

데싱디바 사이트 전체 페이지 이미지 링크 추출

```
In [93]: import requests
from bs4 import BeautifulSoup

headers = {'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/58.0.3029.11'}
all_product_urls = [] # 전체 제품 링크 저장할 리스트

### 1~8 한 페이지씩 방문 ###
for i in range(1,9):
    url = f'https://dashingdiva.co.kr/category/%EB%84%A4%EC%9D%BC/82/?page={i}'
    response = requests.get(url, headers=headers) # request 요청

    if response.status_code == 200:
        soup = BeautifulSoup(response.text, "lxml")
        prdImg = soup.select('div.thumbnail > div > a')
        print(f'{i}번째 페이지 제품 수 :', len(prdImg)) # 페이지 당 제품 수

        # 이미지 링크 추출
        for img_link in prdImg:
            href = img_link['href']
            href = 'https://dashingdiva.co.kr' + href
            all_product_urls.append(href) # 이미지 링크 리스트에 추가

    else:
        print('페이지를 가져올 수 없습니다.')

print()
print(f'총 제품 수는 {len(all_product_urls)}')
print('===== 제품 urls =====')
print(all_product_urls[:3])
```

1번째 페이지 제품 수 : 24
 2번째 페이지 제품 수 : 24
 3번째 페이지 제품 수 : 24
 4번째 페이지 제품 수 : 24
 5번째 페이지 제품 수 : 24
 6번째 페이지 제품 수 : 24
 7번째 페이지 제품 수 : 24
 8번째 페이지 제품 수 : 20

총 제품 수는 188

===== 제품 urls =====

```
[ 'https://dashingdiva.co.kr/product/매직프레스-21-겨울-네일-팁-33/7973/category/82/display/1/', 'https://dashingdiva.co.kr/product/매직프레스-쉬머-퀵팅아몬드/7894/category/82/display/1/', 'https://dashingdiva.co.kr/product/글레이즈-21-겨울-젤네일-33/7977/category/82/display/1/' ]
```

!! 추가 !!

feature 추가 :

- 리뷰 수(단계별) : 성공
- 별점 : 성공
- 원가 : 성공

원가

```
In [53]: # 원가 : div.price-box > ul > li.dis-product-custom
# 할인율 : div.price-box > ul > li.ec-sale-rate
import requests
from bs4 import BeautifulSoup

headers = {'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/58.0.3029.11'}
url = 'https://dashingdiva.co.kr/product/%EB%A7%A4%EC%A7%81%ED%94%84%EB%A0%88%EC%8A%A4-%EB%8A%98%EC%9D%B4%EB%86%8Dpick-%ED%95%'

response = requests.get(url, headers=headers) # request 요청

if response.status_code == 200:
    soup = BeautifulSoup(response.text, "lxml")
```

```

real_price = soup.select('div.price-box > ul > li.dis-product-custom')[0].get_text(strip=True) # 원가
now_price = soup.select('div.price-box > ul > li.prd-price')[0].get_text(strip=True) # 판매가

if real_price != now_price:
    print('할인 상품')
    print('원가 :', real_price)
    print('판매가 :', now_price)
else:
    print('no 할인 제품')
    print('가격 :', now_price)

else:
    print('페이지를 가져올 수 없습니다.')

```

no 할인 제품
가격 : 10,800원

별점

```

In [77]: # score :
data_product_code = soup.select('#prdDetail > ul > li > a > span')[0]['data-product-code']
review_url = f'https://review9.cre.ma/dashingdiva.co.kr/products/reviews?product_code={data_product_code}&iframe_id=crema-prod'
r_response = requests.get(review_url, headers=headers)

if r_response.status_code == 200:
    soup2 = BeautifulSoup(r_response.text, "lxml")
    review_score = soup2.select('div.products_reviews_summary_v2__score > span.products_reviews_summary_v2__score_text')[0].get_text()
    print(review_score)
else:
    print(0)

```

4.9

리뷰 수

```

In [91]: score_great = soup2.select('div.products_reviews_summary_v2__rcontent > ul > li:nth-child(1) > div.products_reviews_summary_v2__score')
score_good = soup2.select('div.products_reviews_summary_v2__rcontent > ul > li:nth-child(2) > div.products_reviews_summary_v2__score')
score_normal = soup2.select('div.products_reviews_summary_v2__rcontent > ul > li:nth-child(3) > div.products_reviews_summary_v2__score')
score_soso = soup2.select('div.products_reviews_summary_v2__rcontent > ul > li:nth-child(4) > div.products_reviews_summary_v2__score')

```

```
score_bad = soup2.select('div.products_reviews_summary_v2__rcontent > ul > li:nth-child(5) > div.products_reviews_summary_v2__')
print(score_great, score_good, score_normal, score_soso, score_bad)
```

9 1 0 0 0

전체 상품 정보 뽑기

In [103...

```
import requests
from bs4 import BeautifulSoup
import time

product_types = []
product_names = []
summary_lst = []
keywords_lst = []
product_codes = []
discount_prices = []
origin_prices = []
scores = []
score_great = []
score_good = []
score_normal = []
score_soso = []
score_bad = []
reviews_cnt = [score_great, score_good, score_normal, score_soso, score_bad]

start_time = time.time()

## 크롤링 시작
for idx, url in enumerate(all_product_urls):
    response = requests.get(url, headers=headers)    # request 요청
    if idx % 25 == 0:
        print(f'===== {idx}번째 크롤링 시작 =====')

    if response.status_code == 200:
        soup = BeautifulSoup(response.text, "lxml")

        ## 종류, 제품명 추출
        product_info = soup.select('div.headingArea > h2')
```

```

if product_info:      # 리스트에 요소가 있는지 확인
    product_info_txt = product_info[0].get_text() # 첫 번째 요소의 텍스트 추출
    product_type = product_info_txt.split('-')[0]
    product_name = product_info_txt.split('-')[1]
    product_types.append(product_type.strip())
    product_names.append(product_name.strip())
else:
    product_types.append(None)
    product_names.append(None)

## 상세설명 추출
summary = soup.select('span.summary')
if summary:
    summary = summary[0].get_text()
    summary_lst.append(summary)
else:
    summary_lst.append(None)

## 키워드 추출
keywords = soup.select('div.headingArea > div.simple')
keywords = keywords[0].get_text()
keywords_lst.append(keywords)

## 상품 코드 추출
if len(keywords)==0:
    product_code = soup.select('div.infoArea tr:nth-child(6) > td > span')
    if product_code:
        product_code = product_code[0].get_text()
        product_codes.append(product_code)
    else:
        product_codes.append(None)
else:
    product_code = soup.select('div.infoArea tr:nth-child(7) > td > span')
    if product_code:
        product_code = product_code[0].get_text()
        product_codes.append(product_code)
    else:
        product_codes.append(None)

## 상품 가격 추출
origin_price = soup.select('div.price-box > ul > li.dis-product-custom')[0].get_text(strip=True) # 원가

```

```

discount_price = soup.select('div.price-box > ul > li.prd-price')[0].get_text(strip=True) # 판매가
if real_price == origin_price:
    discount_prices.append(None)
    origin_prices.append(origin_price)
else:
    discount_prices.append(discount_price)
    origin_prices.append(origin_price)

## 별점
data_product_code = soup.select('#prdDetail > ul > li > a > span')[0]['data-product-code']
review_url = f'https://review9.cre.ma/dashingdiva.co.kr/products/reviews?product_code={data_product_code}&iframe_id=cr
r_response = requests.get(review_url, headers=headers)
if r_response.status_code == 200:
    soup2 = BeautifulSoup(r_response.text, "lxml")
    review_score = soup2.select('div.products_reviews_summary_v2__score > span.products_reviews_summary_v2__score_text
    scores.append(review_score)
else:
    print('!! ERROR - 리뷰 페이지 요청 실패 !! - data-product-code:', data_product_code)

## 단계별 리뷰 수
for num in range(1,6):
    review_cnt = soup2.select(f'div.products_reviews_summary_v2__rcontent > ul > li:nth-child({num}) > div.products_re
    reviews_cnt[num-1].append(review_cnt)

else:
    print('!! ERROR - 페이지를 가져올 수 없습니다. !!')

print('- 크롤링 완료 -')

end_time = time.time()
execution_time = end_time - start_time
print(f"코드 실행 시간: {execution_time} 초")

```

```

===== 0번째 크롤링 시작 =====
===== 25번째 크롤링 시작 =====
===== 50번째 크롤링 시작 =====
===== 75번째 크롤링 시작 =====
===== 100번째 크롤링 시작 =====
===== 125번째 크롤링 시작 =====
===== 150번째 크롤링 시작 =====
===== 175번째 크롤링 시작 =====
- 크롤링 완료 -
코드 실행 시간: 667.1760718822479 초

```

데이터프레임으로 저장, to_csv

```

In [104... import pandas as pd
products_info = pd.DataFrame({'product_type':product_types, 'product_name':product_names,
                             'summary':summary_lst, 'keywords':keywords_lst,
                             'product_code':product_codes, 'discount_price':discount_prices,
                             'origin_price':origin_prices, 'scores_rate':scores,
                             'great_cnt(5)':score_great, 'good_cnt(4)':score_good, 'normal_cnt(3)':score_normal,
                             'soso_cnt(2)':score_soso, 'bad_cnt(1)':score_bad})

products_info.to_csv('dashingdiva.csv', index=False)

```

```

In [105... import pandas as pd
df = pd.read_csv('dashingdiva.csv')
df.info()
# summary - 실제로 없음
# keyword - 실제로 없음
# product_code - [2+1]상품은 제품코드 미포함

```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 188 entries, 0 to 187
Data columns (total 13 columns):
#   Column                Non-Null Count  Dtype
---  -
0   product_type          188 non-null    object
1   product_name          188 non-null    object
2   summary               185 non-null    object
3   keywords              165 non-null    object
4   product_code          185 non-null    object
5   discount_price        98 non-null     object
6   origin_price          188 non-null    object
7   scores_rate           188 non-null    object
8   great_cnt(5)          188 non-null    int64
9   good_cnt(4)           188 non-null    int64
10  normal_cnt(3)         188 non-null    int64
11  soso_cnt(2)           188 non-null    int64
12  bad_cnt(1)            188 non-null    int64
dtypes: int64(5), object(8)
memory usage: 19.2+ KB
```

In [106... `df.head()`

Out[106]:

	product_type	product_name	summary	keywords	product_code	discount_price	origin_price	scores_rate	great_cnt(5)	good_cnt(4)
0	매직프레스	[2+1] 겨울 네일 팁 ~33%	NaN	33% 할인, 겨울 신상 2+1,신상 네일 2개 구매 시 1종 증정	NaN	21,600원	0원	5.0	177	4
1	매직프레스	쉬머 켈팅(아몬드)	영롱한 자석젤 위에 3D 엠보로 모양을 낸 켈팅 네일	늘이닝 PICK, 켈팅, 자석젤, 마그네틱, 3D	MDR3W045PA	14,800원	14,800원	5.0	9	0
2	글레이즈	[2+1] 겨울 젤 네일 ~33%	NaN	33% 할인, 겨울신상 2+1,신상 네일 2개 구매 시 1종 증정	NaN	29,600원	0원	5.0	24	0
3	매직프레스	드리미 블랙(코핀)	고급스러운 실버 별빛과 어우러진 마그네틱 네일	늘이닝 PICK, 23.11 출시, 실버 별, 잉크, 틴트, 마블, 마그네틱, 자석젤	MDR3W036CF	NaN	10,800원	5.0	3	0
4	매직프레스	스노우 볼(레귤러 라운드)	스노우볼과 눈꽃 스톤으로 겨울을 표현한 프리미엄 네일	크리스마스, 스노우볼, 스톤, 눈꽃	MDR3W040PR	14,800원	14,800원	5.0	2	0



In [107...]

df[df.product_code.isna()]

Out[107]:

	product_type	product_name	summary	keywords	product_code	discount_price	origin_price	scores_rate	great_cnt(5)	good_cnt(4)	n
0	매직프레스	[2+1] 겨울 네일 팁 ~33%	NaN	33% 할인, 겨울 신상 2+1,신상 네일 2개 구매 시 1 종 증정	NaN	21,600원	0원	5.0	177	4	
2	글레이즈	[2+1] 겨울 젤 네일 ~33%	NaN	33% 할인, 겨울신상 2+1,신상 네일 2개 구매 시 1 종 증정	NaN	29,600원	0원	5.0	24	0	
6	글레이즈	[2+1] 글레이즈 BEST ~33%	NaN	33% 할인, 가을 신상 2+1,신상 네일 2개 구매 시 1 종 증정	NaN	29,600원	0원	4.9	185	10	



In []: