

```
In [1]: # python selenium 설치하기
!pip install selenium
```

```
Requirement already satisfied: selenium in ./opt/anaconda3/lib/python3.9/site-packages (4.15.2)
Requirement already satisfied: trio~=0.17 in ./opt/anaconda3/lib/python3.9/site-packages (from selenium) (0.23.1)
Requirement already satisfied: trio-websocket~=0.9 in ./opt/anaconda3/lib/python3.9/site-packages (from selenium) (0.11.1)
Requirement already satisfied: certifi>=2021.10.8 in ./opt/anaconda3/lib/python3.9/site-packages (from selenium) (2021.10.8)
Requirement already satisfied: urllib3[socks]<3,>=1.26 in ./opt/anaconda3/lib/python3.9/site-packages (from selenium) (1.26.7)
Requirement already satisfied: idna in ./opt/anaconda3/lib/python3.9/site-packages (from trio~=0.17->selenium) (3.2)
Requirement already satisfied: sniffio>=1.3.0 in ./opt/anaconda3/lib/python3.9/site-packages (from trio~=0.17->selenium) (1.3.0)
Requirement already satisfied: attrs>=20.1.0 in ./opt/anaconda3/lib/python3.9/site-packages (from trio~=0.17->selenium) (21.2.0)
Requirement already satisfied: outcome in ./opt/anaconda3/lib/python3.9/site-packages (from trio~=0.17->selenium) (1.3.0.post0)
Requirement already satisfied: exceptiongroup>=1.0.0rc9 in ./opt/anaconda3/lib/python3.9/site-packages (from trio~=0.17->selenium) (1.1.3)
Requirement already satisfied: sortedcontainers in ./opt/anaconda3/lib/python3.9/site-packages (from trio~=0.17->selenium) (2.4.0)
Requirement already satisfied: wsproto>=0.14 in ./opt/anaconda3/lib/python3.9/site-packages (from trio-websocket~=0.9->selenium) (1.2.0)
Requirement already satisfied: PySocks!=1.5.7,<2.0,>=1.5.6 in ./opt/anaconda3/lib/python3.9/site-packages (from urllib3[socks]<3,>=1.26->selenium) (1.7.1)
Requirement already satisfied: h11<1,>=0.9.0 in ./opt/anaconda3/lib/python3.9/site-packages (from wsproto>=0.14->trio-websocket~=0.9->selenium) (0.14.0)
```

```
In [3]: !pip install pandas numpy tqdm selenium webdriver_manager beautifulsoup4
```

```
Requirement already satisfied: pandas in ./opt/anaconda3/lib/python3.9/site-packages (1.3.4)
Requirement already satisfied: numpy in ./opt/anaconda3/lib/python3.9/site-packages (1.20.3)
Requirement already satisfied: tqdm in ./opt/anaconda3/lib/python3.9/site-packages (4.62.3)
Requirement already satisfied: selenium in ./opt/anaconda3/lib/python3.9/site-packages (4.15.2)
Collecting webdriver_manager
  Downloading webdriver_manager-4.0.1-py2.py3-none-any.whl (27 kB)
Requirement already satisfied: beautifulsoup4 in ./opt/anaconda3/lib/python3.9/site-packages (4.10.0)
Requirement already satisfied: python-dateutil>=2.7.3 in ./opt/anaconda3/lib/python3.9/site-packages (from pandas) (2.8.2)
Requirement already satisfied: pytz>=2017.3 in ./opt/anaconda3/lib/python3.9/site-packages (from pandas) (2021.3)
Requirement already satisfied: trio-websocket~=0.9 in ./opt/anaconda3/lib/python3.9/site-packages (from selenium) (0.11.1)
Requirement already satisfied: trio~=0.17 in ./opt/anaconda3/lib/python3.9/site-packages (from selenium) (0.23.1)
Requirement already satisfied: urllib3[socks]<3,>=1.26 in ./opt/anaconda3/lib/python3.9/site-packages (from selenium) (1.26.7)
Requirement already satisfied: certifi>=2021.10.8 in ./opt/anaconda3/lib/python3.9/site-packages (from selenium) (2021.10.8)
Requirement already satisfied: packaging in ./opt/anaconda3/lib/python3.9/site-packages (from webdriver_manager) (21.0)
```

Collecting python-dotenv

Downloading python_dotenv-1.0.0-py3-none-any.whl (19 kB)

Requirement already satisfied: requests in ./opt/anaconda3/lib/python3.9/site-packages (from webdriver_manager) (2.26.0)

Requirement already satisfied: soupsieve>1.2 in ./opt/anaconda3/lib/python3.9/site-packages (from beautifulsoup4) (2.2.1)

Requirement already satisfied: six>=1.5 in ./opt/anaconda3/lib/python3.9/site-packages (from python-dateutil>=2.7.3->pandas) (1.16.0)

Requirement already satisfied: attrs>=20.1.0 in ./opt/anaconda3/lib/python3.9/site-packages (from trio~0.17->selenium) (21.2.0)

Requirement already satisfied: idna in ./opt/anaconda3/lib/python3.9/site-packages (from trio~0.17->selenium) (3.2)

Requirement already satisfied: sniffio>=1.3.0 in ./opt/anaconda3/lib/python3.9/site-packages (from trio~0.17->selenium) (1.3.0)

Requirement already satisfied: exceptiongroup>=1.0.0rc9 in ./opt/anaconda3/lib/python3.9/site-packages (from trio~0.17->selenium) (1.1.3)

Requirement already satisfied: outcome in ./opt/anaconda3/lib/python3.9/site-packages (from trio~0.17->selenium) (1.3.0.post0)

Requirement already satisfied: sortedcontainers in ./opt/anaconda3/lib/python3.9/site-packages (from trio~0.17->selenium) (2.4.0)

Requirement already satisfied: wsproto>=0.14 in ./opt/anaconda3/lib/python3.9/site-packages (from trio-websocket~0.9->selenium) (1.2.0)

Requirement already satisfied: PySocks!=1.5.7,<2.0,>=1.5.6 in ./opt/anaconda3/lib/python3.9/site-packages (from urllib3[socks]<3,>=1.26->selenium) (1.7.1)

Requirement already satisfied: h11<1,>=0.9.0 in ./opt/anaconda3/lib/python3.9/site-packages (from wsproto>=0.14->trio-websocket~0.9->selenium) (0.14.0)

Requirement already satisfied: pyparsing>=2.0.2 in ./opt/anaconda3/lib/python3.9/site-packages (from packaging->webdriver_manager) (3.0.4)

Requirement already satisfied: charset-normalizer~2.0.0 in ./opt/anaconda3/lib/python3.9/site-packages (from requests->webdriver_manager) (2.0.4)

Installing collected packages: python-dotenv, webdriver_manager

Successfully installed python-dotenv-1.0.0 webdriver_manager-4.0.1

In [4]:

```
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()

# 크롤링할 사이트 무신사 호출
# url = "https://www.musinsa.com/app/"
# driver.get(url)
# time.sleep(1)

# 무신사 스토어 > 상의(무신사 추천순) 해당 url오픈
```

```

html = urlopen("https://www.musinsa.com/categories/item/001")
bsObject = BeautifulSoup(html, "html.parser")

print(bsObject)

# 상품들의 정보가 담긴 li_box 모두 가져오기
item_list = bsObject.findAll('li', {'class': 'li_box'})
print(item_list)

# 빈 리스트
brand = []
name = []
price = []
number = []
number_box = 0 # 순위 카운트

for item in item_list:

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

    # 상품명 담기
    name_box = item.find('a', {'class': 'img-block'})
    if name_box != None:
        name_text = name_box.get('title')
        if name_text != None:
            name.append(name_text.strip())
        else:
            name.append('N/A')
    else:
        name.append('N/A')

    # 가격 담기
    price_box = item.find('p', {'class': 'price'})
    if price_box != None:
        price_text = price_box.get_text().split()
        if len(price_text) == 1:
            price.append(price_text[0])
        elif len(price_text) == 2:
            price.append(price_text[1])
    else:
        price.append('N/A')

    # 순위 담기
    number_box += 1
    number.append(str(number_box))

# 확인
print(brand)
print(name)
print(price)

```

```

<!DOCTYPE html>

<html lang="ko">
<head>
<script type="text/javascript">
    var dataLayer = dataLayer || [];
    </script>
<!-- Google Tag Manager -->

```

```

<script>(function(w,d,s,l,i){w[l]=w[l]||[];w[l].push({'gtm.start':
    new Date().getTime(),event:'gtm.js'});var f=d.getElementsByTagNameNam
e(s)[0],
    j=d.createElement(s),dl=l!='dataLayer'?'&l='+l:'';j.async=true;j.src=
    'https://www.googletagmanager.com/gtm.js?id='+i+dl;f.parentNode.inser
tBefore(j,f);
    })(window,document,'script','dataLayer','GTM-TXDSFSF');</script>
<!-- End Google Tag Manager -->
<script>
    dataLayer.push({
        mainInfo: {
            'email': ""
        }
    });
</script>
<title>상의 | 무신사 추천 브랜드</title>
<meta charset="utf-8"/>
<meta content="IE=edge" http-equiv="X-UA-Compatible">
<meta content="width=device-width, initial-scale=1" name="viewport"/>
<meta content="무신사에서 상의 관련 상품 리스트와 쇼핑 기획에 대한 자세한 정보를 확인하세요." na
me="description"/>
<!-- 페이스북 오픈 그래프 메타태그 -->
<meta content="648837775133366" id="fbAddId" property="fb:app_id">
<meta content="website" id="fbOgType" property="og:type">
<meta content="https://www.musinsa.comnull?_gf=A" id="fbOgUrl" property="og:u
rl"/>
<meta content="상의 | 무신사 추천 브랜드" id="fbOgTitle" property="og:title"/>
<meta content="https://image.msscdn.net/mfile_s01/fb/share_musinsa.png" id="f
bOgImage" property="og:image"/>
<meta content="무신사에서 상의 관련 상품 리스트와 쇼핑 기획에 대한 자세한 정보를 확인하세요." id
="fbOgDescription" property="og:description"/>
<!--// 페이스북 오픈 그래프 메타태그 -->
<link href="//image.msscdn.net/favicon.ico?20200213" rel="shortcut icon" type
="image/x-icon"/>
<link href="//static.msscdn.net/platform/css/common.css" rel="stylesheet"/>
<link href="//image.msscdn.net/ui/musinsa/resources/common/css/icon.min.css?1
697673600000" rel="stylesheet"/>
<link href="//static.msscdn.net/ui/build/pc/css/common.css" rel="stylesheet"/
>
<link href="//static.msscdn.net/ui/build/pc/css/review.css" rel="stylesheet"/
>
<link href="//static.msscdn.net/skin/musinsa/css/magazine_common.css?16976736
000000" rel="stylesheet" type="text/css">
<link href="//static.msscdn.net/skin/musinsa/css/store_common.css?16976736000
00" rel="stylesheet" type="text/css">
<link href="//static.msscdn.net/skin/musinsa/css/layout.min.css?169767360000
0" rel="stylesheet" type="text/css">
<link href="//static.msscdn.net/skin/musinsa/css/guide.min.css?1697673600000"
rel="stylesheet" type="text/css">
<link href="//static.msscdn.net/skin/musinsa/css/style.min.css?1697673600000"
rel="stylesheet" type="text/css"/>
<link href="//static.msscdn.net/skin/musinsa/css/sub.css?1697673600000" rel
="stylesheet" type="text/css"/>
<link href="//static.msscdn.net/skin/musinsa/css/media_query.css?169767360000
0" rel="stylesheet" type="text/css"/>
<link href="//static.msscdn.net/skin/musinsa/css/new.css?1697673600000" rel
="stylesheet" type="text/css"/>
<link href="//static.msscdn.net/static/display/front/css/pc/brandshop.css?169
7673600000" rel="stylesheet" type="text/css"/>
<script src="//static.msscdn.net/mfile_outsrc/js/vendor/jquery-1.11.1.min.js?
1697673600000" type="text/javascript"></script>
<script src="//static.msscdn.net/mfile_outsrc/js/vendor/jquery.easing.js?1697
673600000" type="text/javascript"></script>
<script src="//static.msscdn.net/skin/musinsa/js/jquery.cycle.all.js?16976736

```

```

<a href="//www.musinsa.com/app/goods/726408" name="goods_link" title="[Mmlg]
MMLG HF-T (WHITE)">
<strong class="txt_reserve"></strong>
    [Mmlg] MMLG HF-T (WHITE)
</a>

</p>
<p class="price">
<del>42,000원</del>
    25,200원
</p>
<p class="membership" onclick="viewMemberPrice('sPrice_726408', '726408',
'0');">MEMBERSHIP PRICE<span class="fr">▼</span></p>
<div class="member_price" id="sPrice_726408" style="display:none; z-index: 99
99;">
<p>무신사 회원가<span class="fr"><a href="#" onclick="viewMemberPrice('sPrice_726
408', '726408', '0');return false;">X</a></span></p>
<ul>
<li class="pertinent" style="border-bottom:1px solid #ddd;">
<span>&gt; </span>
        비회원
        <span class="txt_price_member">25,200원</span></li>

</ul>
</div>
<p class="point">
<span class="img-score"><span class="bar" style="width:94%"></span></span>
<span class="count">5,519</span>
</p>
<p class="txt_cnt_like" data-goodsno="726408"></p>
</div>
</div>
<div class="option">
<p class="option_btn">
<a data-goods-no="726408" href="#pol726408" onclick="viewOptionBox(726408,
0); return false;">OPTION ▼</a>
</p>
<div class="list_pol_goods" id="pol726408" style="display:none;"></div>
</div>
</li>, <li class="li_box" data-no="567733">
<!-- 상단 좌측 라벨 -->
<div class="icon_new">SALE 30%</div>
<!-- 상단 우측 라벨 -->
<div class="icon_group">
<ul>
<li class="icon_man sight_out" title="남성">남성</li>
</ul>
</div>
<div class="li_inner">
<div class="list_img">
<a class="img-block" href="//www.musinsa.com/app/goods/567733" name="goods_li
nk" title="[UNISEX] 45수 특양면 옆트임 오버핏 반팔티 (화이트)">

</img></a>
</div>
<div class="article_info">
<p class="item_title"><a href="//www.musinsa.com/brands/teeaz">티아즈</a></p>
<p class="list_info">
<a href="//www.musinsa.com/app/goods/567733" name="goods_link" title="[UNISE
X] 45수 특양면 옆트임 오버핏 반팔티 (화이트)">
<strong class="txt_reserve"></strong>
    [UNISEX] 45수 특양면 옆트임 오버핏 반팔티 (화이트)
</a>

</p>

```

[illegible]

[illegible]

In [5]:

```
# 데이터 정제
import pandas as pd
data = {'순위': number, '브랜드': brand, '상품명': name, '가격' + '(' + '할인가' + ')'
df=pd.DataFrame(data)
df
```

Out[5]:

순위		브랜드	상품명	가격(할인가)
0	1	굿라이프웍스	N/A	38,000원

순위		브랜드	상품명	가격(할인가)
1	2	니티드	N/A	39,600원
2	3	트래블	N/A	47,500원
3	4	게인스보로	N/A	41,500원
4	5	필루미네이트	N/A	46,000원
...
127	128	낮띵이즈베럴댄미	FLOWER BED TEE_white	18,500원
128	129	랭글러	코듀로이 웨스턴 셔츠 카키	41,650원
129	130	마리떼	W EMBROIDERY TEE navy	36,000원
130	131	엠엠엘지	[Mmlg] MMLG HF-T (WHITE)	25,200원
131	132	티아즈	[UNISEX] 45수 특양면 옆트임 오버핏 반팔티 (화이트)	14,700원

132 rows × 4 columns

```
In [6]: # index를 순위로 바꾸기
df2 = df.set_index('순위')
df2
```

Out [6]:

순위		브랜드	상품명	가격(할인가)
1	2	굿라이프웍스	N/A	38,000원
2	3	니티드	N/A	39,600원
3	4	트래블	N/A	47,500원
4	5	게인스보로	N/A	41,500원
5	...	필루미네이트	N/A	46,000원
...
128	129	낮띵이즈베럴댄미	FLOWER BED TEE_white	18,500원
129	130	랭글러	코듀로이 웨스턴 셔츠 카키	41,650원
130	131	마리떼	W EMBROIDERY TEE navy	36,000원
131	132	엠엠엘지	[Mmlg] MMLG HF-T (WHITE)	25,200원
132	...	티아즈	[UNISEX] 45수 특양면 옆트임 오버핏 반팔티 (화이트)	14,700원

132 rows × 3 columns

```
In [ ]:
```