

lazada

June 24, 2021

```
[2]: import requests
import time
import random
from random import randint
import requests
from bs4 import BeautifulSoup
import pandas as pd
import json
import urllib.parse
from time import sleep
from openpyxl.workbook import Workbook
```

```
[3]: finaldata = []
```

```
[4]: keywords = ['lego 42121',
                 'lego 76160',
                 'lego 76191',
                 'lego 71372',
                 'lego 10941',
                 'lego 11013',
                 'lego 76383',
                 'lego 10951',
                 'lego 41936',
                 'lego 42123',
                 'lego 43191',
                 'lego 60287',
                 'lego 60291',
                 'lego 60304',
                 'lego 71736',
                 'lego 31115',
                 'lego 76901',
                 'lego 11001',
                 'lego 10872',
                 'lego 10920',
                 'lego 10915',
                 'lego 75551',
                 'lego 71742',
```

```
'lego 71360',  
'lego 71368',  
'lego 76165']
```

```
[5]: def getCurrentEpochTime():  
      milliseconds = int(round(time.time() * 1000))  
      return milliseconds
```

```
[6]: ses = requests.Session()  
mmstat = "https://gm.mmstat.com/fsp.1.1"  
headers0 = {'user-agent': 'Mozilla/5.0 (Windows NT 6.0; Win64; x64) AppleWebKit/  
→537.36 (KHTML, like Gecko) Chrome/51.0.2704.79 Safari/537.36 Edge/14.14393',  
            'accept-language': 'en-GB,en;q=0.5',  
            'accept-encoding': 'gzip, deflate, br',  
            'content-type': 'text/plain; charset=UTF-8',  
            'content-length': '958',  
            'origin': 'https://www.lazada.sg',  
            'referer': 'https://www.lazada.sg/',  
            'te': 'trailers'}  
REQUESTBODY = b'{"gmkey":"OTHER","gokey":  
→"c1=24cb9fab896eae862cd1e05be46c159&delay=0&hash=&last_pos=0%252C0&msg=Daraz-mm-web%2520%2  
→lazada.  
→sg%252Fcatalog%252F&patch_ver=-&pid=punish-page&query=_keyori%253Dss%2526from%253Dinput%252  
→0%2520(Windows%2520NT%25206.  
→0%253B%2520Win64%253B%2520x64)%2520AppleWebKit%252F537.  
→36%2520(KHTML%252C%2520like%2520Gecko)%2520Chrome%252F51.0.2704.  
→79%2520Safari%252F537.36%2520Edge%252F14.  
→14393&referrer=&rel=&scr=1920x1080&spm_a=&spm_b=&title=&tracker_ver=4.0.  
→0&type=11&ua=Mozilla%252F5.0%2520(Windows%2520NT%25206.  
→0%253B%2520Win64%253B%2520x64)%2520AppleWebKit%252F537.  
→36%2520(KHTML%252C%2520like%2520Gecko)%2520Chrome%252F51.0.2704.  
→79%2520Safari%252F537.36%2520Edge%252F14.  
→14393&uid=24cb9fab896eae862cd1e05be46c159","logtype":"2"}'  
res0 = ses.post(mmstat, headers=headers0, data=REQUESTBODY)  
# print(res0.headers)  
sca = res0.cookies.get('sca')
```

```
[7]: url1 = 'https://sg.mmstat.com/v.gif?  
→logtype=1&title=Sorry%2C%20we%20have%20detected%20unusual%20traffic%20from%20your%20network  
→&pre=&scr=1920x1080&_p_url=https%3A%2F%2Fwww.lazada.  
→sg%2Fcatalog%2F%3F_keyori%3Dss%26from%3Dinput%26page%3D1%26q%3Dlego%2Bsuper%2Bheroes%2B7616  
→0.0.0.  
→12ea720bk56Fp8&category=&uidaplus=&aplus&yunid=&&trid=0b1195ac16242090565304668e18a4&asid=A  
→15.6&jsver=aplus_std&pver=0.7.  
→11&ps_i=AtvbWCF0BFxy5LNEYujErjyi9S0vs0YE&pc_i=dd1LzJIaYkuiQ2qD4w76tNGWBDqdu383&_p_ref=&_p_u  
headers0.update({"cookie": sca})  
aptResp = ses.get(url1, headers=headers0)
```

```
atpsida = aptResp.cookies.get('atpsida')
cna = aptResp.cookies.get('cna')
```

```
[8]: apt2_url = 'https://sg.mmstat.com/v.gif?
↳logtype=1&title=lego%20super%20heroes%2076160%20-%20Buy%20lego%20super%20heroes%2076160%20a
↳lazada.sg&pre=https%3A%2F%2Fwww.lazada.
↳sg%2Fcatalog%2F%3F_keyori%3Dss%26from%3Dinput%26page%3D1%26q%3Dlego%2Bsuper%2Bheroes%2B7616
↳lazada.
↳sg%2Fcatalog%2F%3F_keyori%3Dss%26from%3Dinput%26page%3D1%26q%3Dlego%2Bsuper%2Bheroes%2B7616
↳+ \
    '&spm-cnt=a2o42.searchlist.0.0.12ea5675du9NP3&cfgver=1.
↳0&apls&asid=AQAAAAA67xdgHA3hZAAAAAAcxAAg7RzJQg==&p=1&o=winVista&b=edge14&s=1920x1080&w=edg
↳15.6&jsver=apls_int&pver=0.7.
↳11&ps_i=AtvbWCF0BFxy5LNEYUjErjyi9S0vs0YE&pc_i=dd1LzJIaYkuiQ2qD4w76tNGWBDqdu383&lzd_pg_type=
↳lazada.sg&p_meta_robots=index%2Cfollow&p_canonical=https%3A%2F%2Fwww.
↳lazada.
↳sg%2Fcatalog%2F&_hng=SG%257Cen-SG%257CSGD%257C702&>tag=1&stag=-1&lstag=-1&_slog=0'
headers0.update({"cookie": "sca="+sca+";"+"atpsida="+atpsida+";"})
apt2Res = ses.get(apt2_url, headers=headers0)
```

```
[9]: x5sec_url = ''https://www.lazada.sg/catalog/____tmd____/slide?
↳slidedata=%7B%22a%22%3A%22X82Y__be31bf05f5bc16c4a763907b48e1c0af%22%2C%22t%22%3A%2224cb9fab
headers2 = {
    'authority': 'www.lazada.sg',
    'cache-control': 'max-age=0',
    'sec-ch-ua': '"Chromium";v="92", " Not A;Brand";v="99", "Google Chrome";
↳v="92"',
    'sec-ch-ua-mobile': '?0',
    'upgrade-insecure-requests': '1',
    'user-agent': 'Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36 (KHTML,
↳like Gecko) Chrome/92.0.4515.51 Safari/537.36',
    'accept': 'text/html,application/xhtml+xml,application/xml;q=0.9,image/
↳avif,image/webp,image/apng,*/*;q=0.8,application/signed-exchange;v=b3;q=0.9',
    'sec-fetch-site': 'same-origin',
    'sec-fetch-mode': 'navigate',
    'sec-fetch-user': '?1',
    'sec-fetch-dest': 'document',
    'referrer': 'https://www.lazada.sg/catalog/?
↳_keyori=ss&from=input&page=1&q=lego+super+heroes+76160&sort=priceasc',
    'accept-language': 'en-US,en;q=0.9',
}
```

[10]:

```

cust_cookie1 = '_uab_collina=162420906464852886455375;␣
↳t_uid=dd1LzJIaYkuiQ2qD4w76tNGWBDqdu383;␣
↳t_sid=AtvbWCF0BFxy5LNEYujErjyi9S0vs0YE; utm_channel=NA;␣
↳cna=qGRWGaAGK1kCacXoPfgWqu5o'
newCookie1 = cust_cookie1+"t_fv="+str(getCurrentEpochTime())+";"+cna+"cna=";
headers2.update({"cookie": newCookie1})
res1 = requests.request("GET", x5sec_url, headers=headers2)

x5sec = res1.cookies.get('x5sec')

```

```

[11]: csr_url = "https://member.lazada.sg/user/api/getCsrfToken"
res2 = ses.get(csr_url)
tb_token = res2.cookies['_tb_token_']
anon_id = res2.cookies['anon_uid']
lzd_sid = res2.cookies['lzd_sid']

cust_cookie2 = 't_uid=dd1LzJIaYkuiQ2qD4w76tNGWBDqdu383;
↳cna=qGRWGaAGK1kCacXoPfgWqu5o;hng=SG|en-SG|SGD|702;client_type=desktop;'
newCookie2 = cust_cookie2+"_tb_token_" + tb_token + ";" + "anon_uid=" + anon_id + ";"
↳"+ "lzd_sid=" + \
    lzd_sid + ";" + "t_fv=" + str(getCurrentEpochTime()) + \
    ";" + "Domain=.lazada.sg; Path=/;" + x5sec

url = "https://www.lazada.sg/catalog/?
↳_keyori=ss&from=input&page=1&q=lego+super+heroes+76160&sort=priceasc"

payload = {}
headers3 = {
    'authority': 'www.lazada.sg',
    'cache-control': 'max-age=0',
    'sec-ch-ua': '"Chromium";v="92", " Not A;Brand";v="99", "Google Chrome";
↳v="92"',
    'sec-ch-ua-mobile': '?0',
    'upgrade-insecure-requests': '1',
    'user-agent': 'Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36 (KHTML,␣
↳like Gecko) Chrome/92.0.4515.51 Safari/537.36',
    'accept': 'text/html,application/xhtml+xml,application/xml;q=0.9,image/
↳avif,image/webp,image/apng,*/*;q=0.8,application/signed-exchange;v=b3;q=0.9',
    'sec-fetch-site': 'same-origin',
    'sec-fetch-mode': 'navigate',
    'sec-fetch-user': '?1',
    'sec-fetch-dest': 'document',
    'referrer': 'https://www.lazada.sg/catalog/?
↳_keyori=ss&from=input&page=1&q=lego+super+heroes+76160&sort=priceasc',
    'accept-language': 'en-US,en;q=0.9'
}

```

```

}
headers3.update({"cookie": newCookie2})

```

```

[12]: def getProductsDetails(keywords):
    for keyword in keywords:
        keyword_index = keywords.index(keyword)+1
        # print('Index', index)
        keyword_data = []

        try:
            new_string = urllib.parse.quote_plus(keyword)
            keyword_url = 'https://www.lazada.sg/catalog/?q='+str(new_string)

            headers3.update({'referer':keyword_url})
            final_response = requests.request("GET", keyword_url,
            ↪headers=headers3, data=payload)
            print("working on Keyword: ",keyword)
            sleep(10)
            # print(final_response.text)
            soup = BeautifulSoup(final_response.content, "html.parser")
            scripts = soup.select("script")
            pageResults = None
            for script_tag in scripts:
                if "window.pageData" in str(script_tag):
                    pageData = str(script_tag).replace("<script>", "").replace(
                        "</script>", "").replace("window.pageData =", "").
                    ↪rsplit(";", 1)[0].strip()
                    pageData_json = json.loads(pageData)
                    # print(pageData_json)
                    pageResults= pageData_json['mods']["listItems"]
                else:
                    pass
            listItems = pageResults
            for item in listItems:
                try:
                    item["originalPriceShow"]
                    originalPrice = item['originalPriceShow']
                except:
                    originalPrice = 0
                try:
                    item["discount"]
                    discount = item["discount"]
                except:
                    discount = 0

```

```

        # keyword_data.append({"sku":keyword_index,"prod_Name":␣
↪item["name"],"Keyword":keyword, "prod_Id": item["nid"], "prod_url":␣
↪item["productUrl"],
        #
        "image": item["image"], "priceShow":␣
↪item["priceShow"], "originalPrice": originalPrice, "discount": discount})
        keyword_data.append({"sku":keyword_index,"prod_Name":␣
↪item["name"],"Keyword":keyword, "noOfResults": len(listItems),"lowestPrice":␣
↪item["priceShow"]})
        prices=[x["lowestPrice"].split("$")[1].strip() for x in␣
↪keyword_data]
        # print(min(prices))
        price_index =prices.index(min(prices))
        finaldata.append(keyword_data[price_index])

    except:
        raise RuntimeError (f"There has been an error..")
    return

```

```
[ ]: getProductsDetails(keywords=keywords)
```

```
[ ]: final_df = pd.DataFrame(finaldata)
final_df.to_excel("lazada-sg-products.xlsx", index=False)
print("We are Done ... ")

```