WEB SCRAPPING

Suppose you want some information from a website? Let's say a paragraph on Donald Trump! What do you do? Well, you can copy and paste the information from Wikipedia to your own file. But what if you want to get large amounts of information from a website as quickly as possible? Such as large amounts of data from a website to train a Machine Learning algorithm? In such a situation, copying and pasting will not work! And that's when you'll need to use Web Scraping.

Web scraping, web harvesting, or web data extraction is data scraping used for extracting data from websites. Web scraping software may directly access the World Wide Web using the Hypertext Transfer Protocol or a web browser.

Web scraping is an automatic method to obtain large amounts of data from websites. Most of this data is unstructured data in an HTML format which is then converted into structured data in a spreadsheet or a database so that it can be used in various applications. There are many different ways to perform web scraping to obtain data from websites.

How Web Scrapers Work?

Web Scrapers can extract all the data on particular sites or the specific data that a user wants. Ideally, it's best if you specify the data you want so that the web scraper only extracts that data quickly. For example, you might want to scrape an Amazon page for the types of juicers available, but you might only want the data about the models of different juicers and not the customer reviews.

So, when a web scraper needs to scrape a site, first the URLs are provided. Then it loads all the HTML code for those sites and a more advanced scraper might even extract all the CSS and Javascript elements as well. Then the scraper obtains the required data from this HTML code and outputs this data in the format specified by the user. Mostly, this is in the form of an Excel spreadsheet or a CSV file, but the data can also be saved in other formats, such as a JSON file.

Web scrapping means to pull out the necessary data from a website.

In [1]:
```
#importing the libraries

import requests #used to send http request to the specified URL using get,post,put methods.
                #We can retrieve or push the data to the server using request library.
                #It works as request-response protocol.
                #In this assign we r only retrieving the data from the website

import bs4      #Use for pulling data out of html or xml files.
```

In [2]:
```
request1=requests.get('https://www.flipkart.com/boat-rockerz-255-pro-258-asap-charge-upto-40-hours-playback-bluetooth-headset/p/itm9d3a2c5e5356a?pid=ACCFZ95M5JTZQH3F&lic
request1
```

Out[2]:
```
<Response [200]>
```

In [3]:
```
request1.content  #will show the content in request1, that we,ve got. We will get a raw html doc
```

Out[3]:
```
b'<!doctype html><html lang="en"><head><link href="https://rukminim1.flixcart.com" rel="preconnect"/><link rel="stylesheet" href="//static-assets-web.flixcart.com/fk-p-
linchpin-web/fk-cp-zion/css/app_modules.chunk.905c37.css"/><link rel="stylesheet" href="//static-assets-web.flixcart.com/fk-p-linchpin-web/fk-cp-zion/css/app.chunk.4fb0
ab.css"/><meta http-equiv="Content-type" content="text/html; charset=utf-8"/><meta http-equiv="X-UA-Compatible" content="IE=Edge"/><meta property="fb:page_id" content="
102988293558"/><meta property="fb:admins" content="658873552,624500995,100000233612389"/><meta name="robots" content="noodp"/><link rel="shortcut icon" href="https:///w
ww/promos/new/20150528-140547-favicon-retina.ico"/><link type="application/opensearchdescription+xml" rel="search" href="/osdd.xml?v=2"/><meta property="og:type" conten
t="website"/><meta name="og_site_name" property="og:site_name" content="Flipkart.com"/><link rel="apple-touch-icon" sizes="57x57" href="/apple-touch-icon-57x57.png"/><l
ink rel="apple-touch-icon" sizes="72x72" href="/apple-touch-icon-72x72.png"/><link rel="apple-touch-icon" sizes="114x114" href="/apple-touch-icon-114x114.png"/><link re
l="apple-touch-icon" sizes="144x144" href="/apple-touch-icon-144x144.png"/><link rel="apple-touch-icon" href="/apple-touch-icon-57x57.png"/><meta name="twitter:card" co
ntent="app"><meta name="twitter:site" content="@flipkart"><meta name="twitter:creator" content="@flipkart"><meta name="twitter:title" content="boAt Rockerz 255 Pro+ /25
8 Pro+ with ASAP Charge and upto 40 Hours Playback Bluetooth Headset Price in India - Buy boAt Rockerz 255 Pro+ /258 Pro+ with ASAP Charge and upto 40 Hours Playback Bl
uetooth Headset Online - boAt : Flipkart.com"><meta name="twitter:description" content="Shop for electronics, apparels & more using our Flipkart app Free shipping & COD
."><meta property="twitter:image" content="http://rukmini1.flixcart.com/image/300/300/l31x2fk0/headphone/a/s/h/-original-image9ehehz8amg2.jpeg"/><meta name="twitter:app
:country" content="in"><meta name="al:ios:app_name" content="Flipkart"><meta name="al:ios:app_store_id" content="742044692"><meta name="twitter:app:name:iphone" content
="Flipkart"><meta name="twitter:app:id:iphone" content="742044692"><meta name="twitter:app:url:iphone" content="http://dl.flipkart.com/dl/home?"><meta name="twitter:app
:name:ipad" content="Flipkart"><meta name="twitter:app:id:ipad" content="742044692"><meta name="twitter:app:url:ipad" content="http://dl.flipkart.com/dl/home?"><meta na
me="twitter:app:name:googleplay" content="Flipkart"><meta name="twitter:app:id:googleplay" content="com.flipkart.android"><meta name="twitter:app:url:googleplay" conten
t="http://dl.flipkart.com/dl/home?"><style>#container {\n\t\t\theight: 100%;\n\t\t}</style>  <link rel="stylesheet" href="///fk-cp-zion/css/Product.chunk.5723d0.css"/><
title>boAt Rockerz 255 Pro+ /258 Pro+ with ASAP Charge and upto 40 Hours Playback Bluetooth Headset Price in India - Buy boAt Rockerz 255 Pro+ /258 Pro+ with ASAP Charg
e and upto 40 Hours Playback Bluetooth Headset Online - boAt : Flipkart.com</title><meta name="og_title" property="og:title" content="boAt Rockerz 255 Pro+ /258 Pro+ wi
th ASAP Charge and upto 40 Hours Playback Bluetooth Headset Price in India - Buy boAt Rockerz 255 Pro+ /258 Pro+ with ASAP Charge and upto 40 Hours Playback Bluetooth H
eadset Online - boAt : Flipkart.com"/><meta name="Keywords" content="boAt Rockerz 255 Pro+ /258 Pro+ with ASAP Charge and upto 40 Hours Playback Bluetooth Headset Price
, Buy boAt Rockerz 255 Pro+ /258 Pro+ with ASAP Charge and upto 40 Hours Playback Bluetooth Headset Online, boAt Rockerz 255 Pro+ /258 Pro+ with ASAP Charge and upto 40
Hours Playback Bluetooth Headset Features"/> <meta name="Description" content="Buy boAt Rockerz 255 Pro+ /258 Pro+ with ASAP Charge and upto 40 Hours Playback Bluetooth
Headset for Rs.3990 Online, Also get boAt Rockerz 255 Pro+ /258 Pro+ with ASAP Charge and upto 40 Hours Playback Bluetooth Headset Specs &amp;amp; Features. Only Genuine Pr
oducts. 30 Day Replacement Guarantee. Free Shipping. Cash On Delivery!"/><meta property="og:description" content="Buy boAt Rockerz 255 Pro+ /258 Pro+ with ASAP Charge a
nd upto 40 Hours Playback Bluetooth Headset for Rs.3990 Online, Also get boAt Rockerz 255 Pro+ /258 Pro+ with ASAP Charge and upto 40 Hours Playback Bluetooth Headset S
pecs &amp;amp; Features. Only Genuine Products. 30 Day Replacement Guarantee. Free Shipping. Cash On Delivery!"/>  <link rel="canonical" href="https://www.flipkart.com/boat
-rockerz-255-pro-258-asap-charge-upto-40-hours-playback-bluetooth-headset/p/itm9d3a2c5e5356a"/> <link rel="alternate" href="android-app://com.flipkart.android/flipkart/
de_pp_seg_--_ITM9D3A2C5E5356A"/>  <meta name="og_image" property="og:image" content="http://rukmini1.flixcart.com/image/300/300/l31x2fk0/headphone/a/s/h/-original-image
9ehehz8amg2.jpeg"/>  <meta name="og_url" property="og:url" content="https://www.flipkart.com/boat-rockerz-255-pro-258-asap-charge-upto-40-hours-playback-bluetooth-heads
et/p/itm9d3a2c5e5356a"/><meta name="msvalidate.01" content="F4EEB3A0AFFDD385992A06E6920C0AC3"/> <!--[if IE 9 ]>\n\t<script nonce="7488550108402701482">\n\t\twindow.brow
serVersion = \'IE9\';\n\t\tdocument.documentElement.className += \' ie9\';\n\t</script>\n\t<![endif]--><!--[if !IE]><!--><script nonce="7488550108402701482">if (/*@cc_o
n!@*/false) {\n\t\t\twindow.browserVersion = \'IE10\';\n\t\t\tdocument.documentElement.className += \' ie10\';\n\t\t}</script><!--<![endif]-->   <script nonce="74885501
08402701482">!function(){if("PerformancePaintTiming" in window){var e=window.__perf={e:[]};e.o=new PerformanceObserver(function(n){e.e=e.e.concat(n.getEntries())});try{
e.o.observe({entryTypes:["paint"]})}catch(e){}}}();\n\n\t\tif (requestAnimationFrame && performance && performance.mark) {\n\t\t\trequestAnimationFrame(function () {\n\
t\t\t\twindow.first_paint = performance.now();\n\t\t\t\tperformance.mark("first_paint");\n\t\t\t});\n\t\t}\n\n        // web-vitals polyfill for FID\n        !function(
n,e){var t,o,i,c=[],f={passive:!0,capture:!0},r=new Date,a="pointerup",u="pointercancel";function p(n,c){t||(t=c,o=n,i=new Date,w(e),s())}function s(){o>=0&&o<i-r&&(c.f
orEach(function(n){n(o,t)}),c=[])}function l(t){if(t.cancelable){var o=(t.timeStamp>1e12?new Date:performance.now())-t.timeStamp;"pointerdown"==t.type?function(t,o){fun
```

t": "android-app://com.flipkart.android/flipkart/de_sq_seg_-search.flipkart.com-_{search_term_string}",\n "query-input": "required name=search_term_string"\n }]\n}</script><script nonce="7488550108402701482" id="is_script">window.__INITIAL_STATE__ = {};</script> <link data-chunk="NavMenu" rel="stylesheet" href="//static-assets-web.flixcart.com/fk-p-linchpin-web/fk-cp-zion/css/NavMenu.chunk.10ffe7.css"> <script id="__LOADABLE_REQUIRED_CHUNKS__" type="application/json">[121]</script><script id="__LOADABLE_REQUIRED_CHUNKS___ext" type="application/json">{"namedChunks":["NavMenu"]}</script>\n<script async data-chunk="NavMenu" src="//static-assets-web.flixcart.com/fk-p-linchpin-web/fk-cp-zion/js/NavMenu.chunk.3a120f66.js"></script><script nonce="7488550108402701482">window.omniture={tpv:[],tl:[]};window.omniture.trackPageView=function(){window.omniture.tpv.push(arguments)};window.omniture.trackLink=function(){window.omniture.tl.push(arguments)}\n\n if(window.location.host === \'www.flipkart.com\' || /\\.store\\.flipkart\\.com/i.test(window.location.host)){\n window.s_account = "flipkart-prd"; //omniture account id\n }else{\n window.s_account = "flipkart-prd-test";\n }\n window.s_isDesktop = true;</script><script nonce="7488550108402701482">function isMobile(){\n\t\tif (!!(navigator.userAgentData && navigator.userAgentData.mobile) || navigator.userAgent.indexOf("Mobile") !== -1){\n\t\t\treturn true;\n\t\t}\n\t\telse {\n\t\t\treturn false;\n\t\t}\n\t}\n\n\tif (\'serviceWorker\' in navigator && !isMobile()) {\n\t\tnavigator.serviceWorker.register(\'/sw.js\').then(function(registration) {\n\t\t\tconsole.log(\'ServiceWorker registration successful with scope: \', registration.scope);\n\t\t\tnavigator.onLine && navigator.serviceWorker.controller && navigator.serviceWorker.controller.postMessage(\'BANDWIDTH_COMPUTE\');\n\t\t\tnavigator.serviceWorker.addEventListener(\'message\', function (event) {\n\t\t\t\ttry {\n\t\t\t\t\tconst bandwidthData = JSON.parse(event.data);\n\t\t\t\t\tif (bandwidthData && bandwidthData.data) {\n\t\t\t\t\t\twindow.bandwidth = bandwidthData.data.downlink + \'_\' + bandwidthData.data.computeMethod + \'_\' + bandwidthData.data.effectiveType + \'_\' + bandwidthData.data.type;\n\t\t\t\t}\n\t\t\t}catch(err) {\n\t\t\t\tconsole.error(\'JSON.parse failed:\', err);\n\t\t\t}\n\t\t});\n\t})\n\t.catch(function(err) {\n\t\tconsole.error(\'ServiceWorker registration failed: \', err);\n\t});\n} else if(\'serviceWorker\' in navigator && isMobile()){\n navigator.serviceWorker.getRegistration().then(function(registration) {\n var serviceWorkerUnregistered=false;\n if(registration) {\n registration.unregister();\n serviceWorkerUnregistered=true;\n }\n serviceWorkerUnregistered && window.location.reload();\n });\n }</script> <script src="///www/linchpin/batman-returns/omni/omni16.js" defer="defer" async nonce="7488550108402701482" id="omni_script"></script> <script nonce="7488550108402701482">window._pxAppId = \'PXgNtTli3A\';</script><script nonce="7488550108402701482" src="///px/gNtTli3A/init.js" async></script></body></html>'

In [4]:
```
#Now that we,ve fetched the content of webpage,hence we'll save it.
#To save the fetched content of a webpage, We use  BeautifulSoup method from bs4 library

soup=bs4.BeautifulSoup(request1.text)

#Till now we,ve fetched the data from a website and saved it.
#Now time for fetching reviews,comments.
```

In [6]:
```
#Fetching Reviews/Comments
#reviews=soup.find_all('div',{'class':'t-ZTKy'});
#for review in reviews:
 #   print(review.get_text() +"\n")

reviews=soup.find_all('div',{'class':'t-ZTKy'});
for review in reviews:
    print(review.get_text() + '\n')
```

1. Material Quality - Not Bad...2. Sound Quality - Awesome...3. Bass - Decent Quality...4. Battery Backup - Very Good...5. Design & Bulid Quality - Very nice...6. Fitting - Oky Type...Overall very nice and Velue for money product...Really i love the product...🤩READ MORE

Battery backup is best and full charge very quickly. Sound quality is best .I love it. You can buy it without any hesitation.READ MORE

Delivered in 20 February.. today is 28 February and i didn't charge it till now still battery percentage is 70%.. i use it daily for online classes, 1 movie and 3 hours For music approx 9-10 hours daily so Battery is very Powerfull.it's housing is made by plastic and neck wire of somewhat soft silicone material cable is flat long and strong enough..Bass is above the best if you use right size of ear buds..even in 90% volume vocals are crystal clear treble is well balanced..Call quality is far...READ MORE

One of the best Bluetooth Boat. 1- Battery Backup Amazing 2- light wait 3-Awesome Sound quality4-Bass good5- Design normal but good not bestREAD MORE

Writing this review after using more than 15 days. 255F pro+ has super sound quality. Bass is decent and fair enough. Most important highlight of this headset is battery life,it's amazing. Runtime depends on usage but definitely it will give you more than any other headset.Should go for it.READ MORE

Nice product and fast delivery . I am happy nice job Flipkart thanksREAD MORE

Awesome battery backup And Great Sound quality . Perfectly fits in the ear. What else you need. Thanks to bOAT & FLIPKART both for this Wonderful Product.READ MORE

This product is acctullly best. best coloty and design.But sound is little low not bed its a perfect but little low. bass is very powerful and clear. battery backup is very best in this price range. definately you purchase this product.READ MORE

Great product., battery backup up to 14 days use , sound quality is great , I am happy with the product 😊😊😊😊READ MORE

I got this product on 5th oct 2021 , today is 15th october 2021.Iam using this product now with out charging from delivery day...The batterry backup is owsomSound quality amd bass are verry nice and morethan worth for your money...But build quality is not pretty good. But it is also value for money in this range...The magnetic lock is not perfect... Magnetic power is little lower.Some features are miss in this headphone ...like in oneplus headphones.. That is automatically off when we...READ MORE

In [ ]:
```
#overall avg rating
avg_ratings  = soup.find('div',{'class':'_2d4LTz'}).get_text();
```

In [ ]:
```
avg_ratings
```

In [ ]:
```
#Individual ratings

individual_ratings= soup.find_all('div',{'class':'_3LWZlK _1BLPMq'})
for ind_rating in individual_ratings:
        print(ind_rating.get_text()+"\n")
```

In [ ]:
```
#tags
#tags=soup.find('tag_name',{'class':'class_name'}).get_text();
#tags
```

In [ ]:
```
#cust_names
customer_names=soup.find_all('p',{'class':'_2sc7ZR _2V5EHH'});
for cust_names in customer_names:
    print(cust_names.get_text() + "\n")
```

In [ ]:
```

```