

# 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]:
```

```
#Import the required libraries
import requests
from bs4 import BeautifulSoup
```

# In [3]:

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

## In [4]:

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

#### In [7]:

```
1 print(web_page)
```

## <Response [200]>

## In [5]:

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

#### In [8]:

```
1 new web page
```

f"in Object?(r(document),r(window),r(f.prototype)):f.prototype.hasOwnP  $roperty(u) \&\&(i(window),i(f.prototype)),a.on(u+"-start",function(t,e) \\ \{v,v\},v\}$ ar n=t[1],r=s(n,"nr@wrapped",function(){function t(){if("function"==ty peof n.handleEvent)return n.handleEvent.apply(n,arguments)}var e={obje ct:t, "function":n}[typeof n];return e?c(e, "fn-", null, e.name||"anonymou  $s"):n});this.wrapped=t[1]=r}),a.on(d+"-start",function(t){t[1]=this.wr$ apped||t[1]|, {}], 6: [function(t,e,n){function r(t,e,n){var r=t[e]; "fu nction"==typeof r&&(t[e]=function(){var t=o(arguments),e={};i.emit(n +"before-start",[t],e); var a; e[m] & e[m] .dt & (a=e[m].dt); var c=r.apply (this,t);return i.emit(n+"start",[t,a],c),c.then(function(t){return i. emit(n+"end",[null,t],c),t},function(t){throw i.emit(n+"end",[t],c), t})})}var i=t("ee").get("fetch"),o=t(25),a=t(24);e.exports=i;var c=win dow,s="fetch-",f=s+"body-",u=["arrayBuffer","blob","json","text","form Data"],d=c.Request,l=c.Response,p=c.fetch,h="prototype",m="nr@contex t";d&&1&&p&&( $a(u, function(t, e) \{r(d[h], e, f), r(l[h], e, f)\}), r(c, "fetch",$ s),i.on(s+"end",function(t,e){var n=this;if(e){var r=e.headers.get("co ntent-length");null!==r&&(n.rxSize=r),i.emit(s+"done",[null,e],n)}else i.emit(s+"done",[t],n)}))},{}],7:[function(t,e,n){var r=t("ee").get("h istory"), i=t("wrap-function")(r); e.exports=r; var o=window.history&&win

#### In [9]:

```
#Create a soup object to parse the webpage using html parser
soup = BeautifulSoup(new_web_page,'html.parser')
```

#### In [21]:

```
#Close the result object
web_page.close()
type(soup)
```

#### Out[21]:

bs4.BeautifulSoup

#### In [14]:

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

F", "LMS ENTERPRISE SUBDOMAIN URL": "https:/u002F/u002Flms.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-qbs9oqqfqkbpacti4bmssfci516mc651.apps.googleuserconte nt.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]:

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

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

#### 2: View the head and title of the webpage

```
In [16]:
```

```
#View the head of the soup object
   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 &amp; 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.cookielaw.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:tru
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]:
   #View the title of the webpage
   print(soup.title)
```

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

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

#### In [20]:

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

follow" target="\_blank"><div class="award-img"><img alt="2018 Gold Win ner" class="lazy gold-winner0" data-src="https://www.simplilearn.com/ice9/assets/home/2018-gold-winner.png" height="9" src="https://www.simplilearn.com/ice9/assets/form\_opacity.png" width="16"/></div><span>2018 Gold Winner</span>For Customer Service by Stevie Awards</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"><img alt="2020 Gold Winner" class = "lazy gold-winner1" data-src="https://www.simplilearn.com/ice9/asset s/home/2020-gold-winner.png" height="9" src="https://www.simplilearn.com/ice9/assets/form\_opacity.png" width="16"/></div><span>2020 Gold Winner</span>For Customer Service by Stevie Awards</a>, <a aria-label="trusted brands" class="lazy" href="https://tldrify.com/zx2" rel="noreferrer nofollow" target="\_blank"><div class="award-img"><img alt="AEGIS GRAHAM BELL AWARDS" class="lazy gold-winner2" data-src="https://www.simplilearn.com/ice9/assets/home/AGBA-2020-winner.png" height="9" src="https://www.simplilearn.com/ice9/assets/form\_opacity.png" width="16"/></div><span>2020 Winner</span>Innovations in Edtech by Aegis Graham Bell Award</a>, <a aria-label="trusted brands" class="l

In [28]:

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

https://trainingindustry.com/top-training-companies/online-learning-li braries/2020-top-online-learning-library-companies/ (https://trainingi ndustry.com/top-training-companies/online-learning-libraries/2020-toponline-learning-library-companies/) https://play.google.com/store/apps/details?id=com.mobile.simplilearn (https://play.google.com/store/apps/details?id=com.mobile.simplilear 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.co m/refer-and-earn) https://www.facebook.com/simplilearn (https://www.facebook.com/simplil earn) https://www.twitter.com/simplilearn (https://www.twitter.com/simplilea https://www.youtube.com/user/Simplilearn (https://www.youtube.com/use r/Simplilearn) https://www.linkedin.com/company/simplilearn (https://www.linkedin.co m/company/simplilearn)

```
In [ ]:
```

1