



올리브영 제품 유사도 분석 및 추천

멋쟁이 사자 조화로운 팀

팀장: 김태욱

팀원: 김주현, 박수진, 서태원, 이정진

Chapter1



분석 주제
및 목표

Chapter2



데이터
수집방법

Chapter3



분석
프로세스

Chapter4



분석
결과

Chapter 1

분석 주제 및 목표



1. 분석주제 및 목표

‘포스트 코로나’ 뷰티 시대를 준비하다

더마코스메틱 향균 · 아이 메이크업과 ‘맞춤형 화장품’ 시장 주목

By 이아름 기자 - 2021년 2월 24일

포스트 코로나로
뷰티시장의 확대와
남녀노소 구분없이 관심도
증가



1. 분석주제 및 목표

스킨/로션/올인원 카테고리 656개의 상품이 등록되어 있습니다.

인기순 | 최근등록순 | 판매수량순 | 낮은 가격순 | 높은 가격순

VIEW 24 36 48



베스트

증정기회
백 1

라네즈

★크림스킨X노르디스크 한정기획★
대용량250ml+100ml(캡핑수건or와셔)
42,000원 **29,400원**
세일 오놀드림



베스트

철벽
보습

약산성
로션

라운드어라운드

★셔누PICK★ 라운드어라운드
그린티 약산성 고보습로션 220ml
25,000원 **19,700원**
세일 중정 오놀드림



베스트

1등
로너

브링그린

★윈다PICK★ 브링그린 티트리 시카
수딩 토너 510ml(대용량)
22,000원 **16,720원**
세일 쿠폰 중정 오놀드림



베스트

정품
증정

아벤느

★한정기획★아벤느 트릭세라 로션
200ml #윤기보습 #윤광로션...
33,000원 **23,100원**
세일 중정 오놀드림

화장품 > 기초 >
스킨/로션/올인원

화장품에 관심 없던 분들을 위한
제품 추천

제품이 너무 많아 고르는게
어려운 분들



베스트

1등
로너

아블립

★한정기획★아블립 트릭세라 로션
200ml #윤기보습 #윤광로션...
33,000원 **23,100원**
세일 중정 오놀드림



베스트

정품
증정

아벤느

★한정기획★아벤느 트릭세라 로션
200ml #윤기보습 #윤광로션...
33,000원 **23,100원**
세일 중정 오놀드림



베스트

정품
증정

아벤느

★한정기획★아벤느 트릭세라 로션
200ml #윤기보습 #윤광로션...
33,000원 **23,100원**
세일 중정 오놀드림



베스트

정품
증정

아벤느

★한정기획★아벤느 트릭세라 로션
200ml #윤기보습 #윤광로션...
33,000원 **23,100원**
세일 중정 오놀드림

1. 분석주제 및 목표

사용하고 있는 제품과 유사한 제품을 리뷰를 통해 추천

라네즈 >

★크림스킨X노르디스크 한정기획★대용량
250ml+100ml(캠핑수건or와인잔)

42,000원 **29,400원** 혜택 정보 ●

세일 오늘드림

368명이 보고있어요

배송정보

일반배송 | 2,500원 (20,000 원 이상 무료배송)
올리브영 배송 | 평균 3일 이내 배송

오늘드림 | 2,500원 또는 5,000원 ⓘ

결제혜택

THE CJ 카드 추가 10% 할인 ⓘ
CJ ONE 포인트 최대 2% 적립 예상 ⓘ

상품을 선택해주세요



1. 분석주제 및 목표

자신의 피부 고민을 적으면 제품 리뷰 기반으로 알맞은 제품 추천

ex) 요새 야근을
많이해서 그런지
여드름이 다시 부쩍
많아져서 고민이에요 ㅠ



Chapter 2

데이터 수집방법



2.데이터 수집방법

(1) Selenium 활용 -> url 추출

```
1 driver = webdriver.Chrome(executable_path='chromedriver.exe')
```

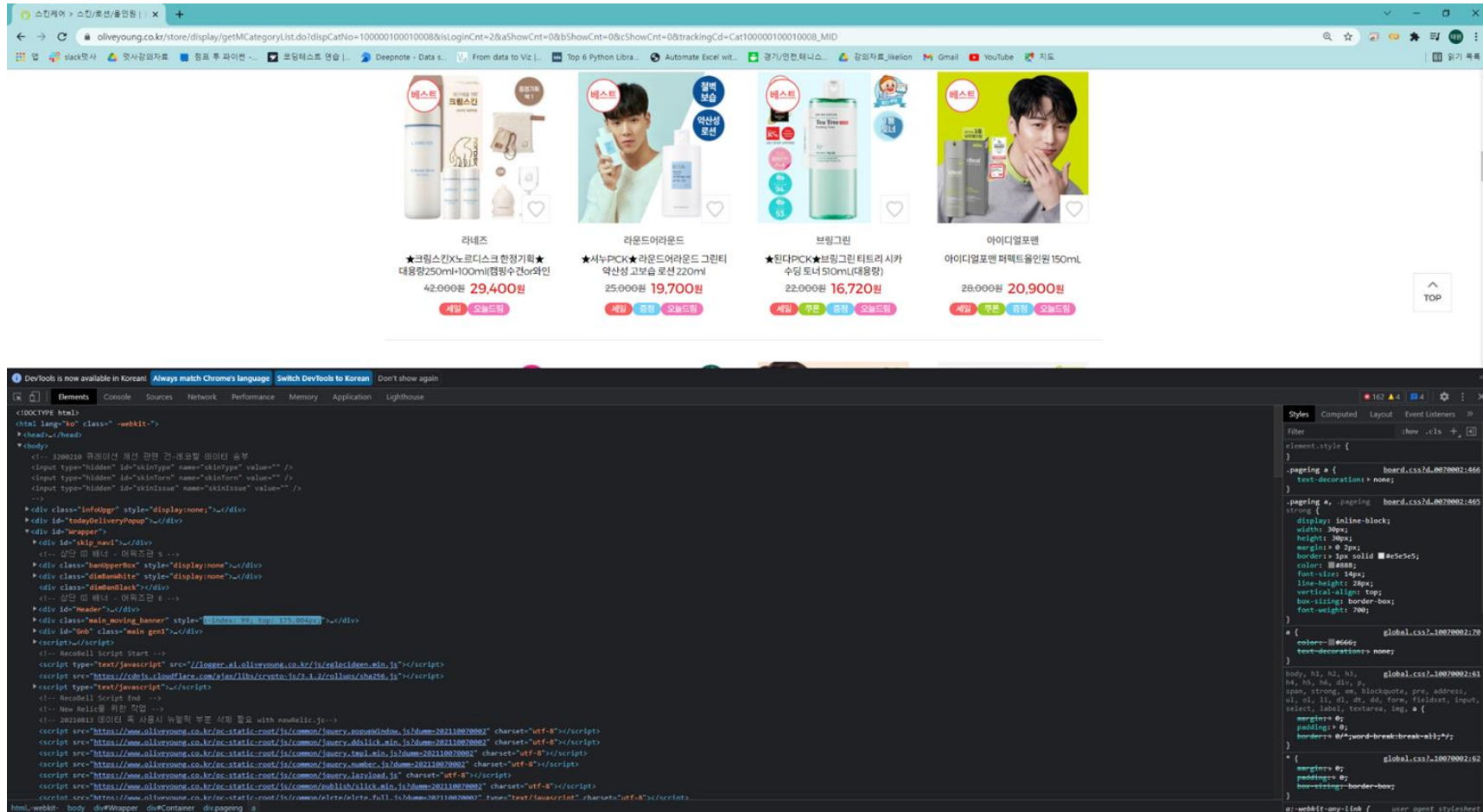
```
1 url_list=[]
```

```
1 page = 1
2
3 while page<=28:
4     try:
5         url = 'https://www.oliveyoung.co.kr/store/display/getMCategoryList.do?dispCatNo=100000100010008&fltDispCatNo=&prdSort=01&pag
6             +str(page)+'&rowsPerPage=24&searchTypeSort=btn_thumb&plusButtonFlag=N&isLoginCnt=0&aShowCnt=0&bShowCnt=0&cShowCnt=0&
7         driver.get(url)
8
9         web_page = BeautifulSoup(driver.page_source, 'html.parser')
10        for urls in web_page.find_all('a',{'class':'prd_thumb goodsList'}):
11            url_list.append(urls['href'])
12        print(str(page)+"번째 페이지 완료")
13        page+=1
14    except:
15        print("오류 발생")
16        break
```

1번째 페이지 완료
2번째 페이지 완료
3번째 페이지 완료
4번째 페이지 완료
5번째 페이지 완료
6번째 페이지 완료
7번째 페이지 완료
8번째 페이지 완료
9번째 페이지 완료
10번째 페이지 완료
11번째 페이지 완료
12번째 페이지 완료
13번째 페이지 완료
14번째 페이지 완료
15번째 페이지 완료
16번째 페이지 완료
17번째 페이지 완료
18번째 페이지 완료
19번째 페이지 완료
20번째 페이지 완료
21번째 페이지 완료
22번째 페이지 완료
23번째 페이지 완료
24번째 페이지 완료
25번째 페이지 완료
26번째 페이지 완료
27번째 페이지 완료
28번째 페이지 완료

2.데이터 수집방법

(1) Selenium 활용 -> url 추출



2.데이터 수집방법

(2) urlopen, requests.get 활용

```
1 #제품명, 브랜드 끌어오기
2 product_name = []
3 product_brand = []
4 for i in range(1,29):
5     url = 'https://www.oliveyoung.co.kr/store/display/getMCategoryList.do?dispCatNo=100000100010008&fltDispCatNo=&prdSort=01&pageIndex='
6         +str(i)+'&rowsPerPage=24&searchTypeSort=btn_thumb&plusButtonFlag=N&isLoginCnt=0&aShowCnt=0&bShowCnt=0&cShowCnt=0&trackingCd'
7     web = urlopen(url)
8     web_page = BeautifulSoup(web, 'html.parser')
9     product_name += web_page.find_all('p', {'class': 'tx_name'})
10    product_brand += web_page.find_all('span', {'class': 'tx_brand'})
11
12 product_name_real=[]
13 product_brand_real=[]
14
15 for i in range(len(product_name)):
16     product_name_real.append(product_name[i].get_text())
17     product_brand_real.append(product_brand[i].get_text())
```

```
1 # 상품별 전체 리뷰 수 끌어오기
2 product_total_review = []
3 check = 1
4
5 for i in product_link:
6     web = requests.get(i, headers=headers).text
7     web_page = BeautifulSoup(web, 'html.parser')
8     review_count = web_page.find('a', {'class': 'goods_reputation'})
9     review_count = review_count.find('span').get_text()
10    product_total_review.append(review_count[1:len(review_count)-1])
11    print(check, "완료")
12    check+=1
```

2.데이터 수집방법

(3) json 활용

https://www.oliveyoung.co.kr/store/goods/getGdasNewListJson.do?goodsNo=A000000131210&gdasSort=05&itemNo=all_search&pageIndex=1&colData=&keywordGdasSeqs=&type=&point=&hashTag=&optionValue=&cTypeLength=0

```
{ "gdasList":
[{"sysRegDtime":null,"sysRegId":null,"sysModDtime":null,"sysModId":null,"rowsPerPage":0,"pageIndex":1,"pageCalc":0,"systemDefaultLanguage":"","dbLocaleLanguage":"ko","state":null,"defaultCntryCd":"","sysRegrNm":null,"sysModrNm":null,"fieldNames":null,"texts":null,"widths":null,"excelTitle":null,"excelYn":null,"excelCurPage":null,"excelCnt":null,"excelTcnt":null,"viewCode":null,"gdasSeq":2645020,"goodsNo":"A000000131210","ordNo":null,"ordGoodsSeq":"1","lgcGoodsNo":null,"titNm":null,"gdasScrVal":10,"prcScrVal":null,"qualScrVal":null,"divScrVal":null,"gdasCont":"솔직한 리뷰 입니다.우선 티르티르 제품은 처음 써봅니다.우선 티르제품을 사용하려고 했으나 좋은 기회에 올영체험단으로 선정되어 사용하게 되었습니다.우선 약건성과 건성 사이 입니다.속이 건조하며 피부가 얇은편이나 피부에 관심이 많아 피부는 꽤 좋은편이며 평소 화장품에 관심이 많아 샤넬부터 시작해서 정말비싼 고가의 라인은 거의 써보는 편입니다.우선!본격적으로 건성피부에 좋으며 순합니다.하루는 손으로 덜어서 써보았습니다.쫄쫄 진득한 우유색 같지만 일반 토너제형이며 피부속 바르는 순간 흡수력은 좋으나 바른후 끈적임이 있습니다.끈적임을 싫어하시는 분은 비추이며 5분정도 지나면 매끈함으로 변합니다.저는 평소에 쫄쫄한 느낌을 굉장히 좋아해서 로션과 오일을 항상 섞어 바릅니다.둘쨏날에는 화장솜에 발라보았습니다.화장솜에 걸들며 흡수가 안되는 토너가 있으나 지금 이 제품은 그렇지 않아 화장솜에 흡수시켜 바르는 것이 좋은것 같습니다.얼굴에 바른후 피부속 당김이 없었으며 보습에 좋았습니다.맛! 그리고 토너후 화장했을때 밀림이나 속당김이 없었습니다.티르제품은 처음사용해보는데 다른 라인도 사용해볼 생각입니다.우선!!! 단점을 굳이 말하라고 한다면 .....흠개인적으로 찾지 못했습니다.계속해서 사용할듯 합니다.건성분을 추천이며 복합성까지도 괜찮을듯 싶어요.", "purMbrYn":"N", "qryCnt":0, "dispStatCd":"10", "dispProcDtime":1581167002000, "dispDspsId":null, "confStatCd":"10", "confProcDtime":1581167002000, "confDspsId":null, "mbrNo":"M0000004356083", "photoGdasYn":"Y", "bestGdasYn":"N", "bestGdasApprId":null, "bestGdasConfDtime":null, "divAempScrVal":null, "imgDispYn":null, "imgNonDispCausCd":null, "siteNo":null, "goodsTpCd":null, "dclDtime":null, "dclCausCd":null, "dclCausNm":null, "dclCont":null, "dclStatNm":null, "dspsId":null, "dspsNm":null, "procCont":null, "procDtime":null, "gdasSctCd":"50", "gdasTpCd":"30", "useYn":null, "pntPayYn":null, "onlMdEmpNo":null, "shrtGdasCont":"속당김 잔주름으로 고민이신분들 추천", "gdasLen":null, "evtNo":null, "evalReply":null, "itemNo":"001", "lgcGoodsNoList":null, "genSctCd":"F", "ageCd":30, "skinTypeCd":"A02", "skinToneCd":null, "skinTroubleCd":null, "offLineGoodsNo":null, "gdasDclModYn":null, "goodsNm":"티르티르 밀크스킨 기획세트(150ml+20ml)", "brndNm":null, "imgPathNm":null, "dspPathNm":null, "thnlPathNm":null, "ordDate":null, "sysRegDate":null, "divFinDate":null, "rplSeq":0, "rplCont":null, "rplSvRegDtime":null, "anslPshlDate":null, "searchMonth":null, "mobileYn":null, "startDate":null, "endDate":null, "modYn":null, "gdasPrhbPatCnt":null, "ratl
```


2.데이터 수집방법

(3) json 활용

```
1 # 리뷰 수집 코드
2 check = 1
3 product_review = []
4
5 # 제품코드 반복
6 for goodsNum in product_num:
7     try:
8         json_url = 'https://www.oliveyoung.co.kr/store/goods/getGdasNewListJson.do?goodsNo={}&gdasSort=05&itemNo=all_search&pageIndex='
9         response = requests.get(json_url, headers = headers).text
10        review_json = json.loads(response)
11        review = []
12
13        ## 나중에 Data Frame에서 리뷰가 아예없는 데이터는 삭제시키기 위해서 리뷰만 if문을 작성.
14        if len(review_json['gdasList']) != 0:
15            for i in range(len(review_json['gdasList'])):
16                if review_json['gdasList'][i]['gdasCont'] == None or review_json['gdasList'] == None:
17                    review.append('없음')
18                    continue
19                #review.append(re.sub('[\n\r]', '', review_json['gdasList'][i]['gdasCont']))
20            else:
21                review.append(review_json['gdasList'][i]['gdasCont'].replace('\r', '').replace('\n', ''))
22            product_review.append(review)
23
24            print(check, "완료")
25            check+=1
26        else:
27            product_review.append('없음')
28            print(check, "미완성")
29            check+=1
30            pass
31    except:
32        product_review.append('없음')
33        print(check, "미완성")
34        check+=1
35        pass
36
```

```
1 len(product_review)
```

656

```
1 product_review
```

[['■복합성 피부입니다. 평소 초가을부터 극건성으로 슬슬시동거는 타입인데요 그래서 전부터 한번 써보고 싶었던제품이었어요 현대 생리 이전 트러블도 나는 편이라서한달동안 여러개의 스킨을 돌아가면서 골라쓰는 피분데제 예상엔 너무 모공을 막거나 기름진않을지 걱정이였거든요 닥토보다는 닥토 하시고 한번더 바르는 용도로쓰는게 좋을거 같아요수분으로 닥고 얼굴의 기름기는 지용성으로 닥는 느낌 으로 써도 굳이겠네여사진에 보시다시피 화장솜으로 발랐는데촉촉하게 흡수가 되고 몇초지나면 번들거림도 없어여많이 건조하신 분들도이 제품과 크림만 쓰셔도 겨울 거뽀히 버틸거 같아요화장품 많이 쓴다고 좋은건 아니라 들어서여름처럼 가볍게 쓸땐 세안 후 이 제품 사용 후엔크림만 발라도 되겠네여향은 흔히 크림에서 나는 향과 스킨의 향이섞인 향 같네요 특별한 향료로 포장된 느낌은아니예요 그래서 더 좋네요꾸준히 구매할거 같아요 ●사용품 노르디스크는 텐트를 본적이 있어서 궁금했던 브랜드였어요호기심으로 구매했는데타올은 보통 세안타올크기보다 정사각형정도로 작은편이지만 간단히 걸어두고 쓰기 좋을거 같아요 막 머리감고 샤워할때 말고간단히 세안할때 쓰는건 괜찮아요.메쉬 주머니도 다른 용도로 써도 좋겠어여와인잔 너무 이쁜데여생각보다 맘에 들어서 더 살까? 고민중입니다. 케이스가 있어서 파손 위험도 적고분리되니 이동보관도 편하구여콜라보 좋네여 ㅎㅎ 2탄 3탄 콜라보 기대하고 있을게영 🍷', '노르디스크 캠핑용 수건 기획세트 출시 기념으로 세일하길래 마침 쓰던 토너도 다 썼고 해서 얼른 구매해봤습니다ㅎㅎ 세일해서 29400원이었고요, 저는 탐리뷰어 쿠폰 사용해서 저렴하게 구매할 수 있었어요!증정으로 받은 수건 짧게 리뷰하자면 행주로 쓰기에 아주 아늑마촉이예요 쿠리나 시대에 캐피은 무스ㅎ 퍼히 진 부에서 애용중입니다

2.데이터 수집방법

(3) json 활용 -> 최종 Data Frame

oliveyoung														
num	name	brand	review	skin1	skin2	skin3	score	review_skin	product_total_review	price				
0	A000000157890	★크림스킨X노르디스크 한정기획★대용량 250ml+100ml(캠핑수건or와인잔)	라네즈	['■복합성 피부 부입니다. 평소 가을부터 건성으로 술시동 거는 타입인데요 그래서 ...	['건성'에 좋아요', '건성'에 좋아요', '건성'에 좋아요', '건성'에 좋아요', '건성'에 좋아요', '건성'에 좋아요', '건성'에 좋아요', '건성'에 좋아요', '...'아 요']	['주름/미백'에 좋아 요', '보습'에 좋아 요', '자극'없이 순해 요', '자극'없이 순해 요', '자극'없이 순해 요', '자극'없이 순해 요', '자극'없이 순해 요', '자극'없이 순해 요', '...'아 요']	[오공이에 요', '자극'없이 순해 요', '보습'에 좋아 요', '자극'없이 순해 요', '자극'없이 순해 요', '자극'없이 순해 요', '자극'없이 순해 요', '자극'없이 순해 요', '...'아 요']	[5.0, 5.0, 4.0, 5.0, 5.0, 5.0, 5.0, 5.0, 5.0, ...]	['복합성', '월톤', '각질', '다크서클', '모공', '미백', '민감성',...]	1,211	29,400원	https://www.oliveyoung.co.kr/store/goods/ge		
1	A000000156235	★셔누PICK★라운드어라운드 그린티 약산성 고보습 로션 220ml	라운드어라운드	[' 30대 초반 남편피부타입 : 중건성(유분약간 올라옴).모공적은 편 약건성, 건성분들은...	['복합성'에 좋아 요', '건성'에 좋아 요', '복합성'에 좋아 요', '보습'에 좋아 요', '지성'에 좋아 요', '...'아 요',...]	['보습'에 좋아 요', '보습'에 좋아 요', '보습'에 좋아 요', '보습'에 좋아 요', '보습'에 좋아 요', '보습'에 좋아 요', '보습'에 좋아 요', '보습'에 좋아 요', '...'아 요']	['자극'없이 순해 요', '자극'없이 순해 요', '자극'없이 순해 요', '자극'없이 순해 요', '자극'없이 순해 요', '자극'없이 순해 요', '자극'없이 순해 요', '자극'없이 순해 요', '...'아 요']	[5.0, 5.0, 5.0, 5.0, 4.0, 5.0, 5.0, 5.0, 5.0, ...]	['복합성', '쿨톤', '모공', '블랙헤드', '탄력', '트러블', '피지와 다...]	48	19,700원	https://www.oliveyoung.co.kr/store/goods/ge		

Chapter 3

분석 프로세스



3. 분석 프로세스

Process 1.

추출한 데이터를 기반으로
데이터 프레임 생성



Process 2.

review 데이터와 skin type
데이터 코사인 유사도 분석
제품고민을 Input 받아 유사한
제품 추천하는 모델 완성



Process 3.

피부고민 Input을 리뷰
데이터와 cosine유사도
분석을 하여 알맞은 제품
추천하는 모델 완성

3. 분석 프로세스

```
array([[1.          , 0.76950519, 0.18062096, ..., 0.87946553, 0.2671233 ,
        0.9327413 ],
       [0.76950519, 1.          , 0.63802587, ..., 0.93585572, 0.63060354,
        0.93047241],
       [0.18062096, 0.63802587, 1.          , ..., 0.4599285 , 0.87504451,
        0.38583242],
       ...,
       [0.87946553, 0.93585572, 0.4599285 , ..., 1.          , 0.51064552,
        0.97568627],
       [0.2671233 , 0.63060354, 0.87504451, ..., 0.51064552, 1.          ,
        0.43036198],
       [0.9327413 , 0.93047241, 0.38583242, ..., 0.97568627, 0.43036198,
        1.          ]])
```

1. 리뷰 코사인 유사도 벡터화

2. 스킨 코사인 유사도 벡터화

3. 코로나 공식사이트에 셀레니움 활용하여 단계 추출

review 코사인 유사도

```
1 # review 명사값만 추출
2 review_list = []
3 for i in range(len(df_new['review'])):
4     nouns = twitter.nouns(df_new['review'][i])
5     nouns = ' '.join(nouns)
6     review_list.append(nouns)
```

```
1 df_cosine=pd.DataFrame()
2 df_cosine['review_noun'] = review_list
```

```
1 #리뷰 데이터 코사인 유사도
2 count_vect_category = CountVectorizer(min_df=0, ngram_range=(1,2))
3 review = count_vect_category.fit_transform(df_cosine['review_noun'])
4 review_cosine = cosine_similarity(review,review)
```

skin 코사인 유사도

```
1 skin1_list=[]
2 for i in range(len(df_new['skin1'])):
3     nouns = twitter.nouns(df_new['skin1'][i])
4     nouns = ' '.join(nouns)
5     skin1_list.append(nouns)
```

```
1 df_cosine['skin1_noun'] = skin1_list
```

```
1 count_vect_category = CountVectorizer(min_df=0, ngram_range=(1,2))
2 skin1 = count_vect_category.fit_transform(df_cosine['skin1_noun'])
3 skin1_cosine = cosine_similarity(skin1,skin1)
```

```
1 skin1_cosine
```

3. 분석 프로세스

1. 리뷰 코사인 유사도 벡터화

2. 스킨 코사인 유사도 벡터화

3. 코로나 공식사이트에 셀레니움 활용하여 단계 추출

별점 => 상위 10개의 리뷰 데이터 평점 평균

```
1 score_num = df_new['score'][0].replace('[', '').replace(']', '').replace("'", '').split(',')
2 score_num
```

```
: ['5.0', ' 5.0', ' 4.0', ' 5.0', ' 5.0', ' 5.0', ' 5.0', ' 5.0', ' 5.0', ' 5.0']
```

```
1 list=[]
2 for i in range(len(score_num)):
3     list.append(float(score_num[i]))
4 sum(list)
```

```
: 49.0
```

```
1 score_sum=[]
2 for i in range(len(df_new)):
3     score_num = df_new['score'][i].replace('[', '').replace(']', '').replace("'", '').split(',')
4     list=[]
5     for j in range(len(score_num)):
6         list.append(float(score_num[j]))
7     score_sum.append(sum(list)/len(list))
8 score_sum[:5]
```

```
: [4.9, 4.9, 4.3, 4.8, 4.7]
```

```
1 df_new['score_mean'] = score_sum
```

skin1,2,3 합쳐주기

```
1 df_count = pd.read_excel('oliveyoung_skin_count.xlsx')
```

```
1 df_count.head()
```

	Unnamed: 0	num	skin1_noun	skin2_noun	skin3_noun
0	0	A000000157890	[[('건성', 9), ('복합', 1)]]	[[('보습', 9), ('주름', 1), ('미백', 1)]]	[[("자극없이순해요", 9), ("보통이에요", 1)]]
1	1	A000000156235	[[('건성', 5), ('복합', 4), ('지성', 1)]]	[[('보습', 10)]]	[[("자극없이순해요", 10)]]
2	2	A000000141704	[[('복합', 5), ('지성', 4), ('건성', 1)]]	[[('진정', 10)]]	[[("보통이에요", 7), ("자극없이순해요", 3)]]
3	3	A000000016686	[[('건성', 6), ('복합', 4)]]	[[('보습', 10)]]	[[("자극없이순해요", 9), ("자극이느껴져요", 1)]]
4	4	A000000137180	[[('복합', 6), ('건성', 4)]]	[[('진정', 5), ('보습', 5)]]	[[("자극없이순해요", 10)]]

```
1 del df_count['Unnamed: 0']
```

```
1 df_final = pd.merge(df_new, df_count, on='num')
```

3. 분석 프로세스

최종 DF 완성

```
1 df_final.head()
```

	num	name	brand	review	skin1	skin2	skin3	score	review_skin	product_total_review	price	url	score_mean	skin1_noun	skin2_noun	skin3_noun
0	A000000157890	★크림스킨X노르디스크 한정기획★대용량 250ml+100ml(캠핑수건or와인잔)	라네즈	[■복합성 피부 부입니다. 평소 조가울부터 극 건성의로 술시동거는 타입인데요 그래서 ... [주름/미백에 좋아요', '건성'에 좋아 요', '보습'에 좋아 요', '자극 없이 손해 요', '자극 없이 손해 요', '이제 요', '자극 없이 손해 요']	([('원톤', 6), ('잡티', 6), ('트러블', 6), ('건성', 6), ('미백...')])	[[('보습', 9), ('주름', 1), ('미백', 1)]]	[[("자극없이 손해요", 9), ("보통이에 요", 1)]]	5.0, 5.0, 4.0, 5.0, 5.0, 5.0, 5.0, 5.0, 5.0,	((원톤, 6), (잡티, 6), (트러블, 6), (건성, 6), (미 백...	1,211	29,400원	https://www.oliveyoung.co.kr/store/goods/getGo...	4.9	['건성', 9], (['복합', 1))]	[('보습', 9), ('주름', 1), ('미백', 1)]	["자극없이 손해요", 9), ("보통이에 요", 1)]

1. 리뷰 코사인 유사도 벡터화
2. 스킨 코사인 유사도 벡터화
3. 코로나 공식사이트에 셀레니움 활용하여 단계 추출

3. 분석 프로세스

1. input에 제품명 입력하여 유사도가 높은 제품 같이 보여주기

1. 리뷰 코사인 유사도 벡터화

2. 스킨 코사인 유사도 벡터화

3. 코로나 공식사이트에 셀레니움 활용하여 단계 추출

```
1 product_simi_co = (  
2     +review_cosine *1  
3     +skin1_cosine*0.3  
4     +skin2_cosine*0.3  
5     +np.repeat([df_new['score_mean'].values], len(df_new), axis=0)*0.1  
6 )  
7  
8 product_simi_co_sorted_ind = product_simi_co.argsort()[::-1]  
9  
10 def find_simi_product(df, sorted_ind, product_name, top_n=10):  
11  
12     product_title = df[df['name'].str.contains(product_name)]  
13     product_index = product_title.index.values  
14     similar_indexes = sorted_ind[product_index, :(top_n)]  
15     similar_indexes = similar_indexes.reshape(-1)  
16  
17     df_print = df.iloc[similar_indexes]  
18  
19     for i in similar_indexes:  
20         print("-----")  
21         print("\n")  
22         print("✓제품명 : ", df_print["name"][i])  
23         print("✓브랜드 : ", df_print["brand"][i])  
24         print("✓가격 : ", df_print["price"][i])  
25         print("✓별점 : ", df_print["score_mean"][i], " (리뷰수 : ", df_print["product_total_review"][i], "건)")  
26  
27         print("\n♥이런 분들께 추천드려요♥\n")  
28  
29         print("피부타입 : ", df_print['skin1_noun'][i].replace("{", "").replace("}", "").replace("'", "").split(',')[0].split(':')[0].st  
30         print("피부고민 : ", df_print['skin2_noun'][i].replace("{", "").replace("}", "").replace("'", "").split(',')[0].split(':')[0].st  
31         print("자극도 : ", df_print['skin3_noun'][i].replace("{", "").replace("}", "").replace("'", "").replace(' ', "").split(',')[0].sp  
32  
33         print("\n♥이런 분들이 많이 사용하셨어요♥\n")  
34  
35  
36         for j in range(3):  
37             print("{}위 : ".format(j+1), df_print['review_skin'][i][j][0])  
38  
39         print("\n바로 구매해 보세요 ♥ ", df_print["url"][i])  
40         print("\n")  
41  
42     #return df.iloc[similar_indexes]  
43
```


3. 분석 프로세스

2. input(피부고민) text를 리뷰데이터에 포함시켜 비슷한 제품 추천

1. 리뷰 코사인 유사도 벡터화

2. 스킨 코사인 유사도 벡터화

3. 코로나 공식사이트에 셀레니움 활용하여 단계 추출

```
1 place_simi_co = review_cosine
2
3
4
5 # 아래 place_simi_co_sorted_ind 는 그냥 바로 사용하면 됩니다.
6 sorted_ind = place_simi_co.argsort()[::-1]
7
8
9 # 최종 구현 함수
10 def find_simi_review(df, sorted_ind, place_name, top_n=10):
11     nouns = twitter.nouns(place_name)
12     nouns = ' '.join(nouns)
13     df.loc[len(df)] = ['', '', '', nouns, '', '', '', '', '', '', '', '']
14
15     # 리뷰 텍스트 데이터 간의 텍스트 피쳐 벡터라이징
16     count_vect_review = CountVectorizer(min_df=2, ngram_range=(1,2))
17     place_review = count_vect_review.fit_transform(df['review'])
18
19     # 리뷰 텍스트 간의 코사인 유사도 따지기
20     place_simi_review = cosine_similarity(place_review, place_review)
21     sorted_ind = place_simi_review.argsort()[::-1]
22
23
24     similar_indexes = sorted_ind[len(df)-1, :(top_n+1)]
25     # return df.iloc[similar_indexes]
26     df_print = df.iloc[similar_indexes]
27
```

```
28
29 for i in similar_indexes:
30     if i==623:
31         pass
32     else:
33         print("-----")
34         print("\n")
35         print("✓제품명 : ", df_print["name"][i])
36         print("✓브랜드 : ", df_print["brand"][i])
37         print("✓가격 : ", df_print["price"][i])
38         print("\n바로 구매해 보세요 \n ☞ ", df_print["url"][i])
39         print(" (리뷰수 : ", df_print["product_total_review"][i], "건)")
40         print("\n")
41     df.drop(df.tail(1).index, inplace=True)
```

Chapter 4

분석 결과

OLIVE  YOUNG



4. 분석 결과

Input data : 사용하고 있는 제품명

output data : 추천제품 출력

추천 제품명
브랜드
가격
별점

피부타입별 추천

많이 사용한 사람들의 피부타입

제품 링크

```
find_simi_product(df_final, product_simi_co_sorted_ind, input("★ 인생템을 입력해주세요 ★ \n\n▶ "), 3)
```

★ 인생템을 입력해주세요 ★

▶ 노르디스크

✓제품명 : ★크림스킨X노르디스크 한정기획★대용량250ml+100ml(캠핑수건or와인잔)
✓브랜드 : 라네즈
✓가격 : 29,400원
✓별점 : 4.9 (리뷰수 : 1,211 건)

♥이런 분들께 추천드려요♥

피부타입 : 건성
피부고민 : 주름
자극도 : 보통이에요

♥이런 분들이 많이 사용하셨어요♥

1위 : 원톤
2위 : 잡티
3위 : 트러블

바로 구매해 보세요

👉 <https://www.oliveyoung.co.kr/store/goods/getGoodsDetail.do?goodsNo=A000000157890&dispCatNo=100000100010008&trackingCd=Cat100000100010008.Small&curatation&egcode&rccode&egrankcode>

✓제품명 : 라네즈 크림스킨 150ml 기획
✓브랜드 : 라네즈
✓가격 : 28,000원
✓별점 : 4.8 (리뷰수 : 1,449 건)

♥이런 분들께 추천드려요♥

피부타입 : 건성
피부고민 : 보습
자극도 : 자극없이순해요

♥이런 분들이 많이 사용하셨어요♥

1위 : 미백
2위 : 잡티
3위 : 건성

바로 구매해 보세요

👉 <https://www.oliveyoung.co.kr/store/goods/getGoodsDetail.do?goodsNo=A000000147482&dispCatNo=100000100010008&trackingCd=Cat100000100010008.Small&curatation&egcode&rccode&egrankcode>

4. 분석 결과

Input data : 사용하고 있는 제품명

output data : 추천제품 출력

추천 제품명
브랜드
가격
별점

피부타입별 추천

많이 사용한 사람들의 피부타입

제품 링크

```
find_simi_product(df_final, product_simi_co_sorted_ind, input("★ 인생템을 입력해주세요 ★ \n\n▶ "), 3)
```

★ 인생템을 입력해주세요 ★

▶ 노르디스크

✓제품명 : ★크림스킨X노르디스크 한정기획★대용량250ml+100ml(캠핑수건or와인잔)
✓브랜드 : 라네즈
✓가격 : 29,400원
✓별점 : 4.9 (리뷰수 : 1,211 건)

♥이런 분들께 추천드려요♥

피부타입 : 건성
피부고민 : 주름
자극도 : 보통이에요

♥이런 분들이 많이 사용하셨어요♥

1위 : 원톤
2위 : 잡티
3위 : 트러블

바로 구매해 보세요

👉 https://www.oliveyoung.co.kr/store/goods/getGoodsDetail.do?goodsNo=A000000157890&dispCatNo=100000100010008&trackingCd=Cat100000100010008_Small&curatation&egcode&rccode&egrankcode

✓제품명 : 라네즈 크림스킨 150ml 기획
✓브랜드 : 라네즈
✓가격 : 28,000원
✓별점 : 4.8 (리뷰수 : 1,449 건)

♥이런 분들께 추천드려요♥

피부타입 : 건성
피부고민 : 보습
자극도 : 자극없이순해요

♥이런 분들이 많이 사용하셨어요♥

1위 : 미백
2위 : 잡티
3위 : 건성

바로 구매해 보세요

👉 https://www.oliveyoung.co.kr/store/goods/getGoodsDetail.do?goodsNo=A000000147482&dispCatNo=100000100010008&trackingCd=Cat100000100010008_Small&curatation&egcode&rccode&egrankcode

4. 분석 결과

Input data : 사용하고 있는 제품명

output data : 추천제품 출력

추천 제품명
브랜드
가격
별점

피부타입별 추천

많이 사용한 사람들의 피부타입

제품 링크

```
find_simi_product(df_final, product_simi_co_sorted_ind, input("★ 인생템을 입력해주세요 ★ \n\n▶ "), 3)
```

★ 인생템을 입력해주세요 ★

▶ 노르디스크

✓제품명 : ★크림스킨X노르디스크 한정기획★대용량250ml+100ml(캠핑수건or와인잔)
✓브랜드 : 라네즈
✓가격 : 29,400원
✓별점 : 4.9 (리뷰수 : 1,211 건)

♥이런 분들께 추천드려요♥

피부타입 : 건성
피부고민 : 주름
자극도 : 보통이에요

♥이런 분들이 많이 사용하셨어요♥

1위 : 원톤
2위 : 잡티
3위 : 트러블

바로 구매해 보세요

👉 <https://www.oliveyoung.co.kr/store/goods/getGoodsDetail.do?goodsNo=A000000157890&dispCatNo=100000100010008&trackingCd=Cat100000100010008.Small&curatation&egcode&rccode&egrankcode>

✓제품명 : 라네즈 크림스킨 150ml 기획
✓브랜드 : 라네즈
✓가격 : 28,000원
✓별점 : 4.8 (리뷰수 : 1,449 건)

♥이런 분들께 추천드려요♥

피부타입 : 건성
피부고민 : 보습
자극도 : 자극없이순해요

♥이런 분들이 많이 사용하셨어요♥

1위 : 미백
2위 : 잡티
3위 : 건성

바로 구매해 보세요

👉 <https://www.oliveyoung.co.kr/store/goods/getGoodsDetail.do?goodsNo=A000000147482&dispCatNo=100000100010008&trackingCd=Cat100000100010008.Small&curatation&egcode&rccode&egrankcode>

4. 분석 결과

Input data : 피부고민 관련 내용

ex) 요새 야근을 많이해서 그런지 여드름이 다시 부쩍
많아져서 고민이에요 ☹

output data : 추천제품 출력

ex) 제품명 : 프리메라 맨오가니언스 트리트먼트
브랜드 : 프리메라
가격 : 30,000원

구매링크 : URL

```
In [13]: find_simi_place(df,sorted_ind, input(), 6)
```

요새 야근을 많이해서 그런지 여드름이 다시 부쩍 많아져서 고민이에요 ☹

✓제품명 : 프리메라 맨오가니언스 트리트먼트 워터 180ML
✓브랜드 : 프리메라
✓가격 : 30,000원

바로 구매해 보세요

➡ https://www.oliveyoung.co.kr/store/goods/getGoodsDetail.do?goodsNo=A000000138008&dispCatNo=100000100010008&trackingCd=Cat100000100010008_Small&curation&egcode&rccode&egrankcode
(리뷰수 : 16 건)

✓제품명 : 알랑 인트랄 액티브 스킨클라잉 로션 100ml
✓브랜드 : 알랑
✓가격 : 95,000원

바로 구매해 보세요

➡ https://www.oliveyoung.co.kr/store/goods/getGoodsDetail.do?goodsNo=A000000154626&dispCatNo=100000100010008&trackingCd=Cat100000100010008_Small&curation&egcode&rccode&egrankcode
(리뷰수 : 1 건)

✓제품명 : 막단자로부터 컨트롤레이 티트리먼트 모이스처라이저 100ml (대용량 티트리 약산성 보습크림)
✓브랜드 : 막단자로부터
✓가격 : 43,000원

바로 구매해 보세요

➡ https://www.oliveyoung.co.kr/store/goods/getGoodsDetail.do?goodsNo=A000000128322&dispCatNo=100000100010008&trackingCd=Cat100000100010008_Small&curation&egcode&rccode&egrankcode
(리뷰수 : 201 건)

✓제품명 : 막단자로부터 컨트롤레이 티트리먼트 토너 120ml (티트리 약산성 토너)
✓브랜드 : 막단자로부터
✓가격 : 19,000원

바로 구매해 보세요

➡ https://www.oliveyoung.co.kr/store/goods/getGoodsDetail.do?goodsNo=A000000128107&dispCatNo=100000100010008&trackingCd=Cat100000100010008_Small&curation&egcode&rccode&egrankcode
(리뷰수 : 278 건)

✓제품명 : 더바디산 티트리 오일 10ML
✓브랜드 : 더바디산
✓가격 : 13,500원

바로 구매해 보세요

➡ https://www.oliveyoung.co.kr/store/goods/getGoodsDetail.do?goodsNo=A000000138982&dispCatNo=100000100010008&trackingCd=Cat100000100010008_Small&curation&egcode&rccode&egrankcode
(리뷰수 : 151 건)

4. 분석 결과

Input data : 피부고민 관련 내용

ex) 요새 야근을 많이해서 그런지 여드름이 다시 부쩍
많아져서 고민이에요 ☹

output data : 추천제품 출력

ex) 제품명 : 프리메라 맨오가니언스 트리트먼트
브랜드 : 프리메라
가격 : 30,000원

구매링크 : URL

```
In [13]: find_simi_place(df,sorted_ind, input(), 6)
```

요새 야근을 많이해서 그런지 여드름이 다시 부쩍 많아져서 고민이에요 ☹

✓제품명 : 올망 인트랄 엑티브 스킨클레이징 로션 100ml
✓브랜드 : 올망
✓가격 : 95,000원

바로 구매해 보세요

➡ https://www.oliveyoung.co.kr/store/goods/getGoodsDetail.do?goodsNo=A000000154626&dispCatNo=100000100010008&trackingCd=Cat100000100010008_Small&curation&egcode&rccode&egrancode
(리뷰수 : 1 건)

✓제품명 : 약당자료를 컨도플에이 티트리먼트 모이스처라이저 100ml (대용량 티트리 약산성 보습크림)
✓브랜드 : 약당자료를
✓가격 : 43,000원

바로 구매해 보세요

➡ https://www.oliveyoung.co.kr/store/goods/getGoodsDetail.do?goodsNo=A000000128322&dispCatNo=100000100010008&trackingCd=Cat100000100010008_Small&curation&egcode&rccode&egrancode
(리뷰수 : 201 건)

✓제품명 : 약당자료를 컨도플에이 티트리먼트 토너 120ml (티트리 약산성 토너)
✓브랜드 : 약당자료를
✓가격 : 19,000원

바로 구매해 보세요

➡ https://www.oliveyoung.co.kr/store/goods/getGoodsDetail.do?goodsNo=A000000128107&dispCatNo=100000100010008&trackingCd=Cat100000100010008_Small&curation&egcode&rccode&egrancode
(리뷰수 : 278 건)

✓제품명 : 더바디산 티트리 오일 10ML
✓브랜드 : 더바디산
✓가격 : 13,500원

바로 구매해 보세요

➡ https://www.oliveyoung.co.kr/store/goods/getGoodsDetail.do?goodsNo=A000000138982&dispCatNo=100000100010008&trackingCd=Cat100000100010008_Small&curation&egcode&rccode&egrancode
(리뷰수 : 151 건)

4. 분석 결과

Input data : 피부고민 관련 내용

ex) 요새 야근을 많이해서 그런지 여드름이
많아져서 고민이에요 □

output data : 추천제품 출력

ex) 제품명 : 프리메라 맨오가니언스 트리트먼트
브랜드 : 프리메라
가격 : 30,000원

구매링크 : URL

```
In [13]: find_simi_place(df,sorted_ind, input(), 6)
```

요새 야근을 많이해서 그런지 여드름이 다시 부쩍 많아져서 고민이에요 ♪

✓제품명 : 프리메라 맨오가니언스 트리트먼트 워터 180ML
✓브랜드 : 프리메라
✓가격 : 30,000원

바로 구매해 보세요

➡ https://www.oliveyoung.co.kr/store/goods/getGoodsDetail.do?goodsNo=A000000138008&dispCatNo=100000100010008&trackingCd=Cat100000100010008_Small&curation&egcode&rccode&egrankcode
(리뷰수 : 16 건)

✓제품명 : 프리메라 맨오가니언스 트리트먼트 워터 180ML
✓브랜드 : 프리메라
✓가격 : 30,000원

바로 구매해 보세요

➡ https://www.oliveyoung.co.kr/store/goods/getGoodsDetail.do?goodsNo=A000000138008&dispCatNo=100000100010008&trackingCd=Cat100000100010008_Small&curation&egcode&rccode&egrankcode
(리뷰수 : 16 건)

✓제품명 : 닥터자료를 컨트론헤이 티트리먼트 토너 120ml (티트리 약산성 토너)
✓브랜드 : 닥터자료들
✓가격 : 43,000원

바로 구매해 보세요

➡ https://www.oliveyoung.co.kr/store/goods/getGoodsDetail.do?goodsNo=A000000128322&dispCatNo=100000100010008&trackingCd=Cat100000100010008_Small&curation&egcode&rccode&egrankcode
(리뷰수 : 201 건)

✓제품명 : 닥터자료들 컨트론헤이 티트리먼트 토너 120ml (티트리 약산성 토너)
✓브랜드 : 닥터자료들
✓가격 : 19,000원

바로 구매해 보세요

➡ https://www.oliveyoung.co.kr/store/goods/getGoodsDetail.do?goodsNo=A000000128107&dispCatNo=100000100010008&trackingCd=Cat100000100010008_Small&curation&egcode&rccode&egrankcode
(리뷰수 : 278 건)

✓제품명 : 더바디샵 티트리 오일 10ML
✓브랜드 : 더바디샵
✓가격 : 13,500원

바로 구매해 보세요

➡ https://www.oliveyoung.co.kr/store/goods/getGoodsDetail.do?goodsNo=A000000138982&dispCatNo=100000100010008&trackingCd=Cat100000100010008_Small&curation&egcode&rccode&egrankcode
(리뷰수 : 151 건)

향후 보완 과제

■ 보완 할 것들

코사인 유사도 뿐만 아니라, 클러스터링등 머신러닝 모델을 이용해 추천 시스템을 만들어보면 더 좋을 것 같다.

전체리뷰수도 가중치를 주어 추천모델을 보완 필요

가장 합리적인 가중치 값을 찾는 방법에 대해 추가적인 고민 필요

더 많은 리뷰 데이터를 이용해 신뢰도 향상 필요
- (제품별 리뷰수 최대 10개만을 이용)

댓글 알바와 실제 내돈내산 리뷰의 분류를 통해 리뷰데이터의 신뢰성을 높이는것 필요

출력값을 gui를 활용하여 시각화 필요

참고 사이트 및 출처

1. 올리브영 온라인 쇼핑몰

<https://www.oliveyoung.co.kr/store/main/main.do?oy=0>

Thank You



Q&A

1. 감사합니다