K-Data Team Project

유사색상 기반 립스틱 추천시스템

전상기, 이지수, 윤지성 , 표정호, 김성중

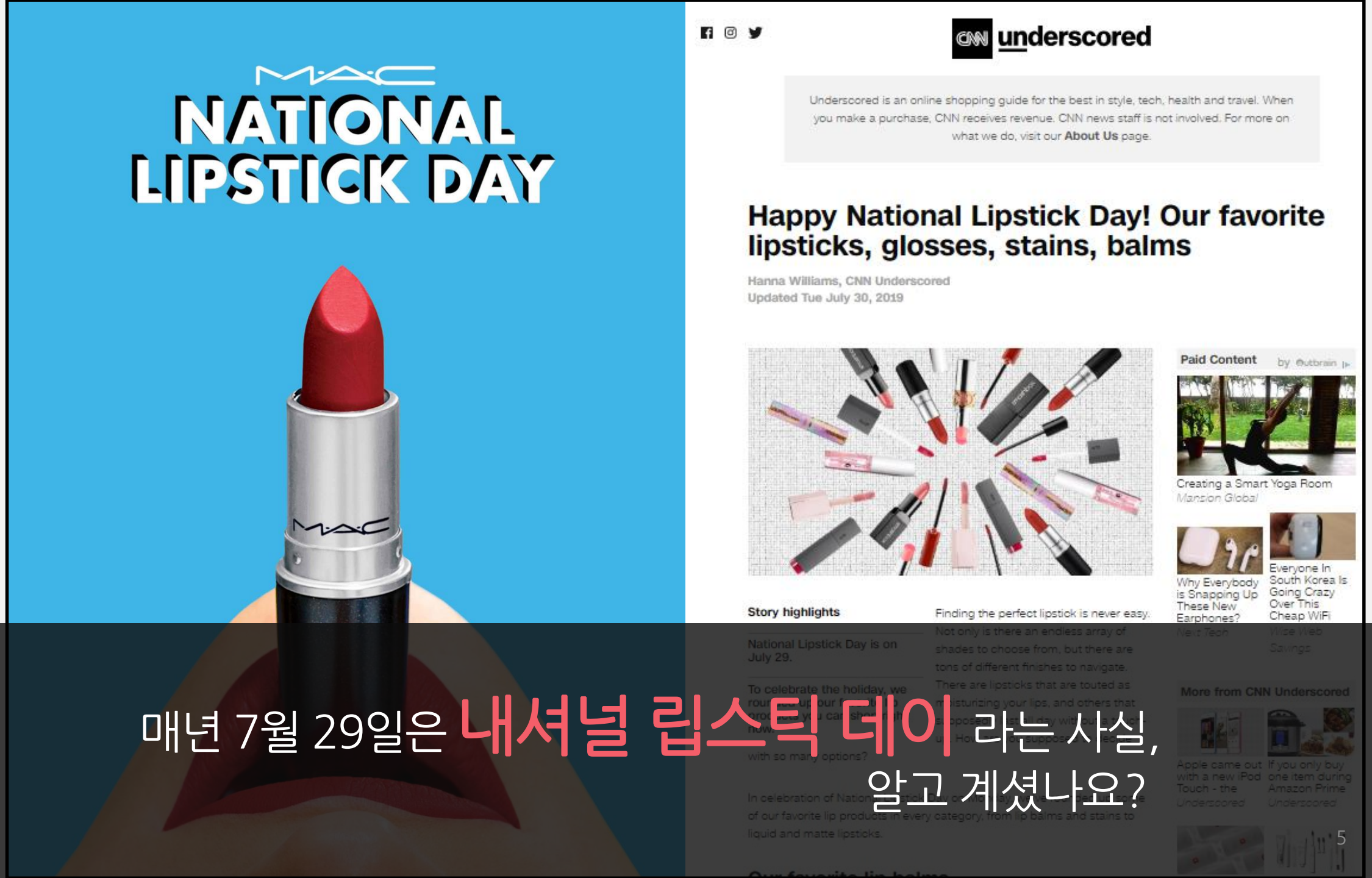
"The most beautiful makeup for a woman is passion. But cosmetics are easier to buy."

– Yves Saint Laurent

K - Data Team Project

CONTENTS

01 K - Data Team Project
주제 선정 배경



매년 7월 29일은 **내셔널 립스틱 데이** 라는 사실,
알고 계셨나요?

주제 선정 배경

립스틱 선택의 어려움

: 비슷한 색상도 브랜드 간 **표현 방식**이 다름

: 같은 색상일지라도 브랜드에 따라 **가격**의 차이가 존재함

: 자신이 사용하는 색과 유사한 색의 저가 제품을 찾기 어려움

: 원하는 색상을 찾는다고 해도
어떤 브랜드에 있는지 파악하고 방문하는데 **시간낭비**가 발생함

주제 선정 배경

MAC 매트 립스틱

RGB (225, 44, 48)

레이디 데인저
비비드한 코랄 레드
[매트]

NARS 래비싱레드

RGB (219, 48, 41)

NEW
래비싱 레드
브라이트 핑크 코랄

동일하거나 비슷한 색상이라도
회사마다 표현 방법이 다름 » **RGB 값**을 분석하여 유사도 파악

브랜드 간 유사색상의 립스틱 가격 차이 존재

-> 가격 비교를 통한 **경제적 소비 가능**

주제 선정 배경

시간 단축 및 단순화

매장 방문 시간 단축

방문할 매장을
찾는 시간 단축

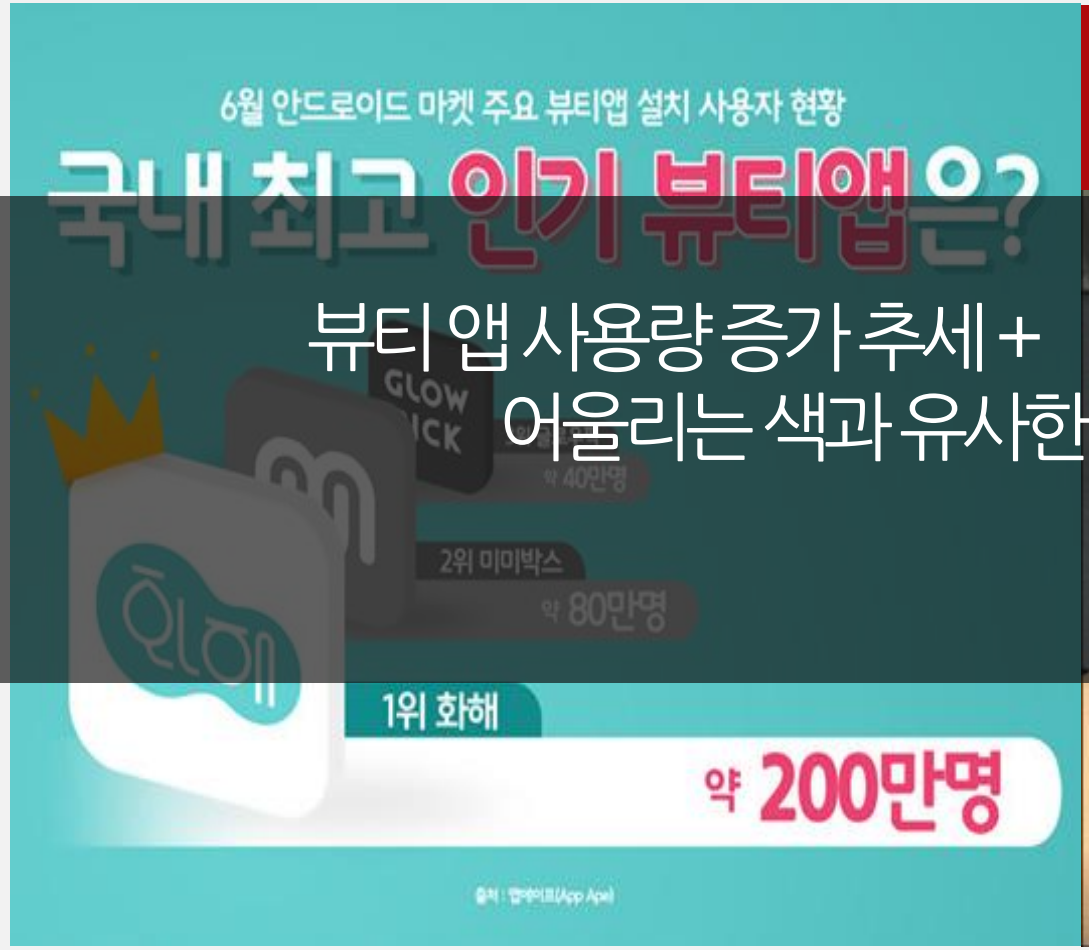
탐색 범위 감소

검색 결과의
제품만 test

검색의 단순화

RGB 자체를
기준으로 검색

주제 선정 배경



뷰티 앱 사용량 증가 추세 +
어울리는 색과 유사한 색상을 찾는 심리적 요인
-> **시장 경쟁력**

02

K - Data Team Project

데이터 수집 / 전처리

주요 사용 모듈

BeautifulSoup

HTML 데이터 추출

제품정보 및
리뷰 수집에 사용

MySQL

수집한 데이터를
DB에 저장,
이후 활용

Selenium

브라우저를 조정하여
웹을 테스트하는 용도
(원래 기능)

자바 스크립트 코드를
처리하는 용도 or
버튼 클릭 기능

한글 형태소 분석기
-> Twitter

Word Cloud

데이터 수집 및 전처리

데이터 수집 및 전처리

데이터 수집

제품정보 (1658개)
: Brand, Product, Color, Price, Image, RGB 저장
(1658개, 20개 사 이상)

리뷰 (13331개)
: Brand, Product, Review 저장

키워드 4개
: 발색력, 발림성, 지속력, 수분감

+

전처리

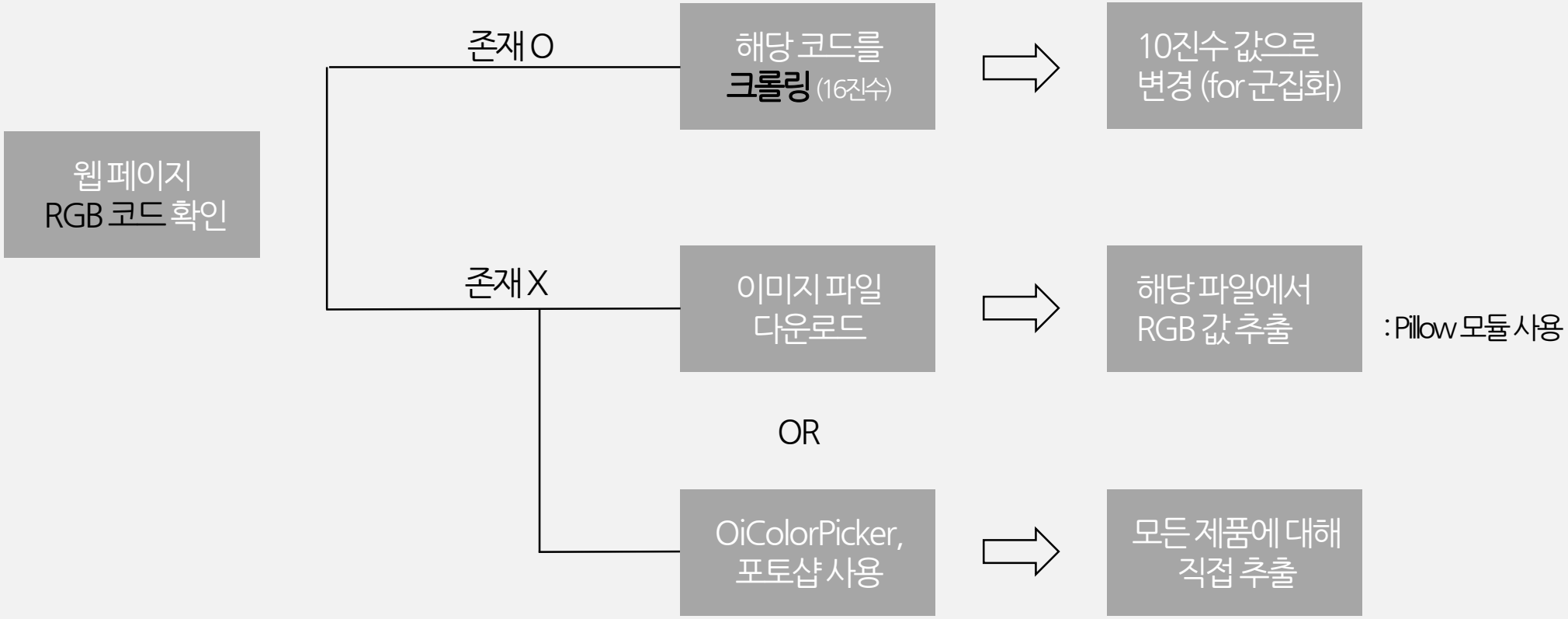
① **제품 정보**

- 군집화를 위한 **RGB** 값 추출
- **결측 값** 처리(수정)

② **리뷰 데이터**

- **중복 리뷰** 제거
- 리뷰 데이터 처리

RGB 값 구하기


웹 페이지
RGB 코드 확인
존재 O
해당 코드를
크롤링 (16진수)
10진수 값으로
변경 (for 군집화)
존재 X
이미지 파일
다운로드
해당 파일에서
RGB 값 추출
: Pillow 모듈 사용
OR
OiColorPicker,
포토샵 사용
모든 제품에 대해
직접 추출

02 제품 정보 전처리

RGB 값 구하기

```
true,"sppShort":true}" data-product-id="PROD60284" data-misc-flag=
"0" data-pr-product-id="PROD60284" data-inv-status="1">
  ▼<div class="product__detail">
    ▶<script type="application/ld+json">…</script>
    ▼<div class="product__sku-details" id="product-detail-attach">
      ▶<header class="product__header">…</header>
      ▶<div class="product__images js-shade-trigger" style="width:
      100%;">…</div>
      ▼<footer class="product__footer">
        ▶<div class="view-all-shades view-all-shades--mobile hide-
        xlarge-up js-trigger-mobile-shade-selector">…</div>
        ▼<div class="product__product-details-shade">
          ▼<div class="product__product-details-shade-smoosh shade-
          picker__color-wrapper">
            ▼<a class="js-product__link-to-spp js-trigger-mobile-
            shade-selector" href="/product/13854/60284/makeup/powder-
            kiss-lipstick#/shade/디보티드_투_칠리">
              <div class="shade-picker__color-texture  lazyloaded"
              data-bg="/media/export/cms/products/smoosh/
              mac_smoosh_S4K031.jpg" style="background-color:
              rgb(177, 60, 50); background-image: url("/media/
              export/cms/products/smoosh/mac_smoosh_S4K031.jpg");">
              </div> == $0
            </a>
          </div>
          <div class="product__product-details-shade-name" style=
          "color:#b13c32;">디보티드 투 칠리</div>
```

RGB 코드가 존재하면

값을 DataFrame에 저장

```
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.image as mpimg
from PIL import Image

# 샘플 그림을 그립시다.
plt.style.use("default")
jpg_img_arr = mpimg.imread('/Users/frhyme/Downloads/google2.0.0.
jpg_IMG = Image.open('/Users/frhyme/Downloads/google2.0.0.jpg')
print(type(jpg_img_arr))# 얘는 np.array
print(type(jpg_IMG))# 얘는 PIL.JpegImagePlugin.JpegImageFile' 오
print((jpg_img_arr == np.array(jpg_IMG)).mean())# 다행히 np.arra

height, width, layer = jpg_img_arr.shape

f, axes = plt.subplots(2, 2, figsize=(8, 8*height/width))
## original img plotting
axes[0][0].imshow(jpg_img_arr[:, :, :]), axes[0][0].axis('off')
axes[0][0].set_xticks([]), axes[0][0].set_yticks([])# 이걸 하지
# Red, Green, Blue로 구분하여 표현. colormap 또한, 그 형식에 맞추
# 실제 그림을 보면 색깔별로 어느 정도 구분되어 있는 것을 알 수 있
cmaps = [plt.cm.Reds, plt.cm.Greens, plt.cm.Blues]
for i in range(1, 4):
```

값이 없으면

Pillow 모듈로 직접 구함

02 제품 정보 전처리

결측 값 / 부가 정보

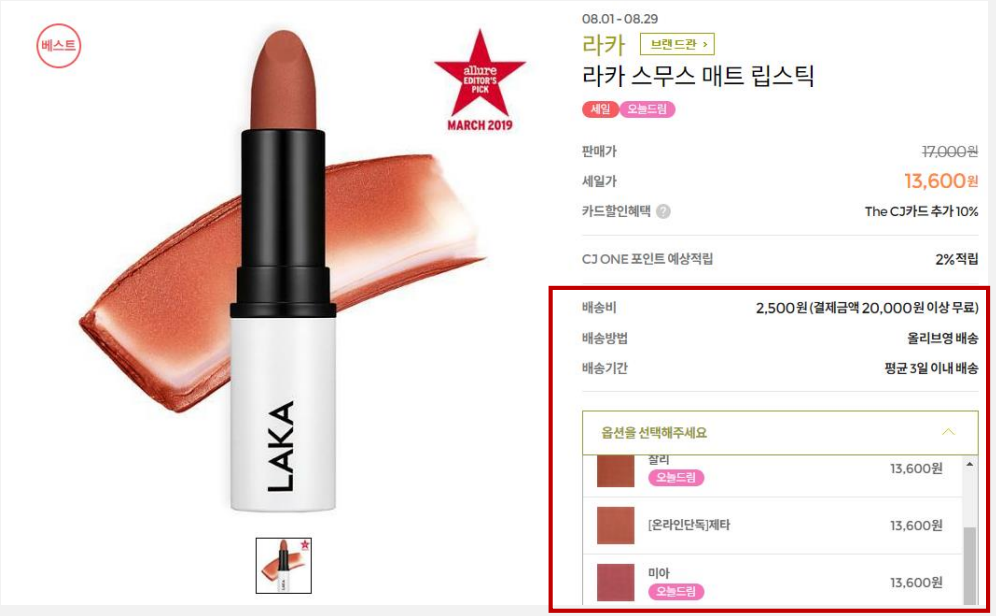
결측 값 처리

11,000	https://images-kr	202	38	29	ca261d
11,000	https://images-kr	220	74	75	dc4a4b
11,000	https://images-kr	219	15	27	db0f1b
11,000	https://images-kr	186	17	22	ba1116
11,000	https://images-kr	179	1	35	b30123
11,000	https://images-kr				
11,000	https://images-kr	177	1	50	b10132
11,000	https://images-kr	143	2	37	8f0225
11,000	https://images-kr	168	72	60	a8483c
11,000	https://images-kr	142	37	33	8e2521
11,000	https://images-kr	131	50	57	833239
11,000	https://images-kr	180	87	70	b45746
11,000	https://images-kr	192	110	116	c06e74
11,001	https://images-kr	132	36	37	842425
11,000	https://images-kr	174	55	47	ae372f
11,000	https://images-kr	196	58	45	c43a2d

RGB 값이 누락되어 있을 경우

Pillow or **수작업**으로 처리

부가 정보 제거



배송정보, 품절 상품, 행사 태그 등

불필요한 정보 제거

02 리뷰 전처리

리뷰 데이터

중복 리뷰 제거

★★★★★
jepy*** | 민감성
2019.07.22
0

08 러브 송
립스틱 색이 이쁘네요~~~ !!!!! 발색도 좋고 지속력도 좋아요!! ㅎ
구매할게요~~ 감사합니다!!!! 많이 파세요~~~~~^^

★★★★★
jepy*** | 민감성
2019.07.22
0

05 그레이스
립스틱 색이 이쁘네요~~~ !!!!! 발색도 좋고 지속력도 좋아요!! ㅎ
구매할게요~~ 감사합니다!!!! 많이 파세요~~~~~^^

데이터의 신뢰성을 위해 **삭제**

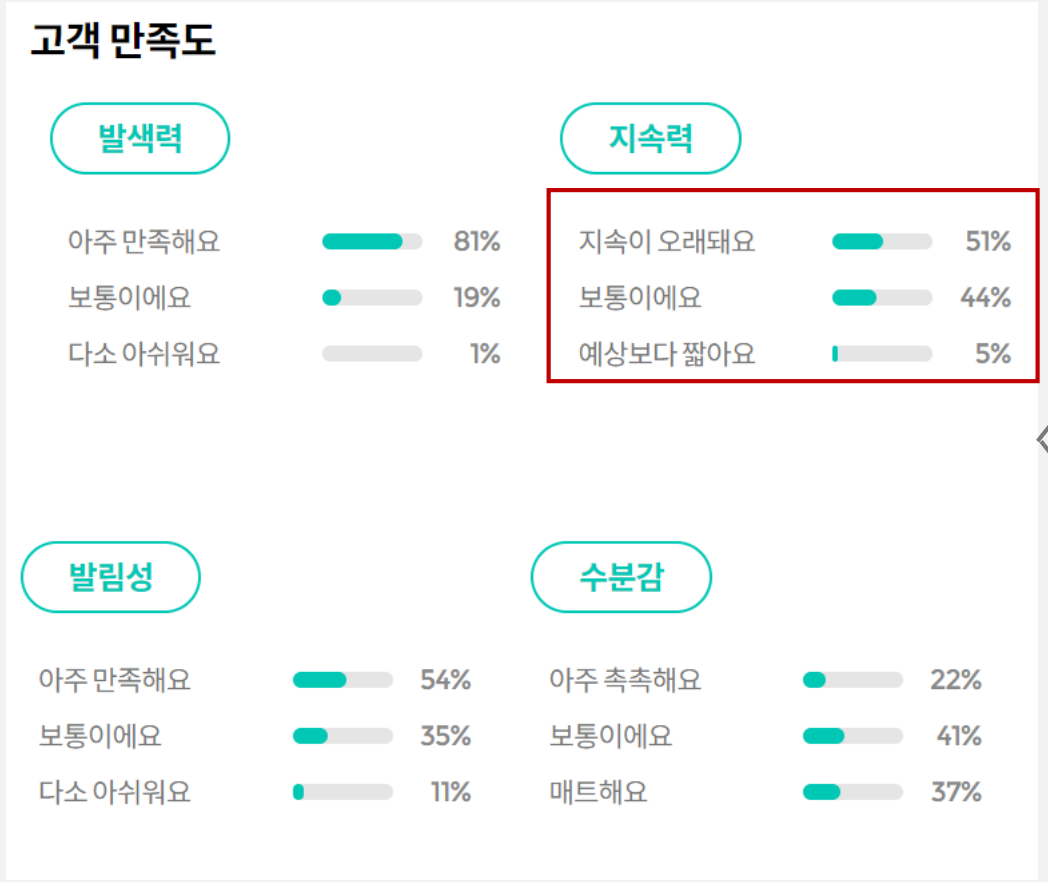
명사, 형용사 추출

```
[('지속', 'Noun'), ('력', 'Suffix'), ('이', 'Josa'), ('조금', 'Noun'),
iner'), ('외', 'Noun'), ('에는', 'Josa'), ('만족합니다', 'Adjective'),
게', 'Adjective'), ('잘', 'Verb'), ('구매', 'Noun'), ('했', 'Verb'), ('
--------------------------------
[('얼마', 'Noun'), ('전', 'Noun'), ('에', 'Josa'), ('오프라인', 'Noun')
순', 'Noun'), ('으로', 'Josa'), ('미니', 'Noun'), ('사이즈', 'Noun'), (
uffix'), ('이', 'Josa'), ('있었는데', 'Adjective'), ('별로', 'Noun'), (
a'), ('써', 'Verb'), ('봤는데', 'Verb'), ('너무', 'Adverb'), ('좋은', '
ticle'), ('저', 'Noun'), ('는', 'Josa'), ('입술', 'Noun'), ('이', 'Josa
a'), ('안', 'VerbPrefix'), ('쓰거든요', 'Verb'), ('ㅠㅠ', 'KoreanPartic
('바르니까', 'Verb'), ('자연', 'Noun'), ('스럽게', 'Josa'), ('발색', 'N
('도', 'Josa'), ('없고', 'Adjective'), ('지속', 'Noun'), ('도', 'Josa')
b'), ('파세요', 'Verb'), ('ㅋㅋㅋㅋㅋ', 'KoreanParticle')]
--------------------------------
[('예상', 'Noun'), ('했던', 'Verb'), ('색상', 'Noun'), ('은', 'Josa'),
('들어요', 'Verb'), ('.^^', 'Punctuation')]
```

단어 별 개수 count 후, **불용어** 삭제
(2000 여개의 단어 대상)

02 리뷰 전처리

키워드 수집



고객 만족도 항목 O

4개 키워드 (발색력, 발림성, 지속력, 수분감)
평가 항목을 기준으로 -1, 0, 1로
수치화 하여 저장

고객 만족도 항목 X

리뷰 사이트 '**언니들의 파우치**',
맥 **리뷰 스냅샷** 등 참고

키워드 수집

리뷰 스냅샷

이 제품을 추천하는 이유

자연스러운 메이크업 연출 (2)
강한 컬러 (2)
오래 지속되는 (2)
컬러 그대로 발색 (2)
크리미한 (2)

더 많이

www.maccosmetics.co.kr - 매트 립스틱

라카(LAKA)

스무스 매트 립스틱 [로이]

★ 4.0

가격 17,000원 용량 3.5 g

장점 매트립치고 굉장히 부드럽게 입술주름을 메꿔주고 일단 컬러가 옴총나게 예쁩니다!!!

단점 아무래도 매트립인만큼 각질부각이 살짝 있어요!

추천 일반 매트립보다는 각질부각이 덜하고 스머징했을 때 더 예쁜 립스틱이에요 :)

www.unpa.me - 라카 스무스 매트 립스틱

03 활용 모델

활용 모델 / 분석 결과

활용 모델

활용모델

K - means

유클리드 거리 사용
: 맨해튼, 민코우스키 등의 거리보다 적합

Elbow Method
: 적절한 K값 계산

K=4로 1차 군집화
: 특정군집 내의 데이터를 추천의 기준으로 사용함

KNN

Pearson 상관계수 사용

반복적인 시행을 통해 적절한 **K** 값 도출

동률 제거를 위해 K 값을 **홀수**로 설정

03 활용 모델
K - means

DataFrame

Color	Price	Image	R	G	B	RGB
Fuchsia/Orange 107	50,000	https://www.yslbeautykr.com/dw/image/v2/	234	65	22	ea4116
Red/Blue 105	50,000	https://www.yslbeautykr.com/dw/image/v2/	233	62	71	e93e47
Green/Blue 106	50,000	https://www.yslbeautykr.com/dw/image/v2/	221	44	34	dd2c22
104 쥬 다트락시옹 (새틴 핫 레드)	44,000	https://www.yslbeautykr.com/dw/image/v2/	189	23	25	bd1719
101 메이크 잇 번 (샤인 퓨어 레드)	44,000	https://www.yslbeautykr.com/dw/image/v2/	193	7	46	c1072e
102 레디 투 시듀스	44,000	https://www.yslbeautykr.com/dw/image/v2/	173	43	45	ad2b2d
04 루즈 베르밀리옹	44,000	https://www.yslbeautykr.com/dw/image/v2/	184	52	75	b8344b
17 로즈 다일라	44,000	https://www.yslbeautykr.com/dw/image/v2/	234	94	107	ea5e6b
13 르 오랑지	44,000	https://www.yslbeautykr.com/dw/image/v2/	213	29	55	d51d37
19 푸시아	44,000	https://www.yslbeautykr.com/dw/image/v2/	161	24	96	a11860
36 코랄 레장드	44,000	https://www.yslbeautykr.com/dw/image/v2/	235	105	92	eb695c
49 로즈 트로피칼	44,000	https://www.yslbeautykr.com/dw/image/v2/	202	102	174	ca66ae
51 코랄 어바인	44,000	https://www.yslbeautykr.com/dw/image/v2/	235	85	71	eb5547
52 루쥬 로즈	44,000	https://www.yslbeautykr.com/dw/image/v2/	246	98	96	f66260
70 르 뉘	44,000	https://www.yslbeautykr.com/dw/image/v2/	203	132	123	cb847b
72 루쥬 바이닐	44,000	https://www.yslbeautykr.com/dw/image/v2/	146	23	41	921729
74 오렌지 엘렉트로	44,000	https://www.yslbeautykr.com/dw/image/v2/	229	35	23	e52317
82 루쥬 프로보케이션	44,000	https://www.yslbeautykr.com/dw/image/v2/	194	14	62	c20e3e
01 매드 누드	45,000	https://www.yslbeautykr.com/dw/image/v2/	216	78	91	d84e5b
02 대즐링 푸시아	45,000	https://www.yslbeautykr.com/dw/image/v2/	225	53	109	e1356d
04 익스포징 코랄	45,000	https://www.yslbeautykr.com/dw/image/v2/	229	30	53	e51e35
05 딜리리어스 오랑쥬	45,000	https://www.yslbeautykr.com/dw/image/v2/	227	2	20	e30214
06 루나틱 레드	45,000	https://www.yslbeautykr.com/dw/image/v2/	175	17	32	af1120
43 로즈 리브 고쉬	44,000	https://www.yslbeautykr.com/dw/image/v2/	208	60	74	d03c4a
45 루쥬 턱시도	44,000	https://www.yslbeautykr.com/dw/image/v2/	195	15	42	c30f2a

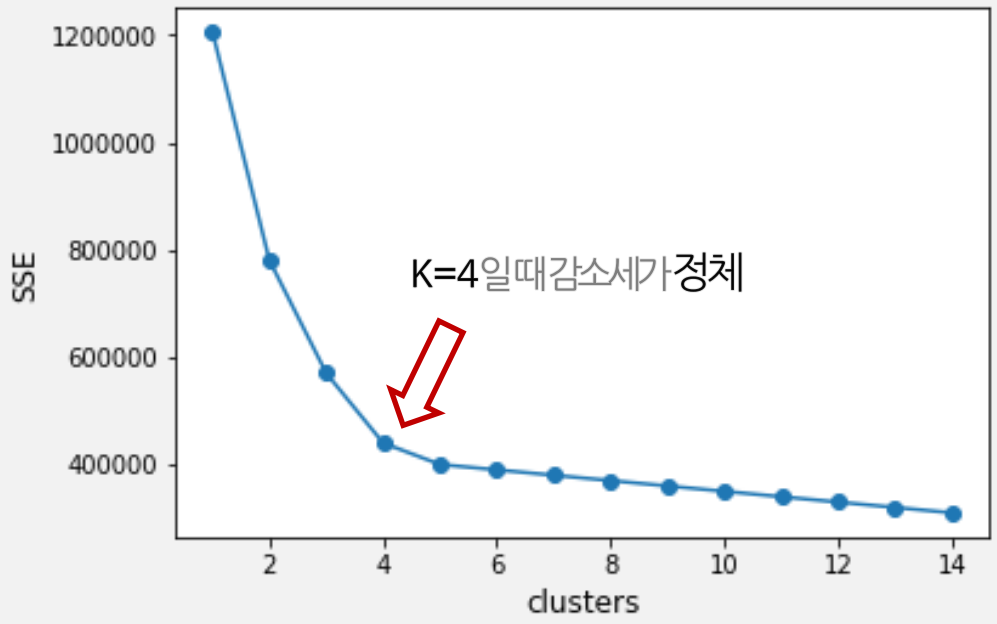
① 수집한 RGB 값을
각각 R, G, B로 **분할**하여
새로운 Column에 저장

② 분할한 값을 List에 저장(2중 list)
⇨ np.array 형식으로 **변환**

③ K-means를 반복 시행하여
적절한 K값 찾기

K - means

적절한 K 값 찾기: Elbow method

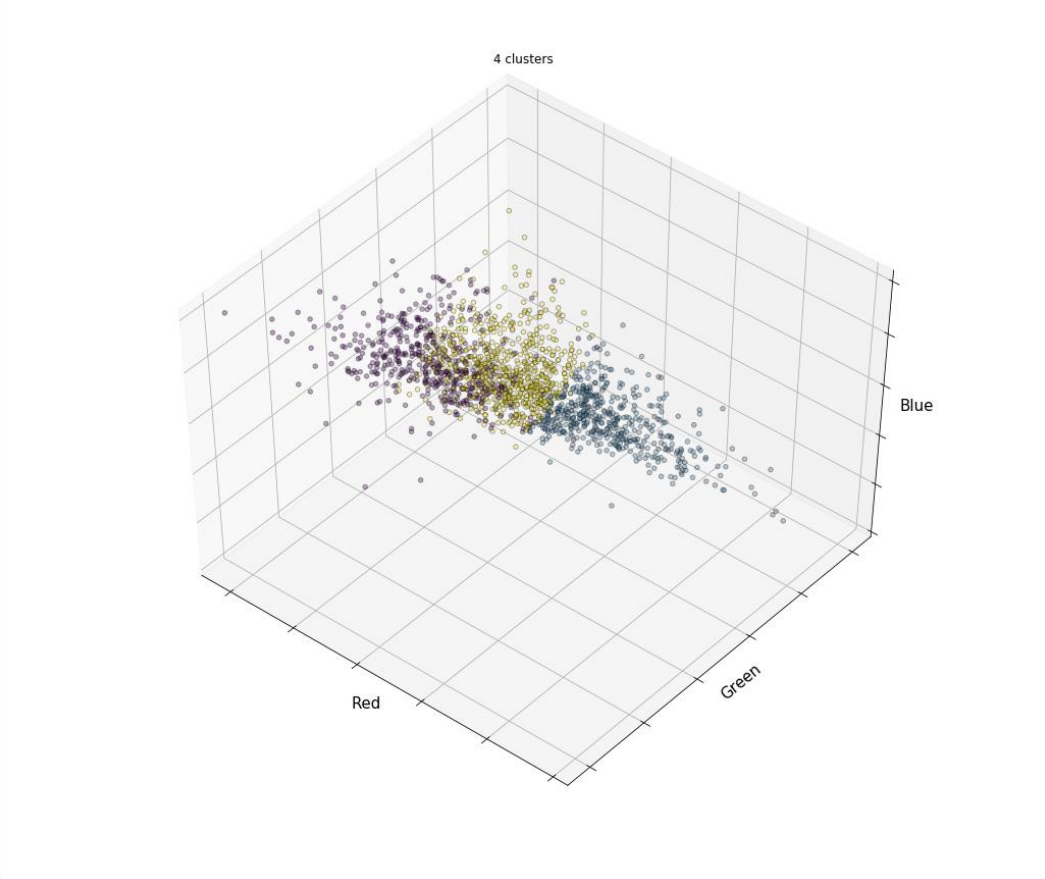


“ Elbow Method 란? ”

- 클러스터의 수(K) 를 순차적으로 늘려가며 기울기가 **완만해지는 곳**을 찾음
- K의 값을 **지정**해줘야 하는 K - means의 한계점을 Trial and Error로 보완
- 오차제곱합(SSE)과 클러스터의 개수를 기준으로 가장 **적절한 군집의 수**를 평가

K - means

K - means (첫번째 군집화)



K-means를 통해 첫번째 군집화

총 1658개 데이터를 4개의 군집으로 분류 (K=4)

⇨ 이후 K=6으로 두번째 군집화 (feat. Elbow method)

입력되는 순서에 따라 군집이 변하는 DBSCAN보다
클러스터링에 적합하다고 판단

03 활용 모델
Cosine Similarity

```
def get_recommendations(title, cosine_sim=cosine_sim):
    # 선택한 립스틱의 타이틀로부터 해당되는 인덱스를 받아옴
    idx = indices[title]

    # 모든 립스틱에 대해서 해당 립스틱과의 유사도를 구합니다.
    sim_scores = list(enumerate(cosine_sim[idx]))

    # 유사도에 따라 립스틱들을 정렬합니다.
    sim_scores = sorted(sim_scores, key=lambda x: x[1], reverse=True)

    # 가장 유사한 7개의 립스틱을 받아옵니다.
    sim_scores = sim_scores[1:8]

    # 가장 유사한 7개의 립스틱의 인덱스를 받아옵니다.
    movie_indices = [i[0] for i in sim_scores]

    # 가장 유사한 7개의 립스틱의 제목을 리턴합니다.
    return df.iloc[movie_indices]

df = pd.read_excel('./products.xls', encoding='utf-8')

tmp = []
for a, b, c, d, e in zip(df['Price'],df['colorpower'],df['spread'],df['ke
    tmp.append([a,b,c,d,e ])
cosine_sim = cosine_similarity(tmp)

indices = pd.Series(df.index, index=df['Color'])

get_recommendations('Fuchsia/Orange 107')
```

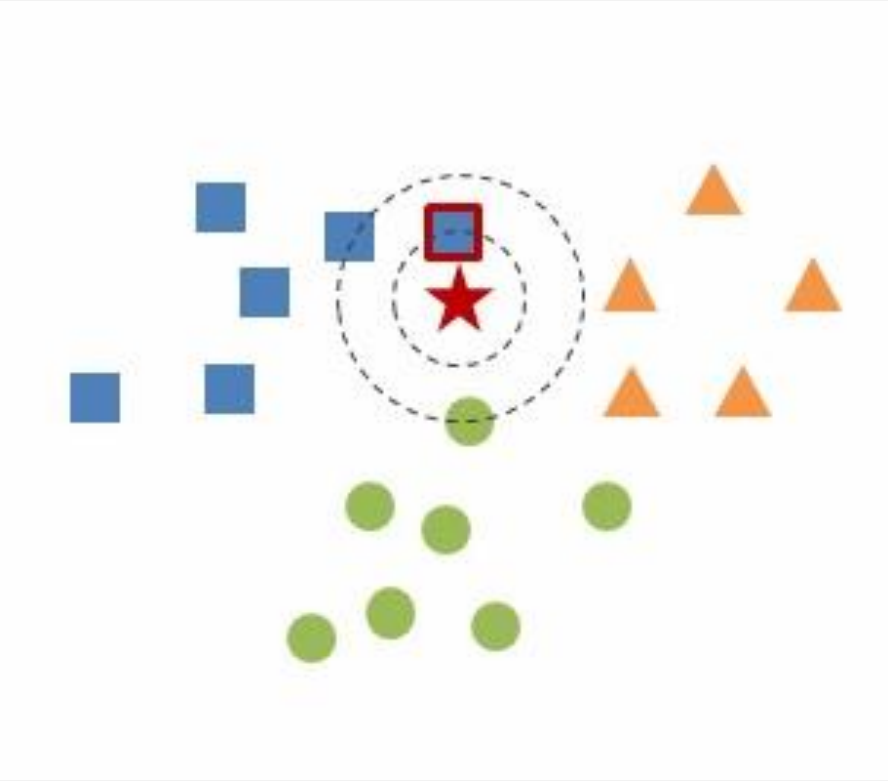
① 같은 군집 내의 립스틱들을 데이터베이스에서 가져옴

② 유사도 계산 (Cosine Similarity)
* Feature: 키워드 4개 + RGB
(발색력, 지속력, 발림성, 수분감)

③ 가장 유사한 7개 립스틱의 정보를 받아 옴

KNN

KNN (K-nearest neighbor)

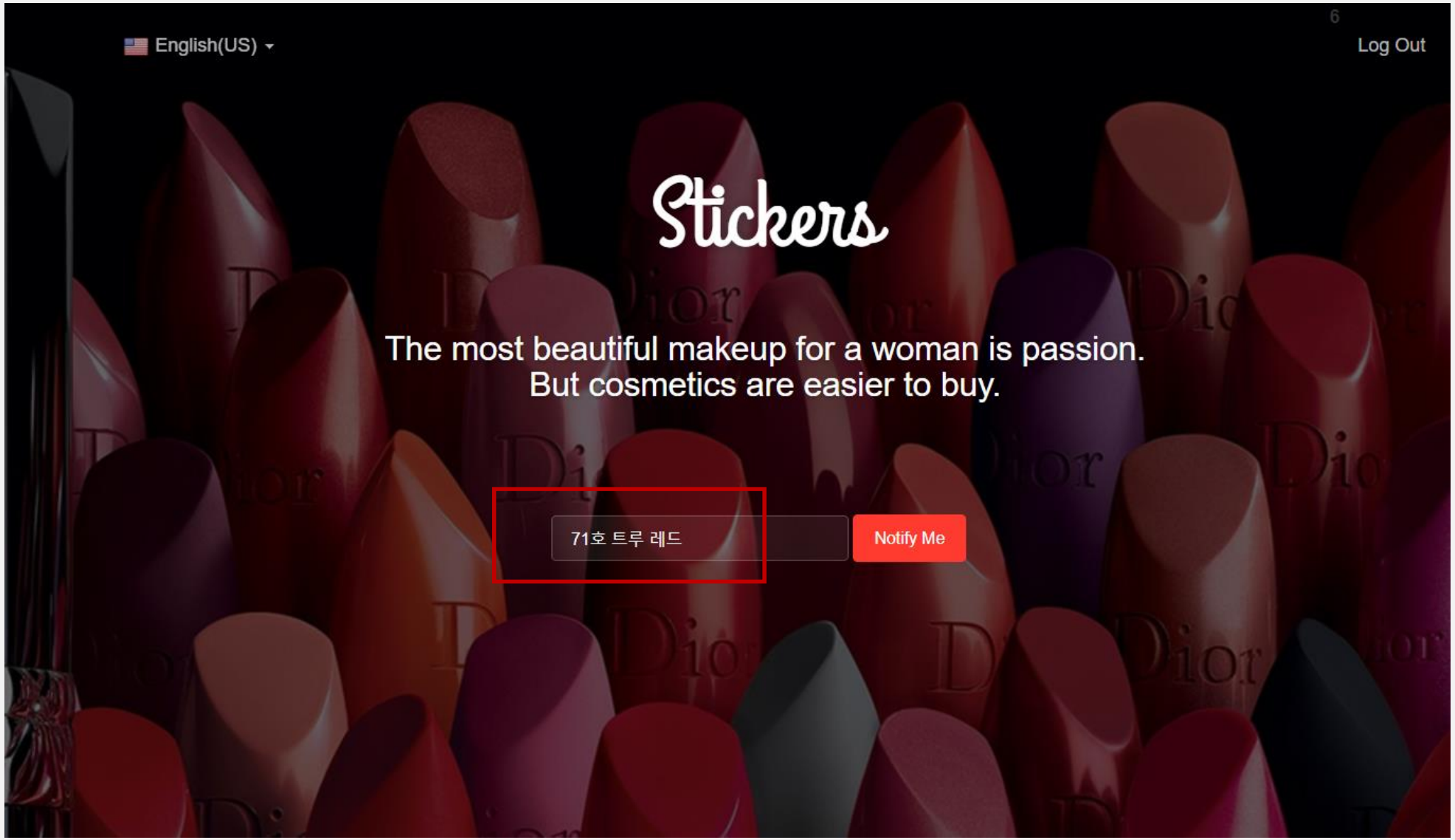


- 사용자 기반 추천 (개인화)
  - ⇨ Collaborative Filtering 모델
- 구매 이력: **평점**, 사용자, 상품명
  - ⇨ 사용자의 만족도와 직결됨
- 소비패턴이 유사한 사용자 찾기 (K명)
  - ⇨ 각각의 선호도가 가장 높은 제품을 **추천**

03 분석 결과

활용 모델 / 분석 결과

웹 Interface


English(US)
Log Out
Stickers
The most beautiful makeup for a woman is passion.
But cosmetics are easier to buy.
71호 트루 레드
Notify Me

웹 Interface

Home

My Recommend

My Page

Stickers

Find Products

Enter Email Address

립스테리
06 엠마

15400원 라카 5 Comment

A small river named Duden flows by their place and supplies it with the necessary regelialia.

Read More >

수퍼 러스트러스 립스틱 기획세트
525 기획세트

14000원 레브론 5 Comment

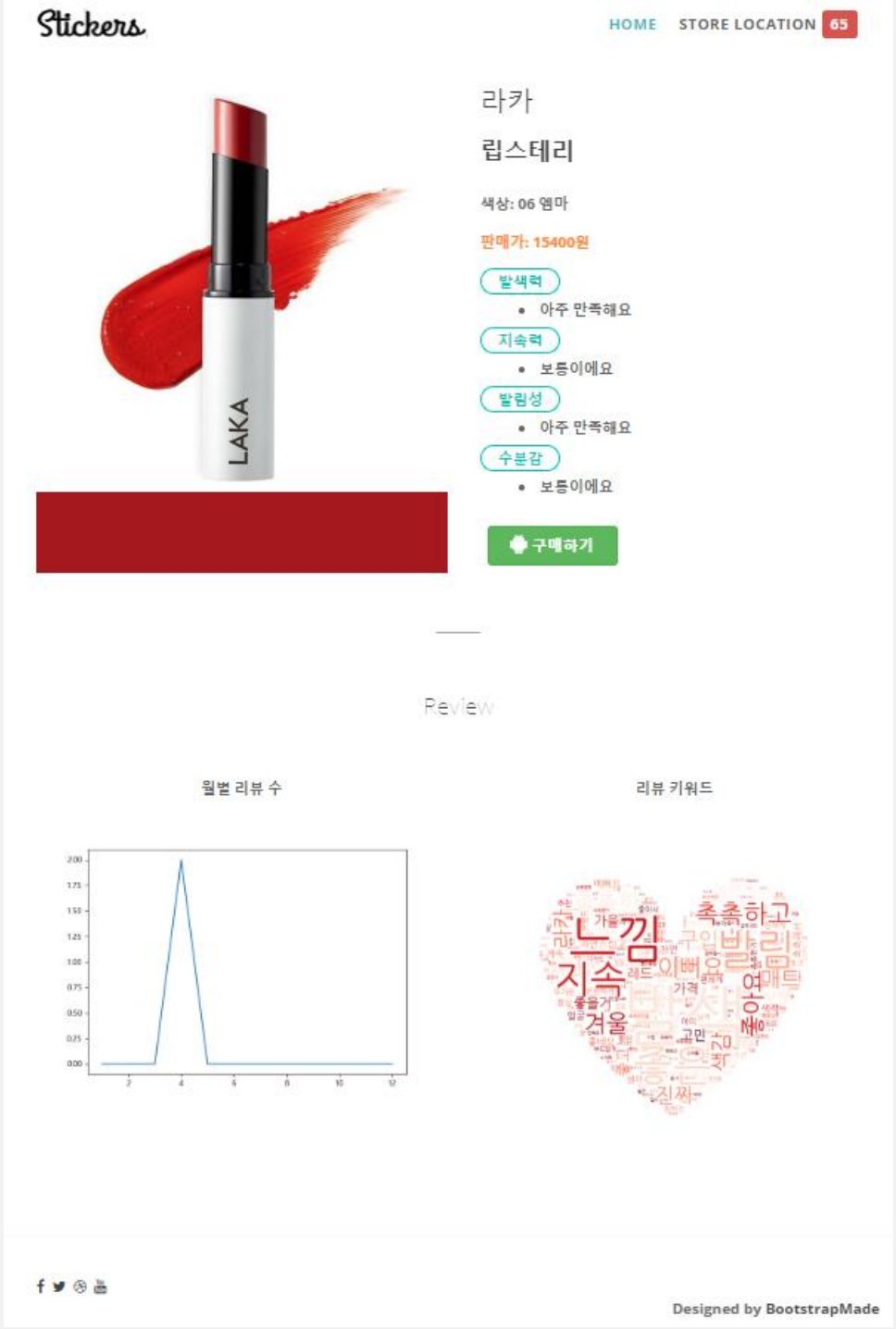
A small river named Duden flows by their place and supplies it with the necessary regelialia.

Read More >

립코스터
BE 01 빈티지 칠리

14400원 웨이크메이크 5 Comment

A small river named Duden flows by their place and supplies it with the necessary regelialia.



결과 제품 비교

Input	Output
	
뿌빠 아이엠 매트 립스틱	라카 립스테리
71호 트루 레드	06 엠마
23,000 원	15,400원
(155, 23, 28)	(164, 23, 29)

My Recommend: 개인화 추천

Home

My Recommend

My Page

Stickers

Find Products

Enter Email Address

립스테리
06 엠마

15400원 라카 5 Comment

A small river named Duden flows by their place and supplies it with the necessary regelialia.

Read More >

수퍼 러스트러스 립스틱 기획세트
525 기획세트

14000원 레브론 5 Comment

A small river named Duden flows by their place and supplies it with the necessary regelialia.

Read More >

립코스터
BE 01 빈티지 칠리

14400원 웨이크메이크 5 Comment

A small river named Duden flows by their place and supplies it with the necessary regelialia.

My Recommend: 개인화 추천

Home

My Recommend

My Page

레트로 매트 리퀴드 립컬러 메탈릭 포일드 매트

34000원 맥 5 Comment

A small river named Duden flows by their place and supplies it with the necessary regelialia.

Read More >

새틴 립스틱 모카 크림

31000원 맥 5 Comment

A small river named Duden flows by their place and supplies it with the necessary regelialia.

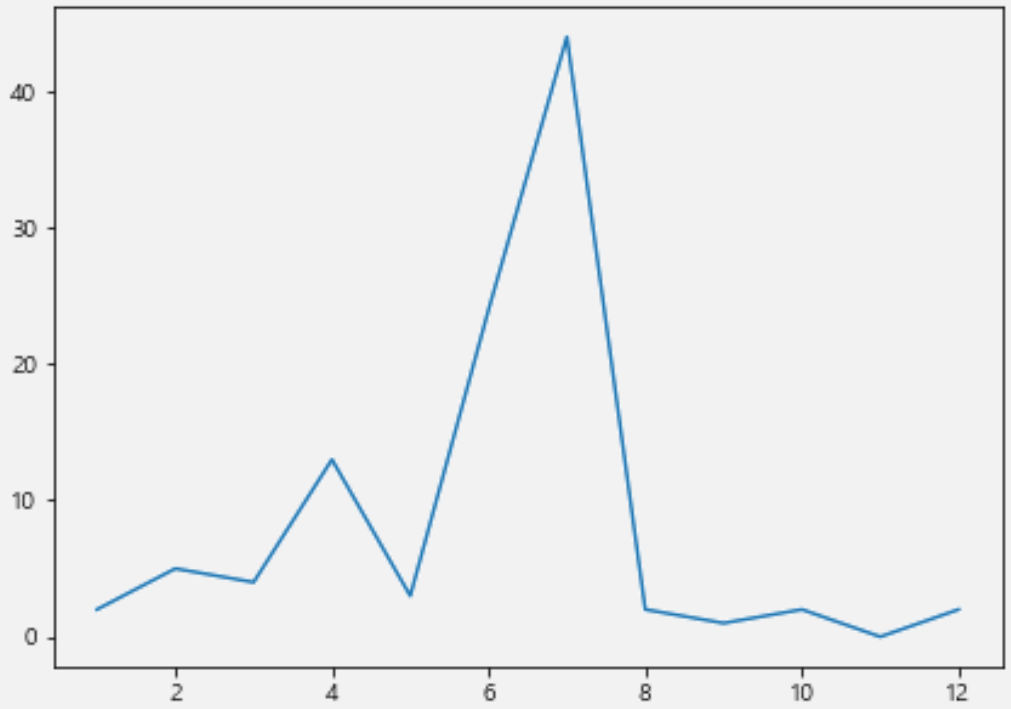
Read More >

퓨어 컬러 엔비 립스틱 누드 무드

41000원 에스티로더 5 Comment

A small river named Duden flows by their place and supplies it with the necessary regelialia

부가 정보



‘ 마몽드 트루컬러립스틱 05 그레이스 ’의 예시

‘입생로랑 루쥬 볼륍떼 샤인 12 ’의 예시

리뷰 개수로 판매 경향 분석

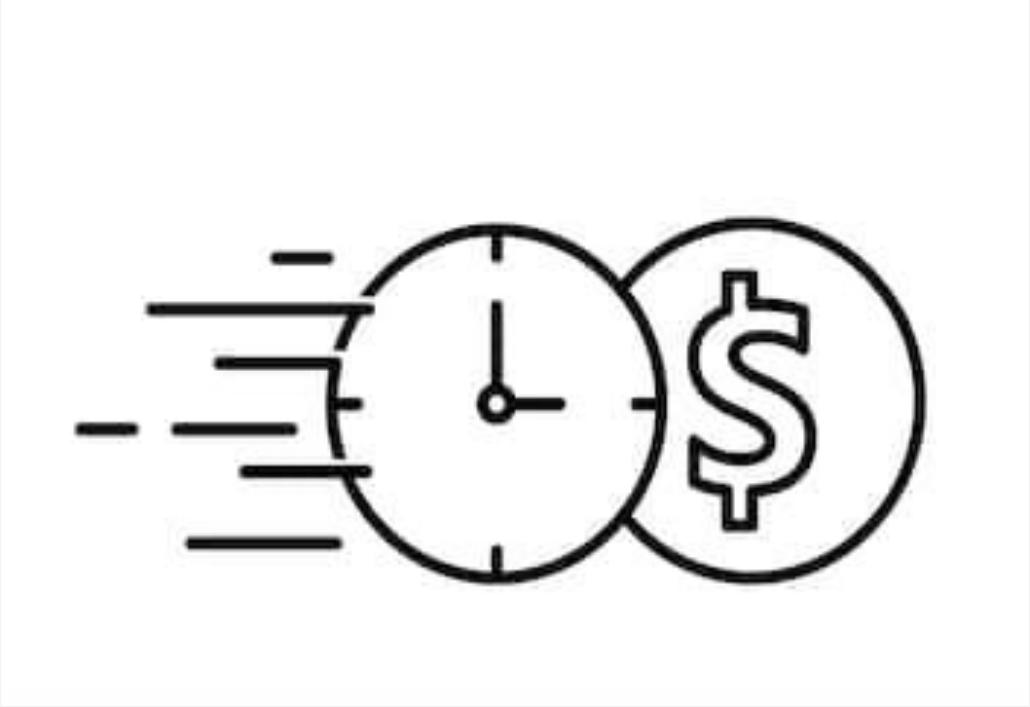
시즌 별 **트렌드 예측** 가능

사용자 리뷰의 워드 클라우드

또 다른 제품 정보 제공 가능

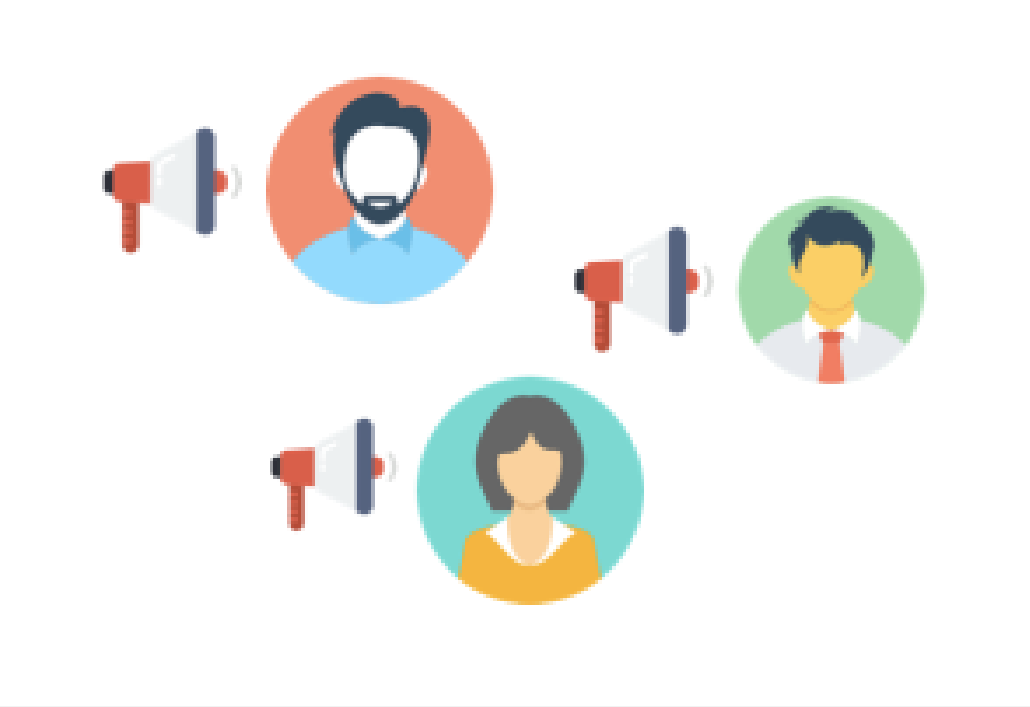
04 K - Data Team Project
기대효과 / 활용방안

04 기대효과 / 활용방안



탐색 비용 감소

빠른 구매 가능



개인화 서비스

자신 만을 위한 맞춤 서비스


다양한 색조 제품에
활용 가능

새로운 색상 개발에 활용

04 기대효과 / 활용방안

헬스 앤 뷰티 스토어 & 브랜드 매장 에서 활용 가능

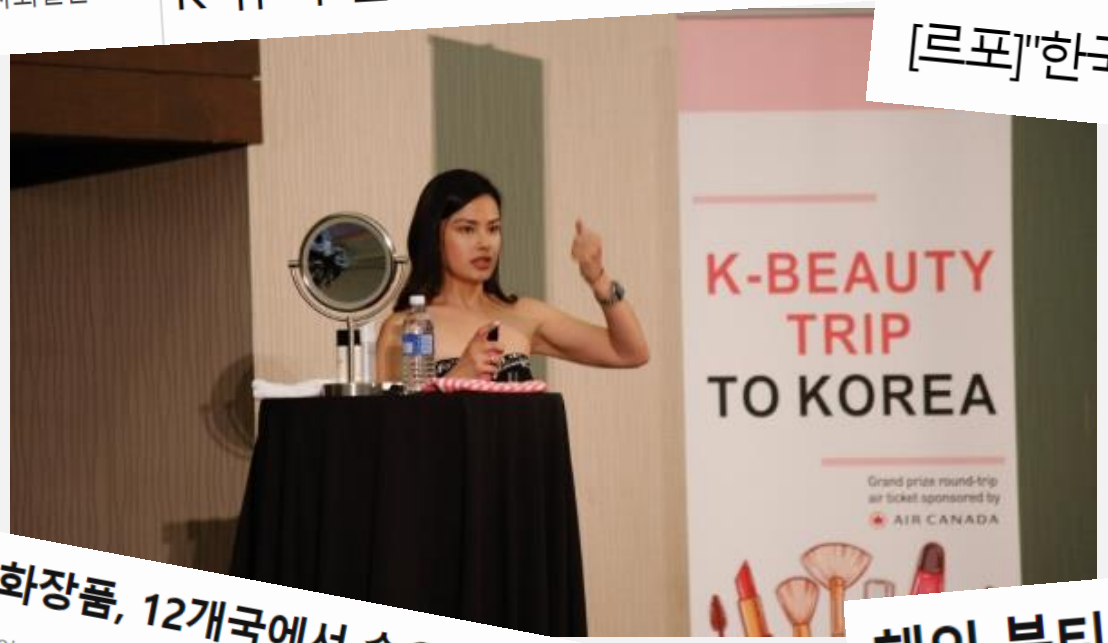
04 기대효과 / 활용방안

Beauty #kbeauty

K-BEAUTY 잘 나가요

사회일반

K-뷰티 붐 타고 세계 화장품 트렌드 선도



[르포]"한국드라마보고 립스틱 샀어요"...베트남 한류박람회 가보니

BEST OF
K-BEAUTY
AWARDS 2018

BEST OF K-BEAUTY
SOKO GLAM
AWARD WINNER

한국화장품, 12개국에서 수요 급등 품목 1위로 선정

해외 바이어들 조사_가성비 및 높은 품질력으로 글로벌 여성들 사이에 한국산 화장품 인지도 확산

해외 뷰티 유튜버(Youtuber)도 'K-beauty'에 관심 집중!

공통적으로 '무결점 피부 표현'을 꼽아_스킨케어, 밝은 립컬러, 외모지상주의까지 언급해

감사합니다

Q & A