

# Data X Berkeley

## Plaksha Web scraping assignment

### Assignment problem

In this week's assignment you should extract live weather data in Mohali for the next 15 days:

<https://weather.com/en-IN/weather/tenday/l/d253814420804f28d49f7a75130e71b797c3c93f9ded842eaf6769e3784da352>

- Task scrape
  - day | period (as Wed 11 | Day, Wed 11 | Night etc.)
  - the temperature for the period
  - the short description (e.g. Mostly Clear, Sunny etc.)
  - the long weather description (e.g. Showers late at night. Low 12°C. Winds SE at 10 to 15 km/h. Chance of rain 40%).

Store the scraped data strings in a Pandas DataFrame

**Hint:** The weather information is found in a div tag with `class_='DailyForecast--DisclosureList--nosQS'`

Your DataFrame should be similar to the below screenshot (with the same columns):

	day	temp	short_desc	desc
0	Wed 11   Day	14°	Cloudy	Cloudy. High 14°C. Winds SW and variable.
1	Wed 11   Night	11°	Cloudy	Skies clearing late. Low 11°C. Winds ESE and variable.

### Assignment solution

#### 1. Importing Libraries

```
1 import requests
2 import bs4 as bs
3 import pandas as pd
```

#### 2. Getting Request

##### user-agent

When you're sending a request to a webpage (no matter if it comes from your computer, iphone, or Python's request package), then you also include a user-agent. This let's the webserver know how to render the contents for you. You can also send user-agent information via a request (to specify who you are for example, or to disguise that you're an automated scraper).

Find your machine's / browser's true user agent by Googling: <https://www.google.com/search?q=find+my+user+agent&oq=find+my+user+agent>

```
1 headers = {'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/109.0.0.0 Safari/537.36'}
2 source = requests.get('https://weather.com/en-IN/weather/tenday/l/d253814420804f28d49f7a75130e71b797c3c93f9ded842eaf6769e3784da352', headers=headers)
```

```
1 source
<Response [200]>
```

```
1 source.headers
{'Content-Type': 'text/html; charset=utf-8', 'X-DNS-Prefetch-Control': 'off', 'X-Frame-Options': 'SAMEORIGIN', 'X-Download-Options': 'noopen', 'X-Content-Type-Options': 'nosniff', 'X-XSS-Protection': '1; mode=block', 'webcakes-app-name': 'daily', 'X-UA-Compatible': 'IE=edge'}
```

```
'webcakes-app-version': '1.0.0-06bdfdc2cc', 'webcakes-region': 'sjc03', 'ETag': 'W/"1044c5-1kLXH122ep2wLSWs4UZ0cWmfblQ"', 'X-Akamai-Transformed': '9 - 0 pmb=mTOE,4mRUM,1', 'Content-Encoding': 'gzip', 'Expires': 'Fri, 20 Jan 2023 06:31:29 GMT', 'Cache-Control': 'max-age=0, no-cache', 'Pragma': 'no-cache', 'Date': 'Fri, 20 Jan 2023 06:31:29 GMT', 'Transfer-Encoding': 'chunked', 'Connection': 'keep-alive, Transfer-Encoding', 'Vary': 'Accept-Encoding', 'Set-Cookie': 'speedpin=4G; expires=Fri, 20-Jan-2023 07:01:29 GMT; path=/; domain=.weather.com; secure, ci=TWC-Connection-Speed=4G&TWC-Locale-Group=GLS&TWC-Device-Class=desktop&X-Origin-Hint=PROD-IBM-Daily&TWC-Network-Type=wifi&TWC-GeoIP-Country=US&TWC-GeoIP-Lat=39.0438&TWC-GeoIP-Long=-77.4879&Akamai-Connection-Speed=1000+&TWC-Privacy=usa-va&TWC-GeoIP-DMA=511&TWC-GeoIP-City=ASHBURN&TWC-GeoIP-Region=VA; path=/; domain=.weather.com; secure, ak_bmsc=B44832DB7FB4CEBF798C85C07E7590D9~00000000000000000000000000000000~YAAQxzoiF6ZvXM2FAQAAAgjgzRLmkHAQpzChPUP5sawMGYxC+hyQQM8YF0i Domain=.weather.com; Path=/; Expires=Fri, 20 Jan 2023 08:31:28 GMT; Max-Age=7199; HttpOnly, bm_mi=B0453B9AD9A1C26B9818C9D57E040BBD~YAAQxzoiF6dvXM2FAQAAAgjgzRKKQcm4q/jTfTnOiuumuITyKHrejMe8yJtwudRfwM8NGrIfTAxYXh+gUC10Rfry2uRK Domain=.weather.com; Path=/; Expires=Fri, 20 Jan 2023 06:31:29 GMT; Max-Age=0; Secure', 'Server-Timing': 'cdn-cache; desc=REVALIDATE, edge; dur=161, origin; dur=776, intid; desc=b18636200161b61b', 'Access-Control-Max-Age': '86400', 'Access-Control-Allow-Credentials': 'false', 'Access-Control-Allow-Headers': '*', 'Access-Control-Allow-Methods': 'GET,POST,OPTIONS', 'Access-Control-Allow-Origin': '*', 'Property-id': 'TWC-WX-Prod', 'TWC-PATH-LOCALE': '/en-IN/', 'TWC-Subs': 'none', 'TWC-Unit': 'u', 'TWC-Privacy': 'usa-va', 'TWC-GeoIP-Region': 'VA', 'TWC-GeoIP-City': 'ASHBURN', 'TWC-GeoIP-DMA': '511', 'TWC-GeoIP-LatLong': '39.0438,-77.4879', 'TWC-GeoIP-Country': 'US', 'TWC-Device-Class': 'desktop', 'TWC-Locale-Group': 'GLS+', 'TWC-Connection-Speed': '4G', 'Strict-Transport-Security': 'max-age=15552000; preload', 'TWC-AK-Req-ID': '97b606', 'X-Origin-Hint': 'PROD-IBM-Daily'}
```

### 3. Scraping

```
1 soup = bs.BeautifulSoup(source.content, features = 'html.parser')
```


Creating dictionary for the 4 columns to be created in the dataframe


```
1 data = {'day':[], 'temp':[], 'short_desc':[], 'desc':[]}
```

The required data such as day, temperature, short description and long description is scraped by inspecting the weather site and finding the relevant class name.





I used the following four highlighted elements to scrape the data.

Fri 20 | Day

**17°**  12%

 NNW 12 km/h

Cloudy. High 17°C. Winds NNW at 10 to 15 km/h.

 Humidity <b>70%</b>	 UV Index <b>2 of 10</b>
 Sunrise <b>07:20</b>	 Sunset <b>17:47</b>

Using html attribute 'class' and for loop to find all the 15 days (and night) data related to the weather.

```
1 for d in soup.find_all(class_ = 'DailyContent--daypartName--3emSU'):
2     data['day'].append(d.text)

1 for t in soup.find_all(class_ = 'DailyContent--temp--1s3a7'):
2     data['temp'].append(t.text)

1 for sd in soup.find_all(class_ = 'DailyContent--weatherIcon--Kh7Ru'):
2     data['short_desc'].append(sd.text)

1 for des in soup.find_all(class_ = 'DailyContent--narrative--3Ti6_'):
2     data['desc'].append(des.text)
```

### 4. Output

```
1 pd.DataFrame(data = data)
```

	day	temp	short_desc	desc	
0	Fri 20   Day	17°	Cloudy	Cloudy. High 17°C. Winds NNW at 10 to 15 km/h.	
1	Fri 20   Night	9°	Partly Cloudy Night	Partly cloudy. Low 9°C. Winds N and variable.	
2	Sat 21   Day	20°	Sunny	Generally clear. High 20°C. Winds W and variable.	
3	Sat 21   Night	8°	Clear Night	Generally clear. Low 8°C. Winds N and variable.	
4	Sun 22   Day	21°	Sunny	Generally clear. High 21°C. Winds NW and varia...	
5	Sun 22   Night	9°	Mostly Clear Night	Generally clear. Low 9°C. Winds NNE and variable.	
6	Mon 23   Day	20°	Partly Cloudy	Partly cloudy. High 20°C. Winds NW and variable.	
7	Mon 23   Night	11°	Scattered Showers Night	Showers late at night. Low 11°C. Winds NNE and...	
8	Tue 24   Day	20°	Partly Cloudy	Partly cloudy. High 20°C. Winds SE at 10 to 15...	
9	Tue 24   Night	12°	Scattered Thunderstorms Night	Scattered thunderstorms. Low 12°C. Winds ESE a...	
10	Wed 25   Day	19°	Thunderstorm	Thunderstorms. High 19°C. Winds SE at 10 to 15...	
11	Wed 25   Night	12°	Thunderstorm	Thunderstorms. Hazy. Low 12°C. Winds ESE and v...	
12	Thu 26   Day	19°	Scattered Thunderstorms	Scattered thunderstorms. High 19°C. Winds SSE ...	
13	Thu 26   Night	12°	Thunderstorm	Thunderstorms. Low 12°C. Winds ESE and variabl...	
14	Fri 27   Day	18°	Scattered Thunderstorms	Scattered thunderstorms. High 18°C. Winds S at...	
15	Fri 27   Night	11°	Scattered Thunderstorms Night	Scattered thunderstorms. Low 11°C. Winds NE an...	
16	Sat 28   Day	18°	Scattered Showers	Showers. High 18°C. Winds SSW at 10 to 15 km/h...	
17	Sat 28   Night	10°	Scattered Showers	Showers. Low 10°C. Winds ESE and variable. Cha...	
18	Sun 29   Day	15°	Scattered Showers	Showers. High 15°C. Winds WSW and variable. Ch...	
19	Sun 29   Night	8°	Rain	Rain. Low 8°C. Winds NE and variable. Chance o...	
20	Mon 30   Day	17°	Sunny	Generally clear. High 17°C. Winds WNW at 10 to...	
21	Mon 30   Night	7°	Clear Night	Generally clear. Low 7°C. Winds NNW and variable.	
22	Tue 31   Day	17°	Sunny	Generally clear. High 17°C. Winds WNW at 10 to...	
23	Tue 31   Night	7°	Mostly Clear Night	Generally clear. Low 7°C. Winds NNW and variable.	
24	Wed 01   Day	17°	Sunny	Generally clear. High 17°C. Winds WNW at 10 to...	
25	Wed 01   Night	7°	Mostly Clear Night	Generally clear. Low 7°C. Winds NNW and variable.	
26	Thu 02   Day	18°	Sunny	Generally clear. High 18°C. Winds WNW at 10 to...	
27	Thu 02   Night	7°	Clear Night	Generally clear. Low 7°C. Winds NNW and variable.	
28	Fri 03   Day	19°	Sunny	Generally clear. High 19°C. Winds WNW at 10 to...	
29	Fri 03   Night	9°	Partly Cloudy Night	Partly cloudy. Low 9°C. Winds NNE and variable.	

1