

A close-up photograph of a person's face. They are applying a white cream or lotion to their cheek with a silver-colored brush. The person has light-colored hair and is wearing a grey t-shirt with yellow stripes. In the background, there is a large green palm leaf.

GLOWPICK 기반 제품 추천 웹서비스

테스트하기



목차

1. 프로젝트 주제 소개
2. 데이터 소개 / 탐색 / 시각화
3. 데이터 전처리 과정
4. 적용한 분석 기법 및 모델 소개
5. 추천 알고리즘
6. 모델을 활용한 웹 서비스에 대한 소개
7. 후속 과제



1. 프로젝트 주제 소개

“코로나 19로 인해 마스크를 지속적으로 착용함으로써 피부 케어에 대한 관심도가 높아짐”

B 비즈니스워치 2020.05.06.

이젠 마스크가 일상...화장품도 변신 중

한국콜마는 코로나19 사태는 물론 미세먼지 등으로 마스크 사용이 일상화하고 있는 만큼 앞으로 소비자 니즈가 더욱 확대될 것으로 예상하고 있다. ◇ 아모레·유한킴...

#코로나19 #화장품 #아모레퍼시픽 #LG생활과학 #애경 #마스크

뷰티업계, 피부 손상·트러블 완화 및 위생용품 집중



/사진=이명근 기자 qwe123@

코로나19 확산과 함께 마스크와 손소독제를 비롯한 위생용품 사용이 급증하고 있다. 그러자 화장품 업계도 맞춤형 대응에 나서고 있다.

특히 화장품이 묻지 않는 마스크나 마스크 착용에 따른 피부 트러블 완화 화장품들이 큰 인기를 끌고 있으며, 새롭게 위생용품 사업도 강화하고 있다.

I know, I know..... but 2021.05.14.

코로나로 바뀐 러시아 화장품 시장동향

□ 색조화장 판매량 감소, 스킨케어, 핸드크림 판매 증가 □ 2020년 대한민국, 러시아 화장품 수입 제2국 달성 코로나 기간 동안의 화장품 산업 코로나로 인해 러시아 비즈니스 전체가 큰 타격을 입었고 그중 화장품...

#러시아 #색조화장 #판매량 #감소 #스킨케어 #핸드크림

| 코로나19가 가져온 몽골 화장품 시장의 새로운 트렌드(#몽골 #화장품시장)



제품은 너무 많은데..

나에게 맞는 제품은 뭘까?

1. 프로젝트 주제 소개

사용자가 원하는 화장품 종류 , 피부타입 , 성별 , 나이를 고려해 화장품을 추천해주는 웹서비스!!



1. 프로젝트 주제 소개

대상 피부고민이 있는 고객, 제품이 너무 많아 고르는 게 어려운 고객

목적 그 고민을 해소하기 위해 사용자에게 필요한 제품을 추천

방법 글로우픽에 있는 화장품 정보 및 리뷰 데이터를 활용하여 추천 알고리즘 작성

→ 본인의 정보(나이, 피부 고민 등)를 적으면 제품 정보 기반 맞춤 제품 추천



2. 데이터 소개 / 탐색 / 시각화

The screenshot shows a product page for 'Waterfull Hyaluronic Acid Toner'. Key elements include:

- Image:** A blue bottle of toner with a pump dispenser, highlighted with a red box.
- Product Info:** Includes brand (주미소), name (워터풀 히알루론산 토너), size (250ml), and price (20,000원).
- Rating:** 4.65 stars from 156 reviews, highlighted with a red box.
- Review Summary:** 156 reviews, average rating 4.65, 95% satisfaction.
- Review Distribution:** A chart showing the distribution of star ratings: 5 stars (110), 4 stars (38), 3 stars (7), 2 stars (1), 1 star (0).
- Search Bar:** Input field for searching products.
- Filter Options:** Buttons for 전체 (All), 긍정 (Positive), 부정 (Negative), and star ratings (5, 4, 3, 2, 1).
- Recent Activity:** 최근 등록순 (Recent registration).
- Review System:** 글린 리뷰 시스템이 작동 중 (The Green Review System is running).
- Comments:** A comment by a user named 조프 (Jopf) with 53 reviews, highlighted with a red box. The comment reads: "조금 미끈거리지만 얼굴에 바르고나면 쫀쫀해요! 이벤트로 받았던 것 같은데 용량도 많고 잘 쓰고 있습니다. 냄새도 거의 무향이고 자극도 없는 것 같아요".
- Review Details:** A note indicating the review was written by a platform member (평가단).

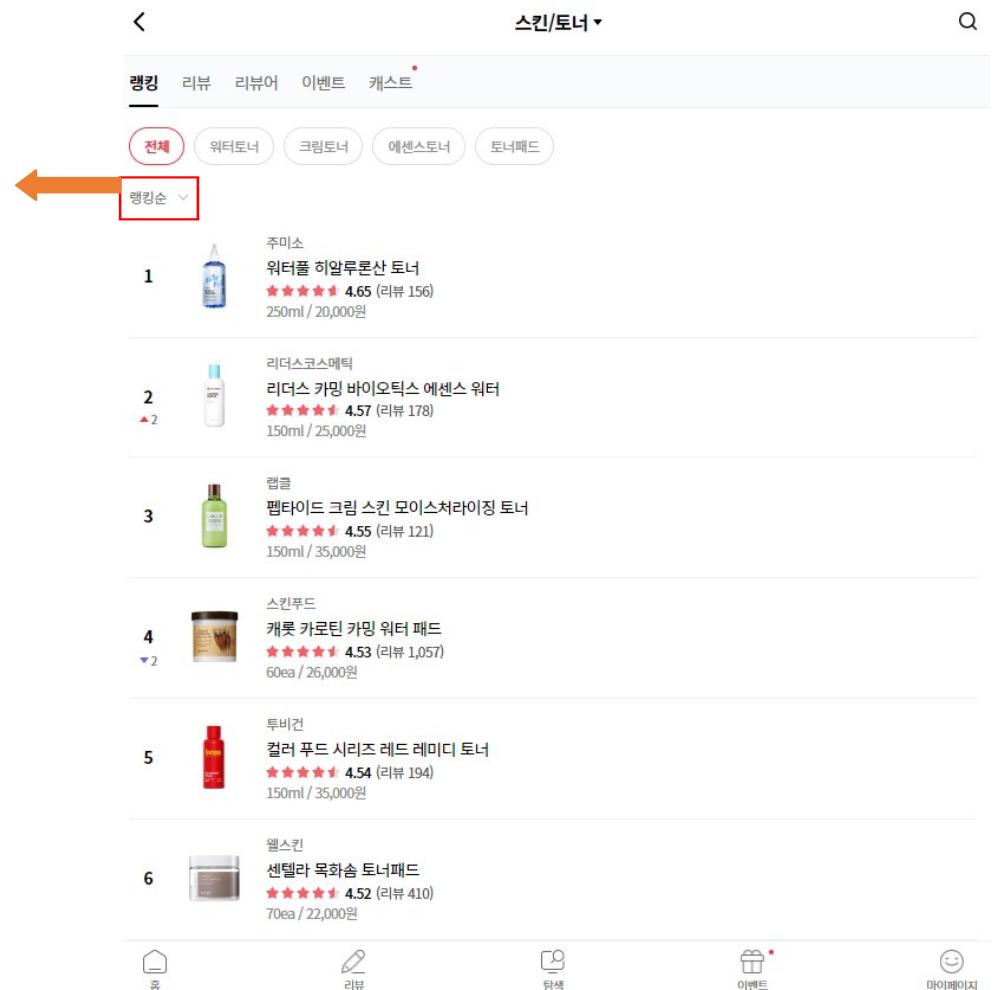
brand	name	bottle	price_num	star	product_info	product_type	ingredient	danger_num	rank_num
age	skintype	gender	rating	review	offer	소분류	대분류	image_link	



추출한 데이터의 columns : 총 19열
약 2만 5천 행

2. 데이터 소개 / 탐색 / 시각화

- 소분류 별로 20개씩 보여줌
- 랭킹 순으로 가져옴으로써
이미 소비자들이 많이 사용하고 있는 제품 추천
(검증된 제품)



The screenshot shows a product ranking page for toners. The top navigation bar includes a back arrow, a search bar, and a dropdown for 'Skin/Toner'. Below the navigation is a sub-navigation bar with categories: 전체 (All), 워터토너 (Water Toner), 크림토너 (Cream Toner), 에센스토너 (Essence Toner), and 토너패드 (Toner Pad). A red box highlights the 'Ranking' dropdown menu. The main content area displays six products in a list format:

Rank	Brand/Type	Product Name	Rating	Reviews	Price
1	주미소	워터풀 히알루론산 토너	★★★★★	4.65 (리뷰 156)	250ml / 20,000원
2	리더스코스메틱	리더스 카밍 바이오틱스 에센스 워터	★★★★★	4.57 (리뷰 178)	150ml / 25,000원
3	캡글	펩타이드 크림 스킨 모이스처라이징 토너	★★★★★	4.55 (리뷰 121)	150ml / 35,000원
4	스킨푸드	캐롯 가로틴 카밍 워터 패드	★★★★★	4.53 (리뷰 1,057)	60ea / 26,000원
5	투비건	컬러 푸드 시리즈 레드 레미디 토너	★★★★★	4.54 (리뷰 194)	150ml / 35,000원
6	엘스킨	센텔라 목화솜 토너패드	★★★★★	4.52 (리뷰 410)	70ea / 22,000원

At the bottom of the page are icons for Home, Review, Toner, and Event.

2. 데이터 소개 / 탐색 / 시각화

수집 항목 우선 순위(피부 고민에 초점)

1. **스킨케어**
2. 클렌징
3. 마스크/팩
4. 썬케어
5. 배쓰&바디

진행 속도 보면서 점차 카테고리 늘리려고 했으나 시간 부족으로 스킨케어만 진행

카테고리	브랜드	스토어
 스킨케어		 페이스메이크업
 립메이크업		 아이메이크업
 컨투어링		 뷰티풀
 클렌징		 마스크/팩
 선케어		 배쓰&바디
 헤어		 네일
 생활용품		 렌즈
 프레그란스		 디바이스
 이너뷰티		 베이비
 남성화장품		

2. 데이터 소개 / 탐색 / 시각화

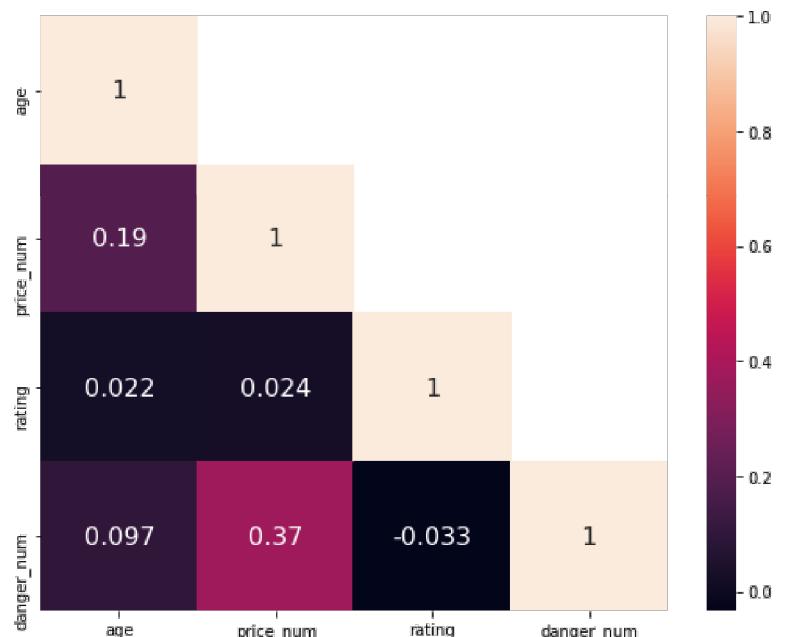
- 추출 방법 : 셀레니움을 사용한 동적 웹 크롤링

	brand	name	bottle	price_num	star	product_info	product_type	ingredient	danger_num	rank_num	age	skintype	gender	rating	review	offer
0	주미소	워터풀 히 알루론산 토너	250	20000	4.65	피부 갈증의 해답 히알루론산 수분길토너- 저분자부터 고분자까지 탄탄하게 채워주는 강...	보습 수분공 급 저자극	정제수Water;Aqua피부건강제, 용제부틸렌글라이콜Butylene Glycol...	0	1	21	복합성	여	5	조금 미끈거리 지만 얼굴에 바 르고나면 쫀쫀 해요! 이벤트로 받았던 것 같은 데 용량도 많...	0
1	주미소	워터풀 히 알루론산 토너	250	20000	4.65	피부 갈증이 해답 히알루론산 수분길토너- 저분자부터 고분자까지 탄탄하게 채워주는 강...	보습 수분공 급 저자극	정제수Water;Aqua피부건강제, 용제부틸렌글라이콜Butylene Glycol...	0	1	31	복합성	여	4	친구가 있어서 써봤는데요 사 용감이 쫀쫀한 게 건성인피부 에 수분보호막 을 형성해주는 것 같...	0
2	주미소	워터풀 히 알루론산 토너	250	20000	4.65	피부 갈증의 해답 히알루론산 수분길토너- 저분자부터 고분자까지 탄탄하게 채워주는 강...	보습 수분공 급 저자극	정제수Water;Aqua피부건강제, 용제부틸렌글라이콜Butylene Glycol...	0	1	16	민감성	여	5	자극적이지 않 은 제품 얼굴이 되게 쫀쫀해지 는거 같고 자극 적이지 않아서 좋았어요	0
3	주미소	워터풀 히 알루론산 토너	250	20000	4.65	피부 갈증의 해답 히알루론산 수분길토너- 저분자부터 고분자까지 탄탄하게 채워주는 강...	보습 수분공 급 저자극	정제수Water;Aqua피부건강제, 용제부틸렌글라이콜Butylene Glycol...	0	1	39	복합성	여	4	좋습니다 미끌 거리지만 되직 한게 오히려 수 분을 잘잡네요	0
4	주미소	워터풀 히 알루론산 토너	250	20000	4.65	피부 갈증의 해답 히알루론산 수분길토너- 저분자부터 고분자까지 탄탄하게 채워주는 강...	보습 수분공 급 저자극	정제수Water;Aqua피부건강제, 용제부틸렌글라이콜Butylene Glycol...	0	1	27	복합성	여	5	살짝 미끌거리 는 제형이라서 수분도 묵직하 게 채워주는느 낌이고 용기도 위생적이고 편...	0

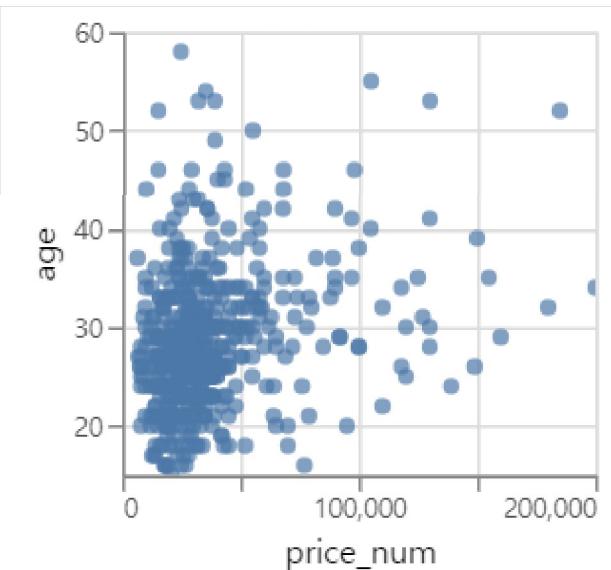
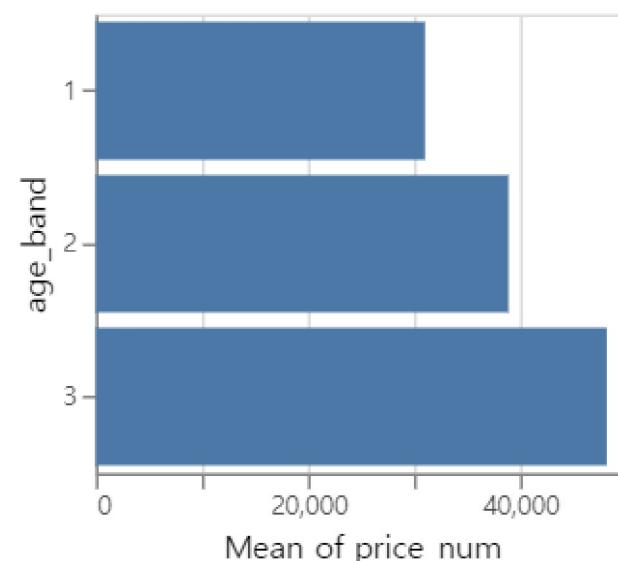
2. 데이터 소개/탐색/시각화

데이터 시각화

Seaborn heatmap



Lux



-> 나이정보 입력 안받기로 결정

2. 데이터 소개 / 탐색 / 시각화

총 리뷰중 성별

남

442
명

여

24,632

-> 성별정보 입력 안받기로 결정

2. 데이터 소개 / 탐색 / 시각화

전처리 1

- 결측치 처리 및 scaler

```
# 데이터 전처리_part1

def data_preprocess(df_list):
    dataframe_list = []
    for df in df_list:
        for j in range(len(df['price_num'])):
            if df['price_num'][j] == '가격미정':
                df['price_num'][j] = None

        # 성분정보 결측치 처리
        df['ingredient'] = df['ingredient'].fillna('성분정보없음')
        # 제품 가격 결측치 처리
        df['price_num'] = df['price_num'].apply(
            lambda x : int(x.replace(',','').replace(' ','')) if (bool(x) == True) else x)
        df['price_num'] = df['price_num'].fillna(df['price_num'].mean())
        df['danger_num'] = df['danger_num'].apply(lambda x : 0 if (x == '없음') else x)
        df['danger_num'] = df['danger_num'].apply(lambda x : int(x))
        # 가격, 위험성분 column scoring and Scaler
        min_max_sc = MinMaxScaler()
        df['price_num_score'] = min_max_sc.fit_transform(df[['price_num']])
        df['danger_num_score'] = min_max_sc.fit_transform(df[['danger_num']])
        df['price_num'] = df['price_num'].astype(int)
        # 리뷰 결측치 처리
        df['review'] = df['review'].fillna('리뷰없음')

    dataframe_list.append(df)
    return dataframe_list
```

3. 데이터 수집 및 전처리 과정

전처리 2 - 리뷰 토큰화

```
# 데이터 전처리_part2
# 리뷰 토큰화 - 향후 장고에서 제품 추천 알고리즘 적용 시 데이터 처리 속도 향상을 위하여 사전 작업 진행.
okt = konlpy.tag.Okt()
def tokenize(df_list):
    dataframe_list = []
    for df in df_list:
        col_to = []
        for item in df['review']:
            col_to.append(okt.morphs(item))
        review_to = []
        for j in col_to:
            review_to.append(' '.join(j))

        df['review'] = review_to
        dataframe_list.append(df)
    return dataframe_list
```

3. 데이터 수집 및 전처리 과정

TF-IDF

- 결측치 처리 및 scaler

```
# TF-IDF 적용
def tfidf_vect_func(df, target_col, target_row):

    tfidf_vect = TfidfVectorizer()
    feature_vect = tfidf_vect.fit_transform(df[target_col])

    #코사인 유사도 분석
    similarity_simple_pair = cosine_similarity(feature_vect[target_row], feature_vect)
    result_list = similarity_simple_pair.tolist()[0]
    df[f'{target_col}_result'] = result_list

return df
```

3. 데이터 수집 및 전처리 과정

CountVectorizer

- 제품 특성 유사도 분석

```
#count_vectorizer 적용
def count_vect_func(df,target_col, target_row):
    count_vect_category = CountVectorizer(min_df=0, ngram_range=(1,2))
    feature_vect = count_vect_category.fit_transform(df[target_col])

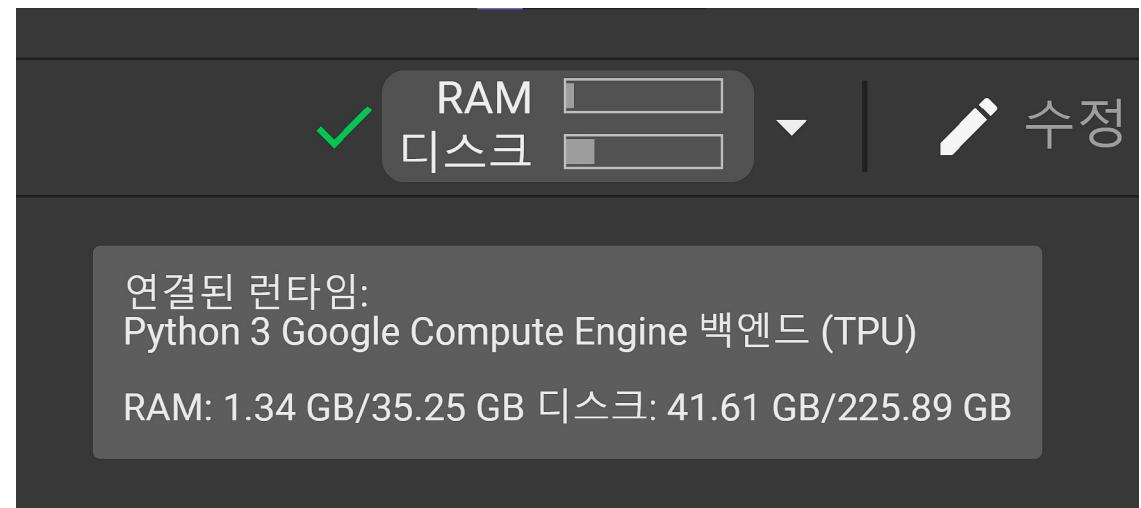
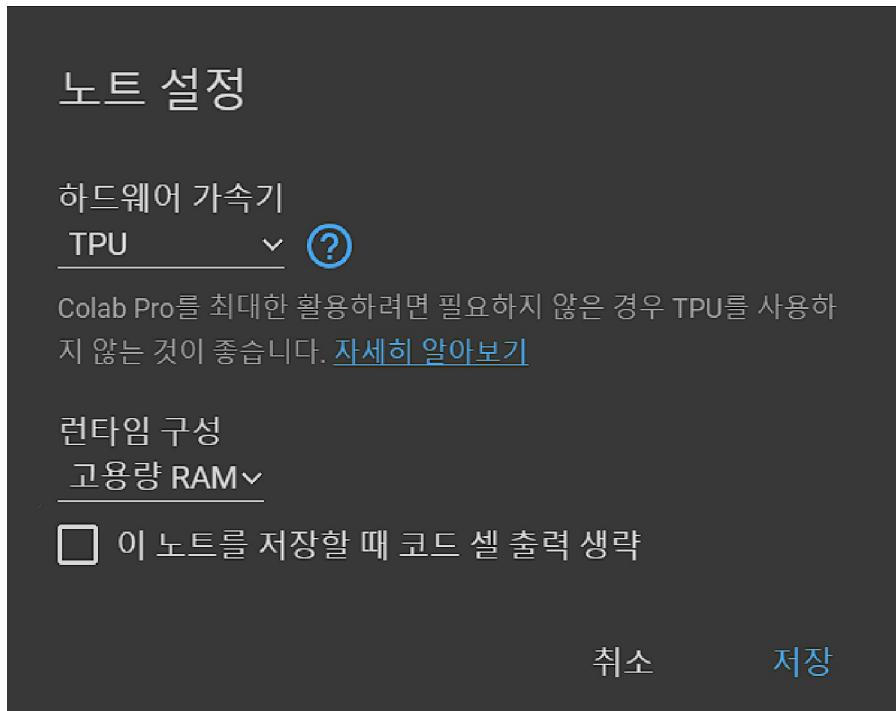
    #코사인 유사도 분석
    similarity_simple_pair = cosine_similarity(feature_vect[target_row], feature_vect)
    result_list = similarity_simple_pair.tolist()[0]
    df[f'{target_col}_result'] = result_list

return df
```

4. 적용한 분석 기법 및 모델 소개

BERT

-colab pro 결제를 통한 고용량 RAM 및 TPU 사용



4. 적용한 분석 기법 및 모델 소개

BERT

```
# ready to fine tuning with review data of naver shopping and glowpick
reviews_df = pd.read_table('/content/drive/MyDrive/Colab Notebooks/naver_glowpick.txt', names=['ratings', 'reviews'])
reviews_df = reviews_df.drop_duplicates('reviews')
reviews_df = reviews_df.reset_index(drop=True)
reviews_df['ratings'] = reviews_df['ratings'].apply(pd.to_numeric)
reviews_df['label'] = reviews_df['ratings'].apply(lambda x : 1 if x > 3 else 0)

for dirname, _, filenames in os.walk('/content/drive/MyDrive/Colab Notebooks/data'):
    for filename in filenames:
        data = pd.read_excel(f'/content/drive/MyDrive/Colab Notebooks/data/{filename}')
        sentiment_list=[]
        file_name = filename.replace('_complete.xlsx', '')
        for num in range(len(data)):
            sentiment_list.append(predict_sentiment(data['review'][num], tokenizer, model))
            clear_output(wait=True)
            print(f'{file_name}:{np.round(num/len(data),2)*100}%', flush=True)
        data['sentiment'] = np.array(sentiment_list)
        new_file_name = filename.replace('complete','sentiment')
        data.to_excel(f'{new_file_name}', encoding='utf-8')
```

5. 추천 알고리즘

피부타입	좋은 성분	주의해야할 성분
지성(여드름) 피부	글리콜린산/살리실산/난 옥시놀-9/녹차/ 위치 하젤/레몬/감파/멘톨클로로필/ 알라토인/ 티트리/감초/징크 옥사이트/칼렌 둘라 추출물/ 설퍼/트리클로란/티타늄 옥사이드	트리글리세라이드/팔마티산염/미리스틴산/ 스테아르산염/스테아린산/코코넛오일/ 시어버터/바세린/옥시벤존/메톡시시나메이트
건성(노화) 피부	히아루론산/글리세린/프로필렌 글라이콜/1,3-부 틸렌 글라이콘/소디움PCA/ 비타민E/비타민A/비타민C/콜레겐/엘라스틴/아 보카도 오일/이브닝 프라임 로즈 오일 오토밀 단백질/콩 추출물/카모마일/오이/복숭아/ 해조 추출물/상백피 추출물/코직산/ 알부틴/포토씨 추출물/베타카로틴/시어버터/파 일워트 추출물/비타민B 복합체/판테놀	알코올/진흙/계면활성제 멘톨/페파민트
민감성 피부	비타민K/비타민F/호스트체스트넛 추출물/카모마 일/알로에/콘플라워/알란토인/ 해조 추출물/티타늄 옥사이드	알코올/계면활성제/멘톨/페파민트/유칼립투 스/아로마오일/고농도 과일산(AHA,BHA) 오렌지/딸기/레몬/레티놀/옥시벤존/메톡시 시 나메이트/ 그 밖에 자신의 피부에 맞지 않는 알레르기 물 질 등

- 건성, 지성, 민감성 :
좋은성분 가산점, 주의성분 필터링**

- 복합성**
**- 건성 + 지성 주의해야할 성분 필터
링**

- 중성**
- 가산점,필터링 반영X

3. 주의해야 할 성분 필터링

```
#피해야할 성분이 있는 행 삭제
del_list = '|'.join(good_oily_list) #피부타입 입력값에 따른 bad리스트
df_result['del'] = df_result['ingredient'].str.contains(del_list)
del_index = df_result[df_result['del']==True].index
df_result.drop(del_index, axis=0, inplace=True)
df_result.drop(['del'], axis=1, inplace=True)
df_result
```

5. 추천 알고리즘

4. 겹치는 리뷰 개수 → 가산점 반영

```
#겹치는 리뷰 갯수 셈 후 열 추가  
num_count_dict = df_result['name'].value_counts().to_dict()  
df_result['num_count'] = df_result['name'].apply(lambda x : num_count_dict[x])  
  
#겹치는 행 삭제  
df_result = df_result.drop_duplicates(['name'])  
  
#겹치는 리뷰 갯수 스케일링 후 가산점  
df_result['num_count_score'] = min_max_sc.fit_transform(df_result[['price_num']])  
df_result['new_result'] = df_result.apply(lambda x : x['new_result']+ (0.01*x['num_count']), axis=1)
```

5. 추천 알고리즘

5. 가격 낮은 경우 가산점(사용자 선택 시)

```
def weight_sum (df_result,question_1,question_2,min_max_sc):  
  
    #가격이 중요할 경우 가중치 부여  
    if question_1 == 'q1_price' :  
        df_result['price_num_score'] = min_max_sc.fit_transform(df_result[['price_num']])  
        df_result['new_result'] = df_result.apply(lambda x : x['new_result']+ ( 0.01*(1- x['price_num_score'])), axis=1)  
  
    #성분이 중요할 경우 가중치 부여  
    if question_2 == 'q2_ingre':  
        df_result['danger_num_score'] = min_max_sc.fit_transform(df_result[['danger_num']])  
        df_result['new_result'] = df_result.apply(lambda x : x['new_result']+ ( 0.01*(1-x['danger_num_score'])), axis=1)  
  
    df_result.sort_values(by='new_result',ascending=False,inplace=True)  
    df_result_final = df_result  
  
    return df_result_final
```

5. 추천 알고리즘

6. 추천성분 및 감성분석 가산점

```
#추천성분 가산점
good_list = '|'.join(good_oily_list) #피부타입 입력값에 따른 good리스트
df_result['good'] = df_result['ingredient'].str.contains(good_list)
df_result['new_result'] = df_result.apply(lambda x : x['new_result']+ ( 0.01*x['good']), axis=1)

#감성분석 가산점
df_result['new_result'] = df_result.apply(lambda x : x['new_result']+ ( 0.01*x['sentiment']), axis=1)
```

6. 모델을 활용한 웹 서비스에 대한 소개

와이어 프레임



#한눈에 #플로틀리
#워드 #예쁘게
#한눈에 #예쁘게 #워드
#한눈에 #들어오는 #한눈에 #한눈에
#예쁘게 #플로틀리

원하는 태그를 모두 입력해주세요!

뒤로

확인

This wireframe shows a list of hashtags related to skincare and beauty ('#한눈에 #플로틀리', '#워드', '#예쁘게', etc.) and a text input field for users to enter their own hashtags ('원하는 태그를 모두 입력해주세요!').

마지막으로 더 정확한 결과를 위해 정보를 입력해주세요!

피부타입 자성 건성 복합성

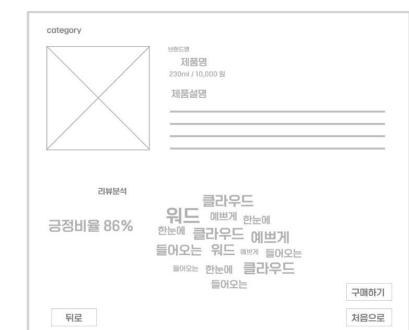
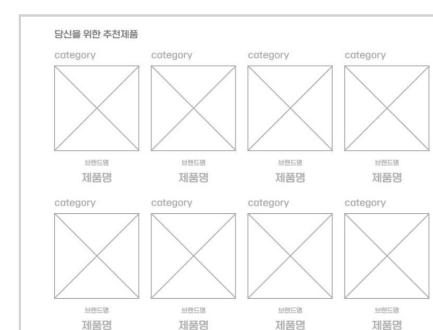
성별 남 여 (데이터상 떠보고 추후 결정)

나이 세

뒤로

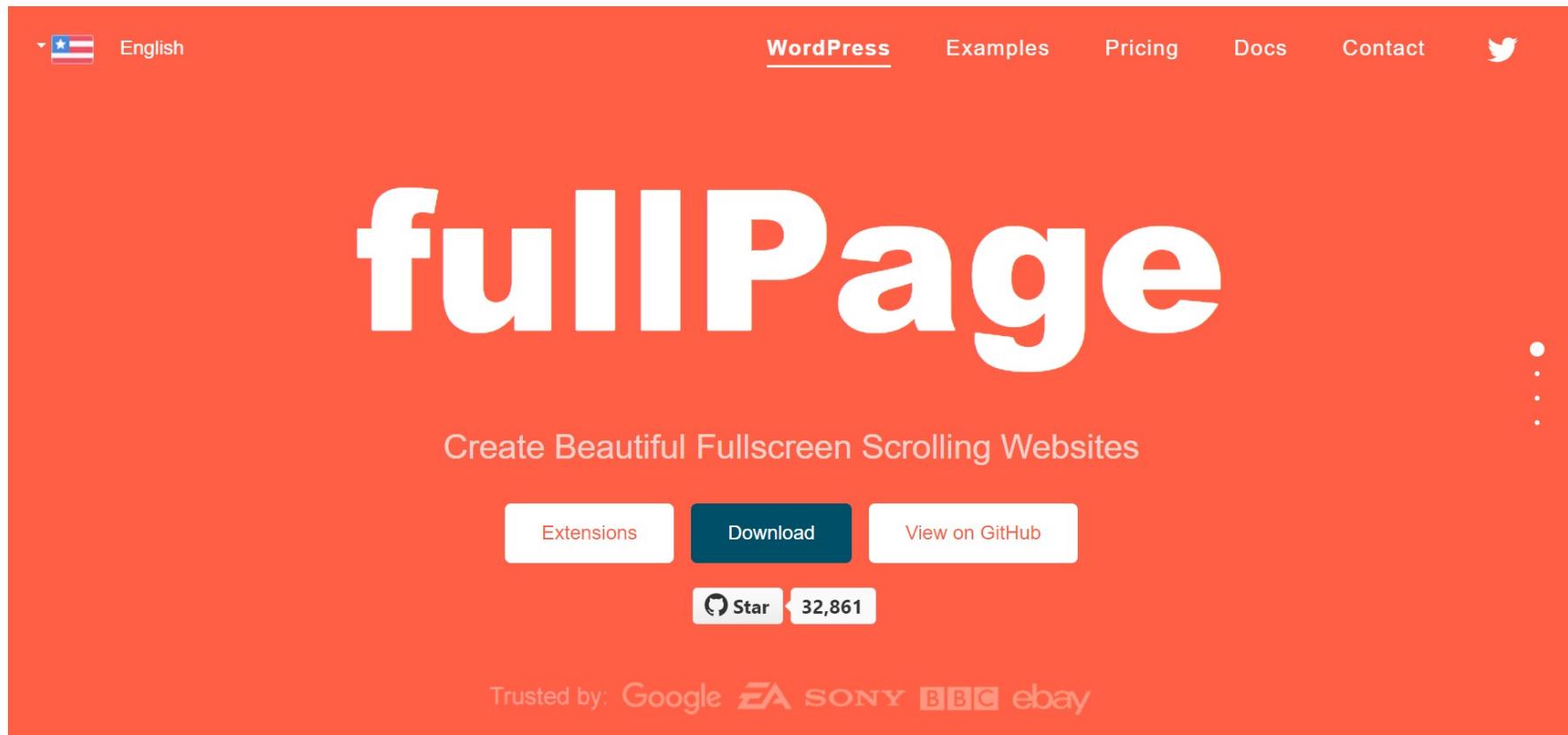
확인

This wireframe asks for specific user information to refine results: skin type ('피부타입'), gender ('성별'), age ('나이'), and gender ('성별'). It also includes a note that the gender selection is based on data and will be updated later ('(데이터상 떠보고 추후 결정)').



6. 모델을 활용한 웹 서비스에 대한 소개

fullpage.js 활용



6. 모델을 활용한 웹 서비스에 대한 소개

사용자 입력 페이지



6. 모델을 활용한 웹 서비스에 대한 소개

사용자 입력 페이지



6. 모델을 활용한 웹 서비스에 대한 소개

사용자 입력 페이지



6. 모델을 활용한 웹 서비스에 대한 소개

사용자 입력 페이지



6. 모델을 활용한 웹 서비스에 대한 소개

사용자 입력 페이지

정확한 결과를 위해 정보를 입력해주세요

어떤 카테고리의 제품을 찾으시나요?

스킨/토너 ▼

피부 타입을 알려주세요.

- 건성
- 지성
- 중성
- 복합성
- 민감성

전송

6. 모델을 활용한 웹 서비스에 대한 소개

사용자 입력 페이지

원하는 태그를 아래 이미지에서 골라서 입력해 주세요



수분공급

결과보기

6. 모델을 활용한 웹 서비스에 대한 소개

결과 페이지

분석완료!

AI가 당신에게 맞는 추천제품을 찾아냈어요



마녀공장
비피다 바이옴 앰플 토너

[자세히 보기](#)



더랩 바이 블랑드
올리고 히알루론산 7멀티 포
뮬라 패드

[자세히 보기](#)



프로주
모이스처 크리미 토너

[자세히 보기](#)



웰스킨
센텔라 목화솜 토너패드

[자세히 보기](#)



닥터자르트
2세대 시카페어 토너

[자세히 보기](#)



바이오헐 보
판테놀 시카 블레미쉬 논코메
도 더마 패드

[자세히 보기](#)



디어클레어스
서풀 프레퍼레이션 페이셜 토
너

[자세히 보기](#)



브링그린
사철쑥 카밍 약산성 토너 패드

[자세히 보기](#)

6. 모델을 활용한 웹 서비스에 대한 소개

제품 상세 페이지



제품명 :
비피다 바이옴 앰플 토너

가볍게 닦아내기만 해도 앰플 케어 효과를 선사하는 토너- PHA 성분으로 피부에 쌓이기 쉬운 각질과 노폐물을 부드럽게 정리- 5중 히알루론산으로 촘촘하게 채워 물 셀 틈 없이 꽉 찬 수분- 빠른 진정 효과와 탁월한 모공 수렴 작용

브랜드 : 마녀공장
카테고리 : 워터토너
가격 : 25,000 원

긍정리뷰 비율 :
90.0%

7. 후속 과제

1. 상품분류 좀더 다양하게 데이터 수집해서 사용해보기
2. 다른 뷰티사이트의 화장품 정보들도 가져와 사용해보기
3. 데이터 수집과정 간소화 및 통합
4. 적절한 weight 값 연구

A close-up photograph of a woman's hands. She is wearing a white t-shirt and a grey sweatshirt. In her right hand, she holds a white skincare bottle with a silver cap, pouring a small amount of cream onto her left hand. A green orchid flower is visible in the lower-left foreground. The background is blurred, showing tropical leaves.

myskincare 시스템이 종료되었습니다!!

감사합니다