In [1]:
```
!pip install selenium
```

```
Requirement already satisfied: selenium in /home/user/anaconda3/li
b/python3.7/site-packages (4.11.2)
Requirement already satisfied: trio~=0.17 in /home/user/anaconda3/l
ib/python3.7/site-packages (from selenium) (0.22.2)
Requirement already satisfied: certifi>=2021.10.8 in /home/user/ana
conda3/lib/python3.7/site-packages (from selenium) (2023.11.17)
Requirement already satisfied: urllib3[socks]<3,>=1.26 in /home/use
r/anaconda3/lib/python3.7/site-packages (from selenium) (2.0.7)
Requirement already satisfied: trio-websocket~=0.9 in /home/user/an
aconda3/lib/python3.7/site-packages (from selenium) (0.11.1)
Requirement already satisfied: outcome in /home/user/anaconda3/lib/
python3.7/site-packages (from trio~=0.17->selenium) (1.3.0.post0)
Requirement already satisfied: sortedcontainers in /home/user/anaco
nda3/lib/python3.7/site-packages (from trio~=0.17->selenium) (2.1.
0)
Requirement already satisfied: sniffio in /home/user/anaconda3/lib/
python3.7/site-packages (from trio~=0.17->selenium) (1.3.0)
Requirement already satisfied: attrs>=20.1.0 in /home/user/anaconda
3/lib/python3.7/site-packages (from trio~=0.17->selenium) (23.1.0)
Requirement already satisfied: idna in /home/user/anaconda3/lib/pyt
hon3.7/site-packages (from trio~=0.17->selenium) (2.8)
Requirement already satisfied: exceptiongroup>=1.0.0rc9; python_ver
sion < "3.11" in /home/user/anaconda3/lib/python3.7/site-packages
(from trio~=0.17->selenium) (1.1.3)
Requirement already satisfied: pysocks!=1.5.7,<2.0,>=1.5.6; extra =
= "socks" in /home/user/anaconda3/lib/python3.7/site-packages (from
urllib3[socks]<3,>=1.26->selenium) (1.7.1)
Requirement already satisfied: wsproto>=0.14 in /home/user/anaconda
3/lib/python3.7/site-packages (from trio-websocket~=0.9->selenium)
(1.2.0)
Requirement already satisfied: importlib-metadata; python_version <
"3.8" in /home/user/anaconda3/lib/python3.7/site-packages (from att
rs>=20.1.0->trio~=0.17->selenium) (0.23)
Requirement already satisfied: h11<1,>=0.9.0 in /home/user/anaconda
3/lib/python3.7/site-packages (from wsproto>=0.14->trio-websocket~=
0.9->selenium) (0.14.0)
Requirement already satisfied: zipp>=0.5 in /home/user/anaconda3/li
b/python3.7/site-packages (from importlib-metadata; python_version
< "3.8"->attrs>=20.1.0->trio~=0.17->selenium) (0.6.0)
Requirement already satisfied: typing-extensions; python_version <
"3.8" in /home/user/anaconda3/lib/python3.7/site-packages (from h11
<1,>=0.9.0->wsproto>=0.14->trio-websocket~=0.9->selenium) (4.7.1)
Requirement already satisfied: more-itertools in /home/user/anacond
a3/lib/python3.7/site-packages (from zipp>=0.5->importlib-metadata;
python_version < "3.8"->attrs>=20.1.0->trio~=0.17->selenium) (7.2.
0)
```

In [2]:
```
!pip install pandas numpy tqdm selenium webdriver_manager beautiful
soup4
```

```
Requirement already satisfied: pandas in /home/user/anaconda3/lib/p
ython3.7/site-packages (0.25.1)
Requirement already satisfied: numpy in /home/user/anaconda3/lib/py
thon3.7/site-packages (1.17.2)
Requirement already satisfied: tqdm in /home/user/anaconda3/lib/pyt
hon3.7/site-packages (4.36.1)
Requirement already satisfied: selenium in /home/user/anaconda3/li
b/python3.7/site-packages (4.11.2)
Requirement already satisfied: webdriver_manager in /home/user/anac
onda3/lib/python3.7/site-packages (4.0.1)
Requirement already satisfied: beautifulsoup4 in /home/user/anacond
a3/lib/python3.7/site-packages (4.8.0)
Requirement already satisfied: python-dateutil>=2.6.1 in /home/use
r/anaconda3/lib/python3.7/site-packages (from pandas) (2.8.0)
Requirement already satisfied: pytz>=2017.2 in /home/user/anaconda
3/lib/python3.7/site-packages (from pandas) (2019.3)
Requirement already satisfied: trio-websocket~=0.9 in /home/user/an
aconda3/lib/python3.7/site-packages (from selenium) (0.11.1)
Requirement already satisfied: urllib3[socks]<3,>=1.26 in /home/use
r/anaconda3/lib/python3.7/site-packages (from selenium) (2.0.7)
Requirement already satisfied: certifi>=2021.10.8 in /home/user/ana
conda3/lib/python3.7/site-packages (from selenium) (2023.11.17)
Requirement already satisfied: trio~=0.17 in /home/user/anaconda3/l
ib/python3.7/site-packages (from selenium) (0.22.2)
Requirement already satisfied: packaging in /home/user/anaconda3/li
b/python3.7/site-packages (from webdriver_manager) (19.2)
Requirement already satisfied: python-dotenv in /home/user/anaconda
3/lib/python3.7/site-packages (from webdriver_manager) (0.21.1)
Requirement already satisfied: requests in /home/user/anaconda3/li
b/python3.7/site-packages (from webdriver_manager) (2.22.0)
Requirement already satisfied: soupsieve>=1.2 in /home/user/anacond
a3/lib/python3.7/site-packages (from beautifulsoup4) (1.9.3)
Requirement already satisfied: six>=1.5 in /home/user/anaconda3/li
b/python3.7/site-packages (from python-dateutil>=2.6.1->pandas) (1.
12.0)
Requirement already satisfied: wsproto>=0.14 in /home/user/anaconda
3/lib/python3.7/site-packages (from trio-websocket~=0.9->selenium)
(1.2.0)
Requirement already satisfied: exceptiongroup; python_version < "3.
11" in /home/user/anaconda3/lib/python3.7/site-packages (from trio-
websocket~=0.9->selenium) (1.1.3)
Requirement already satisfied: pysocks!=1.5.7,<2.0,>=1.5.6; extra =
= "socks" in /home/user/anaconda3/lib/python3.7/site-packages (from
urllib3[socks]<3,>=1.26->selenium) (1.7.1)
Requirement already satisfied: sortedcontainers in /home/user/anaco
nda3/lib/python3.7/site-packages (from trio~=0.17->selenium) (2.1.
0)
Requirement already satisfied: idna in /home/user/anaconda3/lib/pyt
hon3.7/site-packages (from trio~=0.17->selenium) (2.8)
Requirement already satisfied: sniffio in /home/user/anaconda3/lib/
python3.7/site-packages (from trio~=0.17->selenium) (1.3.0)
Requirement already satisfied: outcome in /home/user/anaconda3/lib/
python3.7/site-packages (from trio~=0.17->selenium) (1.3.0.post0)
Requirement already satisfied: attrs>=20.1.0 in /home/user/anaconda
3/lib/python3.7/site-packages (from trio~=0.17->selenium) (23.1.0)
Requirement already satisfied: pyparsing>=2.0.2 in /home/user/anaco
nda3/lib/python3.7/site-packages (from packaging->webdriver_manage
r) (2.4.2)
Requirement already satisfied: chardet<3.1.0,>=3.0.2 in /home/user/
anaconda3/lib/python3.7/site-packages (from requests->webdriver_man
ager) (3.0.4)
```

```
Requirement already satisfied: h11<1,>=0.9.0 in /home/user/anaconda
3/lib/python3.7/site-packages (from wsproto>=0.14->trio-websocket~=
0.9->selenium) (0.14.0)
Requirement already satisfied: importlib-metadata; python_version <
"3.8" in /home/user/anaconda3/lib/python3.7/site-packages (from att
rs>=20.1.0->trio~=0.17->selenium) (0.23)
Requirement already satisfied: typing-extensions; python_version <
"3.8" in /home/user/anaconda3/lib/python3.7/site-packages (from h11
<1,>=0.9.0->wsproto>=0.14->trio-websocket~=0.9->selenium) (4.7.1)
Requirement already satisfied: zipp>=0.5 in /home/user/anaconda3/li
b/python3.7/site-packages (from importlib-metadata; python_version
< "3.8"->attrs>=20.1.0->trio~=0.17->selenium) (0.6.0)
Requirement already satisfied: more-itertools in /home/user/anacond
a3/lib/python3.7/site-packages (from zipp>=0.5->importlib-metadata;
python_version < "3.8"->attrs>=20.1.0->trio~=0.17->selenium) (7.2.
0)
```

In [9]:
```
import csv
import pandas as pd
import numpy as np
import time
import requests

from tqdm import tqdm
from selenium import webdriver
from webdriver_manager.chrome import ChromeDriverManager
from selenium.webdriver.common.by import By
from bs4 import BeautifulSoup
from selenium.webdriver.common.keys import Keys
from itertools import repeat  # 반복 가능한 데이터를 다루는데 도움을 주는 도구 제공
# repeat 함수 : 지정된 값을 무한히 반복하는 이터레이터 생성
from urllib.request import urlopen  # URL을 열고 읽는데 사용되는 함수 제공
# urlopen 함수 : 주어진 URL을 열고 해당 URL에서 읽은 데이터 반환
import re  # 정규 표현식을 사용하여 문자열을 처리하는 함수 제공

# 드라이버 생성 설치 경로 정확하게 기재 (드라이버 로드)
chromedriver = '/home/user/다운로드'

# 드라이버 사용하겠다.
driver = webdriver.Chrome()
class Item:
    def __init__(self, brand, name, price, link, like):
        self.brand = brand
        self.name = name
        self.price = price
        self.link = link
        self.like = like

    def get_info(self):
        print('brand: ', self.brand)
        print('name: ', self.name)
        print('price: ', self.price)
        print('link: ', self.link)
        print('like: ', self.like)
        print('')

def get_upper_category():
    main_url = urlopen('https://www.musinsa.com/app/')
    _main_page = BeautifulSoup(main_url, 'html.parser')

    upper_category = [category
                      for category in _main_page.findAll('strong',
{'class':'title'})]

    print('<전체 카테고리 목록>')
    for category in upper_category:
        print(category.get_text())
    return upper_category

def get_lower_category(upper_category):
    main_category = input('어느 카테고리를 검색할까요? ')
    for category_searcher in upper_category:
        if main_category == category_searcher.get_text():
            lower_category_info = [category
                                   for category in category_searcher.par
ent.parent.parent.findAll('li')]
```

```
    lower_category=[]
    for s in lower_category_info:
        tmp_category = []
        tmp = s.get_text().split()
        category_name = ''
        for n in range(0, len(tmp) - 1):
            category_name+=tmp[n]
        tmp_category = [s.a.attrs['href'], category_name]
        lower_category.append(tmp_category)

    print('<하위 카테고리 목록>')
    for c in lower_category:
        print(c[1])
    return lower_category

def get_item_page_url(lower_category):
    sub_category = input('어느 카테고리를 검색할까요? ')
    for s in lower_category:
        if s[1] == sub_category:
            item_url = 'https://www.musinsa.com/app/' + s[0]
    return item_url

def get_item_page(item_url, i):
    item_url += 'https://www.musinsa.com/ranking/best?period=now&ag
e=ALL&mainCategory=&subCategory=&leafCategory=&price=&golf=false&ki
ds=false&newProduct=false&exclusive=false&discount=false&soldOut=fa
lse&page='
    item_url += str(i)
    item_url += '&viewType=small&priceMin=&priceMax='
    return item_url

def get_item_info(Item_list, item_url):
    item_page = BeautifulSoup(urlopen(item_url), 'html.parser')

    item_list = item_page.findAll('li',{'class':'li_box'})
    for item in item_list:
        brand = ""
        name = ""
        price = ""
        link = ""
        like = "0"

        _brand = item.findAll('p', {'class':'item_title'})
        if len(_brand) == 1:
            brand = _brand[0].get_text()
        elif len(_brand) == 2:
            brand = _brand[1].get_text()

        _name = item.find('p', {'class':'list_info'}).get_text().sp
lit()
        if item.find('strong', {'class':'txt_reserve'}) != None:
            for n in range(2, len(_name)):
                name += _name[n] + ' '
        else:
            for n in _name:
                name += n + ' '

        _price = item.find('p',{'class':'price'}).get_text().split
()
        if len(_price) == 1:
```

```
            price = _price[0]
        elif len(_price) == 2:
            price = _price[1]

        item_link = item.find('p', {'class':'list_info'}).a.attrs
['href']
        link = 'https://www.musinsa.com/app/' + item_link

        if item.find('p', {'class':'txt_cnt_like'}) != None:
            like = ""
            tmp_like = item.find('p', {'class':'txt_cnt_like'}).get
_text().split()
            _like = tmp_like[0].split(',')
            for l in _like:
                like += l


        item = Item(brand, name, price, link, like)
        Item_list.append(item)

Item_list = []
upper_category = get_upper_category()
lower_category = get_lower_category(upper_category)
item_page_url = get_item_page_url(lower_category)
for i in range(1, 6):
    current_page = get_item_page(item_page_url, i)
    get_item_info(Item_list, current_page)

for item in Item_list:
    item.get_info()

session_obj = requests.Session()
response = session_obj.get(url, headers={"User-Agent": "Mozilla/5.
0"})
```

```
---------------------------------------------------------------------
--------
HTTPError                                 Traceback (most recent ca
ll last)
<ipython-input-9-4d2d0272d112> in <module>
    131
    132 Item_list = []
--> 133 upper_category = get_upper_category()
    134 lower_category = get_lower_category(upper_category)
    135 item_page_url = get_item_page_url(lower_category)

<ipython-input-9-4d2d0272d112> in get_upper_category()
     39
     40 def get_upper_category():
---> 41     main_url = urlopen('https://www.musinsa.com/app/')
     42     _main_page = BeautifulSoup(main_url, 'html.parser')
     43

~/anaconda3/lib/python3.7/urllib/request.py in urlopen(url, data, t
imeout, cafile, capath, cadefault, context)
    220     else:
    221         opener = _opener
--> 222     return opener.open(url, data, timeout)
    223
    224 def install_opener(opener):

~/anaconda3/lib/python3.7/urllib/request.py in open(self, fullurl,
data, timeout)
    529         for processor in self.process_response.get(protoco
l, []):
    530             meth = getattr(processor, meth_name)
--> 531             response = meth(req, response)
    532
    533         return response

~/anaconda3/lib/python3.7/urllib/request.py in http_response(self,
request, response)
    639         if not (200 <= code < 300):
    640             response = self.parent.error(
--> 641                 'http', request, response, code, msg, hdrs)
    642
    643         return response

~/anaconda3/lib/python3.7/urllib/request.py in error(self, proto, *
args)
    567         if http_err:
    568             args = (dict, 'default', 'http_error_default')
+ orig_args
--> 569             return self._call_chain(*args)
    570
    571 # XXX probably also want an abstract factory that knows whe
n it makes

~/anaconda3/lib/python3.7/urllib/request.py in _call_chain(self, ch
ain, kind, meth_name, *args)
    501         for handler in handlers:
    502             func = getattr(handler, meth_name)
--> 503             result = func(*args)
    504             if result is not None:
    505                 return result
```

```
~/anaconda3/lib/python3.7/urllib/request.py in http_error_default(s
elf, req, fp, code, msg, hdrs)
    647 class HTTPDefaultErrorHandler(BaseHandler):
    648     def http_error_default(self, req, fp, code, msg, hdrs):
--> 649         raise HTTPError(req.full_url, code, msg, hdrs, fp)
    650
    651 class HTTPRedirectHandler(BaseHandler):

HTTPError: HTTP Error 403: Forbidden
```

In [ ]: