

Module-10 Assignment - Submission by Diwakar Sinha

```
In [1]: # Importing Libraries
from bs4 import BeautifulSoup
import requests
import pandas as pd
```

Q1) Go to the website

https://www.marketwatch.com/investing/stock/tsla?mod=search_symbol to scrape the below items

- Current price of the stock
- Close price of the stock
- Analyst rating of the stock

Ans1) Stock price extraction

```
In [2]: url = 'https://www.marketwatch.com/investing/stock/tsla?mod=search_symbol'
page = requests.get(url)
print(page)
```

<Response [200]>

```
In [3]: soup = BeautifulSoup(page.text, 'lxml')
```

```
In [4]: current_price = soup.find('bg-quote',{'class':'value'}).text
print("The current price of the stock is:",current_price)
```

The current price of the stock is: 239.09

```
In [5]: close_price = soup.find('td',{'class':'table__cell u-semi'}).text
print("The close price of the stock is:",close_price)
```

The close price of the stock is: \$238.82

```
In [6]: rating = soup.find('li',{'class':"analyst__option active"}).text
print("The analyst rating of the stock is:",rating)
```

The analyst rating of the stock is: Hold

Q2) Go to the website <https://www.icc-cricket.com/rankings/mens/team-rankings/test> to scrape the table given on the website.

Ans2) Table extraction

```
In [7]: url = 'https://www.icc-cricket.com/rankings/mens/team-rankings/test'
page = requests.get(url)
print(page)
```

<Response [200]>

```
In [8]: soup = BeautifulSoup(page.text, 'lxml')
```

```
In [9]: table = soup.find('div',{ 'class': 'rankings-block__container full rankings-table'})  
table
```

```

Out[9]: <div class="rankings-block__container full rankings-table">
<table class="table">
<thead>
<tr class="table-head">
<th class="table-head__cell u-text-right">Pos</th>
<th class="table-head__cell rankings-table__team">
<span class="u-hide-mobile">Team</span>
<span class="u-show-mobile">T</span>
</th>
<th class="table-head__cell u-center-text">
<span class="u-hide-mobile">Matches</span>
<span class="u-show-mobile">M</span>
</th>
<th class="table-head__cell u-center-text">
<span class="u-hide-mobile">Points</span>
<span class="u-show-mobile">P</span>
</th>
<th class="table-head__cell u-text-right rating">
<span class="u-hide-mobile">Rating</span>
<span class="u-show-mobile">R</span>
</th>
</tr>
</thead>
<tbody>
<tr class="rankings-block__banner">
<td class="rankings-block__banner--pos">1</td>
<td class="rankings-block__banner--team-name">
<span class="flag-30 rankings-block__banner--flag IND"></span>
<span class="u-hide-phablet">India</span>
<span class="u-show-phablet">IND</span>
</td>
<td class="rankings-block__banner--matches">29</td>
<td class="rankings-block__banner--points">3,434</td>
<td class="rankings-block__banner--rating u-text-right">
118
<span class="circle-branding circle-branding--top"></s
pan>
<span class="circle-branding circle-branding--bottom"></span>
</td>
</tr>
<tr class="table-body" data-team-id="15">
<td class="table-body__cell table-body__cell--position u-text-right">2</td>
<td class="table-body__cell rankings-table__team">
<span class="flag-15 table-body_logo AUS"></span>
<span class="u-hide-phablet">Australia</span>
<span class="u-show-phablet">AUS</span>
</td>
<td class="table-body__cell u-center-text">30</td>
<td class="table-body__cell u-center-text">3,534</td>
<td class="table-body__cell u-text-right rating">118</td>
</tr>
<tr class="table-body" data-team-id="11">
<td class="table-body__cell table-body__cell--position u-text-right">3</td>
<td class="table-body__cell rankings-table__team">
<span class="flag-15 table-body_logo ENG"></span>
<span class="u-hide-phablet">England</span>
<span class="u-show-phablet">ENG</span>
</td>
<td class="table-body__cell u-center-text">43</td>
<td class="table-body__cell u-center-text">4,941</td>
<td class="table-body__cell u-text-right rating">115</td>
</tr>
<tr class="table-body" data-team-id="19">
<td class="table-body__cell table-body__cell--position u-text-right">4</td>

```

```

<td class="table-body__cell rankings-table__team">
<span class="flag-15 table-body_logo SA"></span>
<span class="u-hide-phablet">South Africa</span>
<span class="u-show-phablet">SA</span>
</td>
<td class="table-body__cell u-center-text">21</td>
<td class="table-body__cell u-center-text">2,182</td>
<td class="table-body__cell u-text-right rating">104</td>
</tr>
<tr class="table-body" data-team-id="16">
<td class="table-body__cell table-body__cell--position u-text-right">5</td>
<td class="table-body__cell rankings-table__team">
<span class="flag-15 table-body_logo NZ"></span>
<span class="u-hide-phablet">New Zealand</span>
<span class="u-show-phablet">NZ</span>
</td>
<td class="table-body__cell u-center-text">23</td>
<td class="table-body__cell u-center-text">2,291</td>
<td class="table-body__cell u-text-right rating">100</td>
</tr>
<tr class="table-body" data-team-id="20">
<td class="table-body__cell table-body__cell--position u-text-right">6</td>
<td class="table-body__cell rankings-table__team">
<span class="flag-15 table-body_logo PAK"></span>
<span class="u-hide-phablet">Pakistan</span>
<span class="u-show-phablet">PAK</span>
</td>
<td class="table-body__cell u-center-text">25</td>
<td class="table-body__cell u-center-text">2,304</td>
<td class="table-body__cell u-text-right rating">92</td>
</tr>
<tr class="table-body" data-team-id="13">
<td class="table-body__cell table-body__cell--position u-text-right">7</td>
<td class="table-body__cell rankings-table__team">
<span class="flag-15 table-body_logo SL"></span>
<span class="u-hide-phablet">Sri Lanka</span>
<span class="u-show-phablet">SL</span>
</td>
<td class="table-body__cell u-center-text">27</td>
<td class="table-body__cell u-center-text">2,123</td>
<td class="table-body__cell u-text-right rating">79</td>
</tr>
<tr class="table-body" data-team-id="21">
<td class="table-body__cell table-body__cell--position u-text-right">8</td>
<td class="table-body__cell rankings-table__team">
<span class="flag-15 table-body_logo WI"></span>
<span class="u-hide-phablet">West Indies</span>
<span class="u-show-phablet">WI</span>
</td>
<td class="table-body__cell u-center-text">28</td>
<td class="table-body__cell u-center-text">2,154</td>
<td class="table-body__cell u-text-right rating">77</td>
</tr>
<tr class="table-body" data-team-id="22">
<td class="table-body__cell table-body__cell--position u-text-right">9</td>
<td class="table-body__cell rankings-table__team">
<span class="flag-15 table-body_logo BAN"></span>
<span class="u-hide-phablet">Bangladesh</span>
<span class="u-show-phablet">BAN</span>
</td>
<td class="table-body__cell u-center-text">19</td>
<td class="table-body__cell u-center-text">873</td>
<td class="table-body__cell u-text-right rating">46</td>
</tr>

```

```

<tr class="table-body" data-team-id="18">
<td class="table-body__cell table-body__cell--position u-text-right">10</td>
<td class="table-body__cell rankings-table__team">
<span class="flag-15 table-body_logo ZIM"></span>
<span class="u-hide-phablet">Zimbabwe</span>
<span class="u-show-phablet">ZIM</span>
</td>
<td class="table-body__cell u-center-text">7</td>
<td class="table-body__cell u-center-text">223</td>
<td class="table-body__cell u-text-right rating">32</td>
</tr>
</tbody>
</table>
</div>

```

```
In [10]: table.find_all('th')
```

```

Out[10]: [<th class="table-head__cell u-text-right">Pos</th>,
<th class="table-head__cell rankings-table__team">
<span class="u-hide-mobile">Team</span>
<span class="u-show-mobile">T</span>
</th>,
<th class="table-head__cell u-center-text">
<span class="u-hide-mobile">Matches</span>
<span class="u-show-mobile">M</span>
</th>,
<th class="table-head__cell u-center-text">
<span class="u-hide-mobile">Points</span>
<span class="u-show-mobile">P</span>
</th>,
<th class="table-head__cell u-text-right rating">
<span class="u-hide-mobile">Rating</span>
<span class="u-show-mobile">R</span>
</th>]

```

```

In [11]: headers = []
for i in table.find_all('th'):
    col_name = str.strip(i.text)
    col_name = str.replace(col_name, '\n', '')
    headers.append(col_name)
headers

```

```
Out[11]: ['Pos', 'TeamT', 'MatchesM', 'PointsP', 'RatingR']
```

```

In [12]: final_df = pd.DataFrame(columns=headers)
final_df

```

```
Out[12]:
```

Pos	TeamT	MatchesM	PointsP	RatingR
-----	-------	----------	---------	---------

```

In [13]: for i in table.find_all('tr')[1:]:
    row_data = i.find_all('td')
    rows = [t.text for t in row_data]
    length = len(final_df)
    final_df.loc[length] = rows

```

```
In [14]: final_df
```

Out[14]:

	Pos	TeamT	MatchesM	PointsP	RatingR
0	1	\n\nIndia\nIND\n	29	3,434	\n 118\n ...
1	2	\n\nAustralia\nAUS\n	30	3,534	118
2	3	\n\nEngland\nENG\n	43	4,941	115
3	4	\n\nSouth Africa\nSA\n	21	2,182	104
4	5	\n\nNew Zealand\nNZ\n	23	2,291	100
5	6	\n\nPakistan\nPAK\n	25	2,304	92
6	7	\n\nSri Lanka\nSL\n	27	2,123	79
7	8	\n\nWest Indies\nWI\n	28	2,154	77
8	9	\n\nBangladesh\nBAN\n	19	873	46
9	10	\n\nZimbabwe\nZIM\n	7	223	32

In [15]:

```
# cleaning up \n
final_df = final_df.replace('\n', ' ', regex=True)
final_df
```

Out[15]:

	Pos	TeamT	MatchesM	PointsP	RatingR
0	1	India IND	29	3,434	118 ...
1	2	Australia AUS	30	3,534	118
2	3	England ENG	43	4,941	115
3	4	South Africa SA	21	2,182	104
4	5	New Zealand NZ	23	2,291	100
5	6	Pakistan PAK	25	2,304	92
6	7	Sri Lanka SL	27	2,123	79
7	8	West Indies WI	28	2,154	77
8	9	Bangladesh BAN	19	873	46
9	10	Zimbabwe ZIM	7	223	32

Q3) Using Selenium search for the term 'Bali' and scrap 10 images from google.com.

In [16]:

```
# Import necessary Libraries
from selenium import webdriver
from selenium.webdriver.common.keys import Keys
import time
import warnings
warnings.filterwarnings('ignore')

# Set up the WebDriver (using Chrome)
driver = webdriver.Chrome('D:\chromedriver\chromedriver.exe') # Set the correct path
```

In [17]:

```
# Open Google
url= "https://www.google.com"
driver.get(url)
```

```
# Find the search bar element and enter the search term
search_box = driver.find_element_by_name("q")
search_term = "Bali"
search_box.send_keys(search_term)
search_box.send_keys(Keys.RETURN)
```

```
In [18]: # Find and click on the "Images" tab
images_tab = driver.find_element_by_link_text("Images")
images_tab.click()
```

```
In [19]: # Find image elements
image_elements = driver.find_elements_by_css_selector(".rg_i")

# Define the number of images to scrape
num_images_to_scrape = 10
images_downloaded = 0

for img in image_elements:
    # Get the image URL from the data-src attribute
    image_url = img.get_attribute("data-src")

    # Skip elements without a valid image URL
    if not image_url:
        continue

    # Download the image using the requests library
    import requests
    response = requests.get(image_url)

    if response.status_code == 200:
        # Specify the file path to save the image
        file_name = f"ds_bali_image_{images_downloaded+1}.jpg"

        with open(file_name, "wb") as file:
            file.write(response.content)

        print(f"Downloaded {file_name}")

        images_downloaded += 1
        if images_downloaded >= num_images_to_scrape:
            break

# Close the WebDriver
driver.quit()
```

```
Downloaded ds_bali_image_1.jpg
Downloaded ds_bali_image_2.jpg
Downloaded ds_bali_image_3.jpg
Downloaded ds_bali_image_4.jpg
Downloaded ds_bali_image_5.jpg
Downloaded ds_bali_image_6.jpg
Downloaded ds_bali_image_7.jpg
Downloaded ds_bali_image_8.jpg
Downloaded ds_bali_image_9.jpg
Downloaded ds_bali_image_10.jpg
```

```
In [ ]:
```