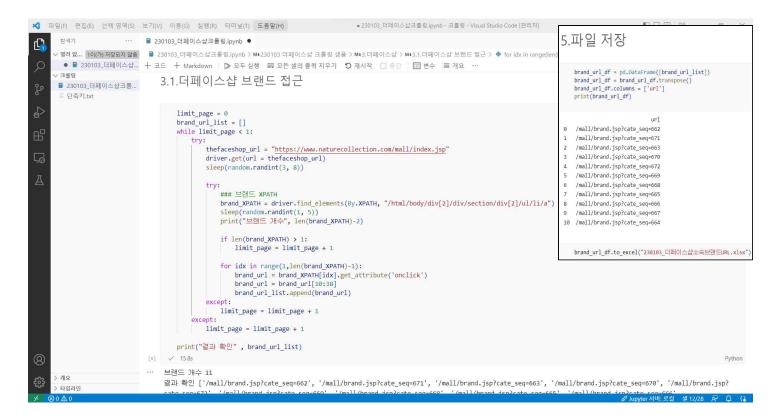
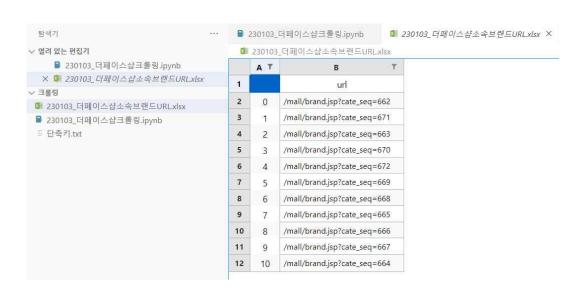
크롤링 - 더페이스샵

1. 소속 브랜드 URL 추출

(코드)



(결과물)

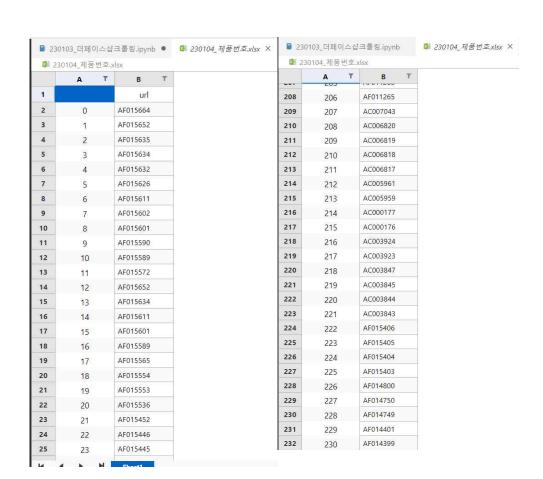


2. 제품 번호 추출

(코드)

```
▲ 파일(F) 편집(E) 선택 영역(S) 보기(V) 이동(G) 실행(R) 터미널(T) 도움말(H)
                                                                                 • 230103_더페이스샵크를링 6.3. 파일 저장 (제품 번호)
      🛢 230103 더페이스샵크톨링.jpynb > M4230103 더페이스샵크롤링 > M46.제품 정보 가져오기 > M46.1.드라이버 오픈 > 🍨 product nu
                                                                                                        product_num_df = pd.DataFrame([product_num_list])
      + 코드 + Markdown | ▶ 모두실행 ➡ 모든셀의출력지우기 5 재시작 □ 중단 | 屆 변수 ※ 개요 …
                                                                                                        product_num_df = product_num_df.transpose()
                                                                                                        product_num_df.columns = ['url']
                                                                                                        print(product_num_df)
              brand_url_list = []
              product_num_list = []
                                                                                                             ur1
              for idx in range(3): #len(brand_df)
                                                                                                    0
                                                                                                         AF015664
                  print(idx+1, "========
                                                                                                         AF015652
                  brand_url = "https://www.naturecollection.com/m" + brand_df['url'][idx]
                                                                                                         AF015635
                                                                                                    2
                                                                                                    3
                                                                                                         AF015634
                     driver.get(url = brand_url)
                                                                                                         AF015632
                     brand_url_list.append(brand_url)
                     sleep(random.randint(3,7))
                                                                                                    226
                                                                                                         AF014800
                     print("yes")
                                                                                                    227 AF014750
                                                                                                    228 AF@14749
                         ## 클릭(카테고리)
                                                                                                    229
                                                                                                         AF014401
                         category_button_XPATH = driver.find_elements(By.XPATH, "/html/body/section/div/di
                                                                                                    230 AF014399
                         print("카테고리:", len(category_button_XPATH))
                                                                                                    [231 rows x 1 columns]
                         for bidx in range(len(category_button_XPATH)):
                            category_name = category_button_XPATH[bidx].text
                             if "헤어" in category_name:
                                                                                                        product_num_df.to_excel("230104_제품번호.xlsx")
                                continue
                             elif "메이크얼" in category_name:
                                continue
                                category_button_XPATH[bidx].click()
                                sleep(random.randint(2,5))
                                    ## 제품 번호 추출
                                    product_num_XPATH = driver.find_elements(By.XPATH, "/html/body/section/div/div/section/div[5]/div[1]/div/ul/li/figure/a")
                                    print(len(product_num_XPATH))
```

(결과물)



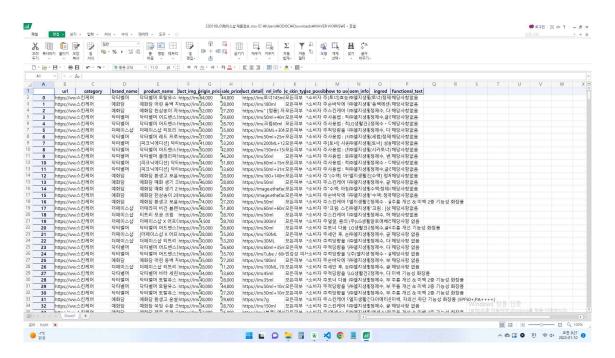
3. 제품 상세정보 추출

(코드)

3.3. 상세 페이지 접속

```
product_url_list = []
category_list = []
brand_name_list = []
product_name_list = []
product_img_list = []
origin_price_list = []
sale_price_list = []
# sale_time_list = []
product_detail_img_list = []
## 구매정보 클릭 후, 수집되는 리스트
ml_info_list = []
REC_skin_type_list = []
use_possible_list = []
how_to_use_list = []
oem_list = []
ingred_list = []
functional_test_list = []
plen = len(product_num_df);
for idx in range(plen):
    product_url = "https://www.naturecollection.com/mall/product/product-view.jsp?dpid=" + product_num_df['url'][idx]
    try:
       driver.get(url = product_url)
        ## 해당 URL 주소
        product_url_list.append(product_url)
        sleep(random.randint(3, 7))
```

(결과)



4. 제품 리뷰 정보 추출

(코드)

3.4. 리뷰 페이지 접속

(결과)

Α	В	C	D	E	F	G	Н	1	J	K	L	M	N
	url	user_ID	user_spec	rating	eview_dat	view_cote	nts						
0	https://ww	NONE	NONE	NONE	NONE	NONE							
1	https://ww	NONE	NONE	NONE	NONE	NONE							
2	https://ww	KA_1****	30대	5	23.01.04	너무 좋아	요잘 쓸게요	2~감사합니	다나한테 질	말 맞아요~7	디인에게도	강추하고싶	어요이 제
3	https://ww	KA_1****	30대	5	23.01.07	겨울철 닥	터벨머 시키	 리커버리	크림을 애용	하는데 저림	취하게 잘싰	네요. 덤으로	및 어드밴리
4	https://ww	KA_1****	30대	5	23.01.03	닥터벨머!	는 순하지만	기능도 좋	아서 제가 지	인들한테 =	추천하는 저	베품인데요!^	^토너 바
5	https://ww	11ya****	30대 ^^	4	22.12.09	닥터벨머	어드밴스드	시카 펩타	이트 앰플이	처음 나왔	을 때 써 보	고 좋아 꾸긴	돈히 쓰다?