



Assignment 01: Scrape the Simplilearn Website Homepage

The comments/sections provided are your cues to perform the assignment. You don't need to limit yourself to the number of rows/cells provided. You can add additional rows in each section to add more lines of code.

If at any point in time you need help on solving this assignment, view our demo video to understand the different steps of the code.

Happy coding!

1: View and print the Simplilearn web page content in a good format

In [2]:

```
1 #Import the required libraries
2 import requests
3 from bs4 import BeautifulSoup
```

In [3]:

```
1 #Provide the url for web scraping
2 url = 'http://www.simplilearn.com/'
```

In [4]:

```
1 #Access the url with the request object
2 web_page = requests.get(url)
```

In [7]:

```
1 print(web_page)
```

<Response [200]>

In [5]:

```
1 #Scrape the webpage content
2 new_web_page = web_page.content
```


In [14]:

```
1 #View the content of the soup object
2 print(soup)
```

```
F", "LMS_ENTERPRISE_SUBDOMAIN_URL": "https:\u002F\u002Fflms.simplilearn.c
om", "ICE9_APIBASE_ENTERPRISE_SUBDOMAIN_URL": "https:\u002F\u002Fenterpr
ise.simplilearn.com\u002Fapi\u002Fv1\u002Findex", "AMPLITUDE_USER_TYP
E": "Freemium", "EXPLORE_COURSE_MOBILE": "exploreCourseMobile", "SPLITIT_R
OW_GATEWAY_KEY": "92dbfbce-6a8e-444d-9c3e-2b98e505feaf", "SPLITIT_US_GAT
EWAY_KEY": "1ceb7b77-f694-41fc-b40d-f864085129d1", "SPLITIT_ENABLED_COUN
TRY_IDS": [ "ALL_EXCEPT_INDIA" ], "SPLITIT_ENABLED_COUNTRY_CODES": [ "ALL_EX
CEPT_INDIA" ], "SPLITIT_US_ENTITY_COUNTRY_CODES": [ "CA", "GB", "MX" ], "SPLIT
IT_US_ENTITY_COUNTRY_IDS": [14, 35, 36], "FOR_FREE": "For FREE", "BASE_ACCOU
NTS_LOGIN_URL": "https:\u002F\u002Faccounts.simplilearn.com\u002Fuser\u
002Fmanage\u002Fprocess-google-one-tap-response", "GOOGLE_WEB_CLIENT_I
D": "908267265821-qbs9oggfkgbpacti4bmssfc1516mc651.apps.googleusercontent.com", "TOC_DEFAULT_LINK_COUNT": 5, "IND_SERVICE_TAX": 18, "DIRECT_CALLS_
US_NUM": "+1-844-842-4145", "DIRECT_CALLS_IN_NUM": "+91 80684 52482", "ITI
L_CATEGORY_ID": 2, "US_COUNTRY_ID": 34, "IN_COUNTRY_ID": 6, "SG_COUNTRY_ID":
213, "IN_COUNTRY_PHONE_CODE": "91", "APPLICATION_FROM_TYPE_SALES": "sale
s", "APPLICATION_FROM_TYPE_MARKETING": "online", "PAGE_TYPE_COHORT": "coho
rt", "PAGE_TYPE_PURDUE": "purdue", "PAGE_TYPE_UNIVERSITY": "university", "P
AGE_TYPE_CLASSIC": "classic", "SKILLUP_HOMEPAGE_POPUP": "freemium_homepag
e_popup", "FREEMIUM COURSE POPUP": "freemium course popup", "PAID COURSE
```

In [15]:

```
1 #use the prettify() method to view a well formatted output (with html tags)
2 print(soup.prettify())
```

```
</div>

</h4>
<div class="info-list">
  <span class="lazy">
    5
    <!-- -->
    Months
  </span>
  <span class="lazy">
    Cohort Starts :
    <!-- -->
    09
    <!-- -->
    <!-- -->
    Jan
    <!-- -->
    ,
    <!-- -->
    2022
  </span>
```

2: View the head and title of the webpage

In [16]:

```
1 #View the head of the soup object
2 print(soup.head)
```

```
<head>
<meta content="text/html; charset=utf-8" http-equiv="Content-Type"/>
<meta content="IE=edge" http-equiv="X-UA-Compatible"/>
<title>World's #1 Online Bootcamp & Certification Course Provider
| Simplilearn</title>
<link href="https://ssl.google-analytics.com/" rel="dns-prefetch"/>
<link href="https://stats.g.doubleclick.net/" rel="dns-prefetch">
<link href="https://www.google.com/" rel="dns-prefetch"/>
<link href="https://cdn.cookiecutter.org/" rel="dns-prefetch"/>
<link href="https://js-agent.newrelic.com" rel="preconnect"/>
<script type="text/javascript">
    ;window.NREUM||(NREUM={});NREUM.init={privacy:{cookies_enabled:true
e}};window.NREUM||(NREUM={}),__nr_require=function(t,e,n){function r
(n){if(!e[n]){var i=e[n]={exports:{}};t[n][0].call(i.exports,function
(e){var i=t[n][1][e];return r(i||e)},i,i.exports)}return e[n].exports}
if("function"==typeof __nr_require)return __nr_require;for(var i=0;i<
n.length;i++)r(n[i]);return r}({1:[function(t,e,n){function r(t){try
{c.console&&console.log(t)}catch(e){}}var i,o=t("ee"),a=t(24),c={};try
{i=localStorage.getItem("__nr_flags").split(","),console&&"function"==
```

In [18]:

```
1 #View the title of the webpage
2 print(soup.title)
```

```
<title>World's #1 Online Bootcamp & Certification Course Provider
| Simplilearn</title>
```

3: Print all the href links present in the Simplilearn webpage

In [20]:

```
1 #Find all the links present in the webpage
2 print(soup.findAll('a'))
```

```
follow" target="_blank"><div class="award-img"></div><span>2018
Gold Winner</span><p>For Customer Service by Stevie Awards</p></a>, <a
aria-label="trusted brands" class="lazy" href="https://stevieawards.co
m/sales/2020-stevie%C2%AE-award-winners" rel="noreferrer nofollow" tar
get="_blank"><div class="award-img"></div><span>2020 Gold Win
ner</span><p>For Customer Service by Stevie Awards</p></a>, <a aria-la
bel="trusted brands" class="lazy" href="https://tldrify.com/zx2 " rel
="noreferrer nofollow" target="_blank"><div class="award-img"></div><span>2020 Winner</span><p>Innovations in Edtech by Ae
gis Graham Bell Award</p></a>, <a aria-label="trusted brands" class="l
```

In [28]:

```
1 for tag in soup.findAll('a', href=True):  
2     print(tag['href'])
```

```
https://trainingindustry.com/top-training-companies/online-learning-libraries/2020-top-online-learning-library-companies/ (https://trainingindustry.com/top-training-companies/online-learning-libraries/2020-top-online-learning-library-companies/)  
https://play.google.com/store/apps/details?id=com.mobile.simplilearn (https://play.google.com/store/apps/details?id=com.mobile.simplilearn)  
https://apps.apple.com/app/simplilearn/id963042747?ls=1 (https://apps.apple.com/app/simplilearn/id963042747?ls=1)  
https://www.simplilearn.com/refer-and-earn (https://www.simplilearn.com/refer-and-earn)  
https://www.facebook.com/simplilearn (https://www.facebook.com/simplilearn)  
https://www.twitter.com/simplilearn (https://www.twitter.com/simplilearn)  
https://www.youtube.com/user/Simplilearn (https://www.youtube.com/user/Simplilearn)  
https://www.linkedin.com/company/simplilearn (https://www.linkedin.com/company/simplilearn)
```

In []:

```
1
```