

**Webscraping**

## Basic html page

```
<!DOCTYPE html>
<html>
<head>
  <title>Web Page!</title>
  <style>
    body {background-color: powderblue;}
    h1   {color: blue;}
    p    {color: red;}
  </style>
  <link rel="stylesheet" href="styles.css">
  <script>
    document.getElementById("demo").innerHTML = "Hello JavaScript!";
  </script>
</head>
<body>
  <h1>A Very Bold Header</h1>
  <div style="background-color:lightblue">
    <p>This is a paragraph.</p>
  </div>
</body>
</html>
```

## nyc weather history

<http://w1.weather.gov/data/obhistory/KNYC.html>  
(<http://w1.weather.gov/data/obhistory/KNYC.html>)

```
In [1]: knyc_link = 'http://w1.weather.gov/data/obhistory/KNYC.html'
```

```
In [2]: import requests

        knyc_page = requests.get(knyc_link)
        knyc_page
```

```
Out[2]: <Response [200]>
```

```
In [4]: # need to parse some html!  
        from bs4 import BeautifulSoup
```

```
In [5]: knyc_soup = BeautifulSoup(knyc_page.content)
```

```
In [6]: # first 1000 characters more legibly
print(knyc_soup.prettify()[:1000])
```

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0 Transitional//EN">
<html>
  <head>
    <meta content="Leon Minton" name="Author"/>
    <title>
      National Weather Service : Observed Weather for past 3 Days : New York Cit
y, Central Park
    </title>
    <link href="/images/weather/fcicons/main.css" rel="stylesheet" type="text/cs
s"/>
  </head>
  <body background="/images/weather/fcicons/gray_background.gif" bgcolor="#ffff
ff" leftmargin="0" marginheight="0" marginwidth="0" topmargin="0">
    <table background="/images/weather/fcicons/topbanner.jpg" border="0" cellpad
ding="0" cellpadding="0" cellspacing="0" width="670">
      <tr>
        <td align="right" height="19">
          <a href="http://weather.gov">
            <span class="nwslink">
              weather.gov
            </span>
          </a>
        </td>
      </tr>
    </table>
    <table border="0" cellpadding="0" cellspacing="0" width="670">
      <tr valign="top">
        <td rowspan="2">
          <a href="http://www.noaa.gov">
            <img alt="NOAA logo - Click to go to the NOAA homepage" border="0" heigh
t="78" s
```

```
In [7]: # print the 4rd table in the page
print(knyc_soup.find_all('table')[3])
```

Time (edt)			Wind (mph)			Vis. (mi.)			Weather			Sky Cond.																																
			Temperature (°F)			Relative Humidity			Wind Chill (°F)			Heat Index (°F)			Pressure			Precipitation (in.)																										
			Air			Dwpt			6 hour			altimeter (in)			sea level (mb)			1 hr			3 hr			6 hr																				
04			15:51			W 14 G 37			10.00			Overcast			0VC080			60			35			39%			NA			NA			29.76			1006.7								
04			14:51			Vrbl 7 G 26			10.00			Mostly Cloudy			BKN110			63			35			35%			NA			NA			29.74			1006.0								
04			13:51			Vrbl 5			10.00			Overcast			0VC090			66			36			75			64			33%			NA			NA			29.70			1004.7		
04			12:51			Vrbl 3			10.00			Overcast			FEW080 SCT090 OVC110			70			36			29%			NA			NA			29.67			1003.7								
04			11:51			Vrbl 2			10.00			Overcast			0VC090 SCT090 OVC110			70			36			29%			NA			NA			29.67			1003.7								
04			10:51			Vrbl 1			10.00			Overcast			0VC090 SCT090 OVC110			70			36			29%			NA			NA			29.67			1003.7								
04			09:51			Vrbl 1			10.00			Overcast			0VC090 SCT090 OVC110			70			36			29%			NA			NA			29.67			1003.7								
04			08:51			Vrbl 1			10.00			Overcast			0VC090 SCT090 OVC110			70			36			29%			NA			NA			29.67			1003.7								
04			07:51			Vrbl 1			10.00			Overcast			0VC090 SCT090 OVC110			70			36			29%			NA			NA			29.67			1003.7								
04			06:51			Vrbl 1			10.00			Overcast			0VC090 SCT090 OVC110			70			36			29%			NA			NA			29.67			1003.7								
04			05:51			Vrbl 1			10.00			Overcast			0VC090 SCT090 OVC110			70			36			29%			NA			NA			29.67			1003.7								
04			04:51			Vrbl 1			10.00			Overcast			0VC090 SCT090 OVC110			70			36			29%			NA			NA			29.67			1003.7								
04			03:51			Vrbl 1			10.00			Overcast			0VC090 SCT090 OVC110			70			36			29%			NA			NA			29.67			1003.7								
04			02:51			Vrbl 1			10.00			Overcast			0VC090 SCT090 OVC110			70			36			29%			NA			NA			29.67			1003.7								
04			01:51			Vrbl 1			10.00			Overcast			0VC090 SCT090 OVC110			70			36			29%			NA			NA			29.67			1003.7								
04			00:51			Vrbl 1			10.00			Overcast			0VC090 SCT090 OVC110			70			36			29%			NA			NA			29.67			1003.7								
04			23:51			Vrbl 1			10.00			Overcast			0VC090 SCT090 OVC110			70			36			29%			NA			NA			29.67			1003.7								
04			22:51			Vrbl 1			10.00			Overcast			0																													

04	11:51	NA	10.00	Mostly Cloudy	SCT080	BKN095	70	37	30%	NA	NA	29.68	1004.0
04	10:51	Calm	10.00	Fair	CLR		68	37	32%	NA	NA	29.69	1004.5
04	09:51	W 7	10.00	Fair	CLR		67	40	37%	NA	NA	29.70	1004.7
04	08:51	Calm	10.00	Mostly Cloudy	BKN100		64	37	37%	NA	NA	29.70	1004.9
04	07:51	NE 5	10.00	Partly Cloudy	SCT100		64	39	69	63	40%	NA	NA
04	06:51	Vrbl 7	10.00	Overcast	OVC110		64	40	41%	NA	NA	29.70	1004.7
04	05:51	Calm	10.00	Overcast	OVC110		64	42	45%	NA	NA	29.68	1004.1
04	04:51	Calm	10.00	NA	NA		64	42	45%	NA	NA	29.66	1003.4
04	03:51	Vrbl 7	10.00	A Few Clouds	FEW100		66	43	43%	NA	NA	29.64	1002.9
04	02:51	Vrbl 6	10.00	Light Rain	FEW060	OVC095	68	45					

		44%	NA	NA	29.64	1002.7
	04		01:51	Vrbl 3	10.00	
	Partly Cloudy	SCT070	68	56		
	76	68	65%	NA	NA	29.63
	2.4					100
	04		00:51	SW 10 G 21	10.0	
	0					
	Mostly Cloudy	BKN095	70	56		
		61%	NA	NA	29.65	1003.0
	03		23:51	Vrbl 3	10.00	
	Fair	CLR	71	50		
		47%	NA	NA	29.64	1002.9
	03		22:51	SW 8	10.00	
	Fair	CLR	71	49		
		46%	NA	NA	29.66	1003.4
	03		21:51	Vrbl 6	10.00	
	Fair	CLR	71	49		
		46%	NA	NA	29.68	1004.1
	03		20:51	SW 6	10.00	
	Fair	CLR	74	48		
		40%	NA	NA	29.68	1004.0
	03		19:51	Vrbl 5	10.00	
	Fair	CLR	75	48		
	80	74	39%	NA	NA	29.67
	3.9					100
						3.9
	03		18:51	Vrbl 5	10.00	
	Fair	CLR	77	46		
		33%	NA	78	29.67	1003.8
	03		17:51	Vrbl 6 G 16	10.00	
	Fair	CLR	79	48		
		34%	NA	79	29.67	1003.9



03	16:51	SW 7	10.00	A Few Clouds	FEW080	79	48			34%	NA	79	29.68	1004.0
03	15:51	W 6	10.00	Fair	CLR	79	46			31%	NA	79	29.69	1004.4
03	14:51	Vrbl 6 G 18	10.00	A Few Clouds	FEW060	78	48			35%	NA	78	29.73	1005.7
03	13:51	SW 9	10.00	Fair	CLR	74	52	74	57	46%	NA	NA	29.74	1006.1
03	12:51	SW 5	10.00	Fair	CLR	72	54			53%	NA	NA	29.76	1007.0
03	11:51	W 7	10.00	Fair	CLR	68	53			59%	NA	NA	29.80	1008.0
03	10:51	W 8	10.00	Fair	CLR	64	52			65%	NA	NA	29.82	1008.9
03	09:51	Calm	10.00	Fair	CLR	60	53			78%	NA	NA	29.83	1009.1
03	08:51	SW 8	10.00	Fair	CLR	59	52			78%	NA	NA	29.85	1009.9
03	07:51	NA	10.00	Partly Cloudy	SCT100	57	51							

64	54	81%	NA	NA	29.85	1010.0
			0.04			
03	06:51	Calm	10.00			
Mostly Cloudy	BKN110	55	51			
		87%	NA	NA	29.85	1009.9
03	05:51	Vrbl 3	10.00			
Light Rain	FEW060 SCT090 OVC110	55	51			
		87%	NA	NA	29.85	1009.8
			0.03			
03	04:51	Vrbl 6	7.00			
Light Rain	SCT041 BKN050 OVC060	56	49			
		77%	NA	NA	29.86	1010.3
			0.01	0.01		
03	03:51	Vrbl 5	10.00			
Light Rain	OVC070	59	45			
		60%	NA	NA	29.86	1010.2
03	02:51	Vrbl 5	10.00			
Overcast	OVC075	63	41			
		45%	NA	NA	29.86	1010.4
03	01:51	Vrbl 3	10.00			
Light Rain	OVC080	63	41			
69	63	45%	NA	NA	29.91	1011.8
03	00:51	SW 7	10.00			
Overcast	OVC080	65	39			
		39%	NA	NA	29.92	1012.3
02	23:51	NA	10.00			
Overcast	OVC095	67	38			
		35%	NA	NA	29.92	1012.2
02	22:51	SW 5	10.00			
Mostly Cloudy	BKN100 BKN120	66	37			

		34%	NA	NA	29.94	1013.0
	02		21:51	Vrbl 6	10.00	
			Fair	CLR	66	37
		34%	NA	NA	29.94	1013.1
	02		20:51	SW 7	10.00	
			Partly Cloudy	SCT120	68	37
		32%	NA	NA	29.94	1012.9
	02		19:51	Vrbl 7	10.00	
			Fair	CLR	69	38
		32%	NA	NA	29.92	1012.2
	02		18:51	Vrbl 3	10.00	
			Fair	CLR	71	33
		25%	NA	NA	29.90	1011.6
	02		17:51	W 8	10.00	
			Fair	CLR	72	34
		25%	NA	NA	29.92	1012.4
	02		16:51	W 8	10.00	
			Fair	CLR	73	34
		24%	NA	NA	29.92	1012.4
	02		15:51	W 8 G 16	10.00	
			Fair	CLR	73	35
		25%	NA	NA	29.94	1013.0
	02		14:51	Vrbl 5	10.00	
			Fair	CLR	71	34
		26%	NA	NA	29.96	1013.8
	02		13:51	Vrbl 3	10.00	
			Fair	CLR	71	35
		27%	NA	NA	29.98	1014.2

02	12:51	NW 9	10.00
Fair	CLR	70	35
		28%	NA
		NA	30.00
			1015.0

bgcolor="#eeeeee" valign="top">02				11:51	Vrbl 6	10.00
Fair	CLR	68	38			
		33%	NA			
		NA	30.02			
			1015.6			

bgcolor="#f5f5f5" valign="top">02				10:51	NA	10.00
Fair	CLR	64	36			
		35%	NA			
		NA	30.03			
			1015.9			

bgcolor="#eeeeee" valign="top">02				09:51	N 8 G 16	10.00
Fair	CLR	61	35			
		38%	NA			
		NA	30.02			
			1015.6			

bgcolor="#f5f5f5" valign="top">02				08:51	NA	10.00
Fair	CLR	57	37			
		47%	NA			
		NA	30.01			
			1015.4			

bgcolor="#eeeeee" valign="top">02				07:51	W 5	10.00
Fair	CLR	53	38			
		56	51			
		57%	NA			
		NA	30.00			
			1014.9			

bgcolor="#f5f5f5" valign="top">02				06:51	NA	10.00
Fair	CLR	52	38			
		59%	NA			
		NA	29.98			
			1014.1			

bgcolor="#eeeeee" valign="top">02				05:51	Vrbl 6	10.00
Fair	CLR	52	37			
		57%	NA			
		NA	29.94			
			1012.8			

bgcolor="#f5f5f5" valign="top">02				04:51	W 7	10.00
Fair	CLR	52	37			
		57%	NA			
		NA	29.90			
			1011.6			

bgcolor="#eeeeee" valign="top">02				03:51	Vrbl 3	10.00
Fair	CLR	53	37			

		55%	NA	NA	29.88	1010.9	
02	02:51	NA	10.00	Fair	CLR	54	39
		57%	NA	NA	29.86	1010.2	
02	01:51	NA	10.00	Fair	CLR	56	43
63	56	62%	NA	NA	29.84	1009.5	
			0.01				
02	00:51	NW 10 G 18	10.00	Fair	CLR	58	46
		65%	NA	NA	29.82	1008.8	
01	23:51	Vrbl 7	10.00	Overcast	OVC085	59	45
		60%	NA	NA	29.80	1008.2	
			0.01				
01	22:51	Vrbl 3	10.00	Overcast	OVC085	61	47
		60%	NA	NA	29.78	1007.6	
01	21:51	W 9	10.00	Overcast	OVC080	62	48
		60%	NA	NA	29.77	1007.1	
01	20:51	W 5	10.00	Overcast	OVC075	62	48
		60%	NA	NA	29.75	1006.5	
01	19:51	Vrbl 3	10.00	Fair	CLR	63	47
66	61	56%	NA	NA	29.73	1005.8	
01	18:51	Vrbl 6	10.00	Mostly Cloudy	BKN041 BKN050	65	48
		54%	NA	NA	29.72	1005.3	

01	17:51	Vrbl 7		
	10.00			
Partly Cloudy	SCT041	SCT050		
	65	50		
	59%	NA		
	NA	29.70		
		1004.9		
01	16:51	Calm		
	10.00			
Mostly Cloudy	FEW035	SCT045		
	BKN080	65		
	51	61%		
	NA	NA		
	29.69	1004.6		
D a t e	Time (edt)	Wind (mph)		
			Vis. (mi.)	Weather
Air	Dwpt	Max.		
			Min.	Relative Humidity
Wind Chill (°F)	Heat Index (°F)	altimeter (in.)		
			sea level (mb)	1 hr
3 hr	6 hr			
		6 hour	Pressure	
Precipitation (in.)				

```

In [8]: # extract data from the 4th table in the page into a dataframe

import pandas as pd

data_table = knyc_soup.find_all('table')[3]

table_rows = data_table.find_all('tr') # get rows from table

data = []
for idx, tr in enumerate(table_rows):
    if idx < 3 :                               # skip header rows
        continue
    td = tr.find_all('td')                     # get table cells
    row = [elem.text for elem in td]            # pull text from cells
    data.append(row)                            # add to dataset

pd.DataFrame(data).head()

```

```

Out[8]:

```

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
0	04	15:51	W 14 G 37	10.00	Overcast	OVC080		60	35		39%	NA	NA	29.76	1006.7			
1	04	14:51	Vrbl 7 G 26	10.00	Mostly Cloudy	BKN110		63	35		35%	NA	NA	29.74	1006.0			
2	04	13:51	Vrbl 5	10.00	Overcast	OVC090		66	36	75	64	33%	NA	NA	29.70	1004.7		
3	04	12:51	Vrbl 3	10.00	Overcast	FEW080 SCT090 OVC110		70	36		29%	NA	NA	29.67	1003.7			
4	04	11:51	NA	10.00	Mostly Cloudy	SCT080 BKN095		70	37		30%	NA	NA	29.68	1004.0			

## Laguardia Weather History Summary

<https://www.wunderground.com/history/daily/us/ny/new-york-city/KNYC/date/2018-12-3?cm ven=localwx history>  
(<https://www.wunderground.com/history/daily/us/ny/new-york-city/KNYC/date/2018-12-3?cm ven=localwx history>).

```
In [9]: wu_link = 'https://www.wunderground.com/history/daily/us/ny/new-york-city/KLGA'
```



Civil Twilight	5:20 AM	8:20 PM
Nautical Twilight	4:43 AM	9:02 PM
Astronomical Twilight	4:03 AM	9:43 PM
Moon: waxing gibbous	4:37 PM	4:30 AM

## Daily Observations

Time	Temperature	Dew Point	Humidity	Wind	Wind Speed	Wind Gust	Press
12:51 AM	71 °F	55 °F	57 %	SW	10 mph	0 mph	29.59
1:51 AM	69 °F	54 °F	58 %	WNW	9 mph	0 mph	29.57
2:51 AM	67 °F	44 °F	44 %	N	10 mph	0 mph	29.58
3:51 AM	66 °F	43 °F	43 %	NNE	10 mph	0 mph	29.59
4:51 AM	64 °F	43 °F	46 %	ESE	8 mph	0 mph	29.61
5:51 AM	62 °F	44 °F	52 %	N	3 mph	0 mph	29.63
6:51 AM	61 °F	44 °F	54 %	NE	6 mph	0 mph	29.65
7:51 AM	65 °F	37 °F	36 %	NNE	7 mph	0 mph	29.65
8:51 AM	65 °F	36 °F	34 %	E	7 mph	0 mph	29.65
9:51 AM	67 °F	35 °F	31 %	NE	5 mph	0 mph	29.64
10:51 AM	68 °F	38 °F	33 %	WNW	9 mph	0 mph	29.64
11:51 AM	71 °F	36 °F	28 %	WNW	12 mph	0 mph	29.62
12:51 PM	69 °F	37 °F	31 %	NW	12 mph	0 mph	29.61
1:51 PM	66 °F	36 °F	33 %	NW	20 mph	30 mph	29.63
2:51 PM	63 °F	35 °F	35 %	NNW	18 mph	35 mph	29.68

```
Elements Console Sources Network Performance Memory Application Security Audits
<div _ngcontent-app-root-c234 class="region-content-top"></div>
<section _ngcontent-app-root-c148 class="section"></section>
<section _ngcontent-app-root-c148 class="section"></section>
<section _ngcontent-app-root-c148 class="section"></section>
<div _ngcontent-app-root-c234 class="region-content-main">
  <div _ngcontent-app-root-c234 class="row">
    ::before
    <div _ngcontent-app-root-c234 class="small-12 columns has-sidebar"></div>
    <div _ngcontent-app-root-c234 class="columns small-12"></div>
    <div _ngcontent-app-root-c234 class="small-12 columns has-sidebar"></div>
    <div _ngcontent-app-root-c234 class="columns small-12"></div>
    <div _ngcontent-app-root-c234 class="small-12 columns has-sidebar">
      <div _ngcontent-app-root-c234 class="row">
        ::before
        <div _ngcontent-app-root-c234 class="columns small-12">
          <div _ngcontent-app-root-c203 class="lib-city-history-observation _ngcontent-app-root-c234 _ngghost-app-root-c203">
            <div _ngcontent-app-root-c203">
              <div _ngcontent-app-root-c203 class="observation-title">Daily Observations</div>
              <div _ngcontent-app-root-c203 class="observation-table ng-star-inserted">
                <table _ngcontent-app-root-c203 mat-table mat-sort aria-label="table of contents" class="mat-table cdk-table mat-sort ng-star-inserted" role="grid">
                  <thead role="rowgroup"></thead>
                  <tbody role="rowgroup">
                    <tr _ngcontent-app-root-c203 role="row" mat-row class="mat-row cdk-row ng-star-inserted">
                      <td _ngcontent-app-root-c203 role="gridcell" mat-cell class="mat-cell cdk-cell cdk-column-dateString mat-column-dateString mat-table-sticky ng-star-inserted" style="left: 0px; position: sticky; z-index: 1;">
                        <div _ngcontent-app-root-c203 class="ng-star-inserted">12:51 AM</div>
                      </td>
                      <td _ngcontent-app-root-c203 role="gridcell" mat-cell class="mat-cell cdk-cell cdk-column-temperature mat-column-temperature ng-star-inserted"></td>
                      <td _ngcontent-app-root-c203 role="gridcell" mat-cell class="mat-cell cdk-cell cdk-column-dewPoint mat-column-dewPoint ng-star-inserted"></td>
                      <td _ngcontent-app-root-c203 role="gridcell" mat-cell class="mat-cell cdk-cell cdk-column-humidity mat-column-humidity ng-star-inserted"></td>
                      <td _ngcontent-app-root-c203 role="gridcell" mat-cell class="mat-cell cdk-cell cdk-column-windcardinal mat-column-windcardinal ng-star-inserted"></td>
                      <td _ngcontent-app-root-c203 role="gridcell" mat-cell class="mat-cell cdk-cell cdk-column-windSpeed mat-column-windSpeed ng-star-inserted"></td>
                      <td _ngcontent-app-root-c203 role="gridcell" mat-cell class="mat-cell cdk-cell cdk-column-windGust mat-column-windGust ng-star-inserted"></td>
                      <td _ngcontent-app-root-c203 role="gridcell" mat-cell class="mat-cell cdk-cell cdk-column-pressure mat-column-pressure ng-star-inserted"></td>
                      <td _ngcontent-app-root-c203 role="gridcell" mat-cell class="mat-cell cdk-cell cdk-column-precipRate mat-column-precipRate ng-star-inserted"></td>
                      <td _ngcontent-app-root-c203 role="gridcell" mat-cell class="mat-cell cdk-cell cdk-column-condition mat-column-condition ng-star-inserted"></td>
                    </tr>
                  </tbody>
                </table>
              </div>
            </div>
          </div>
        </div>
      </div>
    </div>
  </div>
</div>
```

```
In [10]: # get the page  
import requests  
wu_page = requests.get(wu_link)  
wu_page
```

```
Out[10]: <Response [200]>
```

```
In [11]: from bs4 import BeautifulSoup
wu_soup = BeautifulSoup(wu_page.content)
```

```
In [12]: print(wu_soup.prettify()[:1000])
```

```
<!DOCTYPE html>
<html itemscope="" itemtype="http://schema.org/Organization" lang="en" prefix
="og: http://ogp.me/ns#">
  <head itemscope="" itemtype="http://schema.org/WebSite">
    <meta charset="utf-8"/>
    <title>
      New York City, NY Weather History | Weather Underground
    </title>
    <base href="/" />
    <meta content="width=device-width, initial-scale=1" name="viewport"/>
    <meta content="IE=edge,chrome=1" http-equiv="X-UA-Compatible"/>
    <link href="//widgets.outbrain.com" rel="dns-prefetch"/>
    <link href="//odb.outbrain.com" rel="dns-prefetch"/>
    <link href="//c.amazon-adsystem.com" rel="dns-prefetch"/>
    <link href="//s.amazon-adsystem.com" rel="dns-prefetch"/>
    <link href="//aax.amazon-adsystem.com" rel="dns-prefetch"/>
    <link href="//partner.googleadservices.com" rel="dns-prefetch"/>
    <link href="//tpc.googlesyndication.com" rel="dns-prefetch"/>
    <link href="//pagead2.googlesyndication.com" rel="dns-prefetch"/>
    <link href="//h.nexac.com" rel="dns-prefetch"/>
    <link href="/
```

```
In [13]: # the table we want doesn't exist! culprit: javascript
wu_soup.find_all('table', class_=lambda value: value and value.startswith("mat-ta
ble"))
```

```
Out[13]: []
```

```
In [14]: # get the text from the page
wu_text = wu_soup.get_text()

# clean up the whitespace
import re
wu_text = re.sub(r'\n+', '\n', wu_text.strip())
print(wu_text[-1000:])
```

```
nTime=timing.domainLookupEnd-timing.domainLookupStart;api.connectTime=timing.c
onnectEnd-timing.connectStart;api.requestTime=timing.responseEnd-timing.reques
tStart;api.initDomTreeTime=timing.domInteractive-timing.responseEnd;api.loadEv
entTime=timing.loadEventEnd-timing.loadEventStart}return api},printTable:funct
ion(opts){var table={};var data=this.getTimes(opts)||{};Object.keys(data).sort
().forEach(function(k){table[k]={ms:data[k],s:+(data[k]/1e3).toFixed(2)}});con
sole.table(table)},printSimpleTable:function(){this.printTable({simple:true
e}})};function isNumeric(n){return!isNaN(parseFloat(n))&&isFinite(n)}if(typeof
module!="undefined"&&module.exports){module.exports=window.timing}})(typeof w
indow!="undefined"?window:{});
```

```
    window.addEventListener("load",function(){setTimeout(function(){newrelic
&&timing&&timing.getTimes())&&Object.keys(timing.getTimes()).forEach(function
(i){newrelic.setCustomAttribute("timing"+i.charAt(0).toUpperCase()+i.slice(1),
timing.getTimes()[i])}}),0)});
```

## Need to actually render page to process scripts!

```
In [15]: # need to install chromedriver: https://sites.google.com/a/chromium.org/chromedriver/home
from selenium.webdriver.chrome.options import Options
from selenium import webdriver

chrome_options = Options()
chrome_options.add_argument("--headless")

driver = webdriver.Chrome(options=chrome_options)
```

```
In [16]: # this will actually render the page
driver.get(wu_link)
```

```
In [18]: # two ways to find the table we want
wu_table = driver.find_element_by_class_name('mat-table')
#wu_table = driver.find_element_by_id('history-observation-table')
```

```
In [19]: # text in the table
wu_table.text
```

```
Out[19]: 'Time\nTemperature\nDew Point\nHumidity\nWind\nWind Speed\nWind Gust\nPressure\nPrecip.\nCondition\n12:51 AM 71 F 55 F 57 % SW 10 mph 0 mph 29.59 in 0.0 in Cloudy\n1:51 AM 69 F 54 F 58 % WNW 9 mph 0 mph 29.57 in 0.0 in Mostly Cloudy\n2:51 AM 67 F 44 F 44 % N 10 mph 0 mph 29.58 in 0.0 in Cloudy\n3:51 AM 66 F 43 F 43 % NNE 10 mph 0 mph 29.59 in 0.0 in Fair\n4:51 AM 64 F 43 F 46 % ESE 8 mph 0 mph 29.61 in 0.0 in Cloudy\n5:51 AM 62 F 44 F 52 % N 3 mph 0 mph 29.63 in 0.0 in Cloudy\n6:51 AM 61 F 44 F 54 % NE 6 mph 0 mph 29.65 in 0.0 in Mostly Cloudy\n7:51 AM 65 F 37 F 36 % NNE 7 mph 0 mph 29.65 in 0.0 in Mostly Cloudy\n8:51 AM 65 F 36 F 34 % E 7 mph 0 mph 29.65 in 0.0 in Partly Cloudy\n9:51 AM 67 F 35 F 31 % NE 5 mph 0 mph 29.64 in 0.0 in Fair\n10:51 AM 68 F 38 F 33 % WNW 9 mph 0 mph 29.64 in 0.0 in Fair\n11:51 AM 71 F 36 F 28 % WNW 12 mph 0 mph 29.62 in 0.0 in Partly Cloudy\n12:51 PM 69 F 37 F 31 % NW 12 mph 0 mph 29.61 in 0.0 in Mostly Cloudy\n1:51 PM 66 F 36 F 33 % NW 20 mph 30 mph 29.63 in 0.0 in Mostly Cloudy\n2:51 PM 63 F 35 F 35 % NNW 18 mph 35 mph 29.68 in 0.0 in Mostly Cloudy\n3:51 PM 61 F 36 F 39 % WNW 26 mph 35 mph 29.70 in 0.0 in Cloudy / Windy'
```

```
In [24]: # extracting text into a dataframe
wu_data = []
for tr in wu_table.find_elements_by_css_selector('tr'):
    tmp_row = []
    for th in tr.find_elements_by_css_selector('th'):
        tmp_row.append(th.text.strip())
    for td in tr.find_elements_by_css_selector('td'):
        tmp_row.append(td.text.strip())
    wu_data.append(tmp_row)
df_wu = pd.DataFrame(wu_data[1:], columns=wu_data[0])
df_wu.head()
```

Out[24]:

	Time	Temperature	Dew Point	Humidity	Wind	Wind Speed	Wind Gust	Pressure	Precip.	Condition
0	12:51 AM	71 F	55 F	57 %	SW	10 mph	0 mph	29.59 in	0.0 in	Cloudy
1	1:51 AM	69 F	54 F	58 %	WNW	9 mph	0 mph	29.57 in	0.0 in	Mostly Cloudy
2	2:51 AM	67 F	44 F	44 %	N	10 mph	0 mph	29.58 in	0.0 in	Cloudy
3	3:51 AM	66 F	43 F	43 %	NNE	10 mph	0 mph	29.59 in	0.0 in	Fair
4	4:51 AM	64 F	43 F	46 %	ESE	8 mph	0 mph	29.61 in	0.0 in	Cloudy

In [25]: `# visualize the rendered table, still missing some stuff, need to debug`  
`wu_table.screenshot('./images/test1.png')`

Out[25]: True

Time	Temperature	Dew Point	Humidity	Wind	Wind Speed	Wind Gust	Pressure	Pre
12:51 AM	71 °F	55 °F	57 %	SW	10 mph	0 mph	29.59 in	0.0
1:51 AM	69 °F	54 °F	58 %	WNW	9 mph	0 mph	29.57 in	0.0
2:51 AM	67 °F	44 °F	44 %	N	10 mph	0 mph	29.58 in	0.0
3:51 AM	66 °F	43 °F	43 %	NNE	10 mph	0 mph	29.59 in	0.0
4:51 AM	64 °F	43 °F	46 %	ESE	8 mph	0 mph	29.61 in	0.0
5:51 AM	62 °F	44 °F	52 %	N	3 mph	0 mph	29.63 in	0.0
6:51 AM	61 °F	44 °F	54 %	NE	6 mph	0 mph	29.65 in	0.0
7:51 AM	65 °F	37 °F	36 %	NNE	7 mph	0 mph	29.65 in	0.0
8:51 AM	65 °F	36 °F	34 %	E	7 mph	0 mph	29.65 in	0.0
9:51 AM	67 °F	35 °F	31 %	NE	5 mph	0 mph	29.64 in	0.0
10:51 AM	68 °F	38 °F	33 %	WNW	9 mph	0 mph	29.64 in	0.0
11:51 AM	71 °F	36 °F	28 %	WNW	12 mph	0 mph	29.62 in	0.0
12:51 PM	69 °F	37 °F	31 %	NW	12 mph	0 mph	29.61 in	0.0
1:51 PM	66 °F	36 °F	33 %	NW	20 mph	30 mph	29.63 in	0.0

 See our [Privacy Policy](#) to learn about the types of data we collect and how we use and share it. We collect data from your browser 