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 [17]:

```
knyc_link = 'http://w1.weather.gov/data/obhistory/KNYC.html'
```

In [18]:

```
import requests

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

Out[18]:

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

In [20]:

```
# need to parse some html!
from bs4 import BeautifulSoup
```

In [21]:

```
knyc_soup = BeautifulSoup(knyc_page.content)
```

In [22]:

```
# 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="http://www.srh.noaa.gov/weather/images/fcicons/main.css" rel="ST
YLESHEET" type="text/css"/>
 </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" 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" b
```

In [23]:

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

```
<table border="0" cellpadding="2" cellspacing="3" width="670"><tr align="cente
r" bgcolor="#b0c4de"><th rowspan="3" width="17">D<br/>a<br/>t<br/>e</th><th ro
wspan="3" width="32">Time<br/>(est)</th>
<th rowspan="3" width="80">Wind<br/>(mph)</th><th rowspan="3" width="40">Vis.<
br/>(mi.)</th><th rowspan="3" width="80">Weather</th><th rowspan="3" width="6
5">Sky Cond.</th>
<th colspan="4">Temperature (ºF)</th><th rowspan="3" width="65">Relative<br/>H
umidity</th><th rowspan="3" width="80">Wind<br/>Chill<br/>(°F)</th><th rowspan
="3" width="80">Heat<br/>Index<br/>(°F)</th><th colspan="2">Pressure</th><th c
olspan="3">Precipitation (in.)</th></tr>
<tr align="center" bgcolor="#b0c4de"><th rowspan="2" width="45">Air</th><th ro
wspan="2" width="26">Dwpt</th><th colspan="2">6 hour</th>
<th rowspan="2" width="40">altimeter<br/>(in)</th><th rowspan="2" width="40">s
ea level<br/>(mb)</th><th rowspan="2" width="24">1 hr</th>
<th rowspan="2" width="24">3 hr</th><th rowspan="2" width="30">6 hr</th></tr>
<tr align="center" bgcolor="#b0c4de"><th width="26">Max.</th><th width="26">Mi
n.</th></tr><tr align="center" bgcolor="#eeeeee" valign="top"><td>03</td><td a
lign="right">13:51</td><td>NA</td><td>10.00</td><td align="left">A Few Clouds
</td><td>FEW050</td><td>53</td><td>33</td>
<td></td><td></td><td>47%</td><td>NA</td><td>NA</td><td>29.66</td><td>1003.7</
td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#f5f5f5" valign
="top"><td>03</td><td align="right">12:51</td><td>NA</td><td>10.00</td><td ali
gn="left">Overcast</td><td>OVC049</td><td>52</td><td>34</td>
<td>55</td><td>52</td><td>50%</td><td>NA</td><td>NA</td><td>29.67</td><td>100
3.8</td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#eeeeee" v
align="top"><td>03</td><td align="right">11:51</td><td>NA</td><td>10.00</td><t
d align="left">Overcast</td><td>OVC037</td><td>52</td><td>37</td>
<td></td><td></td><td>57%</td><td>NA</td><td>NA</td><td>29.67</td><td>1003.9</
td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#f5f5f5" valign
="top"><td>03</td><td align="right">10:51</td><td>NA</td><td>10.00</td><td ali
gn="left">A Few Clouds</td><td>FEW037</td><td>54</td><td>39</td>
<td></td><td></td><td>57%</td><td>NA</td><td>NA</td><td>29.67</td><td>1003.9</
td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#eeeeee" valign
="top"><td>03</td><td align="right">09:51</td><td>NA</td><td>10.00</td><td ali
```

```
="top"><td>03</td><td align="right">09:51</td><td>NA</td><td>10.00</td><td ali
gn="left">Fair</td><td>CLR</td><td>54</td><td>41</td>
<td></td><td></td><td>62%</td><td>NA</td><td>NA</td><td>29.67</td><td>1003.8</
td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#f5f5f5" valign
="top"><td>03</td><td align="right">08:51</td><td>NA</td><td>10.00</td><td ali
gn="left">Fair</td><td>CLR</td><td>54</td><td>44</td>
<td></td><td></td><td>69%</td><td>NA</td><td>NA</td><td>29.66</td><td>1003.3</
td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#eeeeee" valign
="top"><td>03</td><td align="right">07:51</td><td>NA</td><td>10.00</td><td ali
gn="left">Fair</td><td>CLR</td><td>54</td><td>46</td>
<td></td><td></td><td>75%</td><td>NA</td><td>NA</td><td>29.63</td><td>1002.6</
td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#f5f5f5" valign
="top"><td>03</td><td align="right">06:51</td><td>NA</td><td>10.00</td><td ali
gn="left">Fair</td><td>CLR</td><td>53</td><td>49</td>
<td>55</td><td>53</td><td>86%</td><td>NA</td><td>NA</td><td>29.61</td><td>100
2.0</td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#eeeeee" v
align="top"><td>03</td><td align="right">05:51</td><td>NA</td><td>10.00</td><t
d align="left">Partly Cloudy</td><td>SCT013</td><td>54</td><td>51</td>
<td></td><td></td><td>90%</td><td>NA</td><td>NA</td><td>29.60</td><td>1001.4</
td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#f5f5f5" valign
="top"><td>03</td><td align="right">04:51</td><td>NA</td><td>5.00</td><td alig
n="left"> Fog/Mist</td><td>OVC004</td><td>54</td><td>53</td>
<td></td><td></td><td>97%</td><td>NA</td><td>NA</td><td>29.60</td><td>1001.5</
td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#eeeeee" valign
="top"><td>03</td><td align="right">03:51</td><td>NA</td><td>3.00</td><td alig
n="left"> Fog/Mist</td><td>OVC004</td><td>54</td><td>53</td>
<td></td><td></td><td>97%</td><td>NA</td><td>NA</td><td>29.60</td><td>1001.6</
td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#f5f5f5" valign
="top"><td>03</td><td align="right">02:51</td><td>NA</td><td>3.00</td><td alig
n="left"> Fog/Mist</td><td>OVC006</td><td>55</td><td>53</td>
<td></td><td></td><td>93%</td><td>NA</td><td>NA</td><td>29.61</td><td>1001.6</
td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#eeeeee" valign
="top"><td>03</td><td align="right">01:51</td><td>NA</td><td>8.00</td><td alig
n="left">Overcast</td><td>OVC008</td><td>55</td><td>52</td>
<td></td><td></td><td>90%</td><td>NA</td><td>NA</td><td>29.61</td><td>1001.9</
td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#f5f5f5" valign
="top"><td>03</td><td align="right">00:51</td><td>NA</td><td>8.00</td><td alig
n="left">Overcast</td><td>OVC007</td><td>55</td><td>52</td>
<td>56</td><td>54</td><td>90%</td><td>NA</td><td>NA</td><td>29.62</td><td>100
```

```
2.0</td><td></td><td></td><td>0.03</td></tr><tr align="center" bgcolor="#eeeee
e" valign="top"><td>02</td><td align="right">23:51</td><td>NA</td><td>9.00</td
><td align="left">Overcast</td><td>OVC008</td><td>55</td><td>53</td>
<td></td><td></td><td>93%</td><td>NA</td><td>NA</td><td>29.62</td><td>1002.3</
td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#f5f5f5" valign
="top"><td>02</td><td align="right">22:51</td><td>NA</td><td>9.00</td><td alig
n="left">Overcast</td><td>OVC010</td><td>55</td><td>53</td>
<td></td><td></td><td>93%</td><td>NA</td><td>NA</td><td>29.62</td><td>1002.2</
td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#eeeeee" valign
="top"><td>02</td><td align="right">21:51</td><td>NA</td><td>7.00</td><td alig
n="left">Overcast</td><td>OVC007</td><td>55</td><td>54</td>
<td></td><td></td><td>96%</td><td>NA</td><td>NA</td><td>29.63</td><td>1002.5</
td><td>0.01</td><td>0.03</td><td></td></tr><tr align="center" bgcolor="#f5f5f
5" valign="top"><td>02</td><td align="right">20:51</td><td>NA</td><td>1.75</td
><td align="left"> Fog/Mist</td><td>OVC009</td><td>55</td><td>54</td>
<td></td><td></td><td>96%</td><td>NA</td><td>NA</td><td>29.64</td><td>1003.0</
td><td>0.01</td><td></td><td></td></tr><tr align="center" bgcolor="#eeeeee" va
lign="top"><td>02</td><td align="right">19:51</td><td>NA</td><td>2.50</td><td
align="left"> Fog/Mist</td><td>OVC008</td><td>56</td><td>54</td>
<td></td><td></td><td>93%</td><td>NA</td><td>NA</td><td>29.64</td><td>1002.8</
td><td>0.01</td><td></td><td></td></tr><tr align="center" bgcolor="#f5f5f5" va
lign="top"><td>02</td><td align="right">18:51</td><td>NA</td><td>1.25</td><td
align="left"> Light Rain Fog/Mist</td><td>OVC007</td><td>54</td><td>53</td>
<td>55</td><td>52</td><td>97%</td><td>NA</td><td>NA</td><td>29.63</td><td>100
2.7</td><td>0.02</td><td></td><td>0.07</td></tr><tr align="center" bgcolor="#e
eeeee" valign="top"><td>02</td><td align="right">17:51</td><td>NA</td><td>1.75
</td><td align="left"> Fog/Mist</td><td>OVC010</td><td>54</td><td>53</td>
<td></td><td></td><td>97%</td><td>NA</td><td>NA</td><td>29.63</td><td>1002.7</
td><td>0.01</td><td></td><td></td></tr><tr align="center" bgcolor="#f5f5f5" va
lign="top"><td>02</td><td align="right">16:51</td><td>NA</td><td>1.50</td><td
align="left"> Fog/Mist</td><td>OVC007</td><td>53</td><td>52</td>
<td></td><td></td><td>96%</td><td>NA</td><td>NA</td><td>29.64</td><td>1002.8</
td><td>0.01</td><td></td><td></td></tr><tr align="center" bgcolor="#eeeeee" va
lign="top"><td>02</td><td align="right">15:51</td><td>NA</td><td>0.75</td><td
align="left"> Light Rain Fog/Mist</td><td>VV006</td><td>53</td><td>52</td>
<td></td><td></td><td>96%</td><td>NA</td><td>NA</td><td>29.65</td><td>1003.3</
td><td>0.02</td><td>0.03</td><td></td></tr><tr align="center" bgcolor="#f5f5f
5" valign="top"><td>02</td><td align="right">14:51</td><td>NA</td><td>1.25</td
```

```
><td align="left"> Light Rain Fog/Mist</td><td>OVC005</td><td>54</td><td>52</t
d>
<td></td><td></td><td>93%</td><td>NA</td><td>NA</td><td>29.66</td><td>1003.4</
td><td>0.01</td><td></td><td></td></tr><tr align="center" bgcolor="#eeeeee" va
lign="top"><td>02</td><td align="right">13:51</td><td>NA</td><td>1.25</td><td
align="left"> Fog/Mist</td><td>OVC004</td><td>53</td><td>52</td>
<td></td><td></td><td>96%</td><td>NA</td><td>NA</td><td>29.67</td><td>1003.9</
td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#f5f5f5" valign
="top"><td>02</td><td align="right">12:51</td><td>NA</td><td>1.00</td><td alig
n="left"> Fog/Mist</td><td>OVC003</td><td>52</td><td>51</td>
<td>52</td><td>46</td><td>97%</td><td>NA</td><td>NA</td><td>29.70</td><td>100
4.9</td><td>0.01</td><td></td><td>0.20</td></tr><tr align="center" bgcolor="#e
eeeee" valign="top"><td>02</td><td align="right">11:51</td><td>NA</td><td>1.25
</td><td align="left"> Fog/Mist</td><td>OVC003</td><td>52</td><td>51</td>
<td></td><td></td><td>97%</td><td>NA</td><td>NA</td><td>29.72</td><td>1005.6</
td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#f5f5f5" valign
="top"><td>02</td><td align="right">10:51</td><td>NA</td><td>1.50</td><td alig
n="left"> Fog/Mist</td><td>OVC004</td><td>50</td><td>49</td>
<td></td><td></td><td>96%</td><td>NA</td><td>NA</td><td>29.76</td><td>1007.0</
td><td>0.04</td><td></td><td></td></tr><tr align="center" bgcolor="#eeeeee" va
lign="top"><td>02</td><td align="right">09:51</td><td>NA</td><td>1.00</td><td
align="left"> Light Rain Fog/Mist</td><td>OVC004</td><td>49</td><td>48</td>
<td></td><td></td><td>97%</td><td>NA</td><td>NA</td><td>29.80</td><td>1008.3</
td><td>0.06</td><td>0.15</td><td></td></tr><tr align="center" bgcolor="#f5f5f
5" valign="top"><td>02</td><td align="right">08:51</td><td>NA</td><td>1.50</td
><td align="left"> Light Rain Fog/Mist</td><td>OVC004</td><td>48</td><td>47</t
d>
<td></td><td></td><td>96%</td><td>NA</td><td>NA</td><td>29.82</td><td>1008.8</
td><td>0.04</td><td></td><td></td></tr><tr align="center" bgcolor="#eeeeee" va
lign="top"><td>02</td><td align="right">07:51</td><td>NA</td><td>2.50</td><td
align="left"> Light Rain Fog/Mist</td><td>OVC005</td><td>47</td><td>45</td>
<td></td><td></td><td>93%</td><td>NA</td><td>NA</td><td>29.84</td><td>1009.8</
td><td>0.05</td><td></td><td></td></tr><tr align="center" bgcolor="#f5f5f5" va
lign="top"><td>02</td><td align="right">06:51</td><td>NA</td><td>2.00</td><td
align="left"> Rain Fog/Mist</td><td>OVC004</td><td>46</td><td>44</td>
<td>46</td><td>44</td><td>93%</td><td>NA</td><td>NA</td><td>29.88</td><td>101
1.0</td><td>0.05</td><td></td><td>0.29</td></tr><tr align="center" bgcolor="#e
eeeee" valign="top"><td>02</td><td align="right">05:51</td><td>NA</td><td>4.00
```

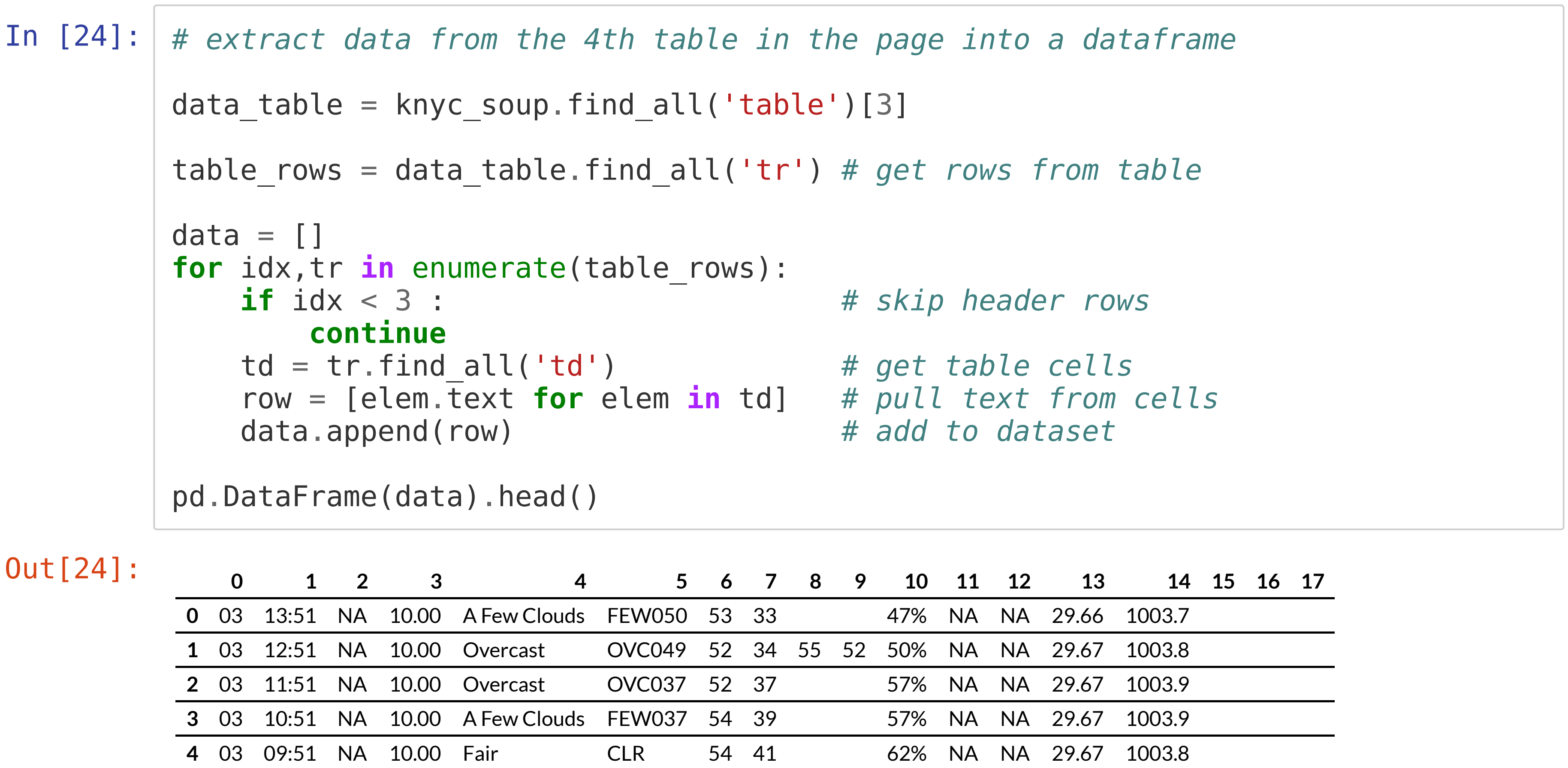
```
</td><td align="left"> Light Rain Fog/Mist</td><td>OVC004</td><td>45</td><td>4
4</td>
<td></td><td></td><td>97%</td><td>NA</td><td>NA</td><td>29.90</td><td>1011.6</
td><td>0.02</td><td></td><td></td></tr><tr align="center" bgcolor="#f5f5f5" va
lign="top"><td>02</td><td align="right">04:51</td><td>NA</td><td>3.00</td><td
align="left"> Light Rain Fog/Mist</td><td>OVC004</td><td>45</td><td>43</td>
<td></td><td></td><td>93%</td><td>NA</td><td>NA</td><td>29.94</td><td>1013.1</
td><td>0.06</td><td></td><td></td></tr><tr align="center" bgcolor="#eeeeee" va
lign="top"><td>02</td><td align="right">03:51</td><td>NA</td><td>1.75</td><td
align="left"> Rain Fog/Mist</td><td>OVC005</td><td>44</td><td>43</td>
<td></td><td></td><td>96%</td><td>NA</td><td>NA</td><td>29.98</td><td>1014.3</
td><td>0.12</td><td>0.16</td><td></td></tr><tr align="center" bgcolor="#f5f5f
5" valign="top"><td>02</td><td align="right">02:51</td><td>NA</td><td>3.00</td
><td align="left"> Light Rain Fog/Mist</td><td>OVC004</td><td>44</td><td>43</t
d>
<td></td><td></td><td>96%</td><td>NA</td><td>NA</td><td>30.00</td><td>1015.0</
td><td>0.02</td><td></td><td></td></tr><tr align="center" bgcolor="#eeeeee" va
lign="top"><td>02</td><td align="right">01:51</td><td>NA</td><td>4.00</td><td
align="left"> Light Rain Fog/Mist</td><td>OVC005</td><td>44</td><td>42</td>
<td></td><td></td><td>93%</td><td>NA</td><td>NA</td><td>30.03</td><td>1015.9</
td><td>0.02</td><td></td><td></td></tr><tr align="center" bgcolor="#f5f5f5" va
lign="top"><td>02</td><td align="right">00:51</td><td>NA</td><td>5.00</td><td
align="left"> Light Rain Fog/Mist</td><td>OVC005</td><td>43</td><td>41</td>
<td>43</td><td>41</td><td>93%</td><td>NA</td><td>NA</td><td>30.06</td><td>101
6.9</td><td></td><td></td><td>0.03</td></tr><tr align="center" bgcolor="#eeeee
e" valign="top"><td>01</td><td align="right">23:51</td><td>NA</td><td>4.00</td
><td align="left"> Fog/Mist</td><td>OVC006</td><td>42</td><td>40</td>
<td></td><td></td><td>92%</td><td>NA</td><td>NA</td><td>30.10</td><td>1018.3</
td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#f5f5f5" valign
="top"><td>01</td><td align="right">22:51</td><td>NA</td><td>8.00</td><td alig
n="left"> Light Rain</td><td>OVC013</td><td>42</td><td>39</td>
<td></td><td></td><td>89%</td><td>NA</td><td>NA</td><td>30.14</td><td>1019.7</
td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#eeeeee" valign
="top"><td>01</td><td align="right">21:51</td><td>NA</td><td>10.00</td><td ali
gn="left">Overcast</td><td>BKN017 OVC025</td><td>42</td><td>37</td>
<td></td><td></td><td>82%</td><td>NA</td><td>NA</td><td>30.16</td><td>1020.5</
td><td></td><td>0.03</td><td></td></tr><tr align="center" bgcolor="#f5f5f5" va
lign="top"><td>01</td><td align="right">20:51</td><td>NA</td><td>10.00</td><td
```

```
align="left"> Light Rain</td><td>SCT024 OVC039</td><td>41</td><td>37</td>
<td></td><td></td><td>86%</td><td>NA</td><td>NA</td><td>30.18</td><td>1021.1</
td><td>0.02</td><td></td><td></td></tr><tr align="center" bgcolor="#eeeeee" va
lign="top"><td>01</td><td align="right">19:51</td><td>NA</td><td>7.00</td><td
align="left"> Light Rain</td><td>OVC050</td><td>42</td><td>36</td>
<td></td><td></td><td>79%</td><td>NA</td><td>NA</td><td>30.18</td><td>1021.1</
td><td>0.01</td><td></td><td></td></tr><tr align="center" bgcolor="#f5f5f5" va
lign="top"><td>01</td><td align="right">18:51</td><td>NA</td><td>7.00</td><td
align="left"> Light Rain</td><td>BKN060 OVC080</td><td>42</td><td>37</td>
<td>46</td><td>42</td><td>82%</td><td>NA</td><td>NA</td><td>30.18</td><td>102
1.2</td><td>0.02</td><td></td><td>0.02</td></tr><tr align="center" bgcolor="#e
eeeee" valign="top"><td>01</td><td align="right">17:51</td><td>NA</td><td>10.0
0</td><td align="left">Mostly Cloudy</td><td>BKN080</td><td>43</td><td>31</td>
<td></td><td></td><td>63%</td><td>NA</td><td>NA</td><td>30.19</td><td>1021.5</
td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#f5f5f5" valign
="top"><td>01</td><td align="right">16:51</td><td>NA</td><td>10.00</td><td ali
gn="left">Fair</td><td>CLR</td><td>44</td><td>32</td>
<td></td><td></td><td>63%</td><td>NA</td><td>NA</td><td>30.20</td><td>1021.7</
td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#eeeeee" valign
="top"><td>01</td><td align="right">15:51</td><td>NA</td><td>10.00</td><td ali
gn="left">Fair</td><td>CLR</td><td>45</td><td>30</td>
<td></td><td></td><td>56%</td><td>NA</td><td>NA</td><td>30.20</td><td>1021.8</
td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#f5f5f5" valign
="top"><td>01</td><td align="right">14:51</td><td>NA</td><td>10.00</td><td ali
gn="left">Fair</td><td>CLR</td><td>45</td><td>31</td>
<td></td><td></td><td>58%</td><td>NA</td><td>NA</td><td>30.20</td><td>1021.8</
td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#eeeeee" valign
="top"><td>01</td><td align="right">13:51</td><td>NA</td><td>10.00</td><td ali
gn="left">A Few Clouds</td><td>FEW031</td><td>46</td><td>31</td>
<td></td><td></td><td>56%</td><td>NA</td><td>NA</td><td>30.19</td><td>1021.7</
td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#f5f5f5" valign
="top"><td>01</td><td align="right">12:51</td><td>NA</td><td>10.00</td><td ali
gn="left">Fair</td><td>CLR</td><td>45</td><td>31</td>
<td>45</td><td>37</td><td>58%</td><td>NA</td><td>NA</td><td>30.21</td><td>102
2.1</td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#eeeeee" v
align="top"><td>01</td><td align="right">11:51</td><td>NA</td><td>10.00</td><t
d align="left">Fair</td><td>CLR</td><td>44</td><td>31</td>
<td></td><td></td><td>60%</td><td>NA</td><td>NA</td><td>30.23</td><td>1022.9</
```

```
td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#f5f5f5" valign
="top"><td>01</td><td align="right">10:51</td><td>NA</td><td>10.00</td><td ali
gn="left">Partly Cloudy</td><td>SCT025</td><td>44</td><td>32</td>
<td></td><td></td><td>63%</td><td>NA</td><td>NA</td><td>30.24</td><td>1023.1</
td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#eeeeee" valign
="top"><td>01</td><td align="right">09:51</td><td>NA</td><td>10.00</td><td ali
gn="left">Fair</td><td>CLR</td><td>42</td><td>32</td>
<td></td><td></td><td>68%</td><td>NA</td><td>NA</td><td>30.24</td><td>1023.2</
td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#f5f5f5" valign
="top"><td>01</td><td align="right">08:51</td><td>NA</td><td>10.00</td><td ali
gn="left">Fair</td><td>CLR</td><td>40</td><td>32</td>
<td></td><td></td><td>73%</td><td>NA</td><td>NA</td><td>30.24</td><td>1023.1</
td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#eeeeee" valign
="top"><td>01</td><td align="right">07:51</td><td>NA</td><td>10.00</td><td ali
gn="left">Fair</td><td>CLR</td><td>38</td><td>32</td>
<td></td><td></td><td>79%</td><td>NA</td><td>NA</td><td>30.23</td><td>1023.0</
td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#f5f5f5" valign
="top"><td>01</td><td align="right">06:51</td><td>NA</td><td>10.00</td><td ali
gn="left">Fair</td><td>CLR</td><td>37</td><td>33</td>
<td>40</td><td>36</td><td>86%</td><td>NA</td><td>NA</td><td>30.21</td><td>102
2.2</td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#eeeeee" v
align="top"><td>01</td><td align="right">05:51</td><td>NA</td><td>9.00</td><td
align="left">Fair</td><td>CLR</td><td>37</td><td>32</td>
<td></td><td></td><td>82%</td><td>NA</td><td>NA</td><td>30.20</td><td>1021.7</
td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#f5f5f5" valign
="top"><td>01</td><td align="right">04:51</td><td>NA</td><td>10.00</td><td ali
gn="left">A Few Clouds</td><td>FEW015</td><td>38</td><td>32</td>
<td></td><td></td><td>79%</td><td>NA</td><td>NA</td><td>30.18</td><td>1021.2</
td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#eeeeee" valign
="top"><td>01</td><td align="right">03:51</td><td>NA</td><td>10.00</td><td ali
gn="left">Fair</td><td>CLR</td><td>39</td><td>33</td>
<td></td><td></td><td>79%</td><td>NA</td><td>NA</td><td>30.15</td><td>1020.3</
td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#f5f5f5" valign
="top"><td>01</td><td align="right">02:51</td><td>NA</td><td>10.00</td><td ali
gn="left">Fair</td><td>CLR</td><td>39</td><td>33</td>
<td></td><td></td><td>79%</td><td>NA</td><td>NA</td><td>30.14</td><td>1020.0</
td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#eeeeee" valign
="top"><td>01</td><td align="right">01:51</td><td>NA</td><td>10.00</td><td ali
```

```
gn="left">Mostly Cloudy</td><td>BKN043</td><td>40</td><td>33</td>
<td></td><td></td><td>77%</td><td>NA</td><td>NA</td><td>30.15</td><td>1020.1</
td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#f5f5f5" valign
="top"><td>01</td><td align="right">00:51</td><td>NA</td><td>10.00</td><td ali
gn="left">Overcast</td><td>OVC043</td><td>40</td><td>34</td>
<td>41</td><td>39</td><td>79%</td><td>NA</td><td>NA</td><td>30.14</td><td>101
9.7</td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#eeeeee" v
align="top"><td>30</td><td align="right">23:51</td><td>NA</td><td>10.00</td><t
d align="left">Overcast</td><td>OVC044</td><td>40</td><td>34</td>
<td></td><td></td><td>79%</td><td>NA</td><td>NA</td><td>30.13</td><td>1019.6</
td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#f5f5f5" valign
="top"><td>30</td><td align="right">22:51</td><td>NA</td><td>10.00</td><td ali
gn="left">Overcast</td><td>OVC044</td><td>40</td><td>33</td>
<td></td><td></td><td>77%</td><td>NA</td><td>NA</td><td>30.13</td><td>1019.4</
td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#eeeeee" valign
="top"><td>30</td><td align="right">21:51</td><td>NA</td><td>10.00</td><td ali
gn="left">Overcast</td><td>OVC045</td><td>40</td><td>33</td>
<td></td><td></td><td>77%</td><td>NA</td><td>NA</td><td>30.11</td><td>1018.8</
td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#f5f5f5" valign
="top"><td>30</td><td align="right">20:51</td><td>NA</td><td>10.00</td><td ali
gn="left">Overcast</td><td>SCT017 BKN029 OVC039</td><td>40</td><td>34</td>
<td></td><td></td><td>79%</td><td>NA</td><td>NA</td><td>30.10</td><td>1018.6</
td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#eeeeee" valign
="top"><td>30</td><td align="right">19:51</td><td>NA</td><td>9.00</td><td alig
n="left">Overcast</td><td>OVC015</td><td>40</td><td>34</td>
<td></td><td></td><td>79%</td><td>NA</td><td>NA</td><td>30.10</td><td>1018.6</
td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#f5f5f5" valign
="top"><td>30</td><td align="right">18:51</td><td>NA</td><td>8.00</td><td alig
n="left"> Light Rain</td><td>OVC016</td><td>39</td><td>35</td>
<td>43</td><td>39</td><td>86%</td><td>NA</td><td>NA</td><td>30.09</td><td>101
7.9</td><td></td><td></td><td>0.02</td></tr><tr align="center" bgcolor="#eeeee
e" valign="top"><td>30</td><td align="right">17:51</td><td>NA</td><td>8.00</td
><td align="left">Overcast</td><td>OVC017</td><td>40</td><td>35</td>
<td></td><td></td><td>83%</td><td>NA</td><td>NA</td><td>30.07</td><td>1017.5</
td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#f5f5f5" valign
="top"><td>30</td><td align="right">16:51</td><td>NA</td><td>9.00</td><td alig
n="left"> Light Rain</td><td>BKN018 OVC041</td><td>40</td><td>35</td>
<td></td><td></td><td>83%</td><td>NA</td><td>NA</td><td>30.05</td><td>1016.8</
```

```
td><td>0.02</td><td></td><td></td></tr><tr align="center" bgcolor="#eeeeee" va
lign="top"><td>30</td><td align="right">15:51</td><td>NA</td><td>10.00</td><td
align="left"> Light Rain</td><td>FEW034 OVC048</td><td>41</td><td>32</td>
<td></td><td></td><td>70%</td><td>NA</td><td>NA</td><td>30.04</td><td>1016.5</
td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#f5f5f5" valign
="top"><td>30</td><td align="right">14:51</td><td>NA</td><td>10.00</td><td ali
gn="left"> Light Rain</td><td>OVC043</td><td>42</td><td>30</td>
<td></td><td></td><td>62%</td><td>NA</td><td>NA</td><td>30.03</td><td>1016.1</
td><td></td><td></td><td></td></tr><tr align="center" bgcolor="#b0c4de"><th ro
wspan="3">D<br/>a<br/>t<br/>e</th><th rowspan="3">Time<br/>(est)</th>
<th rowspan="3">Wind<br/>(mph)</th><th rowspan="3">Vis.<br/>(mi.)</th><th rows
pan="3">Weather</th><th align="CENTER" rowspan="3">Sky Cond.</th>
<th rowspan="2">Air</th><th rowspan="2">Dwpt</th><th>Max.</th><th>Min.</th><th
rowspan="3" width="65">Relative<br/>Humidity</th><th rowspan="3" width="80">Wi
nd<br/>Chill<br/>(°F)</th><th rowspan="3" width="80">Heat<br/>Index<br/>(°F)</
th><th rowspan="2">altimeter<br/>(in.)</th><th rowspan="2">sea level<br/>(mb)
</th>
<th rowspan="2">1 hr</th><th rowspan="2">3 hr</th><th rowspan="2">6 hr</th></t
r>
<tr align="center" bgcolor="#b0c4de"><th colspan="2">6 hour</th></tr><tr align
="center" bgcolor="#b0c4de">
<th colspan="4">Temperature (ºF)</th><th colspan="2">Pressure</th><th colspan
="3">Precipitation (in.)</th></tr></table>
```



In [24]:

```
# extract data from the 4th table in the page into a dataframe

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[24]:

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
0	03	13:51	NA	10.00	A Few Clouds	FEW050	53	33			47%	NA	NA	29.66	1003.7			
1	03	12:51	NA	10.00	Overcast	OVC049	52	34	55	52	50%	NA	NA	29.67	1003.8			
2	03	11:51	NA	10.00	Overcast	OVC037	52	37			57%	NA	NA	29.67	1003.9			
3	03	10:51	NA	10.00	A Few Clouds	FEW037	54	39			57%	NA	NA	29.67	1003.9			
4	03	09:51	NA	10.00	Fair	CLR	54	41			62%	NA	NA	29.67	1003.8			

central park 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 [25]:

```
wu_link = 'https://www.wunderground.com/history/daily/us/ny/new-york-city/KNYC/date/2018-12-3?cm_ven=localwx_history'
```

```
In [26]: # get the page
         wu_page = requests.get(wu_link)
         wu_page

Out[26]: <Response [200]>
```

In [28]:

```
wu_soup = BeautifulSoup(wu_page.content)
```

In [30]:

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

```
<!DOCTYPE html>
<html>
 <head>
  <title>
   Central Park, NY History | Weather Underground
  </title>
  <meta charset="utf-8"/>
  <meta content="IE=edge,chrome=1" http-equiv="X-UA-Compatible"/>
  <meta content="width=device-width, initial-scale=1, maximum-scale=1" name="v
iewport"/>
  <meta content="general" name="rating"/>
  <meta content="no-referrer-when-downgrade" name="referrer"/>
  <meta content="app-id=486154808, affiliate-data=at=1010lrYB&amp;ct=website_w
u" name="apple-itunes-app"/>
  <meta content="325331260891611" name="fb_app_id"/>
  <meta content="width=device-width, initial-scale=1, maximum-scale=1" name="f
b_channel_url"/>
  <meta content="Weather Underground" property="og:site_name"/>
  <meta content="article" property="og:type"/>
  <meta content="Weather Underground provides local &amp; long range weather f
orecasts, weather reports, maps &amp; tropical weather conditions for location
s worldwide." name="description"/>
  <meta content="false" name="wui-member-logged-in"/>
```

In [29]:

```
# the table we want doesn't exist! culprit: javascript
wu_soup.find_all('div',class_='tablesaw-sortable')
```

Out[29]:

```
[]
```

In [31]:

```python
# 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',text.strip())
print(text[:1000])
```

```
Central Park, NY History | Weather Underground
  //<![CDATA[
  window.webpackManifest = {"0":"city-history-module.271410e14d01eca31253.j
s","1":"video-module.94791ee0736f33f5568c.js","2":"health-module.76922905d5d15
54bc3c6.js","3":"hurricane-module.b45b656d6deabd837533.js","4":"city-today-mod
ule.127cf5745059112cf3ac.js","5":"city-ten-day-module.7ed9ac78f02c8aa5e3d4.j
s","6":"city-hourly-module.bc5753f0cb1d1866d6d8.js","7":"precipitation-module.
df41624c9e83a1f6dd81.js","8":"city-history-calendar-module.2988780f63ffe141a66
6.js","9":"city-severe-module.70817f8d824b13fbde08.js","10":"article-page-modu
le.7c507107e75698855672.js","11":"page-module.0d95aaa43f5a7a267c2c.js","12":"m
ember-mydevices-module.6fcf4e3f03eb5d5965ab.js","13":"landing-purpleair-modul
e.72a1382f54cb1a732850.js","14":"test-module.c6d3df57d62511f4345f.js","15":"hu
rricane-storm-module.4fa5ad48c4b52827cfe2.js","16":"wundermap-module.e45823992
2d9a2289e79.js","17":"homepage-module.483ec2af1ec142a28e05.js","18":"cat-six-a
rticle-mo
```

Need to actually render page to process scripts!

In [32]:

```
# need to install chromedriver
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 [33]:

```
# this will actually render the page
driver.get(wu_link)
```

In [47]:

```
# two ways to find the table we want
wu_table = driver.find_element_by_class_name('city-history-observation')
#wu_table = driver.find_element_by_id('history-observation-table')
```

In [48]:

```
# text in the table
wu_table.text
```

Out[48]:

```
'Daily Observations\nTime Temperature Dew Point Humidity Wind Wind Speed Wind
Gust Pressure Precip.\n12:51 AM\n55 F 52 F 89 %\n0 mph 0 mph 29.4 in 0.0 in\n
1:51 AM\n55 F 52 F 89 %\n0 mph 0 mph 29.4 in 0.0 in\n2:51 AM\n55 F 53 F 93 %\n
0 mph 0 mph 29.4 in 0.0 in\n3:38 AM\n54 F 53 F 97 %\n0 mph 0 mph 29.4 in 0.0 i
n\n3:51 AM\n54 F 53 F 97 %\n0 mph 0 mph 29.4 in 0.0 in\n4:51 AM\n54 F 53 F 97
%\n0 mph 0 mph 29.4 in 0.0 in\n5:01 AM\n54 F 53 F 97 %\n0 mph 0 mph 29.4 in 0.
0 in\n5:28 AM\n54 F 52 F 93 %\n0 mph 0 mph 29.4 in 0.0 in\n5:51 AM\n54 F 51 F
90 %\n0 mph 0 mph 29.4 in 0.0 in\n6:51 AM\n53 F 49 F 86 %\n0 mph 0 mph 29.4 in
0.0 in\n7:51 AM\n54 F 46 F 75 %\n0 mph 0 mph 29.5 in 0.0 in\n8:51 AM\n54 F 44
F 69 %\n0 mph 0 mph 29.5 in 0.0 in\n9:51 AM\n54 F 41 F 62 %\n0 mph 0 mph 29.5
in 0.0 in\n10:51 AM\n54 F 39 F 57 %\n0 mph 0 mph 29.5 in 0.0 in\n11:51 AM\n52
F 37 F 57 %\n0 mph 0 mph 29.5 in 0.0 in\n12:51 PM\n52 F 34 F 50 %\n0 mph 0 mph
29.5 in 0.0 in\n1:51 PM\n53 F 33 F 47 %\n0 mph 0 mph 29.5 in 0.0 in'
```

In [49]:

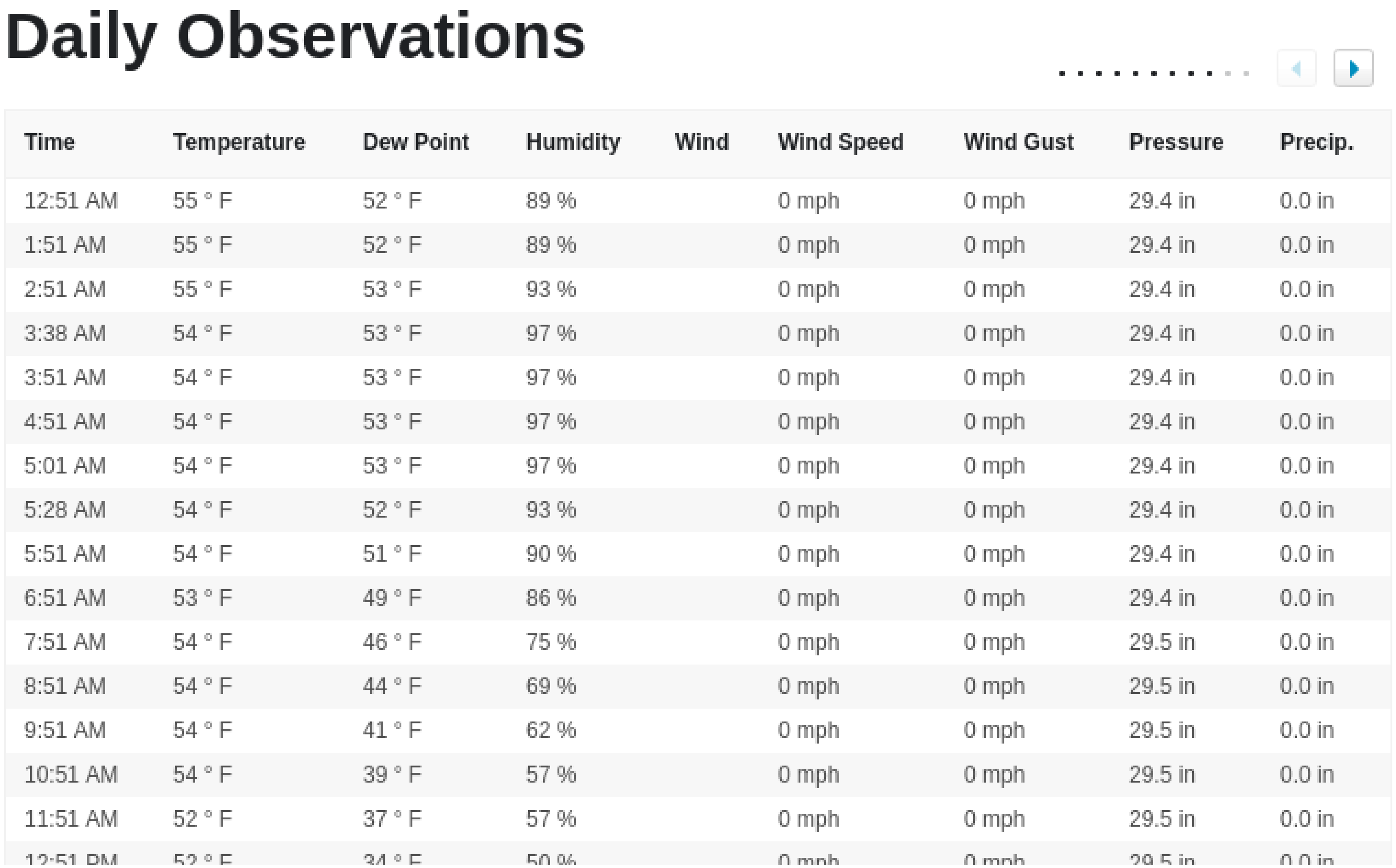
```python
# extracting text into a datafram
wu_data = []
for tr in wu_table.find_elements_by_css_selector('tr'):
    tmp_row = []
    for td in tr.find_elements_by_css_selector('td'):
        tmp_row.append(td.text.strip())
    wu_data.append(tmp_row)
df_wu = pd.DataFrame(wu_data)
df_wu.head()
```

Out[49]:

	0	1	2	3	4	5	6	7	8	9	10
0	None	None	None	None	None	None	None	None	None	None	None
1	12:51 AM	55 F	52 F	89 %		0 mph	0 mph	29.4 in	0.0 in		
2	1:51 AM	55 F	52 F	89 %		0 mph	0 mph	29.4 in	0.0 in		
3	2:51 AM	55 F	53 F	93 %		0 mph	0 mph	29.4 in	0.0 in		
4	3:38 AM	54 F	53 F	97 %		0 mph	0 mph	29.4 in	0.0 in		

```
In [52]: # visualize the rendered table, still missing some stuff, need to debug
         wu_table.screenshot('./images/test1.png')

Out[52]: True
```



Daily Observations

Time	Temperature	Dew Point	Humidity	Wind	Wind Speed	Wind Gust	Pressure	Precip.
12:51 AM	55 ° F	52 ° F	89 %		0 mph	0 mph	29.4 in	0.0 in
1:51 AM	55 ° F	52 ° F	89 %		0 mph	0 mph	29.4 in	0.0 in
2:51 AM	55 ° F	53 ° F	93 %		0 mph	0 mph	29.4 in	0.0 in
3:38 AM	54 ° F	53 ° F	97 %		0 mph	0 mph	29.4 in	0.0 in
3:51 AM	54 ° F	53 ° F	97 %		0 mph	0 mph	29.4 in	0.0 in
4:51 AM	54 ° F	53 ° F	97 %		0 mph	0 mph	29.4 in	0.0 in
5:01 AM	54 ° F	53 ° F	97 %		0 mph	0 mph	29.4 in	0.0 in
5:28 AM	54 ° F	52 ° F	93 %		0 mph	0 mph	29.4 in	0.0 in
5:51 AM	54 ° F	51 ° F	90 %		0 mph	0 mph	29.4 in	0.0 in
6:51 AM	53 ° F	49 ° F	86 %		0 mph	0 mph	29.4 in	0.0 in
7:51 AM	54 ° F	46 ° F	75 %		0 mph	0 mph	29.5 in	0.0 in
8:51 AM	54 ° F	44 ° F	69 %		0 mph	0 mph	29.5 in	0.0 in
9:51 AM	54 ° F	41 ° F	62 %		0 mph	0 mph	29.5 in	0.0 in
10:51 AM	54 ° F	39 ° F	57 %		0 mph	0 mph	29.5 in	0.0 in
11:51 AM	52 ° F	37 ° F	57 %		0 mph	0 mph	29.5 in	0.0 in
12:51 PM	52 ° F	34 ° F	50 %		0 mph	0 mph	29.5 in	0.0 in

```
In [ ]:
```