

Web Scraping

Requests library

Web scraping starts with sending HTTP requests, such as POST or GET, to a website's server, which returns a response containing the needed data

```
In [31]: #pip install requests
```

```
In [ ]: pip install BeautifulSoup
```

Requests library provides easy methods for sending HTTP GET and POST requests.
For example, the function to send an HTTP Get request is named as get()

```
In [2]: import requests
url="https://oxylabs.io/"
response = requests.get(url)
#print(response.text)
```

But this library has a limitation in that it does not parse the extracted HTML data, i.e., it cannot convert the data into a more readable format for analysis.

Also, it cannot be used to scrape websites that are written using purely JavaScript.

Beautiful Soup

Beautiful Soup is a Python library that works with a parser to extract data from HTML and can turn even invalid markup into a parse tree. However, this library is only designed for parsing and cannot request data from web servers in the form of HTML documents/files. For this reason, it is mostly used alongside the Python Requests Library. Note that BeautifulSoup makes it easy to query and navigate the HTML, but still requires a parser.
The following example demonstrates the use of the html.parser module, which is part of the Python Standard Library.

Part 1 – Get the HTML using Requests

```
In [ ]: import requests
url="https://oxylabs.io/blog"
response = requests.get(url)
```

Part 2 – Find the element

```
In [3]: from bs4 import BeautifulSoup
soup = BeautifulSoup(response.text, 'html.parser')
print(soup.title)
```

<title>Innovative Proxy Service to Gather Data at Scale | Oxylabs</title>

Active Stock Data Using Web Scraping

```
In [16]: from bs4 import BeautifulSoup
import requests
```

```
In [17]: url="https://finance.yahoo.com/most-active"
response=requests.get(url)
response.text
```

from bs4 import BeautifulSoup

import requests

[illegible]

[illegible]


```

01:
02:
03:
04:
05:
06:
07:
08:
09:
10:
11:
12:
13:
14:
15:
16:
17:
18:
19:
20:
21:
22:
23:
24:
25:
26:
27:
28:
29:
30:
31:
32:
33:
34:
35:
36:
37:
38:
39:
40:
41:
42:
43:
44:
45:
46:
47:
48:
49:
50:
51:
52:
53:
54:
55:
56:
57:
58:
59:
60:
61:
62:
63:
64:
65:
66:
67:
68:
69:
70:
71:
72:
73:
74:
75:
76:
77:
78:
79:
80:
81:
82:
83:
84:
85:
86:
87:
88:
89:
90:
91:
92:
93:
94:
95:
96:
97:
98:
99:
100:
101:
102:
103:
104:
105:
106:
107:
108:
109:
110:
111:
112:
113:
114:
115:
116:
117:
118:
119:
120:
121:
122:
123:
124:
125:
126:
127:
128:
129:
130:
131:
132:
133:
134:
135:
136:
137:
138:
139:
140:
141:
142:
143:
144:
145:
146:
147:
148:
149:
150:
151:
152:
153:
154:
155:
156:
157:
158:
159:
160:
161:
162:
163:
164:
165:
166:
167:
168:
169:
170:
171:
172:
173:
174:
175:
176:
177:
178:
179:
180:
181:
182:
183:
184:
185:
186:
187:
188:
189:
190:
191:
192:
193:
194:
195:
196:
197:
198:
199:
200:
201:
202:
203:
204:
205:
206:
207:
208:
209:
210:
211:
212:
213:
214:
215:
216:
217:
218:
219:
220:
221:
222:
223:
224:
225:
226:
227:
228:
229:
230:
231:
232:
233:
234:
235:
236:
237:
238:
239:
240:
241:
242:
243:
244:
245:
246:
247:
248:
249:
250:
251:
252:
253:
254:
255:
256:
257:
258:
259:
260:
261:
262:
263:
264:
265:
266:
267:
268:
269:
270:
271:
272:
273:
274:
275:
276:
277:
278:
279:
280:
281:
282:
283:
284:
285:
286:
287:
288:
289:
290:
291:
292:
293:
294:
295:
296:
297:
298:
299:
300:
301:
302:
303:
304:
305:
306:
307:
308:
309:
310:
311:
312:
313:
314:
315:
316:
317:
318:
319:
320:
321:
322:
323:
324:
325:
326:
327:
328:
329:
330:
331:
332:
333:
334:
335:
336:
337:
338:
339:
340:
341:
342:
343:
344:
345:
346:
347:
348:
349:
350:
351:
352:
353:
354:
355:
356:
357:
358:
359:
360:
361:
362:
363:
364:
365:
366:
367:
368:
369:
370:
371:
372:
373:
374:
375:
376:
377:
378:
379:
380:
381:
382:
383:
384:
385:
386:
387:
388:
389:
390:
391:
392:
393:
394:
395:
396:
397:
398:
399:
400:
401:
402:
403:
404:
405:
406:
407:
408:
409:
410:
411:
412:
413:
414:
415:
416:
417:
418:
419:
420:
421:
422:
423:
424:
425:
426:
427:
428:
429:
430:
431:
432:
433:
434:
435:
436:
437:
438:
439:
440:
441:
442:
443:
444:
445:
446:
447:
448:
449:
450:
451:
452:
453:
454:
455:
456:
457:
458:
459:
460:
461:
462:
463:
464:
465:
466:
467:
468:
469:
470:
471:
472:
473:
474:
475:
476:
477:
478:
479:
480:
481:
482:
483:
484:
485:
486:
487:
488:
489:
490:
491:
492:
493:
494:
495:
496:
497:
498:
499:
500:
501:
502:
503:
504:
505:
506:
507:
508:
509:
510:
511:
512:
513:
514:
515:
516:
517:
518:
519:
520:
521:
522:
523:
524:
525:
526:
527:
528:
529:
530:
531:
532:
533:
534:
535:
536:
537:
538:
539:
540:
541:
542:
543:
544:
545:
546:
547:
548:
549:
550:
551:
552:
553:
554:
555:
556:
557:
558:
559:
560:
561:
562:
563:
564:
565:
566:
567:
568:
569:
570:
571:
572:
573:
574:
575:
576:
577:
578:
579:
580:
581:
582:
583:
584:
585:
586:
587:
588:
589:
590:
591:
592:
593:
594:
595:
596:
597:
598:
599:
600:
601:
602:
603:
604:
605:
606:
607:
608:
609:
610:
611:
612:
613:
614:
615:
616:
617:
618:
619:
620:
621:
622:
623:
624:
625:
626:
627:
628:
629:
630:
631:
632:
633:
634:
635:
636:
637:
638:
639:
640:
641:
642:
643:
644:
645:
646:
647:
648:
649:
650:
651:
652:
653:
654:
655:
656:
657:
658:
659:
660:
661:
662:
663:
664:
665:
666:
667:
668:
669:
670:
671:
672:
673:
674:
675:
676:
677:
678:
679:
680:
681:
682:
683:
684:
685:
686:
687:
688:
689:
690:
691:
692:
693:
694:
695:
696:
697:
698:
699:
700:
701:
702:
703:
704:
705:
706:
707:
708:
709:
710:
711:
712:
713:
714:
715:
716:
717:
718:
719:
720:
721:
722:
723:
724:
725:
726:
727:
728:
729:
730:
731:
732:
733:
734:
735:
736:
737:
738:
739:
740:
741:
742:
743:
744:
745:
746:
747:
748:
749:
750:
751:
752:
753:
754:
755:
756:
757:
758:
759:
760:
761:
762:
763:
764:
765:
766:
767:
768:
769:
770:
771:
772:
773:
774:
775:
776:
777:
778:
779:
780:
781:
782:
783:
784:
785:
786:
787:
788:
789:
790:
791:
792:
793:
794:
795:
796:
797:
798:
799:
800:
801:
802:
803:
804:
805:
806:
807:
808:
809:
810:
811:
812:
813:
814:
815:
816:
817:
818:
819:
820:
821:
822:
823:
824:
825:
826:
827:
828:
829:
830:
831:
832:
833:
834:
835:
836:
837:
838:
83
```


[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

```
Out[29]: 25
```

Stock Prices

```
In [30]: for listing in stock_table.find_all('tr'):
           price=listing.find('td',attrs='aria-label':"Price (Intraday)")
           prices.append(price.text)
```

```
In [32]: for listing in stock_table.find_all('tr'):
          change=listing.find('td',attr={'aria-label':"Change"})
          changes.append(change.text)

In [33]: len(changes)
```

```
Out [33]: 25
```

total_volume

```
In [34]: for listing in stock_table.find_all('tr'):
          total_volume+=listing.find('td',attrs={'aria-label':"Volume"}).text
```

```
In [36]: for listing in stock_table.find_all('tr'):
          percent_change=listing.find('td', attrs={'aria-label':"% Change"})
          percent_changes.append(percent_change.text)

In [37]: len(percent_changes)
```

```
Out[37]: 25
```

Market_caps

```
In [38]: for listing in stock_table.find_all('tr'):
```

```
market_caps.append(market_cap.text)

In [39]: len(market_caps)

Out[39]: 25
```

```
import pandas as pd

In [41]: a=[{"symbol":"codes",
             "Name":names,
             "Price":prices,
             "Change":changes,
             "% Change":percent_changes,
```

```

"Volume":total_volumes,
"Market Capital":market_caps
}

In [42]: dfepd.DataFrame(
{
"Symbol":codes,
"Name":names
})
```

```

"Change":changes,
"% change":percent_changes,
"Volume":total_volumes,
"Market Capital":market_caps
}
}

```

Symbol	Name	Price	Change	% Change	Volume	Market Capital
0 TWTR	Twitter, Inc.	45.08	-0.77	-1.68%	252,397M	34,422B
1 BAC	Bank of America Corporation	37.57	-1.25	-3.22%	79,032M	302,997B
2 AAPL	Apple Inc.	165.29	-5.11	-3.00%	75,329M	2,697T

3	AMD	Advanced Micro Devices, Inc.	93.00	-4.66	-4.79%	72.4822M	151,144.28
4	WFC	Wells Fargo & Company	46.35	-2.19	-4.51%	65.583M	176,204.28
5	NVDA	NVIDIA Corporation	212.58	+9.45	+4.26%	56.91M	529,749.98
6	T	AT&T Inc.	19.54	+0.12	+0.62%	52.7077M	139,966.68
7	AAI	American Airlines Group Inc.	19.00	+0.05	+0.26%	42.8222M	12,334.48
8	SOFI	SoFi Technologies Inc.	7.36	-0.35	-4.60%	42.1598M	5,805.48

9	PBR	Petróleo Brasileiro S.A. - Petrobras	14.71	-0.03	-0.21%	42.065M	101.164B
10	NIO	NIO Inc.	19.65	-0.77	-3.77%	43.778M	35.945B
11	F	Ford Motor Company	15.48	-0.03	-0.19%	39.313M	62.211B
12	WBD	Warner Bros. Discovery, Inc.	24.88	-1.12	-4.31%	37.862M	12.49B
13	PGC	PG&E Corporation	12.30	-0.55	-4.28%	37.224M	24.445B

14	INTC	Intel Corporation	45.67	+1.34	-2.85%	36.182M	186.731B
15	C	Citigroup Inc.	50.93	+0.78	+1.56%	34.043M	100.458B
16	SWN	Southwestern Energy Company	8.19	-0.07	-0.85%	33.52M	9.141B
17	ITUB	Ita Unionbank Holding S.A.	5.54	-0.02	-0.36%	32.21M	54.181B
18	TELL	Tellurian Inc.	6.19	+0.40	+6.91%	30.185M	3.205B

19	BBD	Banco Bradesco S.A.	4,500.00	-0.0500	-1.10%	30.49M	47,838
20	DAL	Delta Air Lines, Inc.	42.36	+1.34	+3.27%	29,129M	27,107B
21	MSFT	Microsoft Corporation	279.83	-7.79	-2.71%	28,037M	2,098T
22	VALE	Vale S.A.	19.48	-0.43	-2.16%	25,586M	94,376B
23	AMC	AMC Entertainment Holdings, Inc.	18.02	-0.51	-2.75%	25,334M	9,258
24	CCL	Carnival Corporation & plc	10.66	-0.06	-0.56%	26,337M	22,009B

```
In [44]: df.to_csv("active_stock_data.csv")

In [45]: stocks=pd.read_csv("active_stock_data.csv")

In [47]:
```

Unnamed: 0	Symbol	Name	Price	Change	% Change	Volume	Market Capital
20	DAL	Delta Air Lines, Inc.	42.36	1.34	+3.27%	29,129M	27,1078
21	MSFT	Microsoft Corporation	279.83	-7.79	-2.71%	28,037M	2,098T
22	VALE	Vale S.A.	19.48	-0.43	-2.16%	25,586M	94,3768

23	23	AMC	AMC Entertainment Holdings, Inc.	18.02	-0.51	-2.75%	25.34M	9.58
24	24	CCL	Carnival Corporation & plc	19.56	0.05	+0.26%	25.317M	22.991B

0	TWTR	Twitter, Inc.	45.08	-0.77	-1.68%	252.397M	34,422B
1	BAC	Bank of America Corporation	37.57	-1.25	-3.22%	79.032M	302.997B
2	APPL	Apple Inc.	165.29	-5.11	-3.00%	75.323M	26.97T
3	AMLD	Advanced Micro Devices, Inc.	93.06	-4.68	-4.79%	72.482M	151.442B
4	WFC	Wells Fargo & Company	46.35	-2.19	-4.51%	65.583M	176.204B

6	T	AT&T Inc.	19.54	+0.12	+0.62%	52.707M	139.9668
7	AAL	American Airlines Group Inc.	19.00	+0.05	+0.26%	42.822M	12.3448
8	SOFI	Sofi Technologies, Inc.	7.26	-0.35	-4.60%	42.198M	5.8058
9	PBR	Petróleo Brasileiro S.A. - Petrobras	14.71	-0.03	-0.21%	42.065M	101.1648
10	NIO	NIO Inc.	19.65	-0.77	-3.77%	43.778M	35.9458

11	F	Ford Motor Company	15.48	-0.03	-0.19%	39.313M	62.211%
12	WBD	Warner Bros. Discovery, Inc.	24.88	-1.12	-4.31%	37.862M	12.49%
13	PCG	PG&E Corporation	12.30	-0.55	-4.28%	37.224M	24.445%
14	INTC	Intel Corporation	45.67	-1.34	-2.85%	36.182M	186.731%
15	C	Citigroup Inc.	50.93	+0.78	+1.56%	34.043M	100.458%

16	SWN	Southwestern energy Company	8.19	-0.07	-0.85%	33.52M	9.1418
17	ITUB	Itaú Unibanco Holding S.A.	5.54	-0.02	-0.36%	32.21M	54.1818
18	TBDL	Tellerian Inc.	6.19	+0.40	+6.91%	30.185M	3.2098
19	BBD	Banco Bradesco S.A.	4.5000	-0.0500	-1.10%	30.49M	47.8338
20	DAL	Delta Air Lines, Inc.	42.36	+1.34	+3.27%	29.129M	27.1078
21	MSFT	Microsoft Corporation	279.83	-7.79	-2.71%	28.037M	2.0687

22	VALE	Vale S.A.	19.48	-0.43	-2.16%	25.586M	94.3768
23	AMC	AMC Entertainment Holdings, Inc.	18.02	-0.51	-2.75%	25.334M	9.258
24	CCL	Carnival Corporation & plc	19.56	+0.05	+0.26%	25.317M	22.9918