div>
</div>
<div class="fullRow">
<div class="left"> </div>
<div class="right"><a href="https://forecast.weather.gov/product.php?site=MTR&issuedby=MTR&product=AFD&format=CI&version=1&glossary=1">Forecast Discussion</a></div>
</div>
<div class="fullRow">
<div class="left"> </div>
<div class="right">
<a href="MapClick.php?lat=37.7772&lon=-122.4168&unit=0&lg=english&FcstType=kml"></a>
<a href="MapClick.php?lat=37.7772&lon=-122.4168&unit=0&lg=english&FcstType=dwml"></a>
</div>
</div>
</div>
<!-- /About this Forecast -->
</div>
<!--additionalForecast-->
<div class="panel panel-default" id="additionalForecast">
<div class="panel-heading">
<h2 class="panel-title">Additional Resources</h2>
</div>
<div class="panel-body">
<!-- Radar & Satellite Images -->
<div class="subItem" id="radar">
<h4>Radar & Satellite Image</h4>
<a href="http://radar.weather.gov/station/kmux/standard"></a> <a href="https://www.star.nesdis.noaa.gov/GOES/sector.php?sat=G17&or=psw"></a> </div>
</div>
</div>
</div>
```

```

<!-- /Radar & Satellite Images -->
<!-- Hourly Weather Forecast -->
<div class="subItem" id="feature">
<h4>Hourly Weather Forecast</h4>
<a href="MapClick.php?lat=37.7772&amp;lon=-122.4168&amp;unit=0&amp;lg=english&amp;
FcstType=graphical"></a>
</div>
<!-- /Hourly Weather Forecast -->
<!-- NDFD -->
<div class="subItem" id="NDFD">
<h4>National Digital Forecast Database</h4>
<div class="one-sixth-first"><a href="//graphical.weather.gov/sectors/pacsouthwes
t.php?element=MaxT"></a>
<p><a href="//graphical.weather.gov/sectors/pacsouthwest.php?element=MaxT">High Te
mperature</a></p></div><div class="one-sixth-first"><a href="//graphical.weather.g
ov/sectors/pacsouthwest.php?element=Wx"></a>
<p><a href="//graphical.weather.gov/sectors/pacsouthwest.php?element=Wx">Chance of
Precipitation</a></p></div> </div>
<!-- /NDFD -->
</div>
</div>
<!-- /additionalForecast -->
</div>
<!-- /col-md-4 -->
<!-- /right-side-data -->
<script language="javascript">$( document ).ready(function() { load_openlayers_map
(' ', ' ', ' ', '{"centroid_lat":"37.7772","centroid_lon":"-122.4168","lat1":"37.759
5","lon1":"-122.4205","lat2":"37.781","lon2":"-122.426","lat3":"37.7855","lon3":"-
122.3985","lat4":"37.764","lon4":"-122.393"}') });</script></div>
<!-- /row -->
</img></div>
<!-- /PageFormat-Land -->

<footer>
<div class="sitemap-content row" id="sitemap">
<div class="col-xs-12">
<div class="sitemap-columns">
<div class="sitemap-section">
<div class="panel-heading">
<a class="sitemap-section-heading" href="http://alerts.weather.gov">ACTIVE ALERTS
</a>
<button class="menu-toggle pull-right" data-target="#sitemap-1" data-toggle="colla
pse" type="button">
<span class="sr-only">Toggle menu</span>
<span class="icon-bar"></span>
<span class="icon-bar"></span>
<span class="icon-bar"></span>
</button>
</div>
<div class="sitemap-section-body panel-body collapsable collapse" id="sitemap-1">
<ul class="list-unstyled">
<li><a href=" http://alerts.weather.gov">Warnings By State</a></li>
<li><a href=" http://www.wpc.ncep.noaa.gov/ww.shtml">Excessive Rainfall and Winter
Weather Forecasts</a></li>
<li><a href="http://water.weather.gov/ahps/?current_color=flood&amp;current_type=a
ll&amp;fcst_type=obs&amp;conus_map=d_map">River Flooding </a></li>
<li><a href=" http://www.weather.gov">Latest Warnings</a></li>
<li><a href=" http://www.spc.noaa.gov/products/outlook/">Thunderstorm/Tornado Outl
ook </a></li>
<li><a href=" http://www.nhc.noaa.gov/">Hurricanes </a></li>

```

```
<li><a href=" http://www.spc.noaa.gov/products/fire_wx/">Fire Weather Outlooks </a>
</li>
<li><a href=" http://www.cpc.ncep.noaa.gov/products/stratosphere/uv_index/uv_aler
t.shtml">UV Alerts </a></li>
<li><a href=" http://www.drought.gov/">Drought </a></li>
<li><a href="http://www.swpc.noaa.gov/products/alerts-watches-and-warnings">Space
Weather </a></li>
<li><a href=" http://www.nws.noaa.gov/nwr/">NOAA Weather Radio </a></li>
<li><a href=" http://alerts.weather.gov/">NWS CAP Feeds </a></li>
</ul>
</div>
</div>
<div class="sitemap-section">
<div class="panel-heading">
<a class="sitemap-section-heading" href="https://www.weather.gov/wrh/climate">PAST
WEATHER</a>
<button class="menu-toggle pull-right" data-target="#sitemap-2" data-toggle="colla
pse" type="button">
<span class="sr-only">Toggle menu</span>
<span class="icon-bar"></span>
<span class="icon-bar"></span>
<span class="icon-bar"></span>
</button>
</div>
<div class="sitemap-section-body panel-body collapsable collapse" id="sitemap-2">
<ul class="list-unstyled">
<li><a href=" http://www.cpc.ncep.noaa.gov/products/MD_index.shtml">Climate Monito
ring </a></li>
<li><a href="https://www.weather.gov/wrh/climate">Past Weather </a></li>
<li><a href="https://www.weather.gov/wrh/climate">Monthly Temps </a></li>
<li><a href="https://www.weather.gov/wrh/climate">Records </a></li>
<li><a href="https://www.esrl.noaa.gov/gmd/grad/solcalc/sunrise.html">Astronomical
Data </a></li>
<li><a href="https://www.climate.gov/maps-data/dataset/past-weather-zip-code-data-
table">Certified Weather Data </a></li>
</ul>
</div>
</div>
<div class="sitemap-section">
<div class="panel-heading">
<a class="sitemap-section-heading" href="http://www.weather.gov/current">CURRENT C
ONDITIONS</a>
<button class="menu-toggle pull-right" data-target="#sitemap-3" data-toggle="colla
pse" type="button">
<span class="sr-only">Toggle menu</span>
<span class="icon-bar"></span>
<span class="icon-bar"></span>
<span class="icon-bar"></span>
</button>
</div>
<div class="sitemap-section-body panel-body collapsable collapse" id="sitemap-3">
<ul class="list-unstyled">
<li><a href=" https://radar.weather.gov">Radar </a></li>
<li><a href="http://www.cpc.ncep.noaa.gov/products/monitoring_and_data/">Climate M
onitoring </a></li>
<li><a href=" http://water.weather.gov/ahps/">River Levels </a></li>
<li><a href=" http://water.weather.gov/precip/">Observed Precipitation </a></li>
<li><a href="https://www.wpc.ncep.noaa.gov/sfc/sfcobs/sfcobs.shtml">Surface Weathe
r </a></li>
<li><a href="http://www.spc.noaa.gov/obswx/maps/">Upper Air </a></li>
<li><a href=" http://www.ndbc.noaa.gov/">Marine and Buoy Reports </a></li>
<li><a href="http://www.nohrsc.noaa.gov/interactive/html/map.html">Snow Cover </a>
</li>
<li><a href=" http://www.weather.gov/satellite">Satellite </a></li>
<li><a href=" http://www.swpc.noaa.gov/">Space Weather </a></li>
<li><a href="http://www.weather.gov/pr">International Observations</a></li>
</ul>
</div>
```



```
</div>
</div>
<div class="sitemap-section">
<div class="panel-heading">
<a class="sitemap-section-heading" href="http://weather.gov/forecastmaps">FORECAST
</a>
<button class="menu-toggle pull-right" data-target="#sitemap-4" data-toggle="colla
pse" type="button">
<span class="sr-only">Toggle menu</span>
<span class="icon-bar"></span>
<span class="icon-bar"></span>
<span class="icon-bar"></span>
</button>
</div>
<div class="sitemap-section-body panel-body collapsable collapse" id="sitemap-4">
<ul class="list-unstyled">
<li><a href=" http://www.weather.gov/">Local Forecast </a></li>
<li><a href="http://www.weather.gov/pr">International Forecasts</a></li>
<li><a href=" http://www.spc.noaa.gov/">Severe Weather </a></li>
<li><a href=" http://www.wpc.ncep.noaa.gov/">Current Outlook Maps </a></li>
<li><a href="http://www.cpc.ncep.noaa.gov/products/Drought">Drought </a></li>
<li><a href="http://www.weather.gov/fire">Fire Weather </a></li>
<li><a href=" http://www.wpc.ncep.noaa.gov/">Fronts/Precipitation Maps </a></li>
<li><a href=" http://www.nws.noaa.gov/forecasts/graphical/">Current Graphical Fore
cast Maps </a></li>
<li><a href="http://water.noaa.gov">Rivers </a></li>
<li><a href="https://www.weather.gov/marine/">Marine </a></li>
<li><a href="https://ocean.weather.gov/marine_areas.php">Offshore and High Seas</a
></li>
<li><a href=" http://www.nhc.noaa.gov/">Hurricanes </a></li>
<li><a href=" http://aviationweather.gov">Aviation Weather </a></li>
<li><a href="http://www.cpc.ncep.noaa.gov/products/OUTLOOKS_index.shtml">Climatic
Outlook </a></li>
</ul>
</div>
</div>
<div class="sitemap-section">
<div class="panel-heading">
<a class="sitemap-section-heading" href="http://www.weather.gov/informationcente
r">INFORMATION CENTER</a>
<button class="menu-toggle pull-right" data-target="#sitemap-5" data-toggle="colla
pse" type="button">
<span class="sr-only">Toggle menu</span>
<span class="icon-bar"></span>
<span class="icon-bar"></span>
<span class="icon-bar"></span>
</button>
</div>
<div class="sitemap-section-body panel-body collapsable collapse" id="sitemap-5">
<ul class="list-unstyled">
<li><a href=" http://www.spaceweather.gov">Space Weather </a></li>
<li><a href="http://www.weather.gov/briefing/">Daily Briefing </a></li>
<li><a href=" http://www.nws.noaa.gov/om/marine/home.htm">Marine </a></li>
<li><a href="https://www.weather.gov/wrh/climate">Climate </a></li>
<li><a href="http://www.weather.gov/fire">Fire Weather </a></li>
<li><a href=" http://www.aviationweather.gov/">Aviation </a></li>
<li><a href="http://mag.ncep.noaa.gov/">Forecast Models </a></li>
<li><a href="http://water.weather.gov/ahps/">Water </a></li>
<li><a href="https://www.weather.gov/gis">GIS</a></li>
<li><a href=" http://www.nws.noaa.gov/om/coop/">Cooperative Observers </a></li>
<li><a href="https://www.weather.gov/skywarn/">Storm Spotters </a></li>
<li><a href="http://www.tsunami.gov">Tsunami Warning System</a></li>
<li><a href="http://water.noaa.gov/">National Water Center</a></li>
<li><a href="http://www.weather.gov/pr">International Weather</a></li>
</ul>
</div>
</div>
```

```
<div class="sitemap-section">
<div class="panel-heading">
<a class="sitemap-section-heading" href="http://weather.gov/safety">WEATHER SAFETY
</a>
<button class="menu-toggle pull-right" data-target="#sitemap-6" data-toggle="colla
pse" type="button">
<span class="sr-only">Toggle menu</span>
<span class="icon-bar"></span>
<span class="icon-bar"></span>
<span class="icon-bar"></span>
</button>
</div>
<div class="sitemap-section-body panel-body collapsable collapse" id="sitemap-6">
<ul class="list-unstyled">
<li><a href="http://www.weather.gov/nwr/">NOAA Weather Radio</a></li>
<li><a href="http://www.weather.gov/stormready/">StormReady</a></li>
<li><a href="http://www.nws.noaa.gov/om/heat/index.shtml">Heat </a></li>
<li><a href="https://www.weather.gov/safety/lightning">Lightning </a></li>
<li><a href="http://www.nhc.noaa.gov/prepare/">Hurricanes </a></li>
<li><a href="http://www.nws.noaa.gov/om/thunderstorm/">Thunderstorms </a></li>
<li><a href="https://www.weather.gov/safety/tornado">Tornadoes </a></li>
<li><a href="https://www.weather.gov/safety/ripcurrent">Rip Currents </a></li>
<li><a href="https://www.weather.gov/safety/flood">Floods </a></li>
<li><a href="https://www.weather.gov/safety/tsunami">Tsunamis</a></li>
<li><a href="https://www.weather.gov/tsunamiready/">TsunamiReady</a></li>
<li><a href="http://www.weather.gov/om/winter/index.shtml">Winter Weather </a></l
i>
<li><a href="http://www.nws.noaa.gov/om/heat/uv.shtml">Ultra Violet Radiation </a>
</li>
<li><a href="http://www.weather.gov/airquality/">Air Quality </a></li>
<li><a href="http://www.weather.gov/om/hazstats.shtml">Damage/Fatality/Injury Sta
tistics </a></li>
<li><a href="http://www.redcross.org/">Red Cross </a></li>
<li><a href="http://www.fema.gov/">Federal Emergency Management Agency (FEMA) </a>
</li>
<li><a href="http://www.weather.gov/om/brochures.shtml">Brochures </a></li>
<li><a href="http://www.nws.noaa.gov/os/marine/safeboating/">Safe Boating</a></li>
</ul>
</div>
</div>
<div class="sitemap-section">
<div class="panel-heading">
<a class="sitemap-section-heading" href="http://weather.gov/news">NEWS</a>
<button class="menu-toggle pull-right" data-target="#sitemap-7" data-toggle="colla
pse" type="button">
<span class="sr-only">Toggle menu</span>
<span class="icon-bar"></span>
<span class="icon-bar"></span>
<span class="icon-bar"></span>
</button>
</div>
<div class="sitemap-section-body panel-body collapsable collapse" id="sitemap-7">
<ul class="list-unstyled">
<li><a href="http://weather.gov/news">Newsroom</a></li>
<li><a href="http://www.nws.noaa.gov/com/weatherreadynation/calendar.html">Events
</a></li>
<li><a href="http://www.weather.gov/om/brochures.shtml">Pubs/Brochures/Booklets
</a></li>
</ul>
</div>
</div>
<div class="sitemap-section">
<div class="panel-heading">
<a class="sitemap-section-heading" href="http://weather.gov/owlie">EDUCATION</a>
<button class="menu-toggle pull-right" data-target="#sitemap-8" data-toggle="colla
pse" type="button">
<span class="sr-only">Toggle menu</span>
```

```
<span class="icon-bar"></span>
<span class="icon-bar"></span>
<span class="icon-bar"></span>
</button>
</div>
<div class="sitemap-section-body panel-body collapsable collapse" id="sitemap-8">
<ul class="list-unstyled">
<li><a href="http://weather.gov/education">NWS Education Home</a></li>
<li><a href="http://www.nws.noaa.gov/com/weatherreadynation/force.html">Be A Force
of Nature</a></li>
<li><a href=" http://www.education.noaa.gov/Weather_and_Atmosphere/">NOAA Educatio
n Resources </a></li>
<li><a href=" http://www.weather.gov/glossary/">Glossary </a></li>
<li><a href="https://www.weather.gov/jetstream/">JetStream </a></li>
<li><a href=" http://www.weather.gov/training/">NWS Training Portal </a></li>
<li><a href="https://library.noaa.gov/">NOAA Library </a></li>
<li><a href="http://weather.gov/owlie">For Students, Parents and Teachers</a></li>
<li><a href="http://www.weather.gov/owlie/publication_brochures">Brochures </a></l
i>
</ul>
</div>
</div>
<div class="sitemap-section">
<div class="panel-heading">
<a class="sitemap-section-heading" href="http://weather.gov/about">ABOUT</a>
<button class="menu-toggle pull-right" data-target="#sitemap-9" data-toggle="colla
pse" type="button">
<span class="sr-only">Toggle menu</span>
<span class="icon-bar"></span>
<span class="icon-bar"></span>
<span class="icon-bar"></span>
</button>
</div>
<div class="sitemap-section-body panel-body collapsable collapse" id="sitemap-9">
<ul class="list-unstyled">
<li><a href="http://weather.gov/organization">Organization </a></li>
<li><a href="https://www.noaa.gov/NWStransformation">NWS Transformation</a></li>
<li><a href="https://www.weather.gov/media/wrn/NWS-2023-Strategic-Plan.pdf">Strate
gic Plan </a></li>
<li><a href="https://weather.gov/insider">For NWS Employees </a></li>
<li><a href="https://www.weather.gov/international/">International </a></li>
<li><a href="http://www.weather.gov/ncep">National Centers </a></li>
<li><a href=" http://www.weather.gov/tg/">Products and Services </a></li>
<li><a href="http://www.weather.gov/careers/">Careers</a></li>
<li><a href=" http://www.weather.gov/glossary/">Glossary </a></li>
<li><a href="http://weather.gov/contact">Contact Us </a></li>
<li><a href="https://weather.gov/socialmedia">Social Media</a></li>
</ul>
</div>
</div>
</div>
</div>
</div>
<!-- legal footer area -->
<div class="footer-legal">
<div class="col-xs-12 col-sm-2 col-md-2" id="footerLogo">
<a href="//www.usa.gov"></a>
</div>
<div class="col-xs-12 col-sm-4 col-md-4">
<ul class="list-unstyled footer-legal-content">
<li><a href="//www.commerce.gov">US Dept of Commerce</a></li>
<li><a href="//www.noaa.gov">National Oceanic and Atmospheric Administration</a></l
i>
<li><a href="//www.weather.gov">National Weather Service</a></li>
<li><a href="https://www.weather.gov/mtr">San Francisco Bay Area, CA</a></li><li><
br><a href="mailto:w-mtr.webmaster@noaa.gov">Comments? Questions? Please Contact U
```

```
s.</a></br></li> </ul>
</div>
<div class="col-xs-12 col-sm-3 col-md-3">
<ul class="list-unstyled">
<li><a href="https://www.weather.gov/disclaimer">Disclaimer</a></li>
<li><a href="//www.cio.noaa.gov/services_programs/info_quality.html">Information Q
uality</a></li>
<li><a href="https://www.weather.gov/help">Help</a></li>
<li><a href="//www.weather.gov/glossary">Glossary</a></li>
</ul>
</div>
<div class="col-xs-12 col-sm-3 col-md-3">
<ul class="list-unstyled">
<li><a href="https://www.weather.gov/privacy">Privacy Policy</a></li>
<li><a href="https://www.noaa.gov/foia-freedom-of-information-act">Freedom of Info
rmation Act (FOIA)</a></li>
<li><a href="https://www.weather.gov/about">About Us</a></li>
<li><a href="https://www.weather.gov/careers">Career Opportunities</a></li>
</ul>
</div>
</div>
</footer>
</main>
</body>
</html>
```

```
In [21]: seven_day = soup.find(id="seven-day-forecast")
seven_day
```

```

Out[21]: <div class="panel panel-default" id="seven-day-forecast">
<div class="panel-heading">
<b>Extended Forecast for</b>
<h2 class="panel-title">
San Francisco CA </h2>
</div>
<div class="panel-body" id="seven-day-forecast-body">
<div id="seven-day-forecast-container"><ul class="list-unstyled" id="seven-day-fo
recast-list"><li class="forecast-tombstone">
<div class="tombstone-container">
<p class="period-name">Tonight<br/><br/></p>
<p></p>
<p class="short-desc">Mostly Cloudy</p><p class="temp temp-low">Low: 53 °F</p></div>
</li><li class="forecast-tombstone">
<div class="tombstone-container">
<p class="period-name">Tuesday<br/><br/></p>
<p></p><p class="short-desc">Decreasing<br/>Clouds</p><p
class="temp temp-high">High: 64 °F</p></div></li><li class="forecast-tombstone">
<div class="tombstone-container">
<p class="period-name">Tuesday<br/>Night</p>
<p></p><p class="short-desc">Mostly Cloudy</p><p class="temp temp-low">Low: 54
°F</p></div></li><li class="forecast-tombstone">
<div class="tombstone-container">
<p class="period-name">Wednesday<br/><br/></p>
<p></p><p class="short-desc">Par
tly Sunny</p><p class="temp temp-high">High: 64 °F</p></div></li><li class="foreca
st-tombstone">
<div class="tombstone-container">
<p class="period-name">Wednesday<br/>Night</p>
<p></p><p class
="short-desc">Partly Cloudy</p><p class="temp temp-low">Low: 52 °F</p></div></li>
<li class="forecast-tombstone">
<div class="tombstone-container">
<p class="period-name">Thursday<br/><br/></p>
<p><img alt="Thursday: Partly sunny, with a high near 63." class="forecast-icon" s
rc="newimages/medium/bkn.png" title="Thursday: Partly sunny, with a high near 6
3."/></p><p class="short-desc">Partly Sunny</p><p class="temp temp-high">High: 63
°F</p></div></li><li class="forecast-tombstone">
<div class="tombstone-container">
<p class="period-name">Thursday<br/>Night</p>
<p></p><p class="short-desc">Mostly Cloudy</p><p class="temp temp-
low">Low: 54 °F</p></div></li><li class="forecast-tombstone">
<div class="tombstone-container">
<p class="period-name">Friday<br/><br/></p>
<p><img alt="Friday: Mostly sunny, with a high near 64." class="forecast-icon" src
="newimages/medium/sct.png" title="Friday: Mostly sunny, with a high near 64."/></
p><p class="short-desc">Mostly Sunny</p><p class="temp temp-high">High: 64 °F</p>
</div></li><li class="forecast-tombstone">
<div class="tombstone-container">
<p class="period-name">Friday<br/>Night</p>
<p></p><p class="short-desc">Mostly Clear</p><p class="temp temp-low">Lo
w: 51 °F</p></div></li></ul></div>
<script type="text/javascript">
// equalize forecast heights
$(function () {
  var maxh = 0;
  $(".forecast-tombstone .short-desc").each(function () {
    var h = $(this).height();
    if (h > maxh) { maxh = h; }
  });
  $(".forecast-tombstone .short-desc").height(maxh);
});
</script> </div>
</div>
```

```
In [22]: forecast_items = seven_day.find_all(class_="tombstone-container")
```

```
forecast_items
```

```

Out[22]: [<div class="tombstone-container">
  <p class="period-name">Tonight<br/><br/></p>
  <p></
p><p class="short-desc">Mostly Cloudy</p><p class="temp temp-low">Low: 53 °F</p></
div>,
  <div class="tombstone-container">
  <p class="period-name">Tuesday<br/><br/></p>
  <p></p><p class="short-desc">Decreasing<br/>Clouds</p>
<p class="temp temp-high">High: 64 °F</p></div>,
  <div class="tombstone-container">
  <p class="period-name">Tuesday<br/>Night</p>
  <p></p><p class="short-desc">Mostly Cloudy</p><p class="temp temp-low">Low: 54
°F</p></div>,
  <div class="tombstone-container">
  <p class="period-name">Wednesday<br/><br/></p>
  <p></p><p class="short-desc">P
artly Sunny</p><p class="temp temp-high">High: 64 °F</p></div>,
  <div class="tombstone-container">
  <p class="period-name">Wednesday<br/>Night</p>
  <p></p><p class
="short-desc">Partly Cloudy</p><p class="temp temp-low">Low: 52 °F</p></div>,
  <div class="tombstone-container">
  <p class="period-name">Thursday<br/><br/></p>
  <p></p><p class="short-desc">Partly Sunny</p><p class="temp temp-high">High: 63
°F</p></div>,
  <div class="tombstone-container">
  <p class="period-name">Thursday<br/>Night</p>
  <p></p><p class="short-desc">Mostly Cloudy</p><p class="temp temp-
low">Low: 54 °F</p></div>,
  <div class="tombstone-container">
  <p class="period-name">Friday<br/><br/></p>
  <p><img alt="Friday: Mostly sunny, with a high near 64." class="forecast-icon" sr
c="newimages/medium/sct.png" title="Friday: Mostly sunny, with a high near 64."/>
</p><p class="short-desc">Mostly Sunny</p><p class="temp temp-high">High: 64 °F</p>
</div>,
  <div class="tombstone-container">
  <p class="period-name">Friday<br/>Night</p>
  <p></p><p class="short-desc">Mostly Clear</p><p class="temp temp-low">Lo
w: 51 °F</p></div>]

```

```
In [23]: tonight = forecast_items[0]
```

```
tonight
```

```
Out [23]: <div class="tombstone-container">
<p class="period-name">Tonight<br/><br/></p>
<p></p>
<p class="short-desc">Mostly Cloudy</p><p class="temp temp-low">Low: 53 °F</p></d
iv>
```

3. Extracting information from the page

As we can see, inside the forecast item `tonight` is all the information we want. There are four pieces of information we can extract:

- The name of the forecast item — in this case, `Tonight`.
- The description of the conditions — this is stored in the `title` property of `img`.
- A short description of the conditions — in this case, `Mostly Cloudy`.
- The temperature low — in this case, `53 degrees`.

We'll extract the name of the forecast item, the short description, and the temperature first, since they're all similar:

```
In [24]: tonight
```

```
Out [24]: <div class="tombstone-container">
<p class="period-name">Tonight<br/><br/></p>
<p></p>
<p class="short-desc">Mostly Cloudy</p><p class="temp temp-low">Low: 53 °F</p></d
iv>
```

```
In [25]: period = tonight.find(class_="period-name").get_text()
short_desc = tonight.find(class_="short-desc").get_text()
temp = tonight.find(class_="temp").get_text()
print(period)
print(short_desc)
print(temp)
```

```
Tonight
Mostly Cloudy
Low: 53 °F
```

Now, we can extract the `title` attribute from the `img` tag. To do this, we just treat the `BeautifulSoup` object like a dictionary, and pass in the attribute we want as a key:

```
In [26]: tonight
```

```
Out [26]: <div class="tombstone-container">
<p class="period-name">Tonight<br/><br/></p>
<p></p>
<p class="short-desc">Mostly Cloudy</p><p class="temp temp-low">Low: 53 °F</p></d
iv>
```

```
In [27]: img = tonight.find("img")
desc = img['alt']
print(desc)
```

```
Tonight: Mostly cloudy, with a low around 53. South southwest wind around 10 mph.
```

4. Extracting all the information from the page

Now that we know how to extract each individual piece of information, we can combine our knowledge with CSS selectors and list comprehensions to **extract everything at once**.

In the below code, we will:

- Select all items with the class `period-name` inside an item with the class `tombstone-container` in `seven_day`.
- Use a list comprehension to call the `get_text` method on each `BeautifulSoup` object.

```
In [28]: period_tags = seven_day.select(".tombstone-container .period-name")
period_tags
```

```
Out[28]: [<p class="period-name">Tonight<br/><br/></p>,
<p class="period-name">Tuesday<br/><br/></p>,
<p class="period-name">Tuesday<br/>Night</p>,
<p class="period-name">Wednesday<br/><br/></p>,
<p class="period-name">Wednesday<br/>Night</p>,
<p class="period-name">Thursday<br/><br/></p>,
<p class="period-name">Thursday<br/>Night</p>,
<p class="period-name">Friday<br/><br/></p>,
<p class="period-name">Friday<br/>Night</p>]
```

```
In [29]: period_tags = seven_day.select(".tombstone-container .period-name")
periods = [pt.get_text() for pt in period_tags]
periods
```

```
Out[29]: ['Tonight',
'Tuesday',
'TuesdayNight',
'Wednesday',
'WednesdayNight',
'Thursday',
'ThursdayNight',
'Friday',
'FridayNight']
```

As we can see above, our technique gets us each of the period names, in order. We can apply the same technique to get the other three fields:

```
In [30]: short_descs = [sd.get_text() for sd in seven_day.select(".tombstone-container .short_descs")]
temps = [t.get_text() for t in seven_day.select(".tombstone-container .temp")]
descs = [d["title"] for d in seven_day.select(".tombstone-container img")]
print(short_descs)
print(temps)
print(descs)
```

```
['Mostly Cloudy', 'DecreasingClouds', 'Mostly Cloudy', 'Partly Sunny', 'Partly Cloudy', 'Partly Sunny', 'Mostly Cloudy', 'Mostly Sunny', 'Mostly Clear']
['Low: 53 °F', 'High: 64 °F', 'Low: 54 °F', 'High: 64 °F', 'Low: 52 °F', 'High: 63 °F', 'Low: 54 °F', 'High: 64 °F', 'Low: 51 °F']
['Tonight: Mostly cloudy, with a low around 53. South southwest wind around 10 mph. ', 'Tuesday: Mostly cloudy, then gradually becoming sunny, with a high near 64. South southwest wind 8 to 14 mph, with gusts as high as 20 mph. ', 'Tuesday Night: Mostly cloudy, with a low around 54. South southwest wind 6 to 10 mph. ', 'Wednesday: Partly sunny, with a high near 64. South wind 6 to 10 mph. ', 'Wednesday Night: Partly cloudy, with a low around 52. West wind 7 to 10 mph. ', 'Thursday: Partly sunny, with a high near 63.', 'Thursday Night: Mostly cloudy, with a low around 54.', 'Friday: Mostly sunny, with a high near 64.', 'Friday Night: Mostly clear, with a low around 51.']
```

5. Combining our data into a Pandas Dataframe

We can now combine the data into a `Pandas DataFrame` and analyze it.

```
In [31]: import pandas as pd

weather = pd.DataFrame({"period": periods,
                        "short_desc": short_descs,
                        "temp": temps,
                        "desc": descs
                       })

weather
```

```
Out[31]:
```

	period	short_desc	temp	desc
0	Tonight	Mostly Cloudy	Low: 53 °F	Tonight: Mostly cloudy, with a low around 53. ...
1	Tuesday	DecreasingClouds	High: 64 °F	Tuesday: Mostly cloudy, then gradually becomin...
2	TuesdayNight	Mostly Cloudy	Low: 54 °F	Tuesday Night: Mostly cloudy, with a low aroun...
3	Wednesday	Partly Sunny	High: 64 °F	Wednesday: Partly sunny, with a high near 64. ...
4	WednesdayNight	Partly Cloudy	Low: 52 °F	Wednesday Night: Partly cloudy, with a low aro...
5	Thursday	Partly Sunny	High: 63 °F	Thursday: Partly sunny, with a high near 63.
6	ThursdayNight	Mostly Cloudy	Low: 54 °F	Thursday Night: Mostly cloudy, with a low arou...
7	Friday	Mostly Sunny	High: 64 °F	Friday: Mostly sunny, with a high near 64.
8	FridayNight	Mostly Clear	Low: 51 °F	Friday Night: Mostly clear, with a low around 51.

Example 2. Pandas Web Scraping

Pandas makes it easy to scrape a `table` tag on a web page. You can use the function `read_html(url)` to get webpage contents.

It's only suitable for fetching `Table` type data, then let's see what kind of pages meet the conditions?

HTML Tables

HTML tables allow web developers to arrange data into rows and columns:

```
In [32]: %%HTML

<table>
  <tr>
    <th>Company</th>
    <th>Contact</th>
    <th>Country</th>
  </tr>
```

```
<tr>
  <td>Apple</td>
  <td>Tim Cook</td>
  <td>United States</td>
</tr>
<tr>
  <td>Tencent</td>
  <td>Pony Ma</td>
  <td>China</td>
</tr>
</table>
```

Company	Contact	Country
Apple	Tim Cook	United States
Tencent	Pony Ma	China

Open the web version with a browser, F12 view the HTML structure of the web page, you will find that the structure of the eligible web page has a common feature.

If you find `Table` format, you can use `pd.read_html()`

Sina finance - Institutional ownership

Take the aggregate shareholding data of Sina Financial institutions as an example:

```
In [33]: url = 'http://vip.stock.finance.sina.com.cn/q/go.php/vInvestConsult/kind/nbjy/index
df = pd.read_html(url)[0]
df
```

Out [33]:

	股票代码	股票名称	变动人	变动类型	变动股数	成交均价	变动金额(万元)	变动后持股数	变动原因	变动日期	持股种类	与董监高关系	董监高职务
0	2145	中核钛白	沈鑫	购买	3350000	4.02	1346.70	14150000	竞价交易	2024-04-19	A股	本人	非独立董事
1	688660	电气风电	吴改	购买	10200	3.25	3.32	40255	二级市场买卖	2024-04-19	A股	本人	非独立董事
2	688660	电气风电	王勇	购买	10000	3.26	3.26	33248	二级市场买卖	2024-04-19	A股	本人	审计委员会委员
3	688660	电气风电	王红春	购买	8000	3.24	2.59	33153	二级市场买卖	2024-04-19	A股	本人	职工代表监事
4	688660	电气风电	乔银平	购买	12240	3.27	4.00	42540	二级市场买卖	2024-04-19	A股	本人	董事
5	688660	电气风电	刘向楠	购买	9088	3.25	2.95	29376	二级市场买卖	2024-04-19	A股	本人	职工监事
6	688660	电气风电	黄锋锋	购买	11000	3.25	3.58	40658	二级市场买卖	2024-04-19	A股	本人	财务总监
7	688560	明冠新材	闫勇	购买	30000	11.32	33.96	830000	二级市场买卖	2024-04-19	A股	本人	董事
8	603938	三孚股份	孙任靖	购买	5000	12.5	6.25	150409520	二级市场买卖	2024-04-19	A股	本人	董事长
9	688176	亚虹	PAN KE	购买	155000	5.65	87.58	130103577	二级	2024-04-19	A股	本人	董事

股票代码	股票名称	变动人	变动类型	变动股数	成交均价	变动金额(万元)	变动后持股数	变动原因	变动日期	持股种类	与董监高关系	董监高职务	
	医药							市场买卖				长	
10	2145	中核钛白	沈鑫	购买	10800000	3.98	4298.40	10800000	竞价交易	2024-04-18	A股	本人	非独立董事
11	2286	保龄宝	张国刚	购买	8900	5.54	4.93	55900	竞价交易	2024-04-18	A股	本人	副总经理
12	688560	明冠新材	闫勇	购买	200000	11.58	231.60	800000	二级市场买卖	2024-04-18	A股	本人	董事
13	603031	安孚科技	梁红颖	送、转	9450	0	0.00	30450	分红送转	2024-04-18	A股	本人	副总经理
14	603031	安孚科技	刘荣海	送、转	33930	0	0.00	109330	分红送转	2024-04-18	A股	本人	非独立董事
15	603031	安孚科技	王晓飞	送、转	5535	0	0.00	17835	分红送转	2024-04-18	A股	本人	副总经理
16	688311	盟升电子	陈英	回购注销	-10500	16.37	17.19	10500	股权激励回购注销实施	2024-04-18	A股	本人	副总经理
17	688311	盟升电子	毛钢烈	回购注销	-13440	16.37	22.00	13440	股权激励回购注销实施	2024-04-18	A股	本人	非独立董事

	股票代码	股票名称	变动人	变动类型	变动股数	成交均价	变动金额(万元)	变动后持股数	变动原因	变动日期	持股种类	与董监高关系	董监高职务
18	688311	盟升电子	覃光全	回购注销	-14700	16.37	24.06	14700	股权激励回购注销实施	2024-04-18	A股	本人	非独立董事
19	688311	盟升电子	向荣	回购注销	-27300	16.37	44.69	4467288	股权激励回购注销实施	2024-04-18	A股	本人	董事长
20	688176	亚虹医药	PAN KE	购买	300000	5.79	173.70	129948577	二级市场买卖	2024-04-18	A股	本人	董事长
21	688176	亚虹医药	杨明远	购买	18500	5.8	10.73	18500	二级市场买卖	2024-04-18	A股	本人	非独立董事
22	688169	石头科技	张磊	出售	-500	385	19.25	3768	二级市场买卖	2024-04-18	A股	本人	核心技术人员
23	600576	祥源文旅	陈亚文	购买	6000	5.6	3.36	148500	二级市场买卖	2024-04-17	A股	本人	监事
24	2640	跨境通	杨建新	出售	-1756700	2.24	393.50	150859148	竞价交易	2024-04-17	A股	本人	董事长
25	688628	优利德	孙乔	--	--	--	0.00	不详	二级市场买卖	2024-04-17	A股	本人	核心技术人员
26	688560	明冠	闫勇	购买	57070	11.76	67.11	600000	二级市场	2024-04-17	A股	本人	董事

股票代码	股票名称	变动人	变动类型	变动股数	成交均价	变动金额(万元)	变动后持股数	变动原因	变动日期	持股种类	与董监高关系	董监高职务
	新材							场买卖				
27	600576 祥源文旅	陈云钊	购买	9000	5.55	5.00	30300	二级市场买卖	2024-04-16	A股	本人	职工代表监事
28	600576 祥源文旅	王琦	购买	18600	5.37	9.99	88500	二级市场买卖	2024-04-16	A股	本人	职工监事
29	600576 祥源文旅	徐中平	购买	20300	5.42	11.00	306300	二级市场买卖	2024-04-16	A股	本人	非独立董事
30	600576 祥源文旅	詹纯伟	购买	4600	5.35	2.46	147000	二级市场买卖	2024-04-16	A股	本人	监事
31	688183 生益电子	王小平	购买	38000	8.86	33.67	38000	二级市场买卖	2024-04-16	A股	本人	核心技术人员
32	2286 保险宝	王强	购买	6600	5.14	3.39	107400	竞价交易	2024-04-16	A股	本人	非独立董事
33	2286 保险宝	张国刚	购买	10000	5.15	5.15	47000	竞价交易	2024-04-16	A股	本人	副总经理
34	2640 跨境通	杨建新	出售	-2538800	2.17	550.92	152615848	竞价交易	2024-04-16	A股	本人	董事长
35	688560 明冠新材	闫勇	购买	142930	11.16	159.51	542930	二级市场买卖	2024-04-16	A股	本人	董事

股票代码	股票名称	变动人	变动类型	变动股数	成交均价	变动金额(万元)	变动后持股数	变动原因	变动日期	持股种类	与董监高关系	董监高职务	
36	3013	地铁设计	廖景	购买	78000	14.1	109.98	78000	竞价交易	2024-04-16	A股	本人	审计委员会委员
37	605222	起帆电缆	周桂幸	其它	-21000000	14.73	30933.00	63991400	其他	2024-04-16	A股	本人	副董事长
38	301021	英诺激光	张勇	购买	6300	15.02	9.46	12200	竞价交易	2024-04-16	A股	本人	副总经理
39	688538	和辉光电	李凤玲	购买	115000	2.1	24.15	735000	二级市场买卖	2024-04-16	A股	本人	董事会秘书

```
In [34]: df = pd.DataFrame()

for i in range(1, 3):
    url = 'http://vip.stock.finance.sina.com.cn/q/go.php/vInvestConsult/kind/nbjy/i'
    df = pd.concat([df, pd.read_html(url)[0]])
    print("Page %s completed!" % i)

df
```

Page 1 completed!
Page 2 completed!

Out [34]:

	股票代码	股票名称	变动人	变动类型	变动股数	成交均价	变动金额(万元)	变动后持股数	变动原因	变动日期	持股种类	与董监高关系	董监高职务
0	2145	中核钛白	沈鑫	购买	3350000	4.02	1346.70	14150000	竞价交易	2024-04-19	A股	本人	非独立董事
1	688660	电气风电	吴改	购买	10200	3.25	3.32	40255	二级市场买卖	2024-04-19	A股	本人	非独立董事
2	688660	电气风电	王勇	购买	10000	3.26	3.26	33248	二级市场买卖	2024-04-19	A股	本人	审计委员会委员
3	688660	电气风电	王红春	购买	8000	3.24	2.59	33153	二级市场买卖	2024-04-19	A股	本人	职工代表监事
4	688660	电气风电	乔银平	购买	12240	3.27	4.00	42540	二级市场买卖	2024-04-19	A股	本人	董事
...
35	603456	九州药业	沙裕杰	购买	10000	16.05	16.05	90000	二级市场买卖	2024-04-11	A股	本人	执行董事
36	688691	灿芯股份	庄志青	--	--	--	0.00	3092850	NaN	2024-04-11	A股	本人	董事
37	688257	新锐股份	杨汉民	出售	-80000	25	200.00	309608	二级市场买卖	2024-04-11	A股	本人	事业部技术副总经理
38	872190	雷神科技	路凯林	购买	110000	16.36	179.96	643471	竞价交易	2024-04-11	A股	本人	董事
39	688037	芯源微	王绍勇	购买	2980	96.51	28.76	48980	二级市场买卖	2024-04-11	A股	本人	核心技术人员

80 rows × 13 columns