

네이버 쇼핑 크롤링

1. 키워드로 검색하기
2. 스크롤 끝까지 내리기
3. 페이지 번호 가져오기
4. 데이터 수집하기(광고제외)
5. 파일로 저장

웹페이지 접근하기

```
from selenium import webdriver
from selenium.webdriver.common.by import By
from selenium.webdriver.common.keys import Keys
import pandas as pd
import time
import openpyxl

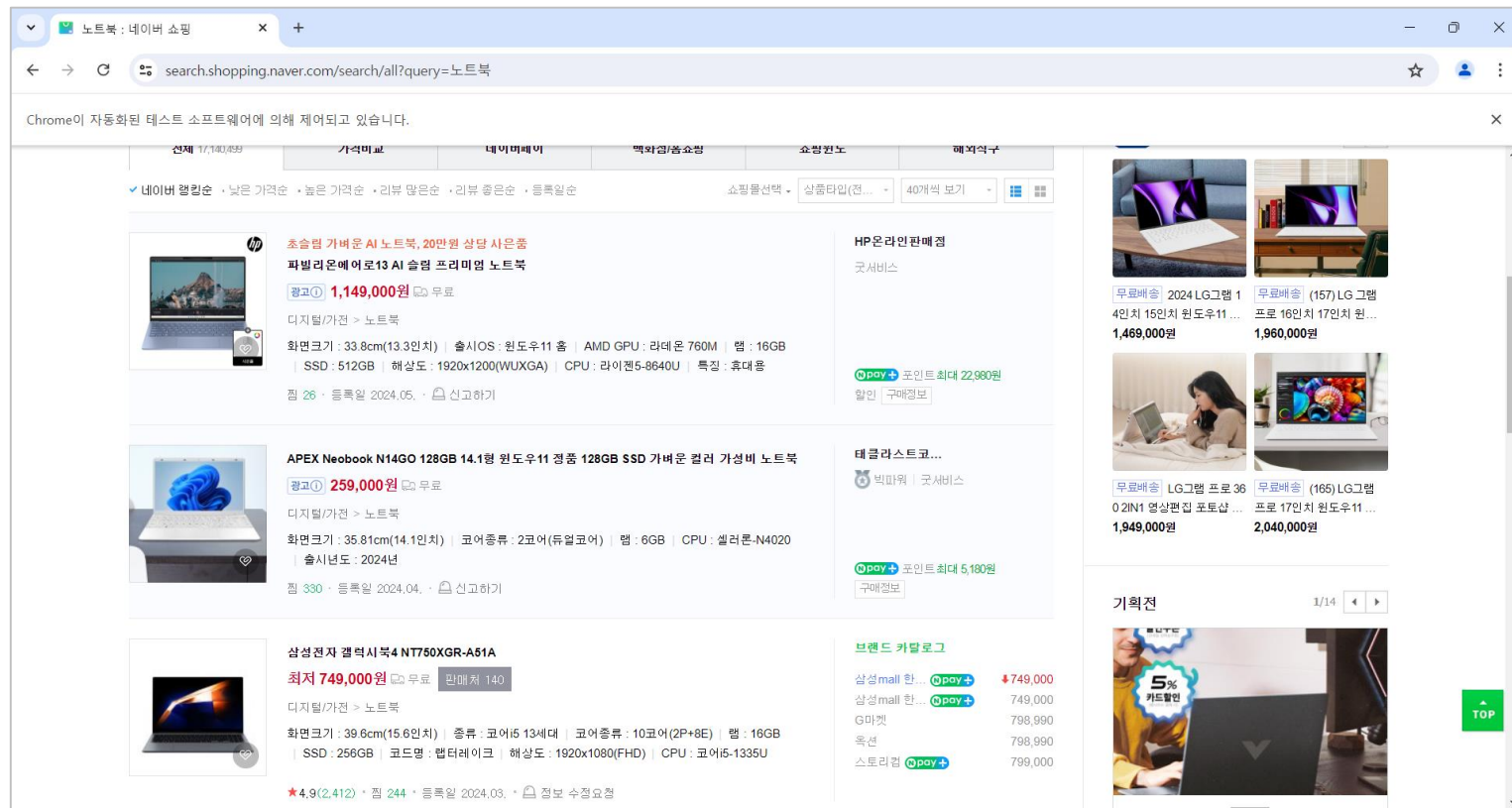
input_search = input('검색할 품목을 입력하세요 --> ')

url = "https://search.shopping.naver.com/search/all?query="+input_search
driver = webdriver.Chrome()
driver.get(url)
driver.maximize_window()
```

스크롤

절대 위치로 스크롤

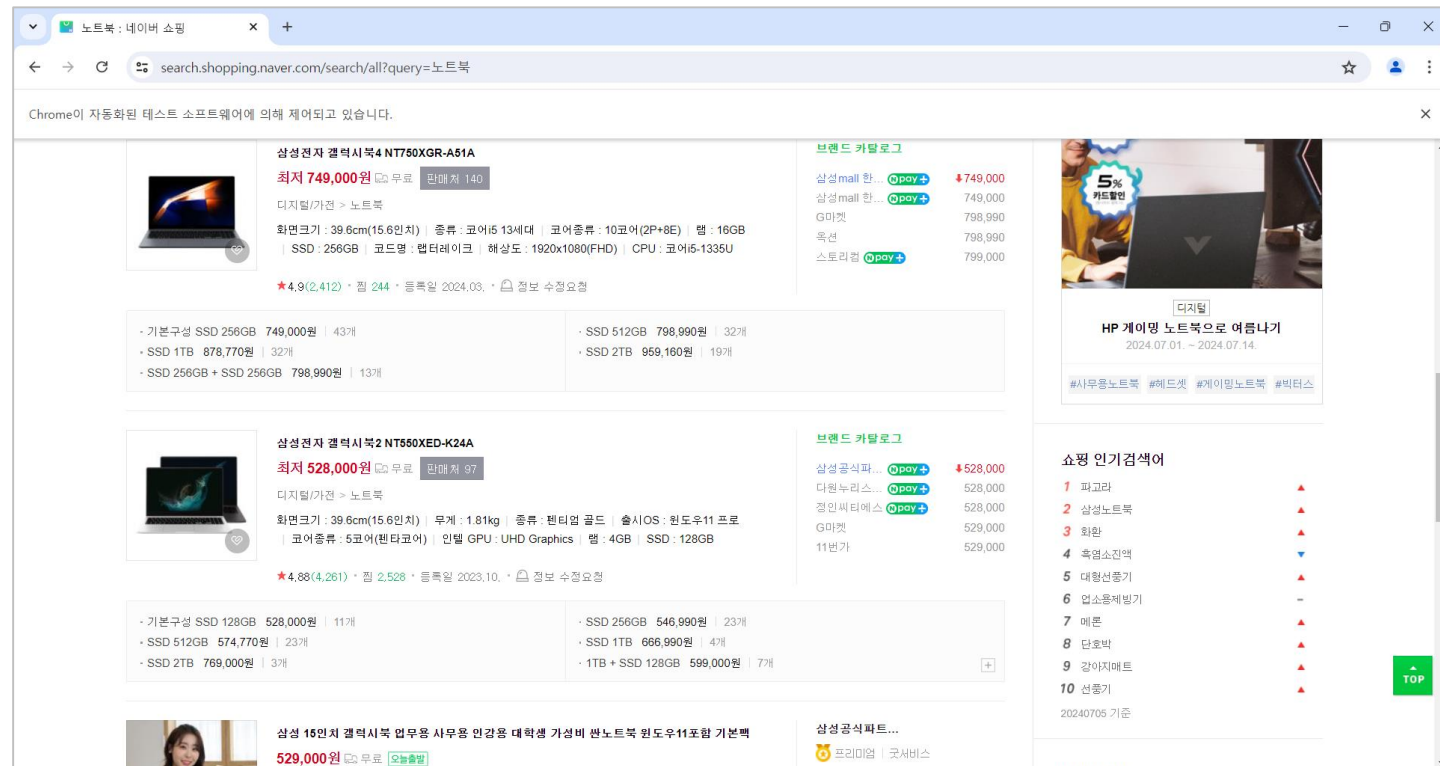
```
driver.execute_script("window.scrollTo(0,500)")
```



스크롤

상대 위치로 스크롤

```
driver.execute_script("window.scrollTo(0,500)")
```



스크롤

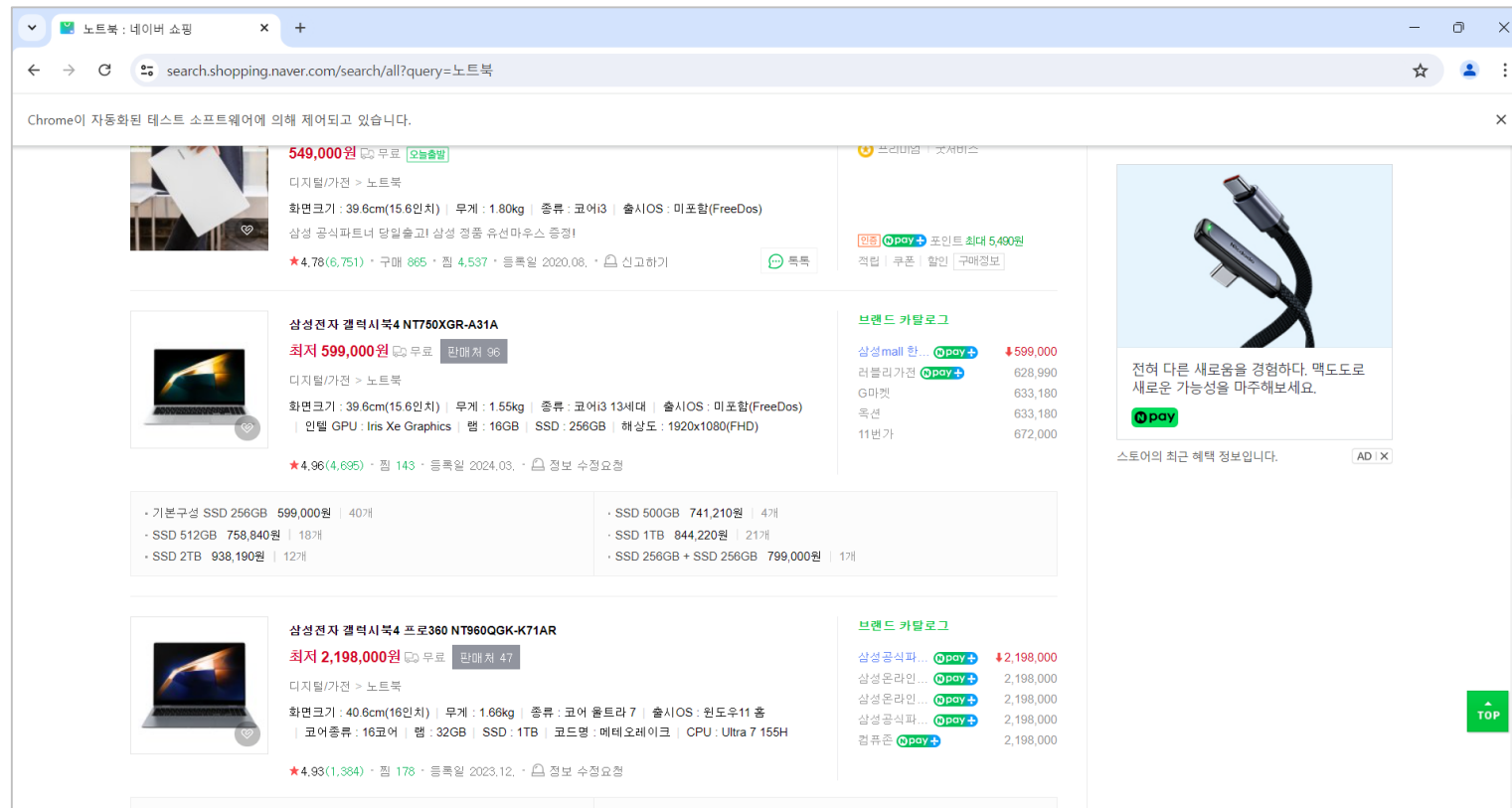
```
# 웹페이지 전체 스크롤 높이  
driver.execute_script("return document.body.scrollHeight")
```

2740

스크롤

페이지 끝까지 스크롤

```
driver.execute_script("window.scrollTo(0, document.body.scrollHeight)")
```



스크롤

```
# 무한스크롤

# 페이지의 스크롤 높이 받아오기
last_height = driver.execute_script("return document.body.scrollHeight")

while True:
    # 스크롤 끝까지 내리기
    driver.execute_script("window.scrollTo(0,document.body.scrollHeight)")

    # 현재 스크롤 위치 받아오기
    current_height = driver.execute_script("return document.body.scrollHeight")

    # 마지막스크롤 = 현재스크롤이면 반복 벗어나기
    if last_height == current_height: break

    # 현재 스크롤을 마지막 스크롤로 업데이트 하기
    last_height = current_height

    time.sleep(1)
```

데이터 수집

2024 LG그램 인텔 Ultra5 16GB 256GB 가벼운 노트북 사무용 대학생 인강용

1,199,000원  무료 오늘출발

디지털/가전 > 노트북

화면크기 : 35.56cm(14인치) | 코어종류 : 14코어(4P+8E+2LE) | 램 : 16GB | CPU : Ultra 5 125H
| 출시년도 : 2024년

동일한 클래스를 가지는 요소가 여러 개

★ 4.91 (581) · 구매 607 · 찜 1,697 · 등록일 2023.08. ·  신고하기

 톡톡

동일한 클래스를 가지는 요소가 여러 개
있을 수도 있고 없을 수도 있다.

데이터 수집

상품명, 상세페이지링크, 최저가, 별점, 리뷰갯수, 등록일

```
html = driver.page_source
soup = BeautifulSoup(html, 'html.parser')

item_list = []
items = soup.select('.product_info_area__xxCTi')
for item in items:
    product_name = item.select_one('.product_title__Mmw2K').text
    link = item.select_one('.product_link__TrAac.linkAnchor').attrs['href']
    price = item.select_one('.price_num__S2p_v').text.replace(',','').replace('원','')
    star_score = item.select_one('.product_grade__IzyU3').text.replace('별점','')
    review_cnt = item.select_one('.product_num__fafa5').text.replace('(', '').replace(')', '').replace(',','')

    # 구매건수
    try:
        purchase = item.select_one('[data-nclick*=purchase] em').text.replace(',','')
    except:
        purchase = ''

    # 등록일
    elements = item.select('.product_etc__LGVaW')
    for element in elements:
        if '등록일' in element.text: reg_date=element.text.replace('등록일 ','')

    item_list.append([product_name, link, price, star_score, review_cnt, purchase, reg_date])
```

데이터 수집

2024 LG그램 인텔 Ultra5 16GB 256GB

1,199,000원  무료 오늘출발

디지털/가전 > 노트북

화면크기 : 35.56cm(14인치) | 코어종류

| 출시년도 : 2024년

★ 4.91 (581) · 구매 607 · 찜 1,697 ·

```
<a target="_blank" rel="noopener" class="product_etc_
_LGVaW linkAnchor" data-nclick="N=a:lst*N.purchase.i:
86636005031,r:10" data-i="86636005031" data-ms="29659
09" href="https://cr.shopping.naver.com/adcr.nhn?x=Gu
AtoeuglzVGbzlrnYxnd%2F%2F%2F...llyPFUpWztUfPMYdg1QmAmn
oJRdq4K9btjkCM0%3D&nvMid=86636005031&catId=5000015
1"> == $0
```

```
::before
```

```
<span>
```

```
<span>
```

```
"구매"
```

```
<em class="product_num__fafe5">607</em>
```

```
</
```

```
</sp
```

```
</a>
```

data-link 속성에 'purchase'가 포함되어 있는 요소의
자손 중 em태그의 text

파일로 저장

```
import time

# 파일명 생성 함수
def filename(input_search):
    save_datetime = time.localtime()
    year = save_datetime.tm_year
    mon = save_datetime.tm_mon
    mday = save_datetime.tm_mday
    hour = save_datetime.tm_hour
    min = save_datetime.tm_min
    sec = save_datetime.tm_sec
    save_datetime = f'{year}{mon:02d}{mday:02d}_{hour:02d}{min:02d}{sec:02d}'
    file_name = f'{input_search}_{save_datetime}'

    return file_name

# 파일 저장
import openpyxl

df = pd.DataFrame(item_list,
                  columns = ['상품명', '상세페이지url', '최저가', '별점', '리뷰수', '구매건수', '등록일'])

df.to_excel(f'{filename(input_search)}.xlsx', index=False)
```

미션

수집할 데이터 개수 입력받아
개수만큼 수집하기