

Selenium 2

- 스크래핑
 - G마켓 베스트 상품 데이터

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

패키지 import

```
In [2]: import requests
from selenium import webdriver
import time
```

1. Scraping

- G마켓 베스트 셀러 페이지 크롤링
- 크롤링한 데이터를 전처리 (원가 데이터가 없는 경우 판매가를 넣음)
- 할인율을 계산한 컬럼을 추가
- 할인율이 50%이상인 상품만 필터링

웹드라이버 실행 및 페이지 이동

```
In [3]: driver = webdriver.Chrome("C:/Myexam/chromedriver/chromedriver.exe")
url = "http://corners.gmarket.co.kr/Bestsellers"
driver.get(url)
```

아이템 엘리먼트들 선택

```
In [4]: items = driver.find_elements_by_css_selector(
".best-list:nth-child(3) > ul > li")
len(items)
```

Out[4]: 200

제목 링크 원가, 판매가 텍스트 데이터를 크롤링

```
In [5]: dict_list = []
for item in items:
    title = item.find_element_by_css_selector(".itemname").text
    link = item.find_element_by_css_selector(".itemname").get_attribute("href")
    o_price = item.find_element_by_css_selector(".o-price").text[:-1]
    s_price = item.find_element_by_css_selector(
".s-price").text.split(" ")[0][:-1]

    data = {
        "title": title,
        "link": link,
        "o_price": o_price,
        "s_price": s_price,
    }

    dict_list.append(data)
```

```
In [6]: # 브라우저 닫기
driver.quit()
```

데이터 프레임으로 만들기

```
In [7]: import pandas as pd
```

```
In [8]: df = pd.DataFrame(dict_list)
df.tail()
```

Out[8]:

	title	link	o_price	s_price
195	[닉스]닉스 홈 맨투맨 슬랙스 청바지 외 특가전	http://item.gmarket.co.kr/Item?goodscode=17601...	25,600	12,800
196	[한상궁김치]100% 국산 열무김치 2kg 열무비빔밥 열무국수에 제격	http://item.gmarket.co.kr/Item?goodscode=17981...	18,600	13,000
197	[돌]DOLE 돌후룻팜40팩(피치+애플+오렌지+망고+파인)	http://item.gmarket.co.kr/Item?goodscode=20449...	17,900	16,650
198	[플라팜]허그 유아 삼푸의자 목욕의자	http://item.gmarket.co.kr/Item?goodscode=66728...	51,800	25,900
199	[홈에디션명진]대용량 롤백 500매x3개 위생백 지퍼백 위생장갑	http://item.gmarket.co.kr/Item?goodscode=21349...	10,580	10,000

값이 없는 o_price를 s_price로 넣기

```
In [9]: nodata_indexs = df["o_price"] == ""
```

```
In [10]: df["o_price"][nodata_indexs].tail()
```

Out[10]: 163
180
182
190
193
Name: o_price, dtype: object

```
In [11]: df["s_price"][nodata_indexs].tail()
```

Out[11]: 163 29,000
180 20,700
182 19,900
190 29,500
193 24,500
Name: s_price, dtype: object

```
In [12]: df[nodata_indexs].tail()
```

Out[12]:

	title	link	o_price	s_price
163	역주행상품 조이풀 천연소가죽 이지샌들	http://item.gmarket.co.kr/Item?goodscode=21421...		29,000
180	[잔슨빌소시지]잔슨빌 폭립 1+1 (참나무BBQ 500g + 갈비맛 500g)	http://item.gmarket.co.kr/Item?goodscode=20895...		20,700
182	파격조건 고메 함박스테이크 152gX10개/핫도그 80gX24봉 중 택1	http://item.gmarket.co.kr/Item?goodscode=21421...		19,900
190	[상하치즈]상하 유기농 어린이 고칼슘치즈 4단계 100매/아기치즈	http://item.gmarket.co.kr/Item?goodscode=21799...		29,500
193	롯데칠성음료 T 롯데칠성 칠성사이다 제로 250ml x 3	http://item.gmarket.co.kr/Item?goodscode=21078...		24,500

값이 없는 o_price를 s_price로 넣기

```
In [13]: df["o_price"][nodata_indexs] = df["s_price"][nodata_indexs]
```

```
In [14]: df.loc[nodata_indexs].tail()
```

Out[14]:

	title	link	o_price	s_price
163	역주행상품 조이풀 천연소가죽 이지샌들	http://item.gmarket.co.kr/Item?goodscode=21421...	29,000	29,000
180	[잔슨빌소시지]잔슨빌 폭립 1+1 (참나무BBQ 500g + 갈비맛 500g)	http://item.gmarket.co.kr/Item?goodscode=20895...	20,700	20,700
182	파격조건 고메 함박스테이크 152gX10개/핫도그 80gX24봉 중 택1	http://item.gmarket.co.kr/Item?goodscode=21421...	19,900	19,900
190	[상하치즈]상하 유기농 어린이 고칼슘치즈 4단계 100매/아기치즈	http://item.gmarket.co.kr/Item?goodscode=21799...	29,500	29,500
193	롯데칠성음료 T 롯데칠성 칠성사이다 제로 250ml x 3	http://item.gmarket.co.kr/Item?goodscode=21078...	24,500	24,500

```
In [15]: df.tail()
```

Out[15]:

	title	link	o_price	s_price
195	[닉스]닉스 홈 맨투맨 슬랙스 청바지 외 특가전	http://item.gmarket.co.kr/Item?goodscode=17601...	25,600	12,800
196	[한상궁김치]100% 국산 열무김치 2kg 열무비빔밥 열무국수에 제격	http://item.gmarket.co.kr/Item?goodscode=17981...	18,600	13,000
197	[돌]DOLE 돌후룻팜40팩(피치+애플+오렌지+망고+파인)	http://item.gmarket.co.kr/Item?goodscode=20449...	17,900	16,650
198	[플라팜]허그 유아 삼푸의자 목욕의자	http://item.gmarket.co.kr/Item?goodscode=66728...	51,800	25,900
199	[홈에디션명진]대용량 롤백 500매x3개 위생백 지퍼백 위생장갑	http://item.gmarket.co.kr/Item?goodscode=21349...	10,580	10,000

할인을 계산해서 컬럼에 추가하기

```
In [16]: df["o_price"] = df["o_price"].apply(
    lambda price: price.replace(".", "").astype("int"))
```

```
In [17]: df["s_price"] = df["s_price"].apply(
    lambda price: price.replace(".", "").astype("int"))
```

```
In [18]: df["sale_rate"] = round((1 - df["s_price"] / df["o_price"]) * 100, 2)
```

```
In [19]: df.tail()
```

Out[19]:

		title	link	o_price	s_price	sale_rate
195		[닉스]닉스 홈 맨투맨 슬랙스 청바지 외 특가전	http://item.gmarket.co.kr/Item?goodscode=17601...	25600	12800	50.00
196	[한상궁김치]	100% 국산 열무김치 2kg 열무비빔밥 열무국수에 제격	http://item.gmarket.co.kr/Item?goodscode=17981...	18600	13000	30.11
197	[돌]	DOLE 돌후룻팜40팩(피치+애플+오렌지+망고+파인)	http://item.gmarket.co.kr/Item?goodscode=20449...	17900	16650	6.98
198		[플라팜]허그 유아 샴푸의자 목욕의자	http://item.gmarket.co.kr/Item?goodscode=66728...	51800	25900	50.00
199	[홈에디션명진]	대용량 롤백 500매x3개 위생백 지퍼백 위생장갑	http://item.gmarket.co.kr/Item?goodscode=21349...	10580	10000	5.48

```
In [20]: df[df["sale_rate"] > 50].sort_values("sale_rate", ascending=False).reset_index(drop=True).head()
```

Out[20]:

		title	link	o_price	s_price	sale_rate
0	[곰짝마스크]	블루본 아이노우 썸머라인 소형 마스크 KF80 KF94	http://item.gmarket.co.kr/Item?goodscode=20954...	100000	30000	70.0
1		반반한샵 가을신상/블라우스/원피스/팬츠/SET	http://item.gmarket.co.kr/Item?goodscode=21782...	33000	9900	70.0
2	[오가닉맘]	오가닉맘 21FW 가을신상내의/출산선물/아동팬티	http://item.gmarket.co.kr/Item?goodscode=18742...	61000	18300	70.0
3		요즘예/가을신상+10%/청바지/슬랙스/면팬츠/밴딩/3XL~	http://item.gmarket.co.kr/Item?goodscode=18984...	33000	9900	70.0
4	[또래친구]	주니어 가을OPEN+여름OFF 균일가 최종 마감전	http://item.gmarket.co.kr/Item?goodscode=67104...	23000	6900	70.0

```
In [ ]:
```