



Assignment 02: Scrape the Simplilearn Website Resource Page

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 webpage content in a good format

In [1]:

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

In [2]:

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

In [5]:

```
1 #Scrape the web page document
2 web_page = requests.get(url)
3 webPage = web_page.content
```

In [6]:

```
1 #Create the soup object and parse the webpage using an html parser
2 soup = BeautifulSoup(webPage, 'html.parser')
```

In [7]:

```
1 #Close the webpage object
2 web_page.close()
```

In [8]:

```
1 #View the soup contents
2 print(soup)
```

```
<!DOCTYPE html>
<html dir="ltr" lang="en-US" xmlns="https://www.w3.org/1999/xhtml">
<head>
<meta content="text/html; charset=utf-8" http-equiv="Content-Type"/>
<meta content="IE=edge" http-equiv="X-UA-Compatible"/>
<title>Resources | Free Resources to shape your Career - 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.cookiecaw.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
```

In [9]:

```
1 #Use the prettify() method to view the well-formatted html output
2 print(soup.prettify())
```

```
(
<!-- -->
293
<!-- -->
)
</span>
</h3>
<a href="https://www.simplilearn.com/resources/data-science-
business-analytics?source=frs_home">
    View All
</a>
<a class="card" href="https://www.simplilearn.com/role-of-mi
crosoft-excel-certification-in-your-career-rar270-article?source=frs_h
ome">
    <div class="card-img">
         (<https://www.simplilearn.com>)

<https://www.simplilearn.com> (<https://www.simplilearn.com>)

[https://www.simplilearn.com/trends-in-digital-marketing-for-2022-webinar?source=frs\\_home\\_carousal](https://www.simplilearn.com/trends-in-digital-marketing-for-2022-webinar?source=frs_home_carousal) ([https://www.simplilearn.com/trends-in-digital-marketing-for-2022-webinar?source=frs\\_home\\_carousal](https://www.simplilearn.com/trends-in-digital-marketing-for-2022-webinar?source=frs_home_carousal))

[https://www.simplilearn.com/upskilling-and-reskilling-pdf?source=frs\\_home\\_carousal](https://www.simplilearn.com/upskilling-and-reskilling-pdf?source=frs_home_carousal) ([https://www.simplilearn.com/upskilling-and-reskilling-pdf?source=frs\\_home\\_carousal](https://www.simplilearn.com/upskilling-and-reskilling-pdf?source=frs_home_carousal))

[https://www.simplilearn.com/non-data-roles-that-need-data-analytics-skills-guide-pdf?source=frs\\_home\\_carousal](https://www.simplilearn.com/non-data-roles-that-need-data-analytics-skills-guide-pdf?source=frs_home_carousal) ([https://www.simplilearn.com/non-data-roles-that-need-data-analytics-skills-guide-pdf?source=frs\\_home\\_carousal](https://www.simplilearn.com/non-data-roles-that-need-data-analytics-skills-guide-pdf?source=frs_home_carousal))

[https://www.simplilearn.com/getting-started-with-aiops-guide-pdf?source=frs\\_home\\_carousal](https://www.simplilearn.com/getting-started-with-aiops-guide-pdf?source=frs_home_carousal) ([https://www.simplilearn.com/getting-started-with-aiops-guide-pdf?source=frs\\_home\\_carousal](https://www.simplilearn.com/getting-started-with-aiops-guide-pdf?source=frs_home_carousal))

[#data-science-and-business-analytics](#)

[#ai-and-machine-learning](#)

[#project-management](#)

[#data-science-and-business-analytics](#)

### 4: Search and print the resource headers of the Simplilearn webpage

In [50]:

```
1 #View all the headers name
2 for tag in soup.findAll(['h1','h2','h3']):
3 print(tag.next_element)
4 #the div information is obtained from the web page (view page source)
```

## Resources

Trends in Digital Marketing for 2022

Upskilling and Reskilling in 2021: A Global Overview

The Rise of the Data-Driven Professional: 6 Non-Data Roles That Need Data Analytics Skills

Getting Started With AIOps

Data Science & Business Analytics

AI & Machine Learning

Project Management

Cyber Security

Cloud Computing

DevOps

Business and Leadership

Quality Management

Software Development

Agile and Scrum

IT Service and Architecture

Digital Marketing

Big Data

Career Fast-track

Enterprise

Other Segments

## 5: Search resource topics

In [55]:

```
1 #View all the resource topics
2 #the class information is obtained from the web page (view page source)
3 for tag in soup.findAll('h3'):
4 print(tag.next_element)
```

Data Science & Business Analytics

AI & Machine Learning

Project Management

Cyber Security

Cloud Computing

DevOps

Business and Leadership

Quality Management

Software Development

Agile and Scrum

IT Service and Architecture

Digital Marketing

Big Data

Career Fast-track

Enterprise

Other Segments

## 6: View the article names and navigate through them

In [56]:

```
1 #Create article object for finding the list (Hint: the class information is obtained from bs4)
2 from bs4 import BeautifulSoup
3 tag_with_article_only = BeautifulSoup(class_="detail-menu")
4 new_soup = BeautifulSoup(webPage, 'html.parser', parse_only=tag_with_article_only)
5
```

In [57]:

```
1 #View the type of the article object
2 type(new_soup)
```

Out[57]:

bs4.BeautifulSoup

In [59]:

```
1 print(new_soup.prettify())
```

```
n.com/ice9/assets/form_opacity.png" title="Free Mind-map: PMP Certification" width="335"/>
</div>
<div class="content_ebook card-content">
 <div>
 <h4 class="content-title">
 Free Mind-map: PMP Certification
 </h4>
 </div>

 Ebook

 Jun 23, 2021

</div>

</div>
<div class="details-list target" id="cyber-security">
 <h3>
```

In [75]:

```

1 #Find all the article names
2 for tag in new_soup.findAll(class_="content-title"):
3 print(tag.next_element)

```

How to Become a Certified Microsoft Excel Professional?

What is Data Collection? Definition, Types, Tools, and Techniques

Free eBook: Secret to Unlocking Tableau's Hidden Potential

Keras vs Tensorflow vs Pytorch: Understanding the Most Popular Deep Learning Frameworks

Top 10 Machine Learning Algorithms You Need to Know in 2021

Artificial Intelligence Career Guide: A Comprehensive Playbook to Becoming an AI Expert

6 Successful Project Estimation Techniques

6 Interesting Tips on Project Management

Free Mind-map: PMP Certification

Are you the CFO of an Organization? How can you ensure Safe Cyber Security?

Top 5 Cyber Security Jobs in India in 2021

Cybersecurity Trends: Where the Industry Is Heading in an Uncertain 2021

Top AWS Solutions Architect Interview Questions and Answers in 2021

AWS Certification Cost and Type of AWS Certification Exam

Getting Started With Multi-Cloud Architecture

DevOps Engineer Job Description: Skills, Roles and Responsibilities

The Four C's of DevSecOps: Code, Container, Cloud, and Cluster

The Impact of DevOps in 2021

30 Common Business Analyst Interview Questions

Top 10 Most Effective Business Analysis Techniques

Skilling for the Digital Economy: A Role-Based Approach

What Is Six Sigma Certification and Reasons to Get Six Sigma Certification

Six Sigma Certification Cost [Black, Green, White, Yellow Belt]

Free Lean Six Sigma Green Belt Exam Practice Test

Everything You Need to Know About Let in Javascript

What Is Salesforce Data Loader?

Blockchain Career Guide: A Comprehensive Playbook To Becoming A Blockchain Developer

The Benefits of Certified Scrum Master (CSM®) Certification

All About a Certified Scrum Master Salary: Top Paying Industries and Countries

Free eBook: Guide To Scrum Methodology

What Is Strategic Implementation? Aspects and Key Steps in the Implementation Process

What is a Service Level Agreement and the Types of SLA

Free eBook: Roadmap to the TOGAF® Certification

7 Best Marketing Campaigns of 2020 and What You Can Learn from Them for 2021

Content Strategist Job Description: Role, Responsibilities, Salary Trends, and More

Digital Marketing Resume Guide: Your Chance to Win the Digital Marketing Manager Job

Data Science vs. Big Data vs. Data Analytics

Working With Azure to Use and Manage Data Lakes

Data Engineer Interview Guide

How to Measure Training Effectiveness?

What's the Difference Between Leadership vs Management?

State of Upskilling in 2021: Insights from Simplilearn's Survey of Learners

We are Expanding: New Office in Raleigh

Simplilearn and People Matters Partner for Research Study Assessing Onboarding  
 Upskilling and Reskilling in 2021: A Global Overview  
 Top SAP Modules list in 2021 [SAP FI, SAP CO, SAP SD, SAP HCM and More]  
 What Is Salesforce Pardot?  
 Free CCNA Routing & Switching Exam Practice Test

In [109]:

```
1 #Find first article name
2 first = new_soup.find(class_="content_article card-content")
```

In [111]:

```
1 first.div.h4.string
```

Out[111]:

'How to Become a Certified Microsoft Excel Professional?'

In [121]:

```
1 #Find next sibling (next article name)
2 first.parent.next_sibling.div.next_sibling.div.h4.string
```

Out[121]:

'What is Data Collection? Definition, Types, Tools, and Techniques'

In [122]:

```
1 #Find previous sibling (previous article name)
2 first.parent.previous_sibling.div.next_sibling.div.h4.string
```

```


AttributeError Traceback (most recent call
last)
<ipython-input-122-6181ccdbeaa2> in <module>
 1 #Find previous sibling (previous article name)
----> 2 first.parent.previous_sibling.div.next_sibling.div.h4.string

AttributeError: 'NoneType' object has no attribute 'div'
```

In [ ]:

```
1
```