



DIOR & YSL

립 제품 리뷰 분석

서론

데이터 수집

전처리

단어 빈도

감성 분석

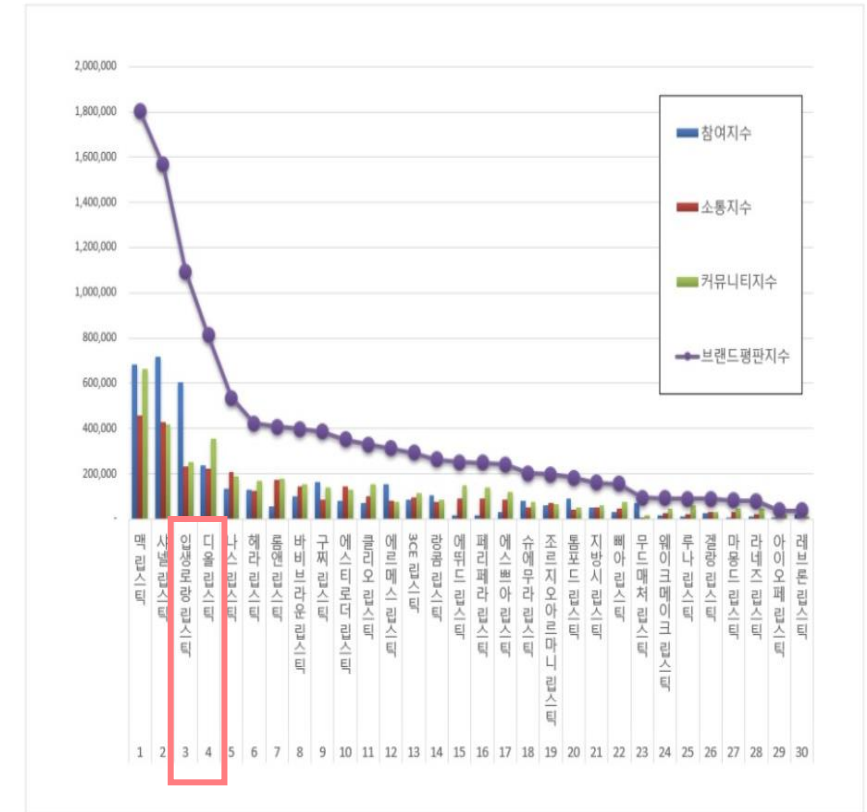
주제 분석

결론

- 주제 : DIOR & YSL 립 제품 리뷰 분석

- 주제 선정 이유 :

평소 화장품에 관심이 많아, 화장품 리뷰 분석을 통해 그 안에 담긴 해당 브랜드 및 제품의 특징을 찾아보고 자본 주제를 선정하게 되었다. 개인차가 심한 화장품의 특성을 고려하여, 가장 개인차가 적고 개인이 여러 개의 립 제품을 사용하는 만큼 리뷰에서 얻을 수 있는 정보량이 많을 것이라고 판단하여 화장품 종류를 립 제품으로 제한했다. 2021년 1월 립스틱 브랜드 평판 순위에서 상위권을 차지한 브랜드 중 리뷰 양이 많은 DIOR 과 YSL(입생로랑)을 선정하여 각 브랜드의 대표 제품에 대한 리뷰 분석을 진행했다.



(출처 : [브랜드평판] 립스틱 브랜드 2021년 1월 빅데이터 분석결과...1위 맥 립스틱, 2위 샤넬 립스틱, 3위 입생로랑 립스틱 (rekorea.net))

■ 서론

■ 데이터 수집

■ 전처리

■ 단어 빈도

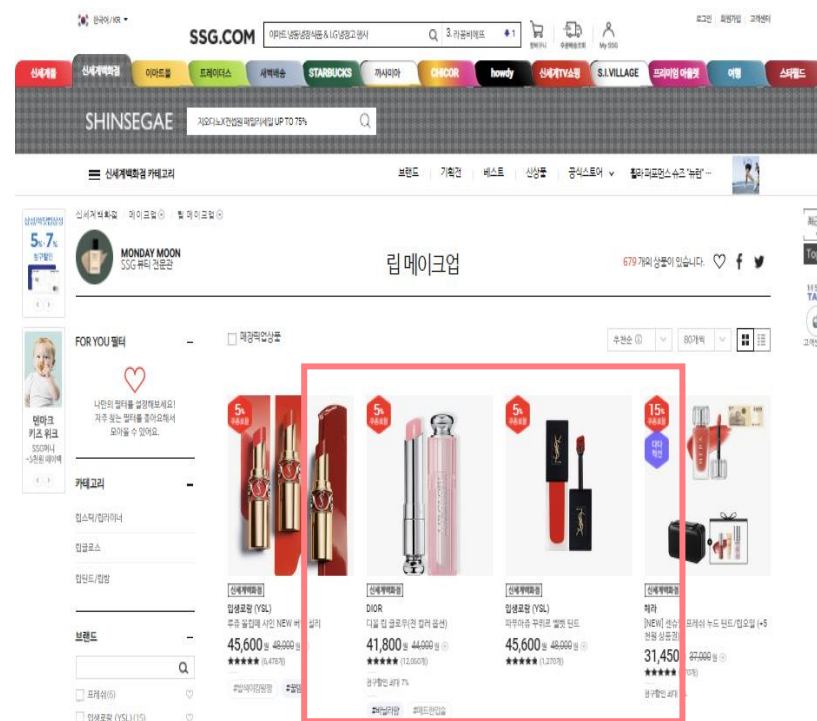
■ 감성 분석

■ 주제 분석

■ 결론

- 데이터 수집 :

화장품을 판매하는 종합 쇼핑몰 중 두 브랜드
대표 립 제품의 리뷰를 가장 많이 포함하고 있는
신세계백화점에서 Selenium을 활용하여 리뷰
웹 스크래핑을 진행했다.



- 전처리 :

clean_text 라는 이름의 사용자 정의 함수를 만들고
구두점 , 불필요한 공백들을 삭제하였고,
no_emoji 라는 이름의 사용자 정의 함수를 만들어,
이모티콘을 삭제하였다.

(참고자료 : <https://ebbnflow.tistory.com/246>)

참고자료 : <https://studyprogram.tistory.com/entry/Python%EC%97%90%EC%84%9C%ED%85%8D%EC%8A%A4%ED%8A%B8-%EC%95%88%EC%9D%98%EC%9D%B4%EB%AA%A8%EC%A7%80emoji%EC%A0%9C%EA%B1%B0%ED%95%98%EA%B8%B0>

```
def clean_text(texts):  
    review = re.sub(r'[@&#%+=()!/~#@^*&*?«»©³³³ª¹º-¼½¾·¸:~¡¢£¥¦§¨ª«¬®¯°±²³´µ¶·¸¹º»¼½¾¿`"]', '', str(texts))  
    review = review.lower()  
    review = re.sub(r'www+', '', review)  
    review = re.sub(r'<[>]+>', '', review)  
    review = re.sub(r'www+', '', review)  
    review = re.sub(r'"\"'+\"www\"'+\", \"\", review)  
    review = re.sub(r'www+$', '', review)  
    return review
```

```
def no_emoji(text):
    emoji_pattern = re.compile("[
        u'\\U0001F600-\\U0001F64F'
        u'\\U0001F300-\\U0001F5FF'
        u'\\U0001F680-\\U0001F6FF'
        u'\\U0001F1E0-\\U0001F1FF'
    ]+", flags=re.UNICODE)
    return emoji_pattern.sub(r'', text)
```

- 형태소 분석 :

Stanza를 사용하여 형태소 분석을 진행하고
형태소가 명사에 해당하는 단어들로 단어 문서 행렬을 만들었다.

■ 서론

■ 데이터 수집

■ 전처리

■ 단어 빈도

■ 감성 분석

■ 주제 분석

■ 결론

- 단어 구름을 만들 때에 립 제품과 관련된 이미지인 입술 이미지와 그 색상을 이용하여 단어 구름을 만들기 위해 ImageColorGenerator와 WordCloud의 mask 파라미터, matplotlib의 imshow를 활용해 코드를 만들어 입술 모양 단어 구름을 만들었다.



단어 구름(CountVectorizer)



단어 구름(TfidfVectorizer)

- 리뷰 특성상 많이 들어갈 수 밖에 없는 "제품", "구입", "색상", "선물", "사용" 등의 단어가 가장 높은 빈도로 나타난 것을 확인할 수 있었다.

■ 서론

■ 데이터 수집

■ 전처리

■ 단어 빈도

■ 감성 분석

■ 주제 분석

■ 결론



DIOR 단어 구름



YSL 단어 구름

- 브랜드별로 나누어 단어 구름을 그려보았다. 앞에서 그렸던 두 단어구름과 마찬가지로 공통적으로 "제품", "구매", "배송", "선물" 같은 단어들이 많이 나타난 것을 확인할 수 있었다.
- DIOR과는 다르게 YSL 단어 구름에서는 "호" 라는 단어가 큰 비중을 차지하는 것을 확인할 수 있는데, 이는 색상 명을 1호, 2호 ... 이런 식으로 사용하여 다양한 색상의 제품을 선보이는 YSL 특징이 담겨 있다고 볼 수 있다.

YSL에 대한 리뷰

0

DIOR에 대한 리뷰

1

- DIOR에 대한 리뷰를 1, YSL에 대한 리뷰를 0으로 두고 감성 분석을 진행했다.

```
model.fit(x_train.A, y_train, epochs = 3)
```

```
Epoch 1/3  
568/568 [=====] - 2s 3ms/step - loss: 0.6765 - accuracy: 0.5786  
Epoch 2/3  
568/568 [=====] - 2s 3ms/step - loss: 0.6174 - accuracy: 0.6886  
Epoch 3/3  
568/568 [=====] - 2s 3ms/step - loss: 0.5826 - accuracy: 0.7039  
<tensorflow.python.keras.callbacks.History at 0x28d36f5bb48>
```

```
model.evaluate(x_test.A, y_test)
```

```
142/142 [=====] - 0s 2ms/step - loss: 0.5801 - accuracy: 0.6991  
[0.5800700187683105, 0.699118971824646]
```

CountVectorizer를 이용한 TDM 감성분석

```
: model.fit(x_train.A, y_train, epochs = 3)
```

```
Epoch 1/3  
568/568 [=====] - 2s 3ms/step - loss: 0.6691 - accuracy: 0.6115  
Epoch 2/3  
568/568 [=====] - 2s 3ms/step - loss: 0.6116 - accuracy: 0.6859  
Epoch 3/3  
568/568 [=====] - 2s 3ms/step - loss: 0.5811 - accuracy: 0.7012  
<tensorflow.python.keras.callbacks.History at 0x28d32af6748>
```

```
: model.evaluate(x_test.A, y_test)
```

```
142/142 [=====] - 0s 2ms/step - loss: 0.5777 - accuracy: 0.6980  
[0.5777093768119812, 0.6980175971984863]
```

TfidfVectorizer를 이용한 TDM 감성분석

- 두 가지 TDM으로 감성분석을 진행하였는데,
둘 다 약 0.7 정도의 아주 유사한 정확도를 얻었다.

■ 서론

■ 데이터 수집

■ 전처리

■ 단어 빈도

■ 감성 분석

■ 주제 분석

■ 결론

DIOR 가중치표	YSL 가중치표
디올	틴트
립밤	호
립글로우	입생로랑
코랄	입생
보습력	12
글로우	9
라즈베리	지속력
자연	색상
은은	레드
핑크	12호

DIOR 가중치표	YSL 가중치표
코랄	호
디올	틴트
립글로우	입생로랑
자연	9
은은	입생
보습	지속력
핑크	12
보습력	색상
글로우	레드
울트라핑크	7

브랜드별 가중치표(TfidfVectorizer) 브랜드별 가중치표(countVectorizer)

- 가중치표를 살펴보아도 거의 유사한 결과가 나왔다는 것을 다시 한번 확인 할 수 있다.

- 브랜드별로 가중치를 비교해보면, DIOR은 립밤과 유사한 해당 제품의 특성에 맞게 "자연", "은은", "보습력" 등의 단어가 가중치로 보였고 주로 핑크, 코랄 계열의 색상명이 가중치로 나왔다.
- YSL은 DIOR의 제품과는 다르게 "틴트" 제형임을 확인 할 수 있고 "지속력"이 강하며 "레드" 계열의 색상이 주력이고 숫자로 호수(색상명)을 이용함을 가중치표로 확인할 수 있다.

■ 서론

■ 데이터 수집

■ 전처리

■ 단어 빈도

■ 감성 분석

■ 주제 분석

■ 결론

- 브랜드를 구분하지 않고, LSA로 전체 주제 분석을 진행했다.

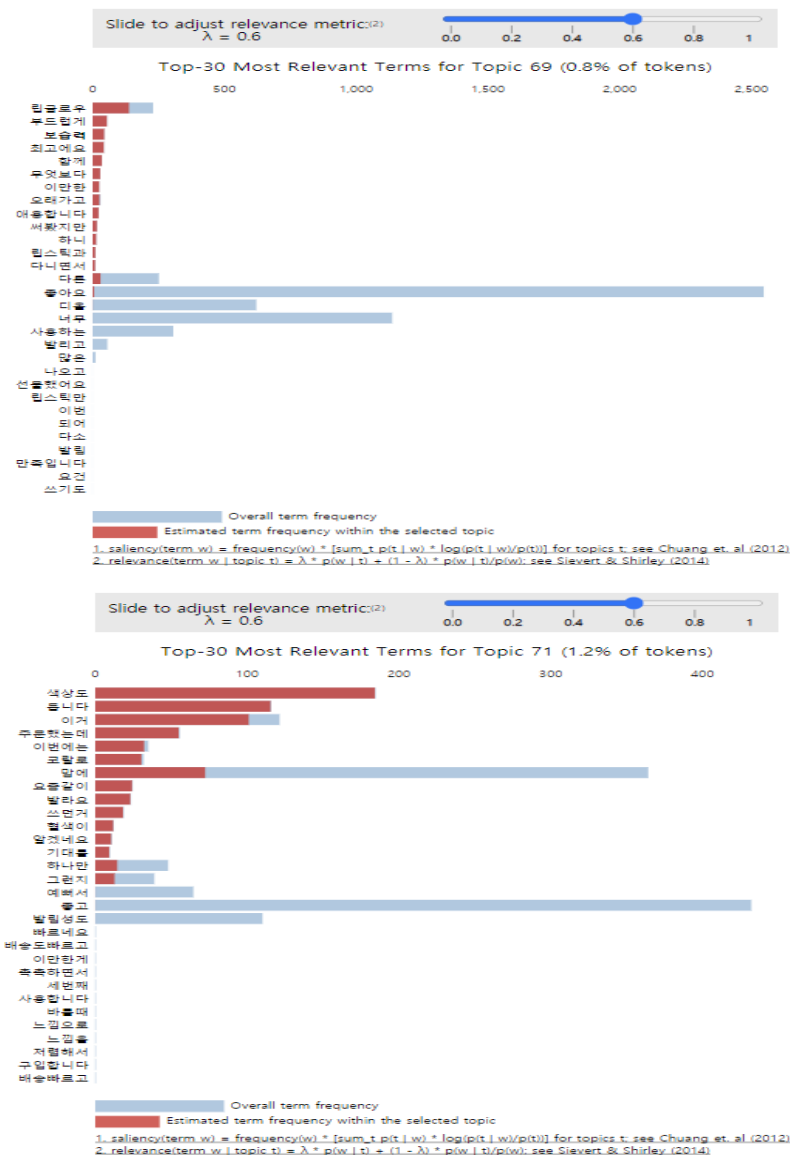
	word	loading
670	유명	0.162799
94	겨울	0.166751
270	라즈베리	0.167059
298	립글로스	0.181220
577	어딕트	0.295703
853	최고	0.319355
48	농 농	0.328428
415	보습력	0.346009
295	립글로우	2.446068
250	디올	7.477682

DIOR과 가장 관련이 높은 주제

	word	loading
679	은	0.243298
527	수	0.267821
713	인생	0.359481
853	최고	0.481804
822	진짜	0.488557
670	유명	0.648447
843	처음	0.822885
725	입생	1.153958
727	입생트랑	3.299883
908	틴트	8.713272

YSL과 가장 관련이 높은 주제

- 전체 주제 분석을 진행하니, 각 브랜드별 특성이 반영된 주제 분석 결과를 얻기 어려웠다.
- 따라서 각 브랜드별 주제 분석을 한번 더 진행하여 어떠한 주제들이 있고 그를 통해 알 수 있는 해당 브랜드 제품의 특징이 무엇인지 알아보았다.



- LDA를 활용하여, DIOR 리뷰에 대한 주제 분석을 진행했다.
- 여러 주제들 중 69번 주제를 보면, "립글로우", "보습력", "부드럽게" 등의 단어들이 묶여 있다. 이는 해당 제품의 제형과 관련된 특징이 나타나는 단어들이 한 주제에 묶인 것으로 해석할 수 있다.
- 이외에도 71번 주제를 보면, "색상도", "코랄로", "혈색이" 등의 단어들이 묶인 것으로 보아 제품 색상에 관련한 특징을 알 수 있는 주제라고 해석할 수 있다.

■ 서론

■ 데이터 수집

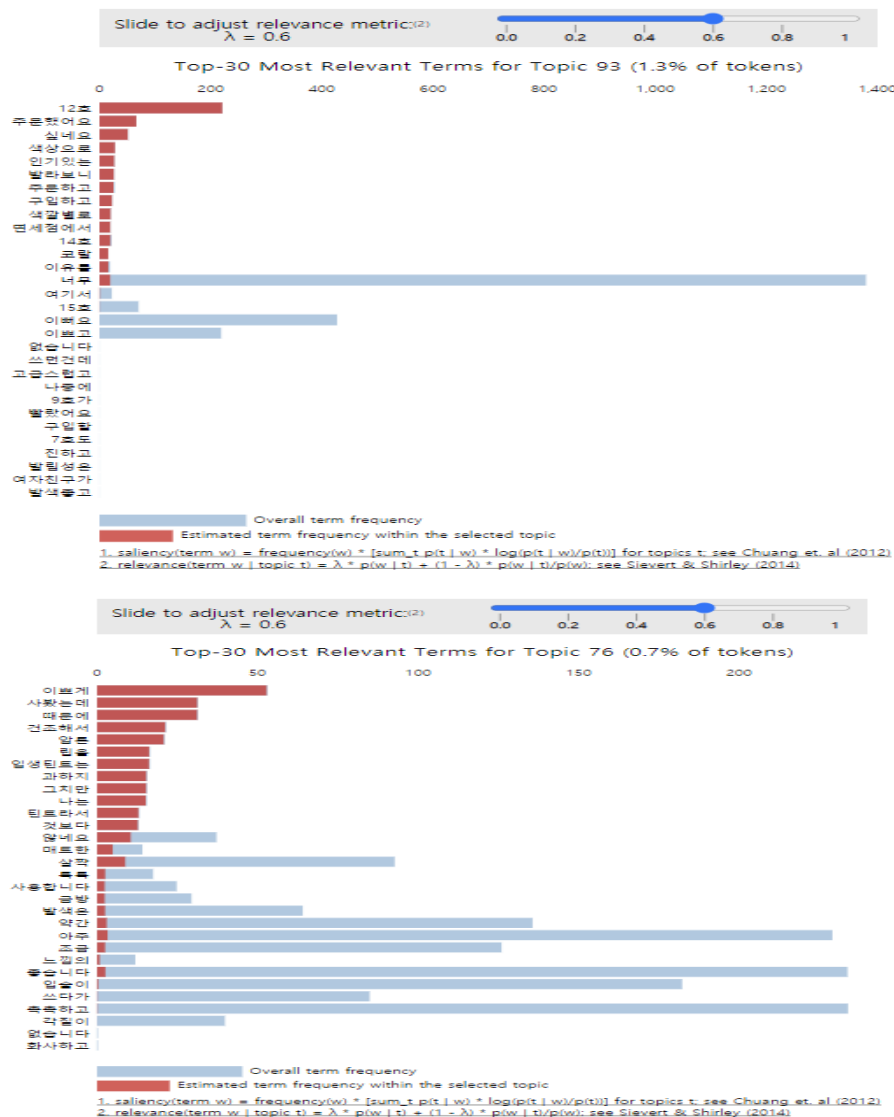
■ 전처리

■ 단어 빈도

■ 감성 분석

■ 주제 분석

■ 결론



- 동일한 방식으로 YSL 리뷰에 대한 주제 분석을 진행했다.
- 93번 주제를 보면, "12호", "색상으로", "색깔", "14호", "코랄" 등의 단어들이 함께 나오는 것을 보아 인기있는 색상에 관련된 단어들이 93번 주제에 있다고 해석할 수 있다.
- 이외에도 76번 주제를 보면, "건조해서", "매트한", "과하지" 등 사용감과 관련된 단어들이 연관되어 나오는 것을 볼 수 있다.

■ 서론

■ 데이터 수집

■ 전처리

■ 단어 빈도

■ 감성 분석

■ 주제 분석

■ 결론

- 단어 구름과 감성 분석을 진행한 후의 가중치표를 통해 DIOR 립 제품은 자연스럽게 은은한 립 글로우에 해당하고, YSL 립 제품은 지속력이 강한 틴트 제품에 해당된다는 서로 구분되는 특징을 확인할 수 있었다.
- LDA를 이용한 브랜드별 주제 분석을 통해 각 제품이 어떠한 특징을 가지고 있는지 더 세분화하여 알 수 있었다.
- 제품의 리뷰는 해당 제품의 특징을 충분히 포함하고 있음을 확인할 수 있었다.

■ 서론

■ 데이터 수집

■ 전처리

■ 단어 빈도

■ 감성 분석

■ 주제 분석

■ 결론

항목	점수	평가근거
서론	2/2	다루고자 하는 주제를 알맞게 설명했음.
데이터 수집	3/3	Selenium을 이용하여 웹 스크래핑을 할 수 있는 함수를 만들어 보다 간결하게 반복문을 활용한 웹 스크래핑을 수행함.
전처리	3/3	이모티콘, 특수문자, 불필요한 공백 등을 삭제하는 사용자 정의 함수를 만들어 전처리를 수행함.
단어 빈도	3/3	주제와 관련된 이미지와 해당 이미지 색상을 이용하여 단어 구름을 만들었음. 이때 ImageColorGenerator와 WordCloud의 mask 파라미터, matplotlib의 imshow를 활용하여 단어 구름을 만들.
감성 분석	2/3	수업 시간에 배운 내용을 활용하여 감성 분석을 수행하고 이를 알맞게 해석함.
주제 분석	2/3	수업 시간에 배운 내용을 활용하여 주제 분석을 수행하고 이를 알맞게 해석함.
결론	2/2	서론 및 본론과 부합하는 결론을 잘 제시함.
합계	17/19	



■ 서론

■ 데이터 수집

■ 전처리

■ 단어 빈도

■ 감성 분석

■ 주제 분석

■ 결론

■ END

감사합니다.