Gmarket

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

In [1]:
```
import pandas as pd
import requests
from bs4 import BeautifulSoup
```

1. URL 찾기

In [3]:
```
url = "https://www.gmarket.co.kr/n/best"
```

2. request > response : str(html)

In [5]:
```
headers = {
    "user-agent": "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebK
    "cookie": "cguid=11692834805054007282000000; pguid=21692834805054007282(
    "accept-language": "ko-KR,ko;q=0.9,en-US;q=0.8,en;q=0.7",
    "sec-ch-ua-platform": "macOS",
}
response = requests.get(url, headers=headers)
response
```

Out[5]:
```
<Response [200]>
```

3. bs > DataFrame

In [7]:
```
dom = BeautifulSoup(response.text, "html.parser")
```

In [8]:
```
# select items
elements = dom.select("#container > div.box__best-list > ul > li")
len(elements)
```

Out[8]:
```
200
```

In [9]:
```
element = elements[1]
```

In [10]:
```
# select item data
item = {
    "title": element.select_one(".box__item-title").text,
    "link": element.select_one("a").get("href"),
    "img": "https:" + element.select_one("img").get("src"),
    "s_price": element.select_one(
        ".box__price-seller > .text__value").text,
}
try:
    item["o_price"] = element.select_one(
        ".box__price-original > .text__value").text
except:
```

```
    item["o_price"] = None
item
```

Out[10]:
```
{'title': '(빽다방) 디지털 금액권 1만원권',
 'link': 'http://item.gmarket.co.kr/Item?goodscode=4113465600&ver=2024091
1',
 'img': 'https://gdimg.gmarket.co.kr/4113465600/still/300?ver=1724905130',
 's_price': '8,900',
 'o_price': '10,000'}
```

In [11]:
```
# make DataFrame
items = []

for element in elements:
    item = {
        "title": element.select_one(".box__item-title").text,
        "link": element.select_one("a").get("href"),
        "img": "https:" + element.select_one("img").get("src"),
        "s_price": element.select_one(
            ".box__price-seller > .text__value").text,
    }
    try:
        item["o_price"] = element.select_one(
            ".box__price-original > .text__value").text
    except:
        item["o_price"] = None
    items.append(item)

df = pd.DataFrame(items)
df.tail(2)
```

Out[11]:

	title	link	img	s_pri
198	(신선집중) NH카드/ 새코롬 감귤 고당도 비가림 하우스감귤 2.5kg 로얄과	http://item.gmarket.co.kr/Item?goodscode=26151...	https://gdimg.gmarket.co.kr/2615106227/still/3...	18,9
199	(4800원/ 무료배송)리바이스 스테디셀러 드로즈 시리즈	http://item.gmarket.co.kr/Item?goodscode=40007...	https://gdimg.gmarket.co.kr/4000728540/still/3...	6,0

In [12]:
```
# 데이터 전처리
df1 = df.copy()
none_idx = df1[df1['o_price'].isnull()].index
df1.loc[none_idx, 'o_price'] = df1['s_price']
df1.tail(2)
```

Out[12]:

	title	link	img	s_pri
198	(신선집중) NH카드/ 새코롬 감귤 고당도 비가림 하우스감귤 2.5kg 로얄과	http://item.gmarket.co.kr/Item?goodscode=26151...	https://gdimg.gmarket.co.kr/2615106227/still/3...	18,9
199	(4800원/ 무료배송)리바이스 스테디셀러 드로즈 시리즈	http://item.gmarket.co.kr/Item?goodscode=40007...	https://gdimg.gmarket.co.kr/4000728540/still/3...	6,0

4. Download Image

In [14]:
```
# make directory
import os

if not os.path.exists("data"):
    os.makedirs("data")
```

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)
```

```
https://gdimg.gmarket.co.kr/3580541703/still/300?ver=1725845882
```

In [17]:
```
# download image
```

In [18]:
```
response = requests.get(img_link)
response
```

Out[18]:
```
<Response [200]>
```

In [19]:
```
with open("data/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("data/test.png")
```

Out[22]:

In [23]:

```python
# 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"data/{filename}.png", "wb") as file:
        file.write(response.content)
```

```
0 1 2 3 4
```

In [24]:

```python
%ls data
```

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

In [25]:

```python
pil.open("data/002.png")
```

Out[25]: