

## Gmarket

- 베스트 상품 200개 데이터 수집
- 상품의 이미지 200개 다운로드

In [1]:

```
import pandas as pd
import requests
from bs4 import BeautifulSoup
```

In [2]:

```
# 1. URL 찾기
```

In [3]:

```
url = "http://corners.gmarket.co.kr/Bestsellers"
```

In [4]:

```
# 2. request > response : str(html)
```

In [5]:

```
response = requests.get(url)
response
```

Out[5]:

```
<Response [200]>
```

In [6]:

```
# 3. bs > DataFrame
```

In [7]:

```
dom = BeautifulSoup(response.text, "html.parser")
```

In [8]:

```
# select items
elements = dom.select("#gBestWrap > div.best-list > ul > li")
len(elements)
```

Out[8]:

```
200
```

In [9]:

```
element = elements[0]
```

In [10]:

```
# select item data
data = {
    "title": element.select_one(".itemname").text,
    "link": element.select_one(".itemname").get("href"),
    "img": element.select_one("img").get("data-original"),
    "o_price": element.select_one(".o-price").text,
    "s_price": element.select_one(".s-price").text.strip().split("\n")[0],
}
data
```

Out[10]:

```
{'title': '[베베숲]NEW 베베숲 프라임 70매 캡 20팩 신규런칭/고평량 도톰한 아기물티슈/7  
년 연속 판매 1위 베베숲',  
'link': 'http://item.gmarket.co.kr/Item?goodscode=3039555146&ver=2023  
0829',  
'img': None,  
'o_price': '정가42,790원',  
's_price': '할인가30,570원 28%'}
```

In [11]:

```
# make DataFrame
data = []
for element in elements:
    data.append({
        "title": element.select_one(".itemname").text,
        "link": element.select_one(".itemname").get("href"),
        "img": "http:" + element.select_one("img").get("src"),
        "o_price": element.select_one(".o-price").text,
        "s_price": element.select_one(".s-price").text.strip().split("\n")[0],
    })
df = pd.DataFrame(data)
df.tail(2)
```

Out[11]:

	title	link	img	o_price
198	[스타벅스] 스타벅스 카페라떼 컵커피 200mlx20 컵+아이스 포장	http://item.gmarket.co.kr/Item? goodscode=15521...	http://gdimg.gmarket.co.kr/1552146498/still/30...	
199	찾아도 찾 아도 끝판 왕 1000개 숨은 그림 찾기 5종세 트 선택구 매	http://item.gmarket.co.kr/Item? goodscode=29799...	http://gdimg.gmarket.co.kr/2979976232/still/30...	정가 40,000 원

In [12]:

```
df.head()
```

Out[12]:

	title	link	img	o_price
0	[베베숲]NEW 베베숲 프라임 70매 캡 20팩 신규런칭/고평 량 도통한 아기 물티슈/...	http://item.gmarket.co.kr/Item? goodscode=30395...	http://gding.gmarket.co.kr/3039555146/still/30...	정가 42,790 원
1	[크리넥스] (15%+클럽 12% 쿠폰적용 가 11180) 크 리넥스 안심클 래식 키친타 올...	http://item.gmarket.co.kr/Item? goodscode=27081...	http://gding.gmarket.co.kr/2708176714/still/30...	정가 21,420 원
2	[프레시누리]청 정우 홍두깨살 바로구이 육전 용 250gX3팩	http://item.gmarket.co.kr/Item? goodscode=22541...	http://gding.gmarket.co.kr/2254199961/still/30...	정가 26,500 원
3	[아디다스](대전 신세계)[New Performance] 아디다스 갤러 시 13종 택1	http://item.gmarket.co.kr/Item? goodscode=27404...	http://gding.gmarket.co.kr/2740413857/still/30...	정가 48,500 원
4	[엘에이]올인원 캡슐세탁세제 (1+2앵글)엘에 이 핑크/그린 30개입x3 (총 90개입)	http://item.gmarket.co.kr/Item? goodscode=25153...	http://gding.gmarket.co.kr/2515395333/still/30...	정가 31,500 원

In [13]:

```
# 4. download image
```

In [14]:

```
# make directory
import os
path = 'data'
if not os.path.exists(path):
    os.makedirs(path)
```

In [15]:

```
%ls data
```

000.png 001.png 002.png 003.png 004.png test.png

In [16]:

```
img_link = df.loc[0, "img"]  
print(img_link)
```

<http://gdimg.gmarket.co.kr/3039555146/still/300?ver=20230829> (<http://gdimg.gmarket.co.kr/3039555146/still/300?ver=20230829>)

In [17]:

```
# download image
```

In [18]:

```
response = requests.get(img_link)  
response
```

Out[18]:

<Response [200]>

In [19]:

```
with open(f"{path}/test.png", "wb") as file:  
    file.write(response.content)
```

In [20]:

```
%ls data
```

000.png 001.png 002.png 003.png 004.png test.png

In [21]:

```
from PIL import Image as pil
```

In [22]:

```
pil.open(f"{path}/test.png")
```

Out[22]:



In [23]:

```
# 5개의 아이템 이미지 다운로드
```

```
for idx, data in df[:5].iterrows():
    filename = "0" * (3 - len(str(idx))) + str(idx)
    print(idx, end=" ")
    response = requests.get(data.img)
    with open(f"{path}/{filename}.png", "wb") as file:
        file.write(response.content)
```

0 1 2 3 4

In [24]:

```
%ls data
```

000.png 001.png 002.png 003.png 004.png test.png

In [25]:

```
pil.open(f"{path}/003.png")
```

Out[25]:

